Arnold Kiv (Ed.)

M^{3}E^2

Machine Learning for Prediction of Emergent Economy Dynamics

Proceedings of the Selected Papers of the Special Edition of International Conference on Monitoring, Modeling & Management of Emergent Economy (M3E2-MLPEED 2020)

Odessa, Ukraine
July 13-18, 2020

Kiv, A. (Ed.): Machine Learning for Prediction of Emergent Economy Dynamics, Proceedings of the Selected Papers of the Special Edition of International Conference on Monitoring, Modeling & Management of Emergent Economy (M3E2-MLPEED 2020), Odessa, Ukraine, July 13-18, 2020, CEUR-WS.org, online (2020)

This volume represents the proceedings of the selected papers of the Special Edition of International Conference on Monitoring, Modeling & Management of Emergent Economy (M3E2-MLPEED 2020), held in Odessa, Ukraine, on July 13-18, 2020. It comprises 23 contributed papers that were carefully peer-reviewed and selected from 45 submissions. The accepted papers present the state-of-the-art overview of successful cases and provides guidelines for future research.

Program Committee

Albert Azaryan, Kryvyi Rih National University, Ukraine
Helena Fidlerová, Slovak University of Technology, Slovakia
Irina Georgescu, Bucharest University of Economics, Romania
Lidiya Guryanova, Simon Kuznets Kharkiv National University of Economics, Ukraine
Nadia Kabachi, Claude Bernard Lyon 1 University, France
Arnold Kiv, Ben-Gurion University of the Negev, Israel
Vitaliy Kobets, Kherson State University, Ukraine
Hennadiy Kravtsov, Kherson State University, Ukraine
Orken Mamyrbayev, Institute of Information and Computational Technologies, Kazakhstan
Lukas Pichl, International Christian University, Japan
Nina Rizun, Gdansk University of Technology, Poland
Abdel-Badeeh M. Salem, Ain Shams University, Egypt
Yaroslav Shramko, Kryvyi Rih State Pedagogical University, Ukraine
Oleksandr Spivakovskiy, Kherson State University, Ukraine
Aleksandr Uchitel, State University of Economics and Technology, Ukraine
Eugene Yakub, Odessa National Economic University, Ukraine

Local Organization Committee Members

Arnold Kiv, Ben-Gurion University of the Negev, Beer Sheva, Israel
Pavlo Hryhoruk, Khmelnytskyi National University, Khmelnytskyi, Ukraine
Inesa Khvostina, Ivano-Frankivsk National Technical University of Oil and Gas, Ivano-Frankivsk, Ukraine
Victoria Solovieva, State University of Economics and Technology, Kryvyi Rih, Ukraine
Vladimir Soloviev, Kryvyi Rih State Pedagogical University, Kryvyi Rih, Ukraine
Serhiy Semerikov, Kryvyi Rih State Pedagogical University, Kryvyi Rih, Ukraine

Additional Reviewers

Vitalina Babenko, V. N. Karazin Kharkiv National University, Ukraine
Hanna Danylchuk, The Bohdan Khmelnytsky National University of Cherkasy, Ukraine
Giedrė Dzemydaitė, Vilnius University, Lithuania
Liliana Horal, Ivano-Frankivsk National Technical University of Oil and Gas, Ukraine
Alexey Hostryk, Odessa National Economic University, Ukraine
Pavlo Hryhoruk, Khmelnytskyi National University, Ukraine
Liubov Kibalnyk, The Bohdan Khmelnytsky National University of Cherkasy, Ukraine

Machine learning of emerging markets in pandemic times

Arnold Kiv[1[0000-0002-0991-2343]], Pavlo Hryhoruk[2[0000-0002-2732-5038]],
Inesa Khvostina[3[0000-0001-5915-749X]], Victoria Solovieva[4[0000-0002-8090-9569]],
Vladimir Soloviev[5,6[0000-0002-4945-202X]], Serhiy Semerikov[5,7,8[0000-0003-0789-0272]]

[1] Ben-Gurion University of the Negev, P.O.B. 653, Beer Sheva, 8410501, Israel
kiv.arnold20@gmail.com
[2] Khmelnytskyi National University, 11 Instytutska Str., Khmelnytskyi, 29016, Ukraine
violete@ukr.net
[3] Ivano-Frankivsk National Technical University of Oil and Gas,
15 Karpatska Str., Ivano-Frankivsk, 76019, Ukraine
inesa.hvostina@gmail.com
[4] State University of Economics and Technology,
16 Medychna Str., Kryvyi Rih, 50005, Ukraine
vikasolovieva2027@gmail.com
[5] Kryvyi Rih State Pedagogical University, 54 Gagarin Ave., Kryvyi Rih, 50086, Ukraine
[6] The Bohdan Khmelnytsky National University of Cherkasy,
81 Shevchenka Blvd., 18031, Cherkasy, Ukraine
vnsoloviev2016@gmail.com
[7] Institute of Information Technologies and Learning Tools of the NAES of Ukraine,
9 M. Berlynskoho Str., Kyiv, 04060, Ukraine
[8] Kryvyi Rih National University, 11 Vitalii Matusevych Str., Kryvyi Rih, 50027, Ukraine
semerikov@gmail.com

Abstract. This is an introductory text to a collection of selected papers from the M3E2 2020 Summer: The Special Edition of International Conference on Monitoring, Modeling & Management of Emergent Economy, which was held in Odessa, Ukraine, on the July 13-18, 2020. It consists of short introduction and some observations about the event and its future.

Keywords: machine learning, prediction of emergent economy dynamics, COVID-19.

1 M3E2 2020 Summer at a glance

Monitoring, Modeling & Management of Emergent Economy (M3E2) is a peer-reviewed international conference focusing on research advances and applications of nonlinear dynamics methods, econophysics and complex systems methodology of emergent economy.

The M3E2 Conference occupies contributions in all aspects of Computational Finance, Economics, Risk Management, Statistical Finance, Trading and Market Microstructure, (Deep) Machine Learning technologies and tools, paradigms and models, relevant to modern financial engineering and technological decisions in the modern age. There is urgent general need for principled changes in postclassic economy elicited by current models, tools, services, networks and IT communication.

M3E2 topics of interest since 2019 [15]:

- Complex cyberphysical systems, synergy, econophysics, economy of agents.
- Mathematical methods, models, informational systems and technologies in economics.
- Monitoring, modeling, forecasting and preemption of crisis in socio-economic systems.
- Models of global transformations.
- Experimental economics.
- The dynamics of emergent markets in post crisis period.
- Management of the state's economic safety and economic safety of economic agents.
- Modeling of hospitality sphere development.
- Prioritized ways of formation of the innovation model of Ukrainian economical development.
- The Global Challenges for Economic Theory and Practice in CEE Countries.
- (Deep) Machine Learning for prediction of emergent economy dynamics.
- Risk Management models in emergent economy.

This volume contains the selected papers presented at M3E2 2020 Summer: The Special Edition of International Conference on Monitoring, Modeling & Management of Emergent Economy held on July 13-18, 2020 in Odessa, Ukraine (fig. 1, 2, 3).

Fig. 1. Joint photo of conference participants: the national quarantine is ended?

There were 45 submissions selected. Each submission was reviewed by at least 2, and on the average 2.3, program committee members. The committee decided to accept 23 papers.

Fig. 2. Dr. Hanna Danylchuk and Prof. Vladimir Soloviev: the social distancing during talk on modelling of cryptocurrency market using fractal and entropy analysis in COVID-19 [3].

2 Proceedings overview

The article “Recurrence plot-based analysis of financial-economic crashes” [22] of Vladimir Soloviev, Oleksandr Serdiuk, Serhiy Semerikov and Arnold Kiv (fig. 3) considers the possibility of analyzing the dynamics of changes in the characteristics of time series obtained on the basis of recurrent plots. The possibility of using the studied indicators to determine the presence of critical phenomena in economic systems is considered. Based on the analysis of economic time series of different nature, the suitability of the studied characteristics for the identification of critical phenomena is assessed. The description of recurrent diagrams and characteristics of time series that can be obtained on their basis is given. An analysis of seven characteristics of time series, including the coefficient of self-similarity, the coefficient of predictability, entropy, laminarity, is carried out. For the entropy characteristic, several options for its calculation are considered, each of which allows the one to get its own information about the state of the economic system. The possibility of using the studied characteristics as precursors of critical phenomena in economic systems is analyzed. Authors demonstrated that the entropy analysis of financial time series in phase space reveals the characteristic recurrent properties of complex systems. The recurrence entropy methodology has several advantages compared to the traditional recurrence entropy defined in the literature, namely, the correct evaluation of the chaoticity level of the signal, the weak dependence on parameters. The characteristics were studied on the basis of daily values of the Dow Jones index for the period from 1990 to 2019 and daily values of oil prices for the period from 1987 to 2019. The behavior of recurrent entropy during critical phenomena in the stock markets of the USA, Germany and France was studied separately. As a result of the study, it was determined that delay time measure, determinism and laminarity can be used as indicators of critical phenomena. It turned out that recurrent entropy, unlike other entropy indicators of complexity, is an indicator and an early harbinger of crisis phenomena. The ways of further research are outlined.

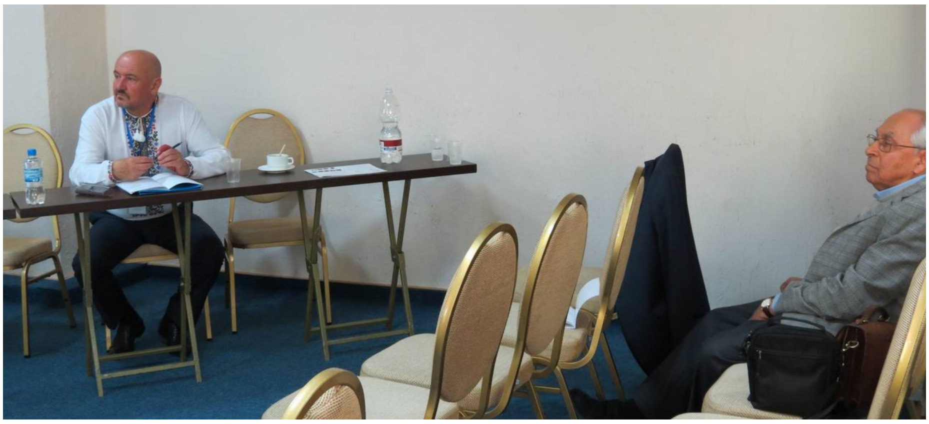


Fig. 3. Prof. Vladimir Soloviev and Prof. Arnold Kiv: the social distancing during talk on recurrence plot-based analysis of financial-economic crashes [22].

The article "Casual analysis of financial and operational risks of oil and gas companies in condition of emergent economy" [14] of Inesa Khvostina, Serhiy Semerikov, Oleh Yatsiuk, Nadiia Daliak, Olha Romanko and Ekaterina Shmeltser is devoted to control the risk that accompanies businesses in their day-to-day operations, and at the same time changing economic conditions make risk management an almost indispensable element of economic life. Selection of the main aspects of the selected phases of the risk management process: risk identification and risk assessment are related to their direct relationship with the subject matter (risk identification to be managed; risk analysis leading to the establishment of a risk hierarchy, and, consequently, the definition of risk control' methods) and its purpose (bringing the risk to acceptable level). It is impossible to identify the basic patterns of development of the oil and gas industry without exploring the relationship between economic processes and enterprise risks. The latter are subject to simulation, and based on models it is possible to determine with certain probability whether there have been qualitative and quantitative changes in the processes, in their mutual influence on each other, etc. The work is devoted to exploring the possibilities of applying the Granger test to examine the causal relationship between the risks and obligations of oil and gas companies. The analysis is based on statistical tests and the use of linear regression models.

The article "Complex networks theory and precursors of financial crashes" [23] of Vladimir Soloviev, Victoria Solovieva, Anna Tuliakova, Alexey Hostryk and Lukáš Pichl presents a systematic analysis of the dynamics of the largest stock markets in the world and cryptocurrency market. According to the algorithms of the visibility graph and recurrence plot, the daily values of stock and crypto indices are converted into a networks and multiplex networks, the spectral and topological properties of which are sensitive to the critical and crisis phenomena of the studied complex systems. This work is the first to investigate the network properties of the crypto index CCI30 and the multiplex network of key cryptocurrencies. It is shown that some of the spectral and topological characteristics can serve as measures of the complexity of the stock and crypto market, and their specific behaviour in the pre-crisis period is used as indicators-precursors of critical phenomena.

Prof. Lukáš Pichl was the Professor of Information Science at Department of Natural Sciences, International Christian University, Mitaka, Tokyo, Japan. Lukáš Pichl, born in 1974, received his PhD from the Graduate University of Advanced Studies in 2000. Since 2001 he taught computer science in the University of Aizu, before joining the International Christian University in 2005. His field of specialization is time series analysis and machine learning algorithms for the analysis of economic data including cryptocurrencies. Prof. Pichl has published a number of papers in international journals; he is a member of ACM and IEEE Computer Society, and currently serves as an associate editor of the Journal of Economic Interaction and Coordination (Springer). Prof. Lukáš Pichl passed away on April 10, 2020 at the age of 46.

Fig. 4. Prof. Vladimir Soloviev: in the memory of Prof. Lukáš Pichl.

The article “Modeling the optimal management of the distribution of profits of an oil and gas company taking into account risks” [13] of Inesa Khvostina, Viktor Oliinyk, Valerii Yatsenko, Liliia Mykhailyshyn and Uliana Berezhnytska discusses the optimal management of the distribution of the net income of an oil and gas company, taking into account risks. The utility function for the investigated enterprise acts as an optimality criterion. The control parameter is the distribution of the shares of net income for its optimal distribution in the selected areas. As a numerical implementation of the proposed algorithm, the activity of a catch oil and gas enterprise in the period 2018-2022 is considered. The optimal distribution of the received net income is given taking into account the discount rate and deductions to the State budget of Ukraine. The proposed algorithm can be used for optimal management of the company’s financial activities.

In the article “Scenario forecasting information transparency of subjects' under uncertainty and development of the knowledge economy” [17] of Hanna Kucherova, Anastasiia Didenko, Olena Kravets, Yuliia Honcharenko and Aleksandr Uchitel the topicality of modeling information transparency is determined by the influence it has on the effectiveness of management decisions made by an economic entity in the context of uncertainty and information asymmetry. It has been found that information transparency is a poorly structured category which acts as a qualitative characteristic of information and at certain levels forms an additional spectrum of properties of the information that has been adequately perceived or processed. As a result of structuring knowledge about the factor environment, a fuzzy cognitive model of information transparency was constructed in the form of a weighted digraph. Structural analysis and scenario forecasting of optimal alternatives of the fuzzy cognitive model made it possible to evaluate the classes of factors, identify their limited relations, establish the centrality of the roles of information transparency and information and communication security in the system built and evaluate their importance when modeling the situation self-development. Information visibility, reliability and availability have been found to have the strongest impact on the system. Taking into account different initial weights of the key factors – information transparency and information and communication security – the study substantiates the strategic ways for economic entities to achieve their goals in the context of uncertainty and information asymmetry, which allows us to use this approach as a tool for strategic management in the information environment.

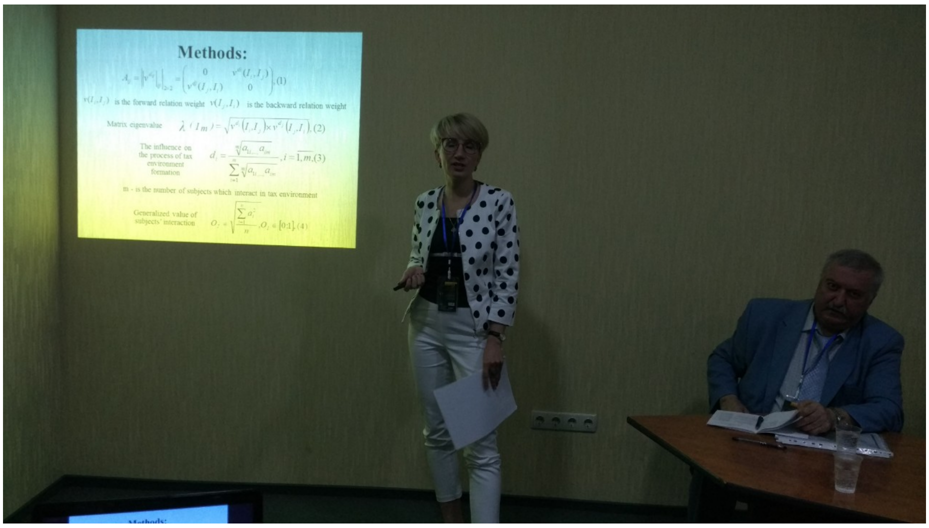


Fig. 5. Prof. Hanna Kucherova and Prof. Sultan Ramazanov: the social distancing during talk on scenario forecasting information transparency of subjects’ under uncertainty and development of the knowledge economy [17].

The article “Use of simulation modeling for predicting optimization of repair works at oil and gas production enterprises” [5] of Iryna Hobyr, Vitalina Babenko, Sofiia Kafka, Yuliia Bui, Oksana Savko and Ekaterina Shmeltser is substantiated the expediency of using the methods of the queuing theory for supporting various business processes at enterprises. The task of mass service for the organization and management of repair services of oil and gas producing enterprises has been set. It is stated that the objective function of solving this task is the sum of costs for maintaining repair crews and the reduction of losses from well downtime. It is proved that to solve the problem of optimizing the process of repair work at the researched enterprises it is necessary to use modeling as one of the effective tools of prediction, which do not require to bring the researched models to a specific form and allows to predict systems in different states and industries. A simulation model of the organization of repair works in wells is given. Testing of simulation models is carried out on the example of the fields of the three largest oil and gas producing enterprises in the Western region of Ukraine, it allowed to establish the laws of distribution of failures of oilfield equipment on fields, the process of which includes operations related to the creation, transformation and implementation of random events, quantities and processes that cause random changes in the state of the system. And also, to make calculations of the optimal between-repairs periods, total costs of well servicing processes, losses from their downtimes and optimum quantity of crews at various options of intensity of failures.

Cluster analysis of the efficiency of the recreational forest use of the region by separate components of the recreational forest use potential is provided in the article “Fuzzy cluster analysis of indicators for assessing the potential of recreational forest use” [16] of Evstakhii Kryzhanivs’kyi, Liliana Horal, Iryna Perevozova, Vira Shyiko, Nataliia Mykytiuk and Maria Berlous. The main stages of the cluster analysis of the recreational forest use level based on the predetermined components were determined. Among the agglomerative methods of cluster analysis, intended for grouping and combining the objects of study, it is common to distinguish the three most common types: the hierarchical method or the method of tree clustering; the K-means Clustering Method and the two-step aggregation method. For the correct selection of clusters, a comparative analysis of several methods was performed: arithmetic mean ranks, hierarchical methods followed by dendrogram construction, K-means method, which refers to reference methods, in which the number of groups is specified by the user. The cluster analysis of forestries by twenty analytical grounds was not proved by analysis of variance, so the re-clustering of certain objects was carried out according to the nine most significant analytical features. As a result, the forestry was clustered into four clusters. The conducted cluster analysis with the use of different methods allows us to state that their combination helps to select reasonable groupings, clearly illustrate the clustering procedure and rank the obtained forestry clusters.

The problem of determining the investment priorities of the national economy development has been actuated in the article “Strategic priorities of innovation and investment development of the Ukraine's economy industrial sector” [24] of Valentyna Stadnyk, Pavlo Izhevskiy, Nila Khrushch, Sergii Lysenko, Galyna Sokoliuk and Tetjana Tomalja. It has been argued that the formation of institutional preferences for activation of industry investment processes should be carried out taking into account the potential ability of each sectoral group enterprises to increase the added value. The scientific and methodical approach for sub-sectors investment attractiveness assessment has been formed on the example of the Ukrainian food industry. It has been recommended to use for this substantiated set of relative performance indexes which are duplicated in aggregate statistical state surveys based on the enterprise's financial statements. It has been formed the recommendations for the investment priorities of food industry development in Ukraine which are based on the appropriate calculations made by the TOPSIS and CRITIC methods. Methods of economic-statistical and comparative analysis were used for structural and dynamic characteristics of the Ukraine industrial enterprises activities. Given that innovation processes should also cover small and medium-sized industrial enterprises, whose resource opportunities are mostly limited, it is proposed to expand them within the framework of a strategic partnership. Graphic modeling methods have been used to visualize the process of building the business structures resource potential on the basis of their strategic partnership. The influence of the motivational environment on the value of organizational relations within the partnership has been formalized.

The article "Modelling of trade relations between EU countries by the method of minimum spanning trees using different measures of similarity" [2] of Hanna Danylchuk, Oksana Ivanylova, Liubov Kibalnyk, Oksana Kovtun, Tetiana Melnyk, Oleksandr Serdiuk and Vladimir Zaselskiy is devoted to the study of changes in relations between the countries of the European Union based on modeling and analysis of the structure of trade relations between countries. The article analyzes the dynamics of exports and imports of goods and services between the countries of the European Union on the basis of data taken for the period from 2006 to 2019. The study is based on one of the methods of cluster analysis, namely – the method of constructing minimal spanning trees. For the analysis the method of visualization of links is defined and the choice of the corresponding graphic representation is substantiated: the display of links using the dendrograms which carry more information in comparison with display of the minimum spanning trees in the form of a planar graph is chosen. Four different methods were used to construct the minimum spanning trees on the basis of which the visualization of links is performed: the Single link method, the averaged link method, the complete links method, and the Ward method. Based on the analysis of the results obtained using each of the methods, the best of them is selected, which is then used throughout the study. As a result of the study, suggests were made about the criteria by which clusters are formed within the European market. Such criteria are both the geographical neighborhood, which means mostly similar climatic conditions, and the common strategy of economic development of the country and the common strategy of behavior in the world market. In addition, a number of countries have been identified that are gradually moving to the use of their own economic strategies, as well as a number of countries seeking to align strategies of behavior in the world market. The influence of such factors as joining the integration union of new member states and global financial crises on the structure of trade relations is substantiated. Changes in the structure of relations between EU countries due to the influence of these factors are simulated. The study is of an applied nature and can be used in the future as a methodological basis for developing effective mechanisms for reformatting trade relations between countries in the context of geoeconomic transformations and global financial crises.

The article "Modeling the assessment of credit risk losses in banking" [18] of Katerina Larionova, Tetyana Donchenko, Andriy Oliinyk, Hennadii Kapinos, Oleg Savenko and Olexander Barmak develops a model of credit risk assessment within the scope of the variability concept that can be used for verification of new methods for borrowers' credit capacity estimation, the acceptable level of credit risk forecasting and its early prediction. It is aimed to be used during the automated banking systems development. The proposed model of credit risk assessment has been tested on the basis of the data from one of the Ukrainian banks. To determine the adequacy of this model has been proved by the comparison analysis of the proposed model with the results obtained by the National Bank of Ukraine methodology.

The article "The impact of COVID-induced shock on the risk-return correspondence of agricultural ETFs" [11] of Andrii Kaminskyi, Maryna Nehrey and Nina Rizun advocate the risk-return correspondence for different investment asset classes forms as one of the pillars of modern portfolio management. This correspondence together with interdependency analysis allows us to create portfolios that are adequate to given goals and constraints. COVID-induced shock unexpectedly generated high uncertainty and turmoil. The article is devoted to the investigation path through shock by agricultural assets (presented by ETFs) in comparison with traditional assets. There were identified three time periods: before the shock, explicitly shock, and post-shock. At the explicit shock period was suggested estimation risk frameworks on the pair indicators: falling depth and recovery ratio. Basic attention focuses on comparison risk-return estimations prior to shock and post-shock. To this end was considered four approaches to risk measurement and were applied to the sample of agricultural ETFs. The results indicated differences in risk changing by the path from before shock to post-shock. Differences arise from choosing the approach of risk measuring. The variability approach reveals much growth of risk of traditional assets, but the Value-at-Risk approach indicates higher risk growth for agricultural ETFs. Combine together with relatively low correlation these estimations provide a clear vision of risk-return frameworks.

The article "Comparative analysis of the attractiveness of investment instruments based on the analysis of market dynamics" [19] of Nataliia Maksyshko, Oksana Vasylieva, Igor Kozin and Vitalii Perepelitsa continues the authors' research on solving the problem of choosing the most attractive investment instrument from a variety of alternatives, based on a comparative analysis of the dynamics for the respective markets. The nature of the dynamics affects the predictability level of the investor's income and is determined by finding out which hypothesis corresponds to the dynamics: the efficient market hypothesis, the fractal market hypothesis and the coherent market hypothesis. The methodology of comparative analysis developed by the authors is based on the use of statistical analysis methods combined with the methods of complex fractal analysis. It makes it possible to reveal the presence of deterministic chaos in the dynamics and to obtain estimates of the long-term memory in time series. The calculated characteristics of the fuzzy set of the memory depth for time series make it possible to draw conclusions about the financial instruments preference for the investor. The methodology developed by the authors is applied to three markets. A comparative analysis of three instruments (gold, EUR/USD currency pair and Bitcoin cryptocurrency) was carried out. The dynamics of prices and profitability for financial instruments in the conditions before the onset of the COVID-19 crisis and during it is considered.

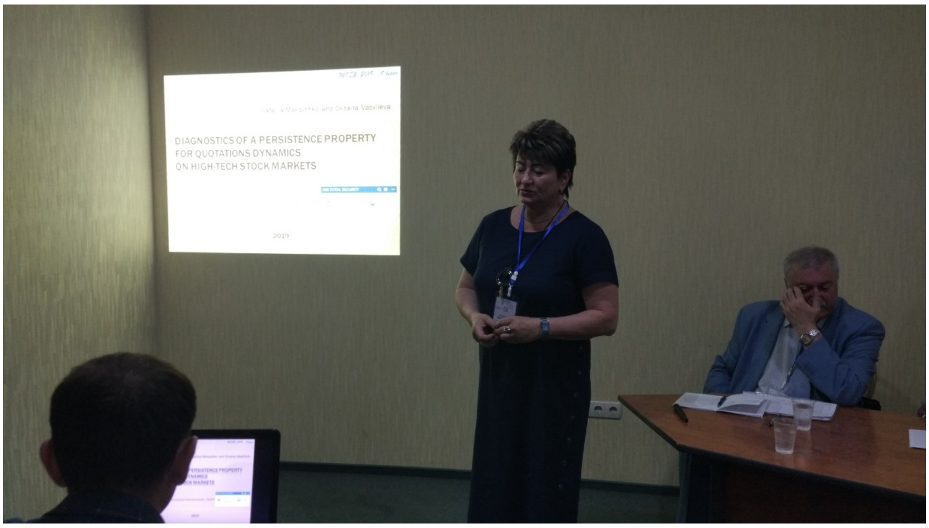


Fig. 6. Prof. Nataliia Maksyshko and Prof. Sultan Ramazanov: the social distancing during talk on comparative analysis of the attractiveness of investment instruments based on the analysis of market dynamics [19].

The article "Assessment of bank's financial security levels based on a comprehensive index using information technology" [12] of Nila Khrushch, Pavlo Hryhoruk, Tetiana Hovorushchenko, Sergii Lysenko, Liudmyla Prystupa and Liudmyla Vahanova considers the issues of assessing the level of financial security of the bank. An analysis of existing approaches to solving this problem. A scientific and methodological approach based on the application of comprehensive assessment technology is proposed. The computational algorithm is presented in the form of a four-stage procedure, which contains the identification of the initial data set, their normalization, calculation of the partial composite indexes, and a comprehensive index of financial security. Results have interpretation. Determining the levels of financial security and the limits of the relevant integrated indicator is based on the analysis of the configuration of objects in the two-scale space of partial composite indexes, which is based on the division of the set of initial indicators by content characteristics. The results of the grouping generally coincided with the results of the banks ranking according to the rating assessment of their stability, presented in official statistics. The article presents the practical implementation of the proposed computational procedure. To automate calculations and the possibility of scenario modeling, an electronic form of a spreadsheet was created with the help of form controls. The obtained results allowed us to identify the number of levels of financial security and their boundaries.

Fig. 7. Prof. Pavlo Hryhoruk before the talk on assessment of bank's financial security levels based on a comprehensive index using information technology [12]

The article "The cryptocurrencies risk measure based on the Laplace distribution" [9] of Petro Hrytsiuk and Tetiana Babych analyzed the daily returns of the most common cryptocurrencies: Bitcoin, Ethereum, XRP, USDT, Bitcoin Cash, Litecoin. It is shown that the asset returns are not normally distributed, but with good precision follow the Cauchy distribution and Laplace distribution. The analytical expressions for risk measure were obtained using the distribution function and the VaR technique. However, the risk assessment of the return obtained on the basis of the Cauchy distribution is twice as high as the risk assessment obtained on the basis of the Laplace distribution. Therefore, the question arises: what distribution law to use to measurement the cryptocurrency risk? The paper shows that the Laplace distribution is the most adequate basis for measuring of cryptocurrencies risk.

The article "Model for assessing and implementing resource-efficient strategy of industry" [21] of Nadiia Shmygol, Francesco Schiavone, Olena Trokhymets, Dariusz Pawliszczy, Viktor Koval, Ruslan Zavgorodniy and Andrii Vorfolomeiev determined that a number of scientists were involved in the development of a balanced system of indicators of the development of the oil and gas sector. Though an urgent scientific problem that needs further consideration is the development of a model of resource efficiency diagnostics in the oil and gas sector of the economy of Ukraine, taking into account the peculiarities of statistical monitoring. The scientific novelty of the paper is: this study improved the model of diagnostics of resource efficiency in oil and gas sector in the economy of Ukraine based on the additive-multiplicative compression of the formed system, which, unlike the existing ones, takes into account their variation while defining weighting coefficients which show the experts' system of preferences. It is reasonable to use the proposed model at the further economic assessment of the consequences of realization of resource-efficient strategy at enterprises of the oil and gas sector of the economy of Ukraine.

In the article “Fuzzy modelling of Big Data of HR in the conditions of Industry 4.0” [10] of Mykola Ivanov, Sergey Ivanov, Olexander Cherep, Nataliia Terentieva, Victoria Maltiz, Iulia Kaliuzhna and Vitaliy Lyalyuk a systematic methodology for analyzing and assessing the effectiveness of human resources based on fuzzy sets using big data technologies is used. Authors analyzed the big data construction method for our chosen approach using Industry 4.0. For the selected fuzzy sets, a set of sequence of procedures in the sequence of the method for assessing the effectiveness of human resources have been identified. Input and output membership functions for data mining have been developed. This article discusses process of building rules of fuzzy logic that allowed us to determine the degree of truth for each condition. The relevance achieved through the development of a methodology that includes eight procedures required for a comprehensive assessment of the economic efficiency of human resources. In this article, an approach to assessing the normative or average values of the performance of official duties by employees of an enterprise in many specialties, educational levels, levels of management, as well as taking into account the description of many positions, descriptions of compliance and interchangeability of positions, assessment of additional characteristics of employees and a description of many additional tasks and their characteristics is presented. The article presents a structural data-mining model for personnel assessment. The results of modeling the assessment of human resources is presented.

Fig. 8. Dr. Alexey Hostryk and Prof. Mykola Ivanov: the social distancing during talk on fuzzy modelling of Big Data of HR in the conditions of Industry 4.0 [10].

The article “Using non-metric multidimensional scaling for assessment of regions' economy in the context of their sustainable development” [8] of Pavlo Hryhoruk, Svitlana Grygoruk, Nila Khrushch and Tetiana Hovorushchenko is devoted to the solving the problems of regions’ socio-economic development as strategic and most important for any country. In particular, the implementation of a new, active role of the region as a subject of sustainable development is important for the direct implementation of current regional policy. An important component of such a policy is the assessment of sustainable development of regions, which contributes to the timely detection of internal and external threats, the development of necessary stabilizing measures to prevent their negative impact, the formation of strategies aimed at sustainable regional systems. The economic system is an important subsystem of the region. The article proposes an approach to assessing the regions’ economic development in the context of ensuring their sustainable development. Authors used the methods of multidimensional nonmetric scaling to solve this problem. The study aims to determine the structure of regions in the context of their sustainable development. Based on non-metric data reflecting the economic development of Ukraine’s regions, two-dimensional space of latent scales was built based on multidimensional measures of proximity between them, and the positioning of regions in this space was carried out. The results received a semantic interpretation, which was improved by using the procedure of rotation of the scale space. The use of multidimensional non-metric scaling confirms its usefulness for the study of economic development of regions in the region and allows for their comparison and dynamics of their structure in the context of sustainable development.

The article “Predicting the economic efficiency of the business model of an industrial enterprise using machine learning methods” [6] of Liliana Horal, Inesa Khvostina, Nadiia Reznik, Vira Shyiko, Natalia Yashcheritsyna, Svitlana Korol and Vladimir Zaselskiy considers the problem of studying the impact of key determinants on the industrial enterprise business model economic efficiency and aims to build an optimal model for predicting the industrial enterprise business model effectiveness using neural boundaries. A system of key determinants key factors has been developed. Significant factors were later used to build neural networks that characterize the studied resultant trait development vector. The procedure for constructing neural networks was performed in the STATISTICA Neural Networks environment. As input parameters, according to the previous analysis, 6 key factor indicators were selected. The initial parameter is determined by economic efficiency. According to the results of the neural network analysis, 100 neural networks were tested and the top 5 were saved. The following types of neural network architectures, multilayer perceptron, generalized regression network and linear network were used. Based on the results of the neural network modeling, 5 multilayer perceptrons of neural network architectures were proposed. According to descriptive statistics, the best model was a multilayer perceptron, with the MLP 6-10-1 architecture, which identifies a model with 6 input variables, one output variable and one hidden layer containing 10 hidden neurons. According to the analysis of the sensitivity of the network to input variables, it was determined that the network is the most sensitive to the variable the share of electricity costs in total costs. According to the results of selected neural networks standard prediction, the hypothesis of the best neural network was confirmed as Absolute res., Squared res, Std. Res for the neural network MLP 6-10-1 reached the optimal value and indicate that the selected model really has small residues, which indicates a fairly high accuracy of the forecast when using it.

The article "Modelling of cryptocurrency market using fractal and entropy analysis in COVID-19" [3] of Hanna Danylchuk, Liubov Kibalnyk, Oksana Kovtun, Arnold Kiv, Oleg Pursky and Galina Berezhna present the results of simulation for cryptocurrency market based on fractal and entropy analysis using six cryptocurrencies in the first 20 of the capitalization rating. The application of the selected research methods is based on an analysis of existing methodologies and tools of economic and mathematical modeling of financial markets. It has been shown that individual methods are not relevant because they do not provide an adequate assessment of the given market, so an integrated approach is the most appropriate. Daily values of cryptocurrency pairs from August 2016 to August 2020 selected by the monitoring and modelling database. The application of fractal analysis led to the conclusion that the time series of selected cryptocurrencies were persistent. And the use of the window procedure for calculating the local Hurst coefficient allowed to detail and isolate the persistant and antipersistant gaps. Interdisciplinary methods, namely Tsallis entropy and wavelet entropy, are proposed to complement the results. The results of the research show that Tsallis entropy reveals special (crisis) conditions in the cryptocurrency market, despite the nature of the crises' origin. Wavelet entropy is a warning indicator of crisis phenomena. It provides additional information on a small scale.

In the article "Econophysics of sustainability indices" [1] of Andriy Bielinskyi, Serhiy Semerikov, Oleksandr Serdiuk, Victoria Solovieva, Vladimir Soloviev and Lukas Pichl the possibility of using some econophysical methods for quantitative assessment of complexity measures: entropy (Shannon, Approximate and Permutation entropies), fractal (Multifractal detrended fluctuation analysis – MF-DFA), and quantum (Heisenberg uncertainty principle) is investigated. Comparing the capability of both entropies, it is obtained that both measures are presented to be computationally efficient, robust, and useful. Each of them detects patterns that are general for crisis states. The similar results are for other measures. MF-DFA approach gives evidence that Dow Jones Sustainability Index is multifractal, and the degree of it changes significantly at different periods. Moreover, authors demonstrate that the quantum apparatus of econophysics has reliable models for the identification of instability periods. Authors conclude that these measures make it possible to establish that the socially responsive exhibits characteristic patterns of complexity, and the proposed measures of complexity allow us to build indicators-precursors of critical and crisis phenomena.

The article "Asymptotic methods in optimization of multi-item inventory management model" [7] of Lidiia Horoshkova, Ievgen Khlobystov, Volodymyr Volkov, Olha Holovan, Svitlana Markova, Alexander Golub and Oleksandr Oliynyk describes the asymptotic methods for optimizing multi-item inventory model. To achieve the objective of the study, formulas of the optimal value of multi-item delivery frequency based on the asymptotic approach under conditions of minor changes in the input parameters have been obtained. The discrete increase in the execution costs and inventory holding costs which depend on the “small parameter” as well as a gradual increase in periodic fluctuations in demand for products have been taken as variable parameters of the system. Easy-to-use analytical formulas for determining optimal order interval when ordering and inventory holding costs, as well as demand meet insufficient changes have been obtained. Testing of the proposed approach to the multi-item inventory model has been carried out on the example of HoReCa regional market segment. The proposed formulas allow to apply the obtained results for optimization and forecasting of decision-making in the system of procurement logistics of a company amid variation of input parameters describing changes of external and internal business environment.

The article “Modeling and prediction of the gas pipelines reliability indicators in the context of energy security of Ukraine” [25] of Halyna Zelinska, Irina Fedorovych, Uliana Andrusiv, Oksana Chernova and Halyna Kupalova based on the many years of experience in operation of the gas transportation system shows that the largest accidents with severe consequences arise due to untimely detection and elimination of gas leaks in underground gas pipelines. The decrease in the reliability of the gas transportation system functioning can be considered from the following two perspectives: the first perspective is the economic one – it leads to an increase in the economic expenses of an enterprise; and the second perspective is the social and environmental one – it results in emergence of a threat to public health, as well as loss of human and natural resources. Hence, the issue of modeling and prediction of the reliability indicators of natural gas transportation via gas pipelines becomes especially urgent because of the requirements for reliable operation of the system. It has been proven that the main problem leading to a decrease in the reliability of the gas transportation process is the significant deterioration of fixed assets, which requires investment of considerable financial resources in the gas transportation system of Ukraine (GTS). The article substantiates that it is possible to increase the reliability of operation of the line section of the main gas pipelines (LSMGP) through a high-quality system of repairs and equipment modernization. The main factor allowing to reduce the number of accidents is considered to be timely detection of damages on gas pipelines and their prediction. It has been determined that the failure rate depends on the diameter and number of lines of a gas pipeline. The authors propose to conduct a comprehensive diagnosis of the process of reliability of gas pipelines together with of their technical and economic indicators, based on the development of a system of measures to improve the safety of gas pipelines in Ukraine. A system of measures has been developed to improve the reliability of gas pipelines operation in Ukraine.

The article “Machine learning approaches for financial time series forecasting” [4] of Vasily Derbentsev, Andriy Matviychuk, Nataliia Datsenko, Vitalii Bezkorovainyi and Albert Azaryan is discusses the problems of the short-term forecasting of financial time series using supervised machine learning (ML) approach. For this goal, authors applied several the most powerful methods including Support Vector Machine (SVM), Multilayer Perceptron (MLP), Random Forests (RF) and Stochastic Gradient Boosting Machine (SGBM). As dataset were selected the daily close prices of two stock index: SP 500 and NASDAQ, two the most capitalized cryptocurrencies: Bitcoin (BTC), Ethereum (ETH), and exchange rate of EUR-USD. As features were used only the past price information. To check the efficiency of these models was made out-of-sample forecast for selected time series by using one step ahead technique. The accuracy rates of the forecasted prices by using ML models were calculated. The results verify the applicability of the ML approach for the forecasting of financial time series. The best out of sample accuracy of short-term prediction daily close prices for selected time series obtained by SGBM and MLP in terms of Mean Absolute Percentage Error (MAPE) was within 0.46-3.71 %. The results are comparable with accuracy obtained by Deep learning approaches.

Fig. 9. The coronavirus sadness [20] of Prof. Andriy Matviychuk before the talk on machine learning approaches for financial time series forecasting [4]

3 Conclusion

The vision of the M3E2 2019 is provides a premier interdisciplinary platform for researchers, practitioners and educators to present and discuss the most recent innovations, trends, and concerns as well as practical challenges encountered and solutions adopted in the fields of emergent economy.

The conference has successfully performing forum to transferring and discussing research result among the researcher, students, government, private sector or industries. Participants and presenters from several countries such as Italy, Israel, Japan, Lithuania, Poland, Ukraine have attended the conference to share their significant contribution in research related to Monitoring, Modeling & Management of Emergent Economy.

The special instalment of M3E2 at the Hot Coronavirus Summer'2020 was organized by:

- Ben-Gurion University of the Negev, Israel (with support of the rector Prof. Chaim Hames)
- Institute of Information Technologies and Learning Tools of the NAES of Ukraine (with support of the director Prof. Valeriy Bykov)
- Ivano-Frankivsk National Technical University of Oil and Gas, Ukraine (with support of the rector Prof. Evstakhii Kryzhanivs'kyi)
- Khmelnytskyi National University, Ukraine (with support of the rector Prof. Mykola Skyba)
- Kryvyi Rih National University, Ukraine (with support of the rector Prof. Mykola Stupnik)
- Kryvyi Rih State Pedagogical University, Ukraine (with support of the rector Prof. Yaroslav Shramko)
- State University of Economics and Technology, Ukraine (with support of the rector Prof. Andrii Shaikan)
- The Bohdan Khmelnytsky National University of Cherkasy, Ukraine (with support of the rector Prof. Olexandr Cherevko)

We are thankful to all the authors who submitted papers and the delegates for their participation and their interest in M3E2 as a platform to share their ideas and innovation. Also, we are also thankful to all the program committee members for providing continuous guidance and efforts taken by peer reviewers contributed to improve the quality of papers provided constructive critical comments, improvements and corrections to the authors are gratefully appreciated for their contribution to the success of the conference.

We hope you enjoy this conference and meet again in more friendly, hilarious, and happiness of further M3E2 2021.

References

1. Bielinskyi, A., Semerikov, S., Serdiuk, O., Solovieva, V., Soloviev, V., Pichl, L.: Econophysics of sustainability indices. In: Kiv, A. (ed.) Machine Learning for Prediction of Emergent Economy Dynamics, Proceedings of the Selected Papers of the Special Edition of International Conference on Monitoring, Modeling & Management of Emergent Economy (M3E2-MLPEED 2020), Odessa, Ukraine, July 13-18, 2020, CEUR-WS.org, online (2020, in press)
2. Danylchuk, H., Ivanylova, O., Kibalnyk, L., Kovtun, O., Melnyk, T., Serdiuk, O., Zaselskiy, V.: Modelling of trade relations between EU countries by the method of minimum spanning trees using different measures of similarity. In: Kiv, A. (ed.) Machine Learning for Prediction of Emergent Economy Dynamics, Proceedings of the Selected Papers of the Special Edition of International Conference on Monitoring, Modeling & Management of Emergent Economy (M3E2-MLPEED 2020), Odessa, Ukraine, July 13-18, 2020, CEUR-WS.org, online (2020, in press)
3. Danylchuk, H., Kibalnyk, L., Kovtun, O., Kiv, A., Pursky, O., Berezhna, G.: Modelling of cryptocurrency market using fractal and entropy analysis in COVID-19. In: Kiv, A. (ed.) Machine Learning for Prediction of Emergent Economy Dynamics, Proceedings of the Selected Papers of the Special Edition of International Conference on Monitoring,

Modeling & Management of Emergent Economy (M3E2-MLPEED 2020), Odessa, Ukraine, July 13-18, 2020, CEUR-WS.org, online (2020, in press)

4. Derbentsev, V., Matviychuk, A., Datsenko, N., Bezkorovainyi, V., Azaryan, A.: Machine learning approaches for financial time series forecasting. In: Kiv, A. (ed.) Machine Learning for Prediction of Emergent Economy Dynamics, Proceedings of the Selected Papers of the Special Edition of International Conference on Monitoring, Modeling & Management of Emergent Economy (M3E2-MLPEED 2020), Odessa, Ukraine, July 13-18, 2020, CEUR-WS.org, online (2020, in press)
5. Hobyr, I., Babenko, V., Kafka, S., Bui, Y., Savko, O., Shmeltser, E.: Use of simulation modeling for predicting optimization of repair works at oil and gas production enterprises. In: Kiv, A. (ed.) Machine Learning for Prediction of Emergent Economy Dynamics, Proceedings of the Selected Papers of the Special Edition of International Conference on Monitoring, Modeling & Management of Emergent Economy (M3E2-MLPEED 2020), Odessa, Ukraine, July 13-18, 2020, CEUR-WS.org, online (2020, in press)
6. Horal, L., Khvostina, I., Reznik, N., Shyiko, V., Yashcheritsyna, N., Korol, S., Zaselskiy, V.: Predicting the economic efficiency of the business model of an industrial enterprise using machine learning methods. In: Kiv, A. (ed.) Machine Learning for Prediction of Emergent Economy Dynamics, Proceedings of the Selected Papers of the Special Edition of International Conference on Monitoring, Modeling & Management of Emergent Economy (M3E2-MLPEED 2020), Odessa, Ukraine, July 13-18, 2020, CEUR-WS.org, online (2020, in press)
7. Horoshkova, L., Khlobystov, I., Volkov, V., Holovan, O., Markova, S., Golub, A., Oliynyk, O.: Asymptotic methods in optimization of multi-item inventory management model. In: Kiv, A. (ed.) Machine Learning for Prediction of Emergent Economy Dynamics, Proceedings of the Selected Papers of the Special Edition of International Conference on Monitoring, Modeling & Management of Emergent Economy (M3E2-MLPEED 2020), Odessa, Ukraine, July 13-18, 2020, CEUR-WS.org, online (2020, in press)
8. Hryhoruk, P., Grygoruk, S., Khrushch, N., Hovorushchenko, T.: Using non-metric multidimensional scaling for assessment of regions' economy in the context of their sustainable development. In: Kiv, A. (ed.) Machine Learning for Prediction of Emergent Economy Dynamics, Proceedings of the Selected Papers of the Special Edition of International Conference on Monitoring, Modeling & Management of Emergent Economy (M3E2-MLPEED 2020), Odessa, Ukraine, July 13-18, 2020, CEUR-WS.org, online (2020, in press)
9. Hrytsiuk, P., Babych, T.: The cryptocurrencies risk measure based on the Laplace distribution. In: Kiv, A. (ed.) Machine Learning for Prediction of Emergent Economy Dynamics, Proceedings of the Selected Papers of the Special Edition of International Conference on Monitoring, Modeling & Management of Emergent Economy (M3E2-MLPEED 2020), Odessa, Ukraine, July 13-18, 2020, CEUR-WS.org, online (2020, in press)
10. Ivanov, M., Ivanov, S. Cherep, O., Terentieva, N., Maltiz, V., Kaliuzhna, I., Lyalyuk, V.: Fuzzy modelling of Big Data of HR in the conditions of Industry 4.0. In: Kiv, A. (ed.) Machine Learning for Prediction of Emergent Economy Dynamics, Proceedings of the Selected Papers of the Special Edition of International Conference on Monitoring, Modeling & Management of Emergent Economy (M3E2-MLPEED 2020), Odessa, Ukraine, July 13-18, 2020, CEUR-WS.org, online (2020, in press)
11. Kaminskyi, A., Nehrey, M., Rizun, N.: The impact of COVID-induced shock on the risk-return correspondence of agricultural ETFs. In: Kiv, A. (ed.) Machine Learning for Prediction of Emergent Economy Dynamics, Proceedings of the Selected Papers of the

Special Edition of International Conference on Monitoring, Modeling & Management of Emergent Economy (M3E2-MLPEED 2020), Odessa, Ukraine, July 13-18, 2020, CEUR-WS.org, online (2020, in press)

12. Khrushch, N., Hryhoruk, P., Hovorushchenko, T., Lysenko, S., Prystupa, L., Vahanova, L.: Assessment of bank's financial security levels based on a comprehensive index using information technology. In: Kiv, A. (ed.) Machine Learning for Prediction of Emergent Economy Dynamics, Proceedings of the Selected Papers of the Special Edition of International Conference on Monitoring, Modeling & Management of Emergent Economy (M3E2-MLPEED 2020), Odessa, Ukraine, July 13-18, 2020, CEUR-WS.org, online (2020, in press)
13. Khvostina, I., Oliinyk, V., Yatsenko, V., Mykhailyshyn, L., Berezhnytska, U.: Modeling the optimal management of the distribution of profits of an oil and gas company taking into account risks. In: Kiv, A. (ed.) Machine Learning for Prediction of Emergent Economy Dynamics, Proceedings of the Selected Papers of the Special Edition of International Conference on Monitoring, Modeling & Management of Emergent Economy (M3E2-MLPEED 2020), Odessa, Ukraine, July 13-18, 2020, CEUR-WS.org, online (2020, in press)
14. Khvostina, I., Semerikov, S., Yatsiuk, O., Daliak, N., Romanko, O., Shmeltser, E.: Casual analysis of financial and operational risks of oil and gas companies in condition of emergent economy. In: Kiv, A. (ed.) Machine Learning for Prediction of Emergent Economy Dynamics, Proceedings of the Selected Papers of the Special Edition of International Conference on Monitoring, Modeling & Management of Emergent Economy (M3E2-MLPEED 2020), Odessa, Ukraine, July 13-18, 2020, CEUR-WS.org, online (2020, in press)
15. Kiv, A., Semerikov, S., Soloviev, V., Kibalnyk, L., Danylchuk, H., Matviychuk, A.: Experimental Economics and Machine Learning for Prediction of Emergent Economy Dynamics. In: Kiv, A., Semerikov, S., Soloviev, V., Kibalnyk, L., Danylchuk, H., Matviychuk, A. (eds.) Experimental Economics and Machine Learning for Prediction of Emergent Economy Dynamics, Proceedings of the Selected Papers of the 8th International Conference on Monitoring, Modeling & Management of Emergent Economy (M3E2 2019), Odessa, Ukraine, May 22-24, 2019. CEUR Workshop Proceedings **2422**, 1–4. http://ceur-ws.org/Vol-2422/paper00.pdf (2019). Accessed 17 Aug 2020
16. Kryzhanivs'kyi, E., Horal, L., Perevozova, I., Shyiko, V., Mykytiuk, N., Berlous, M.: Fuzzy cluster analysis of indicators for assessing the potential of recreational forest use. In: Kiv, A. (ed.) Machine Learning for Prediction of Emergent Economy Dynamics, Proceedings of the Selected Papers of the Special Edition of International Conference on Monitoring, Modeling & Management of Emergent Economy (M3E2-MLPEED 2020), Odessa, Ukraine, July 13-18, 2020, CEUR-WS.org, online (2020, in press)
17. Kucherova, H., Didenko, A., Kravets, O., Honcharenko, Y., Uchitel, A.: Scenario forecasting information transparency of subjects' under uncertainty and development of the knowledge economy. In: Kiv, A. (ed.) Machine Learning for Prediction of Emergent Economy Dynamics, Proceedings of the Selected Papers of the Special Edition of International Conference on Monitoring, Modeling & Management of Emergent Economy (M3E2-MLPEED 2020), Odessa, Ukraine, July 13-18, 2020, CEUR-WS.org, online (2020, in press)
18. Larionova, K., Donchenko, T., Oliinyk, A., Kapinos, H., Savenko, O., Barmak, O.: Modeling the assessment of credit risk losses in banking. In: Kiv, A. (ed.) Machine Learning for Prediction of Emergent Economy Dynamics, Proceedings of the Selected Papers of the Special Edition of International Conference on Monitoring, Modeling & Management of

Emergent Economy (M3E2-MLPEED 2020), Odessa, Ukraine, July 13-18, 2020, CEUR-WS.org, online (2020, in press)

19. Maksyshko, N., Vasylieva, O., Kozin, I., Perepelitsa, V.: Comparative analysis of the attractiveness of investment instruments based on the analysis of market dynamics. In: Kiv, A. (ed.) Machine Learning for Prediction of Emergent Economy Dynamics, Proceedings of the Selected Papers of the Special Edition of International Conference on Monitoring, Modeling & Management of Emergent Economy (M3E2-MLPEED 2020), Odessa, Ukraine, July 13-18, 2020, CEUR-WS.org, online (2020, in press)
20. Semerikov, S., Chukharev, S., Sakhno, S., Striuk, A., Osadchyi, V., Solovieva, V., Vakaliuk, T., Nechypurenko, P., Bondarenko, O., Danylchuk, H.: Our sustainable coronavirus future. E3S Web of Conferences **166**, 00001 (2020). doi:10.1051/e3sconf/202016600001
21. Shmygol, N., Schiavone, F., Trokhymets, O., Pawliszczy, D., Koval, V., Zavgorodniy, R., Vorfolomeiev, A.: Model for assessing and implementing resource-efficient strategy of industry. In: Kiv, A. (ed.) Machine Learning for Prediction of Emergent Economy Dynamics, Proceedings of the Selected Papers of the Special Edition of International Conference on Monitoring, Modeling & Management of Emergent Economy (M3E2-MLPEED 2020), Odessa, Ukraine, July 13-18, 2020, CEUR-WS.org, online (2020, in press)
22. Soloviev, V., Serdiuk, O., Semerikov, S., Kiv, A.: Recurrence plot-based analysis of financial-economic crashes. In: Kiv, A. (ed.) Machine Learning for Prediction of Emergent Economy Dynamics, Proceedings of the Selected Papers of the Special Edition of International Conference on Monitoring, Modeling & Management of Emergent Economy (M3E2-MLPEED 2020), Odessa, Ukraine, July 13-18, 2020, CEUR-WS.org, online (2020, in press)
23. Soloviev, V., Solovieva, V., Tuliakova, A., Hostryk, A., Pichl, L.: Complex networks theory and precursors of financial crashes. In: Kiv, A. (ed.) Machine Learning for Prediction of Emergent Economy Dynamics, Proceedings of the Selected Papers of the Special Edition of International Conference on Monitoring, Modeling & Management of Emergent Economy (M3E2-MLPEED 2020), Odessa, Ukraine, July 13-18, 2020, CEUR-WS.org, online (2020, in press)
24. Stadnyk, V., Izhevskiy, P., Khrushch, N., Lysenko, S., Sokoliuk, G., Tomalja, T.: Strategic priorities of innovation and investment development of the Ukraine's economy industrial sector. In: Kiv, A. (ed.) Machine Learning for Prediction of Emergent Economy Dynamics, Proceedings of the Selected Papers of the Special Edition of International Conference on Monitoring, Modeling & Management of Emergent Economy (M3E2-MLPEED 2020), Odessa, Ukraine, July 13-18, 2020, CEUR-WS.org, online (2020, in press)
25. Zelinska, H., Fedorovych, I., Andrusiv, U., Chernova, O., Kupalova, H.: Modeling and prediction of the gas pipelines reliability indicators in the context of energy security of Ukraine. In: Kiv, A. (ed.) Machine Learning for Prediction of Emergent Economy Dynamics, Proceedings of the Selected Papers of the Special Edition of International Conference on Monitoring, Modeling & Management of Emergent Economy (M3E2-MLPEED 2020), Odessa, Ukraine, July 13-18, 2020, CEUR-WS.org, online (2020, in press)

Recurrence plot-based analysis of financial-economic crashes

Vladimir Soloviev[1,2[0000-0002-4945-202X]], Oleksandr Serdiuk[2[0000-0002-1230-0305]], Serhiy Semerikov[1,3,4[0000-0003-0789-0272]] and Arnold Kiv[5[0000-0002-0991-2343]]

[1] Kryvyi Rih State Pedagogical University, 54 Gagarin Ave., Kryvyi Rih, 50086, Ukraine
vnsoloviev2016@gmail.com, semerikov@gmail.com

[2] The Bohdan Khmelnytsky National University of Cherkasy, 81 Shevchenka Blvd., 18031, Cherkasy, Ukraine
serdyuk@ukr.net

[3] Kryvyi Rih National University, 11 Vitalii Matusevych Str., Kryvyi Rih, 50027, Ukraine

[4] Institute of Information Technologies and Learning Tools of NAES of Ukraine, 9 M. Berlynskoho Str., Kyiv, 04060, Ukraine

[5] Ben-Gurion University of the Negev, P.O.B. 653, Beer Sheva, 8410501, Israel
kiv.arnold20@gmail.com

Abstract. The article considers the possibility of analyzing the dynamics of changes in the characteristics of time series obtained on the basis of recurrence plots. The possibility of using the studied indicators to determine the presence of critical phenomena in economic systems is considered. Based on the analysis of economic time series of different nature, the suitability of the studied characteristics for the identification of critical phenomena is assessed. The description of recurrence diagrams and characteristics of time series that can be obtained on their basis is given. An analysis of seven characteristics of time series, including the coefficient of self-similarity, the coefficient of predictability, entropy, laminarity, is carried out. For the entropy characteristic, several options for its calculation are considered, each of which allows the one to get its own information about the state of the economic system. The possibility of using the studied characteristics as precursors of critical phenomena in economic systems is analyzed. We have demonstrated that the entropy analysis of financial time series in phase space reveals the characteristic recurrent properties of complex systems. The recurrence entropy methodology has several advantages compared to the traditional recurrence entropy defined in the literature, namely, the correct evaluation of the chaoticity level of the signal, the weak dependence on parameters. The characteristics were studied on the basis of daily values of the Dow Jones index for the period from 1990 to 2019 and daily values of oil prices for the period from 1987 to 2019. The behavior of recurrence entropy during critical phenomena in the stock markets of the USA, Germany and France was studied separately. As a result of the study, it was determined that delay time measure, determinism and laminarity can be used as indicators of critical phenomena. It turned out that recurrence entropy, unlike other entropy indicators of complexity, is an indicator and an early precursor of crisis phenomena. The ways of further research are outlined.

Keywords: complex systems, recurrence entropy, indicator-predictor of crashes.

1 Introduction

During last few decades the behavior of global financial system attracted considerable attention. Strong sharp fluctuations in stock prices lead to sudden trend switches in a number of stocks and continue to have a huge impact on the world economy causing the instability in it with regard to normal and natural disturbances [18]. The reason of this problem is the crisis of methodology modeling, forecasting and interpretation of socio-economic realities. The doctrine of the unity of the scientific method states that for the study of events in socio-economic systems, the same methods and criteria as those used in the study of natural phenomena are applicable. Similar idea has attracted considerable attention by the community from different branches of science in recent years [6].

The increasing mathematical knowledge of the complex structures of nonlinear systems has provided successful tools to the understanding of irregular space and temporal behaviors displayed by collected data in all applied sciences. Time series analysis has turned to be a key issue providing the most direct link between nonlinear dynamics and the real world [9]. Among the many methods of analysis of complex nonlinear, non-stationary emergent signals, which are the signals of complex systems, those that adequately reflect the spatial and temporal manifestations of complexity are especially popular [17]. In this case, the search for such quantitative measures of complexity that adequately reflect the dynamics of processes taking place in a complex system is relevant. Financial systems being complex dynamic objects exhibit unexpected critical phenomena, which are most clearly manifested in the form of crashes. Over the past 20 years, these are the global currency crisis of 1998, the collapse of the dotcoms 2001, the global financial crisis of 2008, the European debt crisis of 2012, the Chinese crisis of 2015-2016 and the crisis of the US stock market in early 2019 [16]. For this reason, it is extremely important to highlight such measures of complexity that are sensitive to critical phenomena and can serve as their predictors [2; 21].

In this paper, we will consider the possibilities of new entropy indicators of the systems complexity, calculated in the phase space, and examine their capabilities with respect to the prevention of crisis phenomena.

2 Research methods

2.1 The recurrence plots

In recent years, new quantifiers of nonlinear time series analysis have appeared based on properties of phasespace recurrences [13]. According to stochastic extensions to Taken's embedding theorems the embedding of a time series in phase space can be carried out by forming time-delayed vectors $\vec{X}_n = [x_n, x_{n+\tau}, x_{n+2\tau}, \ldots, x_{n+(m-1)\tau}]$ for each value x_n in the time series, where m is the embedding dimension, and τ is the embedding delay. These parameters are obtained by systematic search for the optimal set. Figure 1 shows a phase portrait of the normalized logarithmic returns of the time series of Bitcoin (BTC) prices for the period July 17, 2010 to August 30, 2019.

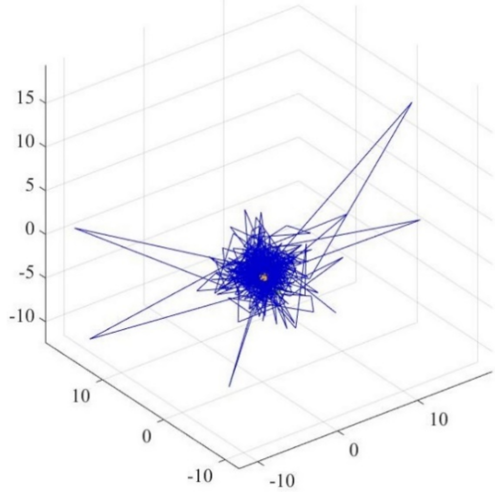


Fig. 1. A phase portrait of the normalized logarithmic returns of the daily values BTC/$ for the period July 17, 2010 to August 30, 2019.

A modern visualization method known as recurrence plot (RP), and is constructed from the recurrence matrix $\vec{R}_{ij}$ defined as $\vec{R}_{ij}(\varepsilon) = \Theta(\varepsilon - \|x_i - x_j\|)$, $x_i \in R$, $i, j = 1,2, \ldots, M$, where x_i and x_j represent the dynamical state at time i and j, Θ is the Heaviside function, M is length of the analyzed time series and ε is the threshold or vicinity parameter, consisting of a maximum distance between two points in a trajectory such that both points can be considered recurrent to each other.

The recurrence plot for the phase portrait of figure 1 is presented in figure 2.

The graphical representation of the RP allows to derive qualitative characterizations of the dynamical systems. For the quantitative description of the dynamics, the small-scale patterns in the RP can be used, such as diagonal and vertical lines. The histograms of the lengths of these lines are the base of the recurrence quantification analysis (RQA) developed by Webber and Zbilut [20] and later by Marwan et al. [13]. Based on the statistical properties of the recurrence plot, a large number of quantifiers have been developed to analyze details of a RP. Many of them deal with statistical properties such as mean size, maximum size, frequency of occurrence of diagonal, vertical or horizontal recurrence lines.

The appearance of the recurrence diagram allows us to judge the nature of the processes occurring in the system, the presence and influence of noise, states of repetition and fading (laminarity), the implementation during the evolution of the system of abrupt changes (extreme events). Thus, a visual assessment of the diagrams can give an idea of the evolution of the studied trajectory. There are two main classes of image structure: typology, represented by large-scale structures, and texture

(texture), formed by small-scale structures. Topology gives a general idea of the nature of the process. A detailed examination of recurrence diagrams allows you to identify small-scale structures – a texture that consists of simple points, diagonal, horizontal and vertical lines. Combinations of vertical and horizontal lines form rectangular clusters of points.

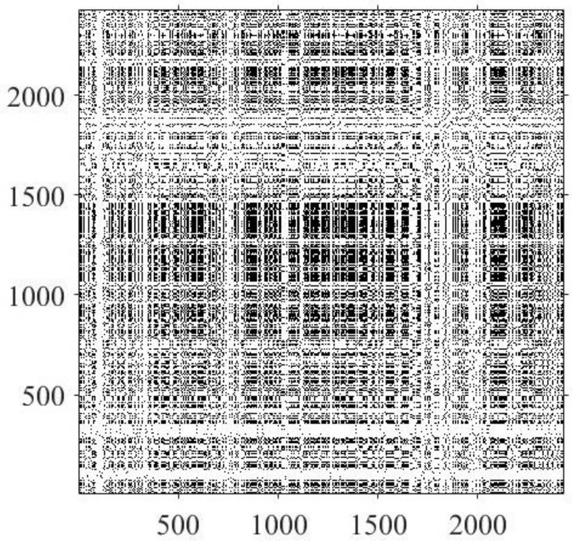


Fig. 2. Recurrence plot of daily values of BTC/$ price fluctuations.

Webber and Zbilut developed a tool for calculating a number of measures based on the calculation of the density of recurrence points and the construction of the frequency distribution of diagonal line lengths: recurrence rate (*RR*), determinism (*DET*), divergence (*DIV*), entropy (*ENTR*), trend (*TREND*), laminarity (*LAM*), trapping time (*TT*). The calculation of these measures in the submatrices of the recurrence chart along the identity line shows the behavior of these measures over time. Some studies of these measures have shown that their application can help identify bifurcation points, chaos-order transitions.

Let R_{ij}=1 if (i, j) are recurrent, otherwise R_{ij}=0; and D_{ij}=1 if (i, j) and $(i$+1, j+1) or (i–1, j–1) are recurrent, otherwise D_{ij}=0.

Now the coefficients of self-similarity and predictability will be, respectively, equal $RR = \frac{1}{N^2}\sum_{i,j=1}^{N} R_{ij}$ and $DET = \frac{\sum_{i,j=1}^{N} D_{ij}}{\sum_{i,j=1}^{N} R_{ij}}$.

To demonstrate the values of the described indicators, we used 3 time series with 1/*f*-noise with α=0.5, α=0.75, α=1.0, a time series of sine values and the same series with mixed values. The time series with 1/*f*-noise are shown in figure 3.

The figure 4 shows the graphs of the RR values for the demonstration time series. As can be seen from the figure for more ordered rows (another words the rows with less noise), the value of RR is higher.

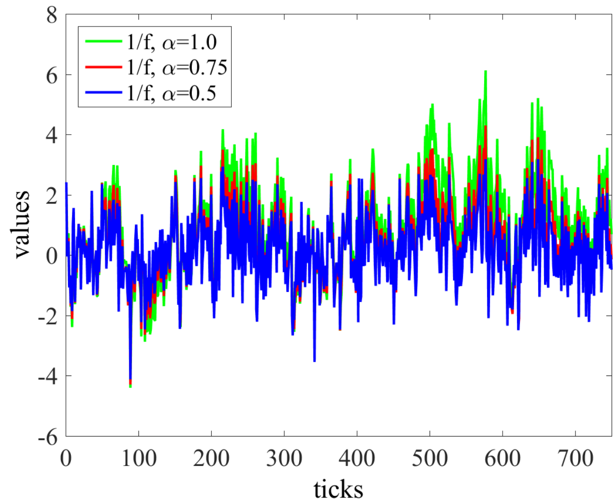


Fig. 3. The time series with 1/*f*-noise with α=0.5, α=0.75, α=1.0, and the time series of sine values and the mixed sin values.

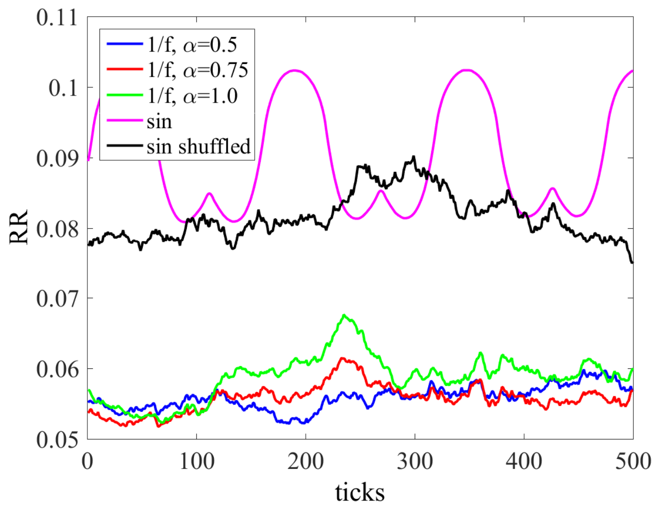


Fig. 4. Self-similarity (*RR*) of 1/*f*-noise and sin time series.

The figure 5 shows the graphs of the *DET* values for the demonstration time series. As can be seen from the figure for more ordered rows (another words the rows with less noise), the value of *DET* is higher.

If N_l is the number of diagonal lines, and l_i is the length of the i-th diagonal line, then the length of the longest diagonal line is determined by the expression

$$L = max(\, l_i; i = 1, \dots, N_l).$$

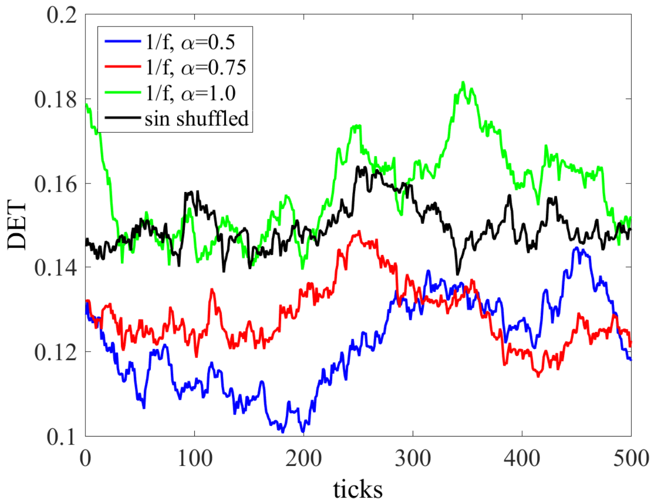


Fig. 5. Determinism (*DET*) of 1/*f*-noise and sin time series.

A diagonal line of length l means that the segment of the trajectory is close for l steps of time to another segment of the trajectory at another time; therefore, these lines are associated with the divergence of the trajectory segments.

The average length of the diagonal line $L = \frac{\sum_{l=l_{min}}^{N\Sigma} lP(l)}{\sum_{l=l_{min}}^{N\Sigma} P(l)}$ is the average time during which the two segments of the trajectory are close to each other, and can be interpreted as the average time of the forecast.

The average length of vertical structures is given by expression $TT = \frac{\sum_{v=v_{min}}^{N\Sigma} vP(v)}{\sum_{v=v_{min}}^{N\Sigma} P(v)}$ and is called the delay time or capture time. Its calculation also requires consideration of the minimum length v_{min}, as in the case of *LAM*. The *TT* estimates the average time that the system will be in a certain state, or how long this state will be captured.

The figures 6 and 7 present graphs of L and *TT* calculation for the demonstration time series.

As the measures can refer to the diagonal and/or horizontal lines on the recurrence map, at the same time, there are vertical lines with appropriate measures.

The total number of vertical length lines in *RP* is given by the histogram

$$P(v) = \sum_{i,j=1}^{N}(1 - R_{i,j})(1 - R_{i,j+v}) \prod_{k=0}^{v-1} R_{i,j+k}.$$

Similar to the definition of determinism, the ratio of recurrent points that form vertical structures to a complete set of recurrent points can be calculated as $LAM = \frac{\sum_{v=v_{min}}^{N\Sigma} vP(v)}{\sum_{v=1}^{N} vP(v)}$. This measure is called laminarity (*LAM*). Laminarity calculations are performed for those v that exceed the minimum length v_{min}. For recurrence maps often take v_{min}=2. The value of *LAM* decreases if *RP* consists of more single recurrent points than vertical structures.

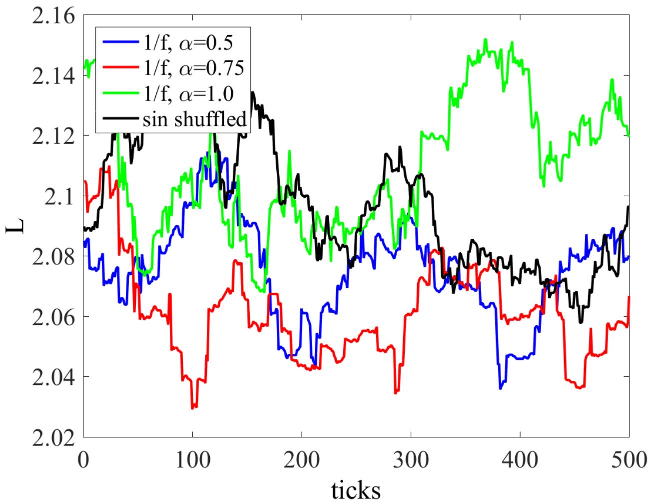


Fig. 6. The diagonal lines measure (*L*) of 1/*f*-noise and sin time series.

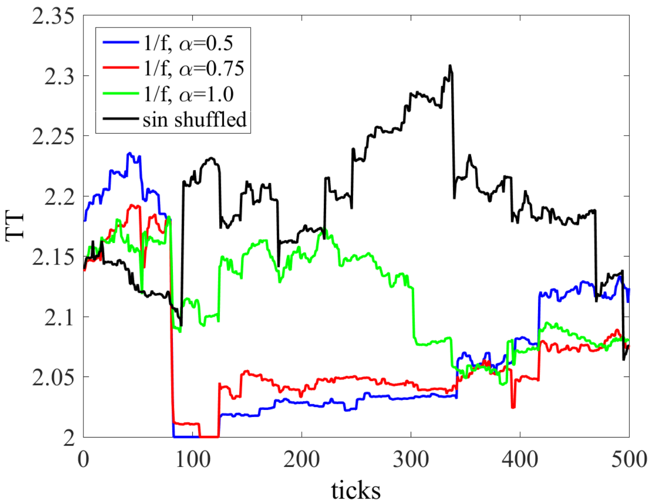


Fig. 7. The delay time measure (*TT*) of 1/*f*-noise and sin time series.

An example of *LAM* calculation for the demonstration time series, is given in figure 8. In contrast to quantitative measures based on diagonal lines, the measures just introduced can be applied to chaos-chaos transitions. The last two parameters characterize two different typical time intervals during which the trajectories are close to ε. Their window dynamics allows the one to track the time component of recurrence maps.

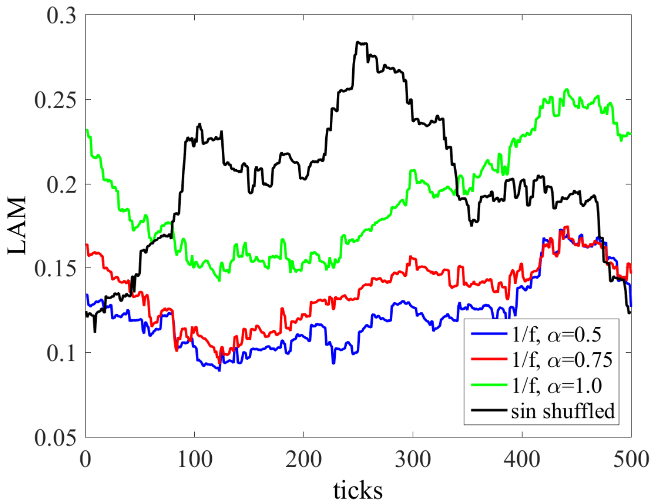


Fig. 8. The laminarity (*LAM*) of daily values of BTC/$ fluctuations and returns of fluctuations.

2.2 The recurrence-based entropy

An important class of recurrence quantifiers are those that try to capture the level of complexity of a signal. As an example, we mention the already known entropy based on diagonal lines statistics. This quantity has been correlated with others dynamical quantifiers as, for example, the largest Lyapunov exponent, since both capture properties of the complexity level of the dynamics. The vertical (horizontal) lines in R_{ij} are associated to laminar states, common in intermittent dynamics [13]. It was reported the use of the distribution of diagonal lines $P(l)$ for a different quantifier of recurrences, based on the Shannon entropy [13]. If we choose a distribution of diagonals $p(l) = P(l)/\sum_{l=1}^{K} P(l)$ for K the maximum length of the diagonal lines, then we get one of the known quantitative indicators of recurrence analysis: $ENTR = -\sum_{l=l_{min}}^{l=l_{max}\Sigma} p(l)\, ln\, p\,(l)$. However, as follows from the analysis of entropy indicators, the results are not always possible to coordinate with the proposed models.

To the pleasure of the researchers, it turned out that depending on the technology of using the properties of the recurrence of the phase space, different types of recurrence entropies are distinguished [7].

2.3 Recurrence probability (period) density entropy

Recurrence probability (or period) density entropy (RPDEn) is useful for characterizing the extent to which a time series repeats the same sequence [1; 11; 15] and is, like the *ENTR* a quantitative characteristic of recurrence analysis. Around each point x_n in the phase space, an ε-neighbourhood (an m-dimensional ball with this radius) is formed, and every time the time series returns to this ball, after having left it, the time difference T between successive returns is recorded in a histogram. This histogram is normalized to sum to unity, to form an estimate of the recurrence period density function $P(T)$. The normalized entropy of this density $H_{norm} = -\sum_{t=1}^{T_{max}\Sigma} P(t)\, ln\, P\,(t) \,/\, \ln T_{max}$ is the RPDEn value, where T_{max} is the largest recurrence value.

2.4 Recurrence entropy

Recent works [3; 12] presents a slightly different technique for calculating recurrence entropy using a novel way to extract information from the recurrence matrix. The authors have generalize these concepts recurrence defining recurrence microstates $F(\varepsilon)$ as all possible cross-recurrence states among two randomly selected short sequences of N consecutive points in a K $(K \geq N)$ length time series, namely $F(\varepsilon)$ are $N \times N$ small binary matrices. The total number of microstates for a given N is $N_{ms} = 2^{N^2}$. The microstates are populated by $\bar{N}$ random samples obtained from the recurrence matrix such that $\bar{N} = \sum_{i=1}^{N_{ms}} n_i$, where n_i is the number of times that a microstate i is observed. For $P_i = n_i/\bar{N}$, the probability related to the microstate i, we define an entropy of the RP associated with the probabilities of occurrence of a microstate as $S(N_{ms}) = \sum_{i=1}^{N_{ms}} P_i\, ln\, P_i$.

3 The recurrence plot-based measures for crash time series

The behavior of the measures described in section 2.1 was conducted on the basis of daily data of the Dow Jones Industrial Average (DJIA), taken for the period from 1990 to 2019, and daily values of the price on the spot oil market, taken for the period from 1987 to 2019, in order to assess the dynamics of changes in the values of indicators in certain critical periods of economic systems, tables of critical and crisis phenomena in the relevant markets were compiled.

Table 1 lists the critical and crisis phenomena in the DJIA for the period under study.

Calculations of investigating measures of complexity were carried out within the framework of a moving (sliding) window algorithm. For this purpose, the part of the time series (window), for which there were measures of complexity, was selected, then the window was displaced along the time series in a one-day increment and the procedure repeated until all the studied series had exhausted. Further, comparing the dynamics of the actual time series and the corresponding measures of complexity, we can judge the characteristic changes in the dynamics of the behavior of complexity with changes in the time series. If this or that measures of complexity behaves in a definite way for all periods of crisis, for example, decreases or increases during the pre-crisis period, then it can serve as an indicator or precursor of such a crisis phenomenon.

Table 1. Critical and crisis phenomena in the DJIA for the period from 01.01.1990 to 01.06.2019.

N	Time period	Duration, days	Falling rate, %
1	17.07.1990-23.08.1990	28	17.21
2	01.10.1997-21.10.1997	15	12.43
3	17.08.1998-31.08.1998	11	18.44
4	14.08.2002-01.10.2002	34	19.52
5	16.10.2008-15.12.2008	42	30.21
6	09.08.2011-22.09.2011	32	11.94
7	18.08.2015-25.08.2015	6	10.53
8	29.12.2015-20.01.2016	16	11.02
9	03.12.2018-24.12.2018	15	15.62

In figure 9 presents the results of calculations the measures *RR*, *DET*, *L*, *TT*, and *LAM* for the DJIA values database. The calculations were carried out for a moving window size of 500 days and a step of 1 day.

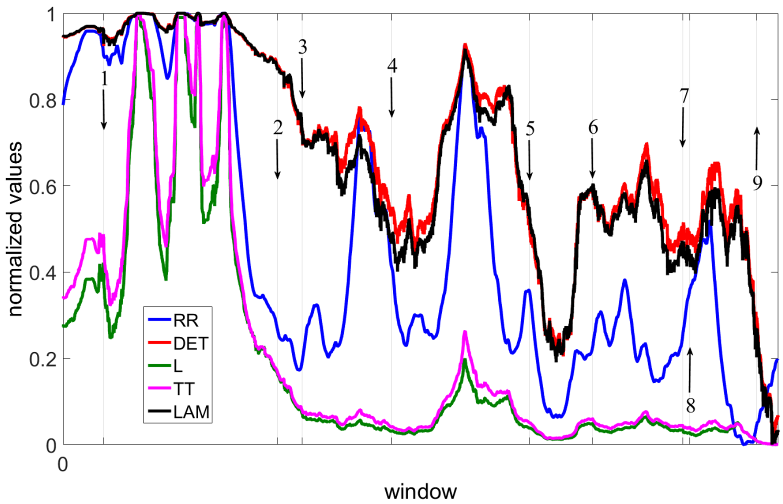


Fig. 9. Window recurrence plot-based measures of complexity for the crashes presented in table 1. The start point of the crash is marked.

The figure shows that the value of the measure of self-similarity (*RR*) in 3 cases out of 9 decreases during the critical phenomenon (events 1, 2, 4, 5), and in 2 cases out of 9 is in the local minimum (events 3, 6). Note that these critical phenomena are quite long with a length of 11 to 42 days, and also have a large percentage of falls: from 12 to 30%. Phenomena 7, 8, 9, which the indicator did not feel, have a shorter duration and a lower percentage of fall.

The determinism (*DET*) measure shows a tendency to fall during all the studied critical phenomena, which, given the stable result, can serve as an indicator of a critical phenomenon. Moreover, in 3 cases out of 9 during the critical phenomenon, the indicator is at the local maximum, after which it begins to decrease. Thus, the *DET* measure can be an indicator of the beginning of a critical phenomenon.

We have noted the same behavior of the determinism and the laminarity measures, the form of graphs of the dynamics of which is very similar. Thus, it is enough to choose only one of the two measures to build economic instruments.

Finally, the form of graphs for the average length of lines *L* and delay time *TT* is the same, which allows using only one of these measures in the research. The values of the measures decrease during 7 of the 9 studied critical phenomena, and during the remaining 2 phenomena the values are at the point of local maximum, after which they begin to fall.

Therefore, based on the analysis of the values of the DJIA index, the following intermediate conclusions were obtained:

1. Only 3 measures out of the 5 studied can be used in research, because there are measures with the same form of graphs, due to which only one can be chosen.
2. The most sensitive to critical phenomena are the measures of *DET*, *L*, *TT*, and *LAM*.

During calculations, we have used a window with a width of 500 points. In order to determine the degree of influence of the window width on the dynamics of the analyzed characteristics, we repeated the calculations using a window width of 250 points. In figure 10 shows graphs of the dynamics of three measures, *RR*, *DET* and *LAM*, obtained for windows with a width of 500 and 250 points. A comparative analysis of the graphs led to the conclusion that although they differ to some extent, however, both graphs retain properties that are essential for our analysis, namely: the starting points of critical phenomena fall into local extremes or areas of decreasing values. This allows us to hypothesize the informativeness of the results, the possibility of using different window widths in the calculation procedure (but not arbitrary width!), and further search for the optimal value of the parameter to minimize calculations while maintaining the informativeness of the result.

Often, time series studies also use not the initial data, but the returns, which for the time series x_i, i=1,..,M, are calculated using the expression $r_i = \frac{x_{i+1}-x_i}{x_i}$. In figure 11 presents the results of calculations the measures *RR*, *DET*, *L*, *TT*, and *LAM* for the DJIA returns database. The calculations were carried out for a moving window size of 500 days and a step of 1 day.

The figure 11 clearly shows approximately the same behavior for all studied indicators. Note the chaotic behavior, that is, the moments of the beginning of critical phenomena exist both on the areas of growth of indicators and on the areas of their decline. Therefore, based on the obtained partial result, we can hypothesize the impossibility of using returns as indicators of critical phenomena for measures based on recurrent plots.

In studying the behavior of the measures described above on the spot oil market prices, critical phenomena were used, the list of which is given in table 2.

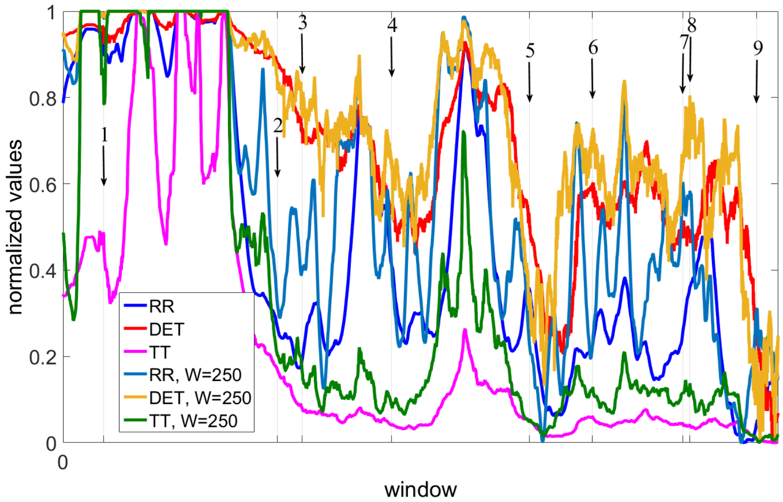


Fig. 10. Window recurrence plot-based measures of complexity for the crashes presented in table 1. The window width is 250 in the window moving procedure. The start point of the crash is marked.

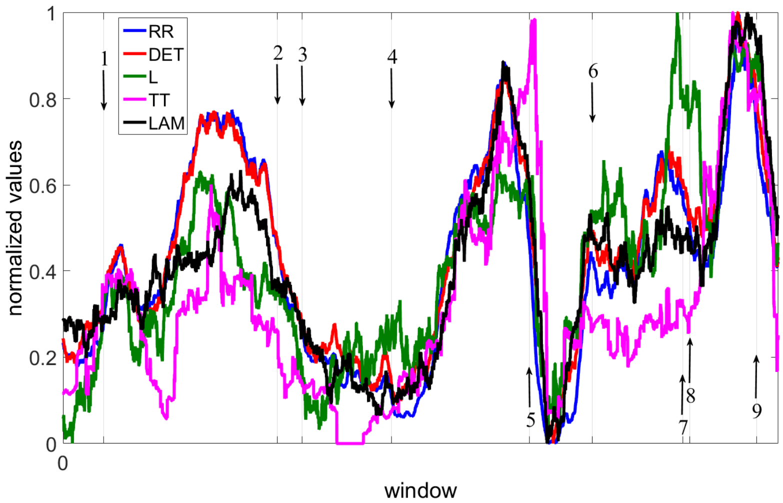


Fig. 11. Window recurrence plot-based measures of complexity for the crashes presented in table 1 using the returns of DJIA. The start point of the crash is marked.

In figure 12 presents the results of calculations the measures *RR*, *DET*, *L*, *TT*, and *LAM* for the spot oil price values database. The calculations were carried out for a moving window size of 500 days and a step of 1 day.

Table 2. Critical and crisis phenomena on the spot oil market for the period from 01.01.1987 to 01.06.2019.

N	Time period	Duration, days	Falling rate, %
1	09.12.1987-21.12.1987	9	18
2	11.10.1990-23.08.1990	8	31
3	17.11.1993-17.12.1993	22	18
4	11.04.1996-05.06.1996	38	22
5	30.09.1998-25.11.1998	40	33
6	07.03.2000-10.04.2000	24	29
7	27.11.2000-20.12.2000	17	29
8	14.09.2001-24.09.2011	6	27
9	12.03.2003-21.03.2003	7	28
10	26.10.2004-10.12.2004	31	28
11	07.08.2006-17.11.2006	73	27
12	03.07.2008-23.12.2008	120	80
13	03.05.2010-25.052010	16	25
14	29.04.2011-17.05.2011	12	15
15	24.02.2012-28.06.2012	87	27
16	06.09.2013-27.11.2013	58	17
17	20.06.2014-29.01.2015	152	59
18	03.11.2015-20.01.2016	52	44
19	03.10.2018-27.12.2018	56	41

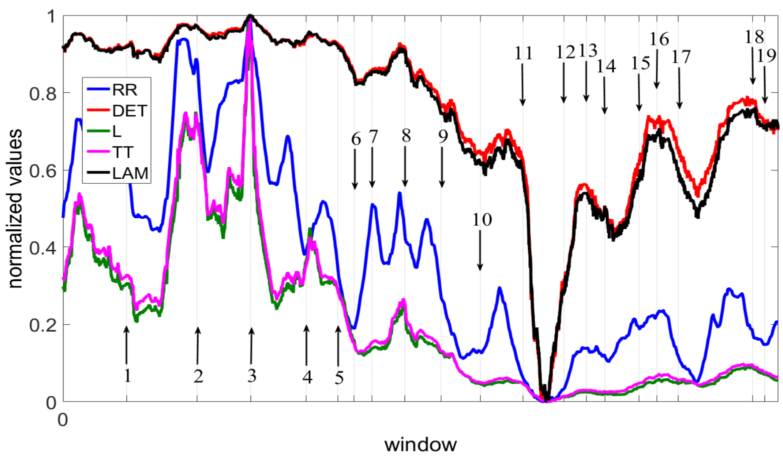


Fig. 12. Window recurrence plot-based measures of complexity for the crashes presented in table 2. The time series of the spot oil price is used. The start point of the crash is marked.

The figure 12 shows that the value of the measure of self-similarity (*RR*) in 9 cases out of 19 decreases during the critical phenomenon, in 5 cases out of 19 is in the local minimum, and in 2 cases out of 19 is in the local maximum. In other cases, the critical phenomenon is on the area of growth of the indicator. Interestingly, the falling rate of critical phenomena, the onset of which goes to the local extremum, is about 30%.

The determinism measure and the laminarity measure, as in the case for DJIA time series, have very similar dynamics. The studied critical phenomena are mainly in the areas of falling indicators: 14 out of 19. Practically the absence of existence indicators in local extrema indicates the impossibility of their use as precursors, but only as indicators of possible critical phenomena.

Measures of average line length (L) and time delay (TT) also coincide in their dynamics. In 10 cases out of 19 at the time of the critical phenomenon, the indicators of measures are decreasing, and in 5 cases out of 19 are in the local maximum. Given that only in one case the indicators of measures are in the local minimum, it is possible to hypothesize the possibility of interpreting the indicators not only as indicators but also as precursors with a short prediction horizon.

Therefore, based on the analysis of the values of the spot oil price, the following intermediate conclusions were obtained:

1. As in previous analysis, the only 3 measures out of the 5 studied can be used in research.
2. The most sensitive to critical phenomena are the measures of L and TT.

In figure 13 presents the results of calculations the measures *RR*, *DET*, *L*, *TT*, and *LAM* for the spot oil price returns database. The calculations were carried out for a moving window size of 500 days and a step of 1 day. Like for the DJIA returns, the figure clearly shows approximately the same behavior for all studied indicators. Note the chaotic behavior, that is, the moments of the beginning of critical phenomena exist both on the areas of growth of indicators and on the areas of their decline. Therefore, based on the obtained result, we confirmed the impossibility of using returns as indicators of critical phenomena for measures based on recurrence plots.

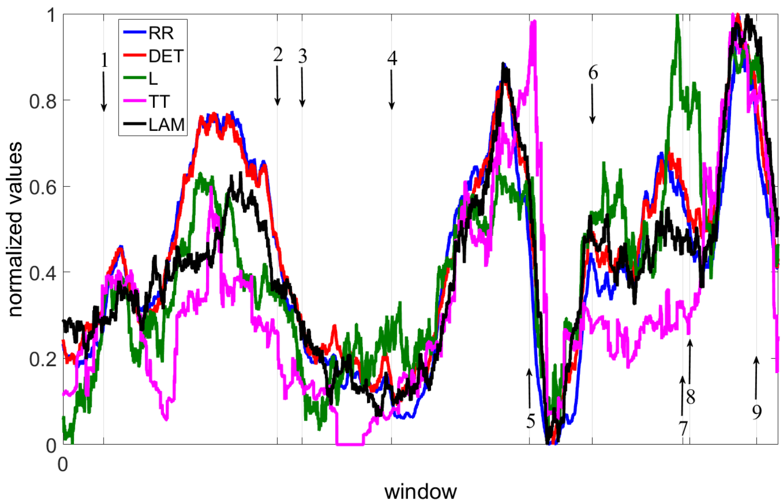


Fig. 13. Window recurrence plot-based measures of complexity for the crashes presented in table 2. The time series of the spot oil price returns is used. The start point of the crash is marked.

4 Recurrence entropy for crash time series

To study the recurrence entropy properties of time series, including periods of crisis, the following databases have been prepared. The first database included fragments of the DJIA for the famous crashes of 1929, 1987, and 2008. In a number of daily values of the DJIA index of 2000 days long, the actual day of the onset of the crash falls at point 1000 (figure 14). In this case, the fixed point of crashes can easily observe the indicator capabilities of entropy measures of complexity.

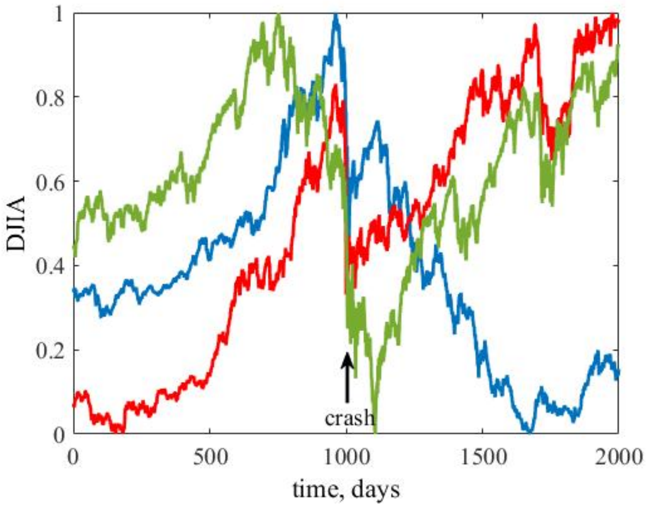


Fig. 14. Fragments of DJIA index with crash at 1000 days.

The following database contains the same length daily values (from March 3, 1990 to August 30, 2019) of the US stock market indices (DJIA), Germany (DAX), France (CAC), used to check the universality of the complexity measure regardless of the index. The index DJIA is also taken for the period from January 1, 1983 to August 30, 2019 in order to cover the crises of 1987 and 1998.

The third database includes the values of daily Bitcoin prices for the entire observation period (from July 17, 2010) and for a shorter period of stabilization of the cryptocurrency market from January 1, 2013 to August 30, 2019.

Calculations of recurrence entropy measures of complexity were carried out within the framework of a moving (sliding) window algorithm [5]. If this or that measures of complexity behaves in a definite way for all periods of crisis, for example, decreases or increases during the pre-crisis period, then it can serve as an indicator or precursor of such a crisis phenomenon.

In Figures 15 and 16 presents the results of calculations RPDEn and RecEn for the first database with a length of 2000 days.

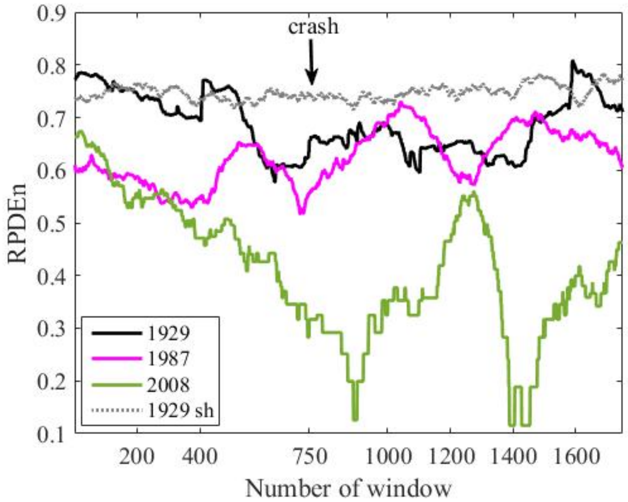


Fig. 15. Window recurrence measures of complexity for the crashes of 1929, 1987, and 2008 using RPDEn procedure. The start point of the crash is marked.

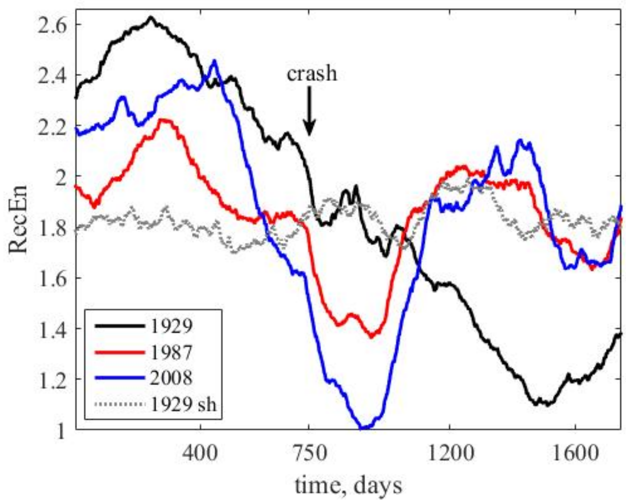


Fig. 16. Window recurrence measures of complexity for the crashes of 1929, 1987, and 2008. a) RPDEn, b) RecEn. The start point of the crash is marked.

The calculations were carried out for a moving window size of 250 days and a step of 1 day. It can be seen from the figure that the recurrence entropy in the pre-crisis period is markedly reduced for all crisis events, which is obviously a precursor of such crisis phenomena. As for RPDEn, such an unambiguous precursor is not observed. Therefore, further we focus on the use of RecEn, leaving for the future a more complete study of RPDEn.

In figure 17 shows the RecEn dynamics for the long index DJIA, which includes the last seven well-known crashes (shown in the figure).

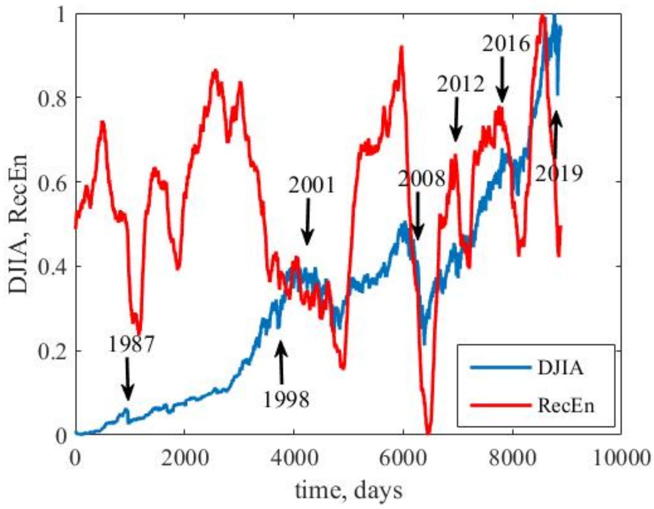


Fig. 17. Comparative dynamics of index DJIA and recurrence entropy RecEn.

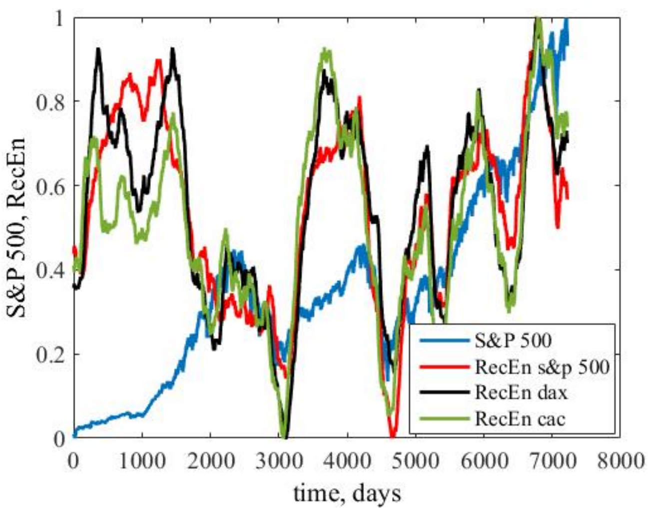


Fig. 18. The dynamics of the stock index S&P 500 and its corresponding and also the indices of the DAX and CAC of recurrence entropy.

Obviously, in this case, RecEn is the precursor of crash events in all these cases. In order to once again verify the universality of RecEn as an indicator-precursor of financial crashes, we examined its dynamics for various stock indices. As an example, the selected indices are the stock markets of the USA (S&P 500), Germany (DAX) and France (CAC) for a comparable period of time (figure 18).

Finally, the analysis of a very volatile cryptocurrency market for BTC/$ data with small window values (50 days) also allows us to identify the main crisis falls in this market (figure 19).

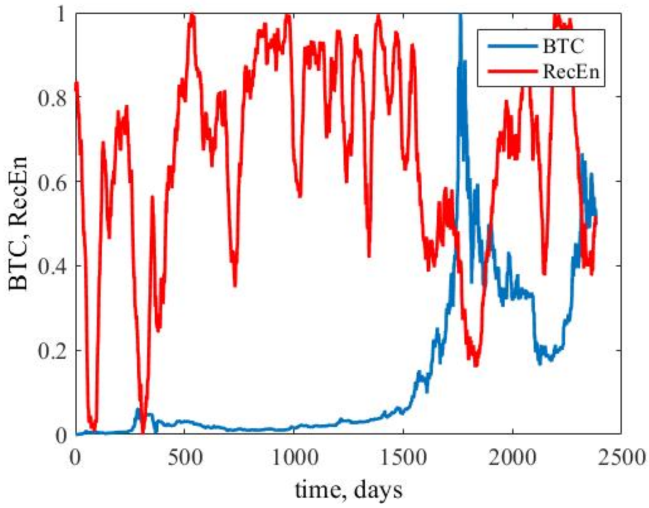


Fig. 19. Comparison of the dynamics of the BTC/$ price with the corresponding recurrence entropy.

The periodization of Bitcoin crises, which we conducted earlier, indicates that recurrence entropy in this case is also a harbinger of crisis phenomena.

5 Conclusion

We have analyzed key measures based on recurrence plots that can be used as indicators and precursors of critical phenomena in complex economic systems. Based on the analysis, several main measures were identified that showed satisfactory results for their use in tools to indicate critical phenomena. Such characteristics were, first of all, the delay time (*TT*) and the average length of the lines on the recurrence plot (*L*), and it was determined that only one of them can be used in studies due to the similarity of the forms of their graphs. A measure of determinism (*DET*) or a measure of laminarity (*LAM*) can also be used to identify critical phenomena. In the future we can focus on the study of the behavior of these characteristics in the analysis of complex economic systems of different nature.

We have demonstrated also that the entropy analysis of financial time series in phase space reveals the characteristic recurrent properties of complex systems. It turned out that recurrence entropy, unlike other entropy indicators of complexity, is an indicator and an early harbinger of crisis phenomena. The recurrence entropy methodology has several advantages compared to the traditional recurrence entropy defined in the literature, namely, the correct evaluation of the chaoticity level of the signal, the weak dependence on parameters. In the future, a thorough comparative analysis of the possibilities of recurrence entropy with other promising types of entropy indicators of complexity should be carried out [4; 8; 10; 14; 19].

References

1. Arunachalam, S.P., Kapa, S., Mulpuru, S.K., Friedman, P.A., Tolkacheva, E.G.: Rotor pivot point identification using recurrence period density entropy. In: 54th Annual Rocky Mountain Bioengineering Symposium, RMBS 2017 and 54th International ISA Biomedical Sciences Instrumentation Symposium 2017, 2017-March
2. Bielinskyi, A., Soloviev, V., Semerikov, S., Solovieva, V.: Detecting stock crashes using Levy distribution. CEUR Workshop Proceedings **2422**, 420–433 (2019)
3. Corso, G., Prado, T., Lima, G., Lopes, S.: A novel entropy recurrence quantification analysis. arXiv:1707.00944v1 [stat.OT] (2017)
4. Danylchuk, H., Derbentsev, V., Soloviev, V., Sharapov, A.: Entropy analysis of dynamics properties of regional stock market. Science and Education a New Dimension. Economics **4**(2), 15–19 (2016_
5. Derbentsev, V., Semerikov, S., Serdyuk, O., Solovieva, V., Soloviev, V.: Recurrence based entropies for sustainability indices. E3S Web of Conferences **166**, 13031 (2020). doi:10.1051/e3sconf/202016613031
6. Diaz, J.F.T.: Evidence of Noisy Chaotic Dynamics in the Returns of Four Dow Jones Stock Indices. Annual Review of Chaos Theory, Bifurcation and Dynamical System **4**, 1–15 (2013)
7. Faure P., Lesne, A.: Estimating Kolmogorov entropy from recurrence plots. In: Webber, G.C.L., Marwan, N. (eds.) Recurrence Quantification Analysis, pp. 45–63. Springer International Publishing, Cham (2015)
8. Gu, R.: Multiscale Shannon Entropy and its application in the stock market. Physica A: Statistical Mechanics and its Applications **484**, 215–224 (2017)
9. Kantz, H., Shreiber, T.: Nonlinear time series analysis, 2nd edn. Cambridge University Press, Cambridge (2004)
10. Lim, J.R.: Rapid Evaluation of Permutation Entropy for Financial Volatility Analysis – A Novel Hash Function using Feature-Bias Divergence". Department of Computer Science, Imperial College of London, London. https://www.doc.ic.ac.uk/teaching/distinguished-projects/2014/j.lim.pdf (2014). Accessed 17 Aug 2020
11. Little, M.A., McSharry, P.E., Roberts, S.J., Costello, D.A.E., Moroz, I.M.: Exploiting nonlinear recurrence and fractal scaling properties for voice disorder detection. BioMedical Engineering OnLine **6**, 23 (2007)
12. Lopes, S., Lima, T., Corso, G., Lima, G., Kurths, J.: Parameter-free quantification of stochastic and chaotic signals. arXiv: 1905.02284v1 [physics.data-au] (2019)
13. Marwan, N., Carmen Romano, M., Thiel, M., Kurths, J.: Recurrence plots for the analysis of complex systems. Phys. Rep. 438(5–6), 237–329 (2007)
14. Pele, T., Lazar, E., Dufour, A.: Information Entropy and Measures of Market Risk. Entropy **19**(5), 1–19 (2017)
15. Rabarimanantsoa, H., Achour, L., Letellier, C., Cuvelier, A., Muir, J.-F.: Recurrence plots and Shannon entropy for a dynamical analysis of asynchronisms in noninvasive mechanical ventilation. Chaos **17**, 013115 (2007)

16. Soloviev, V., Bielinskyi, A., Solovieva, V.: Entropy analysis of crisis phenomena for DJIA index. CEUR Workshop Proceedings **2393**, 434–449 (2019)
17. Soloviev, V.N., Belinskiy, A.: Complex Systems Theory and Crashes of Cryptocurrency Market. Communications in Computer and Information Science **1007**, 276–297 (2019)
18. Sornette, D.: Why Stock Markets Crash: Critical Events in Complex Systems. Princeton University Press, Princeton (2003)
19. Wang, G.-J., Xie, C., Han, F.: Multi-Scale Approximate Entropy Analysis of Foreign Exchange Markets Efficiency. Systems Engineering Procedia **3**, 201–208 (2012)
20. Webber, C.L., Zbilut, J.P.: Recurrence quantification analysis of nonlinear dynamical systems. In: Riley, A., Van Orden. G.C. (eds.) Contemporary Nonlinear Methods for the Behavioral Sciences, pp. 26–94. http://www.nsf.gov/sbe/bcs/pac/nmbs/nmbs.jsp (2005). Accessed 21 Mar 2017
21. Zhou, R., Cai, R., Tong, G.: Applications of entropy in finance: A review. Entropy **15**(11), 4909–4931 (2013)

Casual analysis of financial and operational risks of oil and gas companies in condition of emergent economy

Inesa Khvostina[1[0000-0001-5915-749X]], Serhiy Semerikov[2,3,4[0000-0003-0789-0272]], Oleh Yatsiuk[1[0000-0002-3943-7352]], Nadiia Daliak[1[0000-0002-1599-842X]], Olha Romanko[1[0000-0003-1587-1370]] and Ekaterina Shmeltser[5[0000-0001-6830-8747]]

[1] Ivano-Frankivsk National Technical University of Oil and Gas, 15 Karpatska Str., Ivano-Frankivsk, 76019, Ukraine
[2] Kryvyi Rih State Pedagogical University, 54 Gagarin Ave., Kryvyi Rih, 50086, Ukraine
[3] Kryvyi Rih National University, 11 Vitalii Matusevych Str., Kryvyi Rih, 50027, Ukraine
[4] Institute of Information Technologies and Learning Tools of NAES of Ukraine, 9 M. Berlynskoho Str., Kyiv, 04060, Ukraine
[5] State University of Economics and Technology, 5 Stepana Tilhy Str., Kryvyi Rih, 50006, Ukraine
inesa.hvostina@gmail.com, semerikov@gmail.com, olegstya@gmail.com, nadiya_d82@ukr.net, olgaromanko11@gmail.com, shmelka0402@gmail.com

Abstract. The need to control the risk that accompanies businesses in their day-to-day operations, and at the same time changing economic conditions make risk management an almost indispensable element of economic life. Selection of the main aspects of the selected phases of the risk management process: risk identification and risk assessment are related to their direct relationship with the subject matter (risk identification to be managed; risk analysis leading to the establishment of a risk hierarchy, and, consequently, the definition of risk control' methods) and its purpose (bringing the risk to acceptable level). It is impossible to identify the basic patterns of development of the oil and gas industry without exploring the relationship between economic processes and enterprise risks. The latter are subject to simulation, and based on models it is possible to determine with certain probability whether there have been qualitative and quantitative changes in the processes, in their mutual influence on each other, etc. The work is devoted to exploring the possibilities of applying the Granger test to examine the causal relationship between the risks and obligations of oil and gas companies. The analysis is based on statistical tests and the use of linear regression models.

Keywords: risk, risk identification, casual analysis, causality.

1 Introduction

One of the most important factors that accompany any business activity, including production and commercial activity, is risk. This is because every business operates in a volatile and uncertain environment, in which unforeseen events can occur. Currently, many businesses manage risk by using an early warning system that allows early detection of the threat and initiation of appropriate corrective processes. For businesses, this is a signal to take action to improve the situation [9]. Today, risk is associated with every decision, whether operational, investment or financial. The planning's nature shows that the predicted future values of the variables taken into account are not certain but only probable. Different quantitative parameters assume values only under certain assumptions. The need to monitor risk performance is an indispensable element in achieving the objective of business entities, as the entity, as a market-oriented, earnings-oriented entity, is subject to any management and casual events. Risks that threaten the capital used by the owner or owners mean that the costs associated with realizing that risk, unless they can be transferred to someone else, are borne by the enterprise. Therefore, for the continuity of the company, it is crucial to determine the impact of the adverse event of a casual event on the value of assets and to take appropriate measures to transfer this burden to other entities.

The processes occurring in the operations of oil and gas companies are related to the various forms, frequency and nature of the risks involved and necessitate the study of their cause and effect relationships. Given that risk is an integral part of business entities and a necessary element of economic decision making, the risk of which increases significantly in the conditions of dynamism and instability of the business environment, there is a need to consider risk as an object of managerial influence and its complex analysis.

The purpose of the article is to present the role of identification and measurement of financial risk in the process of managing enterprises of the oil and gas industry by establishing interdependencies between its elements.

2 Background

Risk research has recently received considerable attention. A lot of work is devoted to examining the role and importance of risk for practice. The main issues of the theory and practice of economic risk assessment are outlined in the works of Mohamed Abdel-Basset [1], Bamikole Amigun [2], Vanessa E. Daniel [5], Luca Salvati [32], Qianqian Zhou [35] and others. A wide range of issues related to risk assessment and forecasting are investigated by Olga O. Degtiareva [6], Oleh G. Dzoba [3], Iryna Yu. Ivchenko [12], Oleh Ye. Kuzmin [17; 18], Valentyna V. Lukianova [19], Nazar Yu. Podolchak [31], Liliia I. Rishchuk [25], O. V. Shcherbak [33], Halyna I. Velikoivanenko [21], Valdemar V. Vitlinskyi [4] and others. It is worth noting that most scientists in their scientific work focus on the study of classical methods of risk assessment, while the latest methods, which are the most promising for the functioning of enterprises in a dynamic business environment, remain underutilized.

One of the methods to reduce the risk of an enterprise is its insurance [16]. This method of risk management must be considered in conjunction with other methods of risk diversification. For optimal risk management at an enterprise it is necessary to use portfolio theory [24]. Economic, financial and other parameters of the enterprise's operation act as constraints, and minimization of the enterprise's risk can be used as an optimization criterion.

For the analysis of financial and economic sustainability of the enterprises are used as classical statistical modeling techniques and advanced mathematical tools such as fractal analysis [20] as well as the methods of artificial intelligence [14; 22].

To predict financial time series, artificial intelligence tools are often used, which include as machine learning methods [7; 15].

The works [8; 34] are devoted to a comparative analysis of the complexity of traditional stock indices and social responsibility indices using the example of Dow Jones Sustainability Indices and Dow Jones Industrial Average and opens up new opportunities for investor risk management. A scientific approach to risk assessment taking into account the manifestation of emergent properties and using the method of taxonomy and factor analysis for oil and gas companies is proposed in [13].

In [11], an approach is proposed for assessing the financial efficiency of a business model of an industrial enterprise, where the integral indicator of the financial components of a business model is modeled using the method of fuzzy sets and taxonomic analysis, which will help to more objectively assess the level of financial standing of an industrial enterprise.

3 Methodology

The oil and gas industry of Ukraine is one of the most important components of the fuel and energy complex, but the needs of the domestic economy in oil and gas are only partially met at the expense of its own production. This issue is becoming increasingly relevant today. In order to increase the economic potential of oil and gas companies, it is necessary to infuse investment resources. However, the instability of tax legislation is one of the most significant shortcomings of Ukraine's current tax system, which deter investors. Inconsistency in the application, interpretation and implementation of tax law can lead to litigation, which can ultimately lead to additional taxes, penalties and penalties, and these amounts can be significant. All the above clearly indicates that the oil and gas companies of Ukraine operate in an emergent economy and this fact largely determines the high relevance of our study.

In the ordinary course of business, certain claims are raised against oil and gas companies. If the risk of an outflow of financial resources related to such claims is considered probable, a liability is recognized in the provisions for litigation. If management estimates that the risk of an outflow of financial resources related to such claims is probable or the amount of expenses cannot be estimated reliably, the provision is not recognized and the corresponding amount disclosed in the consolidated financial statements.

There is a claim between businesses and some natural gas suppliers about the volume or prices of the natural gas being supplied and other claims. Management estimates its potential liabilities for such claims at UAH 5890 million (2016: UAH 1380 million; 2017: UAH 3928 million; 2018: UAH 4246 million) [29; 28; 30; 27; 26]. Management cannot reliably estimate the amount of potential losses on these liabilities, if any.

The activities of the oil and gas industry are characterized by a number of financial risks: market risk (including foreign exchange and interest rate risk), concentration risk, credit risk and liquidity risk. Management reviews and aligns its risk management policies to minimize the adverse impact of these risks on the Group's financial performance.

The main categories of financial instruments are presented by structure of financial assets and financial liabilities.

Market risk. Market risks arise from open positions in (a) foreign currencies, (b) interest-bearing assets and liabilities, and (c) investments, all of which are affected by general and specific market changes in condition of emergent economy.

Currency risk. Oil and gas companies operate in Ukraine, and their dependence on foreign exchange risk is mainly determined by the need to purchase natural gas from foreign suppliers, which is denominated in US dollars. The Group also receives foreign currency loans and does not hedge its foreign currency positions.

Dependence on currency risk is presented on the basis of the carrying amount of the respective currency assets and liabilities.

Table 1 provides information on the sensitivity of profit or loss to reasonably possible changes in the exchange rates applied at the reporting date, provided that all other variables remain stable. The risk was calculated only for monetary denominations denominated in currencies other than the functional currency.

Table 1. Profit or loss sensitivity to reasonably possible changes in exchange rates [29; 28; 30; 27; 26]

In millions of Ukrainian Hryvnias	December 31, 2019	December 31, 2018	December 31, 2017	December 31, 2016
US dollar Strengthening by 10%	(2540)	(2865)	(3 400)	(4 957)
US dollar weakening by 10%	2540	2865	3 400	4 957
Euro Strengthening by 10%	225	239	251	299
Euro weakening by 10%	(225)	(239)	(251)	(299)

Granger causality is applied to components of a stationary vector random process. At the heart of the definition is a well-known postulate that the future cannot affect the past.

The essence of the Granger test is that the variable x is causal for the variable y, that is, under the influence of $x \rightarrow y$ changes of x must precede changes of y, not vice versa. Therefore, under the above conditions, it is necessary that the following actions be performed at the same time: the variable x makes a significant contribution to the forecast of y, while the variable does not significantly contribute to the forecast of the variable x [10]. To determine whether x is the cause of y, determine what proportion of the variance of the current value of variable y can be explained by past values of the variable y itself, and whether adding past values of variable x can increase the proportion of explanatory variance. The variable x is the cause of y if x contributes to the prediction of y. In the regression analysis, the variable x will be the cause of y when the coefficients at logs x are statistically significant, but the most commonly investigated cause and effect relationships are two-sided. In other words, the variable x_t is not a Granger cause for the variable u_t if excluding from the model information about the past values of the variable x_t does not impair the predicted value of u_t when used to construct models of both time series. The quality of the forecast in this case is estimated by the standard error. The scheme of model analysis for the presence and direction of causality is shown in figure 1.

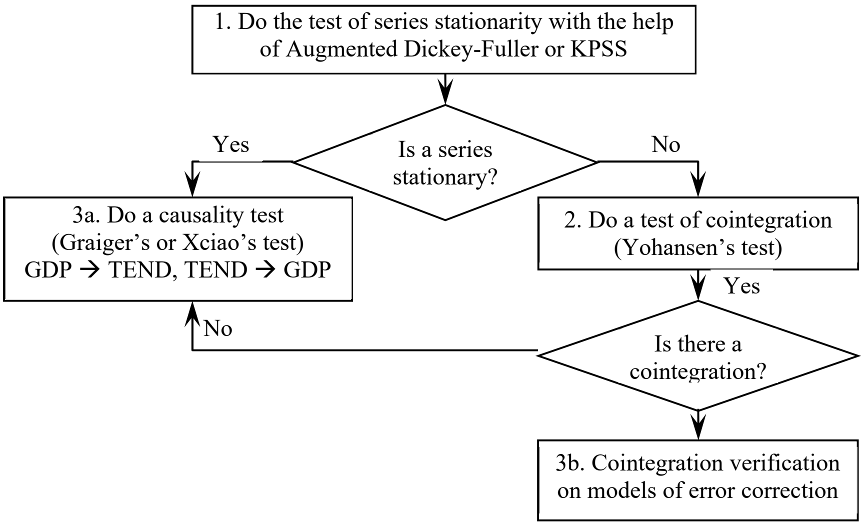


Fig. 1. The scheme of analysis of the model for the presence and direction of causality

To perform this test, three indicators were selected for six NGSUs and Ukrnafa Public Joint-Stock Company (PJSC) for the six months 2016-2019. These include financial risk, operational risk, and contingent and contractual commitments under Chernihivnaftogaz, Poltavanaftogaz, Okhtyrkanaftogaz, Dolynaftogaz, Borislavnaftogaz, Nadvirnaftogaz and Ukrnafa PJSC. Causality testing involves the use of stationary time series. Stationarity is verified in Eviews software, which automatically calculates the required metrics. The functionality of the program proposes to use the Dickey-Fuller and Phillips-Perron test to check the stationarity of a number of selected indicators.

The Dickey-Fuller test is based on the estimation of the parameter $\lambda = \alpha_1 - 1$ of the equation $\Delta Y_t = \lambda Y_{t-1} + \varepsilon_t$, equivalent to the autoregression equation. If the value of the Student's t-statistic for the parameter λ is less than the lower threshold of the DF-statistic, then the null hypothesis $\lambda = 0$ (about the presence of a single root $\alpha_1 = 1$) should be rejected and the alternative about the stationarity of the process Y_t should be accepted [23]. As a result of the Dickey-Fuller test, it was found that even at a significance level of 10%, the hypothesis of stationarity of series should be rejected. To bring the original variables to the stationary form, the transition to the analysis of the second differences of these series was performed. The calculations revealed that the hypothesis of stationarity of a series should be accepted (figure 2).

Null Hypothesis: D(X1,2) has a unit root			
Exogenous: None			
Lag Length: 0 (Automatic - based on SIC, maxlag=1)			
		t-Statistic	Prob.
Augmented Dickey-Fuller test statistic		-3.3939	0.0067
Test critical values:	1% level	-3.1095	
	5% level	-2.0439	
	10% level	-1.5973	

Null Hypothesis: D(X2,2) has a unit root			
Exogenous: None			
Lag Length: 1 (Automatic - based on SIC, maxlag=1)			
		t-Statistic	Prob.
Augmented Dickey-Fuller test statistic		-2.402	0.0319
Test critical values:	1% level	-3.271	
	5% level	-2.082	
	10% level	-1.599	

Null Hypothesis: D(X3,2) has a unit root			
Exogenous: None			
Lag Length: 1 (Automatic - based on SIC, maxlag=1)			
		t-Statistic	Prob.
Augmented Dickey-Fuller test statistic		-4.072	0.0042
Test critical values:	1% level	-3.271	
	5% level	-2.082	
	10% level	-1.599	

Fig. 2. Investigation into the stationarity of the second series differences using the Dickey-Fuller test (series X1 – Financial risk, series X2 – Operational risk, series X3 – Contingent and contractual obligations (Ukrnafta PJSC))

However, there are other tests to check the series for stationarity. Given that the random components of the ADF test can be autocorrelated, have different variances (i.e., heteroskedasticity may be present) and not necessarily normal distributions, compared to the ADF test, the Phillips-Perron test can be used to consider wider classes time series.

Conducting the Phillips-Perron test, which is also present in the Eviews software, shows the same results: the investigated series are non-stationary and the other series differences are stationary (figure 3).

Null Hypothesis: D(X1,2) has a unit root			
Exogenous: Constant			
Bandwidth: 4 (Newey-West automatic) using Bartlett kernel			
		Adj. t-Stat	Prob
Phillips-Perron test statistic		-6.614176	0.0049
Test critical values:	1% level	-5.604618	
	5% level	-3.694851	
	10% level	-2.982813	
Null Hypothesis: D(X2,2) has a unit root			
Exogenous: Constant			
Bandwidth: 4 (Newey-West automatic) using Bartlett kernel			
		Adj. t-Stat	Prob
Phillips-Perron test statistic		-4.209688	0.0315
Test critical values:	1% level	-5.604618	
	5% level	-3.694851	
	10% level	-2.982813	
Null Hypothesis: D(X3,2) has a unit root			
Exogenous: Constant			
Bandwidth: 4 (Newey-West automatic) using Bartlett kernel			
		Adj. t-Stat	Prob
Phillips-Perron test statistic		-4.103371	0.0346
Test critical values:	1% level	-5.604618	
	5% level	-3.694851	
	10% level	-2.982813	

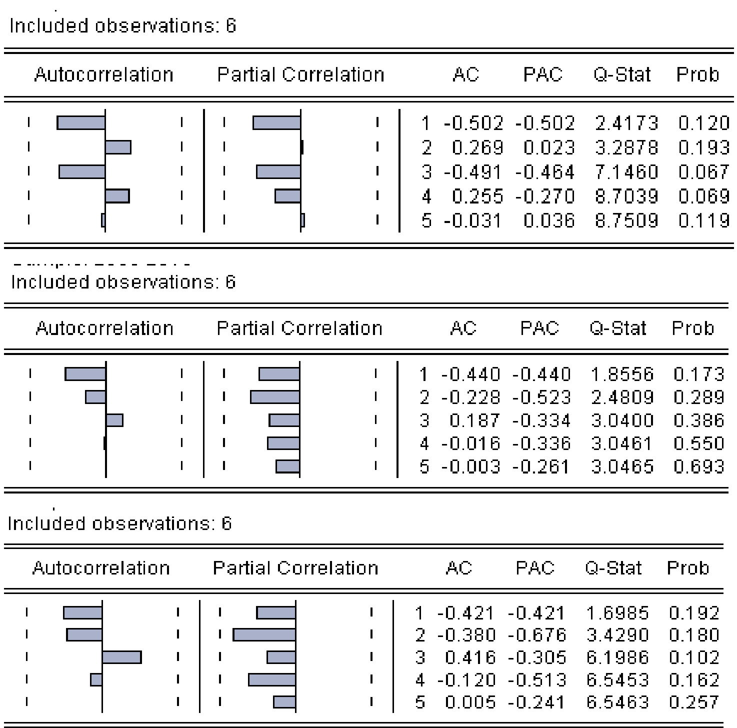
Fig. 3. Investigation into the stationarity of the second series differences using the Phillips-Perron test (series X1 – Financial risk, series X2 – Operational risk, series X3 – Contingent and contractual obligations (Ukrnafta PJSC))

The results obtained by the ADF test can also be verified by visual analysis of the autocorrelogram and partial autocorrelogram (figure 4).

Let's do a Granger causality test. The length of lag p should be chosen from the longest lag, which can still help in predicting. Analysis of the cross correlograms indicates the choice of $p = 2$. In addition, it is confirmed by the well-known rule that the number of lags should not exceed the number of observations divided by 4. Consider the Granger causality for two variables. The model form below is:

$$E_d = \sum_{i=1}^{n} e_{ij} \quad (1)$$

$$E_d = \sum_{i=1}^{n} e_{ij} \quad (2)$$



Included observations: 6

	AC	PAC	Q-Stat	Prob
1	-0.502	-0.502	2.4173	0.120
2	0.269	0.023	3.2878	0.193
3	-0.491	-0.464	7.1460	0.067
4	0.255	-0.270	8.7039	0.069
5	-0.031	0.036	8.7509	0.119

a)

Included observations: 6

	AC	PAC	Q-Stat	Prob
1	-0.440	-0.440	1.8556	0.173
2	-0.228	-0.523	2.4809	0.289
3	0.187	-0.334	3.0400	0.386
4	-0.016	-0.336	3.0461	0.550
5	-0.003	-0.261	3.0465	0.693

b)

Included observations: 6

	AC	PAC	Q-Stat	Prob
1	-0.421	-0.421	1.6985	0.192
2	-0.380	-0.676	3.4290	0.180
3	0.416	-0.305	6.1986	0.102
4	-0.120	-0.513	6.5453	0.162
5	0.005	-0.241	6.5463	0.257

c)

Fig. 4. Self-correlates of the second time series differences: a) Financial risk, b) Operational risk, c) Contingent and contractual obligations (Ukrnafta PJSC)

The absence of a causal relationship from x to y means that when $c_j = 0$ at $j = 1, ..., p$ that is, past values of x do not affect y. The absence of causality from y to x means that $b_j = 0$ at $j = 1, ..., p$.

When the process is stationary, then hypotheses about causality can be tested using F-statistics. The null hypothesis is that one variable is not a Granger cause for another variable.

The results of the test are presented in table 2. Recall that the hypothesis about the causality of this factor is accepted (and the null hypothesis, respectively, is rejected) at a probability of less than 0.05, with a probability greater than 0.05 is accepted null hypothesis.

Table 2. Granger test results

Company	Indicator		
	Financial risk (X1)	**Operational risk (X2)**	**Contingent and contractual obligations (X3)**
Chernihivnaftogaz	The hypothesis is accepted	The hypothesis is accepted	The hypothesis is accepted
Nadvirnaftogaz	The hypothesis is accepted	The hypothesis is rejected (X2→X3)	The hypothesis is accepted
Borislavnaftogaz	The hypothesis is accepted	The hypothesis is accepted	The hypothesis is accepted
Poltavanaftogaz	The hypothesis is accepted	The hypothesis is accepted	The hypothesis is accepted
Dolinanaftogaz	The hypothesis is accepted	The hypothesis is accepted	The hypothesis is rejected (X3→X1)
Okhtyrkanaftogaz	The hypothesis is rejected (X1→X2)	The hypothesis is accepted	The hypothesis is accepted
Ukrnafta	The hypothesis is accepted	The hypothesis is accepted	The hypothesis is accepted

As a result of the causal analysis, it was found that the change in operational risk at Nadvirnanaftogaz is the cause of contingent and contractual obligations, but not vice versa; at Dolynaftogaz a number of contingent and contractual dynamics are the cause of a number of financial risks for Granger; at Okhtyrkanaftogaz financial risk is the cause of operational risk, and the connection is also one-sided. Other hypotheses regarding causality are not accepted.

4 Results

Effective and integrated risk management requires the integration of risk management into the enterprise management process. The market economy creates both opportunities to achieve planned profits, as well as the risk of losses due to adverse changes in the environment of the company and mistakes within the organization.

Many methods and approaches to the risk management process indicate that an important aspect of the company is to look for optimal solutions to existing threats. Decisions on this issue should be taken as a result of the risk management process, in which the placement and awareness of the importance of the risk that threatens the company become a very important starting point.

Risk identification and measurement are important here because they determine the choice of risk control method, which means specific decisions and financial costs. In assessing this, particular attention should be paid to the importance of contingent and contractual obligations, which often go unnoticed and neglected, and omissions of which can be significant.

The obtained results allow us to adjust the policy of activity of oil and gas enterprises for 2 years in advance (lag length) depending on the risks involved and the peculiarities of the socio-economic status of the territories.

The scientific novelty of the study is to formalize the areas of relationships between the risk of oil and gas companies and its elements on the basis of testing by the Granger method. Using the results of the proposed testing allows to determine the direction of causal links between financial risk, which depends on currency risk, that is determined by the need to purchase natural gas from foreign suppliers, operational risk of concentration on revenues from gas transportation and trade payables, as well as contingent and contractual obligations that pose the risk that one party to a financial instrument will cause a financial loss to the other party as a result of a default. This contributes to the definition and coordination of risk management policies to minimize their negative impact on the financial performance of oil and gas companies.

References

1. Abdel-Basset, M., Gunasekaran, M., Mohamed, M., Chilamkurti, N.: A framework for risk assessment, management and evaluation: Economic tool for quantifying risks in supply chain. Future Generation Computer Systems **90**, 489–502 (2019)
2. Amigun, B., Petrie, D., Gorgens, J.: Economic risk assessment of advanced process technologies for bioethanol production in South Africa: Monte Carlo analysis. Renewable Energy **36**(11), 3178–3186 (2011)
3. Andrusiv, U., Kinash, I., Cherchata, A., Polyanska, A., Dzoba, O., Tarasova, T., Lysak, H.: Experience and prospects of innovation development venture capital financing. Management Science Letters **10**(4), 781–788 (2020)
4. Bodnar, T., Gupta, A.K., Vitlinskyi, V., Zabolotskyy, T.: Statistical inference for the beta coefficient. Risks **7**(2), 56 (2019)
5. Daniel, V.E., Florax, R.J.G.M., Rietveld, P.: Flooding risk and housing values: An economic assessment of environmental hazard. Ecological Economics **69**(2), 355–365 (2009)
6. Degtiareva, O.: Analysis of risk and resilience factors in Ukrainian energy sector and integrated assessment of energy security level. Istanbul Journal of Economics and Politics **1**, 73-82 (2015)
7. Derbentsev, V., Matviychuk, A., Soloviev, V.N.: Forecasting of Cryptocurrency Prices Using Machine Learning. In: Pichl, L., Eom, C., Scalas, E., Kaizoji, T. (eds.) Advanced Studies of Financial Technologies and Cryptocurrency Markets, pp. 211-231. Springer, Singapore (2020). doi:10.1007/978-981-15-4498-9_12
8. Derbentsev, V., Semerikov, S., Serdyuk, O., Solovieva, V., Soloviev, V.: Recurrence based entropies for sustainability indices. E3S Web Conferences **166**, 13031 (2020). doi:10.1051/e3sconf/202016613031
9. Dźwigoł, H.: Kontroling w procesie zarządzania współczesnym przedsiębiorstwem. Politechnika Śląska, Gliwice (2011)
10. Fedorova, E.A.: Analysis of the Impact of Financial Integration. Audit and financial analysis 2 (2010)
11. Havrylenko, M., Shiyko, V., Horal, L., Khvostina, I., Yashcheritsyna, N.: Economic and mathematical modeling of industrial enterprise business model financial efficiency estimation. E3S Web Conferences **166** (2020), 13025. doi:10.1051/e3sconf/202016613025
12. Ivchenko, I.Y.: Modeling of economic risks and risk situations. Center for Educational Literature, Kyiv (2007)

13. Khvostina, I., Havadzyn, N., Horal, L., Yurchenko, N.: Emergent Properties Manifestation in the Risk Assessment of Oil and Gas Companies. CEUR Workshop Proceedings **2422**, 157–168 (2019)
14. Khvostina, I., Matviychuk, A., Savina, S., Yatsiuk, O.: Analysis of the Dependence of the Performance of the Ensemble on the Accuracy of its Individual Models. Advances in Economics, Business and Management Research 99 (2019). doi:10.2991/mdsmes-19.2019.70
15. Kiv, A., Semerikov, S., Soloviev, V., Kibalnyk, L., Danylchuk, H., Matviychuk, A.: Experimental Economics and Machine Learning for Prediction of Emergent Economy Dynamics. CEUR Workshop Proceedings **2422**, 1–4 (2019)
16. Kozmenko, O., Oliynyk V.: Statistical model of risk assessment of insurance company's functioning. Investment Management and Financial Innovations 12(2), 189–194 (2015)
17. Kuzmin, O.E, Sytnik, Y.S.: Risks and organizational support of the process of intellectualization of enterprise management systems. Scientific Bulletin of the International Humanities University. Economics and Management Series 23(1), 81–85 (2017)
18. Kuzmin, O.E., Yavorskaya, Zh.B.: Economic evaluation and methods of management risk reduction in enterprise management system. Lviv Polytechnic, Lviv (2014)
19. Lukianova, V.V.: The mechanism of diagnostics of activity of the enterprise considering risk. NAS of Ukraine, Donetsk (2009)
20. Matviychuk, A., Novoseletskyy, O., Vashchaiev, S., Velykoivanenko H., Zubenko, I.: Fractal analysis of the economic sustainability of enterprise. SHS Web of Conferences **65**, 06005 (2019). doi:10.1051/shsconf/20196506005
21. Matviychuk, A., Strelchenko, I., Vashchaiev, S., Velykoivanenko, H.: Simulation of the crisis contagion process between countries with different levels of socio-economic development. CEUR Workshop Proceedings **2393**, 485–496 (2019)
22. Matviychuk, A.: Bankruptcy prediction in transformational economy: discriminant and fuzzy logic approaches. Fuzzy economic review 15(1), 21–38 (2010). doi:10.25102/fer.2010.01.02
23. Menges, G.: Three essays in econometrics - Lectures held in Bombay and Calcutta in September/October 1962. Statistische Hefte **4**(1), 1–37 (1963)
24. Oliinyk, V., Kozmenko O.: Optimization of investment portfolio management. Serbian Journal of Management **14**(2), 373–387 (2019). doi:10.5937/sjm14-16806
25. Petryna, M., Stavnycha, N., Tarayevska, L., Rishchuk, L., Kushlyk, O.: A methodological approach to the evaluation of the effectiveness of innovative projects. E3S Web of Conferences **166**, 13018 (2020)
26. PJSC Ukrnafta Annual Information 2016. https://www.ukrnafta.com/data/Page_FINANCIAL%20RESULTS/ANNUAL_REPORTS/2016.pdf (2017). Accessed 18 Aug 2020
27. PJSC Ukrnafta Annual Information 2017. https://www.ukrnafta.com/data/Ivestor_news/Richna%20informaciya%202017.pdf (2018). Accessed 18 Aug 2020
28. PJSC Ukrnafta Board Report 2018. https://www.ukrnafta.com/data/Investor_docs/Zvit%20Pravlinnya_2018.pdf (2019). Accessed 18 Aug 2020
29. PJSC Ukrnafta Intermediate Information 2019. https://www.ukrnafta.com/data/Investor_docs/Rehularna_promizhna_informatsiya_za_IV_kvartal_2019.pdf (2020). Accessed 18 Aug 2020

30. PJSC Ukrnafta Particular Financial Statements 2017. https://www.ukrnafta.com/data/Ivestor_news/Okrema%20zvitnist%202017.pdf (2018). Accessed 18 Aug 2020
31. Podolchak, N.Y., Kulynyak, I.Y.: The method of complex estimation of risk level for lessor enterprises. Actual Problems of Economics 142(4), 144–157 (2013)
32. Salvati, L., Zitti, M., Ceccarelli, T.: Integrating economic and environmental indicators in the assessment of desertification risk: A case study. Applied Ecology and Environmental Research **6**(1), 129–138 (2008)
33. Shcherbak, O.V.: Risks identification at chemical enterprises and their management. Actual Problems of Economics **119**(5), 145–148 (2011)
34. Soloviev, V., Belinskyi, A., Solovieva, V.: Entropy Analysis of Crisis Phenomena for DJIA Index. CEUR Workshop Proceedings **2393**, 434–449 (2019)
35. Zhou, Q., Mikkelsen, P.S., Halsnæs, K., Arnbjerg-Nielsen, K.: Framework for economic pluvial flood risk assessment considering climate change effects and adaptation benefits. Journal of Hydrology **414-415**, 539–549 (2012)

Complex networks theory and precursors of financial crashes

Vladimir Soloviev[1,2][0000-0002-4945-202X], Victoria Solovieva[3][0000-0002-8090-9569], Anna Tuliakova[4], Alexey Hostryk[5][0000-0001-6143-6797] and Lukáš Pichl[6]

[1] Kryvyi Rih State Pedagogical University, 54 Gagarin Ave., Kryvyi Rih, 50086, Ukraine
[2] The Bohdan Khmelnytsky National University of Cherkasy, 81 Shevchenka Blvd., 18031, Cherkasy, Ukraine
[3] State University of Economics and Technology, 16 Medychna Str., Kryvyi Rih, 50005, Ukraine
[4] Odessa I. I. Mechnikov National University, 2 Dvoryanskaya Str., Odessa, 65082, Ukraine
[5] Odessa National Economic University, 8 Preobrazhenskaya Str., Odessa, 65082, Ukraine
[6] International Christian University, 3-10-2 Osawa, Mitaka-shi, Tokyo, 181-8585, Japan
vnsoloviev2016@gmail.com, vikasolovieva2027@gmail.com, tuliakovaanna@gmail.com, alexeyGostrik@gmail.com, lukas@icu.ac.jp

Abstract. Based on the network paradigm of complexity in the work, a systematic analysis of the dynamics of the largest stock markets in the world and cryptocurrency market has been carried out. According to the algorithms of the visibility graph and recurrence plot, the daily values of stock and crypto indices are converted into a networks and multiplex networks, the spectral and topological properties of which are sensitive to the critical and crisis phenomena of the studied complex systems. This work is the first to investigate the network properties of the crypto index CCI30 and the multiplex network of key cryptocurrencies. It is shown that some of the spectral and topological characteristics can serve as measures of the complexity of the stock and crypto market, and their specific behaviour in the pre-crisis period is used as indicators-precursors of critical phenomena.

Keywords: crypto index, visibility graph, complexity measures of financial crashes.

1 Introduction

The new interdisciplinary study of complex systems, known as the complex networks theory, laid the foundation for a new network paradigm of synergetic [17]. The complex networks theory studies the characteristics of networks, taking into account not only their topology, but also statistical properties, the distribution of weights of individual nodes and edges, the effects of information dissemination, robustness, etc. [1; 6; 10;

20; 21; 22; 31]. Complex networks include electrical, transport, information, social, economic, biological, neural and other networks [1; 6; 10; 20; 18; 22; 24; 31]. The network paradigm has become dominant in the study of complex systems since it allows you to enter new quantitative measures of complexity not existing for the time series [3].

Previously, we introduced various quantitative measures of complexity for individual time series [25; 26; 27; 28; 29]. Significant advantage of the introduced measures is their dynamism, that is, the ability to monitor the time of change in the chosen measure and compare with the corresponding dynamics of the output time series. This allowed us to compare the critical changes in the dynamics of the system, which is described by the time series, with the characteristic changes of concrete measures of complexity. It turned out that quantitative measures of complexity respond to critical changes in the dynamics of a complex system, which allows them to be used in the diagnostic process and prediction of future changes. In [25], we introduced network complexity measures and adapted them to study system dynamics. But networks are rarely isolated. Therefore, it is necessary to take into account the interconnection interaction, which can be realized within the framework of different models [4]. In this paper we will consider it by simulating so-called multiplex networks, the features of which are reduced to a fixed number of nodes in each layer, but they are linked by different bonds.

2 Review

Most complex systems inform their structural and dynamic nature by generating a sequence of certain characteristics known as time series. In recent years, interesting algorithms for the transformation of time series into a network have been developed, which allows to extend the range of known characteristics of time series even to network ones. Recently, several approaches have been proposed to transform time sequences into complex network-like mappings. Three main classes can be distinguished. The first is based on the study of the convexity of successive values of the time series and is called visibility graph (VG) [12]. The second analyzes the mutual approximation of different segments of the time sequence and uses the technique of recurrent analysis [9]. The recurrent diagram reflects the existing repetition of phase trajectories in the form of a binary matrix whose elements are units or zeros, depending on whether they are close (recurrent) with given accuracy or not, the selected points of the phase space of the dynamic system. The recurrence diagram is easily transformed into adjacency matrix, on which the spectral and topological characteristics of the graph are calculated [9]. Finally, if the basis of forming the links of the elements of the graph is to put correlation relations between them, we obtain a correlation graph [9]. To construct and analyze the properties of a correlation graph, we must form adjacency matrix from the correlation matrix. To do this, you need to enter a value which, for the correlation field, will serve as the distance between the correlated agents. So, if the correlation coefficient between the two assets is significant, the distance between them is small, and, starting from a certain critical value, assets can be considered bound on the graph. For an adjacency matrix, this means that they are adjacent to the graph. Otherwise, the assets are not contiguous. In this case, the binding condition of the graph is a prerequisite.

The use of the complexity of recurrent and visibility graph networks to prevent critical and crisis phenomena in stock markets has been considered by us in a recent papers [27; 28]. In this paper, we consider multiplex implementations of these techniques.

The recurrence diagrams for the visualization of phase space recurrences is based on Henri Poincare's idea of the phase space recurrence of dynamical systems. According to Takens's theorem [9], an equivalent phase trajectory that preserves the structure of the original phase trajectory can be recovered from a single observation or time series by the time delay method: $\vec{x}(t) = \left(u_i, u_{i+\tau}, \ldots, u_{i+(m-1)\tau}\right)$, where m – embedding dimension, τ – time delay (real time delay is defined as $\tau \cdot \Delta t$). The recurrence plot shows the existing repetitions in the form of a binary matrix R, where $R_{i,j} = 1$, if $\vec{x}_j$ it is adjacent to the state $\vec{x}_i$, and $R_{i,j} = 0$ otherwise. Neighboring (or recurrent) are states $\vec{x}_j$ that fall into a m-dimensional region with radius ε and center in $\vec{x}_i$. It is clear that parameters m, τ and ε are key when conducting recurrent analysis. The recurrence plot is easily transformed into an adjacency matrix, by which the spectral and topological characteristics of the graph are calculated [28].

The algorithm of the VG is realized as follows. Take a time series $Y(t) = [y_1, y_2, \ldots, y_n]$ of length N. Each point in the time series data can be considered as a vertex in an associative network, and the edge connects two vertices if two corresponding data points can "see" each other from the corresponding point of the time series (fig. 1). Formally, two values y_a of the series (at the time t_a) and y_b (at the time t_b) are connected, if for any other value (y_c, t_c), which is placed between them (that is, $t_a < t_c < t_b$), the condition is satisfied $y_c < y_a + (y_b - y_a)((t_c - t_a)/(t_b - t_a))$. Note that the visibility graph is always connected by definition and also is invariant under affine transformations, due to the mapping method.

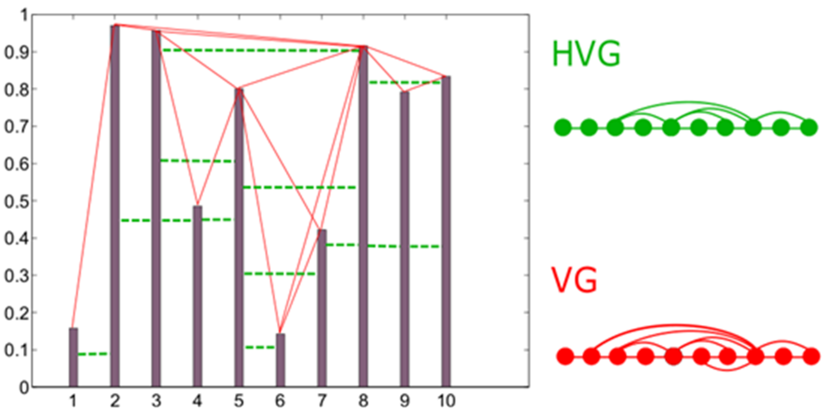


Fig. 1. Illustration of constructing the visibility graph (red lines) and the horizontal visibility graph (green lines) [16].

An alternative (and much simpler) algorithm is the horizontal visibility graph (HVG) [16], in which a connection can be established between two data points a and b, if one can draw a horizontal line in the time series joining them that does not intersect any intermediate data by the following geometrical criterion: y_a, $y_b > y_c$ for all c such that $t_a < t_c < t_b$ (fig. 1).

In multiplex networks, there are two tasks [13]: (1) turn separate time series on the network for each layer; (2) connect the intra-loop networks to each other. The first problem is solved within the framework of the standard algorithms described above. For multiplex networks, the algorithm of the MVG for the three layers is presented in fig. 2.

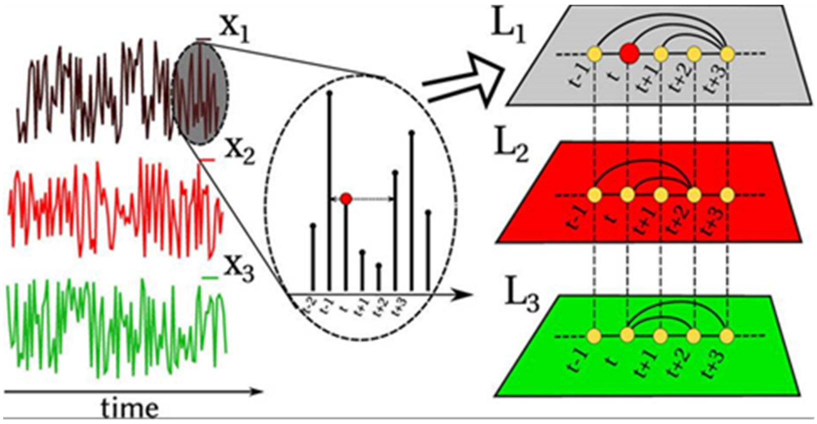


Fig. 2. Scheme for forming bonds between three layers of the multiplex network [13].

The recurrence multiplex network (MCRP) is formed from recurrence plots of individual layers.

2.1 Spectral and topological graph properties

Spectral theory of graphs is based on algebraic invariants of a graph – its spectra. The spectrum of graph G is the set of eigenvalues of a matrix $S_p(G)$ corresponding to a given graph. For an adjacency matrix A of a graph, there exists a characteristic polynomial $|\lambda I - A|$, which is called the characteristic polynomial of a graph $P_G(\lambda)$. The eigenvalues of the matrix A (the zeros of the polynomial $|\lambda I - A|$) and the spectrum of the matrix A (the set of eigenvalues) are called respectively their eigenvalues and the spectrum of graph G. The eigenvalues of the matrix A satisfy the equality $A\bar{x} = \lambda\bar{x}$ ($\bar{x}$ – non-zero vector). Vectors $\bar{x}$ satisfying this equality are called eigenvectors of matrix A (or graph G) corresponding to their eigenvalues.

Another common type of graph spectrum is the spectrum of the Laplace matrix L. The Laplace matrix is used to calculate the tree graphs, as well as to obtain some important spectral characteristics of the graph. In particular, the positive eigenvalues λ_2 is called the index of algebraic connectivity of the graph. This value represents the "force" of the connectivity of the graph component and is used in the analysis of reliability and synchronization of the graph.

Important derivative characteristics are spectral gap, graph energy, spectral moments and spectral radius. The spectral gap is the difference between the largest and the next eigenvalues of the adjacency matrix and characterizes the rate of return of the system to the equilibrium state. The graph energy is the sum of the modules of the eigenvalues of the graph adjacency matrix. The spectral radius is the largest modulus of the eigenvalue of the adjacency matrix. Denote by N_c the value which corresponds to an "average eigenvalue" of the graph adjacency matrix $N_c = -ln\left(N \sum_{i=1}^{N} e^{\lambda_i}\right)$ and is called natural connectivity.

The k-th spectral moment of the adjacency matrix is determined by the expression $m_k(A) = 1/n\sum_{i=1}^{n} \lambda_i^k$, where λ_i is the eigenvalues of the adjacency matrix, n is the vertex of G.

Among the topological measures one of the most important is the node degree k – the number of links attached to this node. For non-directed networks, the node's degree k_i is determined by the sum $k_i = \sum_j a_{ij}$, where a_{ij} – the elements of the adjacency matrix.

To characterize the "linear size" of the network, useful concepts of mean $<l>$ and maximum $l_{\max}$ shortest paths. For a connected network of N nodes, the average path length (ApLen) is equal to $<l>= 2/((n(N-1))\sum_{i>j} l_{ij}$, where l_{ij} – the length of the shortest path between the nodes. The diameter of the connected graph is the maximum possible distance between its two vertices, while the minimum possible is the radius of the graph.

If the average length of the shortest path gives an idea of the whole network and is a global characteristic, the next parameter – the clustering coefficient – is a local value and characterizes a separate node. For a given node m, the clustering coefficient C_m is defined as the ratio of the existing number of links between its closest neighbors to the maximum possible number of such relationships $C_m=2E_m/(k_m(k_m-1))$. Here $k_m(k_m-1)/2$ is the maximum number of links between the closest neighbors. The clustering coefficient of the entire network is defined as the average value C_m of all its nodes. The clustering coefficient shows how many of the nearest neighbors of the given node are also the closest neighbors to each other. He characterizes the tendency to form groups of interconnected nodes – clusters. For real-life networks, the high values of the clustering coefficient are high.

Another feature of the node is the between's. It reflects the role of the node in establishing network connections and shows how many shortest paths pass through this node. Node betweenness σ_m is defined as $\sigma_m = \sum_{i \neq j} B(i, m, j)/B(i, j)$, where $B(i, j)$ – the total number of shortest paths between nodes i and j, $B(i, m, j)$ – the number of shortest paths between i, j those passing through the node m.

One of the main characteristics of the network is the distribution of nodes $P(k)$, which is defined as the probability that the node i has a degree $k_i=k$. For most natural and actual artificial networks there is a power distribution $P(k) \sim 1/k^{\gamma}$, $k\neq 0$, $\gamma > 0$.

Also, important topological characteristics are the vertex eccentricity – the largest distance between m and any other vertex, that is, how far the vertex is far from the other vertices of the graph. The centrality of the vertex measures its relative importance in the graph. At the same time, the farness of a node is defined as the sum of its distances to all other nodes, and its closeness is defined as the backward distance. Thus, the centrality of the node is lower than its total distance to all other nodes.

Another important measure is the link density in the graph, which is defined as the number of links n_e, divided by the expression $n_n(n_n–1)/2$, where n_n is the number of nodes of the graph.

2.2 Multiplex complexity measures

A multilayer/multiplex network is a pair $M=(G, C)$ where $G = \{G_\alpha;\ \alpha \in \{1, \dots, M\}\}$ there is a family of graphs (whether directed or not, weighed or not) $G_\alpha = (X_\alpha, E_\alpha)$, called layers $C = \{E_{\alpha\beta} \subseteq X_\alpha \times X_\beta;\ \alpha, \beta \in \{1,\dots,M\}, \alpha \neq \beta\}$. The latter is a set of links between nodes of different layers G_α and G_β at $\alpha \neq \beta$. Each element E_α is intralayer bonds M in contrast to the elements of each $E_{\alpha\beta}$ $(\alpha \neq \beta)$, called interlayer bonds.

A set of nodes of a layer G_α is denoted $X_\alpha = \{x_1^\alpha, \dots, x_{N_\alpha}^\alpha\}$, and a adjacency matrix as $A^{[\alpha]} = (a_{ij}^\alpha) \in \mathbb{R}^{N_\alpha \times N_\alpha}$, $a_{ij}^\alpha = \begin{cases} 1, (x_i^\alpha, x_j^\alpha) \in E_\alpha, \\ 0 \end{cases}$ for $1 \leq i, j \leq N_\alpha$ and $1 \leq \alpha \leq M$. For an interlayer adjacency matrix we have $A^{[\alpha,\beta]} = (a_{ij}^{\alpha\beta}) \in \mathbb{R}^{N_\alpha \times N_\alpha}$, where $a_{ij}^{\alpha\beta} = \begin{cases} 1, (x_i^\alpha, x_j^\beta) \in E_{\alpha\beta}, \\ 0 \end{cases}$.

A multiplex network is a partial interlayer case and contains a fixed number of nodes connected by different types of links. Multiplex networks are characterized by correlations of different nature [27], which enable the introduction of additional multiplexes.

Let's evaluate the quantitative overlap between the various layers. The average edge overlap obviously equal $\omega = \sum_i \sum_{j>i} \sum_\alpha a_{ij}^{[\alpha]} / M \sum_i \sum_{j>i} (1 - \delta_{0,\sum a_{ij}^{[\alpha]}})$ and determines the number of layers in which this bond is present. Its value lies on the interval $[1/M, 1]$ and equals $1/M$ if the connection (i, j) exists only in one layer, that is, if there is a layer α such that $a_{ij}^{[\alpha]} = 1, a_{ij}^{[\beta]} = 0\ \forall\, \beta \neq \alpha$. If all layers are identical, then $\omega = 1$. Consequently, this measure can serve as a measure of the coherence of the output time series: high values ω indicate a noticeable correlation in the structure of time series.

The total overlap $O^{\alpha\beta}$ between the two layers α and β is defined as the total number of bonds that are shared between the layers α and $\beta: O^{\alpha\beta} = \sum a_{ij}^\alpha a_{ij}^\beta$, where $\alpha \neq \beta$.

For a multiplex network, the vertex degree k is already a vector $k_i = (k_i^{[1]}, \dots, k_i^{[M]})$, where $k_i^{[\alpha]}$ is the degree of the node in the layer, that is, while the elements of the matrix of adjacency for the layer. Specificity of the vector character of the degree of the peak in multiplex networks allows for the introduction of additional interlayer characteristics. One of these is the overlap of the node's degree $o_i = \sum_{\alpha=1}^M k_i^{[\alpha]}$.

The next measure quantitatively describes the interlayer correlations between the degrees of the selected node in two different layers. If, chosen from M the layers of the pair (α, β) characterized by the distribution of degrees $P(k^{[\alpha]}), P(k^{[\beta]})$, the so-called interlayer mutual information is determined by the formula $I_{\alpha,\beta} = \sum\sum P(k^{[\alpha]}, k^{[\beta]})\, log\big((P(k^{[\alpha]}, k^{[\beta]}))/(P(k^{[\alpha]})P(k^{[\beta]}))\big)$, where $P(k^{[\alpha]}, k^{[\beta]})$ is the probability of finding a node degree $k^{[\alpha]}$ in a layer α and degree $k^{[\beta]}$ in a layer β. The higher the $I_{\alpha,\beta}$ value, the more correlated are the distributions of the levels of the two layers, and, consequently, the structure of the time series associated with them. We also find the mean value $I_{\alpha,\beta}$ for all possible pairs of layers – the scalar value $< I_{\alpha,\beta} >$ that quantifies the information flow in the system.

The quantity that quantitatively describes the distribution of the node degree i between different layers is the entropy of the multiplexed degree $S_i = -\sum_{\alpha=1}^{M} k_i^{[\alpha]}/o_i\, ln(k_i^{[\alpha]}/o_i)$. Entropy is zero if all nodes are in the same layer and vice versa, has the maximum value when they are evenly distributed between different layers. That is, the higher the value S_i, the even more links evenly the nodes' connections are distributed between the layers.

A similar magnitude is the multiplex participation coefficient $P_i = M/(M-1)\left[1 - \sum_{\alpha=1}^{M}\left(k_i^{[\alpha]}/o_i\right)^2\right]$. P_i takes values on the interval [0, 1] and determines that homogeneous links of node i are distributed among M the layers. If all links of the node i lie in one layer, $P_i = 0$ and $P_i = 1$ if the node has a precisely defined number of links in each of the M layers. Consequently, the larger the coefficient P_i is, the more evenly distributed the participation of the node in the multiplex.

Obviously, the magnitudes S_i and P_i are very similar. We will show that some of these spectral and topological measures serve as measures of complexity of the system, and the dynamics of their changes allows us to build predictors of crisis situations on financial markets.

2.3 Experimental results and their discussion

In our recent work [26] we investigated the multiplex properties of the most capitalized stock indices for the period from 01/01/1983 to 10/01/2019. Here we will add an analysis of their complex network properties, while also preserving the analysis of multiplex properties for a whiter period of time, but including the COVID-19 crisis [23]: 02/01/2004 – 17/08/2020 [34]. In addition, for the first time, we are examining the network complex properties of the CCI30 crypto index and comparing them with the properties of the multiplex network of the 3 most capitalized cryptocurrencies included in the CCI30 index basket. For the crypto market, the period under consideration corresponds to the time the index was observed: from 07/08/2015 to 17/08/2020 [7]. The time series of daily values of stock market indices were selected as databases, which contained significant changes in the indices, and were identified as crisis phenomena [15]. Daily values of cryptocurrency prices for a period comparable to the CCI30 index were borrowed from [33]. Among the set of stock indexes are the following: SP (S&P500) – USA; DAX (DAX PERFORMANCE-INDEX) – Germany; N225 (Nikkei 225) – Japan; HSI (HANG SENG INDEX) – China; GSPTSE (S&P/TSX Composite index) – Canada.

Up to three selected cryptocurrencies include the following: Bitcoin – BTC; Ethereum – ETH; Litecoin – LTC.

Their comparative dynamics on a relative scale is shown in fig. 3.

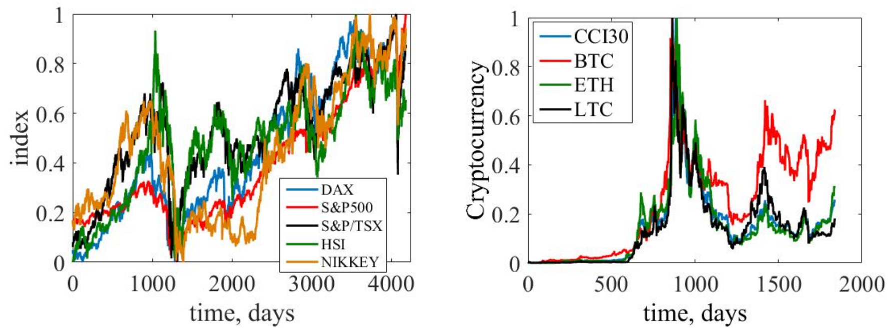


Fig. 3. The dynamics of the daily values of the selected stock market indices (a) and cryptocurrencies and CCI30 (b).

Despite a slight difference in the time series dynamics, the main trends (both rising and falling – actually crises) are observed for all time series.

Crypto index CCI30 refers to a number of emerging crypto indices [32] similar to many stock market indices. CCI30 is an index that is designed to objectively measure the overall growth, daily and long-term movement of the blockchain sector. This allows you to track 30 cryptocurrencies by market cap. This index serves as a tool for passive investors to participate in this asset class and an industry benchmark for investment managers. The main characteristics of the CCI30 index are: 1) versatility; 2) reproducibility; 3) transparent; 4) provides comprehensive coverage of the entire sector; 5) represents the best possible risk-adjusted performance profile. The CCI30 was launched on Jan 1st, 2017. Its starting value is arbitrarily set at 100 on Jan 1st, 2015. Members are automatically selected from the top 30 cryptocurrencies by adjusted market cap and included in the index. The minimum number of constituents required to obtain statistical significance was set at 30. Because using more components will lead to higher fees without significantly improving performance, and using fewer than thirty components will run the risk of reduced performance, lack of diversification, decreased statistical significance, and wasted opportunities to pick the next leader. To date, having hit the top 30 cryptocurrencies, CCi30 captures approximately 90% of the cryptocurrency market capitalization. With this scope, the index statistically represents the entire cryptocurrency market with a confidence level of 99% and a confidence interval of 1.11. The margin of error of the index value as an indicator of the market is just 1.11%. To calculate the weights for each cryptocurrency, you first need to calculate the adjusted market cap. Market capitalization is not calculated as some instantaneous number – the volatility in the cryptocurrency market is such that it too destabilizes the composition of the index. Instead, the CCI30 uses an exponentially weighted moving average of the market cap. Weighted average market cap helps smooth out volatility to give the most accurate portrait of market cap at any given moment.

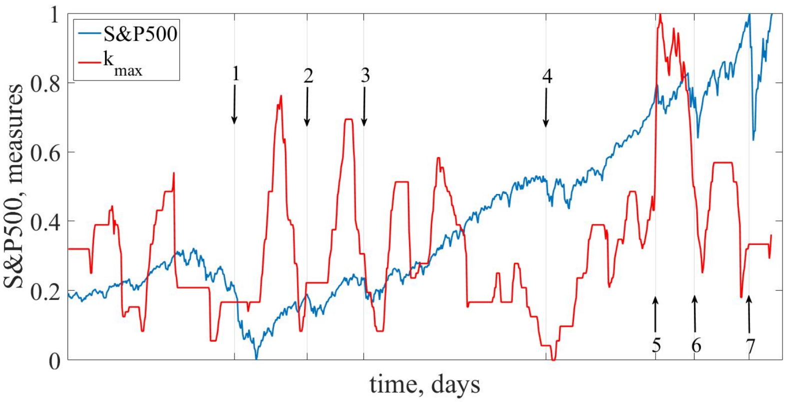
The weight of each constituent cryptocurrency is measured by the square root of its adjusted market capitalization. The square root function was chosen as a hybrid that most accurately weighs the constituents based on the current cryptocurrency market conditions. A simple weighted market cap index would be dominated by the two leading cryptocurrencies, while a slower declining weight, or in a pinch, equal weighting, would give too much weight to the tiny illiquid cryptocurrencies at the bottom of the range. In order to accurately capture market movements, restrictions or minimum values for the weight of cryptocurrencies are not applied. The index is a better investment vehicle than Bitcoin itself, and a much safer approach than trying to pick individual coins. Investing in an index allows you to profit from the unpredictable growth of some cryptocurrencies, while limiting losses associated with the fall of others.

The spectral and topological properties of both stock market indices and the CCI30 index were calculated. The calculations were carried out using the moving window algorithm. In the case of stock indices, the window size was chosen equal to 250 days, and the window offset was equal to 5 days. In the case of a more volatile crypto market, the window size was chosen equal to 100 days.

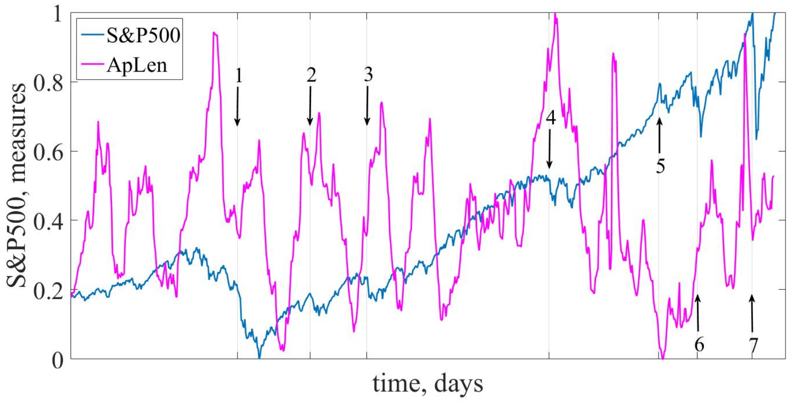
Figure 4, using only two measures as an example, shows their typical dynamics. It follows from the figure 4 that in the pre-crisis period the system has a higher complexity. In this case, node degree takes maximum values, while the average path length, on the contrary, is minimized. After the crisis, the opposite picture is observed.

Calculations of spectral and topological measures by methods of MVG, MCRP were carried out in the following way. For the selected window, the corresponding graphs were constructed and the spectral, topological and multiplex properties were calculated. Next, the window was shifted step by step, for example, one week (5 trading days) and the procedure repeated until the time series were exhausted. Knowing the time of the onset of the crisis and comparing the time series with the dynamics of a certain indicator, it is possible to investigate its dependence on certain the characteristic changes in the stock market: pre-crisis, crisis and post-crisis periods.

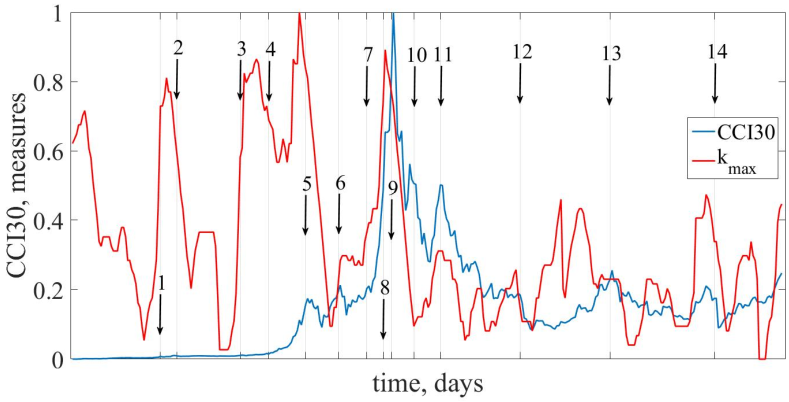
The results of calculations for revived time series of graphs are shown in figures 5-8.



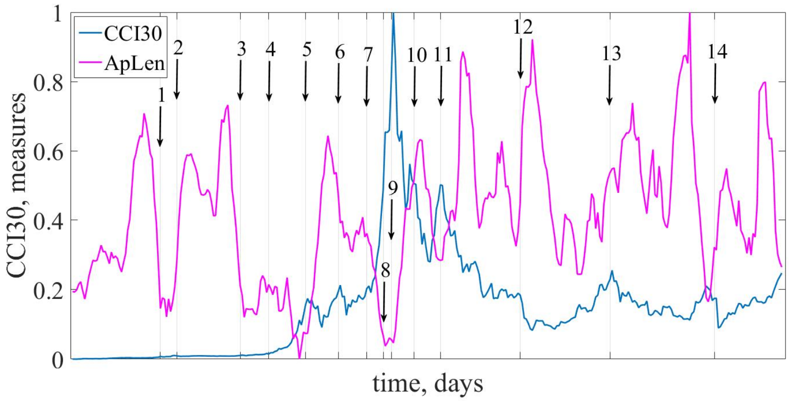
a)



b)



c)



d)

Fig. 4. Window dynamics of spectral $k_{\max}$ and topological ApLen complexity measures for the S&P500 and CCI30 indices.

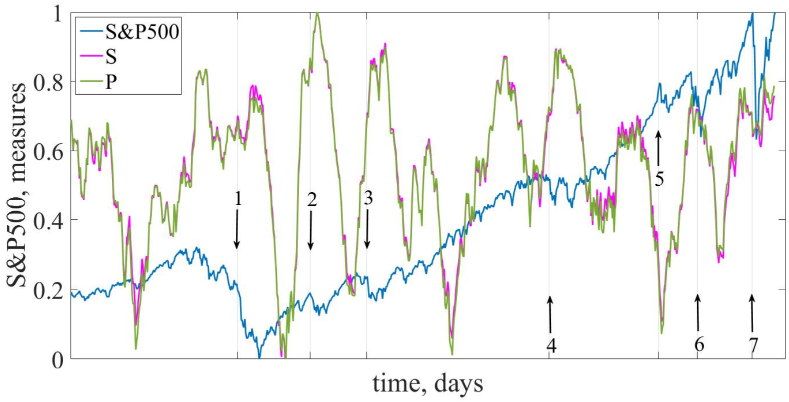


Fig. 5. The entropy of the multiplexed degree and the multiplex participation coefficient calculated for a sliding window (w) of 250 days in increments (Δt) of 5 days. Moments of major stock crises are indicated.

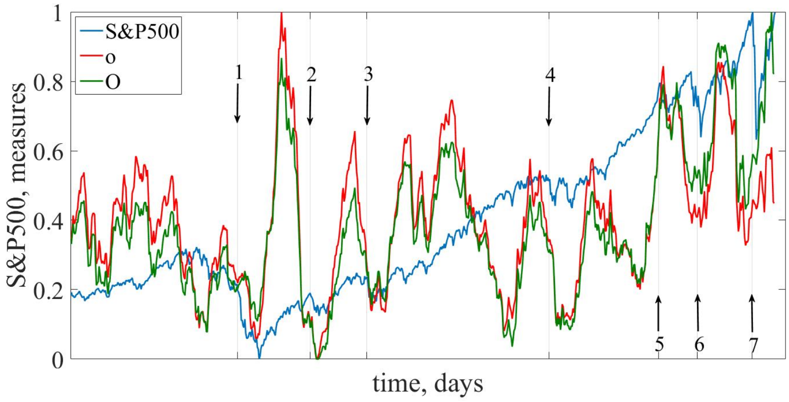


Fig. 6. Nodes degree overlap and the total overlap calculated for a sliding window (w) of 250 days in increments (Δt) of 5 days.

Figures 5-6 show that the pair of parameters S, P in figure 5 is antisymmetric to the pair parameters o, O of figure 6. However, all of them in a characteristic way (decreasing or increasing before the crisis) signal about its possible onset.

Similar calculations were carried out for the CCI30 index (figures 7-8). As in the case of a multiplex visibility graph, multiplex recurrence measures are also informative indicators-predictors of crisis phenomena.

Parameters such as the width of the window w and the step Δt of its displacement along the time series are important. When w is small, the degree of complexity fluctuates noticeably, reacting not only to crises, but also to more or less noticeable fluctuations of the index. On the contrary, with too much window width there is a noticeable smoothing of the appropriate measure and if two crises are at a distance that is smaller w, the indicators of both crises are averaged and less informative. If you choose an oversized parameter Δt, you might miss the actual crisis that distorts the indicator.

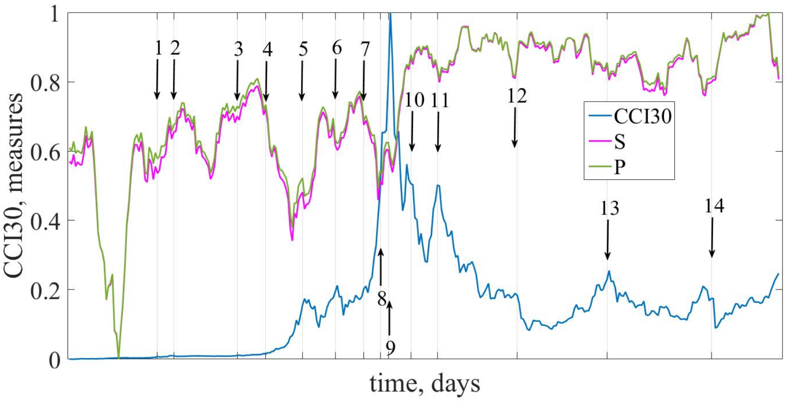


Fig. 7. MVG complexity measures: the entropy of the multiplexed degree and the multiplex participation coefficient.

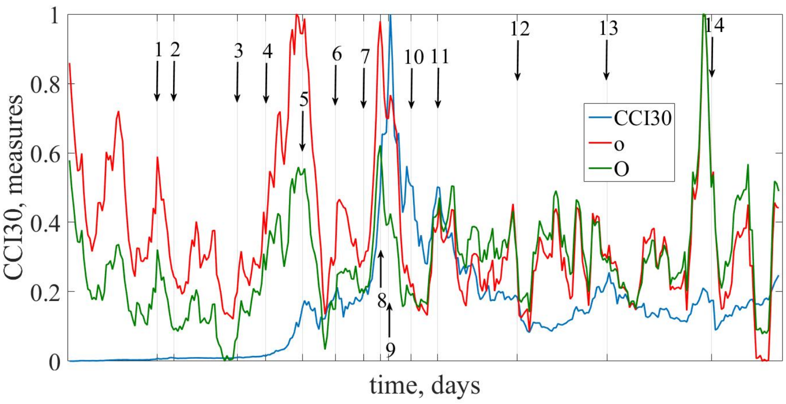


Fig. 8. MVG complexity measures: nodes degree overlaps and the total overlap.

Multiplex networks are actively used to simulate complex networks of different nature: from financial (banks [2], stock market [19], guarantee market [14]) to social [30]. Particular attention should be paid to the work [19], in which the above multiplex measures are analyzed for the subject of correlations with known stock markets crises. Yet there is no systematic analysis of network and multiplex measures and the construction of indicators-predictors of the crisis phenomena in the stock market. Therefore, our studies to some extent fill this gap.

3 Conclusions

We have demonstrated the possibility of studying complex socio-economic systems as part of a network paradigm of complexity. A time series can be represented in an equivalent way – a complex/multiplex network, which has a wide range of characteristics; both spectral and topological, and multiplexed. Examples of known financial crises have shown that most of the network measures can serve as indicators-precursors of crisis phenomena and can be used for possible early prevention of unwanted crises in the financial markets. They are an extension of the already proposed by us and "working" indicators, which use other measures of complexity.

It should be noted that the proposed indicators-precursors do not solve the more general problem of forecasting future values or trends of the stock market. In this way, it is possible to use new approaches or alternative methods based on algorithms of (deep) machine learning [5; 8; 11].

References

1. Avalos-Gaytán, V., Almendral, J.A., Leyva, I., Battiston, F., Nicosia, V., Latora, V., Boccaletti, S.: Emergent explosive synchronization in adaptive complex networks. Physical Review E **97**(4), 042301 (2019)
2. Bargigli, L., di Iasio, G., Infante, L., Lillo, F., Pierobon, F.: The multiplex structure of interbank networks. Quantitative Finance **15**(4), 673–691 (2015)
3. Bianconi, G.: Interdisciplinary and physics challenges in network theory. EPL **111**(5), 56001 (2015)
4. Boccaletti, S., Bianconi, G., Criado, R., del Genio, C.I., Gómez-Gardeñes, J., Romance, M., Sendiña-Nadal, I., Wang, Z., Zanin, M.: The structure and dynamics of multilayer networks. Physics Reports **544**(1), 1–122 (2014)
5. Chatzis, S.P., Siakoulis, V., Petropoulos, A., Stavroulakis, E., Vlachogiannakis, N.: Forecasting stock market crisis events using deep and statistical machine learning techniques. Expert Systems with Applications **112**, 353–371 (2018). doi: 10.1016/j.eswa.2018.06.032
6. Cheng, F., Kovács, I.A., Barabási, A.L.: Network-based prediction of drug combinations. Nature Communications **10**(1), 1197 (2019)
7. Crypto Currency Index CCi30. http://cci30.com (2020). Accessed 17 Aug 2020
8. Derbentsev, V., Matviychuk, A., Soloviev, V.N.: Forecasting of Cryptocurrency Prices Using Machine Learning. In: Pichl, L., Eom, C., Scalas, E., Kaizoji, T. (eds.) Advanced Studies of Financial Technologies and Cryptocurrency Markets, pp. 211–231. Springer, Singapore (2020). doi:10.1007/978-981-15-4498-9_12
9. Donner, R.V., Small, M., Donges, J.F., Marwan, N., Zou, Y., Xiang, R., Kurths, J.: Recurrence-based time series analysis by means of complex network methods. International Journal of Bifurcation and Chaos **21**(4), 1019–1046 (2011)
10. Fortunato, S., Bergstrom, C.T., Börner, K., Evans, J.A., Helbing, D., Milojević, S., Petersen, A.M., Radicchi, F., Sinatra, R., Uzzi, B., Vespignani, A., Waltman, L., Wang, D., Barabási, A.-L.: Science of science. Science **359**(6379), eaao0185 (2018)
11. Kiv, A., Semerikov, S., Soloviev, V., Kibalnyk, L., Danylchuk, H., Matviychuk, A.: Experimental Economics and Machine Learning for Prediction of Emergent Economy Dynamics. CEUR Workshop Proceedings **2422**, 1–4 (2019)

12. Lacasa, L., Luque, B., Ballesteros, F., Luque, J., Nuño, J.C.: From time series to complex networks: The visibility graph. Proceedings of the National Academy of Sciences of the United States of America **105**(13), 4972–4975 (2008)
13. Lacasa, L., Nicosia, V., Latora, V.: Network structure of multivariate time series. Scientific Reports **5**, 15508 (2015) doi:10.1038/srep15508
14. Li, S., Wen, S.: Multiplex Networks of the Guarantee Market: Evidence from China. Complexity **2017**, 9781890 (2017). doi:10.1155/2017/9781890
15. List of stock market crashes and bear markets. https://en.wikipedia.org/wiki/List_of_stock_market_crashes_and_bear_markets (2020). Accessed 17 Aug 2020
16. Luque, B., Lacasa, L., Ballesteros, F., Luque, J.: Horizontal visibility graphs: Exact results for random time series. Physical Review E **80**(4), 046103 (2009). doi:10.1103/PhysRevE.80.046103
17. Malinetskii, G.G., Akhromeeva, T.S.: Self Organization in Complex Systems and New Problems in the Theory of Measurement. Measurement Techniques **59**(6), 577–583 (2016)
18. Markova, O., Semerikov, S., Popel, M.: CoCalc as a learning tool for neural network simulation in the special course "Foundations of Mathematic Informatics". CEUR Workshop Proceedings **2104**, 338–403 (2018)
19. Musmeci, N., Nicosia, V., Aste, T., Di Matteo, T., Latora, V.: The Multiplex Dependency Structure of Financial Markets. Complexity **2017**, 9586064 (2017). doi:10.1155/2017/9586064
20. Newman, M.E.J., Barabási, A.L., Watts, D.: The Structure and Dynamics of Networks. Princeton University Press, Princeton (2006)
21. Prigogine, I.R.: Networks society. Sotsiologicheskie Issledovaniya (1), 24–27 (2008)
22. Riolo, M.A., Newman, M.E.J.: Consistency of community structure in complex networks. Physical Review E **101**(5), 052306 (2020)
23. Semerikov, S., Chukharev, S., Sakhno, S., Striuk, A., Osadchyi, V., Solovieva, V., Vakaliuk, T., Nechypurenko, P., Bondarenko, O., Danylchuk, H.: Our sustainable coronavirus future. E3S Web of Conferences **166**, 00001 (2020). doi:10.1051/e3sconf/202016600001
24. Semerikov, S.O., Teplytskyi, I.O., Yechkalo, Yu.V., Markova, O.M., Soloviev, V.N., Kiv, A.E.: Computer Simulation of Neural Networks Using Spreadsheets: Dr. Anderson, Welcome Back. CEUR Workshop Proceedings **2393**, 833–848 (2019)
25. Soloviev, V., Belinskij, A.: Methods of nonlinear dynamics and the construction of cryptocurrency crisis phenomena precursors. CEUR Workshop Proceedings **2104**, 116–127 (2018)
26. Soloviev, V., Solovieva, V., Tuliakova, A., Ivanova, M.: Construction of crisis precursors in multiplex networks. Advances in Economics, Business and Management Research **99**, 361–366 (2019) doi:10.2991/mdsmes-19.2019.68
27. Soloviev, V., Solovieva, V., Tuliakova, A.: Visibility graphs and precursors of stock crashes. Neuro-Fuzzy Technologies of Modeling in Economy 8, 3–29 (2019). doi:10.33111/nfmte.2019.003
28. Soloviev, V., Tuliakova, A.: Graphodinamical Research Methods for Complexity of Modern Stock Markets. Neuro-Fuzzy Technologies of Modeling in Economy 5, 152–179, (2016)
29. Soloviev, V.N., Belinskiy, A.: Complex Systems Theory and Crashes of Cryptocurrency Market. Communications in Computer and Information Science **1007**, 276–297 (2019)
30. Stephen, C.: Dynamic Phase and Group Detection in Pedestrian Crowd Data Using Multiplex Visibility Graphs. Procedia Computer Science **53**, 410–419 (2015)

31. Vespignani, A.: Twenty years of network science. Nature **558**(7711), 528–529 (2018)
32. Xie, A.: The Ultimate Guide on Cryptocurrency Index Funds. https://www.hodlbot.io/blog/ultimate-guide-on-cryptocurrency-indices (2019). Accessed 25 Oct 2019
33. Yahoo Finance: All Cryptocurrencies Screener. https://finance.yahoo.com/cryptocurrencies (2020). Accessed 17 Aug 2020
34. Yahoo Finance: Stock Market Live, Quotes, Business & Finance News. https://finance.yahoo.com (2020). Accessed 17 Aug 2020

Modeling the optimal management of the distribution of profits of an oil and gas company taking into account risks

Inesa Khvostina[1[0000-0001-5915-749X]], Viktor Oliinyk[2[0000-0001-6251-3846]], Valerii Yatsenko[2[0000-0003-2316-3817]], Liliia Mykhailyshyn[3[0000-0001-7944-5317]] and Uliana Berezhnytska[1[0000-0001-5354-8502]]

[1] Ivano-Frankivsk National Technical University of Oil and Gas, 15 Karpatska Str., Ivano-Frankivsk, 76019, Ukraine
inesa.hvostina@gmail.com, bi_if@ukr.net
[2] Sumy State University, 2 Rymskogo-Korsakova Str., Sumy, 40000, Ukraine
oliynyk.viktor@gmail.com, v.yatsenko@uabs.sumdu.edu.ua
[3] Vasyl Stefanyk Precarpathian National University, 57 Shevchenka Str., Ivano-Frankivsk, 76000, Ukraine
lilya-m@ukr.net

Abstract. This article discusses the optimal management of the distribution of the net income of an oil and gas company, taking into account risks. The utility function for the investigated enterprise acts as an optimality criterion. The control parameter is the distribution of the shares of net income for its optimal distribution in the selected areas. As a numerical implementation of the proposed algorithm, the activity of a catch oil and gas enterprise in the period 2018-2022 is considered. The optimal distribution of the received net income is given taking into account the discount rate and deductions to the State budget of Ukraine. The proposed algorithm can be used for optimal management of the company's financial activities.

Keywords: modeling, optimization, management, profit.

1 Introduction

An existing enterprise of the oil and gas complex in modern conditions should pay special attention to risk management. The main goal of the company is to maximize profit. Achievement of this task will be possible only if all risks affecting the operation of the enterprise are considered. The authors [10; 46] show that the SARS-CoV-2 coronavirus that causes the COVID-19 leads to uncertainty in various areas of economics, finance, risk management, social development, etc. They review scientific materials, according to the study the impact of COVID-19 on global health security, business risk management and other areas.

Risk management needs to be coordinated across the organization. All types of risk, both internal and external, should be managed in a comprehensive manner. One of the approaches to reduce the risk of an enterprise is to predict the occurrence of a risk event and take appropriate measures to reduce it. Here the problem arises in the allocation of appropriate funds to prevent a risk situation. This study allows you to obtain the optimal distribution of cash flows required to reduce possible risky transactions. The total amount of necessary funds is taken as a part of the profit received, i.e. the optimal distribution of the company's profits also takes into account measures to reduce the degree of risk at the enterprise. The ability to optimally manage risky operations allows top management to obtain financial stability, competitiveness and stable profitability of an oil and gas company.

2 Literature review

Socio-economic instability, which is manifested at the present time, suggests that enterprises should devote the most important role to risk management in order to ensure a stable financial and economic condition of the enterprise. Ernst & Young reports the most important types of risks that occur in the activities of an oil and gas company [37; 48]. In [45], an analysis of the activities of oil and gas companies from the Fortune 500 list is carried out: ExxonMobil, Chevron, ConocoPhillips, Baker Hughes, Valero Energy, and Frontier Oil Corporation. The authors deeply analyze the types of risks that exist in the activities of these companies, including those associated with an increase in production capacity. This analysis allowed to diversify the risk portfolio and improve the financial stability of the company.

Many economists pay special attention to the problem of risk management and in their works reflect the basic concepts of the methodology and practice of risk management. To assess the external risks of foreign economic activity of an enterprise in [31], it is proposed to apply a three-dimensional model for assessing and identifying risks, as well as to use a tabular description of the enterprise's activities. The authors in [40] presented a scientific approach for assessing the external economic security of an enterprise. The main mechanisms for managing the anticipated risks and the stages of the implementation of the task are proposed. This approach allows a comprehensive approach to assessing the external economic security of an enterprise.

In [16], the authors consider synergetic models in the management of oil and gas companies. The types of models for describing and predicting oil and gas development are analyzed – exponential, logistic, allometric, Kohlrausch, hyperbolic are analyze. It is proved that synergetic models more fully and adequately describe the processes that take place in the integrated systems of strategic management of oil companies.

The technology of teaching risk management using a business game is considered in [6]. It is believed that in order to improve the qualifications of employees of the enterprise, it is necessary to use interactive methods of training, which will make it possible to further reduce operational and other types of risks. In [49], the authors explore the possibility of applying the theory of dynamic knowledge creation for training corporate employees. To obtain the most productive result, it is proposed to combine the relationship between dynamic knowledge creation and the learning process at the enterprise. This approach leads to a decrease in the operational risk of the enterprise.

The authors of [7] propose a methodology for considering alternative strategies for controlling business processes associated with risk reduction. The characteristics of the methods of risk research, as well as their level of practical use are given. In [5], a discrete dynamic problem of managing the innovation process at an enterprise is considered, taking into account risks. To manage technological innovations at the enterprise, a phase vector of three components is built. The constructed model makes it possible to predict data when constructing optimal processes. The innovation process at the enterprise, described by the discrete dynamic model [3], is investigated. The proposed algorithm assumes the management of parameters of the influence of the activities of the enterprise, as well as the consideration of various risks. The model of operational management of the enterprise activity based on the task of adaptive management has been developed. Optimal management of investment flows of an enterprise is considered in [35].

For enterprise management, an algorithm for modeling a multi-criteria function was proposed [27]. A system of indicators has been built that allows evaluating the activities of an enterprise based on an integrated index. This indicator is based on a partial utility function. The proposed methodological approach takes into account the main characteristics of the enterprise's activities with the possibility of their classification and systematization. In [14], the issues of the emergence of diversification risks at the enterprise are considered. Analysis of factors is carried out. which affect these risks in the economic activities of the enterprise. The systematization of the causes of the emergence of risks and their possible consequences, depending on the strategy of the enterprise is carried out.

In [25], the political risk to security in an oil and gas company is considered. The authors show that there are risks at different levels: transnational, national and the level of human security. Investing in an oil and gas company depends on the risk management strategy. Due to the increased frequency of terrorist attacks in oil and gas companies, it is proposed to pay more attention to the risk of political security and apply an integrated approach to risk management.

In [20], an approach to risk assessment is proposed, taking into account the properties of emergence and re-emergence in the process of risk assessment to prevent and predict the risks of the activities of oil and gas enterprises. Taxonomic procedures and methods of factor analysis were identified, taking into account the properties of emergence, which will allow for a comprehensive assessment of risks and create a foundation for their assessment and forecast of their dynamics.

One of the methods of risk prevention is to eliminate the source of risk through specific measures. A narrow approach is to prevent risk through insurance premiums. The insurer is the initiator of this approach. Insurance companies create an optimization portfolio of financial instruments taking into account risk minimization [21; 36].

In [28], an approach was developed to assess the index of economic stability of an enterprise, based on an integral indicator. For the study, the authors used fractal analysis in combination with elements of the theory of a fuzzy set. The proposed method makes it possible to assess the effective activity of an enterprise in conditions of incomplete information. In [9], the authors suggest paying special attention to risk management issues. This is due to the fact that business decisions are always made in the presence of certain types of risk and optimal risk management leads to their reduction. This is especially true for long-term investment in an oil and gas company, since the amount of investment may exceed US $ 1 billion.

The authors of [11] use Vector Autoregressive (VAR) and Vector Error Correction Models (VECM) to study the impact of a jump in oil prices on stock returns in 12 European countries that import oil. It was shown that the real profitability of stocks depends on the reasons for the jump in oil prices. The performance of most European stock markets is negative in relation to the change in oil prices and is mainly determined by oil supply shocks.

The amount of products produced in the oil and gas complex should be optimal from the point of view of its implementation. If this condition is violated, an imbalance occurs in the country's energy system. In [15], the authors used an improved methodological approach to assess the energy independence of the country and regions. In studies, to analyze the energy dependence of some fear of the world, the coefficient of covering consumption by the production of fuel and energy resources was used. On the basis of the integral indicator of energy independence, a comparative analysis of the regions by the level of energy was carried out. The authors in [8] investigated the volatility of the stock market and the oil market. They considered volatilities of two types: implied and realized volatilities. It is shown that the relationship between volatility is present only for implied volatility.

An in-depth analysis of the financial risk of manufacturing companies is presented in [17]. It is shown that Chinese listed companies have gone bankrupt due to market competition. This happened due to the crisis of financial risks, which becomes more and more complex with the development of society. The authors propose a new concept of financial risk in a manufacturing enterprise, which involves risk management mechanisms.

The work [13] is devoted to a comparative analysis of the complexity of traditional stock indices and social responsibility indices using the example of Dow Jones Sustainability Indices and Dow Jones Industrial Average. A comprehensive assessment of complexity reveals the nature of the effectiveness of social responsibility indicators and opens up new opportunities for investor risk management.

In [18], the risks associated with the redistribution of rental relations between the state, the region and the mining company are considered. The identified risks were grouped using the expert method and their importance was assessed. The results obtained suggest the optimization of the use of natural resources, taking into account the risk minimization strategy. The work [47] considers the counterparty risk associated with its default. The authors propose a model for quantifying credit adjustment without the assumption of independence between default and exposure. An analysis of the activities of corporations to ensure their competitiveness is presented in [24]. In modern conditions, with the advent of new information and innovative technologies, both the level of competition in the financial market and the transformation of risks are increasing. The development and practical implementation of innovative projects should be integrated into a dynamic competitive business environment and rely on advanced information technology.

The authors in their work [26] analyzed the impact of enterprise risk management on the financial performance of 12 companies. It has been shown that the best indicators of profitability, liquidity and productivity improvement of a company occur when there is oversight of the enterprise risk management department by the company's Board of Directors. Depending on the size and complexity of the company, a more complex enterprise risk management framework needs to be employed. Particular attention is paid to the relationship between enterprise risk management and the audit committee, senior management and the chief risk officer [30]. To study the mutual influence, a non-experimental correlation approach was used. Based on correlation and regression analyzes, it is shown that the effectiveness of enterprise risk management depends significantly on the opinion of the risk director, the presence of an audit committee and the support of senior management. In [22; 23], attention was paid to modeling the assessment of the tax consciousness of tax subjects under conditions of uncertainty and risk. Based on the results of the study, the optimal combination of tax structural components of consciousness is calculated to maximize its level and reduce tax risks arising in the process of functioning of the state and taxpayer organizations.

The activity of any enterprise, especially oil and gas, is aimed at increasing the economic potential of the country. One of the main indicators is GDP. Forecasting and control of its components are presented in works [29; 34]. To predict macroeconomic indicators, artificial intelligence methods are widely used, in particular, neural networks and elements of the theory of fuzzy logic [32]. Noteworthy is the original approach to modeling the economic indicators of different countries, when initially all countries are segmented into a number of clusters based on key features, each of which unites countries according to the type of behavior. And already within the framework of each individual cluster, forecasting models of the studied macroeconomic indicators are built. This approach can be used in modeling and forecasting the performance indicators of an oil and gas company.

One of the main criteria for the operation of an oil and gas enterprise is resource security [39], which consists of economic actors, environment and purposeful economic process. Risk analysis and management is carried out in order to optimize national resource security. Based on the research, a comprehensive mechanism for supporting innovative research and the safety of technological resources was developed.

In modern conditions, new information technologies come to the fore in production activities. One of the areas of product implementation is electronic commerce [42]. The authors assess the impact of e-commerce on enterprise profits. A comparative analysis of traditional enterprise trade and electronic commerce is carried out. The work developed a model of price competition between companies. A model of a mixed strategy of a company for the sale of goods is considered. In [44], the Salop model with symmetric product differentiation is considered to analyze the expansion of the e-commerce market. The paper shows that the seller's income increases proportionally with the number of buyers. The analysis of advertising costs increases cubically with the number of buyers. The analysis of the optimal advertising costs in the competitive equilibrium model with free access is carried out. The ability to predict product sales is possible based on the Bass model [33]. This model allows you to get the maximum profit for the enterprise at the best price.

In [41], based on the definition of the integral indicator, a model of socio-economic development is investigated. The research uses the methods of factorial, correlation and regression analyzes, as well as the method of expert estimates. The proposed algorithm allows you to efficiently analyze data and obtain adequate integral indicators, incl. and at the enterprises of the oil and gas complex.

A comprehensive study of enterprise innovation processes is especially relevant in the current Industry 4.0 environment [4]. This paper proposes an algorithm for creating a model for managing innovation at an enterprise based on a multi-criteria optimization problem. In a comprehensive study of enterprise risks, it is necessary to take into account the risks associated with IT technologies [2]. Informatization and computerization of production processes suggests that this type of risk is relevant and its study requires due attention. For the successful operation of an enterprise, it is necessary not only to produce products, but also to find consumers. For the prompt processing of information at the enterprise, it is necessary to have software that allows the creation and processing of large volumes of information. In [19], a technology for storing and processing data is proposed.

In [43], a three-level information system for managing business processes was developed. The web-application includes the necessary functional blocks for e-commerce. This system allows for electronic interaction between sellers and buyers.

Analysis of the scientific, theoretical and practical results of the study shows that it is necessary to conduct additional research on the integrated management of enterprise risks in order to optimize them.

The methodological basis of this research is the following scientific research methods: statistical analysis; correlation and regression analysis; factor analysis; simulation modeling; optimization theory; the Pontryagin maximum principle.

3 Discussions and results

The result of this study is to find the optimal management of the distribution of the net profit of an oil and gas company. The study of historical data allows you to obtain a regression equation that reflects the dependence of the company's net profit $P(t)$ on independent variables. The indicators reflecting the areas of use of net profit act as independent factors. Let's represent the regression equation in the form:

$$P(t) = a_0 + \sum_{i=1}^{n} a_i X_i\,(t) \tag{1}$$

where a_0 – is the free term in the regression equation; a_i – regression coefficients $(i = \overline{1, n})$; $X_i(t)$ – independent variables $(i = \overline{1, n})$.We will assume that the resulting linear regression equation for independent factors is adequate and its coefficients are significance.

Thus, we get that on the basis of historical data it is possible to predict the received profit of the company for future periods. To do this, you need to use the regression equation (1). To obtain the necessary predicted data for independent variables, it is necessary to use the laws of distribution of these variables, as well as external and internal factors of influence on the operation of the enterprise.

The task is to determine the optimal use of net profit for transferring the system from the initial state to a given final state. The objective function is the utility function of the investigated enterprise. The management parameters are the shares of net profit aimed at achieving the assigned task. To solve the problem of optimal management of the distribution of the company's net profit, it is necessary to solve the following system of equations.

1. A system of differential equations:

$$\frac{dX_i(t)}{dt} = \mu_{\ i} X_i(t) + V_i(t)\ ,\ \ (i = \overline{1,n}) \tag{2}$$

where μ_i – rate of change of the i-th studied variable $X_i(t)$; $V_i(t) = \nu_i P(t)$ – management functions; ν_i – control parameters ($0 \le \nu_i \le 1,\ i = \overline{1,n}\ ; \sum_{i=1}^{n} \nu_i \le 1$).

2. Initial conditions:

$$X_i(t_0) = X_i^0 \tag{3}$$

3. The utility function for the enterprise:

$$\int_{t_0}^{T} exp(-\delta\ t)\{1 - \sum_{i=1}^{n} exp[-\nu_i(t)P(t)]\}\, dt + \sum_{i=1}^{n} b_i X_i\,(T) \to max \tag{4}$$

where b_i – coefficients of the variables under study at a finite time T (obtained from the transversality condition); $\delta = ln(1 + i)$ – interest accrual rate; i – interest rate (discount) (%).

As a result, we obtain a mathematical model of the optimal distribution of the company's net profit. It is assumed that the independent variables are transferred from a given initial state to a given final state along an optimal trajectory with respect to the control parameters. To solve the resulting system (1) – (4), we use the Pontryagin maximum principle [1; 38].

Let us find a solution to the problem posed for the case of distribution of net profit in four directions and a linear regression equation. The main areas of use of net profit: X_1 – capital investment; X_2 – deductions for social events; X_3 – costs of monitoring the operation of underground gas storage (UGS); X_4 – deduction to the State budget of Ukraine. In the model, we assume that the deduction to the State Budget of Ukraine is a constant percentage of the company's net profit. In the further listed implementation, we will consider several possible options (30%, 40%,45%).

To solve the problem, we construct the Hamiltonian function:

$$H(t) = \sum_{i=1}^{3} \Psi_i(t)(\mu_i X_i(t) + \nu_i(t)P(t)) + e^{-\delta t}\left(1 - \sum_{i=1}^{3} e^{-\nu_i(t)P(t)}\right) \tag{5}$$

where $P(t) = (1 - \nu_4)\pi(t)$; $\Psi_i(t)$ – auxiliary functions found from the system of differential equations (6):

$$\frac{d\Psi_j(t)}{dt} = -\{\mu_j \Psi_j(t) + (1 - \nu_4)a_j[\sum_{i=1}^{3} \Psi_i\,(t)\nu_i(t) + e^{-\delta t} e^{-\nu_i(t)P(t)}]\}, \left(j = \overline{1,3}\right), \tag{6}$$

Auxiliary functions must satisfy the transversality condition:

$$\Psi_i(T) = -b_i, (i = \overline{1,3}) \quad (7)$$

The optimal control parameters are as follows:

$$\nu_i^* = -\frac{ln(-\Psi_i(t)) + \delta t}{P(t)}, (i = \overline{1,3}) \quad (8)$$

As a numerical implementation of the proposed algorithm, we will consider the activities of a conditional oil and gas enterprise. We will use the results of their activities in 2018 as initial data. Consider the options for distributing the company's net income for a five-year period. To solve this problem, it is necessary to set the numerical values of the independent variables in 2022. The initial data for the calculation are presented in table 1.

Table 1. Oil and gas enterprise activity data, thousand UAH

	Indicators	2018	2022
$\pi(t)$	Net income	329939	1250138
$X_1(t)$	Capital investment	80000	179920
$X_2(t)$	Contributions to social events	70092	105348
$X_3(t)$	Costs of control over the operation of USG	48475	116340
$X_4(t)$	Deduction to the State Budget of Ukraine, (30%)	98981	375041

Consider several options for the optimal distribution of net profit, depending on the fixed amount of distribution to the State Budget of Ukraine and the discount coefficient. For the numerical implementation, we assume that the initial conditions (3) have the following form: $X_i(2018) = 1, (i = \overline{1,3})$. We transform the regression equation (1) to the form:

$$\pi(t) = -0{,}3277 - 0{,}0364X_1 - 1{,}0299X_2 + 2{,}3941X_3 \quad (9)$$

Table 2 shows the optimal distribution of net profit in three areas, taking into account the discount rate equal to zero and deductions to the State Budget of Ukraine of 30% of the net profit received.

Table 2. Dynamics of financial flows of an oil and gas company 2018-2022 ($\delta = 0,\ \nu_4 = 0{,}3$)

Indicators	2018	2019	2020	2021	2022
$\pi(t)$	1	1,305	1,851	2,669	3,789
$X_1(t)$	1	1,141	1,388	1,754	2,248
$X_2(t)$	1	1,059	1,161	1,308	1,503
$X_3(t)$	1	1,155	1,431	1,841	2,400

Table 3 shows the optimal share distribution of net profit for the proposed areas of use without taking into account the discount coefficient when deducting 30%, 40% and 45% of the company's net income to the State Budget of Ukraine.

Table 3. Share distribution of net income of an oil and gas company 2018-2022 ($\delta = 0$)

Indicators	2018	2019	2020	2021	2022
			$\boldsymbol{v_4 = 0,30}$		
$v_1(t)$	0	0,099	0,139	0,145	0,136
$v_2(t)$	0	0,042	0,060	0,063	0,060
$v_3(t)$	0	0,107	0,153	0,160	0,150
			$\boldsymbol{v_4 = 0,40}$		
$v_1(t)$	0	0,099	0,122	0,111	0,094
$v_2(t)$	0	0,047	0,059	0,054	0,046
$v_3(t)$	0	0,166	0,205	0,189	0,159
			$\boldsymbol{v_4 = 0,45}$		
$v_1(t)$	0	0,125	0,177	0,184	0,173
$v_2(t)$	0	0,053	0,077	0,080	0,076
$v_3(t)$	0	0,137	0,194	0,203	0,191

Table 4 shows the optimal distribution of the share distribution of net income according to the proposed variables and the trajectory of financial flows, taking into account the intensity of interest accrual and contributions to the State Budget of Ukraine of 30% of net income.

Table 4. Dynamics of financial flows of an oil and gas company 2018-2022 ($\delta = 0,05$; $v_4 = 0,30$)

Indicators	2018	2019	2020	2021	2022
$\pi(t)$	1	1,399	1,983	2,772	3,789
$X_1(t)$	1	1,213	1,488	1,831	2,248
$v_1(t)$	0,134	0,137	0,125	0,110	0,096
$X_2(t)$	1	1,134	1,263	1,387	1,503
$v_2(t)$	0,137	0,088	0,055	0,034	0,021
$X_3(t)$	1	1,227	1,531	1,919	2,400
$v_3(t)$	0,131	0,145	0,138	0,124	0,110

Tables 1-4 show the optimal distribution of the company's net income in the period 2018-2022 depending on the deductions to the State budget of Ukraine and the discount factor. Analysis of the results obtained shows that the maximum share of deductions from net income excluding discounted flows falls on 2021, and taking into account the discounting factor falls on 2018.

4 Conclusions

From the point of view of economic efficiency, the developed model makes it possible to optimally distribute the profit of the enterprise, within the limited financial resources, to manage environmental risks. The cost of managing environmental risks does not increase compared to traditional profit sharing approaches. The developed model is applicable both for the entire system of enterprise risks that are interconnected and for individual environmental risks and industrial groups. This is achieved by taking into account in the model the amount of damage, and not the amount of risk. The use of the developed model will reduce the costs of risk management by 10-15% in conditions of uncertainty and, accordingly, increase the profit of the enterprise by 15-20%, which will generally increase environmental sustainability. It should be noted that losses from environmental disasters are the largest and most difficult to predict. Therefore, the optimal distribution of profits for risk management and efficiency from investment in safety will not only reduce costs, but also reduce the likelihood of unforeseen consequences that may affect the environment.

In this paper, we consider the problem of optimal management of the company’s net income. The distribution of net income occurs in the most important areas of its use. One of the main directions is the use of the obtained profit to prevent possible damage from various risks. In this setting, a mandatory deduction to the State Budget of Ukraine is assumed. Optimal management of the company’s income is simulated, with these deductions of 30%, 40% or 45%. Tables 2 and 3 show the distribution of net income for the period up to 2022 excluding the discount factor. Table 4 shows the dynamics of financial flows and the share distribution of the net income of the conditional oil and gas company, taking into account the discount factor and deductions to the State Budget of Ukraine in the amount of 30% of the net profit ($\delta = 0{,}05;\ \nu_4 = 0{,}3$).

This mathematical model can be used as a guideline for the adoption by the top management of the enterprise of a long-term plan for its development.

References

1. Aseev, S.M.: On some properties of the adjoint variable in the relations of the Pontryagin maximum principle for optimal economic growth problems. Proceedings of the Steklov Institute of Mathematics **287**(S1), 11–21 (2014). doi:10.1134/S0081543814090028
2. Babenko, V., Lomovskykh, L., Oriekhova, A., Korchynska, L., Krutko, M., Koniaieva, Y.: Features of methods and models in risk management of IT projects. Periodicals of Engineering and Natural Sciences **7**(2), 629–636 (2019). doi:10.21533/pen.v7i2.558
3. Babenko, V., Nakisko, O., Latynin, M., Rudenko, S., Lomovskykh, L., Girzheva, O.: Procedure of Identifying of the Parameters of the Model of Management of Technological Innovations in Economic Systems. 2019 IEEE International Scientific-Practical Conference Problems of Infocommunications, Science and Technology, PIC S&T 2019, **9061259**, 324–328 (2019). doi:10.1109/PICST47496.2019.9061259
4. Babenko, V.: Enterprise Innovation Management in Industry 4.0: Modeling Aspects. Emerging Extended Reality Technologies for Industry 4.0: Early Experiences with

Conception, Design, Implementation, Evaluation and Deployment. Ed. by Jolanda G. Tromp et al. A John Wiley & Sons, Inc., pp. 1-24 (2020). doi:10.1002/9781119654674.ch9

5. Babenko, V.: Formalization of the Model of Management of the Technological Innovations. CEUR Workshop Proceedings **2393**, 595–602 (2019)
6. Bashynska, I., Baldzhy, M., Ivanchenkova, L., Skliar, L., Nikoliuk, O., Tkachuk, G.: Game Risk Management Methods for Investment Portfolio Optimization. International Journal of Recent Technology and Engineering **8**(2), 3940–3943 (2019). doi:10.35940/ijrte.B1729.078219
7. Bashynska, I., Sokhatska, O., Stepanova, T., Malanchuk, M., Rybianets, S., Sobol, O.: Modelling the risks of international trade contracts. International Journal of Innovative Technology and Exploring Engineering **8**(11), 2815–2820 (2019). doi:10.35940/ijitee.K2313.0981119
8. Bašta, M., Molnár P.: Oil market volatility and stock market volatility. Finance Research Letters **26**, 204–214 (2018)
9. Berlin, A.: Managing Political Risk in the Oil and Gas Industries. Oil, Gas & Energy Law Intelligence 2. https://www.ogel.org/article.asp?key=140 (2003). Accessed 17 Aug 2020
10. Chang, C.-L., McAleer, M., Wong, W.-K.: Risk and Financial Management of COVID-19 in Business, Economics and Finance. J. Risk Financial Manag. **13**, 102 (2020)
11. Cunado, J., de Gracia, F.P.: Oil price shocks and stock market returns: Evidence for some European countries. Energy Economics **42**, 365–377 (2014)
12. Derbentsev, V., Matviychuk, A., Soloviev, V.N.: Forecasting of Cryptocurrency Prices Using Machine Learning. In: L. Pichl, C. Eom, E. Scalas, & T. Kaizoji (Eds.) Advanced Studies of Financial Technologies and Cryptocurrency Markets, pp. 211–231 (2020). doi:10.1007/978-981-15-4498-9_12
13. Derbentsev, V., Semerikov, S., Serdyuk, O., Solovieva, V., Soloviev, V.: Recurrence based entropies for sustainability indices. E3S Web of Conferences **166**, 13031 (2020). doi:10.1051/e3sconf/202016613031
14. Donecz, L., Prudnikova, L.: Risks of diversification of the economic activity of the enterprise. Problems and prospects of entrepreneurship development **1**, 16–20 (2011)
15. Dzwigoł, H., Dzwigoł-Barosz, M., Zhyvko, Z., Miskiewicz, R., Pushak, H.: Evaluation of the Energy Security as a Component of National Security of the Country. Journal of Security and Sustainability Issues **8**(3), 307–317 (2019). doi:10.9770/jssi.2019.8.3(2)
16. Fadyeyeva, I., Horal, L.: Application of modern modeling in system of strategic management at oil and gas companies. The Economic Annals-XXI Journal 1-2, 106–109 (2014)
17. Fang, F.: A Study of Financial Risks of Listed Manufacturing Companies in China. Journal of Financial Risk Management **5**, 229–245 (2016). doi:10.4236/jfrm.2016.54022
18. Halynska Y., Telizhenko A.: Risk in the formation of collaboration alliance of the redistribution natural rental income. Problems and Perspectives in Management **14**(4), 181–185 (2016). doi:10.21511/ppm.14(4-1).2016.06
19. Hrabovskyi, Y., Babenko, V., Al'boschiy, O., Gerasimenko, V.: Development of a Technology for Automation of Work with Sources of Information on the Internet. WSEAS Transactions on Business and Economics **17(**25), 231–240 (2020). doi:10.37394/23207.2020.17.25
20. Khvostina, N., Havadzyn, N., Horal, L., Yurchenko, N.: Emergent Properties Manifestation in the Risk Assessment of Oil and Gas Companies. CEUR Workshop Proceedings **2422**, 157–168 (2019)
21. Kozmenko, O., Oliynyk, V.: Statistical model of risk assessment of insurance company's functioning. Investment Management and Financial Innovations **12**(2), 189–194 (2015)

22. Kucherova, H., Pokataieva, O., Bilska, O.: Modelling tax consciousness evaluation in the context of economic development uncertainty. Periodicals of Engineering and Natural Sciences **7**(2), 567–579 (2019). doi:10.21533/pen.v7i2.565
23. Kucherova, H., Serhieieva, L., Bilska, O.: Matrix models for assessing the taxation subjects' interaction under uncertainty of socio-economic processes. CEUR Workshop Proceedings **2422**, 371–384 (2019)
24. Kwilinski, A., Dzwigol, H., Dementyev, V.: Model of Entrepreneurship Financial Activity of the Transnational Company Based on Intellectual Technology. International Journal of Entrepreneurship **24**(1S), 1–5 (2020)
25. Lambrechts, D., Blomquist, L.B.: Political-security risk in the oil and gas industry: the impact of terrorism on risk management and mitigation. Journal of Risk Research **20**(10), 1–18 (2017). doi:10.1080/13669877.2016.1153502
26. Li, L.: A Study on Enterprise Risk Management and Business Performance. Journal of Financial Risk Management **7**(1), 123–138 (2018). doi:10.4236/jfrm.2018.71008
27. Malyarets, L.M., Babenko, V.O., Nazarenko, O.V., Ryzhikova, N.I.: The Modeling of Multi-criteria Assessment Activity in Enterprise Management. Int. J Sup. Chain. Mgt. **8**(4), 997–1004 (2019)
28. Matviychuk, A., Novoseletskyy, O., Vashchaiev, S., Velykoivanenko, H., Zubenko, I.: Fractal analysis of the economic sustainability of industrial enterprise. CEUR Workshop Proceedings **2422**, 455–466 (2019)
29. Matviychuk, A., Strelchenko, I., Vashchaiev, S., Velykoivanenko, H.: Simulation of the Crisis Contagion Process Between Countries with Different Levels of Socio-Economic Development. CEUR Workshop Proceedings **2393**, 485–496 (2019)
30. Mensah, G., Gottwald, W.: Enterprise Risk Management: Factors Associated with Effective Implementation. Social Science Electronic Publishing, Amsterdam (2016)
31. Nesterova, K., Marchenko, V., Lazebnyk, I., Pavlova, V., Burkova, L., Omelchuk, L.: Identification And Assessment Of External Risks Of The Enterprise's Foreign Economic Activity. International Journal of Scientific & Technology Research **9**(2), 4672–4675 (2020)
32. Oliinyk, V., Kozmenko S.: Forecasting and management of gross domestic product. Journal of International Studies **12**(4), 214–228 (2019). doi:10.14254/2071-8330.2019/12-4/14
33. Oliinyk, V., Kozmenko, O., Wiebe, I., Kozmenko, S.: Optimal Control over the Process of Innovative Product Diffusion: The Case of Sony Corporation. Economics and Sociology **11**(3), 265–285 (2018). doi:10.14254/2071- 789X.2018/11-3/16
34. Oliinyk, V.: Optimal Management of GDP Components. Journal of Advanced Research in Law and Economics **IX**/2(32), 603–614 (2018). doi:10.14505/jarle.v9.2(32).24
35. Oliinyk, V.: Optimal Management of the Enterprise's Financial Flows. Journal of Advanced Research in Law and Economics **VIII**/6(28), 1875–1883 (2017). doi:10.14505/jarle.v8.6(28).22
36. Oliynyk, V.: Modeling of the optimal structure of insurance portfolio. Problems and Perspectives in Management **13**(2), 230–234 (2015)
37. Pitatzis, A.: Ernst & Young Report: Top 10 Risk and Opportunities for Oil and Gas Companies, a useful tool for strategic decisions for oil and gas managers. https://energyroutes.eu/2015/11/24/ernst-young-report-top-10-risk-and-opportunities-for-oil-and-gas-companies-a-useful-tool-for-strategic-decisions-for-oil-and-gas-managers/ (2015). Accessed 17 Aug 2020
38. Pontryagin, L.S, Boltyanskii, V.G, Gamkrelidze, R.V., Mishchenko, E.F.: The mathematical theory of optimal processes. Interscience Publishers John Wiley&Sons, New York-London (1962)

39. Prokopenko, O., Omelyanenko, V., Klisinski, J.: Innovation policy development conceptual framework for national resource security providing, Journal of Environmental Management and Tourism **9**(5),1099–1107 (2018)
40. Prokopenko, O.V., Domashenko, M.D., Shkola, V.Y.: Management features of economic security in foreign economic activity of ukrainian machine-building enterprises. Actual Problems of Economics **160**(1), 188–194 (2014)
41. Pursky, O., Dubovyk, T., Gamova, I., Buchatska, I.: Computation algorithm for integral indicator of socio-economic development. CEUR Workshop Proceedings **2393**, 919–934 (2019)
42. Pursky, O., Dubovyk, T., Moroz, I., Buchatska, I., Savchuk, A.: The price competition simulation at the blended trading market. CEUR Workshop Proceedings **2422**, 15–26 (2019)
43. Pursky, O., Selivanova, A., Dubovyk, T., Herasymchuk, T.: Software implementation of e-trade business process management information system. CEUR Workshop Proceedings **2546**, 171–181 (2019)
44. Pursky, O.I., Grynyuk, B.V., Shestopal, D.A.: Planning of advertising costs and vendor number at e-trade market. Actual Problems of Economics **177**(3), 407–413 (2016)
45. Rogers, V.C., Ethridge, J.R.: Enterprise Risk Management in the Oil and Gas Industry: An Analysis of Selected Fortune 500 Oil and Gas Companies' Reaction in 2009 and 2010. American Journal of Business Education **9**(1), 23–30 (2016)
46. Semerikov, S., Chukharev, S., Sakhno, S., Striuk, A., Osadchyi, V., Solovieva, V., Vakaliuk, T., Nechypurenko, P., Bondarenko, O., Danylchuk, H.: Our sustainable coronavirus future. E3S Web of Conferences **166**, 00001 (2020). doi:10.1051/e3sconf/202016600001
47. Slime, B.: Modeling and Quantifying of the Global Wrong Way Risk. Journal of Financial Risk Management **6**, 231–246 (2017). doi:10.4236/jfrm.2017.63017
48. Turn risks and opportunities into results: Exploring the top 10 risks and opportunities for global organizations. https://www.ey.com/Publication/vwLUAssets/The_top_10_risks_and_opportunities_for_global_organizations/$FILE/Business%20Challenge%20main%20report-%20SCORED.pdf (2011)
49. Xu, M.: Design of Enterprise Training Model Based on Dynamic Knowledge Creation. American Journal of Industrial and Business Management **9**(9), 1761–1771 (2019). doi:10.4236/ajibm.2019.99115

Scenario forecasting information transparency of subjects' under uncertainty and development of the knowledge economy

Hanna Kucherova[1[0000-0002-8635-6758]], Anastasiia Didenko[1[0000-0003-1136-403X]], Olena Kravets[1[0000-0002-2980-5238]], Yuliia Honcharenko[1[0000-0003-1567-8618]] and Aleksandr Uchitel[2[0000-0002-9969-0149]]

[1] Classic Private University, 70-b Zhukovskogo Str., Zaporizhzhia, 69002, Ukraine
ekonom.kpu@gmail.com
[2] State University of Economics and Technology,
5 Stepana Tilhy Str., Kryvyi Rih, 50006, Ukraine

Abstract. Topicality of modeling information transparency is determined by the influence it has on the effectiveness of management decisions made by an economic entity in the context of uncertainty and information asymmetry. It has been found that information transparency is a poorly structured category which acts as a qualitative characteristic of information and at certain levels forms an additional spectrum of properties of the information that has been adequately perceived or processed. As a result of structuring knowledge about the factor environment, a fuzzy cognitive model of information transparency was constructed in the form of a weighted digraph. Structural analysis and scenario forecasting of optimal alternatives of the fuzzy cognitive model made it possible to evaluate the classes of factors, identify their limited relations, establish the centrality of the roles of information transparency and information and communication security in the system built and evaluate their importance when modeling the situation self-development. Information visibility, reliability and availability have been found to have the strongest impact on the system. Taking into account different initial weights of the key factors — information transparency and information and communication security — the study substantiates the strategic ways for economic entities to achieve their goals in the context of uncertainty and information asymmetry, which allows us to use this approach as a tool for strategic management in the information environment.

Keywords: information transparency, forecasting, fuzzy cognitive modeling, digraph, factors, relations, strategic management.

1 Introduction

Information transparency is a possibility for any economic stakeholders to track the chain of actions and stages of forming the information content [29] which is important enough to make effective management decisions. On the one hand, information transparency in required scope ensures that its provider can be held accountable for the content and consequences of its disclosure, while on the other it builds trust and reliance on a company and reduces alienation between interaction parties [15]. That is, systemic information transparency determines positive expectations regarding company development, builds up confidence and improves business climate, which is always relevant in the context of competition, economic downturn, changes in socio-political and socio-economic vectors of development.

From the point of view of content understanding, information transparency is interpreted by authors [6] as a requirement, norm, standard or as a goal which, being set, defines the specifics of achieving it. Thus, transparency as one of the 8 major characteristics of the Good governance concept [26] means that decisions taken and their implementation follow rules and regulations. It also means that information is freely available and directly accessible to those who are affected by the decisions and their enforcement. According to the concept, main attributes of information are the scope of information provided and that it is provided in easily understandable forms and media. Another approach can be found in the article [25] whose authors define transparency as perceived quality of the information intentionally shared by its sender. We believe that information transparency is a poorly structured category which nevertheless is a qualitative characteristic of information and at certain levels forms an additional spectrum of properties of the information which has been adequately perceived or processed. Ambiguity of information transparency complicates drawing its overall portrait for comprehensive research and identifying its optimal level for different economic entities and their various purposes.

Furthermore, in economics poorly structured categories are generally studied by means of soft modeling including dynamic cognitive modeling and cognitive mapping. In our case, information transparency is the object of the research, while cognitive modeling of how transparency and its constructs influence each other is its subject. Scientific search is aimed at structuring the information transparency factor environment, identification of strategic changes in its level and the levels of its factors which result in accelerated and decelerated development of the system, extracted from a specific eco-socio-economic environment during t period. Information types, degree of completeness of information, the type of the company's operations, its size, organizational and legal form are decisive factors for characterizing its information transparency.

The openness of the organization's activities and management decisions, disclosure of information in a complete, timely, reliable and understandable form creates space for the formation of transparency in business for the main groups of stakeholders: employees, investors, consumers, suppliers, creditors, government and local authorities, etc. [5; 12; 17; 19; 20].

Own approach to the definition of information transparency of business entities is proposed in the research [34]. The authors of the study understand information transparency as providing free access for all stakeholders to information about various areas of the enterprise's activities, management status, ownership structure, and other data that will be useful for making decisions on interaction with this entity. The authors of the article consider information transparency in two aspects: external and internal.

The external aspect contributes to the formation of a positive image and increases the competitiveness of the enterprise. The internal aspect forms trust within the organization. The authors argue that information transparency is an important tool in the process of preventing and combating economic crime.

Transparency is an integral part of public administration, which is considered in the article [30]. Transparency is determined by the degree of citizens' awareness of important manifestations of the material and procedural aspects of the activities of state authorities and local self-government. The author of [9] notes a close relationship between information transparency of companies and the corporate governance model. A high level of information transparency contributes to the formation of a high-quality model of corporate governance, increasing investment attractiveness. According to the [11] information transparency is a conceptual basis for solving investment risk problems. In the [15] it is noted that the availability of information and transparency helps to strengthen the reliability of the company and reduce misunderstandings between the organization and stakeholders.

Transparency as an openness that allows stakeholders to receive the necessary information for making informed decisions considers in [32]. Availability, timeliness, relevance, and quality are considered characteristics of transparency. At the same time, the lack of transparency is the basis for the development of corruption, it negatively affects economic growth in any industry. Increasing information transparency helps to improve management transparency.

The author of [24] considers transparency as a function that has three variables. This function consists of the owner of the information, the information, and the persons who have access to the information. Transparency is seen as a way of regulation, democracy, and efficiency. Possession of information facilitates control over the entity. At the same time, transparency is not a category that must maximize. The optimal level of transparency is between "highly desirable" and "necessary".

In the article [33] considers two conceptual approaches to transparency. The first approach looks at information transparency, which involves increasing the quantity and quality of information available to stakeholders. The second approach looks at accountability, which manifests itself in the strengthening constraints on employees to hold them accountable for their actions. So, on the one hand, transparency is a means of transmitting information and on the other, it is a tool that restrains the actions of officials. The article substantiates that due to different types of transparency and their economic effects, it is impossible to introduce a single quantitative indicator to measure the level of transparency.

Thus, for economic entities information can be transparent, partially transparent or nontransparent. The main parameters of information transparency management are time, tactical, strategic and technological constrains on the flow of information for a limited or unlimited number of stakeholders. Information flow in the research [6] is characterized by the following parameter tuple:

$$\langle I, IP, IR, IM, IE\rangle, \quad (1)$$

information (I), information provider (IP), information receiver (IR), information medium (IM), and information entity (IE).

It seems reasonable to add information constraints (such as on the scope of information usage) to this set of defined parameters. It is also appropriate to take into account the period with its inherent social and economic trends, which can ensure more accurate and adequate interpretation of results. The authors [6] place emphasis on the importance of subject affiliation when studying the risks of achieving information transparency, which explains different quality of certain level of information transparency for different stakeholders. Information receivers and media also differ in terms of their goals, capabilities and time allocated to the operation, which should also be taken into consideration when assessing the feasibility of increasing or reducing information transparency in order to improve rather than aggravate the situation.

Transparency depends on information type (financial, non-financial, business, public, private etc.) so basic transparency attributes will be further explained through specific characteristics that determine the specialization of a particular type of information. By way of example we can mention corporate information: corporate transparency is defined as widespread availability of relevant, reliable information about periodic performance, financial position, investment opportunities, governance, value, and risk of publicly traded firms [3]. Also, taking into account subjective nature of information (for example, information for external stakeholders) a company's external information transparency determines the degree of completeness of information regarding its own business activities, provided by each company to the market [31].

2 Materials

One of the main attributes of information transparency is *visibility* of information content which in its turn is determined by general availability of information, accessibility of information to third parties and degree of approval to disseminate information [29]. Information visibility is also differentiated according to its degree as visible, partially visible and invisible. In the article [29] the authors argue that high information visibility may result in attention span, produce a flood of information, overwhelming our cognitive and interpretive capabilities, thus rendering the information meaningless or confusing. Moreover, maximum visibility can mask formality of actions or results of the decisions taken, which are presented in the information [29]. On the other hand, insufficient disclosure of information may result in an attempt to compensate for it with some other, forecasted, predicted, guessed, insider information, trustworthiness and reliability of which may not be sufficient to make an effective management decision. The authors of the article [15] point out that transparency arises under the condition of certain balance between the degree of information disclosure and the degree of its perception and understanding by users. Hence, we can make a conclusion that only limited (up to a certain level) visibility ensures high level of transparency, provided that it is the key and quality information that is disclosed, not its noises. In general, the degree of information transparency determines the *associated risk of its disclosure or concealment*. According to the [15], information visibility is determined by *how responsible* its providers are, that is, how consciously they assess the advantages, necessity and need to disclose certain amount and type f information. The attitude and policy of an economic entity in terms of establishing the boundaries of information transparency determine *the degree of confidence* it enjoys as an information provider and medium. *Information availability* depends on how well information reflects actions and decision-making stages in easily recognizable data forms through a reliable storage of those data. In the article [29] it is interpreted as a degree of complexity associated with the extraction and interpretation of information. It is determined by such attributes as possibility for an unlimited number of people to be informed, insignificant amount of time, finances and efforts necessary to find the relevant information, veracious perception of information content. Information availability is ensured by its relevant classifiers, systematizes etc. *Ongoing development of information and communication technologies* has been improving the conditions, forms and methods of storing and transmitting information as accurately as possible, which allows us to expand the possibilities of managing information transparency and its constructs. In order to assess the level of information and communication technologies development, different indices are used, including ICT Development Index, Networked Readiness Index (surveys at the state, business and community levels). Growing complexity of information and communication technologies requires continuous improvement of existing skills on the part of economic entities and drives the need for life-long learning. At the same time, technological advances mean higher associated risks, which urge stakeholders to provide the required *level of information and communication security*.

Adequate and realistic perception of information content by receivers, which is crucial for effective decision making, is determined by *cognitive limitations* of economic entities. Information overload may result in deliberate distancing from all information sources. In addition, dubious quality of information, its shortage or poor relevance cause wrong decisions. We cannot ignore the human factor, i.e. the difference in perception of the same facts by different economic entities, which also contributes to information asymmetry. In general, the above determines *how effectively an economic entity uses the information visible to it*. Information society has been creating new information services and changing approaches to delivering old ones by means of increasing use of modern technologies. Therefore, in order to understand the state of implementing information and communication technologies in society and evaluate the effectiveness of new technologies introduction, there is an urgent need to develop approaches to evaluate knowledge and skills of the service receivers, their social communication skills and skills large-scale data processing [18]. When discussing information transparency, it is absolutely necessary to take into account the role of transparency as an economic entity's accountability. It is not the information itself that is considered, but rather the potential disclosure of that information, which forces stakeholders to “do the right thing”, so information transparency is often quite closely linked to the problem of *corruption*. Hence, we think of information transparency as a vital tool that helps reduce corruption [33]. We find the list of information transparency factors, suggested by the experts, substantial, but offer to add information transparency levels, degrees, and constraints that occur in a real-life environment when it transfers from one state to another. This will help to substantiate the feasibility of achieving a change in specific degrees or levels of factors, informational transparency, in order for stakeholders to achieve their goals. As a result of structuring knowledge on factor loading of information transparency we have identified the following main factors that accelerate and decelerate the development of an economic entity system: the degree of confidence in the medium and source of information, the degree of the economic entity responsibility, the scope of information visibility, the degree of information availability, the degree of information accessibility to third parties, approval to disseminate information, the degree of information reliability, the level of information and communication technologies development, cognitive limitations of the user, the level of information use efficiency, the level of information and communication security, the level of corruption reduction, risks of information disclosure or concealment.

3 Methods

The methodology of information transparency evaluation includes a wide scope of modern methods, from social research to modeling; results of many surveys and observations are studied using econometric tools. The results of critical analysis presented in the article [15] show that there are different approaches to transparency evaluation (market-based, analyst perception-based and accounting-based measures) and researchers invoke different techniques including web content analysis, verbal protocol study, principal component analysis etc. However, it is index approach that has gained popularity in this area of research. In addition, information transparency is defined as a component of a more general indicator, or examined in the context of information type (financial, fiscal, political, economic information). The index method was used in the study [33] to measure Information Transparency Index and the Accountability Transparency Index constructed using a methodology similar to the one used in Transparency International's Corruption Perceptions Index. In the first stage, the researchers selected indicators that determine the index for the maximum period, and normalized them. The resulting index was calculated as a mean of the components of its normalized indicators. The large-scale study [2] provides a transparency index for 194 countries. The indices comprise an aggregate transparency index with two sub-components: economic/institutional transparency and political transparency. By means of correlation and regression analysis using substantiated indicators, the authors showed that information transparency is associated with better human and economic development indicators, higher competition and lower corruption. Political transparency has also been studied separately; the main indicator used over the years has been the Freedom House *Freedom of the Press* index. Its advantage is extensive coverage across countries combined with significant temporal coverage (going back to 1979) [33]. Economic transparency index was presented in the article [10] where it was developed using regression estimation (based on the World Bank and IMF data). The index measures the frequency with which governments update economic data that they make available to the public as a condition of existence (or absence) of laws on freedom of information. The authors took into account the quantity of data released by governments, rather than their timeliness. As a result, they proved that information transparency is directly proportional to effective governance. It is also worth mentioning the Open Budget Index, prepared by the International Budget Partnership, which is aimed at studying fiscal transparency [33]. In addition to these indicators, transparency is studied as a component within larger frameworks. In particular, the World Bank's CPIA indicators have a component on "transparency in the public sector". Information transparency for businesses in terms of obtaining information about government policies and regulations affecting business activities is considered in the Global Competitiveness Report, while freedom of the press as a condition for information transparency is represented in the *CIRI Human Rights* database [33].

One of the priority indicators of the state of information transparency in Ukraine is the Transparency Index of Ukrainian companies' websites; the dynamics of this indicator is shown in figure 1.

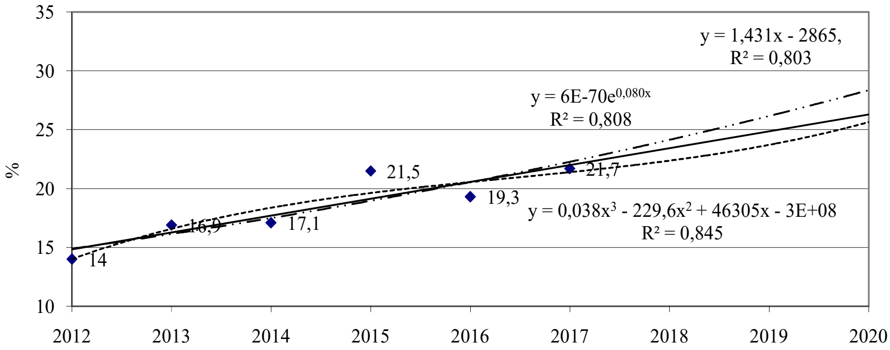


Fig. 1. Dynamics of the Transparency Index of Ukrainian companies' sites.

The results of the analysis of the dynamics and structure of the Transparency Index of Ukrainian companies' websites show that the level of indicators is low however it grows every year, in particular: in 2012 the level of the studied indicator is 14%, in 2013 – 16.9%, in 2014 – 17.1%, in 2015 – 21.5%, in 2016 – 19.3%, in 2017 – 21.7% [35]. The study involved sites of companies that are large taxpayers and leaders of information transparency of the previous period. Transparency companies are in the energy sector, the agricultural sector, communications and transport, metallurgy, and alcohol production. Companies belonging to the mining industry have the worst values. DTEK, SCM, and ArcelorMittal Kryvyi Rih are stable leaders in information transparency.

According to the method of construction, the index is an integrated value of indicators and such categories as content, reporting, navigation, accessibility. The degree of disclosure of the indicator "content" of companies is gradually growing. So it was 6.7% in 2013, 6.3% in 2014, 9% in 2015, 10.3% in 2016, 12.8% in 2017 [35]. In the category of "reporting" companies were assigned 9% in 2012, 13% in 2013, 14% in 2014, 17% in 2015, 12% in 2016, and 16% in 2017. In the reporting category, companies were assigned 9% in 2012, 13% in 2013, 14% in 2014, 17% in 2015, 12% in 2016, and 16% in 2017. Companies scored 3.5 out of 8 possible points in terms of navigation and 6 out of 17 possible points in terms of accessibility. The analysis of indicators also confirmed the importance of the impact of the financial condition of both companies and sectors on the level of information transparency of companies. The experts made the following conclusions about the state of information transparency of companies: it is necessary to ensure systematic updating of information, use modern data visualization technologies, increase accessibility, accelerate disclosure of information on compliance with sustainable development and non-financial reporting, to research information needs of society and the market. Thus, to increase the level of information transparency of Ukrainian companies, the target benchmarks are high availability, visibility of relevant information. Information is presented in an understandable form that fully reveals the principles of sustainable development and the level of corporate social responsibility. Gives information about correlation between the industry indicators and the Information Transparency Index (table 1).

Table 1. The correlation coefficient between the industry indicators and the Information Transparency Index.

Indicator	Industry	Agriculture	Financial and insurance activities	Trade	Transportation and storage	Real estate activiti es
Current liquidity, %	-	0,69	-0,98		0,74	-
Absolute liquidity, %	0,83	0,7	-0,64	0,81	0,99	-
Financial autonomy ratio, %	-0,61	0,59	-0,96	-	-0,68	-0,72
Return on assets, %	0,87	0,78	-	0,7	0,63	-
Return on current assets ,%	0,84	0,78	-	0,7	0,63	0,6
Net margin, %	0,83	0,77	0,51	0,84	0,63	0,53
Return on total assets, %	-	0,77	-0,65	-	0,54	0,55
Total asset turnover	0,91	-	-0,58	-	0,77	0,58
Working capital turnover	-	-	-0,58	-	0,78	-
Receivables turnover ratio	0,8	-	-0,55	-	0,75	-

According to table 1, there is a direct close relationship between the level of information transparency and indicators of the financial condition of industrial and agricultural complexes. There is also a direct but average density relationship between the studied index and the transport and storage sector. Also, the feedback of the average density is shown between the indexes of information transparency and indexes of the financial condition of the financial and insurance sector. The trade sector has a direct close relationship with the indicators of information transparency, however, not for the entire studied financial index, only for absolute liquidity, return on assets and working capital, net margin. Medium-density direct connection with information transparency is the return on assets and return on current assets, net margin, total asset turnover, there is inverse relationship with financial autonomy ratio of the real estate sector. Thus, there is both a qualitative and quantitative link between information transparency and sectoral indicators of the national economy, which determines the feasibility of further research in this area.

The research [13] offers to use a modified Delphi technique in order to study information transparency. This technique implies having rapid reviews from stakeholders, namely researchers, policy makers, industry and health care providers. However, the results of this approach are rather subjective as respondents often change their opinion under the influence of different factors.

The paper [8] proposes four reference models which form a baseline for transparency requirements in information systems. The models cover actors involved in the process of ensuring transparency and information circulation among them, transparency meaningfulness, transparency usefulness for a specific audience, and information quality in transparency. The reference models are further used to create TranspLan, a modeling language for capturing and analyzing information transparency requirements among stakeholders [7]. Thus, a broad range of methodological tools gives researchers quantitative measures of information transparency, the impact it makes and gets. However, each approach or method has inherent risks when used in practice due to assumptions, limitations that are not achievable in real life. In addition, there is no general system of forming information transparency as an economic category since it is impossible to formalize this poorly structured category without considerable impact of subjectivity and assumptions. Considering the importance of the cognitive method when studying poorly structured categories, we propose to analyze information transparency using the methodology of cognitive modeling and impulse processing [17; 23]. The proposed approach is based on structuring knowledge about the object of study, building a cognitive model of object development in the form of a digraph, scenario modeling of the dynamics in the state of the system when it is stable and changing, justification of the option that can help to achieve the goals of the research.

However, the results of previous studies have proved [14] that the level of the values of influencing factors plays a crucial role in substantiating an adequate and high-quality forecast scenario for achieving the required level of information transparency. To solve this problem, we propose to apply the methodology of fuzzy cognitive modeling.

The process of fuzzy cognitive modeling is presented as a sequence of steps. At the first stage, the purpose of modeling and target indicators are determined. In the second stage, there is a substantiation of a fuzzy cognitive map (causal graph) by forming a set of factors of the subject area of research. Factors are divided into four categories: target, which form the purpose of modeling, controlled – these are those that can be directly influenced, intermediate – to describe the subject area, observed – externalities. Factors are selected according to the level of significance and set for each level scale [1; 28]:

$$\tilde{G} =< C, \tilde{R} >, \tag{2}$$

where $C = \{c_1, c_2, \ldots, c_n\}$ – set of factors, $\widetilde{R}$ – fuzzy causal relations on the set C, $\tilde{R} = \{(\mu_R(c_i, c_j)/(c_i, c_j))\}$ – fuzzy set of edges, $c_i, c_j \in C^2$; $\mu_R\{(c_i, c_j)\}$ – the degree of belonging of the edge $\{(c_i, c_j)\}$ to fuzzy set of oriented edges $\widetilde{R}$ [1; 28].

Elements $r_{ij} \in \widetilde{R}$ $(i, j = 1, \ldots, n)$ characterize the direction and degree of intensity (weight) of the impact between the factors c_i and c_j [1; 28]:

$$r_{ij} = r\,(c_i, c_j), \tag{3}$$

where r – indicator of intensity of influence (characteristic function of the relation $\widetilde{R}$), which may have any value in the interval [–1, 1] with considering:

- $r_{ij} = 0$, if the value c_i does not depend on c_j (no influence) [1; 28];
- $0 < r_{ij} < 1$ if the influence c_i to c_j is positive (increasing the value of the factor-cause c_i leads to an increase in the value of the factor-consequence c_j) [1; 28];
- $-1 < r_{ij} < 1$ if the influence c_i to c_j is negative (increasing the value of c_i leads to an decrease in the value of c_j) [1; 28].

The next step is to determine the initial state of the factors and the influence of external factors [1; 28].

In the fourth stage, an adjacency matrix is formed $M = \|m_{ij}\|_{n\times n}$ based on reasonable characteristics of the type of relationship between factors [1; 28].

Suppose a signed digraph with adjacency matrix A determines information transparency and its factor environment. The vertices of the digraph are represented by the set $u_1, u_2, \ldots, u_n$. Every vertex u_i has the value of $v_i(t)$ at discrete moments of time $t = 0, 1, 2, \ldots$. The value of $v_i(t + 1)$ is determined by the value of $v_i(t)$ and information on whether other vertices u_j, adjacent with u_i, have increased or decreased their values at moment t. The change of $p_j(t)$, set by the difference of $v_j(t) - v_j(t - 1)$, is called a pulse if $t > 0$. The initial condition should be specified when $t = 0$. We introduce the following notation:

$$sgn(u_j, u_i) = \begin{cases} 1,\text{if edge is positive,} \\ -1,\text{if edge is negaive,} \\ 0,\text{if edge is missing.} \end{cases}, \tag{4}$$

Analysis of the statistical characteristics of the model, determination of the balance of the system, consonance and dissonance of influence are carried out at the next stage [1; 28]. Detect the indirect interactions of factors on each other in the system, for this purpose convert the initial matrix of the intensity of interactions C into a transitively closed matrix Z, the elements of which are pairs $(z_{ij}, \overline{z_{ij}},)$, where z_{ij} characterizes the power of positive influence, $\overline{z_{ij}}$ – the strength of the negative influence of the i-th factor on the j-th [1; 28].

Based on the Z matrix system characteristics of the fuzzy cognitive map can be calculated (see table 2).

The consonance indicator expresses the degree of confidence in the sign and the strength of the impact (the higher the consonance, the more convincing the expert's opinion) [1; 28]):

p_{ij} – the number of positive influences of the i-th factor on the j-th factor;
n_{ij} – the number of negative influences of the i-th factor on the j-th factor;
k_{ij} – consonance of the influence of one factor on another;

d_{ij} – dissonance of the influence of one factor on another.

Table 2. System characteristics of fuzzy cognitive map [27].

Characteristics	Dissonance of influence	Consonance of influence
Influence of the i-th factor on the j-th	$d_{ij} = 1 - k_{ij}$	$k_{ij} = \dfrac{\lvert p_{ij} - n_{ij} \rvert}{p_{ij} + n_{ij}}$
Influence of the i-th factor on the system	$D_i = \frac{1}{n}\sum_{j=1}^{n} d_{ij}$	$K_i = \frac{1}{n}\sum_{j=1}^{n} k_{ij}$
Influence of the system on the j-th factor	$D_j = \frac{1}{n}\sum_{i=1}^{n} d_{ij}$	$K_j = \frac{1}{n}\sum_{i=1}^{n} k_{ij}$

At this stage, dynamic modeling is performed based on the use of a pulse process according to the following formula [1; 28]:

$$v_i(t+1) = S(v_i(t) + q_i(t+1) + o_i(t+1) + \textstyle\sum_{j=1}^{K} T\,(w_{ij}, p_j(t))), \qquad (5)$$

where $v_i(t)$ – the value of factor c_i at moment t; $v_i(t+1)$ – the value of factor c_i at moment $(t + 1)$; $q_i(t + 1)$ – external influence on c_i at moment $(t + 1)$; $o_i(t + 1)$ – controlling influence on c_i at moment $(t + 1)$; $r_{ij} = r(c_i, c_j)$ – intensity of influence between factors e_i and e_j; $p_j\,(t)$ – change the value of c_j at moment t; T – operation T-norms (the product is used); S – operation S-norms (Lukasevich's S-norm is used).

At the final stage, form a basic set of alternative strategies for system development and justification of priority strategies for achieving targets.

Modern information technologies provide researchers with a possibility to simplify cognition procedure. In particular, to achieve the goal of this research we can use some software products, such as Decision Explorer Application, aimed at providing the user with illustrative cognitive maps which can be further analyzed to better understand the issue under study, links between its factors. Another specialized software product is FCMapper which helps to analyze, model fuzzy cognitive maps, explore behavior of the system and interaction between its factors. FCMapper calculates structural characteristics of the system based on the created digraph, and provides a variety of scenarios. The logistic squashing function is used for standard scenario calculation. It can be written as $f(x)=1/(1+e^{-x})$ [4]. The results obtained for each scenario are compared with the initial one, where the output weights for each vertex are set at the same level, that is, the situation self-development is modeled. The final development of the system in the form of a cognitive model in the absence of external influences can be reduced to a fixed-point attractor, a boundary cycle of repeating binary vectors, or to chaotic or aperiodic attractors [21].

4 Results

As a result of structuring knowledge on factor environment of information transparency we have identified a set of vertices and their cause-effect relations; together they determine the corresponding cognitive model in the form of a weighted digraph (see fig. 2).

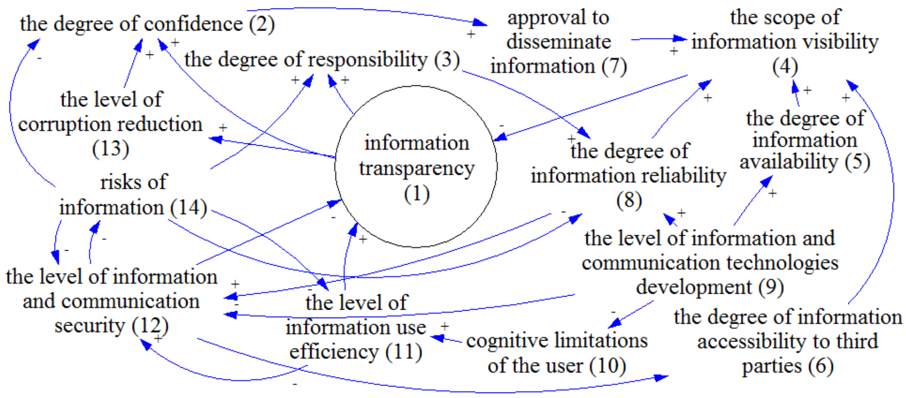


Fig. 2. Cognitive model of information transparency [14].

For the constructed digraph, based on the situation simulation, the following vertices have been found to have the greatest influence on target vertex 1: 8, 9, 11, 12, 14. However, not all of these vertices can be affected in terms of goals and opportunities of economic entities. The following figure shows the results of studying a simulation of the vertex parameter dynamics in the context of system interaction without external influence for $t = 12$, where t is the study period, and the initial values of the vertex parameters $v_i = 0$ (see fig. 3).

The diagram shows the dynamics of the influence those vertices 4, 9, 11, 12 have on the target vertex (1). The diagram shows variable values of vertex 1, but we can see that factors 12 and 4 tend to have a negative influence; factor 11 has a positive influence, while factor 9 does not have a clearly defined tendency during the period under study. Let us now examine the effect of changes in transparency level on vertices 2, 3, 7 and 13 (see fig. 4).

We can see a positive dynamics of the influence that target vertex 1 had on vertices 2, 3, 7 and 13 throughout the period under study. Having analyzed structural characteristics of the cognitive map using FCMapper, we arrived at the following results. One vertex functions as a Transmitter, it is vertex 9, the level of technology development. Other vertices are Ordinary, there are no Receiver vertices in the model. Classification of the cognitive model vertices by these classes helps us better understand the structure of the graph. Having analyzed the cognitive model using the Decision Explorer Application, we identified two key elements of the cause-effect relationship system (based on the calculation of vertices centrality which reflects the strength of relationships between the vertices). The index that shows the proportion of existing connections of potentially possible (density) is calculated to be 0.14. It gets values within [0; 1]. The higher its value is, the more active interaction between the vertices takes place. They are the information and communication security (12) and information transparency (1). Using FCMapper software you can set the initial weights of vertices and investigate their influence on other system indicators. To quantify the system development dynamics, the influence of one vertex on the other is represented by the following set of values: very strong (0.8; 1]; strong (0.6; 0.8]; medium (0.4; 0.6]; weak (0.2; 0.4]; and very weak (0; 0.2]. During the situation self-development modeling (the total number of iterations is 60) the constructed model came to a stable state, which was achieved due to feedbacks, with the maximum number of iterations (51) for the degree of information accessibility to third parties (6) and risks (14), the fewest iterations (25) were required for the availability (5) and technology development (9) vertices. The calculations obtained indicate that the factor with the greatest impact on the system is the scope of information visibility (4). Information reliability (8) and approval to disseminate information (7) are found to be of high influence. The degrees of confidence (2), responsibility (3) and information availability (5) also showed significant influence. Then we analyzed the situation development scenarios under the change of the key elements of the cause-effect relationship system, namely the level of information and communication security (12) and information transparency (1). For the first scenario the initial vertex value was set at a low level (0.1), for the second scenario at a medium level (0.5) and for the third one at a high level (0.9). When the value of vertex 12 is of low and medium level, the most positive changes are observed for the following vertices: information transparency (1), the degree of responsibility (3), the degree of information accessibility to third parties (6), the level of corruption reduction (13) and risks (14). The high value of vertex 12 has the most positive influence on the level of information use efficiency (11). When the value of vertex 12 is of low and medium level, the most negative changes are observed for such vertices as the degree of information reliability (8) and the level of information use efficiency (11) The high value of vertex 12 has the most negative effect on the vertices of information transparency (1), the degree of responsibility (3), the degree of information accessibility to third parties (6), and risks (14). Now, we can have a close look at the influence of information transparency (1) on the cognitive model factors. In general, the level of information transparency does not affect the system significantly, it is 0.32 in the context of the system self-development scenario. At the same time, a decrease in the level of vertex 1 reduces the influence of such factors as the degree of responsibility (3) and the degree of confidence (2) on the system. An increase in the level of vertex 1 results in the growing influence of the corruption reduction factor (13), while low level of vertex 1 causes an increase in the level of corruption (13). A low level of vertex 1 slightly increases such factors as risks (14) and information accessibility to third parties (6), the value of the approval to disseminate information (7) increases when the level of vertex 1 is higher.

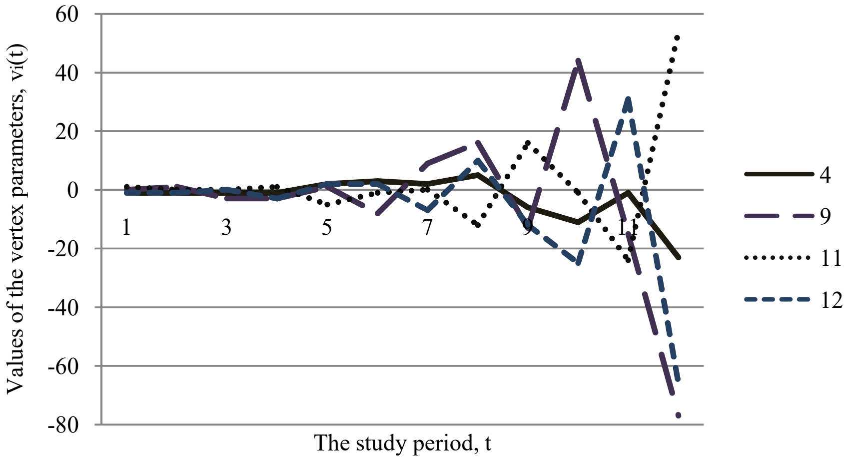


Fig. 3. The dynamics of the value of target vertex 1 under the influence of vertices 4, 9, 11, 12.

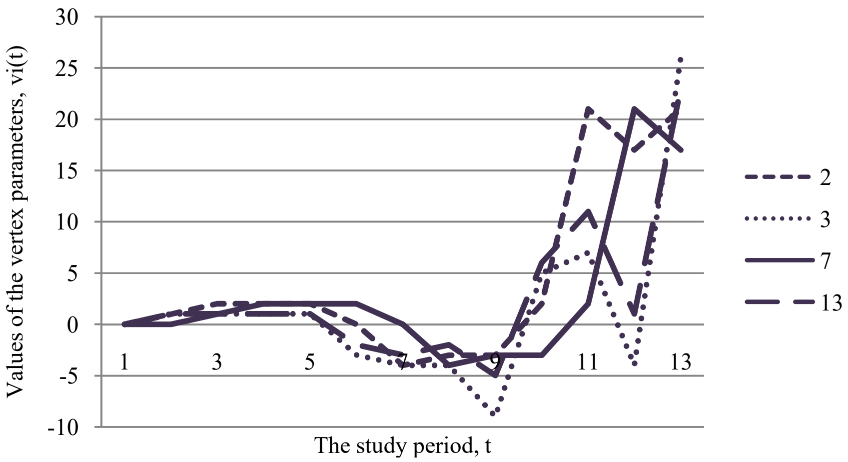


Fig. 4. The dynamics of the influence of target vertex 1 on vertices 2, 3, 7, 13.

Considering that the determining factors are characterized by different initial levels, we will investigate the created cognitive model, taking into account fuzzy logical conclusions regarding their influence on information transparency.

For research, we represent the cognitive model as a fuzzy map. For this, in the digraph, the vertices will be designated as linguistic variables [16]. The set of linguistic variables is characterized by the following parameter tuple:

$$<B_i, T, X, G, M>, \tag{6}$$

where: B_i – linguistic variables, $i = \overline{1,14}$, B_1 – "information transparency", B_2 – "the degree of confidence", B_3 – "the degree of responsibility", B_4 – "the scope of information visibility", B_5 – "the degree of information availability", B_6 – "the degree of information accessibility to third parties", B_7 – "approval to disseminate information", B_8 – "the degree of information reliability", B_9 – "the level of information and communication technologies development", B_{10} – "cognitive limitations of the user", B_{11} – "the level of information use efficiency", B_{12} – "the level of information and communication security", B_{13} – "the level of corruption reduction", B_{14} – "risk of information". T = {"low", "below average", "average", "above average", "high"}; X={information space}; G – procedure for the formation of new terms using logical connective "and", "or"; M – semantic procedure for the formation of fuzzy variables X, and the corresponding fuzzy set for the terms G(T) according to the translation rules fuzzy connective "and", "or".

In this case, T is the terms of these input and output variables of the fuzzy model, which are represented as fuzzy sets $T = \{(\alpha,\ \mu(\alpha)): \alpha \in X,\ \mu(\alpha) \in [0; 1]\}$ $T = \{(\alpha, \mu(\alpha)): \alpha \epsilon X, \mu(\alpha) \epsilon [0; 1]\}$, where α – elements of the set X, $\mu(\alpha)$ – the membership function of a fuzzy set [16]:

$$\mu(\alpha) = \begin{cases} \text{low} & \in [0;\ 0,2) \\ \text{below average} & \in [0,2;\ 0,37) \\ \text{average} & \in [0,37;\ 0,63) \\ \text{above average} & \in [0,63;\ 0,8) \\ \text{high} & \in [0,8;\ 1] \end{cases} \tag{7}$$

We present a linguistic description of the values of the factors and their measured values using the Harrington's desirability functions. The scale value is confined to the closed range of [0, 1]. Zero value corresponds to the worst measured factor state, and one corresponds to the best measured factor value. For a static analysis of the situation, we will calculate the consonance and dissonance of the cognitive map based on the research in [27]. Consonance determines how consistent the presence of factors in the system is. Dissonance determines how well-reasoned the influence of the system on each of the factors. The following table 3 gives system indicators of cognitive model.

The highest values of the consonance of the influence of a factor on the system are such indicators as the degree of confidence, the degree of information availability, the degree of information accessibility to third parties, approval to disseminate information, the level of corruption reduction. The digraph of the interaction of factors is shown below in fig. 5.

The highest values of the consonance of the influence of system on the factor are such indicators as the degree of information availability, the level of information and communication technologies development, cognitive limitations of the user. The digraph of the interaction of factors is shown below in fig. 6.

The highest values of the dissonance of the influence of system on the factor are such indicators as the degree of confidence and approval to disseminate information. Analysis of the dissonance of the influence of the system on the factor revealed the need to increase the degree of confidence in the medium and source of information, and the need to expand the boundaries of permission to disseminate information. All these actions, in general, lead to a decrease in dissonance. The highest values of the dissonance of the influence of factor on the system are such indicators as the level of information and communication technologies development and risk of information.

Table 3. The main system indicators of the cognitive model of information transparency.

Linguistic variables	Consonance of the influence of factor on the system	Consonance of the influence of system on the factor	Dissonance of the influence of factor on the system	Dissonance of the influence of system on the factor	Influencing of the factor on the system	Influencing of the system on the factor
B_1	0,867	0,621	0,133	0,379	0,150	-0,098
B_2	0,908	0,532	0,092	0,468	0,029	0,013
B_3	0,701	0,684	0,299	0,316	0,040	0,025
B_4	0,887	0,625	0,113	0,375	-0,113	0,161
B_5	0,908	1	0,092	0	-0,018	0,036
B_6	0,908	0,823	0,092	0,177	-0,018	-0,052
B_7	0,908	0,532	0,092	0,468	-0,018	0,045
B_8	0,662	0,666	0,338	0,334	0,018	0,068
B_9	0,486	1	0,514	0	0,071	0
B_{10}	0,640	1	0,360	0	0,085	-0,036
B_{11}	0,633	0,870	0,367	0,131	0,107	0,021
B_{12}	0,622	0,803	0,378	0,197	-0,107	0,050
B_{13}	0,908	0,641	0,092	0,359	0,052	-0,011
B_{14}	0,582	0,823	0,418	0,177	-0,107	-0,052

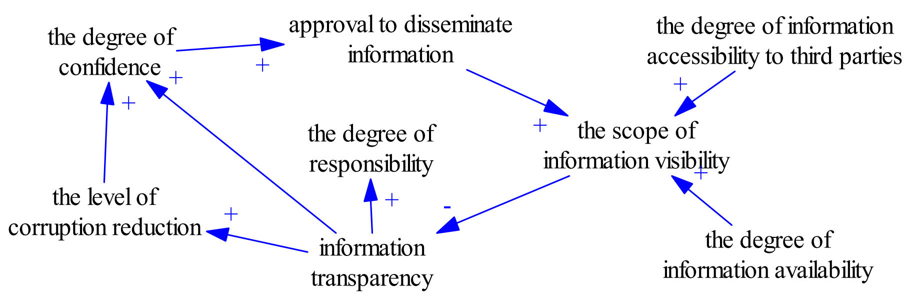


Fig. 5. Digraph of consonance of the influence of factor on the system (slice level 0,7).

Let us give a characteristic to each factor of the system:

B_1 – “information transparency” most strongly influences on the system among other factors (0.15), which is confirmed by the high value of the consonance of the influence of the factor on the system (0.867), but the system as a whole reduces the transparency of information;

B_2 – “the degree of confidence” influences on the system relatively weakly (0.029), the system, in turn, does not influence on a factor considerably (0.013). The dissonance of the system's influence on the factor is sufficiently high (0.468);

B_3 – “the degree of responsibility” - for this indicator, the consonances of the influence of the factor and the system are sufficiently high and have approximately equal values. This factor provides the strengthening of the system. This is an indicator of the prospect of strengthening the system due to the awareness of the need to disclose a certain amount and type of information;

B_4 – “the scope of information visibility” has significant negative influence on the system. But the system reinforces this factor. The value of the consonances of the influence of the factor on the system is quite high (0.887).

B_5 – “the degree of information availability” has a high value of the indicators of the consonances of the factor and the system, their values are approximately equal. It indicates that this factor strengthens the system. The prospect of strengthening is possible by expanding access to information. But the factor has a significant negative impact on the system;

B_6 – “the degree of information accessibility to third parties” weakens the system generally, and so does it. The values of the consonance of the influence of the factor and system are high;

B_7 – “approval to disseminate information” influences negatively on the system. The system does not influence significantly on a factor. The value of the consonance of influence of the factor is high. The value of the dissonance of the system's influence on the factor has a sufficiently high value in comparison with other vertices;

B_8 – “the degree of information reliability” the system influences this factor significantly and factor strengthens the system. The consonance of the factor and the system are equivalent, that is, it is necessary to strengthen the degree of reliability of the information;

B_9 – “the level of information and communication technologies development” - the system has no influence on this factor, but the dissonance of the influence of the factor on the system is of the highest;

B_{10} – “cognitive limitations of the user” the system affects the factor strongly and reduces the user's cognitive skills;

B_{11} – “the level of information use efficiency” has a significant impact on the system. The ratio of the consonances of influence indicates unused opportunities to increase the efficient use of information;

B_{12} – “the level of information and communication security” influences negatively on the system. The ratio of the consonances of influence indicates unused opportunities to increase the level of information and communication security

B_{13} – “the level of corruption reduction” the ratio of the consonances of influence indicates unused opportunities to reducing corruption;

B_{14} – “risk of information” has a significant negative impact on the system. The system diminishes the value of this factor. The dissonance of the influence of a factor on the system is of sufficient importance (0.418).

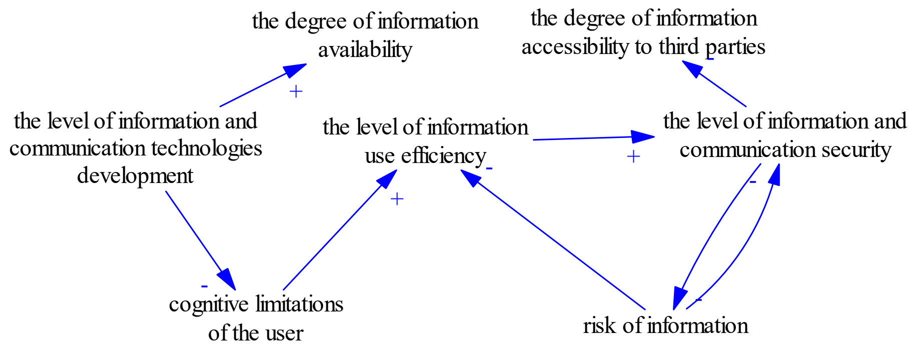


Fig. 6. Digraph of consonance of the influence of system on the factor (slice level 0,7).

A mathematical instrument of impulse processes is used to obtain a forecast of the development of the situation when implementing various alternatives. It allows you to predict the values of factors at discrete times. The "Igla" decision support system was used for modeling [28]. The initial values of the factors and the expected range of the initial values of the factors are determined by the following levels (see Table 4).

Table 4. Initial values of variables and expected area.

Variables	Initial value	Expected values
1. information transparency	"low"	"average"
2. the degree of confidence	"low"	"high"
3. the degree of responsibility	"low"	
4. the scope of information visibility	"average"	
5. the degree of information availability	"average"	
6. the degree of information accessibility to third parties	"average"	
7. approval to disseminate information	"average"	
8. the degree of information reliability	"average"	
9. the level of information and communication technologies development	"low"	
10. cognitive limitations of the user	"average"	
11. the level of information use efficiency	"below average"	-
12. the level of information and communication security	"below average"	-
13. the level of corruption reduction	"low"	"high"
14. risk of information	"high"	-

The program "Igla" has generated over 600 alternatives. The choice of alternatives is carried out per the purpose of the study. Let's consider the selected alternatives in more detail.

The diagram 7 shows the dynamics of the influence that alternatives 522, 537, 558, 567, 570 have on the target factor "information transparency".

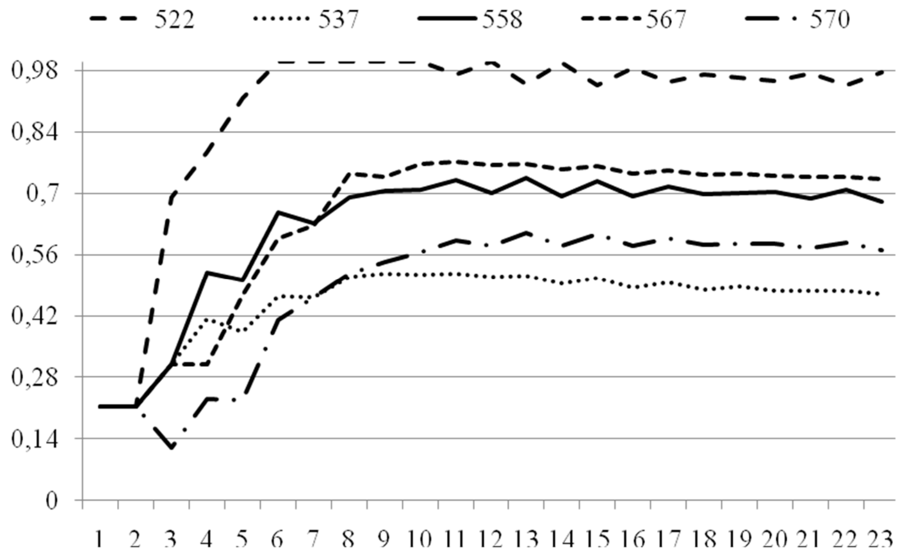


Fig. 7. The dynamics of the value of "information transparency" under the influence of alternatives 522, 537, 558, 567, 570.

The diagram 8 shows the dynamics of the influence that alternatives 574, 525, 542, 575, 541 have on the target factor "information transparency".

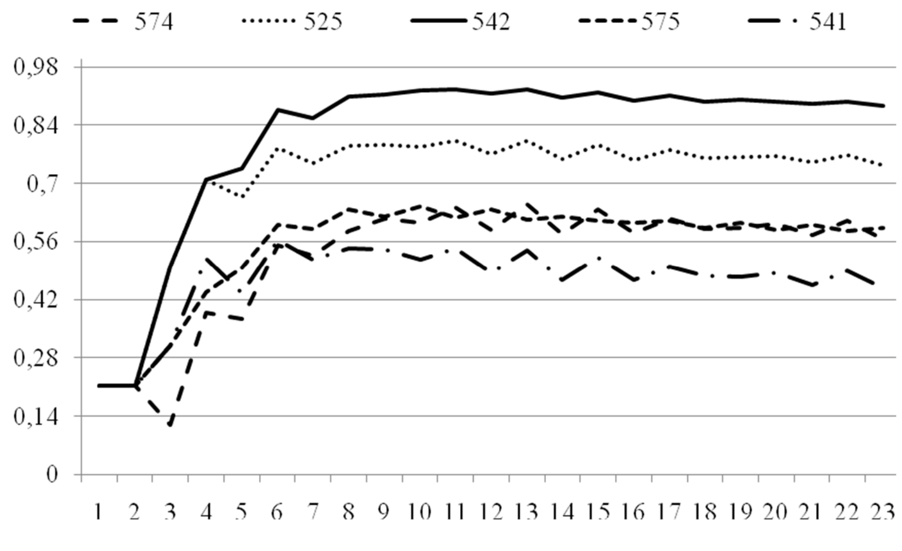


Fig. 8. The dynamics of the value of "information transparency" under the influence of alternatives 574, 525, 542, 575, 541.

Since previous research by specialists has shown that a high level of information transparency leads to an increase in the risk of misuse of information. Therefore, we consider those alternatives in which the level of information transparency corresponds to the average value. These are alternatives to 537 and 541.

The diagram 9 shows the dynamics of the influence that alternatives 537 and 541 have on the factor “the degree of confidence”.

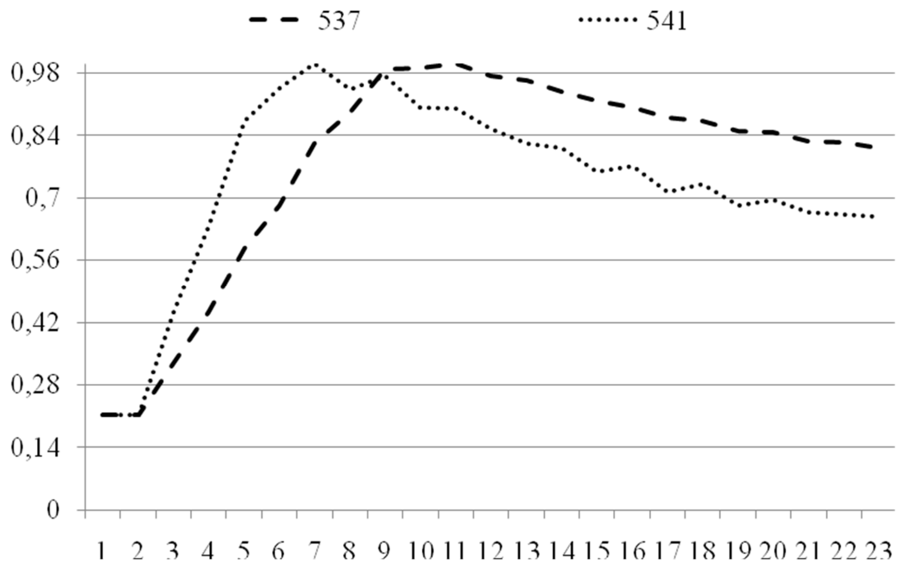


Fig. 9. The dynamics of the value of “the degree of confidence” under the influence of alternatives 537 and 541.

Thus, during the first 10 months, under the influence of both alternatives, the factor “the degree of confidence” gradually increases to a maximum. But in the following periods, under the influence of Alternative 537, the factor takes on a higher level than under the influence of Alternative 541. The diagram 10 shows the dynamics of the influence those alternatives 537 and 541 have on the factor “the level of corruption reduction”.

Thus, under the influence of both alternatives, the factor “the level of corruption reduction” gradually increases to a maximum. But under the influence of Alternative 537, the factor takes on a higher level than under the influence of Alternative 541.

Therefore, the best Alternative is 537, which meets the conditions. It was constructed as follows: in the first step and for one month it is necessary “the scope of information visibility” lower to “below average”, “the degree of information availability” increase to “above average”, “the level of information use efficiency” to leave at the level of “below average”. In general, it is necessary to increase the degree of information reliability and take measures to increase the level of information accessibility, including for third parties. Providing the required level of information reliability remains a difficult issue due to the weak structuring of this category.

Reliability is defined as a feature that allows to characterize the content of information on the presence of errors, distortions, biases, the degree of reflection of reality, the combination of which directly affects management decisions and their effectiveness. It influences the usefulness of financial statements. Its level is revealed due to additional, clarifying characteristics, in particular: "completeness, neutrality, discretion, the prevalence of essence over form and correct presentation" of information. Reliability is determined using other clarifying information, which is explained by the established relationship between information sources.

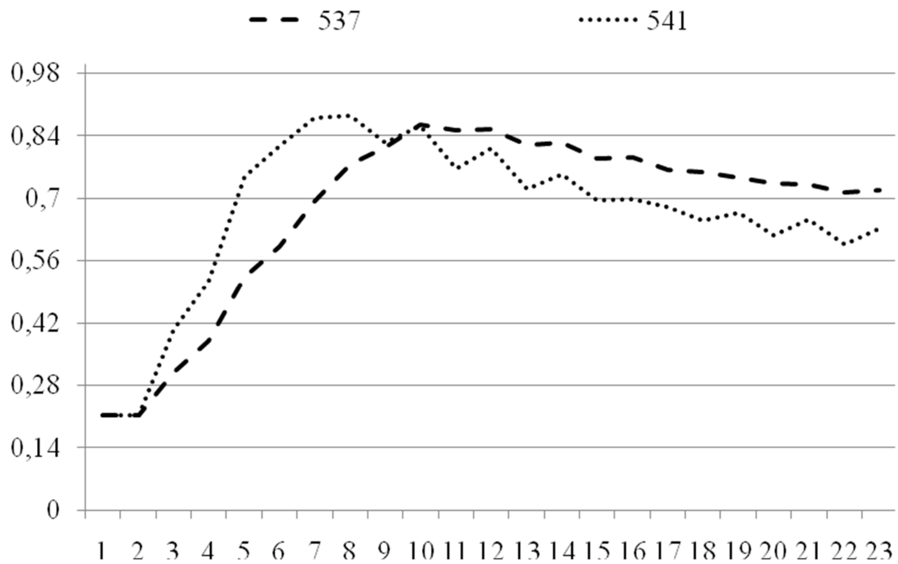


Fig. 10. The dynamics of the value of "the level of corruption reduction" under the influence of alternatives 537 and 541.

Reliability characterizes the authenticity of facts, phenomena, processes in terms of their subjectivity and variability of definition. It is determined according to the scale of the decision-maker by using available information, which may be limited. For its quantitative calculation, scientific methods and empirical research are used. The results of these studies are convincing enough to make the user feel trust to them. The expert verifies the information fact using expert research methods. The interpretation of the results of these studies is influenced by the subjectivity and competence of the evaluator.

Therefore, the results of assessing the reliability of information are taken into account with the assumptions that underlie the applied methods of expert research. First, they are regulated by regulatory and legislative acts for economic experts, both state and non-state. Second, the expert himself accepts the assumptions to generate conclusions.

Thus, the availability of information is a subjective-objective category. It is determined not only by the sphere of financial and economic relations since it is closely related to its reliability, persuasiveness of fact. But accessibility can only exist if there are an object and subject of research. The subject is an expert, based on his knowledge, skills, experience, abilities, considerations, as well as the information he has, makes decisions on the degree of correspondence of the fact being investigated to reality.

This category is not normatively regulated, it cannot be measured quantitatively, and expert methods are used to study it. Experts research this category in the context of a certain area, which is regulated by the relevant regulatory legal acts, and the object of which is the results of the financial and economic activities of the entity. But is it possible to talk about one hundred percent reliability of the results of the entity's activities presented in the financial statements. Since this reporting is made and evaluated by specialists who are subjective, and the influence of the human factor, opportunistic behavior in general, gives rise to the risk of errors, fraud, and the like.

Thus, we can say with a certain degree of probability that all information that has not been fully verified, and is not generated and processed automatically, is partially reliable.

The characteristic principle of reliability is manifested by establishing a minimum deviation between the actual data and the results obtained. It determines the close relationship of the studied category with the comparability of the data, and only together with the usefulness and materiality, the reliability ensures the quality of information.

In [22] it was proved that the reliability is divided into three categories. The first, internal, which reflects the generally accepted characteristics of knowledge about the obtained facts. The second, relative, which indicates the compliance of the facts with the requirements of the information user. The third, absolute, which determines the level of similarity of the facts with the really possible. Thus, information is characterized by the formal features of its construction, by the value for the person and by the reality of the existence of the given facts.

The reliability is determined by three criteria. The first is the validity that is the sources of information confirm the facts or phenomena under investigation. The next one is the consistency that is a fact or phenomenon in its manifestation does not create contradictions with other proven facts or phenomena. The third is the credibility, according to which the sources of information are verified and correspond to realities, and the information carrier is sufficiently protected. Therefore, the level of reliability can be assessed by the results of the ratio of the number of facts that do not correspond to reality to the total number of facts, by calculating the "probability of errors in the transmission of information". It is advisable to highlight the boundaries within which the results are assigned a certain degree of reliability: full or partial, or establish the level at which a fact or phenomenon is determined unreliable. In this case, it is necessary to take into account the specification of information, the sources of its formation, the method of analysis of results, users, the consequences of its use in different scenarios. Thus, information can become, in one situation and for one user, reliable and valuable, and in another situation and for another user – partially reliable, or in general – unreliable, invaluable.

5 Conclusion

It involves structuring knowledge about the factor environment, identifying strategic changes in the level of information transparency and the levels of influence of its factors. Information transparency is defined as a poorly structured category which nevertheless acts as a qualitative characteristic of information, a certain level of which forms an additional spectrum of properties of information which has been adequately perceived or processed. In the course of the study, a cognitive model of information transparency was constructed in the form of a weighted digraph. The results of its structural analysis revealed that the degree of transparency and information and communication security have the most powerful influence on the state of the system. The results showed that higher levels of information and communication security lead to lower risks, lower degree of information accessibility to third parties and information transparency in general. At the same time, only the high level of information and communication security is associated with an increase in the degree of information reliability, the level of visible information use efficiency and information transparency. The analysis of the cognitive model factors which affect the level of information transparency showed that its level is most significantly decreased by the growing level of information and communication security, while it is most significantly increased by the growing efficiency of visible information use, higher level of technological development and reduced scope of information visibility. The results of this study allow us to identify the strategic elements of managing information transparency as a tool for economic entities to achieve their goals in the information environment.

At the same time, it is necessary to take into account the initial levels of factors that influence information transparency. Because the researched categories are poorly structured, the methodology of fuzzy cognitive modeling was used. The fuzziness of the system is manifested in the level variation of factors and the relationships between them, which allows simulating the scenarios of its development under conditions of various levels of initial data.

For static analysis of the situation, the consonance and dissonance of the cognitive model are calculated, it is determined how justified the presence of factors in the system is and how well-argued the influence of the system on each of the factors is. Analysis of the dissonance of the influence of the system on the factor revealed the need to increase the degree of confidence in the medium and source of information, and the need to expand the boundaries of permission to disseminate information. All these actions, in general, lead to a decrease in dissonance. The highest values of the dissonance of the influence of factor on the system are such indicators as the level of information and communication technologies development and risk of information.

A mathematical instrument of impulse processes is used to obtain a forecast of the development of the situation when implementing various alternatives. It allows you to predict the values of factors at discrete times. The “Igla” decision support system was used for modeling scenarios of the situation, more than 600 alternatives were analyzed, in accordance with the purpose of research were chosen the best.

The best alternative is constructed as follows: in the first step and for one month it is necessary the factor “the scope of information visibility” lower to “below average”, "the degree of information availability" increase to "above average", "the level of information use efficiency" to leave at the level of "below average". It is necessary to increase the degree of information reliability and take measures to increase the level of information accessibility, including for third parties. Providing the required level of information reliability remains a difficult issue due to the weak structuring of this category and the associated risks, which are the focus of further research.

References

1. Bakurova, A.V.: Samoorganizacija social'no-jekonomicheskih sistem: modeli i metody (Self-organization of socio-economic systems: models and methods). KPU, Zaporizhzhia (2010)
2. Bellver, A., Kaufmann, D.: Transparenting transparency: initial empirics and policy applications. http://ssrn.com/abstract=808664 (2005). Accessed 25 Aug 2020
3. Bushman, R., Smith, A.: Transparency, financial accounting information, and corporate governance. SSRN Electronic Journal 9, 65–87 (2003)
4. FCMappers. http://www.fcmappers.net (2020). Accessed 25 Aug 2020
5. Havrylenko, M., Shiyko, V., Liliana, L., Khvostina, I., Yashcheritsyna, N.: Economic and mathematical modeling of industrial enterprise business model financial efficiency estimation. E3S Web of Conferences **166**, 13025 (2020). doi:10.1051/e3sconf/202016613025
6. Hosseini, M., Shahri, A., Phalp, K., Ali, R.: Transparency as a Requirement. In: Proc. 21st REFSQ Conference. Poster and Demo Track, Bournemouth University, pp. 222–223 http://ceur-ws.org/Vol-1342/03-Posters.pdf (2015). Accessed 17 Aug 2020
7. Hosseini, M., Shahri, A., Phalp, K.: A modelling language for transparency requirements in business information systems. In: Proc. of the 28th international conference on advanced information systems engineering (CAiSE), pp. 239–254 (2016)
8. Hosseini, M., Shahri, A., Phalp, K.: Four reference models for transparency requirements in information systems. Requirements Engineering **23**(2), 251–275 (2018). doi:10.1007/s00766-017-0265-y
9. Iorhachova, M.I.: Informatsiina prozorist kompanii – skladnyk modeli korporatyvnoho upravlinnia (Information transparency of the company is a component of the corporate governance model). Ekonomika ta suspilstvo 17, 543–550 (2018)
10. Islam, R.: Does more transparency go along with better governance? Economics and Politics **18**(2), 121–167 (2006)
11. Kaminskyi, A.B.: Informatsiina prozorist u diialnosti ISI: teoretychne obgruntuvannia i praktyka realizatsii (Information transparency in CII activities: theoretical justification and implementation practice). Finansy Ukrainy 11, 60–70 (2010)
12. Khvostina, I., Havadzyn, N., Horal, L., Yurchenko, N.: Emergent Properties Manifestation in the Risk Assessment of Oil and Gas Companies. CEUR Workshop Proceedings **2422**, 157–168 (2019)
13. Knottnerus, J.: Promoting transparency of research and data needs much more attention. Journal of Clinical Epidemiology **70**, 1–3 (2016)
14. Kucherova, H., Didenko, A., Kravets O.: Modeling information transparency of economic entities under uncertainty. Advances in Economics, Business and Management Research **99**, 201–205 (2019). doi:10.2991/mdsmes-19.2019.38

15. Kundeliene, K., Leitoniene, S.: Business Information Transparency: Causes and Evaluation Possibilities. Procedia – Social and Behavioral Sciences **213**, 340–344 (2015). doi:10.1016/j.sbspro.2015.11.548
16. Leonenkov, A.: Nechetkoe modelirovanie v srede MATLAB i fuzzy TECH (Fuzzy modeling in MATLAB and fuzzy TECH). BHV-Peterburg, Russia (2005)
17. Los, V.O., Ocheretin, D.V.: Cognitive modeling of tourism development. Business Inform 10(477), 166–173 (2017)
18. Measuring the Information Society Report 2018.1. https://www.itu.int/en/ITU-D/Statistics/Pages/publications/misr2018.aspx (2018). Accessed 17 Aug 2020
19. Oliinyk, V., Kozmenko O.: Optimization of investment portfolio management. Serbian Journal of Management **14**(2), 373–387 (2019). doi: 10.5937/sjm14-16806
20. Olijnyk, O.O.: Naprjamy pidvyshhennja prozorosti dijaljnosti orghanizacij (Areas of increase of transparency of activity of organizations). Socialjno-trudovi vidnosyny: teorija ta praktyka 1, 224–230 (2017)
21. Osoba, O., Kosko, B.: Fuzzy cognitive maps of public support for insurgency and terrorism. Journal of Defense Modeling and Simulation: Applications, Methodology, Technology **14**(1), 17–32 (2017)
22. Pierce, R.: Evaluating information: Validity, reliability, accuracy, triangulation. In: Research Methods in Politics: A Practical Guide. Sage Publications (2008)
23. Roberts, F.: Discrete mathematical models, with applications to social, biological, and environmental problems. Prentice-Hall, Englewood Cliffs (1976)
24. Schauer, F.: Transparency in Three Dimensions, University of Illinois Law Review 4, 1339–1357 (2011)
25. Schnackenberg, A., Tomlinson, E.: Organizational transparency: A new perspective on managing trust in organization-stakeholder relationships. Journal of Management 42.7, 1784–1810 (2016)
26. Sheng, Y.K.: What is Good Governance? United Nations Economic and Social Commission for Asia and the Pacific. United Nations Economic and Social Commission for Asia and the Pacific. https://www.unescap.org/sites/default/files/good-governance.pdf (2009). Accessed 17 Aug 2020
27. Silov, V.B.: Prinjatie strategicheskih reshenij v nechetkoj obstanovke (Making strategic decisions in a fuzzy environment). INPRO-RES, Moscow (1995)
28. Sistema podderzhki prinjatija reshenij "Igla" (Decision support system "Igla"). http://iipo.tu-bryansk.ru/quill/index.html (2020). Accessed 25 Aug 2020
29. Stohl, C., Stohl, M., Leonard, P.: Managing Opacity: Information Visibility and the Paradox of Transparency in the Digital Age. International Journal of Communication 10, 123–137 (2016)
30. Tkach, I.V.: Pryntsypy prozorosti ta vidkrytosti yak nevidiemna skladova publichnoho upravlinnia: teoretychnyi aspekt (The principles of transparency and openness as an integral part of public administration: a theoretical dimension). Ynternauka 12(22), 25–32 (2016)
31. Vaccaro, A.: Privacy, security, and transparency: ICT-related ethical perspectives and contrasts in contemporary firms. Social inclusion: Societal and organizational implications for information systems 208, 245–258 (2006)
32. Vishwanath, T., Kaufmann, D.: Toward transparency: New approaches and their application to financial markets. World Bank Research Observer 16, 41–57 (2001)
33. Williams, A.: A global index of information transparency and accountability. Journal of Comparative Economics **43**, 804–824 (2014). doi:10.1016/j.jce. 2014.10.004
34. Zachosova, N.V., Ghorjachkivsjka, I.V.: Informacijna prozoristj ta zabezpechennja interfejsnoji bezpeky jak elementy poperedzhennja ta protydiji zlochynam u ekonomichnij

sferi na mizhnarodnomu rivni (Information transparency and provision of interface security as elements of the prevention and counteraction to economic crime at the international level). Naukovyj visnyk Uzhghorodsjkogho nacionaljnogho universytetu 8(1), 90–93 (2016)
35. Zinchenko, A., Reznik, N., Saprykina, M.K.: Indeks prozorosti sajtiv ukrayins`ky`x kompanij 2017 (Transparency Index of Ukrainian companies' sites). Yuston, Kyiv (2018)

Use of simulation modeling for predicting optimization of repair works at oil and gas production enterprises

Iryna Hobyr[1[0000-0002-9322-8404]], Vitalina Babenko[2[0000-0002-4816-4579]], Sofiia Kafka[1[0000-0002-6031-383X]], Yuliia Bui[1[0000-0003-0096-2384]], Oksana Savko[1[0000-0002-7282-7731]] and Ekaterina Shmeltser[3[0000-0001-6830-8747]]

[1] Ivano-Frankivsk National Technical University of Oil and Gas,
15 Karpatska Str., Ivano-Frankivsk, 76000, Ukraine
irynagobyr@gmail.com, kafka@i.ua, julia03091980@gmail.com,
oksana_savko19@ukr.net
[2] V. N. Karazin Kharkiv National University, 4 Svobody Sq., Kharkiv, 61022, Ukraine
vitalinababenko@karazin.ua
[3] State University of Economics and Technology,
5 Stepana Tilhy Str., Kryvyi Rih, 50006, Ukraine
shmelka0402@gmail.com

Abstract. There is substantiated the expediency of using the methods of the queuing theory for supporting various business processes at enterprises. The task of mass service for the organization and management of repair services of oil and gas producing enterprises has been set. It is stated that the objective function of solving this task is the sum of costs for maintaining repair crews and the reduction of losses from well downtime. It is proved that to solve the problem of optimizing the process of repair work at the researched enterprises it is necessary to use modeling as one of the effective tools of prediction, which do not require to bring the researched models to a specific form and allows to predict systems in different states and industries. A simulation model of the organization of repair works in wells is given. Testing of simulation models is carried out on the example of the fields of the three largest oil and gas producing enterprises in the Western region of Ukraine, it allowed to establish the laws of distribution of failures of oilfield equipment on fields, the process of which includes operations related to the creation, transformation and implementation of random events, quantities and processes that cause random changes in the state of the system. And also, to make calculations of the optimal between-repairs periods, total costs of well servicing processes, losses from their downtimes and optimum quantity of crews at various options of intensity of failures.

Keywords: oil and gas companies, queuing theory, repair crews, repairs, simulation and prediction, wells.

1 Introduction

Rational organization of maintenance and repair is a determining factor in improving the use of equipment, improving the quality and reducing the cost of production, which generally contributes to the sustainability of the enterprise.

The need to increase the productivity of equipment while maintaining the maximum possible level of reliability in the enterprise and reducing the cost of oil and gas production requires improvement of the system of technical exploitation of equipment.

When assessing the various forms and methods of maintenance and repair, it is necessary to take into account technical and technological, organizational, economic and social factors. The technical and technological factors include the technological complexity of the assessed type of repair works and the technical characteristics of the machinery and equipment required for their performance, qualified personnel requirements. Analysis of the technical feasibility of options for the organization of repair and maintenance is to establish the possibility of work and operations in full compliance with the technical conditions, to identify the advantages and disadvantages of the options for the organization in terms of improving equipment and technology.

Maintenance could be preventive or predictive and it is aimed at applying knowledge of engineering and maintenance capabilities to ensure improved performance and reduction in their rate of failure. Furthermore, this maintenance could include periodic examination in terms of regular walk through to spot change in operating conditions, sounds or temperature. It also includes inspection programs, lubrications among others [13].

Organizational factors include the need for repair services, the estimated volume of specific services, the presence or absence of necessary to perform certain services production capacity, the degree of use of production capacity of repair services, the ability to perform certain services (works) in order to replenish existing production capacity and personnel.

Economic factors include: the level of costs for services and their payment by the consumer; volume of capital investments and sources and conditions of their financing; implementation of certain tax reliefs in relation to enterprises and entrepreneurs providing production and technical services.

To increase the efficiency of oil and gas production enterprises, it is necessary to ensure the continuity of the oil and gas production process, which directly depends on the level, maintenance and repair of wells. With aging equipment and facilities, enterprises face increasing challenges in maintaining equipment reliability and integrity as well as operational safety [8]. That why the organization of special maintenance and repair (current and capital) is associated with the wear of operating equipment and the need for a set of special measures for its restoration. At the same time, the development and operation of oil and gas fields in difficult conditions requires the development of advanced forms of service, which ultimately lead to increased efficiency of oil production, profit growth and cost reduction. That is why increasing the efficiency of repair work on deposits is an urgent problem that has both scientific and practical value.

For oil and gas production enterprises, the optimization and prediction of repair services in wells requires the use of scientific methods to model situations that arise in oil and gas production close to reality. This will allow us to choose possible management strategies using different criteria because when optimizing the operation of the well fund it is important to take into account the economic feasibility of applying one or another strategy depending on the number of repair crews and the duration of the repair period.

2 Features of repair works at oil and gas production enterprises

Oil and gas production enterprises are an integral part of the oil and gas complex of Ukraine. They have a significant social and economic impact on the living standard of the country, providing the population, budgetary and communal institutions with natural gas. Their functioning is accompanied by a complex technological and production cycle and significant risks associated with their activities [9].

The continuity of the oil and gas production process largely depends on the operation, maintenance and repair of wells. The need to organize special maintenance and repair of production wells is associated with the wear and tear of operating equipment and the need to implement a set of special measures for the protection of subsoil. The main task of repair units of oil and gas production enterprises is to maintain in working order the operational fund of wells, to prevent wear and tear of equipment that can cause accidents and downtime of wells. Work on repair maintenance of wells involves current repairs and overhauls of wells and maintenance of operating equipment in the period between regular repairs.

The scientific literature identifies the following reasons that necessitate repair works in oil wells [2; 11; 17]:

- the need to re-equip wells in accordance with the conditions of field development;
- the need for planned periodic stops during long-term operation to prevent equipment, change the mode of operation, method of operation or purpose of the well, geological and technical measures to increase the productivity of the well;
- the need to restore normal operation of wells in the presence of various complications (deposits: salts, paraffin, asphalting-resinous substances, water manifestations, formation of hydrates, corrosion, etc.);
- requirements of technology of rational development of a deposit or formation;
- non-compliance of the construction of wells with the conditions of operation and development of the field, requirements for subsoil and environmental protection;
- occurrence of emergencies or accidents related to well equipment, research instruments, etc.

One of the ways to solve the problem of ensuring the effective organization of any work is to improve the system of their planning, accounting and monitoring. The most important tool of this system is the classifier, which is a systematic list of named objects, each of which is given a unique code. The classification of objects is carried out in accordance with the rules of distribution of a given set of objects into classification groups in accordance with the established signs of their differences or similarities.

Current (underground) repair of wells is called a set of works, which is associated with maintaining well equipment in working order and maintaining a given mode of operation of wells. The following works belong to the current repair [2]:

- equipping the well with underground equipment during commissioning or changing the method of operation;
- ensuring the optimization of operation of wells;
- restoration of serviceability of downhole and mouth equipment;
- cleaning and washing of the lifting column of pipes from paraffin and salt deposits, hydrate and sand plugs.

Overhaul of wells is a set of complex and long-term works related to maintaining the well as a building in working order, i.e. it is a set of works designed to restore the serviceability of the bottomhole zone, the integrity of the cement ring and casing or the well in general for its use as existing or new purpose, elimination of accidents, conservation and liquidation of wells, lowering and raising of equipment for simultaneous and separate operation, as well as for actions on the productive layer and regulation of field development. The average duration of one overhaul is approximately 14 days. Overhaul of wells includes [2]:

- repair and insulation works;
- elimination of leaks in the operating column;
- elimination of accidents;
- transition to other layers and joining layers;
- repair of wells equipped for simultaneous and separate operation;
- works related to drilling;
- processing of the bottomhole zone;
- research and inspection of wells during repairs;
- transfer of wells for use for other purposes;
- commissioning and repair of injection wells; work on injection of radioactive isotopes into the formation;
- conservation and deconservation of wells;
- elimination of wells.

Repair works in wells also include well-repair operations to increase oil and gas extraction from layer. This is a set of works in the well for the injection into the layer of agents that stimulate the flow of physical, chemical or biochemical processes aimed at increasing the coefficient of final oil and gas layer in a given area of the deposit.

In addition, repair works are further subdivided depending on the category of repair complexity. Current repairs, depending on the depth to which the equipment is lowered into the well, and the types of work performed are divided into two categories of repair complexity. The first category, in contrast to the second, contains simpler types of repairs, as well as more complex works at small depths of equipment lowering (up to 700 – 1600 m, depending on the type of works).

Overhauls of wells depending on the volume, nature and complexity of work are also divided into two categories of complexity. Overhauls of wells of the first category of complexity include repairs with a depth of wells up to 1500 m, as well as return to the upper and lower layers and hydrochloric acid treatments. Repairs of the second category of complexity include works with a depth of wells over 1500 m and all other more complex and time-consuming works. Specific categories of complexity are approved by the company.

Current repairs and overhauls of oil and gas production enterprises are carried out by economic and contractual methods. In the case of the economic method, the repair is carried out by oil and gas production enterprises by, respectively, maintenance and wells of specialized shops of current repair of wells or shops of current and overhaul of wells, as well as crews of overhaul of wells.

Current repairs and overhauls are necessary to ensure the continuity of oil and gas production, and they cannot be avoided, as there is a risk of accidents, the elimination of which requires much more time and money.

The reserve for reducing unproductive working time costs of well overhaul and current well repair crews is combining professions, which will allow them to perform part of non-repair work carried out during overhaul and underground well repairs. It will significantly reduce unproductive time due to the absence or malfunction of the lifting unit of the technological fleet of equipment for crews of current repairs and crews of overhaul of wells, as more than 75% of the equipment has already exhausted its service life.

In order to optimize the current repairs of wells and overhauls of wells, it is necessary to involve third-party organizations in their implementation, as the crews of enterprises can not perform all the necessary work on time according to the schedules of current repairs and overhauls of wells. The current trend of involving well overhaul crews in the implementation of current repairs has a negative impact on the volume of well overhauls.

The largest number of current repairs is accounted for oil wells. The main reasons for current repairs of wells are the operation of the pump valves, wear of the steam plunger-bushing, displacement of the bushing. In all methods of well operation, and most of all in the development using rod-deep pumps, in the extraction of oils with a high content of paraffin in the risers and discharge lines, paraffin accumulates, which disrupts the normal operation of wells. A study of the volume of underground repairs, depending on the method of operation, shows that the largest number of underground repairs are performed on wells equipped with rod-deep pumps. This is due to the fact that the rod-deep pumps operate in an aggressive environment with water content of the fields at 60% and above and a significant salt content. Most of the current repairs of wells with rod pumps are related to the need to eliminate breakage of the rods and various seizures of the tubing, because some of the rods work in a corrosive environment in which the metal is corroded.

Optimizing the organization of repair maintenance of wells requires the introduction of a system of planned and preventive repairs and distribution of work between structural units of enterprises and third-party organizations, which, ultimately, will minimize costs and quality work. The combination of repair work with intensification of oil and gas production should be continued, as it allows to achieve time savings through preparatory and final works, well silencing works, lowering and lifting operations, which leads to an increase in total oil and gas production and savings on some types of ancillary work. A significant reserve for improving the organization of repair services is the specialization in the performance of preparatory and final works of one of the teams and the improvement of technical support by updating the fleet of repair equipment. This will minimize downtime of repair crews and increase the share and amount of time of direct repair work in the overall balance of the calendar time of crews of overhaul and current repair of wells.

3 The task of mass service for the organization and management of repair services of oil and gas production enterprises

Making complex managerial decisions involves predicting scenarios of events and consequences of taking certain decisions for the activities of enterprises [18]. Therefore, to ensure undisturbed operation of equipment at enterprises [14], support a variety of business processes [6] and different sectors of the economy [16] the theory of queuing is widely used.

Since an insufficient number of repair crews at oil and gas production enterprises or untimely maintenance can lead to well downtime and significant damage, therefore well repair by crews can be considered as a multi-channel queuing system with applications for repairs as a stream of service requirements. The maintenance procedure consists in the allocation of a repair crew according to the application. The application is considered to be satisfied after the completion of the work of the crew at the facility. There is a queue of applications to be satisfied in the order they are received. The application, which is received when all the repair crews are busy, remains in the system until it is satisfied.

When setting the queuing problem for the organization and management of repair services of oil and gas companies, it is assumed that the system has N repair teams (absolutely reliable service channels). Demand for repair crews is characterized by the flow of applications from individual wells requiring the j^{th} type of work. The intensity of the flow of applications for the repair team from the i^{th} well, requires the j^{th} type of work (the average number of applications per unit time), is λ_{ij}. The duration of the work by the repair team at the i^{th} well, which requires the j^{th} type of work, is t_{ij} days. It is assumed that t_{ij} is a random variable having one or another distribution law. It is also necessary to enter a value $\mu = 1/T_{ij}$ – the intensity of servicing applications, where T_{ij} is the average duration of the crew's work at the facility.

The condition for the existence of a stationary mode of operation of the queuing system (when the queue of service applications is stabilized) is the inequality as follows

$$N > \rho = \lambda/\mu \tag{1}$$

where $\lambda = (b/a)(t/a)^{b-1}$ is the rate of unprocessed applications.

The parameter ρ, which is called the loading of the service channel, plays a pivotal role in any characteristics of the system: the greater is the load and the closer is the performance boundary, the longer is the average time of requirements in the system, that is, there are more downtimes of wells.

The objective function in solving the problem of substantiating the need for repair crews is the average amount of expenses for their maintenance and losses from downtime of wells:

$$C_{\Sigma_i} = (C_i^{const} N_i + C_i^{var} \bar{N}_i) + \sum_{j=1}^{n} Q_j \bar{M}_j, i = \overline{1,m}, \tag{2}$$

where C_i^{const} are fixed costs of an hour work of the i^{th} crew;

N_i is the number of i repair crews;

C_i^{var} are semi-variable costs of an hour work of the i^{th} crew;

$\bar{N}_i$ is the average number of working i teams, equal to ρ;

Q_j are hourly average losses from downtime of the j^{th} well;

We assume that the sequence of moments of applications receipt for repair crews is a Poisson stream, and the duration of their work during well repairs is distributed exponentially. If condition (2) is satisfied, then the average number of idle wells is as follows

$$\bar{M}_j = \frac{\rho^{N_i+1}}{(N_i-1)!(N_i-\rho)^2} p_0 \tag{3}$$

The value p_0 is the probability of simultaneous downtime of all wells, which is determined by the formula:

$$p_0 = \frac{1}{\sum_{k=0}^{N_i-1} \frac{\rho^k}{k!} + \frac{\rho^{N_i}}{N_i!(1-\rho/N_i)}} \tag{4}$$

In fact, the flow of applications is not always subject to Poisson distribution law, or there are no assumptions regarding the law of distribution. In this case, simulation can be applied for the rational management.

4 Prediction of optimizing the repair works using simulation

Simulation is widely used to predict the activities of the enterprise and select an acceptable scenario, as well as to minimize possible risks. It is used to overcome stochasticity [10], to plan the operation of production systems [4; 5], allows you to model the operation of systems in different industries [15] and make optimal management decisions [12].

This approach has considerable flexibility, and the model obtained as a result of a simulation experiment can be brought into line with the system under study [1; 3].

The main reasons for the popularity of the simulation method in relation to queuing systems are:

- high adequacy between the physical essence of the described process and its mathematical model;
- the ability to model a much wider class of systems compared to analytical methods;

- the ability to model the operation of systems with different laws of distribution of numerous random variables;
- modeling the functioning of systems not only in steady but also in transient modes;
- the possibility of obtaining as a result of modeling more meaningful information, including numerical characteristics of the laws of distribution of random variables that describe the operation of the system.

The simulation process includes a significant number of operations related to the formation, transformation and description of the occurrence of random events, random variables and random processes that accompany random changes in the states of the system. Therefore, the trajectories of the motion of a complex system in the set of its states, which are obtained by imitation, are random functions of time. The source material for making any random objects is the so-called random numbers specified by the random number sensor. Random numbers can be considered as possible values x_i of a random variable ξ, approximately described by the normal distribution law in the interval (0; 1). The methods of constructing various complex random objects from random numbers have been sufficiently developed at present and are called the conversion of random numbers.

If there are no free repair crews in the system to meet the next application, a downtime of the i^{th} well arises, and it requires the completion of the j type of work, the duration of which is τ_{ij} in days and depends on the moment the repair crew is free having fulfilled a previous task (application).

The objective function of the mathematical model of the process of providing wells with repair crews is as follows

$$C_{\Sigma} = C_a N + \sum_{t=0}^{T_c} \sum_{j=1}^{n} \sum_{i=1}^{m} C_{ij} \tau_{ij} \to \min, \tag{5}$$

where C_a are average annual maintenance costs for crews;

C_{ij} are the average daily losses from the downtime of the i^{th} well that requires the completion of the j^{th} type of work;

T_c is the moment of the process stabilization.

In case of simulation, the state of the system at the next moment in time depends on its state at the previous moment. At the beginning of the simulation, it is necessary to set the initial state of the system, which is unknown. If we accept that all the repair crews are free at the initial moment, then it takes some time to get an objective picture of the use of the crews. We call it the duration of the process stabilization and assume that the process has reached stabilization if the relative downtimes of the crews have stabilized, that is, the following condition is fulfilled

$$0 < \Delta\tau(T_c) - \Delta\tau(T_{c-1}) < \varepsilon \tag{6}$$

where $\Delta\tau(T_c)$ is an accumulated relative duration of downtime by the crews at the moment T_c:

$$\Delta\tau(T_c)=\frac{\sum_{t=0}^{T_c}\sum_{j=1}^{n}\sum_{i=1}^{m}\tau_{ij}}{T_c} \quad (7)$$

where ε is the predetermined value.

The value ε determines the accuracy of calculations in the simulation process. The reduction in formula (6) is the relative downtime of the crews at the moment of the process stabilization, and the denominator is the same at the previous moment of time. The difference must be non-zero, otherwise, due to the fact that all crews are free and the downtimes are equal to zero level at the beginning of the simulation, it will end without starting.

Due to the fact that the cost of maintaining repair crews is calculated on an annual average, the simulation should last until the time equal to one year. An additional condition is therefore required

$$T_c < T_r \quad (8)$$

where T_r is one year in units of measurement of time modeling.

If condition (8) is not fulfilled, the simulation should be extended to two, three years and, if necessary, further, which should be taken into account when determining the scale of maintenance costs for repair crews.

Expressions (1)-(8) constitute a mathematical model of the well service process, where expression (5) is a goal function and others are constraints.

To solve the problem, we need the following data:

- a list of wells that should be serviced by repair crews;
- types and parameters of laws for the distribution of applications for crews and the duration of repair works for each well;
- losses from well downtime per unit of time;
- average annual maintenance costs for one crew;
- the number of repair crews (the minimum number of crews is determined on the basis of condition (1) and then increases by one at each calculation cycle until the value of the objective function (5) begins to decrease);
- the value ε, the moment of completion of calculations.

5 Algorithm for modeling repair work in wells

In fig. 1 unit 1 simulates the time of the next application for the crew of the j^{th} well. For this, the random number sensor provides a random number, which is uniformly distributed in the interval [0, 1]. Depending on the parameters of the selected law for the distribution of applications for repair crews according to a special program, this random number turns into a random number corresponding to a given distribution law.

Unit 2 chooses the moment of the earliest application. This moment is at the same time the next moment of modeling, that is, the program “goes” at this moment into the real timescale.

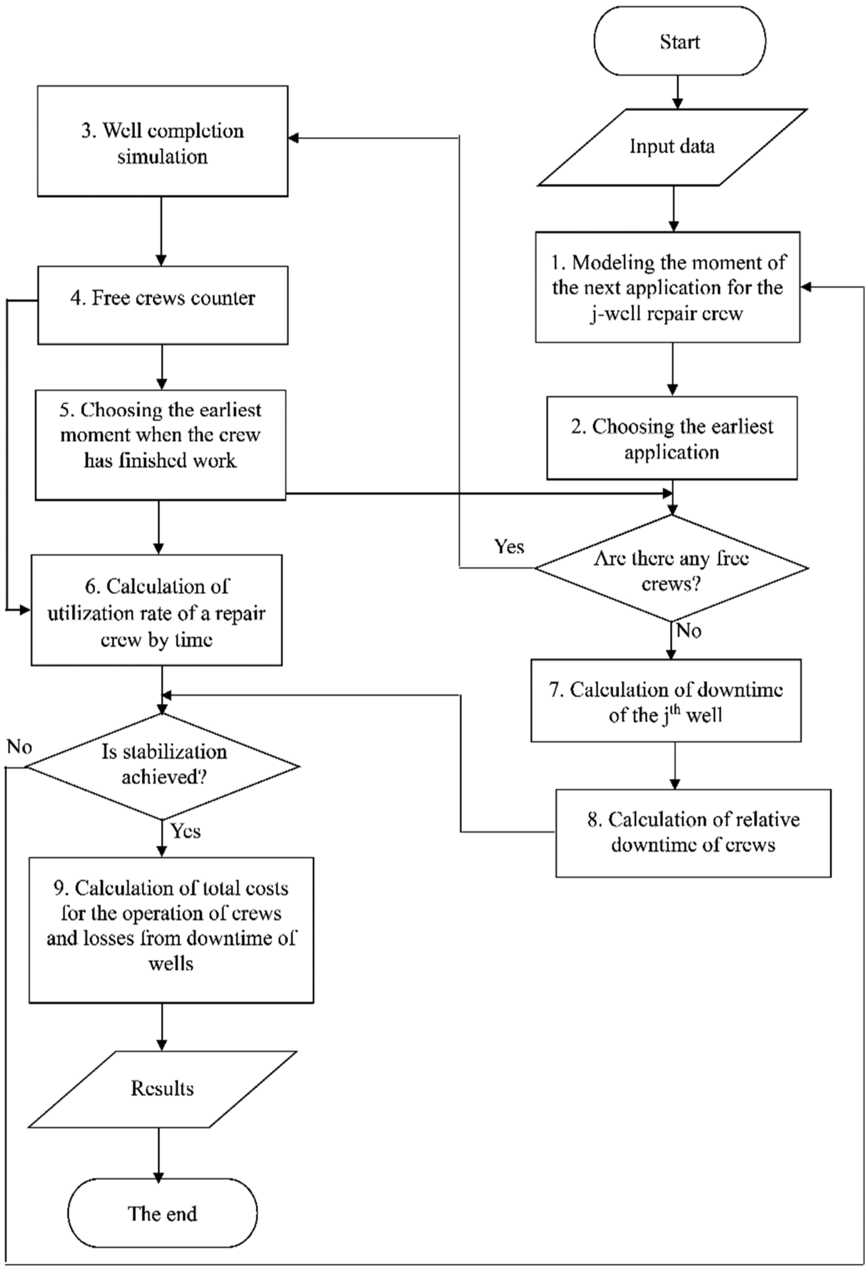


Fig. 1. A simulation model of the organization of repair works in wells

The logical operator determines the availability of free crews. If there is a free crew, it is allocated to satisfy the application in question. The application is considered to be satisfied, and the repair team – busy with repair works until the end of work at the facility, which is modeled by unit 3.

Unit 3 functions similarly to unit 1, but converts a uniformly distributed random number in accordance with the distribution law of the duration of the repair crew work on the j_{th} well. Information on the work of the repair crew is transmitted to unit 4, just like the information being produced by unit 5.

Unit 6 determines the utilization rate of the repair crew in time as the ratio of the number of busy crews to the total number of repair crews N to determine the moment of the process stabilization under the condition similar to condition (6).

When the logical unit determines the absence of crews during the next application for the repair work, the unit 7 comes into operation, which determines the downtime of the well as the difference between the earliest time the crew is free and the moment the application arrived.

Unit 8 calculates the relative duration of well downtime, and the logical unit determines whether stabilization of the process has been achieved under conditions (6) and (8). If the answer to this question is negative, the modeling procedure continues; if the answer is positive, block 9 calculates the total costs for maintaining the crews, losses from well downtime and determines the value of the objective function by summing the cost components (5).

The components of cost and the value of the objective function (5) are printed, after that adjustments are made to the initial data, and the specified number of crews is increased by one. This continues until the value of the objective function decreases. Its increase indicates that the optimal number of repair crews is obtained at the previous stage of modeling.

An exhaustive characteristic of the reliability of oilfield equipment is the law of distribution of time between failure. If the form of this law and its parameters are known, then it is relatively easy to determine any characteristic of the operational reliability of the system. A statistical definition of the law of uptime is associated with significant difficulties.

To set it, according to the statistical data, graphs of statistical functions of the reliability indicator are built. Since the density of distribution more clearly reflects the specific features of the distribution law, then, as a rule, this function is first built so that one could assume the form of the distribution law according to its form.

When constructing the statistical function of the distribution density along the abscissa, the intervals of the statistical series are laid aside. At each interval of the statistical series there is a rectangle, the height of which is equal to the ratio of the frequency to the width of the interval.

When processing statistical material, an important task is the selection of the theoretical distribution law, which expresses its essential properties without an element of randomness.

The theoretical law is chosen, taking also into account:

- the physical nature of failures;
- the shape of the distribution density curve;
- the coefficient of variation.

The value of the coefficient of variation is determined by the formula

$$V = \frac{\sigma}{\bar{t}} \tag{9}$$

where σ is the standard deviation; $\bar{t}$ is the average value of the system uptime, allowing to judge the operating conditions of oilfield equipment.

It is known that if a coefficient of variation is $V < 0,3$, a normal distribution takes place, and if $V > 0,5$, there is a Weibull distribution. The Weibull distribution function has the following form

$$F(t) = 1 - \exp[-(t / a)^b] \tag{10}$$

where a i b are Weibull distribution parameters.

The parameter b can be determined through the coefficient of variation. The parameter a is calculated from the expression $a = \bar{t} / K_b$ or $a = \sigma / C_b$, where K_b and C_b are the coefficients determined with the known coefficient of variation. When $b = 1$, the Weibull distribution transforms to the exponential, and when $b = 2,5 – 3,5$ it is close to the normal one. Therefore, the Weibull distribution is considered as a very flexible law and is widely used in reliability theory. Obviously, the Weibull distribution takes place for equipment failures resulting from its wear and aging, as well as for device failures consisting of series-connected elements.

Thus, the possibilities of using the methods of the theory of queuing in decision-making on the organization and management of repair services at oil and gas production enterprises are substantiated. The proposed simulation model makes it possible to optimize the organization of repair work at oil and gas production enterprises, and its implementation reduces the cost of maintenance of repair crews and reduces losses from well downtime.

6 Application of models of organization of repair work

Practical application of the proposed models of the organization of repair works has been carried out on the example of the fields of the largest oil and gas production enterprises (OGPE) of the public joint-stock company Ukrnafta - OGPE Boryslavnaftogaz, OGPE Dolynanaftogaz and OGPE Nadvirna Naftogaz. Based on the results of processing the field data, there were established the laws for the distribution of failures of oilfield equipment in the fields, described by the Weibull distribution.

So, breaks in tubing in oil and gas fields obey the law of probability of failure-free operation, which is described by the expression

$$P(t) = \exp\left[-\left(\frac{t}{496}\right)^{1,8}\right] \tag{11}$$

Figure 2 shows the probability distribution functions of the failure-free operation and the failure rate of tubing (breaks) at the fields of OGPE Dolynanaftogaz.

The following parameter values are accepted for calculations:

T_{ij} = 24 days – the average duration of a repair crew use at the facility;
C_i^{const} = 51,5 conventional units – fixed costs in the cost of an hour work by the crew;
C_i^{var} = 49,5 conventional units – conditionally variable costs in the cost of an hour work by the crew;
Q_j = 620,3 conventional units – hourly average losses from idle j^{th} well.

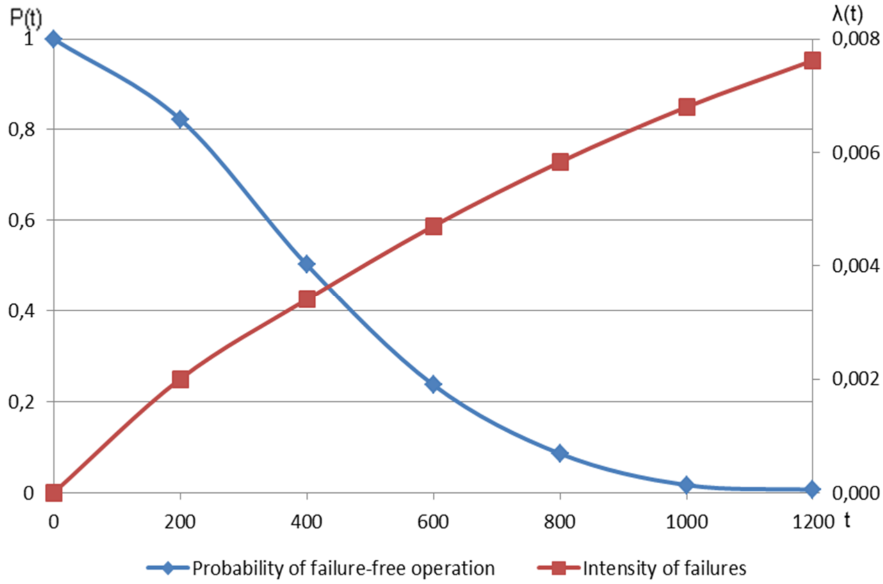


Fig. 2. The probability distribution functions of uptime and failure rate of tubing (breaks) in the fields of OGPE Dolynanaftogaz.

Based on the distribution of the probability of failure-free operation established by law, there are calculated the optimal between-repairs period and failure rates. The calculation results are summarized in Table 1.

Table 1. Optimum between-repairs periods and failure rates.

Optimal between-repairs period, days	Failure rates of the well, λ
226	0,001935
254	0,002125
308	0,002479

Table 2 presents the results of the calculation of costs depending on the number of crews at a failure rate λ = 0,001935.

With an increase in the number of crews from 7 to 11, a sharp (9 crews) decrease in total costs occurs first. The involvement of the 12th crew becomes impractical, since the total costs increase due to the additional costs of maintaining the crews. The total costs include the costs of maintaining crews and oil losses during wells downtime, that is, they can be adjusted due to one or another component.

Table 2. Total costs for well maintenance processes at failure rate λ = 0,001935.

The number of repair crews	7	8	9	10	11	12
Losses from downtime, conv. units	8301,47	1481,36	506,09	197,19	79,65	32,06
The cost of maintaining repair crews, conv. units	675,11	725,63	777,11	827,62	878,13	929,62
Total cost, conv. units	9002,81	2233,22	1308,46	1051,04	984,01	987,90

Table 3 shows the results of cost calculation depending on the number of crews with the failure rate λ = 0,002125.

When analyzing the dependencies (table 3), a similar trend is observed for the dependencies presented in table 2. That is, the optimal number of repair crews is 12. The involvement of another crew is unprofitable. Although losses from downtime continue to decline, at the same time, costs for crews are growing, which, in turn, increases total costs by 2,8%.

Table 3. Total costs of well servicing processes at failure rate λ = 0,002125.

The number of repair crews	7	8	9	10	11	12	13
Losses from downtime, conv. units	13555,70	1804,84	597,40	233,13	94,22	38,86	15,54
The cost of maintaining repair crews, conv. units	701,34	751,85	803,34	853,85	904,36	955,84	1006,36
Total cost, conv. units	14256,07	2556,69	1399,77	1086,01	999,56	993,73	1021,90

Table 4 shows the results of cost calculation depending on the number of crews with the failure rate λ = 0,002479.

Table 4. Total costs of well servicing processes at failure rate λ = 0,002479.

The number of repair crews	8	9	10	11	12	13	14
Losses from downtime, conv. units	8351,01	1664,96	597,40	243,82	103,94	43,71	18,46
The cost of maintaining repair crews, conv. units	801,39	851,91	902,42	953,90	1004,41	1055,90	1106,41
Total cost, conv. units	9152,41	2516,86	1499,82	1197,72	1107,38	1098,64	1123,89

With the failure rate λ = 0,002479, the optimal number of crews is 13, so that total costs decrease by 0,8% due to a decrease in loss from downtime by 57,9%.

7 Determining the optimal number of crews for various failure rates

If the failure rate is λ = 0,001935, the optimal number of crews is 11, the total costs are 984,01 conventional units, if λ = 0,002125 – respectively 12 and 993,73 conventional units, if λ = 0,002479 – respectively 13 and 1098,64 conventional units. A further increase in the number of repair crews is inefficient, although losses from well downtime are reduced, but the costs of their maintaining are growing much faster, which leads to an increase in total costs. With an increase in the number of crews, losses from well downtime are rapidly decreasing, then gradually equalizing. This is due to the fact that at first the losses from downtime are very high, and the costs of maintaining the crews are relatively low.

Fig. 3 and table 5 show that losses from downtime of wells at different λ, when 8 and 9 crews work, are significantly different. In the future, the situation is leveled. This is due to the fact that a larger number of crews repair more wells, reducing their downtime, but at the same time, their maintenance costs increase.

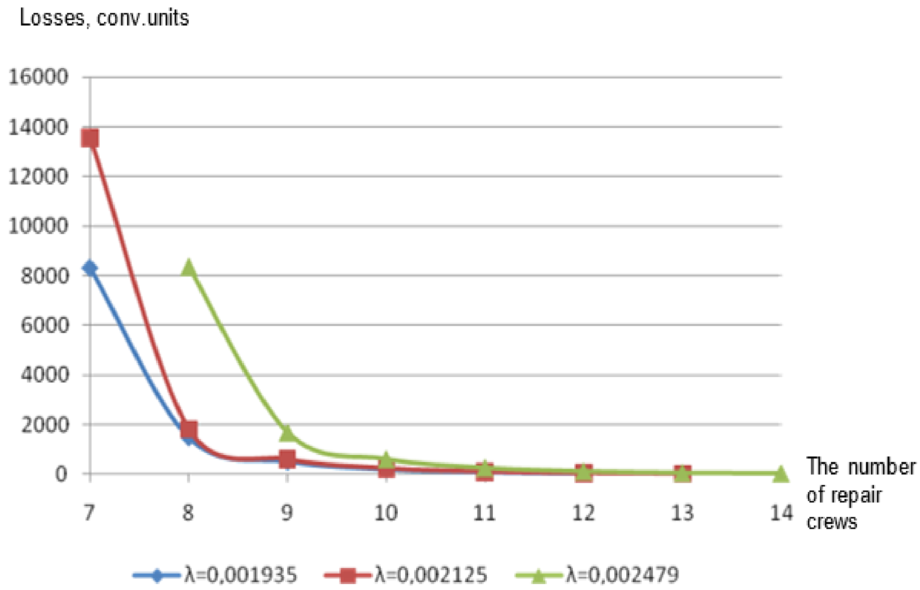


Fig. 3. The losses from downtime of wells at different λ.

Table 5. Losses from downtime of wells at different λ.

The number of repair crews	7	8	9	10	11	12	13	14
λ = 0,001935	8301,47	1481,36	506,09	197,19	79,65	32,06	-	-
λ = 0,002125	13555,7	1804,84	597,40	233,13	94,22	38,86	15,54	-
λ = 0,002479	-	8351,01	1664,96	597,40	243,82	103,94	43,71	18,46

Figure 4 reflects the increase in costs with the increase in the number of crews. It shows the same upward trend in cost if λ increases. If there are 8 crews with λ = 0,001935, the costs make 725,63 conv. units, and at λ = 0,002479 – 801,39 conv. units, i.e. with an increase of λ at 0,000544, costs increase by 9,5% (table 6).

Fig. 5 and table 7 shows that the minimum total costs are in the range from 1013 conv. units up to 1131 conv. units that is, the number of crews varies from 11 to 13.

With the increase in the failure rate, the cost curves shift to the right, which is explained by a large number of repairs, and as a result, there is required an additional number of repair crews [7].

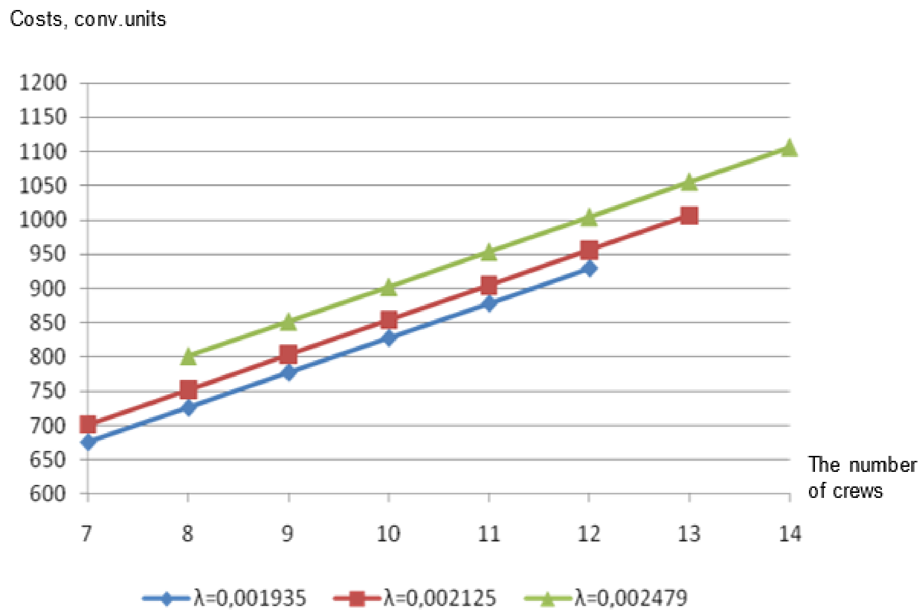


Fig. 4. The costs for maintaining a repair crew at various λ.

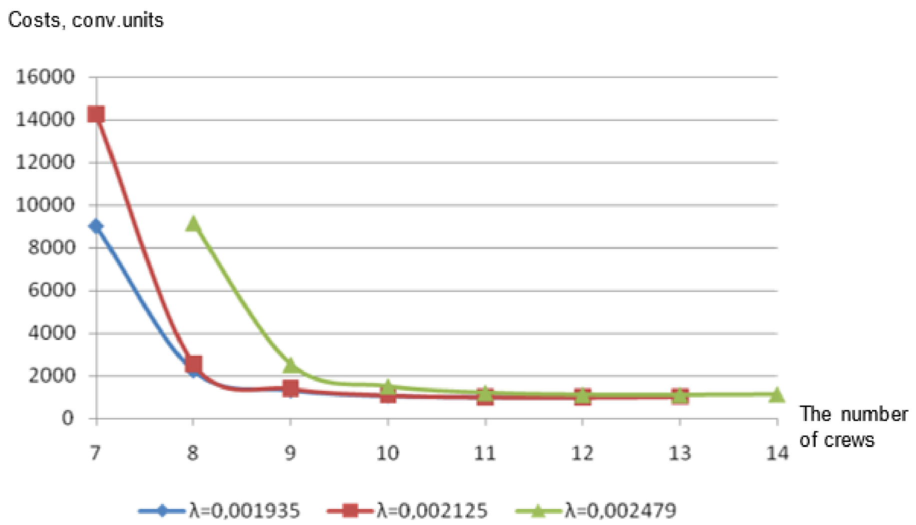


Fig. 5. The total costs at various λ.

Table 6. The total costs of well servicing processes for various λ.

The number of repair crews	7	8	9	10	11	12	13	14
λ = 0,001935	675,11	725,63	777,11	827,62	878,13	929,62	-	-
λ = 0,002125	701,34	751,85	803,34	853,85	904,36	955,84	1006,36	-
λ = 0,002479	-	801,39	851,91	902,42	953,90	1004,41	1055,90	1106,41

Table 7. The total costs of well servicing processes for various λ.

The number of repair crews	7	8	9	10	11	12	13	14
λ = 0,001935	9002,81	2233,22	1308,46	1051,04	984,01	987,90	-	-
λ = 0,002125	14256,1	2556,69	1399,77	1086,01	999,56	993,73	1021,90	-
λ = 0,002479	-	9152,41	2516,86	1499,82	1197,72	1107,38	1098,64	1123,89

8 Conclusion

It is established that the continuity of oil and gas production largely depends on the operation, maintenance and repair of wells, so there is a growing need for their special maintenance and repair, due to wear and tear of operating equipment and the need for special measures to protect the subsoil. Work on repair service of wells involves current repairs and overhauls of wells and care of the operational equipment in the period between the next repairs.

The model of organizing repair works at oil and gas companies has been developed, and its implementation ensures the optimization and prediction of the costs of maintaining repair crews and the reduction of losses from well downtime. Well repair by repair crews is considered as a multichannel queuing system in which the flow of service requests is the flow of repair applications. The maintenance procedure consists in the use of a repair crew according to the application, which is considered completed after the completion of the work of this crew at the facility, and the order of satisfying applications is carried out in the order they are received.

The practical implementation of the proposed models is carried out on the example of the organization of repair work for oil and gas companies OGPE “Dolynaftogaz”, OGPE “Borislavnaftogaz” and OGPE “Nadvirnaftogaz” and confirmed the feasibility of using the methods of queuing in decision-making on the organization and management of repairs. It is established that the increase in the number and intensity of equipment failures requires an increase in the number of repair crews. However, such an increase is not always advisable, since the costs of maintaining the crews begin to exceed the losses from the downtime of the wells.

References

1. Barkalova, T.G., Kokotkina, T.N., Tsaregorodtsev, E.I.: Challenges and Opportunities of Using Simulation Models in Forecasting Socio-economic. Development of the Region International journal of environmental & science education **11**(17), 10661–10671 (2016)

2. Boiko, V.S.: Pidzemnyi remont sverdlovyn. Ivano-Frankivsk (2011)
3. Brumar, B.A.: Modeling and simulation of economic processes. Studies and Scientific Researches - Economic Edition **15**, 322–328 (2010). doi:10.29358/sceco.v0i15.135
4. Fabianova, J., Kacmary, P., Janekova, J.: Operative production planning utilising quantitative forecasting and Monte Carlo simulations. Open Engineering. **9**(1), 613-622 (2019). doi:10.1515/eng-2019-0071
5. Gansterer, M., Almeder, C., Hartl, R.F.: Simulation-based optimization methods for setting production planning parameters. International Journal of Production Economics **151**, 206–213 (2014). doi:10.1016/j.ijpe.2013.10.016
6. Haugg, I.G., Frantz, R.Z., Roos-Frantz, F., Sawickie, S., Zucolotto, B.: Towards optimisation of the number of threads in the integration platform engines using simulation models based on queueing theory. Revista Brasileira de Computação Aplicada **11**(1), 48–58 (2019). doi:10.5335/rbca.v11i1.8784
7. Hobyr, I., Babenko, V., Savko, O., Bui, Yu.: Optimization of arranging of repair works at oil and gas enterprises through use of simulation modeling tools. Advances in Economics, Business and Management Research **99**, 183–188 (2019). doi:10.2991/mdsmes-19.2019.35
8. Khan, R., Mad, A.B., Osman, K. Aziz, M.A.A.: Maintenance Management of Aging Oil and Gas Facilities. In: Márquez, F.P.G. (ed.): Maintenance Management. InTechOpen, London (2019). doi:10.5772/intechopen.82841
9. Khvostina, I., Havadzyn, N., Horal, L., Yurchenko, N.: Emergent Properties Manifestation in the Risk Assessment of Oil and Gas Companies. CEUR Workshop Proceedings **2422**, 157–168 (2019)
10. Liu, R., Xie, X., Yu, K., Hu, Q.: A survey on simulation optimization for the manufacturing system operation. International Journal of Modelling and Simulation. **38**(2), 116-127 (2018). doi:10.1080/02286203.2017.1401418
11. Mazur, I.M.: Mekhanizm zabezpechennia ekonomichnoi efektyvnosti funktsionuvannia naftohazovydobuvnykh pidpryiemstv. Ivano-Frankivsk (2011)
12. Oliveira, J.B., Jin, M., Lima, R.S., Kobza, J.E., Montevechi, J.A.B.: The role of simulation and optimization methods in supply chain risk management: Performance and review standpoints. Simulation Modelling Practice and Theory **92**, 17-44 (2019). doi:10.1016/j.simpat.2018.11.007
13. Olose, E.O.: Effective Maintenance and Reliability Program in the Production of Crude Oil and Natural Gas. International Journal of Scientific & Engineering Research **7**(2), 1048–1056 (2016)
14. Pererva, P.G., Pogorelov, I.N.: Issledovanie sistem massovogo obsluzhivaniya dlya obespecheniya bezpereboynoy rabotyi oborudovaniya na predpriyatiyah. Visnyk Nacionalnogo texnichnogo universytetu “XPI” 8, 187–191 (2011)
15. Saad, S.M., Elsaghier, E.H., Ezaga, D.: Planning and optimising petroleum supply chain. Procedia Manufacturing **17**, 803–810 (2018). doi:10.1016/j.promfg.2018.10.131
16. Shanmugasundaram, S., Umarani, P.: Queuing theory applied in our day to day life. International Journal of Scientific & Engineering Research **6**(4), 533–541 (2015)
17. Standart orhanizatsii Ukrainy 11.2-30019775-168:2010. Sverdlovyny na naftu i haz. Potochnyi ta kapitalnyi remont sverdlovyn. Klasyfikator robit ta poriadok yikh provedennia (2010)
18. Winkowski, C.: Classification of forecasting methods in production engineering. Engineering Management in Production and Services **11**(4), 23–33 (2019). doi:10.2478/emj-2019-0030

Fuzzy cluster analysis of indicators for assessing the potential of recreational forest use

Evstakhii Kryzhanivs'kyi[0000-0001-6315-1277], Liliana Horal[0000-0001-6066-5619], Iryna Perevozova[0000-0002-3878-802X], Vira Shiyko[0000-0002-2822-0641], Nataliia Mykytiuk[0000-0001-3194-3891] and Maria Berlous[0000-0003-2856-9832]

Ivano-Frankivsk National Technical University of Oil and Gas,
15 Karpatska Str., Ivano-Frankivsk, 76019, Ukraine
rector@nung.edu.ua, liliana.goral@gmail.com, perevozova@ukr.net, viraSh@i.ua, nataliamykytiukmmm@gmail.com, masher@i.ua

Abstract. Cluster analysis of the efficiency of the recreational forest use of the region by separate components of the recreational forest use potential is provided in the article. The main stages of the cluster analysis of the recreational forest use level based on the predetermined components were determined. Among the agglomerative methods of cluster analysis, intended for grouping and combining the objects of study, it is common to distinguish the three most common types: the hierarchical method or the method of tree clustering; the K-means Clustering Method and the two-step aggregation method. For the correct selection of clusters, a comparative analysis of several methods was performed: arithmetic mean ranks, hierarchical methods followed by dendrogram construction, K-means method, which refers to reference methods, in which the number of groups is specified by the user. The cluster analysis of forestries by twenty analytical grounds was not proved by analysis of variance, so the re-clustering of certain objects was carried out according to the nine most significant analytical features. As a result, the forestry was clustered into four clusters. The conducted cluster analysis with the use of different methods allows us to state that their combination helps to select reasonable groupings, clearly illustrate the clustering procedure and rank the obtained forestry clusters.

Keywords: cluster analysis, k-means clustering method, forestry, recreation.

1 Introduction

The intensive development of recreation in the world creates motivation to use significant reserves of recreational resources. To expand the use of forest recreational resources, it is necessary to use for this purpose not only nature reserves, but also to involve more and more forests of state forestry farms in this use. The reserves of recreational forest use on the territory of Ukraine are significant. Therefore, there is a need to assess their development on the basis of the classification of forestry areas on many analytical grounds. Taking into account the fact that such classification is a rather time-consuming task, it is proposed to carry out forests clustering with the help of software.

The use of cluster analysis methods is dictated primarily by the fact that they help to build scientifically based classifications, identify internal links between the observed population units. In addition, cluster analysis methods can be used to compress information, which is an important factor in the conditions of constant increase and complication of statistical data flows. That is why this type of statistical analysis is of great importance when analyzing the development of recreational facilities. It should be noted that recently cluster analysis has received considerable attention from domestic and foreign experts in various scientific fields. One of the reasons is that modern science is increasingly relying on classification for its development. Moreover, this process deepens as knowledge specialization grows, which in its turn is based largely on objective classification. Another reason is related to the accompanying deepening of specialized knowledge, the increase in the number of variables, taken into account in the analysis of certain objects.

Clustering of the studied forests will allow the effective management of recreational areas, taking into account the reserves for improving the development of areas for selected components and also to develop at the state level the Strategy of recreational forest use development in Ukraine for the maintenance of the National recreational product competitive in the domestic and world markets. Taking into consideration the fact that each region of Ukraine is characterized by its natural and climatic conditions, ethnic traditions and historical and cultural recreational features, there is a problem of qualitative analysis and assessment of the level of recreational facilities development.

2 Background

The foreign scientists, who studied the issue of recreational forest management, are Simon Bell [2], William M. Murphy [11], Lloyd C. Irland, Darius Adams, Ralph Alig, Carter J. Betz, Chi-Chung Chen, Mark Hutchins, Bruce A. McCarl, Ken Skog and Brent L. Sohngen [6], Nerida Anderson, Rebecca M. Ford, Lauren T. Bennett, Craig Nitschke and Kathryn J. H. Williams [1], Artti Juutinen, Anna-Kaisa Kosenius and Ville Ovaskainen [7], Markus A. Meyer, Joachim Rathmann and Christoph Schulz [10], Tina Gerstenberg, Christoph F. Baumeister, Ulrich Schraml and Tobias Plieninger [5], Kee-Cheo Lee and Kee-Rae Kang [9], Hyun-Kyu Shin and Hong-Chul Shin [12], Yevstakhii Kryzhanivskyi, Liliana Horal, Vira Shyiko, Oleksii Holubchak and Nataliia Mykytiuk [8].

Markus A. Meyer, Joachim Rathmann and Christoph Schulz proved in [10] that visitors cluster along major paths or regions in urban and rural forest, recreation of the local population is highly driven by relaxation, forest structures and demographic factors play a minor role for forest benefits, forest benefits do not strongly vary within the area of the forests, forest management should focus on avoiding nuisances to support forest benefits. They found a weak connection between recreational behavior and demand for specific forest characteristics. For local recreation, we recommend to provide a basic level of highly rated FB and to avoid nuisances rather than designing forests for a desired appearance.

Tina Gerstenberg, Christoph F. Baumeister, Ulrich Schraml and Tobias Plieninger in [5] identified frequencies of activities in urban forests, visualized activity-specific hot routes, and unveiled the contributions of landscape features to recreational use intensity. The hot route maps represent an advancement of existing forest function maps, as they were based on more reliable spatially explicit data on where people move in forests. They used a public participation mapping procedure as a basis for visualizing recreational use intensity. These maps may aid forest managers to tailor management according to residents' forest uses and preferences, prioritize objectives, and prevent conflicts between re-creational user groups, conservationists and representatives of the timber industry. They conclude that urban forest managers may promote outdoor recreation by maintaining large proportions of broadleaved dominated stands. Finally, accessibility to water bodies as well as unique structural compositions – as represented by protected habitats – may enhance recreational use [5].

The purpose of Kee-Cheo Lee and Kee-Rae Kang [9] is to classify the forests by considering the supplier's perspective as well as the user's perspective in order to provide fundamental materials for the operation of the natural recreation forests. A factor analysis was conducted to identify the common characteristics of the selected twelve variables by pre-selection and survey of experts. K-means cluster analysis was conducted among those factors to classify the natural recreation forests in Korea. Four factors were drawn after the factor analysis and the factors were named according to the variables and sizes as 'The use performance and visiting condition factor', 'Education and settlement factor', 'Internal activation factor' and 'Potential factor'. In addition, the cluster analysis of the matrix was conducted for the points of the drawn factors and the final classification consists of five groups. The results of this study may contribute to providing fundamental materials for the operation and management of natural recreation forests. Also, it may act as a reference when investigating the natural recreation forests of Korea. Proposing the classification natural recreation forests could be helpful in selecting the proper recreation forest in the future. Based on the established model, fundamental materials could be provided to improve the profitability of the natural recreation forests by effectively expanding the number of tourists, creating new natural recreation forests and proper maintenance and management [9].

Hyun-Kyu Shin and Hong-Chul Shin in [12] segmented recreational forest's visitors for marketing based on purpose of visit. Using the factor analysis, cluster analysis, cross tab, and t-test to find out different behavioral intention in each cluster, the result elicited some implications. First, 2 clusters were founded and has difference in behavioral intentions. Cluster 1 (married, 200~300 hundred won income) has higher satisfaction, revisit intention, recommendation intention. The result shows that market researcher in recreational forest should approach different marketing strategy and has various facilities, active program. This research needs to survey broad region to generalized result [12].

Thus, having considered the scientific works of both foreign and domestic researchers of the recreational forest management problems and without diminishing their scientific value to improve development of recreational forest management, it is possible to consider and necessarily classify the recreational region for a component that is its own manufacturer [8].

3 Methodology

As it is known, for complex evaluation of every economic process or its components, the methods of integrated indicators calculation are conventionally applied using different economic and mathematical methods and approaches. The complex evaluation is required to define the potential of recreational forest management, considering the development of all its components. Therefore, in [8] we propose to evaluate the potential of recreational forest use by performing the following steps: to identify the recreational forest use potential components; to develop and form a system of quantitative and qualitative indicators (indices) in order to evaluate the efficiency of recreational forest use potential by its component composition; to evaluate the efficiency of recreational forest use of the regional territories by individual components of the recreational forest use potential using certain indicators; to comprehensively evaluate the efficiency of each recreational forest use potential component; to conduct an integrated evaluation of the efficiency of recreational forest use by means of using taxonomic analysis methods and fuzzy set theory; to determine the level of the recreational forest use potential by comparing the integrated indicator value with its standard (critical) values [8]. Based on the previous studies of recreational forest management, the following structural components of recreational forest management potential can be formed: a resource component, social component, economic component, innovation and investment component. Each component of recreational forest use is characterized by a system of performance indicators. According to the above characteristics of each component, the following system of indicators can be proposed, considering the attributes of recreational activity, which are listed in table 1 [8].

Economic and mathematical modelling of evaluation of the recreational forest management potential determined the efficiency of recreational forest use of regional territories by individual components of recreational forest management potential using indicators specified in table 1. A taxonomic method based on determination of taxonomic indicators of each component [8] was used for this stage.

To approve the methodology of assessing the recreational potential of forest use, a typical forestry of the Western region of Ukraine was selected, including 8 forestries. It is worth mentioning that as a result of the underdeveloped information and statistical infrastructure of forestries, it was not possible to calculate a required system of indicators, shown in table 1. However, the taxonomic indicators were calculated based on the actual statistical base on the resource and social components of each forestry. The calculation results of forestry activity were summarized in table 2.

Therefore, based on obtained calculations we can conclude that recreational forest management in Ukraine is low, confirmed by the level of recreational forest management potential (table 2). Of 8 analysed forests only in Forestry 1 the potential level is average, in two forestries the integrated indicator of recreational forest management potential level has been set at a level below average, and the remaining 5 forests have a low level of recreational forest management. Graphically obtained results are shown in figure 1 [8].

Table 1. Evaluation indicators of the recreational forest management potential components.

Component	Indicator	Substantiation
Resource component	Area of recreational territories, km^2	Total area of forestry intended for recreational forest use
	Number of recreational places, quantity	Number of recreational places located on the forestry territory intended for recreational forest management
	The level of attractiveness of natural and recreational resources	The indicator can be evaluated according to the following criteria: exoticism, uniqueness, aesthetics, comfort, etc.
	Quality factor of forest vegetation	It describes the level of recreation applicability
	Exoticism degree (contrast) of recreational territory	It is determined as a contrast ratio degree of the resting place relative to a recreant's permanent residence
Economic component	Proportion of total forestry costs on maintenance of recreational places, %	It shows the proportion of the total costs on maintenance of recreational territories
	Efficiency factor of recreational forest management	It shows attractiveness of recreational forest management
	Wear coefficient of recreational fixed assets (FA)	It characterizes wear level of recreational fixed assets
	Volume of marginal costs for growing 1 ha of recreational forest	They reflect the effect, achieved by improving the forest as a means of labor in recreation sphere
	Capacity of a single recreational load	It shows the maximum permissible number of persons on recreational territory
Social component	Proportion of recreant employees	It shows a proportion of recreant employees in the total number of staff involved in recreational activities
	Recreational capacity	The capacity of recreation centres (resorts, tourist, health, recreational complexes) is a simultaneous number of recreants that can be located in this centre, without disturbing ecological balance within this centre and surrounding territories
	Recreational load per 1 ha of forest	It determines attendance intensity for any segment of the day, during weekends, weekdays
	The average stay of vacationers on the recreational territory, h	It shows an average length of stay of visitors on the recreational territory of forest area
Innovation and invest-	Cost amount on marketing activities of recreational territories	It characterizes the development level of marketing activities

Component	Indicator	Substantiation
ment component	Efficiency of innovation implementation of recreational forest management	It characterizes the innovation level and efficiency of recreational innovation use
	Amount of investments in recreational activity	It shows the amount of investment resources aimed at recreational activities
	Proportion of foreign investments in recreational activities financing	It shows amount of recreational activity financing at the expense of foreign financial sources
	Quantity of the won grants (programs) to finance recreational activities	It characterizes relevance of the recreational sphere development

Table 2. Taxonomic analysis results of recreational forest management of a typical forestry.

Indicator	Forestry 1	Forestry 2	Forestry 3	Forestry 4	Forestry 5	Forestry 6	Forestry 7	Forestry 8
Taxonomic indicator of resource component	1.00	0.51	0.33	0.36	0.32	0.33	0.31	0.33
Taxonomic indicator of social component	1.00	0.56	0.56	0.30	0.30	0.21	0.39	0.39
Taxonomic indicator of economic component	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Taxonomic indicator of innovation and investment component	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Integrated indicator of recreational forest management potential level	0.50	0.27	0.22	0.16	0.16	0.14	0.17	0.18

Thus, according to the results of economic and mathematical modelling of the integrated indicator of recreational forest management potential level, it can be concluded that the recreational forest management potential in Ukraine is low (figure 1), so measures should be taken to improve recreational activity results and develop this industry. As the calculations indicate, first of all, it is urgent to develop economic and innovation investment components of the recreational forest management potential in Ukraine.

Thus, having obtained the results of calculating the integrated indicator of the recreational forest use level in the studied forests, we consider it necessary to conduct a fuzzy cluster analysis of forestry based on the analysis of forest use potential individual indicators for the studied objects. The main stages of cluster analysis of the recreational forest use level by predetermined components are shown in the figure 2.

To implement the clustering process, it is necessary to develop a matrix of observations x_{ij}. In this case, the original set consists of *m* elements described by *n* parameters, and each of its lines can be interpreted as a point or vector placed in *i*-dimensional space with coordinates equal to the value of *n* features for a particular forestry. Thus, in the observation matrix x_{ij} is the value of feature *i* for *j* forestry; *j* – a number of classification objects (forestry); *i* – a number of features of the objects.

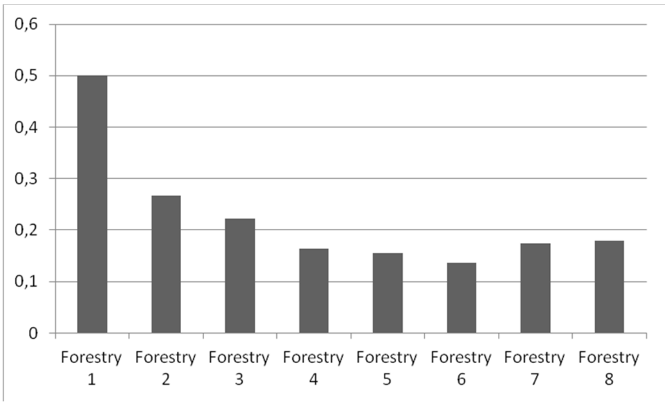


Fig. 1. Integrated indicator of recreational forest management level.

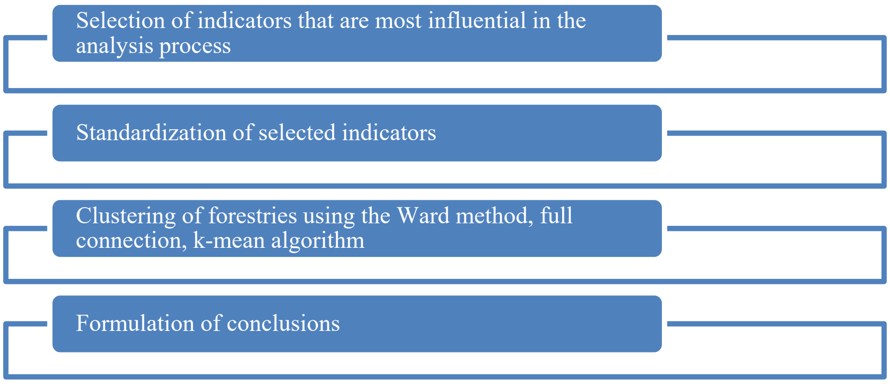


Fig. 2. The main stages of cluster analysis of the recreational forest

Using element multiplicity *w*, described by *n*-signs, each unit can be interpreted as a point of *n*-dimensional space with coordinates equal to the value of n attributes for the analysed unit. Let us represent the matrix as follows:

$$X = \begin{bmatrix} x_{11}x_{12}\ldots x_{1k}\ldots x_{1n} \\ x_{21}x_{22}\ldots x_{2k}\ldots x_{2n} \\ \ldots\ldots\ldots\ldots\ldots\ldots \\ x_{i1}x_{i2}\ldots x_{ik}\ldots x_{in} \\ \ldots\ldots\ldots\ldots\ldots\ldots \\ x_{w1}x_{w2}\ldots x_{wk}\ldots x_{wn} \end{bmatrix} \quad (1)$$

where: *w* is the number of study periods, *n* is the number of indicators of each recreational forest management potential, x_{ik} – indicator value *k* of each specific component for a year ($k = 1 \div n$, $i = 1 \div w$).

As indicators of recreational forest use management level assessment are reflected in various measures, they need to be standardized. One of the most common means of statistical generalization for inhomogeneous populations is the standardization of indicators by the ratio of deviation (x_i) to the unit of standardization. In our case, σ_i is chosen as the standardization unit. These features should be normalized using the following formula:

$$z_{ij} = \frac{x_{ij} - \bar{x}_i}{\sigma_i} \tag{2}$$

when

$$\bar{x}_i = \frac{1}{w} \sum_{i=1}^{w} x_{ij} \tag{3}$$

$$s_k = \left[\frac{1}{w} \sum_{i=1}^{w} (x_{ik} - \bar{x}_k)^2 \right]^{\frac{1}{2}} \tag{4}$$

where: z_{ij} – standardized value of indicator j for the i-th study period; x_{ij} – standardized value of indicator j for the i-th study period; x_j – arithmetic mean of kj indicator; σ_j – standard deviation of k indicator; w – a number of periods.

The main feature of clusters is that objects belonging to one of them are more similar to each other than objects from different clusters. Such a classification with the help of software and computer system STATISTICA, can be performed simultaneously on a fairly large number of analytical features. In our case, clusters will be called geographically concentrated and interconnected by the level of recreational potential of forestry.

Among the agglomerative methods of cluster analysis, which are intended for grouping and combining objects of study, it is common to distinguish three most common types: hierarchical method (I) or the method of tree clustering; K-means Clustering Method (II) and two-step aggregation method (III).

I. Hierarchical clustering is used in the formation of clusters by determining the distances between objects and allows you to graphically visualize the results of the study in the form of a dendrogram. These distances can be determined in one-dimensional or multidimensional space. However, an important step in conducting a cluster analysis is to select the correct method for calculating the distances between the studied objects. The main ways to determine distances are: Euclidean distance, square of Euclidean distances, distance of city squares (Manhattan), Chebyshev distance, power distance.

II. K-means Clustering Method is the most common among non-hierarchical methods of cluster analysis. Unlike hierarchical methods, which did not require prior assumptions about the number of clusters, to be able to use this method it is necessary to have a hypothesis about the most probable number of clusters. K-means Clustering Method builds k clusters located at as large distances from each other as possible. Note that the K-means Clustering Method assumes that the number of clusters includes observations with the closest average value. The method is based on minimizing the sum of the distances squares between each observation and the center of its cluster, i.e. the function. In this case, the choice of the number of clusters is based on the research hypothesis. If it is not present, it is recommended to create 2 clusters, further 3, 4, 5, comparing the received results. The input will be $X_u = \{x_{1u}, x_{2u},\ldots, x_{mu}\}$ – a set of unmarked data; $X_{kl} = \{x_{1l}, x_{2l}, \ldots, x_{pl}\}$ is a set of marked data in the class k, X_l=K_k=1X_{kl}. At the output, we want to obtain separated K sets $\{C_k\}$ K_k = 1 of X_u, which minimizes the objective function in k-means. Set parameters:

1. $t = 0$.
2. Initialization of cluster centers:

$$\mu_k = \frac{1}{|x_k^l|} \sum_{x \in x_k^l} x \tag{5}$$

3. Repeat until convergence:

provide cluster data:

For marked data: $x \in x_k^l$ provide x to the cluster C_k^{t+1}.

For unlabeled data: for $x_{iu} \in x_u$ provide to C_k^{t+1} a cluster obtained from $k = arg\ min_k||x_i^u - \mu_k^t||^2$.

4. Update centers:

$$\mu_k^{t+1} = \frac{1}{|C_k^l|} \sum_{x \in c_k^l} x \tag{6}$$

$t \leftarrow t+1$.

Another component of the algorithm is based on the discrepancy KL, which is a measure of the mismatch between the two probability distributions. Taking into account the K-dimensional probability vector of assignment of clusters p and q corresponding to points respectively x_p and x_q, the discrepancy KL between p and q is given by the formula:

$$KL\ (p\|q) = \sum_{i=1}^{K} p_i \log \frac{p_i}{q_i}, \tag{7}$$

where K is the number of clusters. In this approach, we use a symmetric variant of the discrepancy KL, because we are dealing only with the optimization of the loss function for p and q simultaneously:

$$L_{p,q} = KL(p\|q) + KL(q\|p) \tag{8}$$

Losses are obtained by first fixing p and calculating the discrepancy q with p and vice versa.

The described method makes it possible to automate the process of cluster data analysis, especially if the number of clusters is unknown from the beginning. For this purpose, the model of the neural network-based cluster data analysis system was described on the basis of k-means and KL discrepancy methods.

III. The two-way aggregation method is used in cases when you want to perform simultaneous clustering of objects (columns) and observations (rows) [11].

The key to the adequacy of the economic objects cluster analysis results is a reasonable choice of factors by which the grouping is carried out. Regarding the factor characteristics, we used a four-component system of indicators, which are shown in table 1.

The main purpose of cluster analysis is to break down the set of studied objects and features into homogeneous in the appropriate sense groups or clusters. This means that the task of classifying data and identifying the appropriate structure in it is solved. Methods of cluster analysis can be used in different cases, even when it comes to a simple grouping, and which all comes down to creating groups by the number of similarities.

The need for an objective division of different economic objects into groups exists constantly, because this classification allows you to find methods for effective management of these objects. Methods of cluster analysis allow to solve the following tasks: classification of objects taking into account the features that reflect the essence, nature of objects; verification of the assumptions about the presence of some structure in the studied set of objects, i.e. search for the existing structure; building new classifications for phenomena that have been little studied when it is necessary to establish the existence of relationships within the population and try to introduce a structure into it.

Cluster analysis has certain shortcomings and limitations. In particular, the composition and number of clusters depends on the selected breakdown criteria. When reducing the original data set to a more compact form, certain distortions may occur, and individual features of individual objects may be lost by replacing their characteristics with generalized values of cluster parameters.

When classifying objects, the possibility of the absence of any cluster values in the considered set is often ignored. In the cluster analysis it is considered that: 1) the chosen characteristics allow, in principle, a desirable division into clusters; 2) the units of measurement (scale) are chosen correctly.

The quality criterion of clustering to some extent reflects the following informal requirements: 1) within groups, objects must be closely related; 2) objects of different groups must be far from each other; 3) other things being equal, the distribution of objects by groups must be uniform. The key point in cluster analysis is the choice of metrics (or measures of proximity of objects), which crucially depends on the final version of the objects division into groups with a given algorithm of division.

The task of cluster analysis is to, based on the data of the set X, divide the set of objects G into m (m is an integer) of clusters (subsets) G_1, G_2,…, G_m, so that each object G_j belongs to one and only one subset of the breakdown and that objects belonging to the same cluster are similar, while objects belonging to different clusters are heterogeneous. The solution to the problem of cluster analysis is the breakdowns that satisfy some criterion of optimality. This criterion may be some functionality that expresses the levels of different breakdowns desirability and groups, called the objective function. For further research, it was possible to use the methods of theories of complex systems and equipment made by tools used to examine the necessary systems of complexity, which were used in conventional [4; 3; 14; 13].

Let's perform cluster analysis according to the *K*-means Clustering method described above for each of the selected components (table 3).

Table 3. Substantiation of component's indicator.

Component	Indicator	Substantiation
Resource component	Area of recreational territories, km^2	var2
	Number of recreational sites, quantity	var3
	The level of attractiveness of natural and recreational resources	var4
	Quality factor of forest vegetation	var5
	Exoticism degree (contrast) of recreational territory	var6
Economic component	A proportion of total forestry costs on maintenance of recreational sites, %	var7
	Efficiency factor of recreational forest management	var8
	Wear coefficient of recreational fixed assets (FA)	var9
	Volume of marginal costs for growing 1 ha of recreational forest	var10
	Capacity of a single recreational load	var11
Social component	Proportion of recreant employees	var12
	Recreational capacity	var13
	Recreational load per 1 ha of forest	var14
	The average stay of vacationers on the recreational territory, h	var15
Innovation and investment component	Cost amount on marketing activities of recreational territories	var16
	Efficiency of innovation implementation of recreational forest management	var17
	Amount of investments in recreational activity	var18
	Proportion of foreign investments in recreational activities financing	var19
	Quantity of grants (programs) won to finance recreational activities	var20

To begin with, we will standardize certain input data and summarize the results in table 4.

Table 4. The results of the standardization of the features of the recreational forest use assessment features.

	Var2	Var3	Var4	Var5	Var6	Var7	Var8	Var9	Var10	Var11
1	-0,5669	-0,8472	-1,0818	-1,3149	-0,6656	-1,1348	-1,5557	-0,4606	-1,52518	-0,72738
2	-0,4002	-0,8472	-0,3442	-0,5164	0,10241	-0,5857	0,92388	-0,8844	-0,82125	-0,42176
3	-0,2334	-0,4621	-1,0818	0,28214	0,87045	-0,5857	0,37286	-0,4606	-0,11732	-0,1467
4	0,43351	-0,4621	0,39338	1,08063	1,63849	0,23794	0,64837	-0,0367	0,586608	-0,17726
5	0,60024	-0,4621	1,13097	1,87913	-0,6656	0,51249	0,56572	0,3871	1,290537	-0,11614
6	2,10083	-0,8472	1,13097	1,47988	-0,6656	-0,3112	0,51061	0,81093	0,938573	0,189486
7	-0,0667	-0,4621	-0,3442	0,28214	0,10241	-1,1348	0,81368	0,6414	-0,82125	0,800733
8	0,43351	-0,077	0,39338	0,68139	-0,6656	0,23794	0,89633	0,81093	0,234643	1,411979
9	-1,2338	-0,077	-0,3442	-1,3149	-1,4337	-0,5857	-1,5557	0,25995	-1,17322	-1,64425
10	-0,5669	0,69319	-1,0818	-1,1552	-0,6656	-0,8603	0,92388	1,23476	1,290537	-1,33863
11	1,76737	1,07829	0,39338	-0,5164	0,10241	-0,5857	-1,0047	-1,7321	0,586608	-1,03301
12	0,43351	3,00381	-1,0818	-0,6761	0,87045	0,23794	-1,2802	-2,1559	-0,53968	-0,42176
13	-0,4002	-0,4621	1,13097	0,60154	1,63849	0,51249	-1,2802	-0,0367	1,008966	0,800733
14	-1,5673	0,30808	1,86857	-0,2768	-1,4337	1,61068	0,64837	0,3871	0,586608	1,717603
15	-0,7336	-0,077	-1,0818	-0,5164	0,87045	2,43433	0,37286	1,23476	-1,52518	1,106356

	Var12	Var13	Var14	Var15	Var16	Var17	Var18	Var19	Var20
1	-0,02347	0,20112	-0,5031	-0,25273	-2,4015	-0,15213	-1,24083	-1,68157	-0,60911
2	0,328617	1,039122	0,440211	0,288836	-1,01602	-0,96713	-0,05909	-0,62619	-1,08498
3	-0,37556	1,877124	1,383519	1,371973	0,369462	1,477862	0,531783	0,429187	0,342624
4	-0,02347	-1,47488	-0,5031	1,913541	1,062203	-0,15213	1,122653	0,956876	-0,60911
5	0,328617	-0,63688	0,440211	-0,25273	1,754944	1,477862	1,713523	1,484565	0,342624
6	1,736974	-0,97208	-0,5031	-0,7943	0,369462	-0,15213	-0,64996	0,851338	1,294358
7	-0,72765	0,368721	1,383519	-1,33587	-0,32328	1,477862	-1,24083	0,956876	2,246092
8	-0,37556	0,536321	0,440211	-0,7943	-0,18473	-0,15213	-0,35452	0,7458	1,294358
9	1,736974	1,039122	-0,5031	-1,33587	-1,01602	-0,15213	-0,64996	-1,68157	-0,60911
10	-0,72765	1,039122	0,440211	-0,7943	0,923655	-1,29313	1,713523	0,7458	0,342624
11	-0,02347	0,03352	1,383519	-0,25273	-0,32328	-0,96713	0,531783	-1,25942	-1,08498
12	-0,37556	-1,47488	-1,44641	0,288836	-0,04618	1,477862	-1,24083	-0,41512	-0,89463
13	-1,07974	-0,63688	-1,44641	0,830405	-0,18473	-0,15213	-0,64996	-0,30958	-0,70428
14	-1,78392	-0,13408	0,440211	1,371973	0,369462	-0,96713	-0,05909	-0,20404	-0,60911
15	1,384885	-0,80448	-1,44641	-0,25273	0,646558	-0,80413	0,531783	0,007036	0,342624

In the first stage of the cluster analysis, we find out whether the selected objects of study (Forestris) form "natural clusters". To do this, use the method of hierarchical classification, in which we select the following characteristics: Amalgamation (joining) rule: Complete Linkage, Single Linkage and Ward's method; Distance metric is: Euclidean distances (non-standardized). The obtained clustering results are shown in figures 3-6.

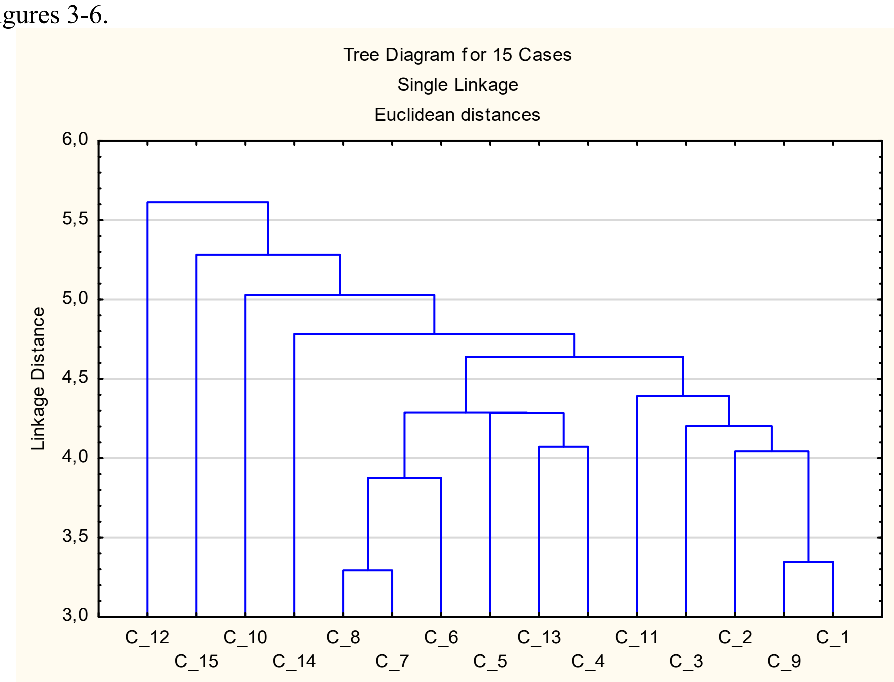


Fig. 3. Tree diagram for 15 forestries (Single Linkage).

Complete Linkage defines a relationship between clusters as the longest distance between two objects in different clusters ("the farthest neighbor"). Distance metric is Euclidean distances is a geometric distance in n-dimensional space and is calculated by the formula:

$$d(x, y) = \sqrt{\sum_{i=1}^{n}(x_i - y_i)^2} \quad (9)$$

From the obtained calculations and the constructed dendrogram it is possible to draw conclusions that the investigated forestries form 5 natural clusters. Let's test the above hypothesis by dividing the original data of K-means clustering into 5 clusters and check the significance of the difference between the obtained groups.

The best results in terms of meaningful interpretation were obtained by using an iterative method of cluster analysis, in particular the K-means clustering algorithm with division into three clusters. After the procedures performed by using the previously mentioned computer program, the results of clustering were obtained, which are shown in figure 6.

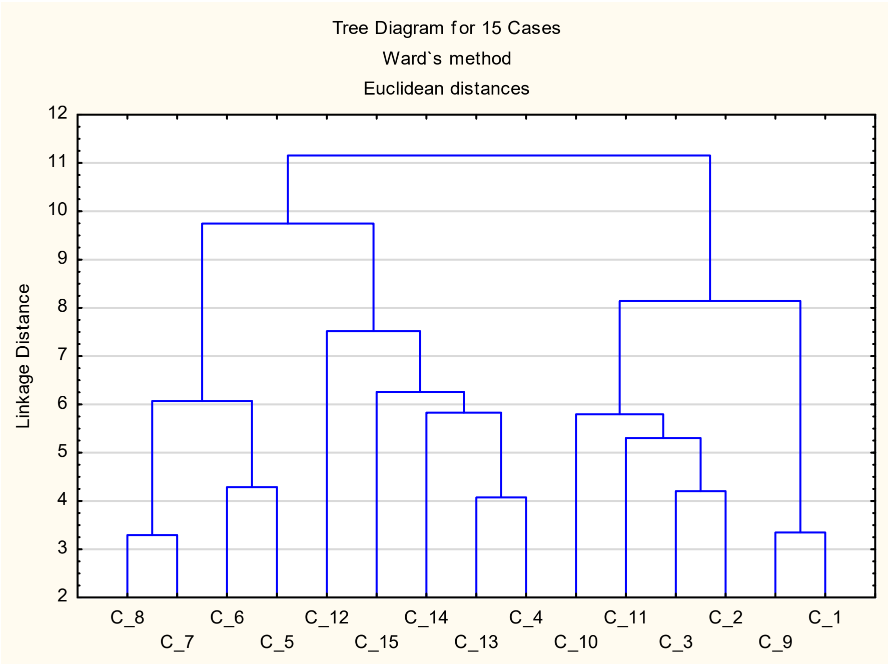


Fig. 4. Tree diagram for 15 forestries (Ward's method).

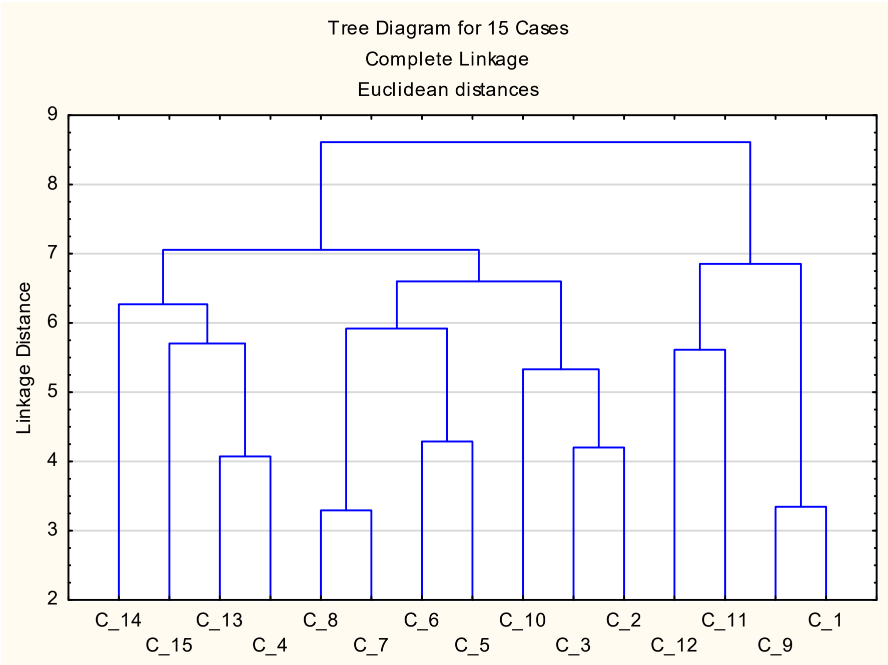


Fig. 5. Tree diagram for 15 forestries (Complete Linkage).

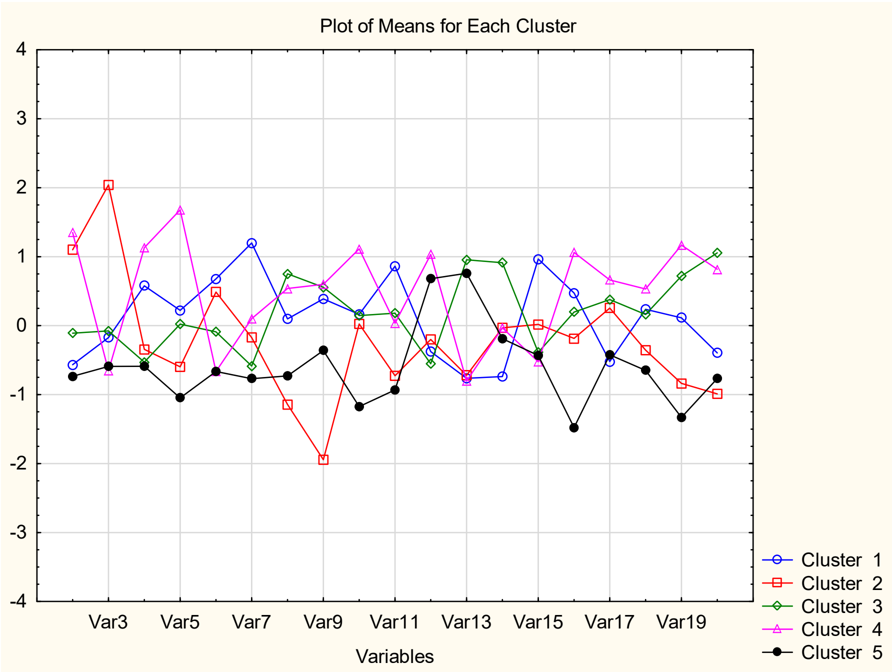


Fig. 6. Average level of normed values of indicators for the selected clusters.

To check the quality of the clustering, a variance analysis was performed, the results of which (table 5) indicate the relative quality of the clustering procedure: intergroup values of variances (Between SS) do not significantly exceed intragroup values (Within SS), except for 9 factors and the level of p- significance reaches the optimal value only for 9 characteristics.

Next, for qualitative clustering in the cluster analysis, we include the 9 most significant features of the previously performed analysis of variance. To implement clustering, we use the method of hierarchical classification, in which we select the following characteristics: Amalgamation (joining) rule: Complete Linkage, Single Linkage and Ward's method; Distance metric is Euclidean distances (non-standardized). The obtained clustering results are shown in the figures 7-9.

From the obtained calculations and the constructed dendrogram we can conclude that the studied forests form 4 natural clusters. Let's test the above hypothesis by dividing the original data of K-means clustering into 4 clusters and check the significance of the difference between the obtained groups.

The best results in terms of meaningful interpretation were obtained by using an iterative method of cluster analysis, in particular the K-means clustering algorithm with division into four clusters. After the procedures performed by using the previously mentioned computer program, the results of clustering are obtained, which are shown in figure 10.

Table 5. Analysis of variance.

Variable	Analysis of Variance (Апробація) Between SS	df	Within SS	df	F	signif. p
Var2	9,01688	4	4,98312	10	4,52371	0,024111
Var3	10,37888	4	3,62112	10	7,16553	0,005449
Var4	6,29275	4	7,70725	10	2,04118	0,164208
Var5	9,85135	4	4,14865	10	5,93648	0,010325
Var6	4,56180	4	9,43820	10	1,20833	0,366127
Var7	8,97487	4	5,02513	10	4,46500	0,025055
Var8	7,08283	4	6,91717	10	2,55987	0,103927
Var9	10,50677	4	3,49323	10	7,51937	0,004596
Var10	6,80881	4	7,19119	10	2,36707	0,122708
Var11	6,76535	4	7,23465	10	2,33783	0,125890
Var12	5,38430	4	8,61570	10	1,56235	0,258010
Var13	10,04167	4	3,95833	10	6,34211	0,008287
Var14	5,62076	4	8,37924	10	1,67699	0,230981
Var15	5,44553	4	8,55447	10	1,59143	0,250834
Var16	9,92733	4	4,07267	10	6,09387	0,009470
Var17	3,19534	4	10,80466	10	0,73934	0,586234
Var18	2,41334	4	11,58666	10	0,52072	0,722951
Var19	11,55609	4	2,44391	10	11,82130	0,000831
Var20	10,15565	4	3,84435	10	6,60426	0,007224

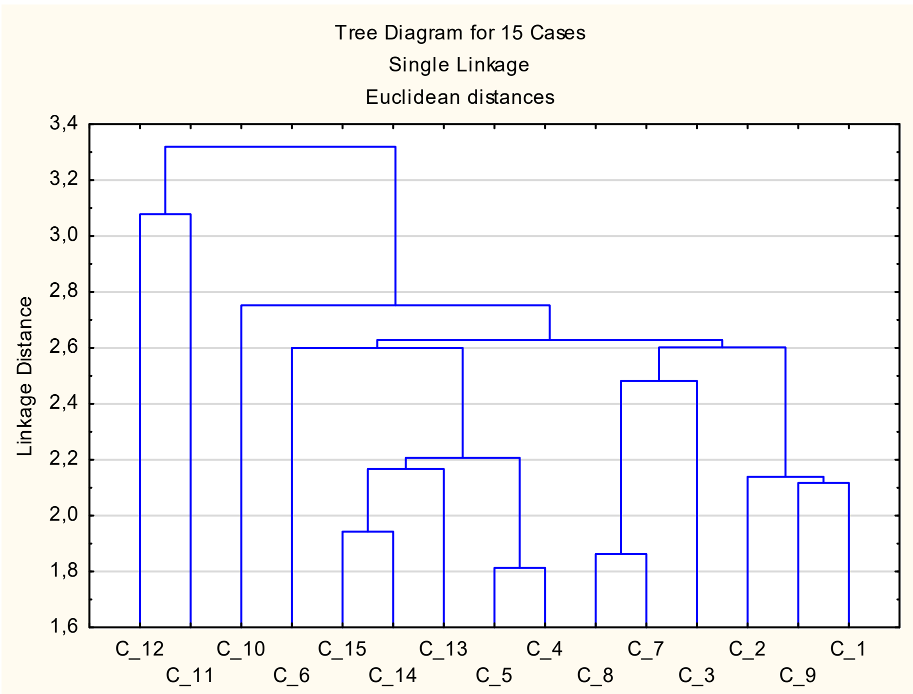


Fig. 7. Tree diagram for 15 forestries (Single Linkage).

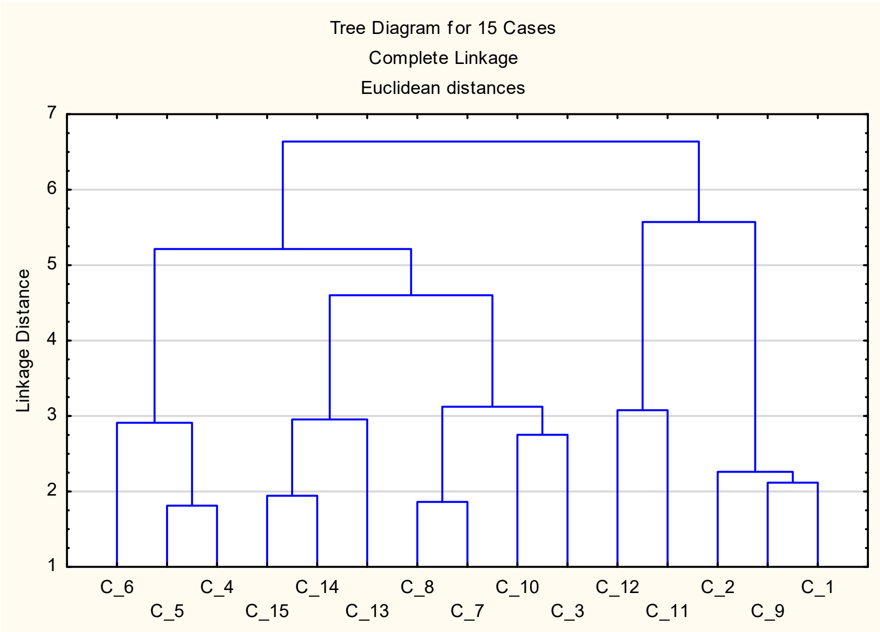


Fig. 8. Tree diagram for 15 forestries (Complete Linkage).

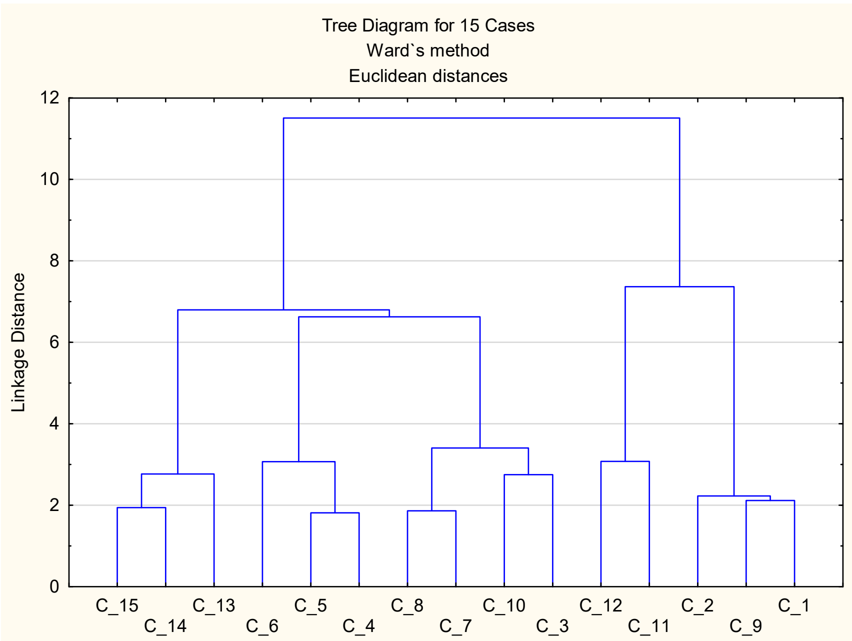


Fig. 9. Tree diagram for 15 forestries (Complete Linkage).

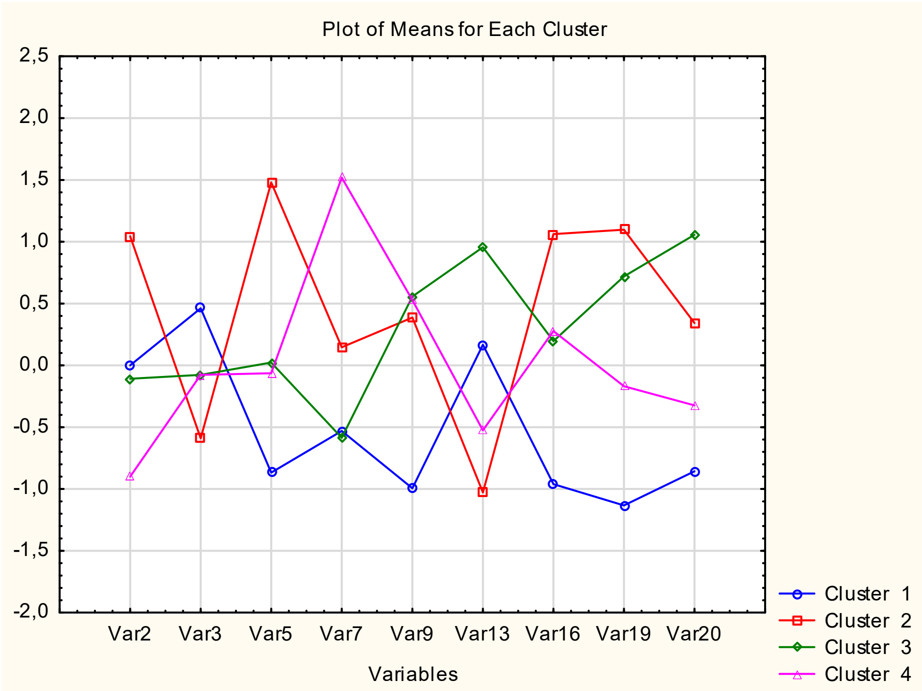


Fig. 10. Average level of normed values of indicators for the selected clusters.

The distance between the clusters, which are selected by K-means Clustering Method, was calculated by a simple Euclidean distance and are presented in table 6.

Table 6. Euclidean distances between clusters.

Cluster Number	Euclidean Distances between Clusters (Апробація) Distances below diagonal Squared distances above diagonal			
	No. 1	No. 2	No. 3	No. 4
No. 1	0,000000	2,445802	1,394819	1,277132
No. 2	1,563906	0,000000	1,068632	1,250228
No. 3	1,181025	1,033747	0,000000	1,106422
No. 4	1,130103	1,118136	1,051866	0,000000

To check the quality of the clustering, a dispersion analysis was performed, the results of which (table 7) indicate the high quality of the clustering procedure: intergroup values of variances (Between SS) significantly exceed intragroup values (Within SS), and the level of p-significance is much better than the normative (0.05).

Also, the contribution to the division of objects into groups is characterized by the values of Fisher’s criterion (F-criterion) and its significance level (p): the higher the values of the first and the smaller the values of the second, the better the clustering. For all parameters, without exception, the significance level approaches 0, which indicates the high statistical significance of the F-criterion. Depending on the levels of these indicators, forestry was grouped into four clusters (table 8).

Table 7. Euclidean distances between clusters.

Variable	Analysis of Variance (Апробація)					
	Between SS	df	Within SS	df	F	signif. p
Var2	9,79064	4	4,209359	10	5,81481	0,011049
Var3	10,46539	4	3,534605	10	7,40210	0,004860
Var5	10,59468	4	3,405316	10	7,77805	0,004073
Var7	10,19347	4	3,806533	10	6,69472	0,006896
Var9	10,47683	4	3,523172	10	7,43423	0,004786
Var13	10,41854	4	3,581461	10	7,27255	0,005172
Var16	10,38962	4	3,610375	10	7,19428	0,005373
Var19	12,47685	4	1,523151	10	20,47867	0,000083
Var20	8,85809	4	5,141908	10	4,30681	0,027826

Table 8. Forestry clusters.

Forestry group	Forestry
1 cluster	1, 2, 9, 11, 12
2 cluster	4, 5, 6
3 cluster	3, 7, 8, 10
4 cluster	13, 14,15

4 Results and conclusion

For the correct selection of clusters, a comparative analysis of several methods was performed: the arithmetic mean, hierarchical methods followed by dendrogram construction, K-means Clustering Method, which refers to reference methods in which the number of groups is specified by the user. The cluster analysis using different methods allows us to state that their combination helps to select reasonable groupings, visually illustrate the clustering procedure and rank the obtained clusters.

Thus, the results of the cluster analysis on 9 analytical grounds confirmed the hypothesis of separation of 4 clusters from 15 forestries. The first cluster is formed by five forestries 1, 2, 9, 11, 12, which are characterized by an average area of recreational territories, biggest number of recreational sites and recreational capacity, lowest quality factor of forest vegetation, proportion of total forestry costs on maintenance of recreational sites, wear coefficient of recreational fixed assets, cost amount on marketing activities of recreational territories, proportion of foreign investments in recreational activities financing, quantity of grants (programs) won to finance recreational activities. The second cluster is formed by three forestries 4, 5, 6. This cluster is characterized by the highest level of recreational territories, quality factor of forest vegetation, cost amount on marketing activities of recreational territories, proportion of foreign investments in recreational activities financing, an average level of recreational capacity and number of recreational sites, lowest level of proportion of total forestry costs on maintenance of recreational sites, wear coefficient of recreational fixed assets, quantity of grants (programs) won to finance recreational activities. The third cluster includes four forestries 3, 7, 8, 10, which have the following characteristics: the highest level of wear coefficient of recreational fixed assets, recreational capacity and quantity of grants (programs) won to finance recreational activities, average area of recreational territories, number of recreational sites and recreational capacity, quality factor of forest vegetation, cost amount on marketing activities of recreational territories and quantity of grants (programs) won to finance recreational activities, lowest proportion of total forestry costs on maintenance of recreational sites. The fourth cluster includes 3 forestries 13, 14, 15 and is characterized by the highest level of the proportion of total forestry costs on maintenance of recreational sites and wear coefficient of recreational fixed assets, lowest number of recreational sites and recreational capacity, quality factor of forest vegetation, recreational capacity, cost amount on marketing activities of recreational territories and quantity of grants (programs) won to finance recreational activities and quantity of grants (programs) won to finance recreational activities, the lowest of recreational sites.

For the proper selection of the clusters, a comparative analysis of several methods was performed: arithmetic mean, hierarchical methods followed by dendrogram construction, K-means method, which refers to the reference methods in which the number of groups is specified by the user. The cluster analysis, using different methods, allows us to state that their combination allows to select reasoned groupings, visually illustrate the clustering procedure and rank the obtained clusters.

The obtained results of clustering will help to develop separate development strategies for each isolated cluster, which will increase the efficiency of recreational areas management in the future. In addition, the results can be used to form an effective model for the development of recreational clusters.

References

1. Anderson, N., Ford, R.M., Bennett, L.T., Nitschke, C., Williams, K.J.H.: Core values underpin the attributes of forests that matter to people. Forestry: An International Journal of Forest Research **91**(5), 629–640 (2018). doi:10.1093/forestry/cpy022
2. Bell, S.: Forest recreation: New opportunities and challenges for forest managers. Rad. Sumar. inst. Izvanredni broj **10**, 155–160 (2005)
3. Bielinskyi, A., Soloviev, V., Semerikov, S., Solovieva, V.: Detecting stock crashes using Levy distribution. CEUR Workshop Proceedings **2422**, 420–433 (2019)
4. Derbentsev, V., Semerikov, S., Serdyuk, O., Solovieva, V., Soloviev, V.: Recurrence based entropies for sustainability indices. E3S Web of Conferences **166**, 13031 (2020). doi:10.1051/e3sconf/202016613031
5. Gerstenberg, T., Baumeister, C.F., Schraml, U., Plieninger, T.: Hot routes in urban forests: The impact of multiple landscape features on recreational use intensity. Landscape and Urban Planning **203**, 103888 (2020). doi:10.1016/j.landurbplan.2020.103888
6. Irland, L.C., Adams, D, Alig, R., Betz, C.J., Chen, C.-C., Hutchins, M., McCarl, B.A., Skog, K., Brent, L.: Assessing Socioeconomic Impacts of Climate Change on US Forests, Wood-Product Markets, and Forest Recreation: The effects of climate change on forests will

trigger market adaptations in forest management and in wood-products industries and may well have significant effects on forest-based outdoor recreation. BioScience **51**(9), 753–764 (2001). doi:10.1641/0006-3568(2001)051[0753:ASIOCC]2.0.CO;2

7. Juutinen, A., Kosenius, A., Ovaskainen, A.: Estimating the benefits of recreation-oriented management in state-owned commercial forests in Finland: A choice experiment. Journal of Forest Economics **20**(4), 396–412 (2014). doi:10.1016/j.jfe.2014.10.003
8. Kryzhanivskyi, Ye., Horal, L., Shyiko, V., Holubchak, O., Mykytiuk, N.: Economic and Mathematical Modelling for Evaluation of Potential Recreational Forest Utilization. Advances in Economics, Business and Management Research **99**, 173–178 (2019)
9. Lee, K.-C., Kang, K.-R.: Classification of Recreation Forests through Cluster Analysis. Journal of the Korean Institute of Landscape Architecture **37**(1), 9–17 (2009)
10. Meyer, M.A., Rathmann, J., Schulz, C.: Spatially-explicit mapping of forest benefits and analysis of motivations for everyday-life's visitors on forest pathways in urban and rural contexts. Landscape and Urban Planning **185**, 83–95 (2019). doi:10.1016/j.landurbplan.2019.01.007
11. Murphy, W.: Forest Recreation in a Commercial Environment. In: Small-scale forestry and rural development: The intersection of ecosystems, economics and society. Proceedings of IUFRO 3.08 Conference, Hosted By Galway-Mayo Institute of Technology, Galway, Ireland, 18-23 June 2006, pp. 347–356
12. Shin, H.-K., Shin, H.-C.: Market Segmentation on Recreational Forest Visitors by Cluster Analysis. The Journal of the Korea Contents Association **10**(3), 364–372 (2010). doi:10.5392/JKCA.2010.10.3.364
13. Soloviev, V., Bielinskyi, A., Solovieva, V.: Entropy Analysis of Crisis Phenomena for DJIA Index. CEUR Workshop Proceedings **2393**, 434–449 (2019)
14. Soloviev, V.N., Belinskiy, A.: Complex Systems Theory and Crashes of Cryptocurrency Market. Communications in Computer and Information Science **1007**, 276–297 (2019)

Strategic priorities of innovation and investment development of the Ukraine's economy industrial sector

Valentyna Stadnyk[0000-0002-2095-3517], Pavlo Izhevskiy[0000-0001-6457-399X], Nila Khrushch[0000-0002-9930-7023], Sergii Lysenko[0000-0001-7243-8747], Galyna Sokoliuk[0000-0003-2589-5376] and Tetjana Tomalja

Khmelnitskiy National University, 11 Instytutska Str. Khmelnitskiy, 29016, Ukraine
stadnyk_v_v@ukr.net, izhevskypavlo@gmail.com, nila.ukr@gmail.com, sigoruk@ukr.net, sokoliukg@gmail.com, sts8383@ukr.net

Abstract. The problem of determining the investment priorities of the national economy development has been actuated. It has been argued that the formation of institutional preferences for activation of industry investment processes should be carried out taking into account the potential ability of each sectoral group enterprises to increase the added value. The scientific and methodical approach for sub-sectors investment attractiveness assessment has been formed on the example of the Ukrainian food industry. It has been recommended to use for this substantiated set of relative performance indexes which are duplicated in aggregate statistical state surveys based on the enterprise's financial statements. It has been formed the recommendations for the investment priorities of food industry development in Ukraine which are based on the appropriate calculations made by the TOPSIS and CRITIC methods. Methods of economic-statistical and comparative analysis were used for structural and dynamic characteristics of the Ukraine industrial enterprises activities. Given that innovation processes should also cover small and medium-sized industrial enterprises, whose resource opportunities are mostly limited, it is proposed to expand them within the framework of a strategic partnership. Graphic modeling methods have been used to visualize the process of building the business structures resource potential on the basis of their strategic partnership. The influence of the motivational environment on the value of organizational relations within the partnership has been formalized.

Keywords: food industry, TOPSIS method, CRITIC method, cluster analysis, strategic partnership, innovative potential, value of partnerships.

1 Introduction

Ensuring sustainable and dynamic development of the country is a key task of the economic policy of each state. This policy is developed taking into account existing and future social (including global) needs and the ability of national economic entities to meet them, relying on their (specific for each) competitive advantages and opportunities for their build-up. The presence of a significant number of such entities in certain industries determines the country's place in the world division of labor. Quite a lot of countries use their natural resources to create competitive advantages, which can be unique and have value in the world market, even with a minimal degree of their processing. However, history shows that such countries, which mostly offer raw materials for the exchange on the world markets, are increasingly lagging over time behind in the level of economic dynamics and quality of life of the population from other countries that develop their ability to create new consumer values.

This ability is especially necessary in the context of the global economic crisis, which is beginning now. The COVID-19 pandemic has become a trigger for this crisis, whose rapid deployment casts doubts on the globalization benefits and resuming discussions about its scale feasibility. And not only in the millions of people life-saving light, but also in the many businesses types inevitable economic losses context, this can disrupt the national economies established structures and threaten their economic security. This crisis pandemic nature has worsened consumer demand in many market segments. In addition, the quarantine restrictions that are taken by the most countries governments for slowing the COVID-19 proliferation pace have led to the numerous logistics chains disruption and this greatly complicated many business processes [20].

As a result, a significant decrease in industrial production is expected (depending on industry specialization). This will lead to the major job losses – especially on enterprises that manufacture products for production systems reproduction of other industries economic entities. Under crisis circumstances, such reproduction is usually complicated – both due to lack of investment resources and because of uncertainty of the future state of consumer markets. And businesses are increasingly unable to produce high value-added products and are focused mainly on natural resources intensive and predatory use, this forms the country's export structure. The presence of such disparities in the modern Ukraine's economy is extremely urgent to overcome them for enhancing the country's ability to withstand the threats of global challenges and maintain an economic security proper level.

The formation of such ability is of paramount importance to Ukraine, because the national economy in the world division of labor still occupies the position of raw material supplement to the economies of the developed countries. However, the current stage of its development is characterized by a change in the vectors and priorities of economic cooperation due to European integration processes. It is important that, in the course of these changes, the Ukrainian economy could take the best positions in the global market from the standpoint of macroeconomic perspectives. This requires a sound approach to determining the directions of structural changes and their implementation through appropriate investment processes, supported at the institutional level through the tools of state economic policy.

Many researchers deal with the problem of the formation and implementation of an effective state economic policy in different market contexts. In particular, in Ukraine, this problem was reflected in the works of Dmytro Lukianenko [12], Oleksandr Bavyko [3], Svitlana Radziyevska [19], Inesa Khvostina [7], Marharyta Sharko [21], Oleksandra Moskalenko [13] and many others. The authors express different views about the instruments and directions of institutional support for the development of the national economy, but it is indisputable that it should ensure the growth of its international competitiveness.

A special interest in this problem arose in recent years, as the crisis phenomena were the reason for that in the socio-economic processes of many developed countries, whose governments supported the dominance of market mechanisms of macroeconomic regulation. Ideally, these mechanisms should ensure capital inflows (inflows of investments) into those sectors of the economy in which greater added value is created, that would ensure the conditions for sustainable socio-economic growth. In particular, this was referred to in the UN Recommendations for those countries (2013) for post-2015 development, which justified the need to transform economies in favor of employment and overall growth based on value added and productivity through industrialization as the central strategy for achieving this transformation [14].

However, Mancur Olson once emphasized that, with strong lobbying, “the choice of state support sectors can create preferences not to those who could use them most effectively from the point of view of the interests of the country” [15]. It can be confirmed by the fact that quite often unjustified preferences were received by individual economic entities in Ukraine at different periods of market transformation. Proceeding from the current political and economic realities in Ukraine, it is important to more reasonably approach the definition of the directions for the formation of such preferences from the point of view of improving the dynamics of overall economic growth.

Indicating the decisive role of the state in regulating economic processes, Justin Yifu Lin focuses on the strategic use of limited resources of the state for the targeted support of certain industries in which a comparative advantage is probable [11]. In this, his position coincides with the position of Michael E. Porter, who highlighted the directions for the formation of these comparative advantages in the context of the national economy (so-called “diamond of competitiveness”) [16]. In its composition, the opportunities of the industrial sector, which in different countries have their structural completeness, take on the prominent place. But most of all it creates the largest number of jobs in it and accumulates the largest share of value added.

Supporting the necessity of developing the industrial sector in each country, Justin Yifu Lin noted that “specific political levers and the institutional framework for generating of optimal industrial policy results” should be defined in appropriate contexts [29, p. 9]. Depending on the tasks to be solved in the course of structural changes in the national economy, as well as on the institutional characteristics of the economic environment, different forces and vectors of leverage should be chosen that will form the economic basis of the regulatory mechanism. In our opinion, the vectors and levers of influence should deal primarily with investment processes and form from the standpoint of the value of investments in certain areas of economic activity for the national economy as a whole.

Despite the variety of such tools, in their totality a significant place is given to those that are aimed at intensifying entrepreneurial activity, especially in areas where large industrial corporations do not see the opportunity to expand mass production, competing with their own kind. It is obvious that in order to maintain the sustainable development of Ukraine’s economy in the global economic crisis context, it is important to prevent further degradation of those areas of economic activity that are potentially capable for increasing the country's production and economic potential. At the same time, the defining emphasis in the regulatory mechanism formation should be its tools and levers general focus on increasing the national economic system ability to compete in the global market.

It is obvious that this ability is formed through the results of the most productive and powerful businesses, able to choose the right strategy and tactics of behavior in the market segment, which they define as promising. And the activities are not so much disparate as integrated, when less powerful participants are involved in business processes.

This approach to development management for economic growth fits, on the one hand, into the concept of sustainable development (in the sense of continuity and complementarity of change processes), and on the other hand it relies on the resource concept (which is most fully formulated by Birger Wernerfelt [31]). and the theory of the firm (conceptually formulated by Ronald H. Coase [4], and expanded by Oliver E. Williamson [32], taking into account the specifics of integration processes between the owners of different types of resources).

Investigating the integration processes motives Tatiana Kolmykova, Olena Lukianykhina, Nataliia Baistriuchenko and Vadym Lukianykhin [8] emphasize that integration expands the participants' resource capabilities and it improves their ability to effective innovation activity and contributes to the national economy innovative development. At the same time, Viktoriia Dergachova and Tetiana Tryhlib [5] point to the necessity of a strategic analysis of the integration different vectors prospects based on the all participants competitive advantages in the integration formation. Similar positions are held by Vitalina Babenko et al [2], who used methods of economic and mathematical modeling to develop forecasts of the economic entities' integration effectiveness in different scenarios.

It is obvious that the greater number of market participants working in mutually beneficial cooperation for the result, the more important role the business processes integrator should play in their spatial and temporal structuring. After all, he will be the initiator of the innovation and investment project, taking on all its risks and rules of the relevant regulatory mechanism. Moreover, his interests will dominate in the "economic growth points" selection and the integrated partner network formation. However, to ensure overall economic growth, these interests should not run counter to the national, in particular, not to increase the scale of Ukraine's exports raw material specialization.

So, the expediency of state support for economic entities innovative development investment projects should be assessed both from a macroeconomic point of view (for these processes effective management in the intersectoral context and taking into account national interests in the implementation of the Association Agreement with the EU) and economic interests market participants (micro level) because this is what determines their strategic choice.

The purpose of the paper is to determine the investment priorities of structural and technological changes in the Ukrainian economy in the context of the implementation of the sustainable development concept.

2 Research methodology

By the methods of logical analysis it has been argued the need of investment priorities argumentation of the national economy development according to the industry-specific based on synthetic indexes which are characterize the financial and economic dynamics of the industry enterprises in terms of their components of liquidity, debt management, productivity, profitability and changes in sales volumes and added value.

The method TOPSIS (Technique for Order Preference by Similarity to an Ideal Solution) has been used for determination the relative deviation from the benchmark by the groups of synthetic indexes of food industry enterprises activity from different sub-sectors. The weighting coefficients of group indexes were determined using the CRITIC (Criteria Importance Through Intercriteria Correlation) method.

Methods of economic-statistical and comparative analysis were used for structural and dynamic characteristics of the Ukraine industrial enterprises activities. Graphic modeling methods have been used to visualize the process of building the business structures resource potential on the basis of their strategic partnership. The influence of the motivational environment on the value of organizational relations within the partnership has been formalized.

3 Results

3.1 Substantiation of priority directions of Ukraine`s economy industrial sector enterprises development investment

The great amount of losses of national economies from the permanent world financial and economic crises and the further aggravation of geo-economic contradictions between countries of the world economy for markets and resources forced to raise again the question of the effectiveness of market mechanisms for ensuring the competitiveness and sustainable development of the national economy as a coherent economic system. Unfortunately, the realities are that national interests are increasingly giving way to the interests of global economic players, which are powerful multinational companies. This confirms the thesis of I. Wallerstein that the development of the global economy is most likely to take place according to the scenario of “fake transformation, superficial transformation” of national economies, whose purpose is “the inviolability of the existing inequality” in the global economic space, because this is in line with interests of the world economy leaders [30, p. 45].

In Ukraine, structural processes in the economy also develop according to the scenario of “fake transformation” in the interests of global economic players. They are interested primarily in expanding its resource base and markets. The main scopes of direct foreign investment are directed in the relevant industries. According to our calculations (made on the basis of official statistics), in the last 5 years only 1.7–2.5 % of the total volume of capital investments is directed into the machine-building industry, which among other industrial sectors should be able to create the largest volumes of value added. In the food industry, which increases the degree of processing of agricultural raw materials (creates value added within Ukraine) it is only 4.21–6.15 % ((table 1, built on data [27]).

Table 1. The structure of capital investment in the industrial sector of Ukraine's economy.

Indicator	Years					
	2014	**2015**	**2016**	**2017**	**2018**	**2019**
Total capital investment, UAH billion:	219.4	273.1	359.2	448.5	578.7	624.0
Investments in the processing industry in general: – volume, billion UAH	42.5	46.2	62.2	73.9	100.9	105.9
– part, %	19.37	16.92	17.32	16.48	17.44	16.97
to the machine-building industry; – volume, billion UAH.	5.5	6.3	7.6	10.3	12.5	11.0
– part in the total CI, %	2.50	2.30	2.13	2.30	2.16	1.76
to the food industry: – volume, billion UAH	13.5	13.5	21.3	18.9	30.2	31.9
– part in the total CI, %	6.15	4.94	5.93	4.21	5.22	5.11

As can be seen from the above data, the dynamics of investment processes in key sectors of the processing industry can not be called such that corresponds the requirements of the national economy sustainable development. Although in absolute terms the volume of capital investment during the study period was constantly growing, the part of investment to the manufacturing decreased from 19.37 % in 2014 to 16.97 % in 2019. And the most noticeable was the investment decrease to the machine-building industry, it was less than 2 % of the total their volume. In 2019 part of investments to the food industry also decreased – by 5.11 % against 5.22 % in the previous year. And compared to 2014 (6.15%) this decrease is even more significant.

At the same time, capital investments to the agricultural production for 5 years ranged from 8.6 to 14.3% of their total volume. That means that investors in Ukraine prefer the development of this industry, rather than processing enterprises. With such dynamics of investment processes, Ukraine loses the opportunity to move away from its raw material orientation. This is logical in terms of investment return – the profitability of agricultural production operating activities is much higher than in the food industry (fig. 1, built on data [28]).

However, while maintaining such dynamics of investment processes, Ukraine loses the opportunity to move away from its raw material orientation. E. Reinert, well-known expert on economic growth, stressed that all rich countries, without exception, became rich in the same way, based on a common strategy for all of them, which is to abandon the raw material orientation of production and exports in favor of manufacturing [18]. After all, this provides a significant increase part of value added in the gross domestic product structure.

It should be emphasized that the development of the food industry should be one of the priorities in Ukraine. Being equipped with the latest technology, it will be the leading link in the chain of creation of consumer values, increasing the value added in the processing of agricultural raw materials of domestic origin. Processing enterprises of the food industry can combine the economic entities of the national economy involved in the whole process of food production in the interval "from raw materials to the final consumer". As the main integrators of the promotion of products to the foreign market, they will contribute to an increase in domestic commodity circulation in Ukrainian agriculture and an increase in external demand for high-quality food products.

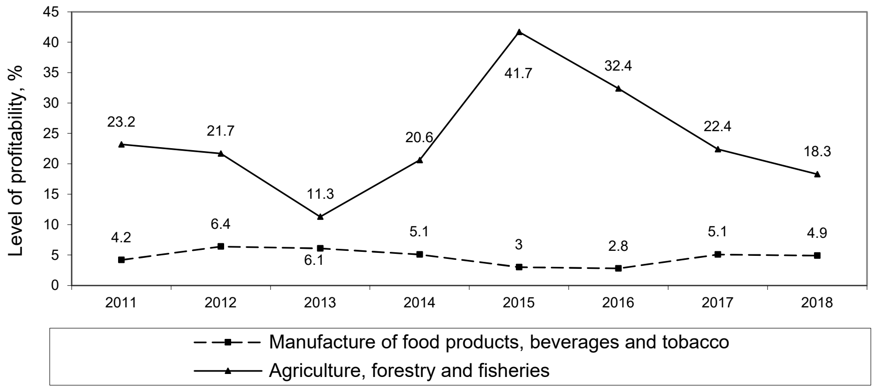


Fig. 1. Comparative dynamics of Ukraine's enterprises operating activity profitability in the spheres of agricultural products production and processing.

Obviously, in order to ensure sustainable economic dynamism, Ukraine must develop its industrial sector by creating new jobs with attractive working conditions. However, the formation of institutional preferences for activating investment processes in industry should be made taking into account the potential ability of enterprises of each branch group to increase the value added, and also to be competitive in the strategic perspective.

Therefore, the determination of the priority and investment attractiveness of the national economy sectors is possible only with the determination of their dynamics and developmental capacity, which is determined by a number of relative performance indicators in international business practice and is duplicated in aggregate statistical state surveys based on the financial statements of enterprises grouped according to sectoral characteristics and value added. The grouping and the procedure for calculating these indicators are given in table 2 [22].

On the basis of grouped indicators, first of all, it is necessary to determine the synthetic indicator of economic and financial efficiency of the national economy of Ukraine in terms of the opportunities for the development of a particular industry that can be ensured by the TOPSIS (Technique for Order Preference by Similarity to an Ideal Solution) method [17; 10] taking into account the following steps:

- definition of the hierarchical structure of the multicriteria problem of the levels of the national economy development;
- normalization of the values of economic and financial indicators;
- substantiation of criteria and features of financial and economic indicators on the basis of assigning weighting coefficients to them using the CRITIC (Criterion Importance Through Intercriteria Correlation) method [1];

- determination of the synthetic measures for economic and financial efficiency using the TOPSIS method.
- linear ordering of the national economy branches.

Table 2. Indicators of evaluation of economic and financial situation.

Indicator	Calculation procedure
Group of financial liquidity	
Current liquidity	Current assets / current liabilities
Instant liquidity	[Current assets – stocks – accrual] / current liabilities
Cash flow	Short-term investments / Current liabilities
Debt burden and the ability to service it	
Total debt obligation	Commitments and provisions for liabilities / assets
Commitment by tangible assets	Fixed assets / long-term liabilities
Coverage of liabilities with financial surplus	[Operating profit + depreciation] / total liabilities
Coverage of interest with operating profit	Operating profit / interest
Productivity	
Total turnover of assets	Revenue / asset summary
Productivity of operating assets	Gross value added / operational assets
Productivity of fixed assets	Gross value added / fixed assets
Labor efficiency	Gross value added / number of employees
Cost Index	Operating expenses / operating income
Profitability	
Operational	Operating profit × 100 / operating income
Operating assets	[Operating profit + depreciation] × 100 / operating assets
Invested capital	Operating profit × 100 / invested capital
ROA (Return on Assets)	(Net profit / average assets) 100
ROE (Return on Equity)	(Net profit / average annual amount of equity) 100
The dynamics of change	
Revenue from sales	($Revenue_t$ – $Revenue_{t-1}$)100/ $Revenue_{t-1}$
Gross value added	(Value $added_t$ – Value $added_{t-1}$)100/ Value added a_{t-1}

At the stage of formation of the hierarchical structure of the evaluation multicriteria task, the following elements are considered: the main criterion of evaluation, the subcriteria, features and objects of evaluation. The main criterion is positioned at the highest level of the hierarchy and contains several subcriteria (individual criteria contain features that describe the objects of evaluation) [10]. Peculiarities create features of development levels of objects, that is, branches of the national economy.

The choice of characteristics of the economic and financial state of the industries is based on objective and statistical analysis, and in order to exclude highly correlated features, the matrix analysis of the inverse correlation R^{-1} is provided (in the case of excessive correlation of certain features relative to other features, the diagonal elements of the matrix R^{-1} will exceed the unit that will be detected in bad numerical conditionality of the matrix R [10].

The value K of economic-financial characteristics for N of statistical units (branches of economy), established on the basis of objective and statistical analysis, are summarized in the matrix of dimensional data ($N \times K$):

$$X = \begin{pmatrix} x_{11} & x_{11} & \dots & x_{1K} \\ x_{21} & x_{22} & \dots & x_{2K} \\ \dots & \dots & \dots & \dots \\ x_{N1} & x_{N2} & \dots & x_{NK} \end{pmatrix}, \quad (1)$$

where x_{ij} (i = 1, ..., N), (j = 1, ..., K) indicates the value of the jth feature in the ith statistical unit.

At the second stage, the normalization of functions is ensured in order to combine them into numeric ranges, and different approaches are used for that [10]. In our case, it is expedient to use the linear normalization approach – zeroed unitarization – by means of which one can compare the stimulator and the disintegrator on the basis of a single dimensionality using transformation formulas:

$$z_{ij} = \frac{x_{ij} - min\{x_{ij}\}}{max\{x_{ij}\} - min\{x_{ij}\}} \text{ (stimulator);} \quad (2)$$

$$z_{ij} = \frac{max\{x_{ij}\} - x_{ij}}{max\{x_{ij}\} - min\{x_{ij}\}} \text{ (destimulator);} \quad (3)$$

Nominals:

$$z_{ij} = \frac{x_{ij} - min\{x_{ij}\}}{nom\{x_{ij}\} - min\{x_{ij}\}}, x_{ij} \leq nom\{x_{ij}\} \quad (4)$$

or

$$z_{ij} = \frac{max\{x_{ij}\} - x_{ij}}{max\{x_{ij}\} - nom\{x_{ij}\}}, x_{ij} \leq nom\{x_{ij}\} \quad (5)$$

Converted values of z_{ij} features are normalized in the interval $\langle 0,1 \rangle$.

The third stage is devoted to the formation of weighting coefficients for economic and financial indicators using the CRITIC method [1], presented in the form of $W = W_1, W_2, \dots W_K$.

Weighting factors in this specified method are determined taking into account the standard deviation of each of the characteristics and coefficients of correlation between the features, and their vectority is as

$$W_j = C_j / \sum_{k=1}^{K} C_k = 1, j = 1,2, \dots, K \quad (6)$$

follows:

$$C_j = s_{j(z)} \sum_{k=1}^{K} (1 - r_{jk}), j = 1,2, \dots, K \quad (7)$$

where: $s_{j(z)}$ – standard deviation, calculated for the normalized values of the jth feature; r_{jk} – correlation coefficient between jth and kth features, for which $\sum_{j=1}^{K} w_j = 1$.

In the future, the product of the normalized values of the features is determined by weight coefficients:

$$z_{ij}^{*} = z_{ij}w_{ij}, i = 1,2,\ldots,N, j = 1,2,\ldots,K. \quad (8)$$

At the fourth stage, we use a reference method for determining the value of the synthetic feature using the classical TOPSIS method for calculating the Euclidean distance to each object of the model:

- development

$$z^{+}: z^{+} = (\max_i(z_{i1}^{*}), \max_i(z_{i2}^{*}), \ldots, \max_i(z_{iK}^{*})) = (z_1^{+}, z_1^{+}, \ldots, z_K^{+}); \quad (9)$$

- antidevelopment

$$z^{-}: z^{-} = (\min_i(z_{i1}^{*}), \min_i(z_{i2}^{*}), \ldots, \min_i(z_{iK}^{*})) = (z_1^{-}, z_1^{-}, \ldots, z_K^{-}) \quad (10)$$

In the case of zeroed unitarization we will receive:

$$z^{+} = \left(\underbrace{1,1,\ldots,1}_{K}\right) \text{or } z^{-} = \left(\underbrace{0,0,\ldots,0}_{K}\right) \quad (11)$$

After defining an example of development and anti-development, the Euclidean distances of each unit of evaluation will be calculated as follows:

- development

$$z^{+}: d_i^{+} = \sqrt{\sum_{j=1}^{K}(z_{ij} - z_j^{+})^2} \quad (12)$$

- antidevelopment

$$z^{+}: d_i^{-} = \sqrt{\sum_{j=1}^{K}(z_{ij} - z_j^{-})^2} \quad (13)$$

The next step is to calculate the synthetic function value S_i:

$$S_i = \frac{d_i^{-}}{d_i^{+} + d_i^{-}}, 0 \leq S_i \leq 1, (i = 1,2,\ldots,N). \quad (14)$$

The smaller the distance of the desired unit from the developmental standard and, accordingly, farther from anti-development, the closer to 1 will be the significance of the synthetic function.

At the final stage, it is proposed to rank the branches of the national economy in accordance with the calculated synthetic values of economic and financial efficiency.

The multidimensional analysis of sub-sectors of the food industry sectors of Ukraine was carried out in accordance with the described methodology. In the State Classifier of Products and Services SC 016-2010 [25], the food industry is included in the section C (processing industry) and includes enterprises producing goods under codes 10 and 11.

Based on the data of the State Statistics Committee [26], the interim calculation of TOPSIS synthetic indicators for the food industry of Ukraine by the formed five groups (see table 1) was carried out for 2 sub-periods: years of 2013-2015 and 2016-2018. Output results made it possible to specify seven groups of food producers that differ in terms of financial liquidity, debt load and maintenance, productivity, profitability, and dynamics of changes determined by intervals. The results of synthetic TOPSIS values and real values form separate clusters that are subject to cluster analysis. It was conducted taking into account individual concentrations (clusters) during the indicated periods. The results of cluster analysis are presented in tables 3 and 4 [22].

Table 3. Ukrainian food industry clusters according to the resalts of calculation synthetic indexes.

Indicator	**Clusters**							**Medium**
	1	***2***	***3***	***4***	***5***	***6***	***7***	
Current liquidity	1.15	1.41	1.44	1.16	1.31	1.2	2.5	1.36
Instant liquidity	0.84	0.95	0.91	0.64	0.72	0.87	1.68	0.88
Cash flow	0.19	0.25	0.17	0.07	0.13	0.16	0.83	0.2
Synthetic, S_1	**0.39**	**0.39**	**0.34**	**0.31**	**0.29**	**0.39**	**0.51**	**0.36**
Total debt obligation	0.54	0.43	0.43	0.63	0.57	0.47	0.29	0.49
Commitment by tangible assets	4.07	6.25	5.67	4.39	3.88	5.94	14.79	5.56
Coverage of liabilities with financial surplus	5.62	10.59	6.29	3.68	2.73	15.11	70.17	10.46
Coverage of interest with operating profit	0.22	0.38	0.27	0.15	0.14	0.48	0.38	0.28
Synthetic, S_2	**0.31**	**0.45**	**0.40**	**0.23**	**0.26**	**0.47**	**0.56**	**0.36**
Total turnover of assets	2.51	1.6	1.43	1.68	1.48	1.86	1.38	1.75
Productivity of operating assets	0.45	0.53	0.4	0.31	0.28	0.73	0.53	0.44
Productivity of fixed assets	0.81	0.9	0.81	0.70	0.71	1.26	1.08	0.87
Labor efficiency	88.98	120.43	119.26	110.97	117.45	219.16	139.98	124.04
Synthetic, S_3	**0.34**	**0.35**	**0.27**	**0.24**	**0.21**	**0.51**	**0.34**	**0.31**
Operational	3.25	7.90	5.52	3.99	3.39	9.67	5.84	5.47
Operating assets	11.91	18.40	11.46	7.16	6.22	31	10.13	13.53
Invested capital	5.16	9.79	6.21	3.66	2.59	15.08	6.8	6.79
ROA (Return on Assets)	15.77	21.79	15.1	11.82	10.72	29.63	17.11	17.06
ROE (Return on Equity)	12.88	18.99	11.8	10.47	7.25	30.68	13.68	14.61
Synthetic, S_4	**0.38**	**0.54**	**0.44**	**0.34**	**0.31**	**0.7**	**0.43**	**0.43**
Revenue from sales	5.93	4.53	2.2	16.61	- 2.5	7.7	-9.9	5.04
Gross value added	6.94	4.27	6.5	14.9	-3.23	6.25	-9.6	4.94
Synthetic, S_5	0.42	0.42	0.42	0.5	0.36	0.42	0.3	0.42
The general synthetic indicator of the food industry	**0.37**	**0.42**	**0.36**	**0.32**	**0.29**	**0.49**	**0.42**	**0.36**

With the change of the time period of research for 2013-2018 compared to the time period of 2012-2017, the distribution of food industries by clusters did not change, but the overall synthetic indicators in each of the clusters deteriorated slightly. First of all, the situation is explained by the unfavorable situation with the export of traditional finished products of the food industry and its compliance with international quality systems, as well as - high interest rates on enterprise lending.

Table 4. Cluster positioning of the Ukrainian foods industry subsectors.

Years/ clusters		**2015-2017**							
		1	***2***	***3***	***4***	***5***	***6***	***7***	**Together**
2012-2014	***1***	10.1; 10.12 10.13; 10.51	11.07			11.01			6
	2	10.91	10.72	10.89			10.52; 10.71		5
	3		10.86; 10.73	10.61; 10.82; 10.83				10.42	6
	4				10.20; 10.41; 10.85	10.39			4
	5		10.81	11.02		10.32; 11.03			4
	6						10.84; 11.05		2
	7							10.31	1
	Together	5	5	5	3	4	4	2	28

Sub-sectors of the food industry belonging to the first cluster (10.11–10.13 – canned meat and meat products, 10.51 – dairy and cheese products, 11.07 – soft drinks, bottled mineral water, 11.01 – alcoholic beverages, distilled, 10.91 – feed products for farm animals) have an average level of efficiency, which is confirmed by a general synthetic index with a value of 0.38. In general, the indicators influencing its final value were characterized by higher (compared with the average in the industry) levels of liquidity and debt, but lower than the average industry profitability and yield. In general, the sub-sector of this cluster needs external sources of investment more than other enterprises of the food industry of Ukraine.

The cluster's fullness varied across different time segments by enterprises of various sub-sectors (during 2013-2015: 10.11–10.13, 10.51, 11.07 and 11.01, during 2016-2018: 10.11–10.13, 10.51, 10.91). The cluster 2 unites enterprises of sub-sectors with a high level of synthetic index – 0.42, which is affected by low debt levels and high profitability. Sub-sectors of this segment have a high level of return and are attractive both for domestic and foreign investment.

However, the subfield format has changed somewhat during the study period. In 2013-2015 it consisted of: 10.91 – food products for livestock, 10.71 and 10.72 – low-moisture bakery products and flour confectionery products, of long-term and short-term storage, 10.89 – other foodstuffs, 10.52 – ice cream. During the years of 2016-2018: 11.07 – soft drinks, bottled mineral water, 10.72 – bakery products, low humidity, confectionery, flour, long-term storage, 10.73 – macaroni, noodles, couscous and similar flour products, 10.81 – sugar.

The cluster group 3 is coherent in its overall level of synthetic indicator of financial and economic efficiency to cluster number 1 (0.37), but the components of such a result differ by the lower level of indicators of liquidity and debt servicing. At the same time, this cluster is characterized by an average level of profitability and a high level of growth in sales and value added, which forms attractive prospects for long-term investment in this direction of food production. The unchanging core of the group consists of the following enterprises: 10.61 – products of the milling industry, 10.82 –

cocoa products, chocolate and sugar confectionery, 10.89 – tea and coffee, processed. During the period of 2013-2015, the group 10.86 was updated with ready-made food products homogenized for children's and dietary foods, 10.73 – macaroni, noodles, couscous and similar flour products, 10.42 – margarine and similar edible fats, and in 2016–2018 they were changed in the sub-sector 10.89 with other food products, 11.02 – grape wines.

The cluster group number 4 combines enterprises with a low level of total synthetic index (0.32) – low liquidity, profitability and high level of indebtedness compared to other branches of the food industry. However, these sub-sectors show a high positive dynamics of sales volumes and value added. This points to their investment attractiveness but not only in increasing their production capacity, but also in their technological upgrading, which will increase their operating profitability or cost-competitiveness. Invariably the group during 2013–2018 includes: 10.20 – fish, crustaceans and shellfish products, processed and preserved, 10.41 – oils and fats, 10.85 – food sets and prepared meals.

The cluster number 5 is characterized by the lowest level of synthetic indicator in the food industry (0.29), which is determined by the lowest rates of practically all components of financial and economic efficiency. To this group in 2013–2015 enterprises included 4 sub-sectors: 10.81 – sugar, 11.02 – grape wines, 10.32 – fruit and vegetable juices, 11.03 – cider and other fruit wines. In 2016–2018, positions 10.32 and 11.03 remained. The listed sub-sectors require significant investments, and in their economic nature they are complex and unattractive for both domestic and foreign investors.

The highest level of efficiency, as well as investment attractiveness is characterized by the sixth group, to which during the years of 2013–2018 the following sub-sectors of the food industry were included: 10.84 – condiments and spices, 11.05 – beer. The total synthetic rate in the group is 0.49.

The cluster number 7 during the analyzed period was characterized by a general synthetic efficiency indicator of 0.42, which in general exceeds the average in the food industry as a whole. However, despite the rather high financial liquidity and low level of indebtedness of the enterprise, the sub-sectors of this cluster practically did not show a positive dynamics in the scopes of sales of goods and value added, that is, they were not attractive for investment. During 2013–2018, this group included a sub-sector of 10.31 – potatoes, processed and preserved, and also in the years of 2016–2018 there was included a sub-sector 10.42 – margarine and similar edible fats.

The obtained results allowed differentiating the food industry into groups with wide possibilities for analysis, on the basis of which their attractiveness for investment was determined. According to the calculations of synthetic indicators and cluster analysis, it was determined that the most priority investment plans for the period include meat processing, oil fat, fish processing, confectionery, milling branch and cereals, a subsector for the production of food sets and ready-made dishes. The development of these sub-sectors is economically profitable both from the standpoint of the owners of capital and from the point of view of the public interest, since it will ensure the growth of the value added within the national economy.

It is obvious that investment processes in the processing industry should provide not only expanded reproduction (or simple scaling) of production systems, but above all their radical renewal. After all, only such production technologies can provide domestic processing enterprises with competitive advantages in the global market, where the parity of major economic players has been established long time ago. In addition, access to the global market is possible only if the production systems correspond to the international quality standards, which requires the introduction of resource-saving and environmentally friendly technologies. Therefore, certain investment priorities in the food industry should be supported by the state if their technological component correspond the requirements of consumer and environmental safety (HACCP standards).

Similar calculations can be performed to identify priority areas of innovation and investment processes in other areas of industrial activity. In addition, the key argument in favor of choosing priority projects, which should provide investment support from the state, should be their ability to increase not only added value but also the consumer value of the final product. Moreover, for a long period, which will form a stable competitive advantage for industrial producers in target market segments. Therefore, advanced production technologies should be a mandatory component of all investment projects in the processing industry.

Their implementation, among other things, should open wider opportunities in the product innovations development. For the food industry, among the latest technologies characteristics, which are preferred for inclusion in the innovation and investment project, is waste-free (increases the product range due to the deep level of raw materials processing). In machine-building complex this characteristic is flexibility, which in combination with standardization provides not only the necessary conditions for modification of prototypes and product line development for target consumers within the selected business strategies, but also reducing the cost of readjustment if necessary to diversify production.

Unfortunately, recent years the Ukraine`s machine-building complex enterprises are increasingly losing markets. Not only their export potential decreases, but also their ability to meet the needs of domestic consumers. As a result, own revenues are not enough for the needs of modernization and technological renewal. Such conditions do not contribute to the innovative activity of the enterprise and cause further technological degradation and loss of competitiveness. The solution to this problem is possible through the involvement of a strategic investor.

But the problem is that in Ukraine investment support for innovative development projects by economic activity state regulation tools is carried out mainly for large enterprises. However, from the standpoint of social security / stability of the state, it is important to create favorable conditions for the development of productive economic activities not only in the format of large-scale, but also using the advantages of small industrial enterprises.

3.2 Development of innovation and investment opportunities for industrial enterprises on the basis of strategic partnership

Small and medium-sized businesses make a significant contribution to the creation of public goods in any country. Its main advantages are high mobility and sensitivity to market changes. However, turnover (market share) and productivity are highest in large enterprises. To assess the contribution of small and medium-sized businesses to the overall results of Ukrainian enterprises economic activity (table 5, is constructed by [9, pp. 153–157 and p. 231]; the data are given without taking into account the temporarily occupied territories of the Crimea, and parts of Donetsk and Luhansk regions).

Table 5. Structural and dynamic indicators of Ukrainian enterprises activity.

Indicators	**Values of indicators by years**					**Changes, 2018-2014**
	2014	**2015**	**2016**	**2017**	**2018**	
The total number of enterprises engaged in economic activity, thousands of units						
In total	341.0	343.4	306.4	338.3	356.0	15.0
of them: large	0.497	0.423	0.383	0.399	0.446	-0.051
medium	15.906	15.203	14.832	14.937	16.057	0.151
small	324.598	327.814	291.15	322.92	339.37	14.772
Part of enterprises engaged in economic activity, % of the total						
Large	0.1	0.1	0.1	0.1	0.1	0
Medium	4.7	4.4	4.9	4.4	4.5	-0.2
Small	95.2	95.5	95.0	95.5	95.4	0.2
The total number of operating industrial enterprises, thousands of units						
In total	42.2	42.6	39.1	42.0	44.4	2.2
of them: large	0.289	0.233	0.208	0.215	0.237	-0.052
medium	4.791	4.691	4.652	4.745	4.866	0.075
small	37.107	37.640	33.695	37.066	39.322	2.215
Part of operating industrial enterprises, % of total						
Large	0.7	0.6	0.5	0.5	0.5	-0.2
Medium	11.3	11.0	12.1	11.3	11.0	-0.3
Small	88.0	88.4	87.4	88.2	88.5	0.5
The total volume of products sold by enterprises, UAH billion						
In total,	4170.7	5159.1	6237.5	7707.9	9206.1	5035.4
of them: by large	1742.5	2053.2	2391.4	2929.5	3515.8	1773.3
by medium	1723.1	2168.8	2668.7	3296.4	3924.1	2201
by small	705.0	937.1	1177.4	1482.0	1766.2	1061.2
Part of products sold by enterprises by their sizes, %						
By large	41.8	39.8	38.3	38.0	38.2	-3.6
By medium	41.3	42.0	42.8	42.8	42.6	1.3
By small	16.9	18.2	18.9	19.2	19.2	2.3
Total volume of sold industrial products, UAH billion						
In total	1546.6	1887.5	2343.0	2862.3	3302.5	1755.9
of them: by large	932.9	1078.2	1232.2	1537.3	1790.4	857.5
by medium	531.5	691.1	921.1	1092.1	1230.3	698.8
by small	82.16	118.1	152.4	188.4	227.6	145.44
Part of industrial products sold by enterprises by their sizes, %						

Indicators	Values of indicators by years					Changes, 2018-2014
	2014	2015	2016	2017	2018	
By large	60.3	57.1	53.4	54.6	55.1	-5.2
By medium	34.4	36.6	40.0	38.7	37.9	3.5
By small	5.3	6.3	6.6	6.7	7.0	1.7
Part of sold industrial products in the total sales of economic activity different types enterprises, %						
In total	37.1	36.6	37.6	37.1	35.9	-1.21
of them: by large	53.5	52.5	51.5	52.5	50.9	-2.61
by medium	30.8	31.9	33.4	34.4	35.4	4.55
by small	11.7	12.6	12.0	13.0	14.0	2.35

From the data given in table 5 it is seen that in the economy of Ukraine there is a steady trend of reducing the efficiency of large industrial enterprises (if the efficiency is understood as the sales of products and services). Thus, the part of products of medium-sized enterprises in 2018 increased to 37.9% compared 34.4% in 2014, and small enterprises – respectively to 7% against 5.3%. The share of industrial output of large enterprises decreased during this period from 60.3% to 55.1%. At the same time, the quantitative structure of enterprises of all types of economic activity of these sizes has changed in the direction of some increase in the share of small enterprises.

In the industrial sector, the structural ratio of industrial sales changed mainly in favor of medium-sized enterprises, while in large enterprises the dynamics of sales was negative. Thus, in 2014, 53.5% of the total sales of products (works, services) were sold at large enterprises, and in 2016 - only 50.9%. Conversely, in medium-sized enterprises this part increased from 30.8% in 2014 to 35.4% in 2018. In small enterprises, the same trend is observed – in 2014-2018, the part of industrial sales increased by 2.35 % – from 11.7 % to 14.0%.

These trends can be explained if the comparison in the same period to analyze the dynamics of the structural relationship of enterprises of different sizes. As can be seen from the table 1 data, the most significant changes affected large enterprises - their total number decreased from 497 in 2014 to 383 in 2016. However, in subsequent years, their number began to grow and in 2018 economic activity was carried out by 447 large enterprises. This is slightly less than in 2014, but positive trends have emerged. A similar trend can be observed in the industrial sector of the national economy – in 2018, large industrial enterprises became 52 units less than in 2014, although in the worst, in 2016, this difference was 81 enterprises.

In small and medium-sized businesses, the number of industrial enterprises has grown significantly in five years, although by 2016 there was a decrease in their number – more than 3.4 thousand small industrial enterprises ceased operations this year. In the following years, their number began to grow and in 2018 even exceeded the number of those operating in 2014 – by 2.2 thousand units. In 2018 middle industrial enterprises were functioned on 75 units more than in 2014.

Thus, these data indicate growth in Ukraine's contribution of small and medium-sized businesses in the overall performance of industrial enterprises. Using a niche strategy, they can provide many consumer needs and create add value to a much greater extent than large enterprises, which often function as raw materials for international corporations. However, the financial results of enterprises in this sector are only 20.8% of the total financial results of all operating enterprises [9, p. 156]. And this indicates a much lower efficiency of small businesses than large enterprises.

Obviously, one of the classic factors of higher efficiency of large enterprises is the larger scale of their activities, which gives economies of scale, as well as the opportunity to choose the best competitive strategies. For small and medium-sized enterprises, higher business efficiency can be achieved only when they do not just copy the existing business, but choose for themselves an innovative development scenario.

However, small industrial enterprises have limited resources to implement business development projects that could ensure its quantitative growth. These opportunities can grow through strategic partnerships. In our opinion, such a partnership will be strategic in the case of the formation of organizational heterarchy – a form of cooperation that provides growth of adaptive properties of participants through a variety of forms and methods of organizational integration. Including, due to cross-sectoral integration processes, as shown by us in previous researches [23].

The term "heterarchy" was introduced into scientific usage by David Stark – as an organization with horizontal or distributed power, which allows it to self-organize. He emphasized that the ability to adapt is stimulated by the organization of diversity within the enterprise – when different organizational principles coexist in a state of active competition within one enterprise. This ensures the development of constructive organizational reflectivity and allows for recombination of resources, recombination of old organizational forms of economic interaction into a new organizational structure with high adaptive properties [24, p. 119].

Describing the features of organizational interaction in the hierarchy, Stark emphasized that this is a new way of organization, which is neither market nor hierarchical. If hierarchies are based on relations of dependence, and the market is based on relations of independence, then hierarchy is based on relations of interdependence. These relations are characterized by a minimum degree of hierarchy and organizational heterogeneity [24, p. 120–121]. Michael T. Hannan also argued that "the ability to adapt is stimulated by the diversity of organizations: a system that has a greater variety of organizational forms is more likely to find a satisfactory solution in the event of changes in external conditions [6].

Strategic partnership is now increasingly recognized as a successful alternative to market competition of small businesses. The organizational development of such partnership is carried out on the principle of competence and resource complementarity in order to maximize the objective function. Through the development of partnerships, it is possible to enhance mutually the resource potential of all participants in the partnership, expanding the ability of competitive business strategies choice.

The key role in determining the direction of enhancing the participants' innovative potential in such a partnership should be played by the producer of the final products – in fact it is in direct contact with the market and it develop the concept of new products in line with the trends of scientific and technological progress.

The solution to these problems belongs to the sphere of marketing and requires adequate information and analytical support. It is advisable to justify and make appropriate decisions by actively interacting with existing/potential partners, involving them in mutually agreed organizational or technological changes. This is shown schematically in fig. 2.

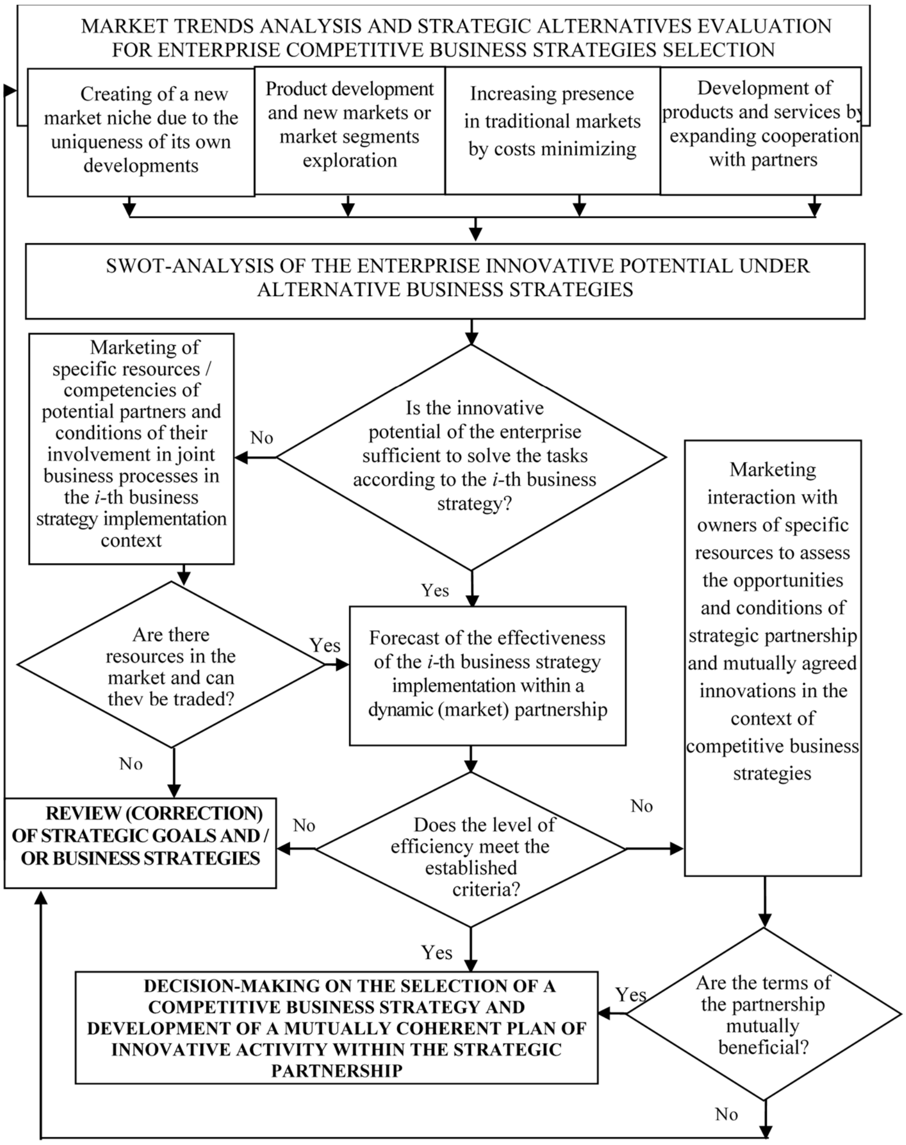


Fig. 2. The model of the enterprise competitive business strategy selection considering the resource potential increasing possibility on the basis of strategic partnership.

Integrated business structures incentives-preferences in the field of innovation need should be formed not only in the organizational space of an individual enterprise, but also in the context of innovative development of all partners – if such a partnership is considered as strategic. To do this, we recommend to developed the motivational potential of engagement marketing, which is based on maintaining the partnerships value.

Communication marketing interaction system will ensure the rapid exchange of information between the participants of the strategic partnership in relation with common goals and strategies, which may also be related to various aspects of innovative potential realizing. It will create favorable conditions for active and effective innovation activity and achievement of the goals set in each strategic period. This also applies to the structural and functional development of the innovation potential of the enterprise – the filling of structural elements with new competences, taking into account the specifics of the selected business strategies.

In the strategic partnership of business structures, it is necessary to cultivate such relationships that form a system of trust. In this case, the partners evaluate positively each other's actions in the process of choosing the directions and scale of organizational change, without looking for threats to their own security. This will eliminate the causes of opportunistic behavior of partners, develop cooperation and mutual support in identifying and solving new problems.

It is extremely important that the integrated business structures participants clearly understand the common benefits, so the value of the strategic partnership. It is advisable to determine the expected value from the development of partnerships by means of the integral efficiency index in cases of interval uncertainty based on the criterion of optimism-pessimism of Leonid Hurwicz – provided that the value of managing these relations is to obtain the potential benefit from increasing the resource potential.

$$E_{ik} = max\{d_{si} \times max\,\mu_{ri}\,(u_i) + (1 - d_{si}) \times min\,\mu_{ri}\,(u_i)\}, \tag{15}$$

where E_{ik} – expected integral efficiency of one enterprise relative to others, which determines the value of relationships in the affiliate network; $\{max\,\mu_{ri}\,(u_i)\}$, $\{min\,\mu_{ri}\,(u_i)\}$ – the maximum and minimum value of the function belonging to the *i*th set of a linguistic variable values; u_i – is a generalized value of the individual dynamics index of each partner in the affiliate network, taking into account the *i*th set of the linguistic variable values; d_{si} – is the average interval value of the E. Harrington scale of dominance estimation, taking into account the *i*th set of the linguistic variable values.

In the organizational space of strategic partnership forms an organizational heterarchy, it expands the possibilities of generating innovative ideas that can give a new impulse to business development. Due to the production of new knowledge, creation of new products and technologies that develop social needs, integrated into the partnership business structures move from a predominantly adaptive scenario of behavior in the market to proactive, maintaining and increasing their competitiveness.

4 Conclusions

Defining strategic priorities for the national economy development is an important element in the formation of state economic policy, the regulatory instruments of which (institutions) should create a motivational environment for overall economic growth. In the set of such tools, an important role is played by those that regulate innovation and investment processes in terms of economic activities and sectoral specialization of economic entities. The paper substantiates that for the national economy sustainable development it is advisable to provide institutional support for investment in those sectors that can provide a high level of added value in the creation of consumer values in strategic perspective. It is also substantiated that it is expedient to use TOPSIS and CRITIC methods to assess investment priorities.

Substantiation of innovation and investment support strategic priorities for the Ukraine's economy industrial sector development was carried out on the example of the food industry, which can significantly increase the total value added, using the agricultural production potential. It covers different sub-sectoral groups in terms of consumer purpose and technological equipment, which determines their different financial and economic performance, and hence different investment prospects.

According to the calculations of synthetic indicators and cluster analysis, it was determined that the most priority investment plans for the period include meat processing, oil fat, fish processing, confectionery, milling branch and cereals, a subsector for the production of food sets and ready-made dishes. The development of these sub-sectors is economically profitable both from the standpoint of the owners of capital and from the public interest point of view, since it will ensure the growth of the value added within the national economy.

It has been noted that in the formation of state economic policy it is advisable to differentiate approaches to the creation of investment preferences for attraction of strategic investors. In particular, such preferences should be defined for the development of the food industry in the context of those sub-sectors that correspond to the dynamics of food markets in Ukraine and increase the opportunities for entry into the European Union market.

Given that innovation processes should also cover small and medium-sized industrial enterprises, whose resource opportunities are mostly limited, it is proposed to expand them within the framework of a strategic partnership. It has been defined the organizational features of such partnership, which is characterized by a blurred hierarchy (heterarchy) and rivalry of organizational principles, which forms the motivational basis for improving the processes of creating consumer values. A graphic model of choosing a competitive business strategy has been built, taking into account the possibility of resource potential building on the basis of strategic partnership. It has been formulated and formalized the author's position on improving the motivational environment of participants in innovation processes in a strategic partnership through monitoring the value of partnerships.

The developed recommendations can be used to expand the directions and methods of the national economy industrial sectors business structures organizational interaction within the plans of strategic partnerships building for the competitive business strategies implementation.

References

1. Adali, E.A., Isik, A.T.: Critic and Maut Methods for the Contract Manufacturer Selection Problem. European Journal of Multidisciplinary Studies **2**(5), 93–101 (2017)
2. Babenko, V.O., Petukhova, V.O., Sidorov, M.V.: Research of strategic prospects of development of integration processes of Ukraine in the global system of international trade. Technological audit and production reserves 3/5(29), 41–48 (2016)
3. Boiko, O., Vasiutkina, N., Bavyko, O., Yermak, S.: The state and ways of Ukrainian food industry development in terms of integration into European economic space. Journal of Hygienic Engineering and Design **27**, 69–86 (2019)
4. Coase, R.H.: The Nature of the Firm. Economica **4**(16), 386–405 (1937). doi:10.1111/j.1468-0335.1937.tb00002.x
5. Dergachova, V.V., Tryhlib, T.V.: Formation of competitive advantages of enterprises in conditions of multi-vector integration orientation. Ekonomichnyj visnyk National Technical University of Ukraine ‘Kyiv Polytechnic Institute’ 10, 81–86 (2013)
6. Hannan, M.T.: Uncertainty, Diversity, and Organizational Change. Economic sociology **2**(2), 73–94 (2001)
7. Khvostina, I., Havadzyn, N., Horal, L., Yurchenko, N.: Emergent properties manifestation in the risk assessment of oil and gas companies. CEUR Workshop Proceedings **2422**, 157–168 (2019)
8. Kolmykova, T., Lukianykhina, O., Baistriuchenko, N., Lukianykhin, V.: International integration in innovative development of economy. Problems and Perspectives in Management **13**(1), 203–207 (2015)
9. Kuznetsova, M. (ed.): State Statistics Service of Ukraine. Activities of large, medium, small and micro enterprises. Statistical collection of 2018. Kiev. http://www.ukrstat.gov.ua (2019). Accessed 26 Jul 2020
10. Lakshmi, M.T., Venkatesan, P.V.: A Comparison of Various Normalization in Techniques for Order Performance by Similarity to Ideal Solution (TOPSIS). International Journal of Computing Algorithm 3, 255–259 (2014)
11. Lin, J.Y.: New Structural Economics: A Framework for Rethinking Development and Policy. World Bank Publications, Washington (2012)
12. Lukianenko, D., Poruchnik, A., Stoliarchuk, Y., Liutak, O.: Globalization of the tourism industry: Scales, levels and institutional formats. Problems and Perspectives in Management **17**(2), 563–574 (2019)
13. Moskalenko, O.: The advanced development and conceptual bases of the modern state’s economic policy. Economic Annals-XXI 1-2(2), 4–7 (2014)
14. New Global Partnership: Eradicate Poverty and Transform Economies Through Sustainable Development. Report of the High_Level Panel of Eminent Persons on the Post-2015 Development Agenda. United Nations Publications, New York (2013)
15. Olson, M.: Autocracy, Democracy and Prosperity. In: Zeckhauser, R.J. (ed.) Strategy and Choice, pp. 131–157. MIT Press, Cambridge (1991)
16. Porter, M.E.: On Competition. Harvard Business School, Boston (1998)
17. Rahim, R., Utama Siahaan, A.P., Wijaya, R.F.: Technique for Order of Preference by Similarity to Ideal Solution (TOPSIS) Method for Decision Support System in Top Management. International Journal of Engineering & Technology 3–4, 290–293 (2018)
18. Reinert, E.S.: How Rich Countries Got Rich and Why Poor Countries Stay Poor. Constable & Robinson, London (2007)
19. Sardak, S., Radziyevska, S., Us, I.: Ukraine’s exports as a global challenge for its future. CEUR Workshop Proceedings **2422**, 84–99 (2019)

20. Semerikov, S., Chukharev, S., Sakhno, S., Striuk, A., Osadchyi, V., Solovieva, V., Vakaliuk, T., Nechypurenko, P., Bondarenko, O., Danylchuk, H.: Our sustainable coronavirus future. E3S Web of Conferences **166**, 00001 (2020). doi:10.1051/e3sconf/202016600001
21. Sharko, M., Gusarina, N., Petrushenko, N.: Information-Entropy Model of Making Management Decisions in the Economic Development of the Enterprises. Advances in Intelligent Systems and Computing **1020**, 304–314 (2020)
22. Stadnyk, V., Izhevskiy, P., Khruhsch, N., Tomalja, T.: Investment Priorities of the National Economy Sectors Development. Advances in Economics, Business and Management Research **95**, 155–160 (2019). doi:10.2991/smtesm-19.2019.31
23. Stadnyk, V., Izhevskiy, P., Zamazii, O., Goncharuk, A., Melnichuk, O.: Factors of enterprises' strategic selection of participation forms in integration formations. Problems and Perspectives in Management **16**(2), 90–101 (2018). doi:10.21511/ppm.16(2).2018.09
24. Stark, D.: The ambiguity of assets and the organization of diversity in post-socialist countries. Economic Sociology **2**(2), 115–129 (2001)
25. State Classifier of Products and Services SC 016-2010. https://dkpp.rv.ua (2010). Accessed 26 Jun 2019
26. State Statistics Service of Ukraine. Activity of enterprises. http://www.ukrstat.gov.ua (2019). Accessed 26 Jun 2019
27. State Statistics Service of Ukraine. Capital investments by the type of industrial activity. http://www.ukrstat.gov.ua (2020). Accessed 17 Aug 2020
28. State Statistics Service of Ukraine. Profitability of operating and all activities of enterprises by type of economic activity with the division into large, medium, small and micro enterprises (2010-2018). http://www.ukrstat.gov.ua (2020). Accessed 17 Aug 2020
29. Lin, J.Y.: The Industrial Policy Revolution I: The Role of Government Beyond Ideology. Palgrave Macmillan, New York (2013)
30. Wallerstein, I.: The Modern World-System I: Capitalist Agriculture and the Origins of the European World-Economy in the Sixteenth Century. University of California Press, Berkeley (2011)
31. Wernerfelt, B.: The Resource-based View of the Firm: Ten Years Later. Strategic Management Journal **16**(3), 171–174 (1995)
32. Williamson, O.E.: The economic institutions of capitalism: firms, markets, relational contracting. Free Press, New York (1985)

Modelling of trade relations between EU countries by the method of minimum spanning trees using different measures of similarity

Hanna Danylchuk[1[0000-0002-9909-2165]], Oksana Ivanylova[2[0000-0002-3711-6117]], Liubov Kibalnyk[1[0000-0001-7659-5627]], Oksana Kovtun[2[0000-0002-0159-730X]], Tetiana Melnyk[2[0000-0002-1513-7304]], Oleksandr Serdiuk[1[0000-0002-1230-0305]] and Vladimir Zaselskiy[3[0000-0002-7517-5433]]

[1] The Bohdan Khmelnytsky National University of Cherkasy,
81 Shevchenka Blvd., Cherkasy, 18031, Ukraine
abdanilchuk@gmail.com, liubovkibalnyk@gmail.com, serdyuk@ukr.net
[2] University of Educational Management, 52A Sichovyh Striltsiv Str., Kyiv, 04053, Ukraine
oksi_iv@i.ua, kovtun.oa71@gmail.com, melnik_tanja@ukr.net
[3] State University of Economics and Technology,
5 Stepana Tilhy Str., Kryvyi Rih, 50006, Ukraine
zaselskiy52@gmail.com

Abstract. The article is devoted to the study of changes in relations between the countries of the European Union based on modeling and analysis of the structure of trade relations between countries. The article analyzes the dynamics of exports and imports of goods and services between the countries of the European Union on the basis of data taken for the period from 2006 to 2019. The study is based on one of the methods of cluster analysis, namely – the method of constructing minimal spanning trees. For the analysis the method of visualization of links is defined and the choice of the corresponding graphic representation is substantiated: the display of links using the dendrograms which carry more information in comparison with display of the minimum spanning trees in the form of a planar graph is chosen. Four different methods were used to construct the minimum spanning trees on the basis of which the visualization of links is performed: the Single link method, the averaged link method, the complete links method, and the Ward method. Based on the analysis of the results obtained using each of the methods, the best of them is selected, which is then used throughout the study. As a result of the study, suggests were made about the criteria by which clusters are formed within the European market. Such criteria are both the geographical neighborhood, which means mostly similar climatic conditions, and the common strategy of economic development of the country and the common strategy of behavior in the world market. In addition, a number of countries have been identified that are gradually moving to the use of their own economic strategies, as well as a number of countries seeking to align strategies of behavior in the world market. The influence of such factors as joining the integration union of new member states and global financial crises on the structure of trade relations is substantiated. Changes in the structure of relations between EU countries due to the influence of these factors are simulated. The study is of an applied nature and can be used in the future as a methodological basis for developing effective mechanisms for reformatting trade relations between countries in the context of geoeconomic transformations and global financial crises.

Keywords: cluster analysis, minimum spanning tree, cluster, dendrogram, European Union, international trade, import, export.

1 Introduction

International trade is one of the oldest and most important forms of implementation of the international division of labour and world economic relations. At the same time, it is the traditional and most developed form of international economic relations.

International trade has grown markedly over the past two centuries [13]. Today, about a quarter of the world's production is exported. This has had a significant impact on the geo-economic structure of the world economy, with the emergence of new centres of economic power and the formation of regional integration associations and coalitions.

The integration of national economies into the world economic system was one of the most important events of the last century. The development and deepening of international trade links have played a role in the formation of integration groupings. At the same time, the processes of integration, in the present circumstances, are conducive to a significant increase in the volume of exports and imports of goods and services between countries. This is particularly true for regional association of countries. Among these, the European Union plays a special role. It is within this integration framework that trade relations have reached the highest stage of their development. At the same time, contemporary threats and challenges have a significant impact on the stability of such links. New members, not always with a high level of economic development, are periodically joining. Or, like the United Kingdom, leaving it [8]. Global financial crises and other external and internal factors also destabilize the situation.

In this context, studies aimed at identifying structural changes in the reformatting of trade relations between countries and the reasons for such changes are of particular importance. One of the modern methods of economic and mathematical modelling that works effectively in such research subjects is cluster analysis and its specific tools: graph theory and minimum spanning tree construction techniques. This article deals with the modelling of trade links between EU countries using spanning tree methods.

2 Literature review

The analysis of the scientific work shows that the researchers possess many developed methods of the modelling trade processes between countries. They target individual countries, integration associations, certain groups of goods and services, etc.

In [9] regression models were used to link trade flows and gross products of all

BRICs countries. The authors have developed a theoretical model that allows us to obtain endogenous macroeconomic models taking into account bilateral trade between all BRIC countries and foreign trade with the rest of the world. The Dixit-Stiglitz model was used to study the effects of international trade [20].

In a study of the dynamics and forecasting of international trade, a gravity model is widely available. For example, [11] tests and compares classical and gravitational trade models with the example of the United Kingdom, with the authors noting the best applicability of the classical model to the United Kingdom.

The study of the impact on international seafood trade of factors such as geographical distance between countries, trade agreements, gross product and other macroeconomic variables using the gravity model is presented in [12].

The extended Frankel gravity model was used by authors [3] in a study of India's bilateral external trade relations with ASEAN countries.

The authors [2] based on the gravity model, consider trade flows between China and its major global partners.

The international trade factor model in the euro area, based on the identification of the interaction between macroeconomic and trade substitutes, is proposed in [5] and proved to be effective for short-term forecasts.

There have recently been publications using network theory to the modelling international trade processes. Thus, the authors [15] consider international trade as a multi-level network.

In such studies, various econophysical methods are used to analyze complex economic systems. For example, in recent years, an analysis of the cryptocurrency market has been conducted using similar methods [18; 19], some of which also underlie the mathematical methods used in this work.

In summary, the above analysis of a sample of contemporary publications proves the relevance of the topic and the interest of scientists in the further development and application of new approaches to the modelling trade links between countries within integration association or globally. The aim of this research is to present the results of the modelling trade links between the countries of the European Union using the minimum spanning tree method and to analyses the changes in the structure of these links as a result of the impact of different social and economic, political factors.

3 Method and materials

It is proposed to assess trade links between countries and their evolution over time using the approach described by Rosario N. Mantegna [10]. This approach is based on constructing a minimum spanning tree through the graph obtained from the cross-correlation matrix and its analysis. Essential in this method is the study of several time series simultaneously.

In research [10], the matching ratio is used to quantify the degree of similarity of the system elements

$$c_{ij} = \frac{\langle Y_i Y_j\rangle - \langle Y_i\rangle\langle Y_j\rangle}{\sqrt{\left(\langle Y_i^2\rangle - \langle Y_i\rangle^2\right)\left(\langle Y_j^2\rangle - \langle Y_j\rangle^2\right)}} \tag{1}$$

where i, j – series index, $Y_i = ln\, P_i\,(t) - ln\, P_i\,(t-1)$ and $P_i(t)$ value of the *i*-series at time *t*. With correlation coefficients c_{ij} a matrix of size $n \times n$ is made. It is known that the correlation coefficient can take values from -1 (totally incorrelated pair) to 1 (fully correlated pair). The matrix of correlation coefficients is a symmetric matrix with units on the main diagonal.

To understand and interpret the topological structure of the investigated system, a generalized metric is used which defined by the formula

$$d(i, j) = \sqrt{2\left(1 - c_{ij}\right)} \tag{2}$$

With this definition $d(i, j)$ numerically satisfies such axioms:

i. $d(i, j) = 0$, if and only if $i = j$, the axiom is fully correlated, it can be obtained by using the same series (series numbers match) and it is almost impossible to obtain for the real market using different series;
ii. $d(i, j) = d(j, i)$ – the second axiom holds because we have a matrix of cross-correlation coefficients and, accordingly, the distance matrix *D* are symmetric by definition;
iii. $d(i,j) \leq d(i,k) + d(k,j)$ – third axiom execution can be verified numerically.

Given the nature of the values of the matrix *C*, the value in the matrix *D* belongs to [0, 2], and, $d(i, j) = 0$ means total time series similarity (in some correlation space the points corresponding to these series coincide), a $d(i, j) = 2$ means the complete opposite of the time series – in the respective space the points are at the maximum distance from each other.

The matrix *D* is used to construct the minimum spanning tree [4; 14]. For an unambiguous definition of taxonomy, a subdominant ultra-metric space is considered to be a finite topological space. In ultramodern space, elements are not placed along a single line, but in a hierarchical tree. The minimum spanning tree reflects the arrangement of the system elements and their optimal connections. The graphically minimal spanning tree is represented as a connected graph consisting of *n* vertices (nodes) and *n*-1 ribs. The minimum spanning tree has the shortest length among all trees based on the sum of the distances between the two elements. In addition, [10] shows that the minimum spanning tree based on economic indicator series reflects hidden information contained in economic time series.

The matrix *D* is also used to construct a hierarchical tree – dendrogram. Dendrogram better reflect the presence of groups of related objects – clusters – in the study system. Since the two trees are based on the same information, they are complementary and visualize each of their specific information, and together they allow for a more detailed analysis of this system.

Both ways of visualizing the connections between points make it possible to examine the presence and mutual location of objects of the system under investigation, which are grouped into clusters. In each cluster, objects are more similar to each other than to the objects of any other cluster, and therefore the distance between objects within the cluster is smaller than those of different objects.

In general, cluster search methods – clustering – form several major families:

1. hierarchical agglomerative methods;
2. hierarchical divisive methods;
3. iterative techniques;
4. methods for determining modal densities;
5. factor methods;
6. condensation methods;
7. graph-based methods.

In the proposed research we use methods belonging to the group of hierarchical agglomerative methods. This group of methods has been chosen for the following reasons. First, the algorithms of the group methods are simple enough to implement due to their iterative nature, since the sample of object pairs is performed sequentially and the entire $n \times n$ matrix is checked for n $(n–1)/2$ pairs. Second, the sequence of clusters is easily visualized by the dendrogram, and the corresponding minimum spanning tree provides additional information. The dendrogram itself for n objects forms in $(n–1)$ steps. Third, the result of a simple construction algorithm is that the interpretation of the results is fairly simple with relatively rich information on the structure of the system and the relationship between its elements.

The easiest way to construct a minimum spanning tree is the following. The points are selected sequentially, starting with the pair of points nearer and gradually changing the pairs in increasing order of the distance between the points in them. When analyzing another pair of points (o_i, o_j), the distance between which d_{ij} can be several:

(1) both points meet for the first time – then a new cluster is formed with the distance between its elements d_{ij};
(2) one of the points already belongs to some cluster – then a second point with distance from other points is assigned to the same cluster d_{ij};
(3) each of the points belongs to different clusters: then the clusters to which the points belong are merged into a large cluster with the distance between the subclusters d_{ij};
(4) both points belong to the same cluster – minimum spanning tree does not change.

From the above distribution of points by clusters, the distance between points is used as a measure of the similarity of the examined objects, with the objects being compared in pairs. This method is the easiest to understand and is called the single-link method [16]. The main disadvantage of the method is the appearance of so-called “chains” – sequences of “elongated” up clusters on dendrogram. This will be seen in the next section.

In the event that, in the search for a cluster to be included, all distances from a given point to the points of the cluster are analyzed and a cluster with the largest number of points within a distance of not more than some threshold distance to a given point is selected, the method of complete links is obtained [17]. Because of the more stringent rule of searching for such objects, the method creates more compact clusters. Although the method has no deficiency in the single-link method, it can also produce an inaccurate structure of the system under consideration, despite the compactness of the found clusters.

In order to eliminate the disadvantages of single-link and complete links methods, the average link method was proposed by the authors [17]. In this method, when searching for the cluster to which the point will be attached, the average distance from the point to the points of each of the clusters is estimated, and the lowest average distance cluster is chosen as the final cluster, but not more than some given threshold distance. A modification of the method can also be compared not with the average distance to the points of the cluster, but with the distance from the given point to the center of mass of the cluster. Note that it is the average link method that is widely used in social sciences.

The Ward's method was also developed to optimize the minimum dispersion within clusters. During the operation of the method, the optimum value of the target function of the sum of the squares of deviations is sought. According to Ward's method, each step into a large cluster consists of those groups of objects for which the sum of the squares of deviations receives the least increment. The method shows a tendency to search for clusters of roughly the same size and hyperspherical shape. The Ward's method is also widely used in the social sciences [1].

We have conducted a study and comparative analysis of economic performance using all four methods to determine the type of economic information that can be obtained by using different similarities in the search for clusters.

The study uses EU Member States' exports and imports of goods and services. The data cover the period from 2006 to 2019. In order to monitor the dynamics of structural changes in the links between countries, the general period was divided into intervals: first – 2006–2010, second – 2011–2015, third – 2015–2019. The analysis used data for all EU Member States at the time of the calculations. All data received from the official statistics site [7].

4 Results and discussion

As part of this approach the tools were implemented and calculations were made in the Matlab.

As mentioned above, the visualization of the results can be performed both using the minimal spanning trees and using the dendrograms. For example, in figure 1 shows an image of a possible minimal spanning tree [6].

The figure shows information about the mutual location of the economic agents under study, the presence of separate clusters, as well as some relationships between clusters. However, due to the fact that the minimal spanning trees mostly do not reflect the closeness of links (or the strength of links), it is very difficult and often impossible to talk about a certain number of clusters, the location of economic agents in clusters, and so on. The use of additional information about the strength of connections in clusters greatly complicates the procedure of visualization of the minimum scanning tree, and often makes it impossible.

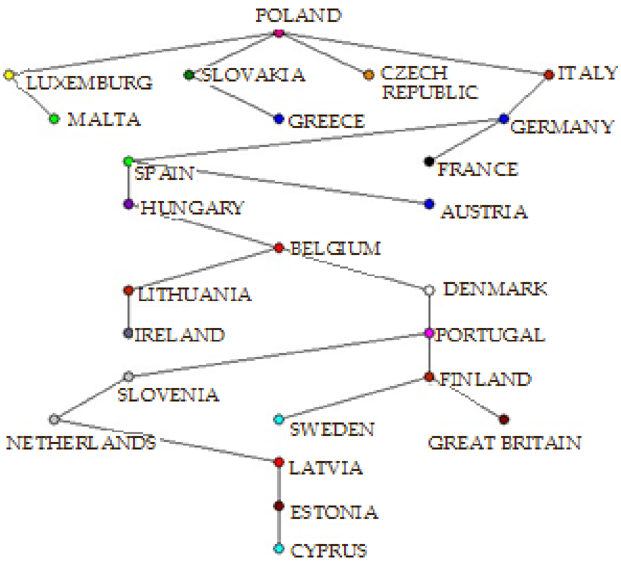


Fig. 1. The minimal spanning tree of the dynamics of exports of EU goods for 2000-2006.

Instead, the dendrogram shows both the closeness of the links between economic agents and, when using the criterion of minimal communication within the cluster, the presence and configuration of clusters. That is why to study the structural shifts and the possibility of their reflection by each of the methods used, we compare the dendrograms obtained for each of the periods by each method. On the dendrograms represented in the article the blue vertical lines from the values of the countries included in the EU before 2000, and the red lines from the countries included in the EU after 2000.

Figures 2-4 shows the dendrograms obtained for the links between the EU Member States in export of goods using the Single link method calculated for the respective periods referred to earlier.

First of all, there is a large cluster with close links, including Austria, Germany, Portugal, Spain, Italy, Belgium, the Netherlands, France, Denmark, Sweden. Note the gradual weakening of Denmark and Sweden with the cluster countries as the time period close to 2019.

Among the EU member countries that have joined since 2000 Slovenia, the Czech Republic, Hungary, Poland, Slovakia are permanently in the cluster and the link with Romania has been strengthened since 2011. In this cluster, the distance between the countries does not exceed 0.5, which corresponds to a correlation coefficient ≈ 0.7.

Over time, Malta and Cyprus have gradually become less connected to other EU countries. Note also that other countries are at a distance of 0.5-0.9 to the core cluster, which corresponds to correlation coefficients from ≈0.7 to ≈0.9. Also significant is the gradual reduction of the distance between these countries and the core cluster in proximity 2019, which can be attributed to the increased coherence in the trade strategy for the EU countries.

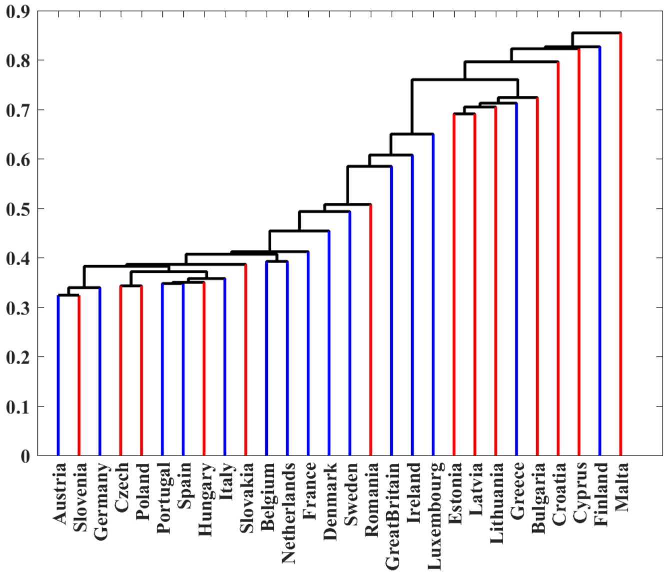


Fig. 2. Dendrogram of Links for Export movements in the European Union for the period 2005-2010, Single link method.

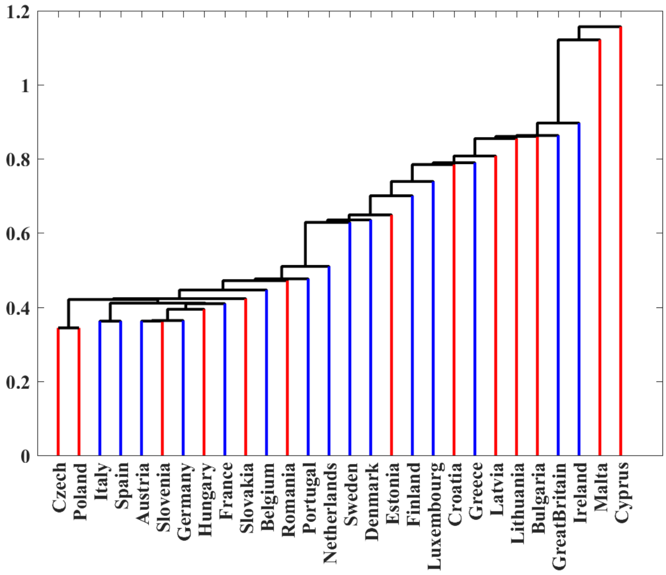


Fig. 3. Dendrogram of Links for EU Export movements in the European Union for the period 2011-2015, Single link method.

The following results were obtained when analyzing the relationship between EU countries in terms of exports of goods using the method of averaging.

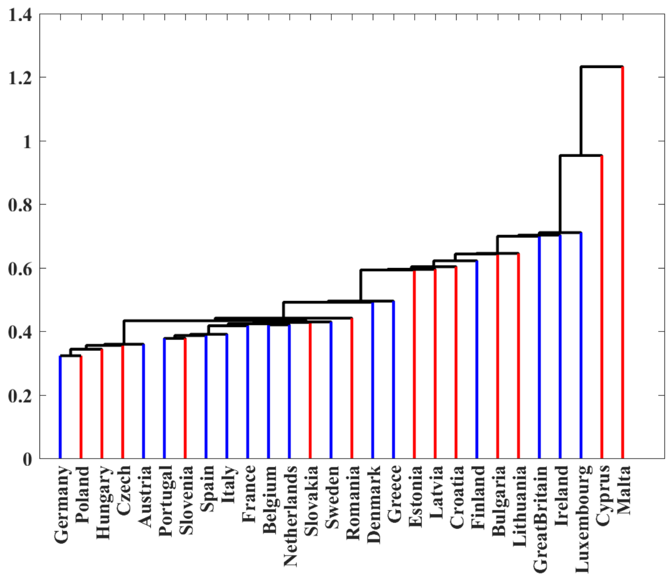


Fig. 4. Dendrogram of Links for Export movements in the European Union for the period 2016-2019, Single link method.

In the first study period there is one powerful and three small clusters with intracluster distances less than 0.5. The powerful cluster includes, inter alia, Belgium, the Netherlands, Italy, Germany, France, and Hungary and Slovakia. Up to three small clusters are: 1) Austria and Slovenia; 2) Czech Republic and Poland; 3) Portugal and Spain. As can be seen, in this case the countries in the small clusters are geographical neighbours and can explain the similar trade strategy of the countries in one cluster. However, inter-cluster links are weak, which means own trade strategies within the cluster.

In the transition to dendrograms received in later periods, this trend towards the location of geographically adjacent countries remains largely unchanged, but also a noticeable tendency to reduce cluster power with strong internal link and weak interlaced links, which even corresponds to anticorrelated behavior of the countries.

Note the existence of a large group of countries identified by this method as countries with a relatively dedicated trade strategy. Another important feature is the continuation of the trend of Cyprus and Malta until full separation of trade strategies from other EU Member States.

Figure 5-7 shows the dendrograms obtained for the relationship between the EU Member States in export using the Complete link method calculated for the respective periods. In the obtained dendrograms, there are 4 to 5 small clusters with distances up to 0.5. In the first time period (2005-2010) there are 4 clusters with this characteristic:

1) Austria and Slovenia; 2) Czech Republic and Poland; 3) Portugal, Spain, Italy, Germany; 4) Belgium, Netherlands, Hungary, France, Slovakia. In this case the principle of geographical link within the clusters and the very weak link between the first named cluster and other EU countries are observed. At the same time, all other countries outside these clusters have weak links, ranging from 0.5 to 1.0.

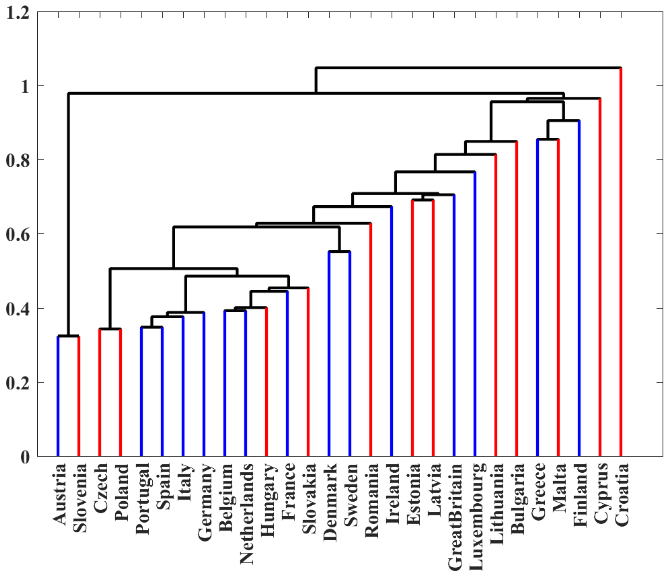


Fig. 5. Dendrogram of Links for Export movements in the European Union for the period 2005-2010, Complete link method.

However, several countries – Croatia, Cyprus, Finland, Greece, Malta and Finland - are virtually on the sidelines because, while Greece, Finland and Malta could be grouped under one cluster, the links between them are so weak that the boundaries of the cluster are extremely blurred.

In the next time period, there is some reformatting of the main clusters (figure 6): 1) Czech Republic, Poland; 2) Italy, Spain; 3) Austria, Slovenia, Germany, Hungary, Slovakia, Belgium, France; 4) Portugal, Romania. The trend towards a joint trade strategy continues, owing to the geographical proximity of countries, as well as the presence of a large group of countries with relatively different trade strategies. Malta and Cyprus are also gradually moving away from the trade strategies of any other country.

In the last time period (figure 7) five clusters with strong links within can be identified again. Growth of cluster capacity is important: 1) Germany, Poland, Hungary; 2) Portugal, Slovenia; 3) Italy, Spain, Austria, France; 4) Belgium, Netherlands; 5) Czech Republic, Slovakia, Sweden, Romania, Denmark. Thus, the number of countries in the observed clusters increased to 16 in the last period, which can be explained by the tendency of Governments to adopt joint trade policies. In the time period under study, countries are grouped into clusters along geographical lines. The trade strategies of Malta and Cyprus can also be further isolated.

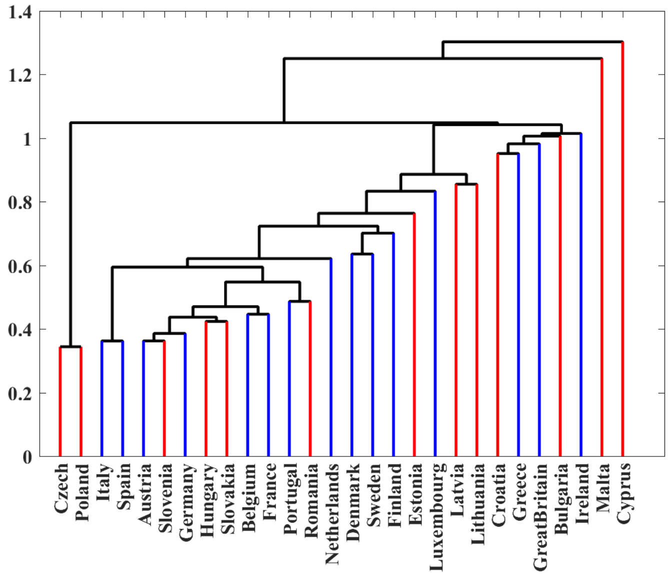


Fig. 6. Dendrogram of Links for Export movements in the European Union for the period 2011-2015, Complete link method.

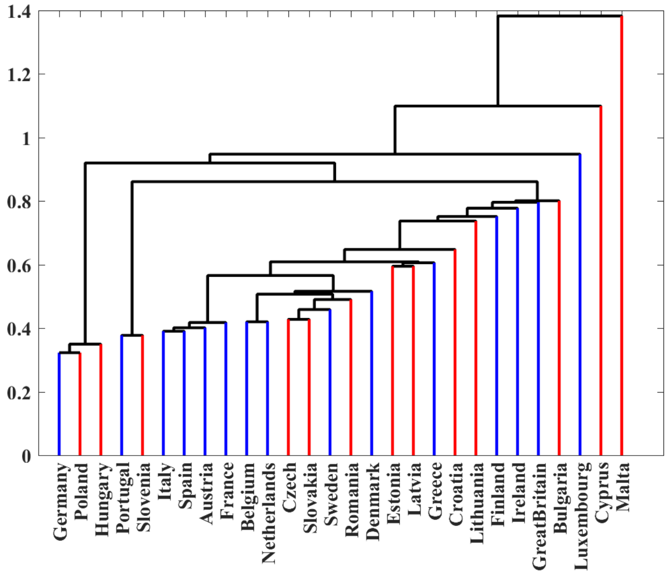


Fig. 7. Dendrogram of Links for Export movements in the European Union for the period 2016-2019, Complete link method.

Figures 8-10 shows the dendrograms obtained for the inter-EU relations for merchandise exports using the Ward's method, calculated for the respective periods.

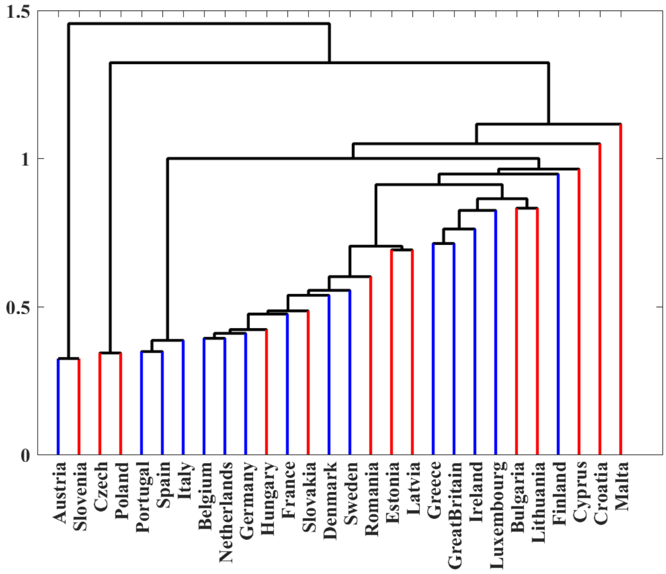


Fig. 8. Dendrogram of Links for Export movements in the European Union for the period 2005-2010, Ward's method.

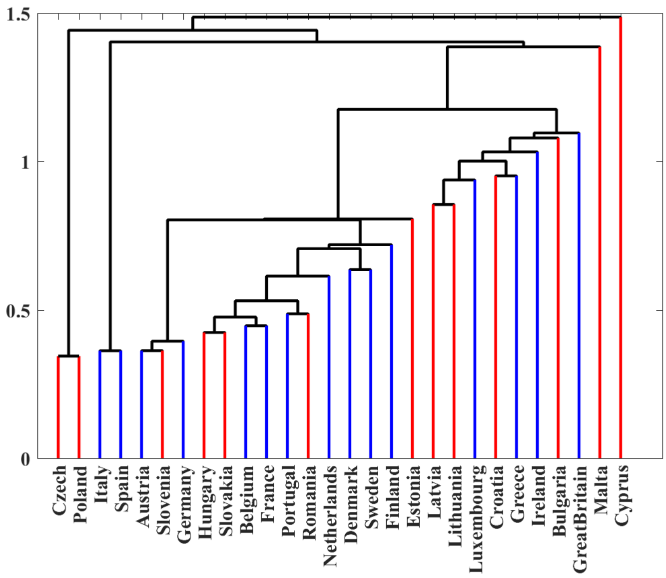


Fig. 9. Dendrogram of Links for Export movements in the European Union for the period 2011-2015, Ward's method.

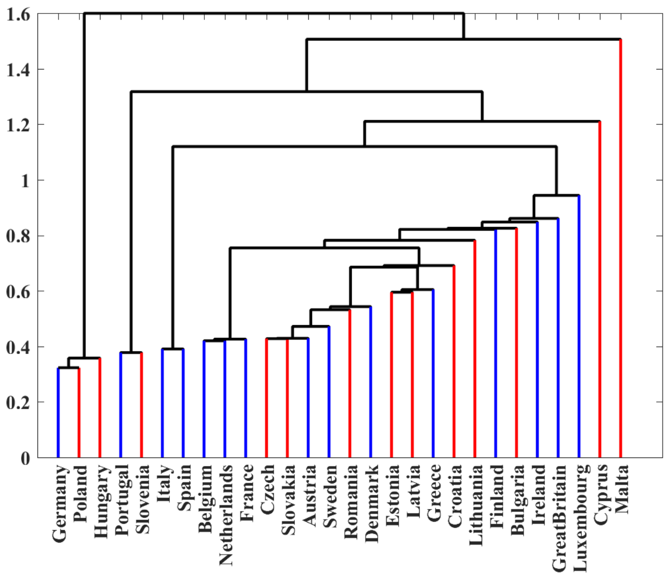


Fig. 10. Dendrogram of Links for Export movements in the European Union for the period 2016-2019, Ward's method.

In the first time period (2006-2010) there are four clusters, three of which are small and the fourth is quite powerful (figure 8). Small clusters are: 1) Austria and Slovenia; 2) Czech Republic and Poland; 3) Portugal, Spain and Italia. The following EU countries have joined cluster 4: Belgium, Netherlands, Germany, Hungary, France, Slovakia. There is also a tendency to join the cluster, but with slightly weaker links, by the following countries: Denmark, Sweden, Romania, Estonia, Latvia. For the results obtained, the principle of geographical proximity continues to be a key one, although it is weaker than the one of complete link. On the other hand, the trend towards Malta and Cyprus' own trade policies continues.

There is some slight reformatting of the clusters when moving to the next period, but the vast majority of countries continue to be linked to the same countries (figure 9). For this time period, there is a slight weakening of the most links within the clusters and a corresponding weakening of the inter-cluster bonds. Malta and Cyprus continue to pursue their own trade strategy. For Portugal and Germany clusters are changing, but it is not possible to tell by the methods used whether these countries have moved to other clusters or have joined countries from other clusters in these clusters.

There is also a significant trend in the results of the weakening of the cluster bonds, which is reflected in the shift of the median value from about 1 side of the increase in distance to 1.5.

In the last time period studied, the number of clusters with strong intra-cluster linkages is 5. The clusters comprise the following countries: 1) Germany, Poland, Hungary, 2) Portugal, Slovenia; 3) Italy, Spain; 4) Belgium, Netherlands, France; 5) Czech Republic, Slovakia, Austria, Sweden. Here, it is clear that the principle of building trade strategies on the basis of the geographical location of countries has been maintained and somewhat strengthened. The trend towards Malta and Cyprus' own strategy, as well as the general trend towards loosening of inter-cluster links, continues. Thus, the following conclusions were drawn from the analysis of the results obtained by the selected methods from the export data of the EU Member States.

The Complete link method and the Ward's method have produced rich and more consistent information than the other two methods. The most detailed information was given by the Complete link method, due to the combination of methods, as already mentioned in the previous section, the advantages of the Single link method and the average link method. Therefore, the Complete link method was chosen for further analysis.

On the clustering of countries and the restructuring of clusters. There are a steady number of clusters during the study periods, ranging from 3 to 5, with close links within countries. There is almost always one more powerful cluster and several smaller ones – 2-3 countries each of clusters. Within clusters, countries are mostly geographically grouped and links between clusters are significantly weaker. This is logically due to the similarity of natural resources and the climate of geographically adjacent countries, which results in a large number of such products, particularly agricultural products. Varying degrees of industrial development within countries adjust their cluster distribution to produce unexpected inter-cluster links. Looking at the EU countries with more advanced economies, one can observe a tendency for smaller economies to join, which is also understandable. Analysis of structural changes towards 2019 shows a trend towards some increase in cluster capacity, may indicate the desire of EU member countries for mutual assistance and common economic policies.

Based on the interim conclusions on the applicability of the chosen methods, analyse the imports of goods by EU Member States over the selected time intervals, using the Complete link method.

In the next step, we analyzed the import of goods from EU member states during the study period using the Complete link method.

From the dendrogram of the links of imports of EU Member States for the period 2005-2010 the presence of four small clusters is clearly visible: 1) France, Spain; 2) Czech Republic, Slovakia, Italy; 3) Austria, Slovenia, Hungary, Poland; 4) Belgium, Netherlands, Portugal. Most of the represented countries are those that joined the EU in the last century, while most of the countries in these clusters have developed economies. The clusters themselves are small, ranging from two to four countries. Linkages between clusters are weaker, but do not exceed 1 (which corresponds to a correlation coefficient of 0.5). As in the case of exports of goods, there is a significant separation of the import strategy for Cyprus and Malta, to which Luxembourg has also acceded, from other EU members. In addition, you can see three more "blurred" clusters, which include most EU members that have remained.

Next, we analyzed the dendrogram of the links of imports of EU Member States goods received over the period 2011–2015. In this case, three clear clusters of low power and one larger, less clearly delineated cluster are visible. The first three are: 1) France, Italy; 2) Germany, Poland; 3) Austria, Slovenia, Hungary. The countries included in the large cluster are: Belgium, Netherlands, Czech Republic, Slovakia, Portugal, Romania, Spain Compared to the previous period, Italy, Spain and Germany have moved to other clusters, while the remaining countries have made little change in their location, and most of them have even strengthened inter-cluster links. Remaining Malta and Cyprus, other countries continue to establish several clusters less than in the previous period continue to create multiple clusters of less clear shape.

At last, we analyzed the dendrogram of the EU Member States import links for the period 2016-2019. Using an intracluster distance criterion of 0.5 or less, three clear clusters are obtained: 1) Germany, Hungary; 2) France, Italy, Austria, Poland, Spain; 3) Czech Republic, Netherlands. However, another clear cluster can be distinguished from intracluster distances of 0.5 Croatia, Romania, Portugal, Estonia, Finland, Belgium, Slovenia. Intracluster links continue to be weak, and there is also a clear trend towards an increase in the number of countries with their own import strategies; in the period under study, there are already four.

Thus, an analysis of imports of goods over the full period shows a tendency to enlarge groups of EU member States with similar import strategies. However, geographical groupings are almost non-existent. This can be explained, in particular, by the different needs of countries in the groups of imported goods, primarily by the industrial groups developed within countries. During the full period, there has been a trend towards clustering, that is to say, an increase in the number of countries with similar import strategies, which could mean greater specialization of countries.

In figures 11-13 presents the dendrograms obtained during the analysis of the dynamics of services' exports by EU member states for the studied periods.

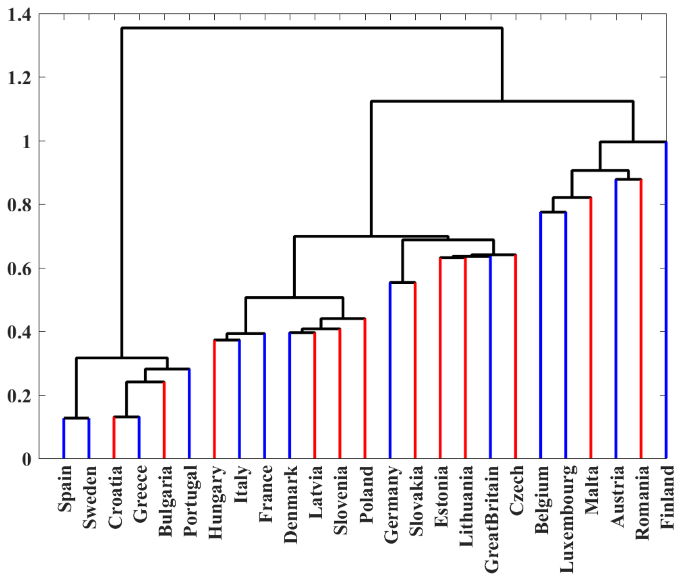


Fig. 11. Dendrogram of Links for service export movements in the European Union for the period 2005-2009, Complete link method.

Figure 11 presents a dendrogram of the EU Member States' relations on the dynamics of services exports for the period 2005-2009. The drawing clearly shows two distinctly strong enough clusters: 1) Spain, Sweden, Croatia, Greece, Bulgaria, Portugal; 2) Hungary, Italy, France, Denmark, Latvia, Slovenia, Poland. Note also those expressed in the clusters: two in each of the clusters. Such a structure could mean a similar strategy for exporting services to a leather cluster with some specialization under clusters. Also important is the very weak link between the identified clusters, which can be explained by the significantly different strategies for service delivery across the different clusters. Compared to the goods market, there are no countries with unique performance in the services export market for the period under study.

The market for services exports is characterized by the consolidation of clusters and the strengthening of intercluster links (figure 12).

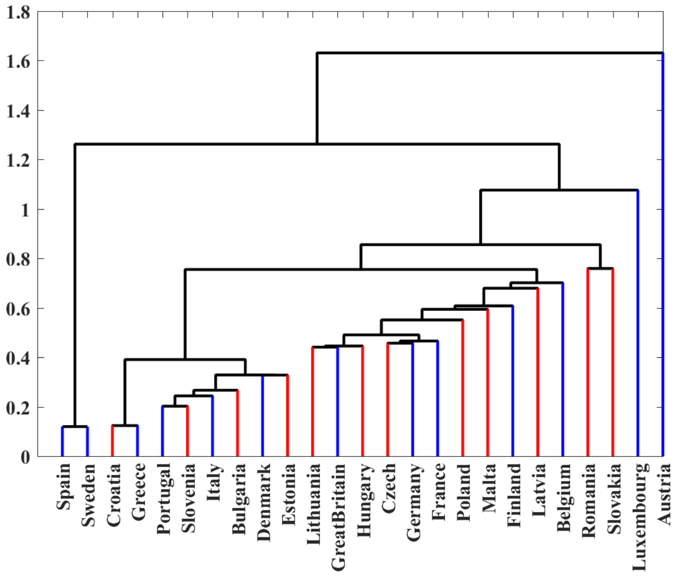


Fig. 12. Dendrogram of Links for service export movements in the European Union for the period 2010-2014, Complete link method.

Four clusters can also be distinguished during this period: 1) Spain, Sweden; 2) Croatia, Greece; 3) Portugal, Slovenia, Italy, Bulgaria, Denmark, Estonia; 4) Lithuania, United Kingdom, Hungary, Czech Republic, Germany, France. In addition to the fact that the number of countries has increased, that has been distributed by clusters in comparison with the previous period – from 13 to 16 – the last of the listed cluster also has several "satellites" that have not been assigned to it, as they have weak links with other countries of the cluster, however, can also be interpreted as belonging to the cluster. There is no territorial unification of countries within the clusters. In addition, Luxembourg and Austria are two countries with unique performance in the services export market.

Figure 13 shows the dendrogram of the links for service exports over the period 2014-2019. The most important distribution of almost all EU Member States by clusters: we have 5 clusters of power 2, 5, 2, 7, 6. Only Finland, Great Britain, Austria are excluded from the clusters. The connections within the clusters do not exceed 0.5. That is, at the end of the study period, we have an increase in the integration of countries into the market of services exports. This may likely indicate a shift in specialization in the EU countries. This view is also supported by the significantly weaker inter-clusters links.

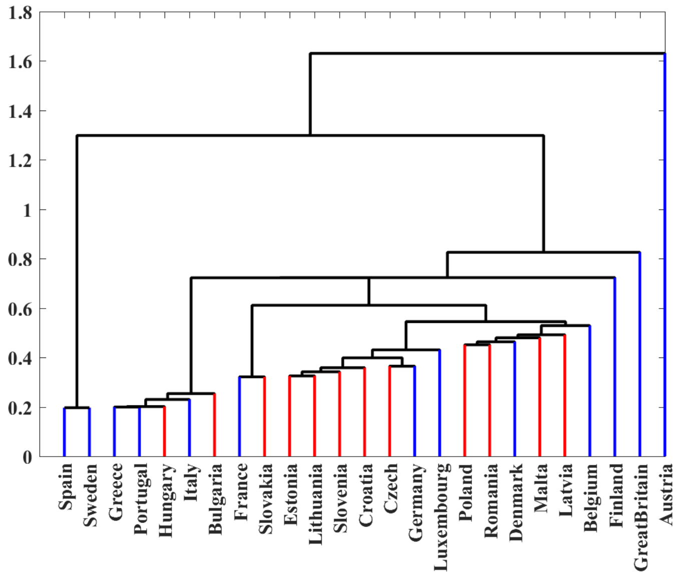


Fig. 13. Dendrogram of Links for service export movements in the European Union for the period 2014-2019, Complete link method.

In summary, the analysis of services exports shows a significant restructuring of the European services market, a tendency of countries to specialize in the market and form large groups of countries with similar dynamics of services exports.

Finally, the latter analysed the behaviour of EU Member States in the services import market.

At first step the dendrogram of the EU Member States' links for imports of services for the period 2005-2009 were analysed. There are four clusters: 1) Germany, Slovenia; 2) Estonia, Latvia, Austria, United Kingdom; 3) Czech Republic, Sweden; 4) Hungary, Slovakia, Spain, Italy, Poland. Dendrogram shows strong links between countries within clusters and much weaker linkages between clusters. Malta and Finland have demonstrated their own behaviour in the services import market. Some of the other countries can be merged into another, more diffuse, cluster, but the inter-cluster links will be significantly weaker, so we did not consider these countries as a separate cluster.

From dendrogram with the links between the countries for the period 2010–2014 were obtained next results. Two clear clusters can be distinguished: 1) Italy, Slovenia;

2) Austria, Germany, Latvia, Poland. Other clusters that could be created have weak inter-cluster links and are therefore not considered. Note the group of countries – Greece, Luxembourg, Lithuania, Malta – with their own, unlike other countries, dynamics of services imports.

The last dendrogram (see the figure 14) with the links for service import movements in the European Union for the period 2015-2019 makes it possible to distinguish three clusters. The first consists of Germany and Spain. The second is from the countries which in turn constitute three subclusters: 1) Italy, Poland; 2) Austria, Slovenia, Latvia; 3) Bulgaria, Portugal. The dendrogram in figure 21 makes it possible to distinguish three clusters. The first consists of Germany and Spain. The second is from the countries which in turn constitute three subclusters: 1) Italy, Poland; 2) Austria, Slovenia, Latvia; 3) Bulgaria, Portugal.

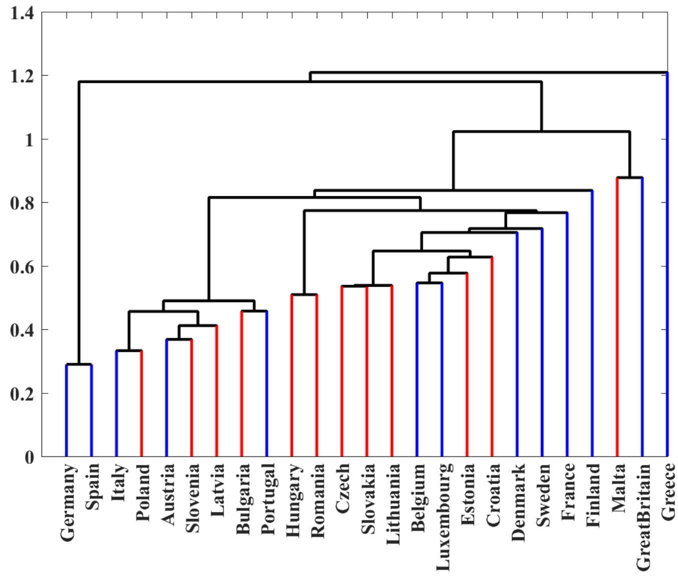


Fig. 14. Dendrogram of Links for the dynamics of imports of services of EU countries for the period 2014-2019, Complete link method.

In summary, the market for services import is characterized by a largely weak correlated behaviour, which is due to the significantly different needs of countries for services. By the end of the period under review, linkages between countries are somewhat weakened and there has been a steady shift from one cluster to another in the pattern of service imports. The behaviour of countries in this market is difficult to characterize and this market requires, first and foremost, further research.

Thus, applying the minimum spanning trees analysis, in particular dendrograms, at successive intervals in the historical time series of exports and imports of goods and services of the EU Member States, it is possible to track changes in the structure of relations between countries that have occurred during this period. In addition, graphical visualization allows quick analysis of the formed clusters, to identify which countries are part of a cluster, i.e., to determine their homogeneity.

5 Conclusions

As a result of the analysis, a number of countries were identified, in particular, Malta, Cyprus, whose links with other EU countries are gradually weakening, which can be explained by the use of its own economic development strategy. Instead, for a large group of countries, there is a strengthening of links and consolidation of clusters over time, which means focusing on a common strategy of economic development. Among the latter countries, there are, first of all, economically powerful EU countries, such as Germany, France, Spain, etc.

In addition, the analysis of relations between countries showed the formation in the existing clusters the smaller clusters of countries, which are located geographically close one by another. This can be explained by additional similar climatic conditions, natural resources and so on.

Thus, the construction of minimum spanning trees and their separate form – dendrograms – allowed to investigate structural and dynamic changes in the trade relations of the EU Member States. On the basis of the analysis made and by comparing the results with the political and economic situation in the market under study. It can be said that the accession of new countries to the EU and the global crises are transforming the trees' structure and that new links between countries are being reformatted and created. The links' structure is also influenced by the economic situation of countries, the natural resources of countries and the degree of development of countries.

References

1. Blashfield, R.K., Morey, L.C.: A comparison of four clustering methods using MMPI Monte Carlo data. Applied Psychological Measurement **4**(1), 57–64 (1980)
2. Caporale, G.M., Sova, A., Sova, R.: Trade flows and trade specialisation: The case of China. China Economic Review **34**, 261–273 (2015). doi:10.1016/j.chieco.2015.03.010
3. Chakravarty, S.L., Chakrabarty, R.: A Gravity Model Approach to Indo-ASEAN Trade-fluctuations and Swings. Procedia – Social and Behavioral Sciences **133**, 383–391 (2014). doi:10.1016/j.sbspro.2014.04.205
4. Cormen, T.H., Leiserson, C.E., Rivest, R.L., Stein, C.: Introduction to Algorithms, 3rd edn. MIT Press, Cambridge (2009)
5. D'Agostino, A., Modugno, M., Osbat, C.: A Global Trade Model for the Euro Area. Finance and Economics Discussion Series 2015-013. Board of Governors of the Federal Reserve System, Washington (2015). doi:10.17016/FEDS.2015.013
6. Danylchuk, H., Kibalnyk, L., Burlaienko, T., Dubinina O.: Modelling of trade relations between EU countries by the method of minimum spanning trees. Advances in Economics, Business and Management Research **95**, 217–221 (2019). doi:10.2991/smtesm-19.2019.43
7. Eurostat. https://ec.europa.eu/eurostat (2020). Accessed 17 Aug 2020

8. Kibalnyk, L.O., Danylchuk, H.B.: Analiz naslidkiv vykhodu Velykobrytanii zi skladu YES merezhno-entropiinymy metodamy (Analysis of the consequences of the UK exit from the EU through network-entropy methods). Mizhnarodni Vidnosyny, seriia Ekonomichny nauky 11. http://journals.iir.kiev.ua/index.php/ec_n/article/view/3196 (2017). Accessed 12 May 2020
9. Makhov, S.A.: Dinamicheskaja model vneshnei torgovli stran BRIKS (BRICS trade dynamic model). Preprinty IPM im. M. V. Keldysha 128 (2016). doi:10.20948/prepr-2016-128
10. Mantegna, R.N.: Hierarchical Structure in Financial Markets. European Physical Journal B **11**(1), 193–197 (1999)
11. Minford, P., Xu, Y.: Classical or Gravity? Which Trade Model Best Matches the UK Facts? Open Economies Review **29**, 579–611 (2018). doi:10.1007/s11079-017-9470-z
12. Natale, F., Borrello, A., Motova, A.: Analysis of the determinants of international seafood trade using a gravity model. Marine Policy **60**, 98–106 (2015). doi:10.1016/j.marpol.2015.05.016
13. Ortiz-Ospina, E., Beltekian, D.: Trade and Globalization. https://ourworldindata.org/trade-and-globalization (2018). Accessed 17 Aug 2020
14. Sedgewick, R., Wayne, K.: Algorithms, 4th edn. Addison-Wesley, New York (2014)
15. Shen, B., Zhang, J., Zheng, Q.: Exploring multi-layer flow network of international trade based on flow distances. arXiv:1504.02361v1 [physics.soc-ph] (2015). Accessed 18 May 2020
16. Sneath, P.: The application of computers to taxonomy. Journal of General Microbiology **17**(1), 201–226 (1957)
17. Sokal, R.R., Michener, C.D.: A statistical method for evaluating systematic relationships. The University of Kansas Scientific Bulletin **XXXVIII**(II, 22), 1409–1438 (1958)
18. Soloviev, V.N., Belinskiy, A.: Complex Systems Theory and Crashes of Cryptocurrency Market. Communications in Computer and Information Science **1007**, 276–297 (2019). doi:10.1007/978-3-030-13929-2_14
19. Soloviev, V.N., Yevtushenko, S.P., Batareyev, V.V.: Comparative analysis of the cryptocurrency and the stock markets using the Random Matrix Theory. CEUR Workshop Proceedings **2546**, 87–100 (2019)
20. Tabachkova, N.A., Krachkovskiy, V.V.: Design of processes of international agglomeration. Scientific Bulletin of the Uzhgorod University, Series International economic relations and world economy 6(3), 74–77 (2016)

Modeling the assessment of credit risk losses in banking

Katerina Larionova[0000-0003-4462-9858], Tetyana Donchenko[0000-0002-6929-4639], Andriy Oliinyk[0000-0003-4144-6052], Hennadii Kapinos[0000-0002-9995-1759], Oleg Savenko[0000-0002-4104-745X] and Olexander Barmak[0000-0003-0739-9678]

Khmelnytskyi National University, 11 Instytutska Str., Khmelnytskyi, 29016, Ukraine
k_larionova@i.ua, donchtv@gmail.com, oleynik_andrey@hotmail.com, p_kapinos@i.ua, savenko_oleg_st@ukr.net, Barmak@gmail.com

Abstract. The article develops a model of credit risk assessment within the scope of the variability concept that can be used for verification of new methods for borrowers' credit capacity estimation, the acceptable level of credit risk forecasting and its early prediction. It is aimed to be used during the automated banking systems development. The proposed model of credit risk assessment has been tested on the basis of the data from one of the Ukrainian banks. To determine the adequacy of this model has been proved by the comparison analysis of the proposed model with the results obtained by the National Bank of Ukraine methodology.

Keywords: credit risk, default risk, credit portfolio, risk management.

1 Introduction

The current state of the financial services market, which is unstable due to relevant external and internal factors influencing it, requires the solution of many problems, one of which is the assessment of risks, especially the risks of the most common lending banking operations. In addition, to maintain the stability of the banking system, which is an urgent task in today's conditions of national economy development, it is important to determine the main areas of research in this area and develop appropriate methodological and scientific support.

The questions of effective strategic and tactical bank management are directly associated with the credit risk, which should be acceptable. Taking into account the fact that it is not possible to avoid credit risk completely, the problem of its estimation is of great relevance, since its accuracy affects the minimization of this risk and the choice of the most effective strategy of its management. The appropriate credit risk estimation and the use of risk management methods can be a guarantee of bank financial firmness and stability.

The objective of the paper is to highlight the fundamental principles and develop a model for credit risk assessment using mathematical tools. To pursue this objective the next tasks should be solved: to distinguish the credit risk notion essence; to define the fundamental principles and to develop a model for credit risk assessment using mathematical models on the basis of the probability theory; to approbate the model of credit risk assessment through the Ukrainian bank data.

The concept of "assessment" is defined as the establishment of the level of the indicator. The combination of categories of "credit risk" and "assessment" is reflected in the definition of "credit risk assessment of financial institutions" as establishing the level of possibility of loss by the bank of partial or full loan to the borrower-financial institution due to bad faith, inability of the borrower own financial resources, low liquidity or insufficiency of collateral for the loan, non-performance of obligations by the guarantor or insurer under the loan insurance agreement. The proposed definition differs from the existing ones by taking into account the primary (financial resources of the financial institution) and secondary (collateral, guarantee and insurance collateral for the loan) sources of loan repayment, which helps to improve credit risk management in the bank. Thus, if there is a risk of improper fulfillment of obligations by the borrower to the bank at the expense of primary sources, the bank may take timely measures to minimize the identified threat by strengthening the collateral for the loan (collateral, surety, loan insurance agreement).

Taking into account the economic literature and practice, it can be concluded that the concept of risk is quite multifaceted and can have different meanings depending on the specific scenarios. However, risk is broadly defined as the uncertainty of a particular event occurrence in future. In the banking business the risk refers to financial losses that may take place during the certain transactions performance. In terms of mathematics the risk can be measured by the probability that an anticipated event will not happen (will happen) and it will lead to negative consequences or losses. It should be emphasized that banking business is rather specific, as it is characterized by high riskiness on the one hand, and assurance of reliability and trust, which are realized through providing certain guarantees to clients – on the other hand. The banking system of Ukraine as well as the world banking practice often faces the problem associated with estimation of credit risk, which can lead to financial losses in the result of default on borrower's commitments. According to the requirements of Basel Committee banks by their own should choose methodologies to estimate and model credit risk in their business [1]. Therefore, the problem of choosing modeling methods for credit risk assessment in banking business is of great relevance and controversy. It is obvious that banks have to create and implement their own risk management systems in order to achieve success in high-risk lending activities. It is these systems which help the management apparatus to identify, estimate and control corresponding risks. Nowadays the process of credit risk management is computer-aided, that means that it is based on the results of estimation, analysis and forecasting, obtained with the help of information systems. The main stages of the credit risk management process can be defined as follows: the identification of objective and subjective reasons for credit risk existence; estimation and forecasting of the risk magnitude; credit risks minimization or optimization; regular control of credit risks through credit operations monitoring.

The decision-making system should not only identify and assess the credit risk, but also determine its justifiability or tolerability in terms of the profitability of lending transaction. Acceptable risk is also possible, since it does not threaten the bank, and the eventual losses are less than expected profit and the amount of the compensation for possible losses reserve, which is specially created for the given transaction. Each banking institution defines the limits of acceptable risk by its own according to the chosen management strategy. However, the risk that exceeds the permissible limits is critical, that means that it is much higher than the amount of expected profits and the reserve created. The upper limit of the critical risk is determined by the bank's fund value, so in case of risk event the bank will not make a profit, as well as will lose its own fund. If the risk magnitude is even higher than the critical risk, the bank goes into bankruptcy which is succeeded by the liquidation procedure and the assets sale.

Credit risk is an important component of any bank portfolio and needs to be controlled and managed regularly. It is necessary to use the modern modeling methods and information technologies in order to reduce credit risk and determine its main indicators (probability of default, credit risk appetite, etc.).

The first who began to develop the credit risk assessment econometric models were Altman (Z-score model), Beaver and Tamar. Since then, the models have been improved by scientists and become more elaborate. Scholars [4; 5; 7; 9] in their papers on credit risk assessment develop and refine models, which help to determine according to the research purpose the probability of default, the level of loan portfolio losses, the variation amount of the loan portfolio value, the probability of non-redemption of nonperforming loans and others.

In recent years, large foreign financial institutions have developed a range of credit risk assessment models that vary in elaboration and application methodologies and are widely recognized throughout the world. *CreditRisk+, CreditMetrics, Moody's, KMV Portfolio Manager, Credit Monitor* and *CreditPortfolioView* (*McKinsey & Co., Inc.*) models are among them. In fact, these models are recommended standards to determine credit risks and the basis for the *VAR* approaches development. They allow to determine with different accuracy level the amount of losses associated with credit risk and to calculate the loan derivative value at risk (*value at risk,* thereinafter *VAR*). The models mentioned above have their own peculiarities, but their purpose is the same – to determine the losses apportionment in the credit risks portfolio and, on this basis, to calculate the expected loss in the portfolio at any confidence interval, the change in the extent of these losses and the amount of funds needed for the portfolio support.

The *VAR* determination models allow banks to estimate the difference between the funds needed for the portfolio support and the amount of capital required by the Basel Accords.

Moody's KMV Portfolio Manager model [13], in particular, provides the probability of default defining with the help of methods used in option pricing based on Black-Scholes and Merton models. The particular feature of *KMV Portfolio Manager* is the fact that it is based on the use of the empirical expected default frequencies (EDF), which is calculated with the help of *KMV Credit Monitor* software, developed by the same company.

The *CreditRisk+* model is aimed exclusively for the default risk assessment and does not estimate losses from other credit events. According to *CreditRisk+* model the default losses are estimated through the simple classification of assets by their size, however, the probability of default for each range falls under the gamma distribution, and then the values in ranges are aggregated into the losses distribution due to the default in all ranges.

The class of macroeconomic models includes the Wilson model [15; 16], which has provided the basis for the *CreditPortfolio View* software product, and is designated for credit risk assessment and has been developed by the McKinsey & Co. Consulting Group.

In recent years, SVM model developed by Bernhard E. Boser, Isabelle M. Guyon, and Vladimir N. Vapnik [3] has become more and more popular in the framework of the defaults forecasting as a nonlinear nonparametric algorithm for classification. Western commercial banks and rating agencies implement it in their business. However, despite its obvious advantages, SVM model has some limitations in loan scoring practice.

The review of existing techniques is extremely important for selection, implementation and adaption of the most appropriate model. At the same time, selecting the most appropriate approach, it is necessary to take into account the available mathematical tools, the nature and the quality of benchmark data, the planning horizon, the study objectives and the specifics of the bank portfolio. The personnel skill level, the extent of adoption and usage the latest IT systems and products in the bank are also of great importance. The usage of foreign models is rather complicated, as they have to be adjusted to the Ukrainian economy realities or the own models for credit risk assessment should be developed.

In Ukraine, the Resolution of the Board of the NBU No 351 [10] proposed a model of credit risk assessment for domestic banks, the application of which allows banks to provide a full and timely assessment of credit risk, which will facilitate the correct calculation of their capital and strengthen financial stability of the banking sector. Resolution of the NBU Board No 351 introduces improved approaches to assessing expected credit risk losses and is based on the Basel Principles of Banking Supervision [2], including the use of three components of credit risk (EAD – risk exposure, PD – probability of default of the debtor, LGD - loss in default). This Resolution and this credit risk assessment model itself have been developed for more than a year in cooperation with the banking community with the involvement of experts from the IMF, the World Bank, the international company Oliver Wyman, USAID.

This model is designed to address a number of significant gaps in the current requirements for credit risk assessment, which allowed banks to significantly underestimate the share of problem loans and the amount of credit risk on assets.

The model of credit risk assessment proposed by the Resolution of the NBU Board No 351 has a number of advantages, in particular:

- determining the amount of credit risk in monetary units, rather than interest, which allows you to estimate the actual amount of loss from lending to a particular borrower;
- taking into account when assessing the credit risk, the quality of credit collateral on the loan and the availability of guarantees, sureties or insurance protection under the loan agreement;

- ease of calculation (to determine the financial condition of borrowers – financial institutions that are large and medium-sized enterprises – only 6 indicators are used for financial institutions, small enterprises – 5 indicators, while in the methods of Fitch Ratings and Standard and Poor's such indicators are about 30);
- availability of criteria and scales for assigning the borrower's creditworthiness to the appropriate risk class;
- assessment of the level of credit risk is carried out on the basis of public information and financial statements of financial institutions.

However, some problems of adequate risk assessment of lending to financial institutions remain unresolved, in particular:

- subjectivity of determining the integral value of credit risk (the methodology does not substantiate exactly how qualitative criteria affect the level of credit risk);
- low formalization, taking into account qualitative criteria for credit risk assessment (there are no clear rules and scales in the methodology to determine which value of the qualitative criterion is low, medium or high);
- to assess the financial condition of borrowers of financial companies and credit unions, as well as enterprises of 15 other types of economic activity, the same logistics models are used, although the specifics of these companies differ significantly. Thus, there is an inconsistency of views on the appropriateness of using different approaches to assessing the risk of lending to financial institutions.

The proposed model allows to apply a complex credit risk assessment using the probability theory apparatus, integral calculations and differential equations, which enables to predict the credit risk level and make effective managerial decisions in risk management.

2 Method

A number of situations in which credit risk has been demonstrated have been investigated in the scope of the relevant study and mathematical models based on probability theory have been developed. In this case, the risk measurement procedure involves constructing of ρ function, which assigns a number to each random allocation:

$$\rho : Y \to R \tag{1}$$

where Y is the set of permissible in the particular problem probable distributions that represent risks; R is the set of real numbers.

The analysis and systematization of scientific sources and practical experience, at the present stage, distinguishes the next concepts in the structure of financial risks measurement procedures [8; 14]: losses in adversity; variability; financial risks measurement in the framework of within the expected utility theory; sensitivity.

The risks measurement according to the concept of losses in adversity is a relatively widespread phenomenon in world practice and the domestic banking system. Thus, it is proposed to estimate the expected losses assessed by the bank, depending on the probability of the borrower default, the amount of default losses and the loan extent at the time of default in order to assess credit risk under Basel II [1]:

$$EL = PD \times LGD \times EAD \quad (2)$$

where EL – expected losses; PD (probability of default) – the probability that the borrower will not be able to fulfill his obligations under the contract; LGD (loss given default) – the percentage of losses as relating to the credit amount that the bank will take at the time of the borrower default; EAD (exposure at default) – the loan amount granted at the moment of default.

Under this approach based on the Internal Rating (IRB) the bank will only assess the probability of default and will use data on losses in case of default and exposure to default provided by a credit or rating agency. The European Union has also adopted the methodology for the risk estimation according to the concept of losses in adversity by the Capital Adequacy Directive [6]. The National Bank of Ukraine (NBU) recommends this method to commercial banks for effective market risk management.

It is necessary in the framework of this paper to consider more precisely the concept of variability. It is based on the identification of so-called “location indicators”, relating to which variability is calculated. As a rule, mathematical expectation, variance, median, mode, quantile can be used as “location indicators”. In the scope of the variability concept loan risk has been constructed, as it is the most widespread and elaborate in the banking activity. Formalized approach to the credit financing has helped to come to conclusion that the lending process, especially at the stage of chargeback and repayment of deposit percentage, can be considered as a functional relationship to the time that is random variable. Therefore, it is reasonable to consider this process as a flow of borrower’s payments at the certain point of time with a corresponding probability. Formalized approach to the credit operations also has allowed us to conclude that credit risk can be observed in three aspects. Firstly, the borrowers do not reimburse the bank loans entirely. Secondly, borrowers do not reimburse bank loans on schedule. Thirdly, borrowers reimburse bank loans entirely, but not on time and by partial amounts at one’s own wish, breaking the loan amortization schedule both in time and amount payments. In practice, as a rule, all three aspects of credit risk can be observed simultaneously. Simultaneous action of all the aspects of credit risk will not be considered, as the situations when the loan is not reimbursed by the borrower and the payments are delayed in time are helpful in the analysis of the credit risk elementary components. It is supposed that the symbiosis of these components will help to simulate the third aspect of the credit risk. It can be argued that the process represents a functional relationship to time, as in case of incomplete credit repayment time latency will be observed. On the one hand, if the borrowers do not reimburse the entire amount of the credit, constructing a model it is necessary to take into account the reduction of the amounts which are returned according to maturity dates on schedule, and, on the other hand, in the case of the complete loan amortization, which is repaid with maturity dates break, it is necessary to consider the payments delay time when the repayment amounts are reimbursed completely.

It is obvious that, in its simplest form, the credit reimbursement process can be considered as a repayment of the credit sum S at the moment of time τ. Taking into account the fact that the credit reimbursement process is continuous in time, can change over time, and moreover, it can be of random character in terms of the sums of loan repayments, it is reasonable to introduce a function of time $\xi(t)$, which determines the loan repayment amount at time t. Thus, the appearance of a payment at time t, referring to the loan reimbursement, is a random variable and the value of the function $\xi(t)$ will be directly proportional to the loan repayment amount at time t. If the function $\xi(t)$ indicates by the flow of loan repayment amounts, each succeeding payment may lead to full or partial loan repayment. Therefore, it is reasonable to use the Boolean function $F(U)$ to define the flag indicator U according to which the loan is reimbursed completely. Correspondingly, if $F(U) = 1$, U possess the "truth" value, that means, that credit has been reimbursed completely, and if $F(U) = 0$, U possesses "false" value, that means, that the loan has not yet been repaid.

First of all, it is necessary to consider the situation of incomplete loan reimbursement. Since the process is of random nature, let's define the random variable μ, which denotes the proportion of the reimbursed loan and, correspondingly, takes the value on the interval [0,1]. Thus, the incomplete loan reimbursement process can be represented by the next formula:

$$\xi(t) = \mu \times S \times F\{t \geq \tau\} \tag{3}$$

where μ is a random variable; S is a loan repayment amount before reimbursement process; $F\{t \geq \tau\}$ is a Boolean function.

It is obvious that the random variable μ has a probabilistic nature and the parameters of its distribution form the variable which characterizes the risk of lending operations. Correspondingly, if the probability $P\{\mu = 1\} = 1$, there is no risk, and if $P\ \{\mu = 0\} = 0$, there is the risk event. The nature of this distribution is rather interesting, problematic and highly discussed among the scientists. In the scope of this paper the distribution of random variable will be considered to be normal. In general, the parameters of this distribution can be investigated:

– on the basis of the mathematical expectation (E) of a random variable:

$$E_{\xi(\tau)} = S \times E_{\mu} \tag{4}$$

where E is the mathematical expectation;

– on the basis of variance:

$$D_{\xi(\tau)} = S^2 \times D_{\mu} \tag{5}$$

where D is a variance.

In this case the payment at the time t is a random variable, correspondingly, $E_{\xi(\tau)}$ and $D_{\xi(\tau)}$ are the average and the variance of the payments flow, $S \cdot E_{\mu}$ is the loan repayment average amount.

Let's consider the next case, when loans are returned, but not on schedule, and there is the certain delay time relating to the loan amortization schedule, which can be defined by δ. Thus, the loan reimbursement process can be represented by the formula:

$$\xi(t) = S \times F\{t \geq \tau + \delta\} \quad (6)$$

It is obvious, that δ is random variable, which cannot be of negative value and the parameters of its distribution characterize the risk of credit operations. Thus, there is no risk in the case of $P\{\delta=0\}=1$, and the risk is very high in the case of $P\{\delta=\infty\}=0$, that means that the loan is not reimbursed. Simulating this process, it has been assumed that as the final result the loan would be reimbursed. The amount of payments at any moment of time $t \geq \tau$ can possess the values 0 or S. Regarding the payments flow, the loan repayment amount 0 or S can appear at different moment of times t with varying probability. Thus, the flow average value and variance will be correspondingly equal to:

$$E_{\xi(t)} = S \times P\{\delta \leq t + \tau\} \quad (7)$$

$$D_{\xi(\tau)} = S^2 \times P\{\delta \leq t - \tau\} \times (1 - P\{\delta \leq t - \tau\}) \quad (8)$$

In this model, the average delay time is δ, and the average loan repayment amount is S, that means, that the model presumes that sooner or later the loan will be reimbursed.

Further the model which would try to take into account all the aspects of credit risk, such as time delays and loan reimbursement payments has been built. The set of random variables possessing values on the interval [0, 1] has been defined as $\{\mu_t, t \geq 0\}$ and the following equation is true for it:

$$P\{\mu_0 > z\} > 0 \text{ to all } z<1 \quad (9)$$

Thus, it is possible to estimate the effect of all the credit risk aspects with the help of the next correspondence:

$$\xi(t) = \mu_{t-\tau} \times S \times F\{t \geq \tau\} \quad (10)$$

In this case, the possibility of loan reimbursement by repeated partial payments is taken into account. In addition, the equation (9) ensures a positive probability of complete loan reimbursement on schedule, since it is not rational to assume that the loan has been issued if there has been the confidence that the borrower would not able to return it. Thus, the next model is built:

$$E_{\xi(t)} = S \times E_{\mu_{t-\tau}} \quad (11)$$

$$D_{\xi(t)} = S^2 \times D_{\mu_{t-\tau}} \quad (12)$$

Correspondingly, the average return share is equal to:

$$\lim_{t\to\infty} E_{\mu_t} \quad (13)$$

Practical calculations for this model face the problem of determining the start values, for example, the probabilistic distribution of the loan repayment random variable μ. We believe that the normal (Gaussian) distribution of a random variable is appropriate for this, however μ should have a positive value and cannot greater than 1, as this will contradict the normal distribution condition. If, however, the given random variable is assumed to be Gaussian with some restrictions, and to characterize the amount of client's available assets with the help of which the loan can be reimbursed, its values which exceed mean, that a client has spare funds and the loan can be completely repaid, and if this random variable is of negative value, a client is insolvent or debtor.

Let's estimate the parameters of these models, since only if these parameters exist the model can be implemented in practice. For example, let's consider the first model and set the distribution of the partial loan repayment random variable μ. It is necessary to choose a set of distributions to which μ can be attributed for this purpose. The normal (Gaussian) distribution, which is defined by the mathematical expectation E and the variance D has been chosen. However, the usage of normal distribution is still controversial. Firstly, μ must be of positive value. Although it is not of great importance, since taking into account the typical values of μ makes the probability of its negative value so slight that it can be neglected. Secondly, the μ partial cannot be greater than 1, and such a probability for a normal distribution with typical values of E and D is very high. So, how this problem can be solved?

It can be assumed that there is a random Gaussian variable λ that characterizes the borrower's ability to repay the loan in order to solve this problem. Let's suppose, for example, that it is the indicator of borrower's available assets at the time of debt repayment. Thus, if λ is greater than 1, a borrower has even more funds than needed for complete loan reimbursement and a bank can receive its assets on schedule. If λ is less than 1, a borrower not only does not have enough money to repay the current loan, but probably has other debts. Thus, the next equity can be truth for any random variable λ, which characterizes the client's ability to reimburse the credit:

$$\mu = min(\,max(\lambda, 0),1) \tag{14}$$

Further it is necessary to define the parameters E and D. It is sufficient to estimate the mathematical expectation of E_μ and the probability $P\{\mu = 0\}$ in order to do this. The usage of the bank balance data, that are turnover and balance data on the corresponding accounts, is the easiest way to do this. It is obvious that this estimation can be obtained by dividing the total amount of delayed loan payments by the total loan amount.

Let's denote the repayment amount of loan debt obligations, the obtaining of which is scheduled at the moment of time t by $P(t)$, and let's indicate the amount of delayed payments, which maturity date at the time t has already passed, by $B(t)$. Thus, the loan reimbursement process can be characterized by the functions $p(u)$,where $P(t)$ is the volume from the repayments amount received by the bank for the time increment u (from time t to time $t+u$), and $b(u)$ is the proportion from the repayments amount $P(t)$, which has not been paid at time t. It is obvious, that:

$$p(0) = b(0) = 0 \tag{15}$$

$$p(u) + b(u) = 1 \tag{16}$$

As it can be seen, these functions do not vary depending on time t and therefore, it is appropriate instead of $P(t)$ and $B(t)$ to take their mean values P and B for the time interval u. Thus, the E_μ can be estimated by the approximate formula:

$$E_\mu = \frac{B}{P} \tag{17}$$

where B is the mean value of the delayed payments amount; P is the mean value of the repayment amount of the debt obligations on the loan.

Formula (17) does not vary depending on the scale of the time measurement and, therefore, is correct, and the B and P variables are defined by the balance sheet. It is obvious, that this is not the only way to estimate E_μ, and, therefore, the question can be controversial. It should be noted that in this case the model parameters have been estimated in simplified way and therefore the initial results of the research can be extended in the direction of detalization, justification and elaboration. Let's try to define the result obtained above more exactly. Let's define the amount of bad debts that have been written off at the cost of the reserve at time t and are nonreturnable – by $Z(t)$ is. Thus, the loan reimbursement process can be characterized by the functions $p(u)$, where $P(t)$ is the volume from the repayments amount received by the bank for the time increment u (from time t to time $t+u$), $b(u)$ is the proportion from the repayments amount $P(t)$, which has not been paid at time t, and $z(u)$ is proportion of the payments amount $P(t)$ that has been written off. It is obvious, that:

$$p_t(0) = b_t(0) = z_t(0) \tag{18}$$

$$p_t(u) + b_t(u) + z_t(u) = 1 \tag{19}$$

If it is assumed that the nature of requirements for debt forgiveness does not depend on the time, they are scheduled for, that means, that the requirements quality are approximately the same, these functions do not vary depending on time t. Thus, the next estimation for the functions $B(t)$ and $Z(t)$ can be obtained:

$$B(t) = \int_0^\infty P(t-u) \cdot g'(u)du \tag{20}$$

$$Z(t) = \int_0^\infty P(t-u) \times z'(u)du \tag{21}$$

The process of loan payments obtaining and writing off in the general case can be represented by a system of differential equations:

$$p'(u) = \alpha \times b(u) \tag{22}$$

$$z'(u) = \beta \times b(u) \tag{23}$$

Thus, taking into account the formulas (9) and (10), the system of equations has been developed:

$$\begin{cases} p'(u) = \alpha \times b(u) \\ z'(u) = \beta \times b(u) \\ p_t(0) = b_t(0) = z_t(0) \\ p_t(u) + b_t(u) + z_t(u) = 1 \end{cases} \tag{24}$$

Let the average values of $P(t)$, $B(t)$, $Z(t)$ to be correspondingly P, B, Z for a certain period of time. Thus, after substituting the mean values in formulas (20), (21) and finding the solution of the obtained system of differential equations, the following approximate formula for E_μ estimation can be developed:

$$E_\mu = \frac{1-\frac{Z}{P}}{1+\frac{B}{P}} \tag{25}$$

where B is the mean value of the delayed payments amount; P is the mean value of the repayments amount of the loan debt obligations; Z is the mean value of the bad debts amount that have been written off at the cost of reserves.

Moreover, investor's ration can be added to the model. Thus, overelaborating of the model gives the possibility to make more extended analysis of the information on bank credit risks elimination:

$$E_\mu = \frac{1-\frac{Z}{P}}{1+\frac{r\times B}{P}} = \frac{P-Z}{P+r\times B} \tag{26}$$

where r is the coefficient that can be defined as a market category, such as market risk.

At the last stage the level of credit risk is determined by E_μ variable interpretation with the help of linguistic characteristics of the bank credit risk levels scale (table 1). The proposed scale is based on the calculation of the integrated index of the financial status of a debtor – a corporate entity class, applying the correction factors specified in paragraph 22 Section II of the NBU Regulation №351. Loan payment by a debtor – a corporate entity in time is adopted as correction factors. However, the following demands should be hold to: if there is a delayed payment from 31 to 60 days – the bank gives a debtor the class that is not higher than 5; if this payment is delayed from 61 to 90 days, the class is not higher than 8; in the case of 91 and more days of delayed payment, a debtor gets the class not higher than 10 [16]. Taking into account the fact that the NBU methodology can define the credit risk losses volume, and the proposed model can determine the loan repayment level, it can be concluded, that estimation scales are of inverse character.

Thus, in our opinion, it is necessary to distinguish the following levels of credit risk: acceptable, moderate and critical.

The mathematical expectation of the credit risk level E_μ acquires a value in the range {0; 1}, and the closer it is to 1, the lower the bank credit risk level is. Thus, the formula for E_μ estimation has been developed. It does not depend on time and uses data that can be obtained from the bank's balance. It is obvious that the result obtained is not the only solution to this problem and can be clarified by different ways, which need to be discussed in further investigations.

Table 1. Characteristics of bank credit risk levels.

Credit risk level	Credit risk bounds	Characteristics
Acceptable	1,00– 0,90	Risk of low level, which can be temporarily ignored. Acceptable payment delay – from 31to 60 days.
Moderate	0,89 – 0,41	Moderate risk level, which should be carefully controlled by bank management. Acceptable payment delay – from 61 to 90 days.
Critical	0,40 – 0	Risk of high level, which can lead to bankruptcy. Acceptable payment delay – from 91 days.

It should be noted that the proposed model has certain advantages and limitations. The main advantages are: an integrated approach to the credit risk assessment modeling, which involves various loan repayment scenarios that are relevant from the aspect of early diagnosis of possible credit risk and the bank's decision of its minimization or taking. Taking into account the high share of problem loans in domestic banks, this advantage is particularly important; the possibility to use this model to predict the acceptable level of credit risk, on the basis of which the optimal credit portfolio can be formed; flexibility in application of this model, as it can be used both for the credit risk assessment of the bank credit portfolio in general, and for its components (for example, for a credit portfolio of corporate clients or individuals); ease of use and implementation of this model in bank information systems.

The main difficulties associated with the integrated credit risk assessment based on the proposed model are: insufficiency or total absence of historical data; the absence or inconsistency of statistical data due to the specifics of the banking business or the credit policy peculiarities. However, this should not prevent the banks from developing and applying the proposed model, the input data for which can be obtained from the open source information or based on the experts' reports. The development and implementation of the own methods and models for credit risk assessment, which will provide risk management with the input data for the formation of a bank credit strategy is an extremely important step for domestic banks. In general, it can be argued that there is a problem of effective modeling of credit risk according to the variability concept. This is primarily associated with the asymmetry of the lending activities process, which results into an asymmetric distribution of random variables and correspondingly increases the margin of error, as in the case of the Gaussian distribution. On the other hand, indicators in this model have a tendency to "dispersion" or "scattering" in different directions, that is not inherent in the risk assessment, since as a rule, deviation is considered under adversity. Despite the fact that there are problems referring model parameters assessment, they have been estimated with some restrictions and can be specified in other ways. As a result, the obtained formulas using the real balance data of the balance can be implemented in the banking systems development.

3 Results and discussion

Let's consider the proposed credit risk assessment model on the example of the Ukrainian joint-stock company commercial bank PrivatBank (JSC CB PrivatBank). The obtained results will be compared with the results obtained by the method proposed by NBU in order to determine the model adequacy.

Let's calculate and assess the integrated credit risk of the Ukrainian JSC CB PrivatBank according to the requirements and methods of the National Bank of Ukraine. The information on the credits' distribution by classes of corporate and private debtors, published by the National Bank of Ukraine on January 1, 2018, serves as an input indicator for the bank's credit risk calculation. According to the statistical reporting of corporate JSC CB PrivatBank we have made the table 2 on the credits distribution by classes of debtors [11; 12;].

Table 2. The credits distribution by classes of debtors in JSC CB PrivatBank.

Date	Class of a corporate debtor, billion UAH									
	1	2	3	4	5	6	7	8	9	10
01.02.18	1,6	0,32	0,2	0,78	0,19	0,1	0,44	0,01	0,003	209,8
01.03.18	1,6	0,38	0,24	0,1	0,22	0,12	0,45	0,02	0,003	175,2
01.04.18	2,0	0,37	0,42	0,22	0,07	0,02	0,45	0,01	0,14	202,9
01.05.18	2,2	0,46	0,45	0,28	0,1	0,04	0,42	0,01	0,19	202,9
01.06.18	2,5	0,47	0,41	0,31	0,15	0,21	0,43	0,005	0,041	202,8
01.07.18	2,7	0,61	0,4	0,5	0,13	0,23	0,42	0,003	0,58	203,2
01.08.18	2,6	0,56	0,49	0,44	0,17	0,25	0,44	0,007	0,044	204,5
01.09.18	2,65	0,58	0,57	0,57	0,26	0,24	0,44	0,008	0,041	206,9
01.10.18	2,7	0,67	0,77	0,73	0,25	0,3	0,43	0,006	0,028	215,2
01.11.18	2,7	0,73	0,85	0,94	0,28	0,27	0,01	0,47	0,04	215,5
01.02.19	3,31	0,74	2,49	0,96	0,57	0,04	0,09	0,03	0,25	216,6
01.03.19	3,27	1,03	2,4	0,97	0,55	0,05	0,09	0,03	0,13	215,9
01.04.19	3,54	0,9	2,4	1,23	0,55	0,1	0,16	0,01	0,21	216,8
01.05.19	3,76	0,96	2,28	0,98	1,22	0,46	0,03	0,01	0,19	216,1
01.06.19	3,94	0,82	2,34	1,05	1,35	0,15	0,09	0,01	0,22	216,4
01.07.19	4,22	0,86	2,58	0,97	1,37	0,05	0,12	0,01	0,25	215,5
01.08.19	3,93	0,63	2,52	0,86	1,63	0,09	0,09	0,01	0,29	214,01
01.09.19	4,01	1,53	2,05	0,46	1,77	0,09	0,02	0,05	0,22	214,47
01.10.19	4,01	1,37	2,33	0,86	1,22	0,13	0,01	0,01	0,25	212,80
01.11.19	3,93	1,61	2,29	0,99	1,15	0,13	0,02	0,01	0,27	214,59
01.12.19	4,31	1,59	2,06	1,19	0,58	0,73	0,02	0,01	0,24	213,26
01.01.20	4,35	1,61	2,22	1,18	0,55	0,03	0,03	0,01	0,22	212,85

The loan debtor's losses given default (LGD) granted to corporate entities of JSC CB PrivatBank according to the scenario of a compromise position of risky lending for 10 months of 2018 and 2019 is calculated in table 3.

Thus, the credit risk of JSC CB PrivatBank for loans granted to corporate entities in the scenario of a compromise position of risk lending for 10 months of 2018 is measured by the losses amount from UAH 175,456 billion at the date of March 1, 2018 to UAH 213,162 billion at the date of January 1, 2020.

Table 3. Credit risk for loans granted to corporate entities according to the scenario of a compromise position of JSC CB PrivatBank risky lending in 2018, 2019.

Date	Bank losses for loans, granted to corporate entities by a borrower's – a corporate entities classes, billion UAH										Total losses, billion UAH
	1 (0,5%)	2 (1,0%)	3 (2%)	4 (4%)	5 (7%)	6 (11%)	7 (18%)	8 (33%)	9 (60%)	10 (100%)	
01.02.18	0,008	0,003	0,004	0,031	0,001	0,011	0,079	0,003	0,002	209,8	209,942
01.03.18	0,008	0,004	0,005	0,004	0,015	0,13	0,081	0,007	0,002	175,2	175,456
01.04.18	0,01	0,004	0,008	0,009	0,005	0,02	0,081	0,003	0,084	202,9	203,124
01.05.18	0,011	0,005	0,009	0,011	0,007	0,004	0,077	0,003	0,114	202,9	203,141
01.06.18	0,012	0,005	0,008	0,012	0,011	0,023	0,077	0,002	0,025	202,8	202,975
01.07.18	0,014	0,006	0,008	0,02	0,009	0,025	0,077	0,001	0,35	203,2	203,836
01.08.18	0,013	0,006	0,01	0,018	0,012	0,028	0,079	0,002	0,026	204,5	204,694
01.09.18	0,014	0,006	0,011	0,023	0,018	0,026	0,079	0,003	0,025	206,9	207,105
01.10.18	0,014	0,007	0,015	0,029	0,018	0,033	0,077	0,002	0,017	215,2	215,43
01.11.18	0,014	0,007	0,017	0,038	0,02	0,03	0,002	0,155	0,024	215,5	215,807
01.02.19	0,017	0,007	0,050	0,038	0,040	0,004	0,016	0,010	0,15	216,6	216,933
01.03.19	0,016	0,010	0,048	0,039	0,039	0,006	0,016	0,010	0,078	215,9	216,162
01.04.19	0,018	0,009	0,048	0,049	0,039	0,011	0,029	0,003	0,126	216,8	217,132
01.05.19	0,019	0,010	0,046	0,039	0,085	0,051	0,005	0,003	0,114	216,1	216,472
01.06.19	0,020	0,008	0,047	0,042	0,095	0,017	0,016	0,003	0,132	216,4	216,779
01.07.19	0,021	0,009	0,052	0,039	0,096	0,006	0,022	0,003	0,15	215,5	215,896
01.08.19	0,020	0,006	0,050	0,034	0,114	0,010	0,016	0,003	0,174	214,01	214,438
01.09.19	0,020	0,015	0,041	0,018	0,124	0,010	0,004	0,017	0,132	214,47	214,851
01.10.19	0,020	0,014	0,047	0,034	0,085	0,014	0,002	0,003	0,15	212,8	213,170
01.11.19	0,020	0,016	0,046	0,040	0,081	0,014	0,004	0,003	0,162	214,59	214,975
01.12.19	0,022	0,016	0,041	0,048	0,041	0,080	0,004	0,003	0,144	213,26	213,658
01.01.20	0,022	0,016	0,044	0,047	0,039	0,003	0,005	0,003	0,132	212,85	213,162

Losses from credit risk on loans granted to corporate entities of JSC CB PrivatBank according to the scenario of the aggressive position of risky lending for the period of 10 months of 2018 and 2019 are represented in the table 4. Thus, the credit risk of JSC CB PrivatBank for loans granted to corporate entities according to the scenario of an aggressive position of risky lending for the research period is measured by the losses amount from UAH 175,565 billion at the moment of March 1, 2018 to UAH 213,351 billion at the date of January 1, 2020.

The calculation of the mathematical expectation of loan repayments according to the proposed model and the analysis of the results are represented in the table 5. Taking into account the results given in the table 5, it can be concluded that the E_μ value is the most closely approximate to zero, that gives evidence of the critical level of JSC CB PrivatBank credit risk in 2018-2019, which corresponds to the results of the NBU methodology. The results obtained with the help of the proposed model and the NBU methodology do not differ significantly by the majority of indicators. Thus, it can be concluded that the proposed model is an effective tool for credit risk assessment in banking business. This statement has been tested on actual data.

Table 4. Credit risk for loans granted to corporate entities according to the scenario of an aggressive position of JSC CB PrivatBank risky lending in 2018, 2019.

Date	Bank losses for loans, granted to corporate entities by a borrower's – a corporate entities classes, billion UAH										Total losses, billion UAH
	1 (0,9%)	2 (1,9%)	3 (3%)	4 (6%)	5 (10%)	6 (17%)	7 (32%)	8 (59%)	9 (99%)	10 (100%)	
01.02.18	0,144	0,006	0,006	0,047	0,019	0,017	0,141	0,006	0,003	209,8	210,189
01.03.18	0,144	0,007	0,007	0,006	0,022	0,02	0,144	0,012	0,003	175,2	175,565
01.04.18	0,18	0,007	0,013	0,013	0,007	0,003	0,144	0,006	0,139	202,9	203,412
01.05.18	0,198	0,009	0,014	0,017	0,01	0,007	0,134	0,006	0,188	202,9	203,479
01.06.18	0,225	0,009	0,012	0,019	0,015	0,036	0,138	0,003	0,041	202,8	203,334
01.07.18	0,243	0,012	0,012	0,03	0,013	0,039	0,134	0,002	0,574	203,2	204,247
01.08.18	0,234	0,011	0,015	0,026	0,017	0,043	0,141	0,004	0,436	204,5	205,427
01.09.18	0,239	0,011	0,017	0,034	0,026	0,041	0,141	0,005	0,041	206,9	207,455
01.10.18	0,243	0,013	0,023	0,044	0,025	0,051	0,138	0,004	0,028	215,2	215,769
01.11.18	0,243	0,014	0,026	0,056	0,028	0,046	0,003	0,28	0,039	215,5	216,235
01.02.19	0,030	0,014	0,075	0,058	0,057	0,007	0,029	0,018	0,248	216,6	217,134
01.03.19	0,029	0,020	0,072	0,058	0,055	0,009	0,029	0,018	0,129	215,9	216,318
01.04.19	0,032	0,017	0,072	0,074	0,055	0,017	0,051	0,006	0,208	216,8	217,332
01.05.19	0,034	0,018	0,068	0,059	0,122	0,078	0,010	0,006	0,188	216,1	216,683
01.06.19	0,035	0,016	0,070	0,063	0,135	0,026	0,029	0,006	0,218	216,4	216,997
01.07.19	0,038	0,016	0,077	0,058	0,137	0,009	0,038	0,006	0,248	215,5	216,127
01.08.19	0,035	0,012	0,076	0,052	0,163	0,015	0,029	0,006	0,287	214,01	214,685
01.09.19	0,036	0,029	0,062	0,028	0,177	0,015	0,006	0,030	0,218	214,47	215,070
01.10.19	0,036	0,026	0,070	0,052	0,122	0,022	0,003	0,006	0,248	212,8	213,384
01.11.19	0,035	0,031	0,069	0,059	0,115	0,022	0,006	0,006	0,267	214,59	215,201
01.12.19	0,039	0,030	0,062	0,071	0,058	0,124	0,006	0,006	0,238	213,26	213,894
01.01.20	0,039	0,031	0,067	0,071	0,055	0,005	0,010	0,006	0,218	212,85	213,351

Table 5. Comparative results of the JSC CB PrivatBank credit risk assessment according to the proposed model and the NBU methodology.

Date	According to the proposed model						According to the NBU methodology [12]	
	B, UAH billion	Z, UAH billion	P, UAH billion	E_μ	Repayments amount, UAH billion	Losses volume, UAH billion	Losses according to the compromise scenario, UAH billion	Losses according to the aggressive scenario, UAH billion
01.02.18	2,936	209,849	212,785	0,0136	2,894	209,891	209,942	210,189
01.03.18	3,216	208,903	212,119	0,0149	3,161	208,958	175,456	175,565
01.04.18	3,557	203,033	206,59	0,0169	3,491	203,099	203,124	203,412
01.05.18	3,991	203,064	207,055	0,0189	3,913	203,142	203,141	203,479
01.06.18	4,435	202,817	207,252	0,0209	4,332	202,920	202,975	203,334
01.07.18	4,936	203,253	208,189	0,0231	4,809	203,380	203,836	204,247
01.08.18	4,971	204,588	209,559	0,0232	4,862	204,697	204,694	205,427
01.09.18	5,314	206,986	212,3	0,0243	5,159	207,141	207,105	207,455
01.10.18	5,810	215,182	220,992	0,0256	5,657	215,335	215,43	215,769
01.11.18	6,194	215,488	221,682	0,0272	6,030	215,652	215,807	216,235

Date	According to the proposed model						According to the NBU methodology [12]	
	B, UAH billion	Z, UAH billion	P, UAH billion	E_μ	Repayments amount, UAH billion	Losses volume, UAH billion	Losses according to the compromise scenario, UAH billion	Losses according to the aggressive scenario, UAH billion
01.02.19	8,23	216,85	225,08	0,0353	7,945	217,135	216,933	217,134
01.03.19	8,39	216,03	224,42	0,0360	8,079	216,341	216,162	216,318
01.04.19	9,06	217,01	226,07	0,0385	8,704	217,366	217,132	217,332
01.05.19	9,70	216,29	225,99	0,0412	9,311	216,679	216,472	216,683
01.06.19	9,75	216,62	226,37	0,0413	9,349	217,01	216,779	216,997
01.07.19	10,18	215,75	225,93	0,0431	9,738	216,192	215,896	216,127
01.08.19	9,76	214,3	224,06	0,0417	9,343	215,327	214,438	214,685
01.09.19	9,98	214,69	224,67	0,0425	9,549	215,121	214,851	215,070
01.10.19	9,94	213,05	222,99	0,0427	9,522	213,468	213,170	213,384
01.11.19	10,13	214,86	224,99	0,0431	9,697	215,293	214,975	215,201
01.12.19	10,49	213,50	223,99	0,0447	10,012	213,978	213,658	213,894
01.01.20	9,98	213,07	223,05	0,0428	9,547	213,953	213,162	213,351

4 Conclusion

The proposed model of credit risk assessment according to the concept of variability is universal and allows us to predict the credit risk level and make effective management decisions due to the means of probability theory, integral calculations and differential equations. This model provides to determine the mathematical expectation E_μ, using data, which are not dependent on time and can be obtained from a banking institution's balance sheet. The model uses a comprehensive approach to credit risk assessment modeling, which involves different ways of loan repayment.

The main advantages of the proposed credit risk assessment model include: a complex approach to modeling credit risk assessment, which involves different kinds of credit repayment; the possibility to use this model for the acceptable credit risk level prediction, on the basis of which the optimal bank credit portfolio can be built; universality of this model (it can be used both for the assessment of bank porfolio credit risk taken as a whole, and for its components); ease of use.

The proposed model for credit risk assessment is one of a set of models, which is primarily can be used during development of automated banking systems, credit risk management systems and expert systems. This model can be also useful for verification of new methods for borrowers' credit capacity assessment and credit risk forecasting.

The model can be included to the so-called "block of models", which is a part of modern decision-making systems. Such systems can be used in banking business, that significantly increases the level of financial management in the area of credit risk management.

References

1. Basel Committee on Banking Supervision (BCBS): Basel II: International Convergence of Capital Measurement and Capital Standards: A revised framework, comprehensive version. http://www.bis.org/publ/bcbsc111.htm (2005). Accessed 27 Aug 2020
2. Basel Committee on Banking Supervision: Revisions to the Standardised Approach for credit risk, 2nd consultation. Bank for International Settlements. http://www.bis.org/bcbs/publ/d347.pdf (2015). Accessed 20 August 2020
3. Boser, B.E., Guyon, I.M., Vapnik, V.N.: A training algorithm for optimal margin classifiers. In: COLT '92: Proceedings of the fifth annual workshop on Computational learning theory, pp. 144–152 (1992). doi:10.1145/130385.130401
4. Cai, R., Zhang, M.: How does credit risk influence liquidity risk? Evidence from Ukrainian banks. Visnyk of the National Bank of Ukraine **241**, 21–33 (2017)
5. Chatterjee, S.: Modelling credit risk. Centre for Central Banking Studies, Bank of England, London. https://www.bankofengland.co.uk/-/media/boe/files/ccbs/resources/modelling-credit-risk (2015). Accessed 29 Jul 2019
6. Council Directive 93/6/EEC of 15 March 1993 on the capital adequacy of investments firms and credit institutions. Official Journal of the European Communities **36**(L 141), 0001–0026
7. Gostrik, O.A., Stepanenko, O.A.: Definition of permissible credit risk of commercial bank with using the method of system dynamics. Scientific Herald of Uzhgorod National University 6(3), 60–63 (2016)
8. Kaminsky, A.B.: Conceptual approaches to the measurement of financial risks. Finance of Ukraine 5, 78–85 (2006)
9. Sheu, H.-J., Liu, C.-L.: The Credit Risk Pricing with Particle Filter Approach. Proceedings of the 9th Joint Conference on Information Sciences, JCIS 2006, CIEF-89
10. The National Bank of Ukraine. Regulation No 351 on approval of the Regulation on the Ukrainian banks determination of the credit risk magnitude for active banking transactions. https://zakon2.rada.gov.ua/laws/show/v0351500-16 (2016). Accessed 17 Aug 2020
11. The National Bank of Ukraine: Statistics. financial statements of Ukrainian banks. https://bank.gov.ua/files/stat/Inc_Exp_Banks_2020-09-01.xlsx (2020). Accessed 22 Aug 2020
12. The National Bank of Ukraine: Statistics. The main indicators of the banks of Ukraine. https://bank.gov.ua/files/stat/Indicators_Banks_2020-09-01.xlsx (2020). Accessed 22 Aug 2020
13. Vasicek, O.A.: Probability of Loss on Loan Portfolio. KMV Corporation, San Francisco. http://www.moodyskmv.com/research/files/wp/Probability_of_Loss_on_Loan_Portfolio.pdf (1987). Accessed 17 Aug 2020
14. Vitlinskyi, V.V., Velykoivanenko, H.I.: Financial risk and methods of its measurement. Finance of Ukraine 12, 13–23 (2000)
15. Wilson, T.: Portfolio Credit Risk (I). Risk **10**(9), 56–61 (1997)
16. Wilson, T.: Portfolio Credit Risk (II). Risk **10**(10), 111–117 (1997)

The impact of COVID-induced shock on the risk-return correspondence of agricultural ETFs

Andrii Kaminskyi[1[0000-0002-6574-8138]], Maryna Nehrey[2[0000-0001-9243-1534]] and Nina Rizun[3[0000-0002-4343-9713]]

[1] Taras Shevchenko National University of Kyiv, 64 Volodymyrska Str., Kyiv, 01026, Ukraine
kaminskyi.andrey@gmail.com
[2] National University of Life and Environment Science of Ukraine, 15 Heroyiv Oborony Str., Kyiv, 03041, Ukraine
marina.nehrey@gmail.com
[3] Gdańsk University of Technology, 11/12 Gabriela Narutowicza Str., Gdańsk, 80-233, Poland
nina.rizun@pg.edu.pl

Abstract. Risk-return correspondence for different investment asset classes forms one of the pillars of modern portfolio management. This correspondence together with interdependency analysis allows us to create portfolios that are adequate to given goals and constraints. COVID-induced shock unexpectedly generated high uncertainty and turmoil. Our paper is devoted to the investigation path through shock by agricultural assets (presented by ETFs) in comparison with traditional assets. There were identified three time periods: before the shock, explicitly shock, and post-shock. At the explicit shock period was suggested estimation risk frameworks on the pair indicators: falling depth and recovery ratio. Basic attention focuses on comparison risk-return estimations prior to shock and post-shock. To this end was considered four approaches to risk measurement and were applied to the sample of agricultural ETFs. The results indicated differences in risk changing by the path from before shock to post-shock. Differences arise from choosing the approach of risk measuring. The variability approach reveals much growth of risk of traditional assets, but the Value-at-Risk approach indicates higher risk growth for agricultural ETFs. Combine together with relatively low correlation these estimations provide a clear vision of risk-return frameworks.

Keywords: exchange traded funds, risk measurement, COVID, shock, portfolio management, agriculture, investment.

1 Introduction

The COVID-19 pandemic has a strong influence on the prices of all financial instruments [29]. Financial markets had shivered at the end of January 2020 and crashed in the middle of March 2020. The shock was extremely forceful. COVID-induced shock hit almost all assets: as traditional assets as alternative assets (including cryptocurrencies). Correspondingly, the shock had an effect on investment portfolio management which led to decreasing portfolio value. Meanwhile, different assets have different dynamics of passing such a turbulent period. Does it necessary to change the asset allocation design of investment portfolios? This question became an actual one for individual and institutional investors. The aim of this paper is to investigate risk-return correspondence "transmission" through the COVID-induced shock for agricultural Exchange Trade Funds (ETFs) and ETN.

Two hypotheses were put forward in our research. First hypothesis conjectures differences of shock parameters for agricultural ETFs and two traditional asset classes such as stocks and bonds (presented in our research by key stock indices). Especially, it was supposed differences in the renewal level. Second, our hypothesis focuses on verification of the assumption that risk is higher aftershock then before the shock. In general, this is a typical effect and we have tried to estimate the level of such risk increasing.

Class of agricultural ETFs one of the significant parts of the commodities ETFs and has its own distinctive features. The first distinctive aspect is that the prices of agricultural production are determined both by market factors (demand in the first place) and the crop yield (production) of a particular agricultural product. The dependence on the yield generates an additional level of lack of correlation of such ETFs with other investment assets, which can be used in investment portfolio forming procedures. The second distinctive feature of the agricultural ETFs is their structuring into ETFs associated with one agricultural product (for example, wheat, rice, livestock, sugar, and others), and associated with a specific fund diversified through different agricultural products. One of the interesting points for analysis concerns the meaningfulness of such features at the time of shock and renewal. Understanding the difference in "risk-return correspondence" in this context will allow a better justification for their using in the portfolio structure.

Our approach involves ETF using. The emergence of ETFs in the early 1990s and their intensive development expanded the portfolio management tools in two ways. First, the essence of the ETF design has allowed expanding the asset classes that can be used in the portfolios. In this regard, it is possible to use ETF connected with non-traditional investments (commodities, gold, private equity, and many others). Such possibilities essentially expand the diversification effect through portfolio construction. As a rule, alternative investments indicate a lower correlation level with others. Secondly, ETFs make it easy to assess the risk and return of the entire portfolio based on their characteristics. In addition, to some extent, with this approach, the task of filling the class with assets can be removed, because ETF diversified funds can be used. The task of portfolio investment, in fact, is more reduced to a strategic allocation. So, we used ETFs for analysis risk-return correspondence for agricultural assets.

It should be noted that we applied a complex view of the notion of "risk measurement". Modern financial risk theory considers different approaches to measure risk. Each approach reflects one or another property of the many-sided notion of "risk". We used three approaches to risk measurement. A first approach based on the classical view for risk measurement at the frameworks of variability. The second approach considers risk from point of view losses in a negative situation. The importance of such an approach is explained by using the regulative risk measure Value-at-Risk (VaR) and coherent risk measure Conditional Value-at-Risk (CVaR). The third approach is based on conception sensitivity. It is logically to use sensitivity analysis in concern both types of traditional assets – stocks and bonds. The results of using such a complex approach are a generalized estimation of risk characteristics changing. Such an approach provides a deeper understanding of investment risk frameworks.

2 Materials and methods

2.1 Risk measurement conception

Risk measurement in the frameworks of portfolio investment can be structured into two blocks. The first block is a risk assessment of an investment asset, considered separately. The second block focuses on assessing the relationship between asset returns and risk through diversification.

The first block of risk assessment supposes to introduce mapping μ which each return of investment asset R (interpreting as random variable) correspond some non-negative number $\mu(R) \subset [0;+\infty]$. The return of investment asset (in this paper – ETF) over a period of time $[t;\ t+1]$ will be expressed through the formula:

$$R_{t,\,t+1} = (P_{t+1} - P_t) \,/\, P_t \tag{1}$$

where P_t and P_{t+1} prices of ETF in USD at times t and $t+1$ correspondingly. $R_{t,\,t+1}$ will be a random variable, because the future price P_{t+1} is unknown. Thereafter R which reflect return through the time is also random variable. Mapping μ which corresponds to some rules interpret as risk measuring.

2.2 Investment risk measures approaches

There are many measures of investment risk present which formalise in mapping μ different logic of risk interpreting [33]. In our research, we have divided risk measuring into three conceptual approaches:

- Variability approach. Such an approach is based on the measurement of return`s variability (volatility). This approach goes back to the papers of H. Markowitz [21] and underlies the models of modern portfolio theory. Critiques of it using in the non-transparency connection between variability indicators and real losses.
- Losses in a negative situation. This more practical and regulative approach. It focuses on measuring possible losses and fulfill capital requirements.
- Sensitivity approach. According to such an approach, the risk is measured as the rate of response for occurring some factors.

Each of the abovementioned approaches had their pros and cons. Our point of that investment risk should be estimated by all these conceptual approaches. It provides multifaceted understanding of investment risk.

The logic of risk measuring leads to properties which reflect "natural properties" of risk. Trying to understand the essence of properties which should be represented in risk measure was formulated in [3]. Authors created the notion of coherent risk measure. Risk measure is coherent if satisfying following properties (axioms):

Axiom 1. Sub-additivity. For all random values presenting asset`s returns R_1 and R_2 we have

$$\mu\ (R_1 + R_2) \leq \mu(R_1) + \mu(R_2) \tag{2}$$

Axiom 2. Positive Homogeneity. For all R and for all $\lambda \geq 0$, we have

$$\mu(\lambda R) = \lambda\mu(R) \tag{3}$$

Axiom 3. Monotonicity: If $R_1 \geq R_2$ for all possible cases then

$$\mu(R_1) \leq \mu(R_2). \tag{4}$$

Axiom 4. Translation Invariance. For all R and for all $\alpha \geq 0$ which interpret as risk-free asset, we have

$$\mu(R+\alpha) = \mu(R)-\alpha. \tag{5}$$

Examples of coherent risk measures are Conditional Value-at-Risk (considered introduced below) [28] and T. Fischer measure [8]. It is necessary to note, that presented approach for coherency is not unique. Other approaches of coherency are considered in [18].

The second block of risk measurement in the portfolio aspect corresponds to estimate interrelations of returns of different asset classes. It can be estimated as average correlation, reducing the value of chosen risk measure for a naïve diversified portfolio or risk value for the portfolio with minimum risk.

Below we try to realize these ideas for agricultural ETFs.

2.3 Risk measurement throughout the period of shock

A financial shock is an exceptionally extraordinary event that affects the entire market. Therefore, the classical approaches to measuring risk may be ineffective and we used the following approach. Based on the analysis of the manifestation of COVID-induced shock, we divided the time interval into three periods. The first period is the “calm” period before the onset of the shock. The shock-related asset price changes began to show in the second half of January 2020. Therefore, we had to take 08/28/2019 to 01/15/2020 as the first period. The role of measuring risk in a given period serves as a benchmark for further changes.

As the second period, we have identified the period 01/16/2020 – 03/31/2020 – the direct manifestation of shock. The manifestation of COVID-induced shock was, in a sense, a classic manifestation of shock. Namely, it had the form Sign of “tick”. At first, the onset of a shock is a gradual fall in asset prices, and then a sharp and deep fall. The shock drop was on 03/17/2020 for the studied assets. After that, a gradual slow price recovery begins. Moreover, at first, after the maximum fall, there is a “rollback”, and then the dynamics stabilize. Thus, as the post-shock period, we have defined the period 04/01/2020 to 08/14/2020.

The use of classical risk measures is not correct due to a sharp fall in a short period. To display risk during a shock period, we have proposed an approach based on two parameters. The first parameter characterizes the depth of the fall, and the second – the level of recovery over a certain period. The parameter that characterizes the depth of the fall is calculated by us as the ratio of the lowest price to the average price for 1,5 months before the start of the shock period. And the second parameter is calculated based on the average stabilization price after the maximum decline. In our case, for calculating average prices, we took the periods 12/01/2019–01/15/2020 and 05/01/2020–06/15/2020.

The logic for calculating the parameters is shown in the fig. 1 for SPY (ETF which correspond to leading stock index S&P500).

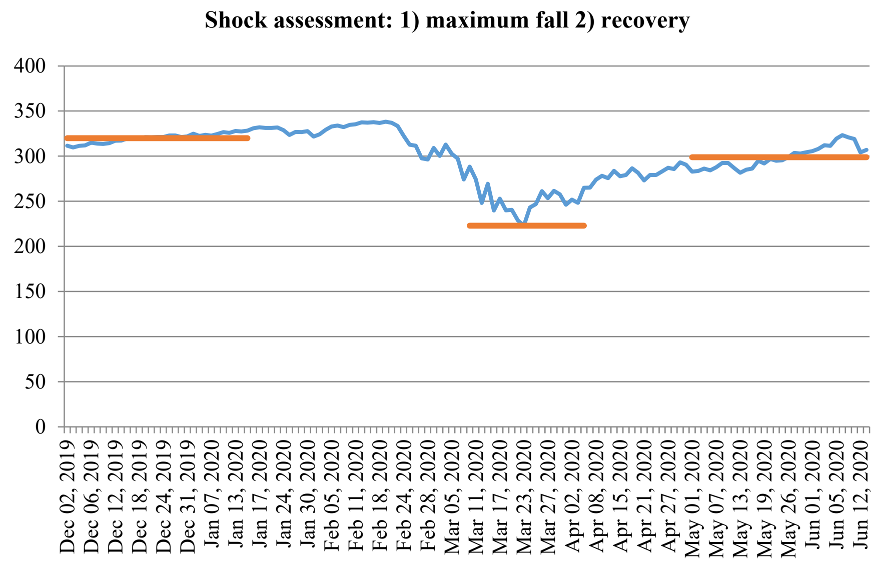


Fig. 1. Parameters of risk during shock period for SPY.

As the third, for this period, we have applied standard approaches to measuring risk. They are compared with the values of these parameters in the first period. The economic sense of the study is in assessing the risk changes as a result of shock.

3 Results and discussion

3.1 Literature review

There has been a lot of academic studies that have addressed agricultural investment and agriculture assets. The last of them are [2; 5; 6; 7; 15; 16; 27; 36].

Martin and Clapp [22] investigated the relationship between agriculture, finance, and the state. In [10] the authors analyzed the relation between the notional value of commodity futures contracts and expected returns on futures contracts.

ETFs as financial instruments investigated in [14] and [32]. Petajisto proposed a method for ETFs mispricings detection [25].

The global challenges caused by COVID have updated crisis and shock research. The analysis of the impact of macroeconomic changes on the financial market was conducted in [1; 11; 17; 19; 24; 26; 30]. Financial security level analysis in order to timely detect and neutralize possible crisis phenomena presents in [9; 13; 20].

Forecasting the dynamics of financial markets during the crisis is studied in [23; 31; 34; 35].

In spite of shortness time after COVID-induced shock, there are a lot of papers described this phenomenon. The uncertainty which have raised from this shock is analyzed in [4].

In new European Banking Study 2020, was quantified COVID induced effects on balance sheets and P&Ls of Europe's 50 largest banks and set out the implications for bank management, governments, and regulators [12].

3.2 Sample of agricultural ETFs

Our sample of agricultural ETFs was created on the base of capitalization level of such financial instrument which traded in the USA which are currently tagged by ETF Database. It is necessary to note that we use term ETF in extend sense which include both instruments which tracking indices: ETF and ETN. Of course, we pay attention for the differences between these instruments, but our main focus for the conceptual essence of tracking indices, after that we did not differentiate ETF and ETN in our paper and use one term ETF.

Agriculture ETFs invest in agriculture commodities including sugar, corn, soybeans, coffee, wheat and other. It can be single commodity fund or diversified fund. We have formed sample (11 components) based on total assets volume by following ETFs (ETN).

CORN. This ETF corresponds to Teucrium Corn Fund which tracks an index of corn futures contracts.

COW. This ETN offers an opportunity for investors to gain exposure to hogs and cattle iShares Global Agriculture Index ETF.

DBA. This ETF corresponds to diversified basket of various agricultural natural resources.

FUD. This is ETN, associated with futures-based index that measures the collateralized returns from a basket of 11 futures contracts from the agricultural and livestock sectors.

JJSF. This is ETN which connected with sugar futures.

NIB. This ETN offers exposure to cocoa futures.

RJA. RJA ETN tracks Rogers International Commodity Index-Agriculture which is consumption-based index of agricultural commodities.

UAG. Exchange-traded note which offers exposure to a number of agricultural commodities, including corn, soybeans, wheat, coffee, cocoa, and other natural resources.

CANE. This ETF offering exposure to the commodity of sugar.

SOYB. This ETF invests in soybean futures contracts.

WEAT. This ETF offers exposure to wheat futures contracts.

The following ETFs were chosen for comparison agricultural ETFs with traditional assets ETFs.

SPDR's **SPY** to model the large-cap public equities, it tracks the Standard & Poor's 500 and is the oldest and largest of all ETFs.

SPDR's **MDY** that tracks the Standard & Poor's 400 to model the mid-cap equities, while being smaller than iShares IJH it has about the same turnover but offers a longer time series.

iShares **IJR** to model the small-cap companies, it tracks the Standard & Poor's 600 index and is much larger and liquid than the corresponding SPDR fund SLY.

iShares **IEF** to model a balanced portfolio of Treasury bonds, the choice of this particular government bond fund is motivated by its duration 7,6 years that is comparable to the duration of other bond funds analyzed in this paper.

iShares **LQD** to model a balanced portfolio of investment-grade corporate bonds, it's one of the oldest bond ETFs and its duration (8,5 years) is approximately the same as for the IEF fund mentioned above, so we can contrast government and corporate bonds.

iShares **TIP** to model inflation-linked bonds, an asset class that should have quite a distinct characteristic, however its duration (7,6 years) aligned to LQD and IEF.

3.3 Measurement of shock characteristics

The measurement of the characteristics of the shock was carried out, as noted above, within the framework of 01/16/2020–03/31/2020 based on two indicators. The first indicator is the depth of the fall (fig. 2). In the context of our work, it can be interpreted as a "measure of risk in shock conditions". The second indicator, the percentage of recovery after a fall, can be interpreted as "profitability in a shock". The economic meaning of this parameter can be interpreted in two directions. First, this is a formal interpretation of the situation to buy assets at a low point and receive income in the recovery process. The second direction concerns the comparison of the falling percentage and the recovery percentage.

Two observations are interesting. The first is that ETFs that match stock indices (especially MDY and IJR) have a deeper fall than most agricultural ETFs. However, the recovery rate is higher. The second observation is that ETFs of bonds did not have a great dip and a recovery rate of about 100%, or even more. The first indicates a high sensitivity of stocks to shock, while bonds are in high demand. Agricultural ETFs are in the middle.

3.4 The variability approach to risk measurement

Table 1 present the comparative analysis which was realized twofold. One side characterizes differences in risk measures prior to and post-shock. The other side characterizes differences of risk measures for alternative and traditional assets. Prior to the shock agricultural ETFs indicate higher values of range than traditional assets (on average close to two times more). After the shock, the widening of the range had concerned both types of assets, but growth of range for traditional assets was essentially more. So, post-shock average ranges for traditional and agricultural ETFs approximately equal. The average growth of ranges in returns was 4% for agricultural ETFs and 5,8% for traditional assets.

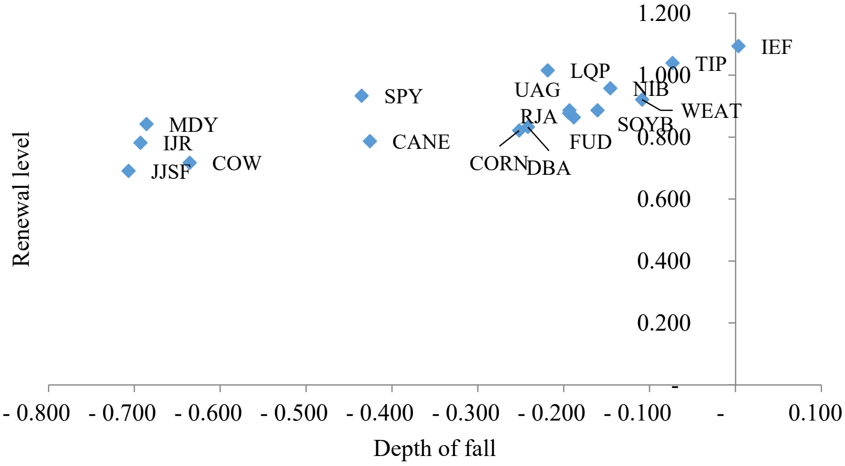


Fig. 2. Depth of fall via renewal level.

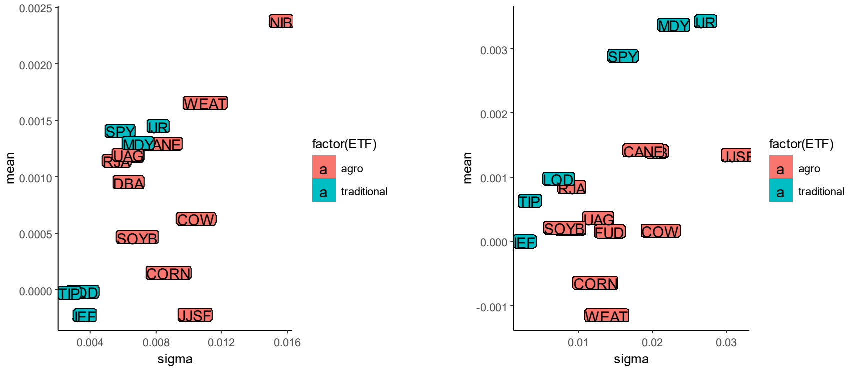
Table 1. Statistical analysis for risk measures. Agricultural ETF.

ETF	min		max		mean		std		skewness		kurtosis	
	Before shock	Post-shock	Befo-re shock	Post-shock	Befo-re shock	Post-shock	Before shock	Post-shock	Before shock	Post-shock	Before shock	Post-shock
Agricultural ETF												
CORN	**-0,025**	-0,031	0,031	**0,037**	0,000	-0,001	0,009	**0,012**	**0,650**	0,142	1,942	**0,948**
COW	**-0,030**	-0,062	0,040	**0,065**	0,001	0,000	0,010	**0,021**	**0,444**	0,289	1,973	**2,048**
DBA	**-0,022**	-0,027	**0,026**	0,024	0,001	0,000	0,006	**0,009**	**0,382**	0,053	**2,925**	0,531
FUD	**-0,020**	-0,051	0,017	**0,044**	0,001	0,000	0,006	**0,014**	**0,006**	-0,185	0,691	**1,729**
JJSF	**-0,051**	-0,073	0,018	**0,085**	0,000	**0,001**	0,010	**0,032**	-1,260	**0,154**	**4,694**	0,342
NIB	**-0,030**	-0,058	0,051	**0,065**	**0,002**	0,001	0,016	**0,020**	**0,262**	0,185	-0,124	**0,642**
RJA	**-0,009**	-0,028	0,021	**0,025**	**0,001**	0,001	0,006	**0,009**	**0,882**	-0,424	**1,525**	0,961
UAG	**-0,012**	-0,052	0,022	**0,056**	**0,001**	0,000	0,006	**0,013**	**0,561**	0,065	0,589	**5,571**
CANE	**-0,015**	-0,052	0,021	**0,041**	0,001	**0,001**	0,008	**0,019**	**0,151**	-0,172	-0,314	**0,401**
SOYB	**-0,014**	-0,025	**0,031**	0,022	**0,000**	0,000	0,007	**0,008**	**0,826**	-0,069	**3,581**	0,765
WEAT	**-0,020**	-0,031	0,030	**0,044**	**0,002**	-0,001	0,011	**0,014**	0,412	**0,498**	-0,365	**0,363**
ETF of traditional assets												
SPY	**-0,018**	-0,046	0,014	**0,067**	0,001	**0,003**	0,006	**0,016**	-0,482	**0,034**	0,840	**2,870**
MDY	**-0,019**	-0,060	0,017	**0,081**	0,001	**0,003**	0,007	**0,023**	-0,188	**0,127**	0,581	**1,219**
IJR	**-0,020**	-0,070	0,025	**0,082**	0,001	**0,003**	0,008	**0,027**	**0,282**	0,092	**0,500**	0,367
IEF	-0,009	**-0,006**	0,009	**0,009**	0,000	**0,000**	**0,004**	0,003	-0,268	**-0,236**	-0,113	**0,707**
LQD	**-0,009**	-0,017	0,008	**0,047**	0,000	**0,001**	0,004	**0,007**	-0,384	**-0,085**	-0,013	**0,244**
TIP	**-0,006**	-0,009	0,006	**0,015**	0,000	**0,001**	0,003	**0,003**	-0,086	**0,434**	-0,263	**0,198**

The situation with standard deviation (std) is similar by essence. Growth of std was for both types of assets, but std for traditional assets demonstrated a faster pace. Average growth of std in returns was 0,68% for agricultural ETFs and 8,1% for traditional assets.

A very interesting difference between agricultural ETFs and traditional assets for average return before and post-shock. They have equal average returns before shock but traditional assets post-shock demonstrated triple higher average returns. At the same time agricultural ETFs shown changing positive returns for negative.

The changing of risk-return correspondence prior to and post-shock is illustrated by Fig. 3. It is very interesting that post-shock traditional assets form exactly efficient frontier at the Markowitz sense.



a) 08/28/2019 – 01/15/2020 b) 04/01/2020 – 08/14/2020

Fig. 3. ETFs risk-return correspondence.

It is interesting results we can identify by analysis of skewness, which indicates divergence from symmetry. Negative skewness indicates a long-left tail of the distribution or the possibility of larger losses than profits. Positive skewness is a desirable characteristic for risk-averse investors. The motivation of that is based on the expected utility theory.

From this point of view, agricultural ETFs have demonstrated higher positive skewness before shock than after. Traditional assets quite the contrary was demonstrated better skewness post-shock. Kurtosis indicators were growth post-shock for traditional assets and were multidirectional for agricultural ETFs.

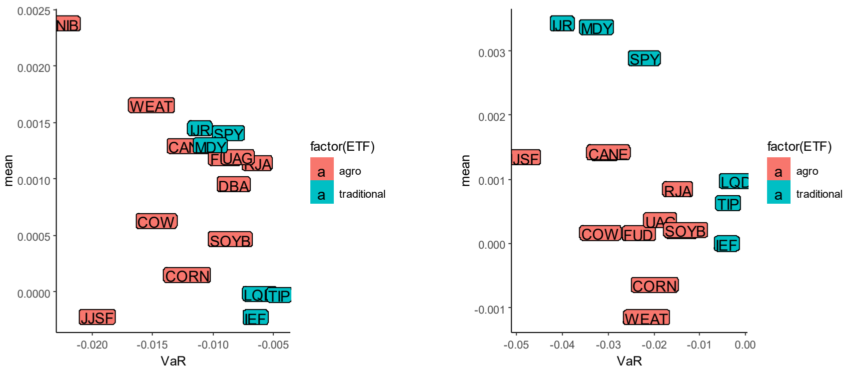
3.5 Risk measurement as losses in a negative situation

This conceptual approach is based on considering measures relating to the interpretation of "negative situation" for the investor. The most popular in this group is Value-at-Risk (VaR), which presents a quantile of the probability distribution function. This quantile corresponding to some level of safety (it maybe 95%, 99%, or 99,9%). The logic of VaR is based on risk covering. If, for example, VaR orients for 95%, then 5% biggest losses will throw off. VaR will cover maximum losses at the framework of 95% possibilities. Risk measure Conditional Value-at-Risk (CVaR) is based on a generalization of VaR. This is the conditional mathematical expectation of losses which higher than VaR (table 2).

Table 2. Risk measurement of ETFs by VaR and CVaR.

ETFs	VaR		CVaR	
	Before shock	Post-shock	Before shock	Post-shock
Agricultural ETFs				
CORN	**-0,012**	-0,020	**-0,017**	-0,028
COW	**-0,015**	-0,032	**-0,020**	-0,049
DBA	**-0,008**	-0,014	**-0,012**	-0,019
FUD	**-0,009**	-0,023	**-0,012**	-0,033
JJSF	**-0,020**	-0,049	**-0,026**	-0,065
NIB	**-0,022**	-0,031	**-0,027**	-0,039
RJA	**-0,006**	-0,015	**-0,008**	-0,022
UAG	**-0,008**	-0,019	**-0,010**	-0,028
CANE	**-0,012**	-0,030	**-0,014**	-0,043
SOYB	**-0,009**	-0,013	**-0,013**	-0,018
WEAT	**-0,015**	-0,022	**-0,018**	-0,024
ETFs of traditional assets				
SPY	**-0,009**	-0,022	-0,013	-0,035
MDY	**-0,010**	-0,033	-0,015	-0,045
IJR	**-0,011**	-0,040	-0,**016**	-0,051
IEF	-0,006	**-0,004**	-0,008	**-0,005**
LQD	-0,006	**-0,002**	**-0,008**	-0,011
TIP	-0,005	**-0,004**	**-0,006**	-0,007

Considering risk measuring for agricultural ETFs we have found that Value-at-Risk and Conditional Value-at-Risk is higher than similar values for traditional assets but not so much. This fact true for both periods prior to and post-shock. Fig. 4 demonstrates the risk-return correspondence between VaR and average returns.



a) 08/28/2019 – 01/15/2020 b) 04/01/2020 – 08/14/2020

Fig. 4. ETFs Value-at-Risk.

Not less excitingly the comparison of changing risk measures values for an approach based on losses in negative situations. In contrast to the results for variability risk measuring here agricultural ETFs indicated higher growth.

It is an interesting conclusion that ratio CVaR/VaR is a good indicator of the distinction of risk. The ratio CVaR/VaR characterizes correspondence between "catastrophic" losses and maximal losses at the frameworks of 95% safety level. This ration became extremely higher for traditional assets than for agricultural ETFs. The changes of CVaR/VaR for agricultural ETFs are negligible in comparison with traditional assets. These values for traditional values had grown 1,6 times on average.

3.6 Risk measurement based on sensitivity approach

Risk measurement at the frameworks of sensitivity analysis provides an opportunity to understand the role of systematic and non-systematic risks. We have chosen for sensitivity analysis SPY as systematic factors. The logic of this choice lies in interpreting the S&P 500 as a leading factor in the stock market. And analysis should provide an answer to the question: How the stock market as a whole affect the return of ETFs? (table 3)

Table 3. Regression analysis.

	SPY beta coefficient		Intercept		R^2		p-value	
	Before shock	Post-shock	Before shock	Post-shock	Before shock	Post-shock	Before shock	Post-shock
Agricultural ETFs								
CORN	0,0214	**0,2134**	0,0014	**0,0030**	0,0010	**0,0266**	**0,7534**	0,1107
COW	**0,0817**	0,0728	0,0014	**0,0029**	**0,0217**	0,0092	0,1498	**0,3487**
DBA	0,0650	**0,6853**	0,0013	**0,0028**	0,0050	**0,1438**	**0,4907**	0,0001
FUD	**0,0718**	0,0045	0,0013	**0,0029**	**0,0061**	0,0000	0,4469	**0,9687**
JJSF	0,0402	**0,3432**	0,0014	**0,0024**	0,0052	**0,4612**	**0,4829**	0,0000
*NIB	-0,0097	**0,2578**	0,0014	**0,0025**	0,0007	**0,1094**	**0,7993**	0,0009
RJA	0,2028	**0,6773**	0,0012	**0,0023**	0,0384	**0,1465**	**0,0545**	0,0001
UAG	0,0138	**0,3302**	0,0014	**0,0028**	0,0002	**0,0682**	**0,8839**	0,0098
CANE	-0,0695	**0,2528**	0,0015	**0,0025**	0,0100	**0,0879**	**0,3295**	0,0032
SOYB	0,0632	**0,8562**	0,0014	**0,0027**	0,0056	**0,1883**	**0,4663**	0,0000
WEAT	0,0866	**0,1820**	0,0013	**0,0031**	**0,0270**	0,0247	0,1075	**0,1242**
ETFs of traditional assets								
MDY	**0,7303**	0,6355	0,0005	**0,0007**	0,7545	**0,8264**	0,0000	0,0000
IJR	**0,5297**	0,5050	0,0006	**0,0012**	0,5524	**0,7376**	0,0000	0,0000
IEF	**-0,6213**	-2,802	0,0013	**0,0029**	0,1538	**0,2170**	**0,0001**	0,0000
LQD	-0,2184	0,9809	0,0014	**0,0019**	0,0176	**0,1990**	**0,1951**	0,0000
TIP	-0,4011	0,**0241**	0,0014	**0,0029**	**0,0348**	0,0000	0,0672	**0,9606**

The main result is very low R-squared indicators. The economic consequence of this is the domination of nonsystematic risks in returns of agro ETFs.

3.7 Correlation analysis

Correlation analysis was provided as inside the sample of agriculture ETF as between traditional assets. It is interesting that agriculture ETFs indicate a very low correlation not only with traditional assets but inside the sample group (table 4). This leads to consideration of portfolio construction directly through agricultural ETFs and through all types of ETFs.

Table 4. Correlation analysis

	Average correlation between sample agriculture ETFs	Average correlation between sample agriculture ETFs and sample of ETFs of traditional assets	Average correlation between sample traditional asset
Before shock	0,31	0,02	0,30
Post-shock	0,33	0,11	0,35

We think that so low correlation can be explained by affecting these ETFs real prices of agricultural products. Not by supply and demand as it appears at the stock market.

4 Conclusion

Risk-return correspondence for different asset classes one of the cornerstones of modern portfolio management. This correspondence together with interdependency analysis allows us to form a portfolio structure that is adequate to given goals and constraints. But “pandemic risk” broke into the investment world and created uncertainty and turmoil. This is a real “black swan” event in terms of Nassim Nicolas Taleb. How much risk investments will involve post-shock? What returns can investors expect? We believe strongly that search answers for these questions will be an actual topic for active research in the nearest future.

Our paper is concentrated on one of such questions. How agricultural commodities expressed by agricultural ETFs pass through COVID-induced shock? How to transform their risk-return correspondence in comparison with traditional assets? The search for the answer was realized through different approaches to risk measurement. First of all was highlighted three time periods: specifically shock period, the quiet period before the shock, and post-shock. It was considered three basic approaches for risk measurement: variability, losses in negative situations, and sensitivity. Correlation analysis also was realized.

Conclusions are the following. Traditional assets (stock indices) demonstrated a higher depth of falling but at the same time higher level of recovery. Indices of bonds not so much falling and then increased in price higher previous level. Agricultural ETFs demonstrated an average level of falling and moderate recovery. The general conclusion lies in increasing risk after shock as for agricultural ETFs as for traditional. It is interesting that risk changing for the first two approaches provides us a discrepancy that is presenting in fig. 5. The variability approach indicated that ranges and standard deviations of traditional asset returns are increased higher. In the meantime, returns of agricultural ETFs demonstrated higher increments in VaR and CVaR. Average returns of agricultural ETFs moved down at the post-shock time but average returns of traditional assets moved up. So, the reaction for shock is different at the frameworks of approaches of risk measuring.

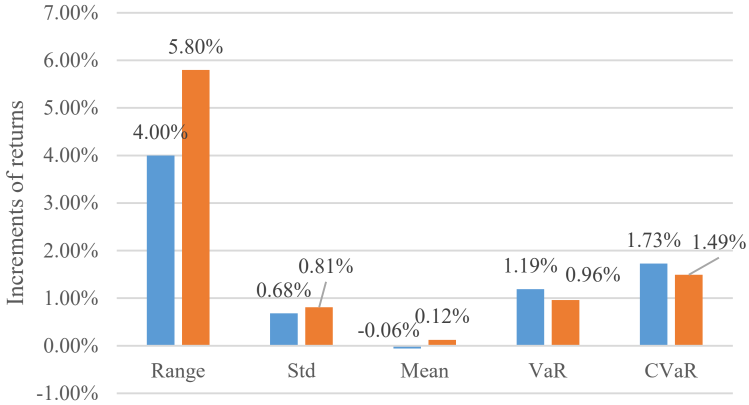


Fig. 5. Growth/falling of risk measures values in absolute increments of returns.

The results of applying sensitivity risk measuring illustrate increasing beta-values to returns of SPY, but R-squared is essentially low as before as after crises. These are confirmed by correlation analysis which shows low correlations. These estimations confirm facts effective diversification between traditional asset classes and alternatives which involved agricultural ETFs.

Summarizing results, it is possible to note differences path of shock and post-shock period for agricultural ETFs and traditional assets.

References

1. Aikman, D., Lehnert, A., Liang, N., Modugno, M.: Financial vulnerabilities, macroeconomic dynamics, and monetary policy. Finance and Economics Discussion Series 2016-055. Board of Governors of the Federal Reserve System, Washington (2016). doi:10.17016/FEDS.2016.055
2. Andreasson, P., Bekiros, S., Nguyen, D.K., Uddin, G.S.: Impact of speculation and economic uncertainty on commodity markets. International review of financial analysis **43**, 115–127 (2016)
3. Artzner, P., Delbaen, F., Eber, J.M., Heath, D.: Coherent measures of risk. Mathematical finance **9**(3), 203–228 (1999)
4. Baker, S., Bloom, N., Davis, S., Terry, S.: COVID-induced economic uncertainty. NBER Working Paper No. 26983. National Bureau of Economic Research, Cambridge (2020). doi:10.3386/w26983
5. Chen, S., Wilson, W., Larsen R., Dahl, B.: Investing in Agriculture as an Asset Class. Agribusiness **31**, 353–371 (2015)

6. Clapp, J.: Responsibility to the rescue? Governing private financial investment in global agriculture. Agriculture and human values **34**(1), 223–235 (2017)
7. Clark, B., Detre, J., D'Antoni, J., Zapata, H.: The role of an agribusiness index in a modern portfolio. Agricultural Finance Review **72**(3), 362–380 (2012)
8. Fischer T.: Risk capital allocation by coherent risk measures based on one-sided moments. Insurance: Mathematics and Economics **32**, 135–146 (2003)
9. Fischer, T., Krauss, C.: Deep learning with long short-term memory networks for financial market predictions. European Journal of Operational Research **270**(2), 654–669 (2018)
10. Hamilton, J., Wu, J.: Effects of index-fund investing on commodity futures prices. International economic review **56**(1), 187–205 (2015)
11. Hau, H., Lai, S.: The role of equity funds in the financial crisis propagation. Review of Finance **21**(1), 77–108 (2017)
12. Holländer, D., Bauer, E., Mrusek, F., Rotermann, B.: European Banking Study 2020: How COVID might affect Europe's banks. https://www.bankinghub.eu/banking/research-markets/european-banking-study-COVID (2020). Accessed 17 Aug 2020
13. Hryhoruk, P., Khrushch, N., Grygoruk, S.: An approach to design a composite index of economic development and identifying the bounds of its levels. Paper presented at the 9th International Conference on Advanced Computer Information Technologies (ACIT), pp. 48–51. IEEE, June 2019. doi:10.1109/ACITT.2019.8779918
14. Jensen, G.R., Mercer, J. M.: Commodities as an Investment. The Research Foundation of CFA Institute Literature Review **6**(2), 1–33 (2011)
15. Johnson, M., Malcolm, B., O'Connor, I.: The Role of Agribusiness Assets in Investment Portfolios. Australasian Agribusiness Review **14**, 11 (2006)
16. Kaminskyi, A., Nehrey, M.: Investment Risk Measurement for Agricultural ETF. Advances in Economics, Business and Management Research **95**, 325–329 (2019)
17. Kaminskyi, A., Versal, N.: Risk management of dollarization in banking: case of post-soviet countries. Montenegrin Journal of Economics **14**(3), 21–40 (2018)
18. Kaminskyi, A.: Financial Risk Modeling. National Taras Shevchenko University, Kyiv (2007)
19. Kuzheliev, M., Zherlitsyn, D., Rekunenko, I., Nechyporenko, A., Nemsadze, F.: The Impact of Inflation Targeting on Macroeconomic Indicators in Ukraine. Banks and Bank Systems **15**(2), 94–104 (2020)
20. Liu, C., Arunkumar, N.: Risk prediction and evaluation of transnational transmission of financial crisis based on complex network. Cluster Computing **22**(2), 4307–4313 (2019)
21. Markowitz, H.: Portfolio Selection: Efficient Diversification of Investment. John Wiley & Sons Ltd, New York (1959)
22. Martin, S., Clapp, J.: Finance for agriculture or agriculture for finance? Journal of Agrarian Change **15**(4), 549–559 (2015)
23. Matviychuk, A.: Fuzzy logic approach to identification and forecasting of financial time series using Elliott wave theory. Fuzzy economic review **11**(2), 51 (2006)
24. Paul, P.: A macroeconomic model with occasional financial crises. Journal of Economic Dynamics and Control **112**, 103830 (2020)
25. Petajisto, A.: Inefficiencies in the pricing of exchange-traded funds. Financial Analysts Journal **73**(1), 24–54 (2017)
26. Racicot, F.É., Théoret, R.: Macroeconomic shocks, forward-looking dynamics, and the behavior of hedge funds. Journal of Banking & Finance **62**, 41–61 (2016)
27. Rasool, A.: Investing in Agribusiness Stocks and Farmland: A Boom or Bust Analysis. Dissertation, Utah State University (2018)

28. Rockafellar, R., Uryasev, S.: Optimization of conditional Value-At-Risk. The Journal of Risk **2**(3), 21–41 (2000)
29. Semerikov, S., Chukharev, S., Sakhno, S., Striuk, A., Osadchyi, V., Solovieva, V., Vakaliuk, T., Nechypurenko, P., Bondarenko, O., Danylchuk, H.: Our sustainable coronavirus future. E3S Web of Conferences **166**, 00001 (2020). doi:10.1051/e3sconf/202016600001
30. Skrypnyk, A., Nehrey, M.: The Formation of the Deposit Portfolio in Macroeconomic Instability. CEUR Workshop Proceedings **1356**, 225–235 (2015)
31. Spelta, A., Flori, A., Pecora, N., Pammolli, F.: Financial crises: uncovering self-organized patterns and predicting stock markets instability. Journal of Business Research (2019, in press). doi:10.1016/j.jbusres.2019.10.043
32. Stoll, H., Whaley, R.: Commodity index investing and commodity futures prices. Journal of Applied Finance **20**(1) (2010)
33. Szegö, G.P.: Risk measures for the 21st century, vol. 1. Wiley, New York (2004)
34. Tiwari, A.K., Albulescu, C.T., Yoon, S.M.: A multifractal detrended fluctuation analysis of financial market efficiency: Comparison using Dow Jones sector ETF indices. Physica A: Statistical Mechanics and its Applications **483**, 182–192 (2017)
35. Uthayakumar, J., Metawa, N., Shankar, K., Lakshmanaprabu, S.K.: Financial crisis prediction model using ant colony optimization. International Journal of Information Management **50**, 538–556 (2020)
36. Valdes, R.: What drives the regional integration of agribusiness stocks? Evidence in worldwide perspective. Paper presented at the 2017 Annual Meeting, Agricultural and Applied Economics Association, Chicago, Illinois, July 30-August 1 (2017). doi:10.22004/ag.econ.258265

Comparative analysis of the attractiveness of investment instruments based on the analysis of market dynamics

Nataliia Maksyshko[0000-0002-0473-7195], Oksana Vasylieva[0000-0002-6332-2707], Igor Kozin[0000-0003-1278-8520] and Vitalii Perepelitsa[0000-0003-2260-5089]

Zaporizhzhia National University, 66 Zhukovskogo Str., Zaporizhzhia, 69600, Ukraine
maxishko@ukr.net, oksanabay@ukr.net, ainc00@gmail.com, vitalijperepelica2@gmail.com

Abstract. The article continues the authors' research on solving the problem of choosing the most attractive investment instrument from a variety of alternatives, based on a comparative analysis of the dynamics for the respective markets. The nature of the dynamics affects the predictability level of the investor's income and is determined by finding out which hypothesis corresponds to the dynamics: the efficient market hypothesis, the fractal market hypothesis and the coherent market hypothesis. The methodology of comparative analysis developed by the authors is based on the use of statistical analysis methods combined with the methods of complex fractal analysis. It makes it possible to reveal the presence of deterministic chaos in the dynamics and to obtain estimates of the long-term memory in time series. The calculated characteristics of the fuzzy set of the memory depth for time series make it possible to draw conclusions about the financial instruments preference for the investor. The methodology developed by the authors is applied to three markets. A comparative analysis of three instruments (gold, EUR/USD currency pair and Bitcoin cryptocurrency) was carried out. The dynamics of prices and profitability for financial instruments in the conditions before the onset of the COVID-19 crisis and during it is considered.

Keywords: gold market, EUR/USD, Bitcoin, statistical analysis, fractal analysis, rescaled range analysis, memory depth, COVID-19.

1 Introduction

Nowadays investors are faced with a wealth of information and investment opportunities. However, the availability of access to global financial markets carries both additional opportunities for profit and new risks. All this requires the development of new modern and effective approaches to assess the investment attractiveness of markets.

The dynamics of investment markets is formed under the influence of many external and internal factors. To understand and explain the nature of this dynamics, scientists have developed and put forward several hypotheses. The most famous of them are: Efficient market hypothesis (EMH) [6; 7], Fractal market hypothesis (FMH) [18] and

Coherent market hypothesis (CMH) [22]. Often, the premises of one hypothesis run counter to the premises of another. For example, the efficient market hypothesis states that prices follow a random walk and the previous price is not related to the subsequent price. To study the effective market, the use of statistical analysis tools is envisaged. Conversely, the presence of fractal dynamics involves long-term memory of the time series and, on the basis of this, the ability to predict the behavior of the system. The Fractal Market Hypothesis involves the use of nonlinear dynamics methods.

In fact, empirical studies show that the nature of the dynamics for a particular market cannot fully satisfy either the EMH requirements or the FMH prerequisites for a long period of time. The characteristics of the dynamics can change over time, which is suggested, for example, by the Coherent Market Hypothesis. Real markets always contain both elements of randomnicity and determinism. Therefore, to understand and evaluate each market, it is important to use all the existing tools of both statistical analysis and nonlinear dynamics methods.

So, development and improvement of the comparative analysis technique of investment markets dynamics in the context of existing market hypotheses, assessment of investment prospects, and developing recommendations on the benefits of investing for different planning horizons are extremely urgent and important tasks. This problem is of particular interest in the conditions for the emergence and extension of COVID-19 crisis.

2 Related work

Each of the aforementioned hypotheses implies appropriate prerequisites for the dynamics of investment markets and uses special methods of diagnostics and analysis.

Consider each of the hypotheses in more detail.

The basis of EMH is the following preconditions [6; 7]: all the information is equally accessible and can be immediately taken into account by the market at a fair price, future prices depend only on the new information, future prices are not related to the previous ones, the impact of the information is linear, market participants are rational and homogeneous (they are equally not risk-averse and have the same investment horizons). Within this hypothesis, linear models, probabilistic calculations and statistical analysis are used.

Main characteristics of a fractal market are [18]: the main thing in the market is not a fair price, but liquidity, prices have a memory of previous values, locally the market is random, but globally – determined, but the dynamics of the market is nonlinear, investors differ in investment horizons. For diagnostics and analysis, nonlinear models, fractal mathematics and chaos theory tools are used.

The Coherent market hypothesis combines the two previous hypotheses and represents a nonlinear statistical model. According to CMH, markets go through four phases: random walk, unstable transition, chaos, and coherence [22].

All the above mentioned hypotheses have been arisen and developed in studies of stock markets [6; 7; 18; 22]. However, the local stock markets assessment of the correspondence to existing hypotheses remains relevant today [1; 2; 3; 8; 9; 17; 16; 23].

But now the scope of their application has considerably expanded, along with stock markets, dynamics analysis is actively conducted for the currency markets [4; 5; 11], deposits [12] and cryptocurrency [13].

Some works are directly related to the use of statistical (for proving the EMH) [1; 2; 3; 4; 9; 17] or fractal (for proving the FHM) [5; 16] analysis tools. Other studies are devoted to a specific analysis tool, for example, Fourier Unit Root Test [8] or Hurst exponent [13]. It should be noted that more interesting relevant and modern are the studies of nonlinear characteristics of dynamics [20; 21].

However, in our opinion, it is important to develop a comprehensive approach to assessing market dynamics, which allows the use of the best diagnostic tools from both statistical and fractal methods. Economic time series often do not represent a classical model of one theory, but combine both stochastic and fractal components. Therefore, the application of different methods to assess the dynamics reveals the patterns of development of economic series from different points of view. The basics of an integrated approach have been outlined in previous works by the authors [14]. This paper uses the main Investment market comparative analysis technique steps from the work [14]. However, the comparison criteria were revised and supplemented. Some criteria have been removed due to their low level of informativeness, some have been added. A distinction was also made as to which criteria should be applied to the time series of prices, which to the time series of profitability, and which could be applied to both types of time series.

This study is especially relevant in connection with the sharp changes in the dynamics of economic systems that are currently occurring during the crisis COVID-19.

It determines the importance of a comprehensive analysis of the investment markets dynamics to highlight their crucial characteristics. These characteristics can be used to compare and select the most attractive investment instruments.

3 Materials and methods

The paper considers the following investment instruments: precious metals market, Forex currency market and cryptocurrency market. All of them are both high-tech and affordable, which means the presence of information technologies and applications for a wide range of individual investors.

The objects of comparative analysis are time series (TS) the daily values of the prices and the profitability of the gold, the currency pair EUR/USD and the Bitcoin for the period from August 2019 to July 2020.

In recent times, humanity has had to face new challenges: the COVID-19 pandemic has made its adjustments in almost every area of life [10; 19]. Unseen before quarantine measures and border closures have shattered logistics chains and dealt a significant blow to the global economy. Unprecedented measures implemented by the governments of many countries of the world have created a new economic reality, with new laws and vectors of development, preconditions and influenced the change in the nature of dynamics in financial markets. In this regard, the hypothesis of a change in the nature of the dynamics of financial instruments arose and the need to take these changes into account when assessing their investment attractiveness. The dynamics of selected investment markets in the form of time series (TS) for two periods is considered: before the beginning of the crisis (from 01.08.2019 to 31.01.2020) and from the conditional beginning of the crisis (from 01.02.2020 to 31.07.2020). This division is due to objective conditions and allows to test the proposed comparison tools under different conditions, tto gain new knowledge about the objects of study: how different markets reacted to changes in external factors, how external factors influenced the dynamics, what new features and properties acquired market dynamics in a crisis situation.

The general scheme of methods used in the study is presented in table 1.

Table 1. Investment market comparative analysis technique steps.

Step	**For TS of the prices**	**For TS of the profitability**
1.	**Visualization of dynamics**	
	Research on the presence and type of trends	-
2.	**Statistical analysis**	
2.1.	Basic numerical characteristics estimation:	
	Coefficient of variation, coefficient of oscillation	Mean, median, standard deviation, skewness, kurtosis
2.2.	Normal distribution tests: • verification for the equality of mean and median; • checking for the matching of the skewness; • checking for the matching of the kurtosis; • Kolmogorov-Smirnov Test; • Shapiro-Wilk test.	
2.3.	Other statistical tests:	
	Breusch-Godfrey serial correlation LM-test, Runs test	–
3.	**Complex fractal analysis**	
3.1.	Deterministic chaos diagnosis:	
	Drifting attractor test, Gilmore's graphic test, construction of pseudo-phase space	
3.2.	Rescaled range analysis (R/S analysis) if Hurst exponent	
	$H \le 0.8$	$H > 0.8$
3.2*.	Construction of delayed (profitability) TS	
3.3.	Method of sequential R/S-analysis:	
	Construction of a fuzzy set of memory depth, calculation of its characteristics: l_{ms}, $l_{\max}$, l_{cg}, H_{entr_L}, $SH(L)$	

In sub-step 2.1 of step 2 basic numerical statistical characteristics estimation selected characteristics that are appropriate to use to compare the TS of each species. Thus, to compare the dynamics of prices, the relative coefficients of variation and oscillations are chosen, and for TS profitability, indicators can be used that are measured in both absolute and relative values (mean, median, standard deviation, skewness, kurtosis).

In step 2.2 of table 1 the compliance with the normal distribution is checked. According to EMH, investment instruments prices already take into account past information, therefore the next price change is influenced only with the new information [7]. Hence, all occurring on the market changes are not related events. It follows from the central limit theorem that the distribution of a large number of random independent variables converges to normal distribution. So, by checking the hypothesis of normal distribution (step 2.2 of figure 1), the hypothesis of an effective market is checked.

The normality of the distribution is analyzed with:

- verification for the equality of mean and median;
- checking for the matching of the skewness;
- checking for the matching of the kurtosis;
- Kolmogorov-Smirnov Test;
- Shapiro-Wilk test.

In addition, there are various statistical tests checking the availability of some characteristics of the weak form of efficient market, such as the independence between events, the stationarity of a time series, the random nature of price changes or the study of variances. In the previous authors' works to check the series for random character of changes was made: constructing regression equations and checking them for statistical significance; checking for auto-correlation; Durbin-Watson Test, Breusch-Godfrey serial correlation LM-test, Augmented Dickey-Fuller Unit Root test (ADF test) and Runs test. In this study, tests were selected that confirmed their effectiveness and proved to be the most informative and indicative criteria for comparative analysis.

In step 2.3, the following tests were carried out and the following methods were applied only to the time series of prices: Breusch-Godfrey serial correlation LM-test, Runs test.

At the first sub-stage of the Complex fractal analysis stage we perform deterministic chaos diagnosis, namely Drifting attractor test, Gilmore's graphic test, construction of pseudo-phase space to the time series of price and profitability.

In step 3.2, the rescaled range analysis (R/S-analysis) is applied and the Hurst exponent is calculated, it allows to determine the presence of memory (persistence) in a TS [18]. The Hurst exponent is a measure to determine the nature of the dynamics of the series: the randomness of events in the series (at H=0.5), the persistence of the series and the presence of memory (at H approaching 1), or antipersistent (at H<0.5). The Hurst exponent is a measure of the trend stability of a series and allows us to determine whether the nature of the dynamics is stochastic or fractal.

In the case when the value of the Hurst exponent indicates the persistence of the series ($H \geq 0.8$), we proceed to step 3.3.

If the Hurst exponent does not show a sufficient level of persistence (H <0.8), then proceed to step 3.2*.

If the Hurst exponent indicates memory availability, then in step 3.3 we use the method of sequential R/S-analysis [15].

The determination of the Hurst exponent is based on the application of the method of the normalized Hurst range and the construction of the R/S trajectory. If TS is characterized by long-term memory, then a number of starting points of the obtained R/S trajectory of this TS form a clear linear trend. At some value of $k = k_0$ R/S-trajectory changes its slope quite sharply, that is, at the point (x_{k0}, y_{k0}) the trajectory receives a significant negative gain in absolute terms – there is a break in the trend and there is no return to the previous trend. It is assumed that at the point k_0 the effect of long-term memory dissipates. In this case, the breakdown of the trend demonstrates the loss of memory of the initial conditions, and also signals (possibly with a lag, i.e. with some delay) the exhaustion of the cycle or quasi-cycle contained in the initial segment of this TS. But, as is known [15], the method of normalized Hurst scope (standard R/S-analysis) provides only the average characteristic of the inertia property (trend resistance) for TS as a whole and does not take into account the changing nature of the dynamics of the indicator.

To overcome this shortcoming, a modified method of fractal analysis was developed [15] the method of sequential R/S analysis (step 3.4). The essence of the method is to sequentially construct a modified computational scheme of R/S-trajectory for the family TS, which are a subset of this TS, but consistently differ from the starting point. The advantage of this method is its ability to take into account the changing nature of the dynamics, to identify the set of cycles (quasi-cycles) that are characteristic of the TS under study, as well as to calculate the lower estimate of memory depth (about the beginning of this TS). The difference in the conditions of application of the method is the absence of significant restrictions on the length of TS.

The result of applying the method of sequential R/S-analysis is to determine not one breakpoint from the trend k_0, but a set of breakpoints from the trend of the family of R/S-trajectories, which reflect the memory loss time of the initial conditions (beginning of the corresponding TS). This allows you to generate a fuzzy set of TS memory depth. Estimating memory depth for a range reflects the uncertainty generated by external and internal influences on the economic system.

The fuzzy set of memory depth for the TS as a whole (is denoted by $L(i)$) has the form

$$L(i) = \{(l, \mu(l)), l \in L^0\}, \quad (1)$$

where $L(i)$ is a fuzzy set of memory depth for TS i, l is the value of the sequence number of the trend change point for TS, and supp$L(i)=L^0 = \{l \in \mathrm{N}, \mu(l)>0\}$.

Based on the analysis of the values of the membership function $\mu(l)$, we can identify the so-called characteristic or significant degrees of membership (for example, $\mu(l)>0.3$), which can be considered uncharacteristic. Restriction on the degree of significance (denote it ε), ie the condition $\mu(l)> \varepsilon$, can be set by an expert.

Based on the analysis of the values of the membership function $\mu(l)$, we can identify the so-called characteristic or significant degrees of membership (for example, $\mu(l)>0.3$), which can be considered uncharacteristic.

The values of memory depth l, which correspond to the values of the membership function $\mu(l)>\varepsilon$, are called ε-valuable [15].

Using the defasification procedure for the selected significant degrees $\mu(l)$, and, if necessary, rounding the calculated value to the nearest whole, we calculate the center of gravity (or gravity) of the set of ε-significant values of memory depth by the formula

$$l_{cg}=[(\Sigma l\cdot\mu(l))/(\Sigma\cdot\mu(l))]. \quad (2)$$

Thus, the obtained predictive information is that the considered TS is characterized by the property of trend resistance over a period of time l_{cg} on average. Depending on the value of l_{cg}, the latter statement in the context of pre-forecast analysis means good preconditions for building a sufficiently reliable forecast of this TS within the forecast horizon l_{cg}.

Recommendations for the forecast horizon (denote it h) can be clarified by using another characteristic of the fuzzy set $L(i)$ of memory depth for TS as a whole – the value of memory depth (denote it l_{ms} (the most significant), which has the largest the value of the membership function $\mu(l)$ of the depth l of the fuzzy set $L(i)$:

$$\mu(l_{ms})= \max\ (\mu(l)). \quad (3)$$

Satisfactory prediction accuracy is then provided when the prediction horizon does not exceed the center of gravity l_{cg} and the most common memory depth value – l_{ms}.

To estimate the property of dynamics uncertainty, the information entropy index of the fuzzy set of memory depth $L(i)$ (H_{entr_L}) is used with respect to the variety of behavior variants of a series of dynamics. It is calculated by the formula:

$$H_{entr\_L}= -\Sigma\ (\mu(l)\cdot log\ \mu(l)). \quad (4)$$

As discussed above, when analyzing the dynamics, it is advisable to analyze not only the entire fuzzy set of memory depth $L(i)$, but also its subset of ε-significant depths, i.e. the set $L^{\varepsilon}(i)$. This somewhat reduces the uncertainty that can be estimated by the value of information entropy by neglecting the values of depths that are not ε-significant, i.e. the indicator

$$H^{\varepsilon}{}_{entr\_L}= -\ \Sigma\ (\mu(l)\cdot log\ \mu(l)),\ \mu(l) > \varepsilon. \quad (5)$$

The redundancy index of the fuzzy memory depth $L(i)$ is also used to characterize TS as a measure of TS noise. It is calculated by the following formula:

$$SH(i)=1-(H^{\varepsilon}{}_{entr\_L}\backslash\ H_{entr\_L}). \quad (6)$$

On the basis of the given numerical characteristics concerning depth of memory of all TS as a whole, it is possible to carry out the comparative analysis of dynamics of the considered TS.

In step 3.2* for the profitability time series with the Hurst exponent close to 0.5, we construct the time series of the delayed profitability by the formula:

$$p^s = \frac{(v_{(t+s)} - v_t)}{v_t} * 100\%, \tag{7}$$

where $v(t)$ – the price of the investment instrument at a day t; s – is a lag value. Then profitability time series is equal to: $P^s(i)=(p^s(i))$, $i \in \{Z, F, B\}$, where Z – TS of gold; F – TS of currency pair EUR/USD, B – TS of cryptocurrency Bitcoin.

According to performed calculations, profitability TS may not have memory (Hurst exponent is close to 0.5), however, with the growth of the time lag, time series of delayed profitability become persistent.

For each of those time series, we calculated the Hurst exponent, until we determine the value of s at which the time series acquires memory (persistence).

4 Results

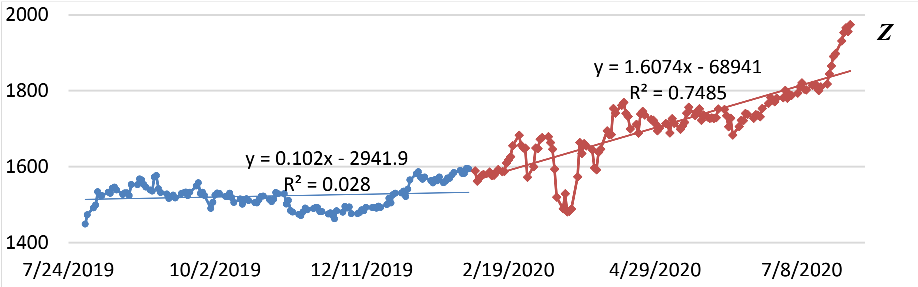
For the comparative analysis, three investment objects have been selected: gold (Z), currency pair EUR/USD (F), and cryptocurrency Bitcoin (B).

The steps and methods described in the previous section were applied to daily prices time series of the gold, the currency pair and the Bitcoin for the period from August 2019 to July 2020. For the comparative analysis of dynamics and studying of the market's reaction character to the changes which have occurred in the markets, time series are divided into two periods: before the beginning of the COVID-19 pandemic (from 01.08.2019 to 01.02.2020) and during its spread (from 01.02.2020 to 01.08.2020).

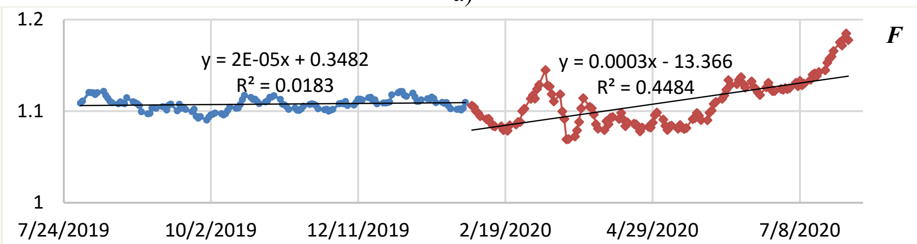
4.1 Application of the comparative analysis methodology to the time series of the prices for the investment instruments

Step 1. Graphical representations of price dynamics are shown in the fig. 1. The figure shows the identified trends in the form of linear trends for each TS. The presence of a significant linear trend was detected only for the time series of gold prices during the pandemic (R^2=0.75). The coefficient of determination greater than 0.5 had Bitcoin before the crisis with a downward linear trend. The linear trend before the crisis for gold and EUR was insignificantly growing, for bitcoin – markedly downward. But with the crisis onset, the direction of price movements became increasing for all three instruments. This is partly due to the significant dollar emission, which occurs as a reaction of the Federal Reserve System to the crisis in the US economy. However, the numerical characteristics of trends are not comparable and cannot serve as indicators for their comparison.

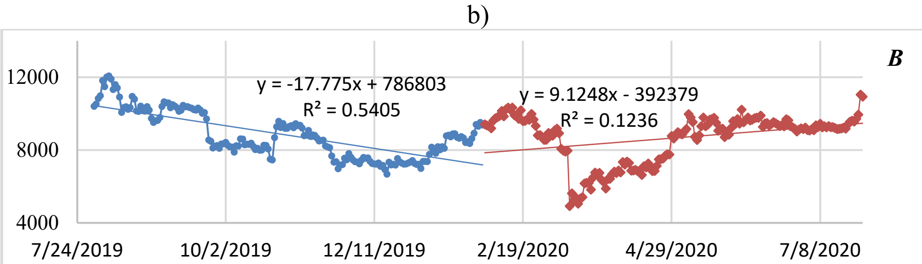
Step 2. At this stage the statistical analysis of time series is performing. Given that the prices of selected investment instruments have different units of measurement, we use relative statistics such as coefficients of variation and oscillations for comparison (table 2).



a)



b)



c)

Fig. 1. Price dynamics for the period from August 2019 to July 2020: a) gold (Z); b) currency pair EUR/USD (F); c) Bitcoin (B).

Table 2. Statistical characteristics of investment instruments prices.

Statistical characteristics	TS								
	All			Before			During		
	P(Z)	P(V)	P(B)	P(Z)	P(V)	P(B)	P(Z)	P(V)	P(B)
Variation coefficient	0.074	0.017	0.152	0.022	0.006	0.146	0.055	0.023	0.158
Oscillations coefficient	0.325	0.104	0.816	0.105	0.028	0.612	0.283	0.104	0.740

Analysis of table 2 shows that the dynamics of each of the three instruments changes significantly with the crisis onset. This confirms the assumption that time series have changed the nature of the dynamics and it is necessary to consider two series of data for each instrument: before the coronavirus pandemic start and during its spread. The largest changes in the dynamics are observed in the market of the currency pair EUR/USD: after a relative calm, quarantine measures were reflected in increased volatility and rapid changes in the direction of price movements. The coefficients of variation and oscillations have increased several times.

Against the background of a noticeable upward trend, the time series of gold prices during the crisis also significantly increased volatility. The smallest increase in these coefficients occurred in the Bitcoin market. However, it should be noted that Bitcoin before the crisis showed strong volatility in contrast to, for example, the TS of currency pair EUR/USD (the coefficient of variation for the currency in the pre-crisis and post-crisis period was 0.006 and 0.023, respectively, compared to 0.146 and 0.158 for cryptocurrency).

The Breusch-Godfrey test (on the correlation between the price values of 1-10 order) is conducted to establish the relationship between the events in the time series. The test results are presented in the table 3.

Table 3. The Breusch-Godfrey test results for TS of prices.

	Gold	EUR/USD	Bitcoin
Before	-	-	Order 7 (p-value = 0.049)
During	-	Order 8 (P-value=0.03) Order 9 (P-value=0.03) Order 10 (P-value=0.04)	Order 1 (P-value=0.047) Order 2 (P-value=0.04) Order 4 (P-value=0.049)

According to table 3, we can say that for TS gold in the two studied periods and EUR/USD in the pre-crisis period, the null hypothesis of no autocorrelation was confirmed. For TS EUR/USD during the pandemic period and for Bitcoin in the two studied periods at a significance level of 0.05, a correlation of certain orders is possible.

Step 3. We turn to the study of the financial instruments dynamics and its comparison by methods of nonlinear dynamics. The table 4 shows the calculated values of the Hurst exponent (H), which determine the level of persistence for the time series of the investment instrument prices and for time series of mixed price values.

Table 4. The value of Hurst exponent (H) for TS of prices.

Hurst exponent	TS								
	All			Before			During		
	P(Z)	P(V)	P(B)	P(Z)	P(V)	P(B)	P(Z)	P(V)	P(B)
H	0.889	0.865	0.917	0.902	0.894	0.925	0.917	0.876	0.938
H mixed	0.538	0.612	0.564	0.601	0.541	0.532	0.623	0.544	0.603

For all series, the Hurst exponent is in the range [8.6; 9.4], from which we can conclude that all series have a memory of previous values. But table 4 shows that the nature of the dynamics during the crisis period is changing. It should be noted that for all financial instruments, the Hurst exponent for the separate (pre-crisis or pandemic) period is higher than the Hurst exponent for the entire (united) period.

The Hurst exponent acquires the highest values for Bitcoin time series in contrast to the lowest H values for series of the currency pair EUR/USD. This means that the dynamics of Bitcoin is described by the laws of nonlinear dynamics, and the influence of randomness on the price formation is small. Recall that H = 1 means a completely deterministic series. Analysis of table 4 leads to the conclusion that all TSs are persistent and have memory. Therefore, we pass to execution of step 3.3 – application of a method of the sequential R/S-analysis.

Let's construct a fuzzy set of memory depth and consider its characteristics.

The fig. 2 shows the fuzzy sets $L(i)$ of memory depth for each TS $i \in \{Z, F, B\}$).

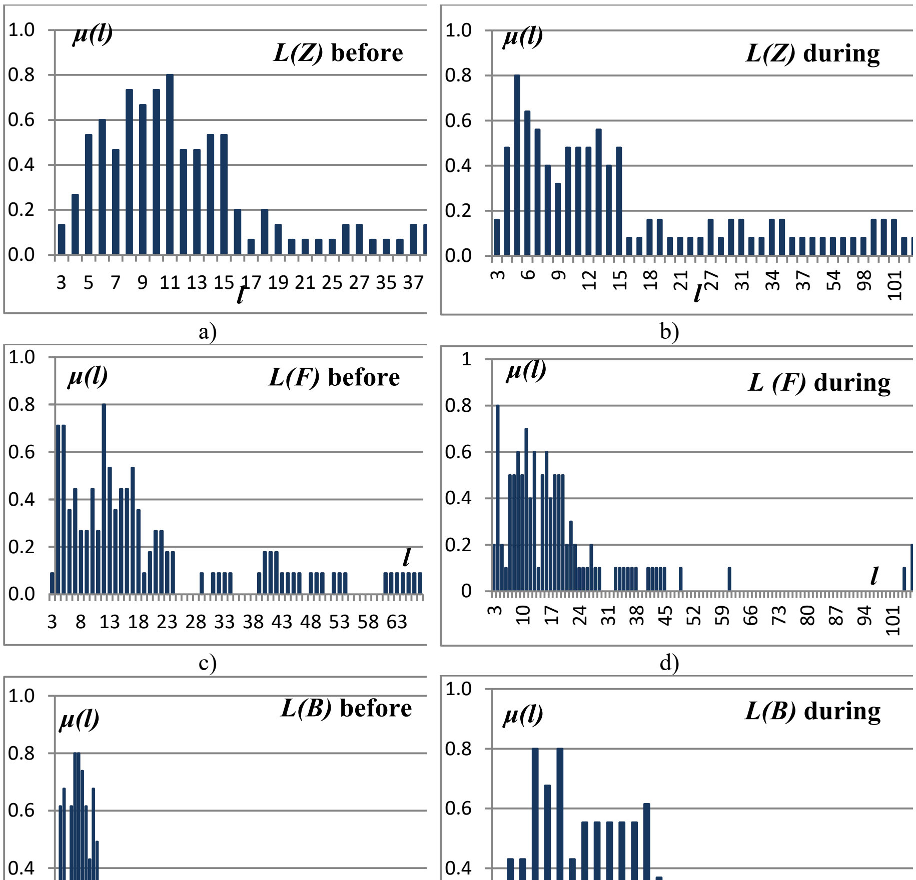


Fig. 2. The fuzzy set of memory depths $L(i)$ for the TS of: a) gold (Z) before; b) gold (Z) during; c) EUR/USD (F) before; d) EUR/USD (F) during; e) Bitcoin (B) before; f) Bitcoin (B) during.

We calculate and compare the characteristics of the depth of memory inherent in the time series of investment instruments, calculated on the basis of their fuzzy sets. The main numerical characteristics of the fuzzy set of memory depth are given in the table 5.

Table 5. The main numerical characteristics of the fuzzy set of memory depth for TS of prices.

Characteristic	**TS**					
	Gold (*Z*)		**EUR/USD (*F*)**		**Bitcoin (*B*)**	
	before	**during**	**before**	**during**	**before**	**during**
l_{max}	38	104	67	108	104	50
l_{ms}	11	5	12	4	8.9	6.8
l_{cg}	12.5	22.3	20.6	20.9	22.4	13.7
H_{entr_L}	10.6	15.9	18.6	17.4	18.2	13.5
Significance ε = 0.3						
l_{max}	15	15	22	22	14	19
l_{ms}	11	11	12	4	8.9	6.8
l_{cg}	9.7	9.2	11.5	12.7	9.1	10.5
H_{entr_L}	5.2	5.7	8.1	7.0	3.9	7.9
$SH(L)$	0.5	0.6	0.6	0.6	0.8	0.4

Consider the characteristics of a fuzzy set of memory depths in the pre-crisis period. From the point of view of reducing uncertainty, the best value of the maximum memory depth l_{max} has gold (the choice from the set is limited by the memory depth 38), this value is relatively good for gold also at ε-significance 0.3. The relatively small value of the center of gravity for the time series of gold, on the one hand, limits the possible forecast horizon, and on the other hand is a consequence of low variability. For the set $L^{\varepsilon}(i)$, the gravity center of gold becomes comparable to that of Bitcoin. The noise level for TS of the gold is the lowest. Despite the fact that the entropy index for the set $L^{\varepsilon}(i)$ is better in TS Bitcoin, we believe that the most stable and predictable in the pre-crisis period is TS of gold.

Given the notable reduction of the fuzzy set at the level of significance $\varepsilon = 0.3$ and the low entropy of Bitcoin with insignificant differences in other indicators, we consider the Bitcoin series to be more stable and attractive for investment than currency.

The introduction of quarantine measures had a negative effect on the most significant memory depth lms: it fell in all series except the gold $L^{\varepsilon}(Z)$ (l_{ms}(before) = l_{ms}(during) = 11). For gold and EUR, the variability and uncertainty of the set $L(i)$ increased significantly (l_{max} from 38 before the crisis to 104 after for gold, l_{max} from 67 before the crisis to 108 during for EUR). Conversely, for Bitcoin these indicators have improved (from 104 to 50). Given the best indicators l_{max}, l_{ms} and H_{entr_L} of the set $L^{\varepsilon}(i)$ for TS of gold, we believe that even after the crisis, this financial instrument remains the most attractive.

4.2 Application of the comparative analysis methodology to the time series of profitability for the investment instruments

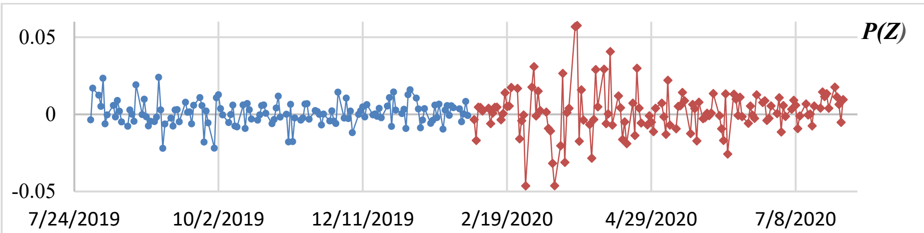
Another important section for studying the market dynamics is the time series of profitability for financial instruments. Profitability is a crucial indicator for every investor. In addition, time series of profitability are indispensable in comparative analysis, as they are not absolute but relative values of the price. Calculate the TS of profitability by this formula for each time series of the price.

$$P(i) = p_t(i),$$

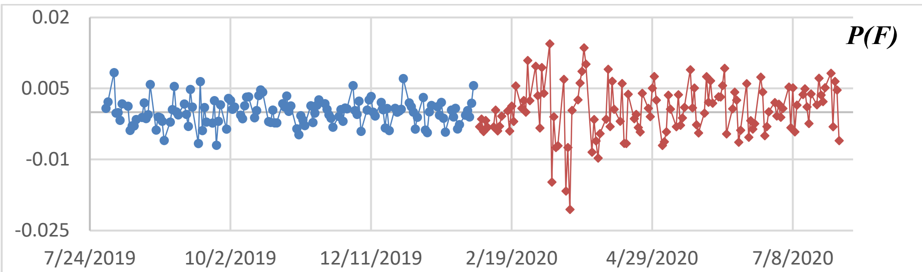
$$p_t(i) = \frac{(v_t\,(i) - v_{(t-1)}(i))}{v_{(t-1)}(i)} * 100\%,$$

where $v_t(i)$ – the price of the investment instrument at a day t, $i \in \{Z, F, B\}$.

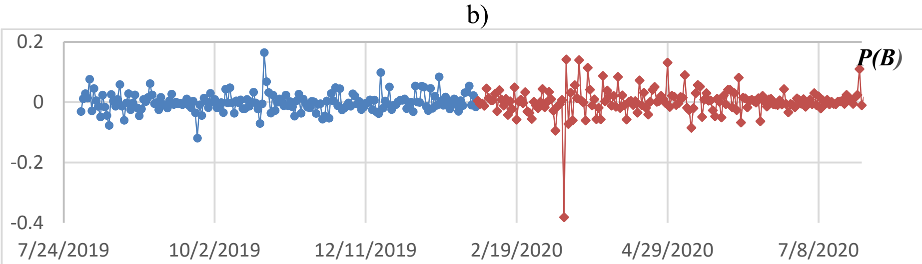
Graphic representation of the obtained series of profitability is shown in the fig. 3.



a)



b)



c)

Fig. 3. Dynamics of profitability time series for a) gold; b) EUR/USD; c) Bitcoin.

The mean values of profitability time series are close to zero, the series differ significantly from each other in terms of variation (standard deviation and range). Since time series of profitability are similar in appearance to a series of random variables, we check them for the normal distribution law according to the five criteria defined in Section 3. The results of the calculation are shown in the table 6.

Table 6. The results of checking the series on the normal distribution law.

Criteria	**TS**					
	Before			**During**		
	P(Z)	***P(V)***	***P(B)***	***P(Z)***	***P(V)***	***P(B)***
1. Verification for the equality of the meanand the median	+	+	+	+	+	+
2. Checking of the skewness	+	+	–	+	+	–
3. Checking of the kurtosis	+	+	–	–	–	–
4. Kolmogorov-Smirnov test	+	+	–	+	+	–
5. Shapiro-Wilk test	p=0.026	+	–	–	p=0.037	–
Mean	0.00072	0.00002	-0.00025	0.00173	0.00047	0.00225
Standard deviation	0.00784	0.00275	0.03158	0.01505	0.00563	0.04638
Range	0.04608	0.01536	0.28352	0.10396	0.03496	0.52353

+ a null hypothesis that a normal distribution not disproved;

– the null hypothesis of a normal distribution disproved.

The time series of profitability of gold and the currency pair EUR / USD before the crisis had the features of randomly distributed values according to all five criteria. After the beginning of the crisis there was an increase kurtosis, Shapiro-Wilk test also showed negative results (for the currency at a significance level of α = 0.01). Bitcoin profitability did not meet the requirements of normally distributed values before or after quarantine measures (except for the mean and median).

The calculation of the Hurst exponent also shows the lack of memory and the random nature of changes in the profitability of the series (H values are in the range [0.57; 0.66]). In this connection, a family of profitability time series with a certain lag was constructed and investigated [18].

The time series of the "delayed" profitability are constructed by the formula (7).

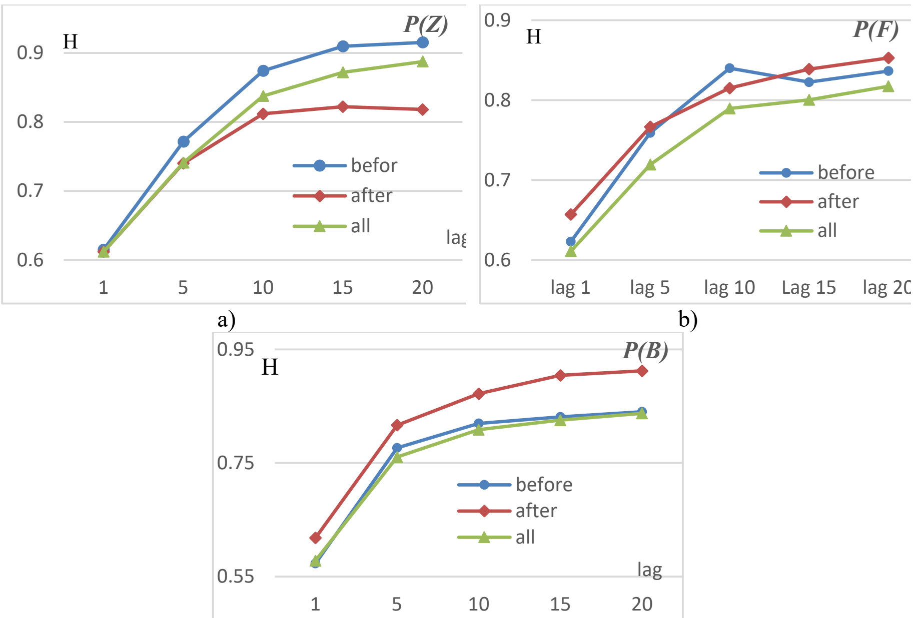
The character of the dynamics of profitability varies depending on the magnitude of the time lag and, as it grows, the time series acquire the properties of persistence (the property of memory). The Hurst exponents for the profitability time series depending on the value of lag is presented in the table 7.

A visual representation of the change in the value of the Hurst exponent for each of the financial instruments is presented in the fig. 4.

The persistence of the united time series of profitability (including data both before the crisis and during quarantine measures) is usually less than the persistence of the divided series. This indicates the different nature of the dynamics of the series before and after the introduction of quarantine measures. Delayed time series of gold profitability in the pre-crisis period acquire persistence faster than the corresponding TS after the introduction of quarantine measures (fig. 4a)). For time series of profitability EUR and Bitcoin the opposite is true (fig. 4b and 4c).

Table 7. The Hurst exponents for the profitability time series depending on the value of lag.

TS	Period	lag 1	lag 5	lag 10	lag 15	lag 20
	before	0.615	0.771	**0.874**	**0.910**	**0.915**
$P(Z)$	after	0.612	0.740	**0.812**	**0.822**	**0.818**
	all	0.612	0.741	**0.837**	**0.872**	**0.888**
	before	0.623	0.759	**0.840**	**0.822**	**0.836**
$P(F)$	after	0.657	0.767	**0.815**	**0.839**	**0.853**
	all	0.611	0.719	**0.789**	**0.800**	**0.817**
	before	0.573	0.776	**0.819**	**0.831**	**0.84**
$P(B)$	after	0.618	**0.816**	**0.872**	**0.904**	**0.912**
	all	0.578	0.760	**0.808**	**0.825**	**0.837**



a) b) c)

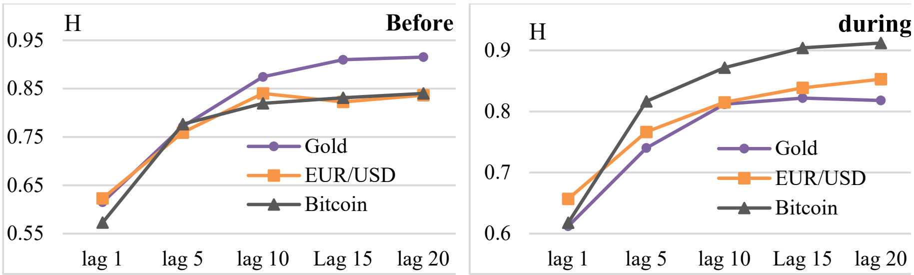
Fig. 4. Hurst exponent depending on the value of lag for: a) $P(Z)$; b) $P(F)$; c) $P(B)$.

A visual representation of the change in the value of the Hurst exponent for two periods is presented in the fig. 5.

In the pre-crisis period, the leaderin the speed of gaining persistence of profitability with increasing lag was gold, after the crisis - Bitcoin.

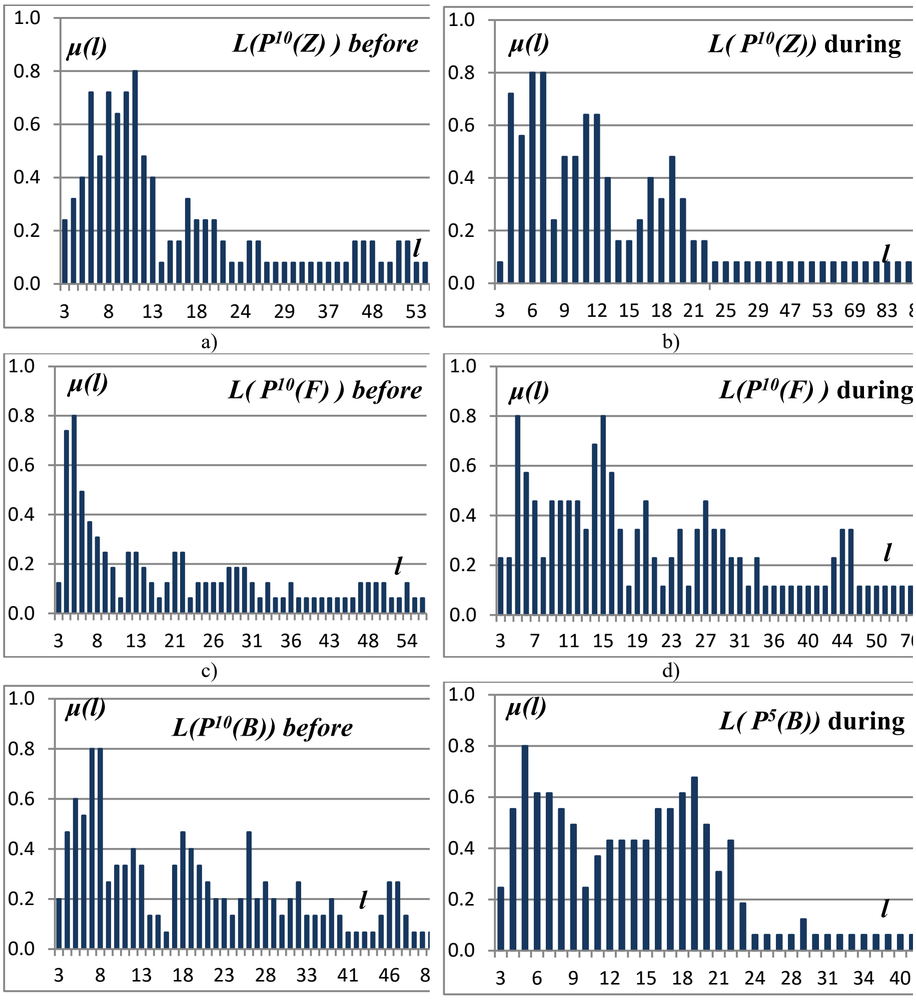
For the received persistent time series, we carry out their system characteristics in their structure of deterministic chaos.

The fig. 6 shows the fuzzy set $L(i)$, $i \in \{Z, F, B\}$ of memory depth for time series of profitability that have memory. We assume that time series have memory, with the Hurst exponent greater than 0.8 (H>0.8).



a) b)

Fig. 5. Hurst exponent depending on the value of the period: a) before crisis; b) after crisis start.



a) b) c) d) e) f)

Fig. 6. The fuzzy set of memory depths $L(i)$ for the TS of i: a) $P^{10}(Z)$ before; b) $P^{10}(Z)$ during; c) $P^{10}(F)$ before; d) $P^{10}(F)$ during; e) $P^{10}(B)$ before; f) $P^{5}(B)$ during.

The main numerical characteristics of the fuzzy set of memory depth are given in the table 8.

Table 8. The main numerical characteristics of the fuzzy set of memory depth for TS of profitability.

Characteristic	**TS**					
	Gold		**EUR/USD**		**Bitcoin**	
	before	**during**	**before**	**during**	**before**	**during**
l_{max}	54	86	63	71	81	46
l_{ms}	11	6.7	5	5.15	7.8	5
l_{cg}	17.5	18.2	20.3	22.0	19.9	14.2
H_{entr_L}	16.0	14.6	16.9	21.7	19.4	13.5
Significance $\varepsilon = 0.3$						
l_{max}	17	20	8	8	47	22
l_{ms}	11	6.7	5	5.15	7.8	5
l_{cg}	9.1	10.4	5.5	19.0	15.3	12.7
H_{entr_L}	4.8	5.8	2.1	10.2	9.7	8.5
$SH(L)$	0.7	0.6	0.9	0.5	0.5	0.4

Due to the decrease in the most significant memory depth (l_{ms}) for the profitability of gold and bitcoin during the crisis, we can say about the emergence of a smaller fractal structure of the time series. For the Euro, this figure has not changed (l_{ms} = 5), but another significant depth of memory l_{ms} = 15 appeared. This is of course a positive signal, but it is offset by a significant increase in the range of the fuzzy set (l_{max} increased from 8 in the pre-crisis period to 45 in the post-crisis period). The best entropy indicators have time series of gold ($SH(L)$ = 4.8 and 5.8 in the pre-crisis and crisis periods, respectively).

5 Conclusion

Comparative analysis technique proposed in the paper integrates the tools and various diagnostic tests to determine the crucial characteristics of each studied market. The presented technique of comparative analysis has been tested on three investment markets: the precious metals market (for example, the gold market), Forex currency market (EUR/USD currency pair) and the cryptocurrency market (Bitcoin). The dynamics of these investment markets is considered in two periods: from 01.08.2019 to 31.01.2020 – pre-crisis period and from 01.02.2020 to 31.07.2020 – crisis period. The division into periods is due to significant changes in the environment in consequence of the COVID-19 pandemic and the introduction of quarantine measures. This allowed not only to compare the dynamics of the three instruments, but also to assess the reaction of markets to the crisis of the economic system.

Time series of the currency pair EUR/USD have the lowest volatility. In the pre-crisis period, the price fluctuated within a narrow range of values. The profitability of

EUR in this period corresponded to the characteristics of the normal distribution law for a random variable. Crisis phenomena in the economy intensified the amplitude of fluctuations and outlined a general upward trend against the background of significant, but short failure. The kurtosis of profitability increased rapidly, and TS of profitability ceased to meet the characteristics of the normal distribution, also and there were heavy "tails" of the distribution. But the features of fractality for this series have remained lower than the corresponding features of gold and bitcoin. And if in the pre-crisis period the most significant depth of memory lms for the price was at a relatively high level (but the entropy index was much worse than the corresponding rate of gold and bitcoin), then during the crisis lms decreased several times. Given the above, we consider the financial instrument EUR/USD to be the least attractive for investment due to the significant share of stochasticity in the dynamics of the instrument. Fractals and, accordingly, memory depth indicators have a small structure for forecasting the daily price data by fractal nonlinear dynamics methods. When working in this market, we recommend using fundamental analysis, follow the news and decisions of the European Central Bank.

Bitcoin is the instrument with the highest volatility, which on the one hand makes it possible to earn additional income, and on the other hand, increases the risks of the investor. In the pre-crisis period, preference was given to short positions, then in the post-crisis period the direction of the trend changed to upward. Statistical analysis showed that the time series of price and profitability of Bitcoin does not fall under the law of normal distribution, the nature of the dynamics is different from random, and the Broisch-Godfrey test could not confirm the absence of first-fourth order autocorrelation. A comprehensive fractal analysis of Bitcoin time series also shows a pronounced fractal dynamics. However, the evaluation of the characteristics of fuzzy memory depth showed that the fractal dynamics of Bitcoin has, firstly, high variability and, secondly, low values of the most significant memory depth lms (fractal structure is manifested in small patterns). Variability can be described as a measure of uncertainty, if we imagine a fractal structure in the form of a tree, where each branch is a new fractal, the fractal tree Bitcoin has a smaller structure compared to gold with many branches and increasing entropy during the crisis.

The main statistical and fractal indicators of gold dynamics occupy an intermediate position between currency and cryptocurrency: the level of stochasticity is lower than in EUR/USD, but the signs of fractality are slightly less than the corresponding signs of Bitcoin. The volatility of the series is also halfway between low-amplitude EUR/USD and high-amplitude cryptocurrency. However, a detailed study of the memory depth set showed that the price of gold has the highest lms (both in the pre-crisis period and during the crisis is 11 days) with low entropy. This makes it possible to use fractal characteristics when predicting the dynamics of gold. Therefore, we consider this tool the most predictable and attractive for investment.

In general, the crisis in the economy significantly affected the dynamics of all three financial instruments, changes occurred in the increase in the amplitude of fluctuations and, to a greater or lesser extent, the emergence of a general upward trend and increasing signs of fractality. However, the analysis of fuzzy memory depth revealed that increasing fractality does not always improve the level of predictability, given the simultaneous increase in the amplitude of changes in the series. The obtained indices of the characteristics of the fuzzy set make it possible to establish a reasonable forecast horizon of the forecast model.

Thus, the results of comparative analysis technique allowed developing practical recommendations to an investor: to compare the markets by their degree of predictability and to determine the параметри прогнозної моделі for each market. The results will also be used in the further development of forecast models for selected investment instruments.

References

1. Ananzeh, I.E.N.: Testing the weak form efficient market hypothesis: Empirical evidence from Jordan. International Business and Management **4**(2), 119-123 (2014)
2. Borges, M.R.: Efficient market hypothesis in European stock markets. European Journal of Finance **16**(7), 711–726 (2010). doi:10.1080/1351847x.2010.495477
3. Chen, C., Metghalchi, M.: Weak form market efficiency: Evidence from the Brazilian stock market. International Journal of Economics and Finance **4**(7), 22–32 (2012)
4. Çiçek, M.A.: Cointegration Test for Turkish Foreign Exchange Market Efficiency. Asian Economic and Financial Review **4**(4), 451–471 (2014)
5. Erokhin, S., Roshka, O.: Application of fractal properties in studies of financial markets. MATEC Web of Conferences **170**, 01074 (2018). doi:10.1051/matecconf/201817001074
6. Fama, E.F.: Efficient capital markets: a review of theory and empirical work. The Journal of Finance **25**(2), 383–417 (1970). doi:10.2307/2325486
7. Fama, E.F.: Efficient capital markets: II. The Journal of Finance **46**(5), 1575–1617 (1991). doi:10.2307/2328565
8. Gümüs, F., Zeren, F.: Analyzing the Efficient Market Hypothesis with the Fourier Unit Root Test: Evidence from G-20 Countries. Ekonomski horizonti **16**(3), 225–237 (2014)
9. Gupta, N., Gedam, A.: Testing of Efficient Market Hypothesis: a study on Indian Stock Market. Journal of Business and Management **16**(8), 28–38 (2014)
10. Hamaniuk, V., Semerikov, S., Shramko, Y.: ICHTML 2020 – How learning technology wins coronavirus. SHS Web of Conferences **75**, 00001 (2020). doi:10.1051/shsconf/20207500001
11. Ibrahim, J., Ghani, H.A.: Weak Form of Foreign Exchange Market in the Organisation for Economic Cooperation and Development Countries: Unit Root Test. International Journal of Business and Management **6**(6), 115–122 (2011)
12. Ivanchenko, I.: Methods for testing the efficiency of the financial market. Financial Analytics: Science and Experience **21**(255), 58–68 (2015)
13. Kristoufek, L.: On Bitcoin markets (in)efficiency and its evolution. Physica A: Statistical Mechanics and its Applications **503**, 257–262 (2018). doi:10.1016/j.physa.2018.02.161
14. Maksyshko N., Vasylieva O.: Investigation of the markets dynamics type for a comparative analysis of the investment instruments attractiveness. Advances in Economics, Business and Management Research **95**, 335–340 (2019). doi:10.2991/smtesm-19.2019.65
15. Maksyshko N.K.: Estimation of the system characteristics economic dynamics based of complex fractal analysis. Bulletin of Zaporizhzhia National University. Economic sciences 2(10), 119–129 (2011)
16. Onali, E., Goddard, J.: Are European equity markets efficient? New evidence from fractal analysis. International Review of Financial Analysis **20**, 59–67 (2011)

17. Onyemachi, M.O.: Weak-form market efficiency, estimation interval and the Nigerian stock exchange: empirical evidence. International Academy of Business Review **3**(1), 42–61 (2016)
18. Peters, E.: Fractal Market Analysis. Applying Chaos Theory to Investment and Analysis. John Wiley & Sons, New York (1994)
19. Semerikov, S., Chukharev, S., Sakhno, S., Striuk, A., Osadchyi, V., Solovieva, V., Vakaliuk, T., Nechypurenko, P., Bondarenko, O., Danylchuk, H.: Our sustainable coronavirus future. E3S Web of Conferences **166**, 00001 (2020). doi:10.1051/e3sconf/202016600001
20. Soloviev, V., Belinskij, A.: Methods of nonlinear dynamics and the construction of cryptocurrency crisis phenomena precursors. CEUR Workshop Proceedings **2104**, 116–127 (2018)
21. Soloviev, V., Serdiuk, O., Semerikov, S., Kohut-Ferens, O.: Recurrence entropy and financial crashes. Advances in Economics, Business and Management Research **99**, 385–388 (2019). doi:10.2991/mdsmes-19.2019.73
22. Vaga, T.: The Coherent Market Hypothesis. Financial Analysts Journal **46**, 36–49 (1990)
23. Wang, X., Lei, T., Liu, Z., Wang, Z.: Long-memory Behavior Analysis of China Stock Market Based on Hurst Exponent. In: 29th Chinese Control And Decision Conference (CCDC), pp. 1709-1712. Chongqing, PEOPLES R CHINA (2017). doi:10.1109/CCDC.2017.7978792

Assessment of bank's financial security levels based on a comprehensive index using information technology

Nila Khrushch[1[0000-0002-9930-7023]], Pavlo Hryhoruk[1[0000-0002-2732-5038]], Tetiana Hovorushchenko[1[0000-0002-7942-1857]], Sergii Lysenko[1[0000-0001-7243-8747]], Liudmyla Prystupa[1[0000-0002-5088-0654]] and Liudmyla Vahanova[2[0000-0002-5812-4597]]

[1] Khmelnytskyi National University, 11 Instytutska Str., Khmelnytskyi, 29016, Ukraine
nila.ukr@gmail.com, violete@ukr.net, tat_yana@ukr.net, sirogyk@ukr.net, prystupala@gmail.com
[2] Khmelnitsky Institute of PJSC "HEI "IAPM" Interregional Academy of Personnel, 101A Myru Ave., Khmelnytskyi, 29015, Ukraine
milascincevaganova@gmail.com

Abstract. The article considers the issues of assessing the level of financial security of the bank. An analysis of existing approaches to solving this problem. A scientific and methodological approach based on the application of comprehensive assessment technology is proposed. The computational algorithm is presented in the form of a four-stage procedure, which contains the identification of the initial data set, their normalization, calculation of the partial composite indexes, and a comprehensive index of financial security. Results have interpretation. Determining the levels of financial security and the limits of the relevant integrated indicator is based on the analysis of the configuration of objects in the two-scale space of partial composite indexes, which is based on the division of the set of initial indicators by content characteristics. The results of the grouping generally coincided with the results of the banks ranking according to the rating assessment of their stability, presented in official statistics. The article presents the practical implementation of the proposed computational procedure. To automate calculations and the possibility of scenario modeling, an electronic form of a spreadsheet was created with the help of form controls. The obtained results allowed us to identify the number of levels of financial security and their boundaries.

Keywords: financial security, bank, composite index, information technology, convolution, scale of financial security levels.

1 Introduction

1.1 Problem description

The financial sector has a leading role in the modern economy, providing financial intermediation between the various actors in the financial market and creating the basic preconditions for social reproduction. A sound and efficient financial sector encourage productive investment, thus supporting innovation and economic growth. Bank credit is also used to finance the needs of households, in particular, to balance the structure of consumption and further investment for future self-sufficiency, sustainability, and development.

The high level of financial security of the banking system determines not only the efficiency of its functioning and economic stability of society but also the financial strength and national security of the state as a whole.

Great Financial Crisis of 2007–09, which accompanied by the emergence of the crisis conditions in the banking system in many countries of the world, challenged the reliability of the entire financial system, caused a real bank bankruptcy. The current trends in the development of bank activity in the countries of the world testify to the intensification of processes that provoke general global instability. The profound, structural nature of these changes shows that the transformation of the world banking system, as well as national banking ones, is an inevitable process with indeterminate consequences. According to the official statistics of the National Bank of Ukraine, the number of banking institutions of domestic owners in Ukraine has halved for the period from 2014 to 2018 [28], which testifies to the ineffectiveness of management work and the inability to withstand financial threats.

The Covid-19 pandemic has become a new challenge for the world economy in general and all its subsystems, including banking one [35]. As has been noted in paper [3], the size, scope, and influence of the Covid-19 crisis are comparable to those of the Great Financial Crisis of 2007–09. As a consequence, long-term rating forecasts have been revised to negative for many banks, especially those with low profitability.

Therefore, currently, the main task of banks is to play their role in stabilizing the economic situation as soon as possible. To do this, they must, above all, have sufficient financial resources to immediately provide companies with the necessary financial support and ensure their survival.

Report of the National Bank of Ukraine on Financial Stability, dated June 2020 [27], states, that the Ukrainian banking sector entered the crisis caused by the pandemic in good shape and with a sufficient margin of safety. Due to the clean-up of the banking sector, the introduction of internationally recognized capital and liquidity requirements, regular stress testing, and other measures to improve the reliability of the sector, the new crisis has not led to disruptions in the banking system. The capital adequacy of banks significantly exceeded the minimum level. Banks can now use its surplus to absorb credit losses and for further lending. Financial institutions have become highly efficient and profitable, so in the end, they generate capital themselves.

The funding structure is also favorable – more than 90% of liabilities are raised in the domestic market. The banking sector today does not depend on the situation on international capital markets. The liquidity of financial institutions is high both in hryvnia and in foreign currencies. At the same time, due to the growth compared to the previous year, allocations to reserves significantly reduced the profitability of the banking sector. There are also losses from the decline in the quality of the loan portfolio, but estimates of their volume can be made only at the end of the year. The report also notes that the current crisis could lead to a breach of the fixed or regulatory capital adequacy ratios of nine banks, which account for 30% of the sector’s assets, of which

25% are two state-owned banks. In two financial institutions, the capital may become negative. These are all institutions that have shown negative results of stress testing in the past but have not taken sufficient measures to address the accumulated issues.

We agree with the statement of the study [7] that the key challenge over the coming period is to keep credit flowing amidst deteriorating credit quality and rising defaults. Critical to supporting this activity are banks' continued ability to raise funds and willingness to provide credit, which in turn would depend on market perceptions of their and their borrowers' solvency.

Under such conditions, the urgent task of the future banking system is to ensure its stability, which is the dominant condition for its steady development and ensuring the stable functioning of the national economy.

Financial security, as a functional part of the bank's economic security, provides an appropriate level of stability of its financial state, which ensures the safety of priority financial interests associated with the implementation of the adopted development strategy and the meet the target competitive positions in the conditions of external and internal threats. Therefore, the banking system faces the challenge of achieving the financial stability of each bank and guaranteeing the financial security of the entire system. The high dynamism of social development in the conditions of the growing openness of the national economy and its integration into the world economy increases the threats and influence of destabilizing factors of the internal and external environment, deepening the financial markets competition prevent the process of realization of strategic directions of banks development in terms of profitability and minimization of risks. This determines the urgency of solving the problem of ensuring the financial security of banks, which will enable them to carry out preventive actions and minimize the negative effects of the crisis in the national banking system.

At present, there are no common approaches among scholars to the definition of the nature and characteristics of the financial security of banks, as well as to the procedures and mechanisms for its quantitative assessment and identification of security level. We agree with the views of scholars [26], according to which the financial security of the bank is considered as it state is characterized by the protection of its financial interests, sufficient volume of resources, the presence of stable growth dynamics of main indicators in the current and prospective periods, which is achieved by using a sound financial strategy, flexibility in the adoption of financial decisions, timely response to the external and internal dangers and threats. Ensuring the financial security of banks should be aimed at neutralizing the structural imbalances in their development and the negative impact of destabilizing internal and external disturbances. The state and level of bank security may have some variability in time, but if the fluctuation of such fluctuations is insignificant, its development is not subject to significant changes. Effective managing the bank's activities necessitate the use of modern tools for identification of bank financial security level and analyzing its dynamics. This allows as objective as possible to describe the current state of security and promptly signal about threats to the bank's financial security depending on the purpose, objectives, and interests of all stakeholders.

1.2 Literature review

The analysis and assessment of the financial security system have long been the subject of research by researchers. Traditionally, these issues were considered within the framework of diagnosing crisis phenomena and assessing the bankruptcy probability. However, recently they have been distinguished in a separate field and it can be stated that the number of scholar's publications and practical developments in this area is increasing. The key issue that arises in assessing financial security is to establish its criterion like a benchmark that can be quantitatively measured by a certain set of initial indicators of the bank's activity. It allows us to conclude about the state or level of bank financial security. As these criteria we may use performance indicators, indicators of the financial condition, including financial viability, minimizing risks, maximizing the cost of equity capital, etc. Additionally, issues of identifying a set of defining indicators that characterize the financial security state are also studied.

An important condition for ensuring the bank's financial security is its financial stability. It is a system category that reflects the prospects for its development. It can be interpreted differently but is always based on indicators that reflect the qualitative and quantitative state of the bank's capital base, the level of liquidity, the quality of assets and liabilities, profitability, and quality of bank management. Financial stability acts both as a prerequisite and as a result of ensuring the financial security of the bank. The issue of assessing financial stability on the example of banks around the world using a variety of tools, in particular, correlation and regression analysis, comprehensive assessment technology, financial analysis, are presented in [2; 8; 12; 14; 29; 40].

The stability of a bank is largely ensured by indicators of liquidity and profitability. They are key components for ensuring the continuity of banking activities, characterize the bank's capacity, especially in times of crisis. On the one hand, they are components of a system of indicators that are used to assess both the banks' financial stability and financial security. On the other hand, many scholars attach special importance to them, considering them both as determinants and as result indicators in the procedures for evaluating the performance of the banking sector. Appropriate approaches are presented in studies [6; 18; 22; 23; 30; 34; 39; 43]. Note that many authors of the presented works paid attention to the evaluation by using the comprehensive assessment technology to design indicators of the bank's financial stability.

Consider the characteristics of the main approaches that focus on assessing the banks' financial security as a separate characteristic of the bank's operation.

At the national level, the methodology presented by the Ministry of Economic Development and Trade of Ukraine [24] is used. It provides a series of indicators of the state of banking security in Ukraine: overdue loan arrears in the total volume of loans of Ukrainian banks; the ratio of bank loans and deposits in foreign currency; the share of foreign capital in the total amount of bank's capital; the ratio of long-term loans and others. An approach based on the use of normative values as the main criteria for assessing financial security is quite common, in particular, in the papers [13; 19; 31; 41]. In our opinion, the disadvantage of this approach is that it does not allow us to establish the level of financial security, but only to state the degree of its provision for individual components. Also, the normative values of bank's effectiveness indicators are designed for some idealized banks and can't objectively reflect the state of real bank's financial security.

The mentioned disadvantages are eliminated in the methods, which use the tools of economics and mathematical modeling to assess the level and state of financial security. In particular, O. A. Sergienko with co-authors [9; 36; 37; 38] considers a set of models that allow to analyze the state and trends of financial market development, to evaluate and analyze the structural elements of bank financial security, to investigate the degree of influence local indicators to the overall financial security level. The result is scenarios for ensuring bank financial security. In our opinion, the given models require rather considerable volumes of the initial data, and the results are based on complicated calculations, which is limited to their practical application. Also, it must be taken into account that ratings are ordinal by their origins, and allow them to use very limited mathematical tools for their correct processing.

Scoring models for assessing the financial security level are given in the papers [32; 44; 46]. The advantages of such models are the simplicity of calculations and the clarity of results interpretation. However, the provision of scores for each characteristic always has some subjectivity, which makes the final evaluation results less reliable. Also, it is indirectly necessary to justify forming an expert group, to carry out agreeing expert views, etc. Among the methods of financial security modeling, of particular note is an approach based on comprehensive assessment technology [1; 4; 15; 16; 17; 20; 21; 25; 45]. Assessment methods presented in these papers differ in the set of indicators, the rules for calculating the final result, and the choice of weight coefficients of the composite index components. An important problem in assessing the financial security level is the choice and justification of the scale, which allows you to identify this level. These issues are highlighted in [16; 20; 25; 36]. However, it should be noted that the author's proposals don't always sufficiently substantiate the number of financial security levels and the boundaries of these.

In the presented researches, procedures of an estimation of financial security differ on the quantity of the initial indicators selected for calculations, ways of data normalization, and their integration into a final indicator of financial security. Although in most cases the technical side of the calculations usually does not require much effort, it should be noted that current issues related to the use of information technology, including specialized software in solving problems of assessing the level of financial security in the scientific literature are covered not enough. The use of such technologies will determine the impact of individual partial indicators on the final result, structuring of the studied objects by the values of the integrated indicator, simplify the recalculation of the result when changing the number of partial indicators, grouping, rules of their normalization, method of convolution into the final result, choice of weight values for initial indicators, etc.

The purpose of this article is to improve the methodology for assessing the level of financial security with the construction and justification of the appropriate scale of levels and their boundaries, as well as a description of the possible use of information technology to automate calculations in this field.

2 Research methodology

2.1 Description of an approach to assess the level of financial security

The basis of the proposed approach is the technology of comprehensive index assessment. The main idea underlying the approach, that is presented in paper [16], is that for determining the classification of the object under study, they are pictured in some multidimensional space of partial composite indexes, calculated by the subset of the aggregate of initial indicators and reflect a certain characteristic of the bank's activities. The number of such composite indexes is determined based on features of the set of initial bank indicators. The constructed structure of the studied objects is aimed at identifying their grouping by the distinguishing characteristics. To quantify the financial security level, a comprehensive index is used, designed by the totality of initial indicators. The number of financial security levels and their boundaries are determined based on the objects' classification obtained.

The calculating procedure for the proposed approach consists of the following steps. At the first stage, a set of initial indicators is shaped, which will be used to assess the financial security level. To simplify calculations, it is recommended to select indicators that can be obtained from the open-access banks' financial statements or that can be obtained based on such information. Also, shaping a training sample for the calculations is carried out in this step. The initial collection is divided into groups of indicators according to certain rules, in particular, based on their content characteristics. The number of groups is determined by the objectives of the study and the characteristics of the indicators.

The second stage carries out the grouping of initial indicators for the essential characteristics. Next, the procedures for data normalization and establishment of weighting factors within each group are executed. Each group can have its own rules for normalization based on the origins and essence of the indicators selected. Typically, this procedure is based on the selected sample taking into account the maximal and minimal values for each indicator.

In this stage, the procedure of data normalization is carried out normalizing with the transformation of their values into an interval from 0 to 1. For each indicator can be selected its own rules of normalization, based on the nature of the selected indicators. Usually, the rationing is carried out according to the formed sample taking into account the largest and smallest values for each indicator (scaling of their values):

$$u_{ij} = 1 - \frac{\left|x_{ij} - x_j^*\right|}{x_{jmax} - x_{jmin}}, \tag{1}$$

where u_{ij} are normalized values of indicators, x_{ij} – initial values of indicators, $x_{jmin} = \min_i x_{ij}$, $x_{jmax} = \max_i x_{ij}$, i=1, ..., m, j=1, ..., n, m – number of objects observed (banks), n – number of initial indicators.

Values x_j^* are identified by the formula:

$$x_j^* = \begin{cases} x_{jmax}, \text{ when } j \text{ is a incentive;} \\ x_{jmin}, \text{ when } j \text{ is a disincentive} \end{cases}. \tag{2}$$

This approach has the disadvantage that changing the sample can lead to a change in its largest and smallest values, and, as a consequence, a change in the sample of normalized values, which can affect the final result.

There are also other approaches to the normalization of indicators by scaling them, in particular, presented in [10]. Provided that there are normative or recommended values that can be used as the best (optimal), the normalization procedure has such form:

a) when x_{ij} is incentive:

$$u_{ij} = \begin{cases} 0, x_{ij} < 0; \\ \frac{x_{ij}}{x_j^*}, 0 \le x_{ij} \le x_j^*;, \\ 1, x_{ij} > x_j^*; \end{cases} \tag{3}$$

b) when x_{ij} is disincentive:

$$u_{ij} = \begin{cases} 0, x_{ij} \le 0; \\ \frac{x_j^*}{x_{ij}}, x_{ij} \ge x_j^*; \ , \\ 1, 0 < x_{ij} < x_j^*; \end{cases} \tag{4}$$

where x_j^* is regulatory (normative) or recommended value selected as an optimal one.

This way of normalization does not depend on the sample but has the disadvantage that not all indicators have such optimal values x_j^*.

Some approaches perform nonlinear normalization of data based on functional dependencies, that are presented, in particular, in [11; 42]. However, they are also focused on the use of optimal values of indicators, which reduces their practical application. Another procedure performed at this stage is to determine the weight of indicators both within each group and for groups. Usually, the weights are either chosen equal or set based on substantive considerations, expert evaluation. Also, quite common is the use of methods of multidimensional statistics, in particular, factor analysis.

In the third stage, the calculation of partial composite indexes for each group and the identification of the grouping of the studied set of objects in the space of new scales. As integration procedure there is usually used additive (formula (5)), multiplicative (formulas (6) and (7)) convolutions or convolutions by the method of distances (formula (8)):

$$Q_A = \sum_{j=1}^{n} w_j u_{ij} \tag{5}$$

$$Q_{M1} = \prod_{j=1}^{n} u_{ij}^{w_j}, \tag{6}$$

$$Q_{M2} = -1 + \prod_{j=1}^{n} (1 + u_{ij}^{w_j}), \tag{7}$$

$$Q_D = 1 - \sqrt{\sum_{j=1}^{n} w_j (1 - u_{ij})^2} \tag{8}$$

where Q_A, Q_{M1}, Q_{M2}, Q_D, are the values of the composite index for additive, multiplicative (formulas (6) and (7)) convolutions or convolutions by the method of distances appropriately, i=1, ..., m; w_j – weighted coefficients of indicators, j=1, ..., n. In this case, weighted coefficients should meet the condition:

$$\sum_{j=1}^{n} w_j = 1. \tag{9}$$

At the fourth stage, the calculation of the comprehensive indicator of financial security is carried out for the whole set of initial indicators. One of the convolutions (5) – (8) is also used for this purpose. After that, taking into account the typology obtained at the previous stage, the number of levels of financial security is determined and their quantitative limits are calculated.

The given procedure contains certain universalism and admits adaptation under the application of specific situations. In particular, this concerns the formation of a training sample, the choice of the initial set of banks indicators, their distribution by essential characteristics, the choice of convolution and weighting coefficients to calculate partial composite indexes and a comprehensive index, the definition of the number of financial security levels and their boundaries. In our view, some of these issues are not significant, such as the choice of grouping form for the initial set of indicators or the convolution form to design both composite and convolution indexes. Solving other issues should be in line with the results obtained by other methods, in particular the identifying financial security level for a selected sample of banks.

We can also use traditional methods of multidimensional statistical analysis, such as factor or component analysis, clustering technology to group banks. However, for the first specified group of methods, there may be difficulties in meaningful interpretation of the latent characteristics obtained. To group objects using cluster analysis technology, there are problems with the interpretation of clusters. We recommend using these methods to reconcile the results obtained by the proposed procedure.

2.2 Using information technology for providing calculations

Consider the possibility of automating calculations by means of spreadsheets. In order to create a user-friendly interface to manage initial data and calculation algorithm, it is advisable to use form controls like Spin Buttons, Scroll Bars, Check Boxes. Option Button, List Box, Combo Box.

In particular, to manage the inclusion of the indicator into the calculations, we can use the form control Check Box. To select the individual components of the calculation algorithm, in particular, the method of data normalization, the method of convolution of partial indicators, it is advisable to use form controls like List Box or Combo Box. In this case, it is necessary to provide for the formation of appropriate lists of elements in any cells of the spreadsheet to their inclusion into the body of List Box or Combo Box. Management of separate values (for example, weights) it is expedient to carry out with using Spin Buttons.

Managing by data and formulas for their processing can be performed using scenario managing tools like what-If analysis. It allows you to create, store, and substitute different sets of values for the output, as well as automatically generate reports on the results of calculations. Such tools are useful in modeling procedures for assessing the level of financial security to change the rules of data processing including their normalization, convolution, etc. and determine the most practical result.

3 Results and discussions

Let us consider the practical use of the approach proposed. To rich this aim, we select a set of initial indicators and shape a training data sample. The information source is the official data of the National Bank of Ukraine [28].

So, for analysis we were selected the following set of indicators: X_1 – index of assets coverage by equity; X_2 – index of assets coverage by authorized capital; X_3 – the share of current deposits in the bank's deposit base, %; X_4 – the share of retail deposits of in the bank's liabilities, %; X_5 – ratio of loans and deposits; X_6 – loan reserve ratio; X_7 – the ratio of assets coverage to liquid assets; X_8 – coefficient of coverage of attracted resources by liquid assets; X_9 – the ratio of liabilities coverage to liquid assets; X_{10} – ROA, %; X_{11} – ROE, %; X_{12} – net interest margin, X_{13} – the ratio of interest expenses to interest income.

The indicators $X_1 – X_6$ reflect the effectiveness of shaping the competitive capacity of banks, the rest of the indicators $X_7 – X_{13}$ – the effectiveness of banks' operational and financial activities.

The selection of two groups of indicators, in this case, has the advantage that the results of grouping can be reflected graphically in a form that is convenient for perception, and therefore subject to a clear and understandable interpretation.

The analysis of the scientific publications showed that for some of the indicators we can determine the optimal (normative or recommended) values. In particular, the author of [5] recommends the following values, presented in table 1.

Table 1. Optimal (recommended value) for indicators.

Indicator	Optimal (recommended value)
X_1 – index of assets coverage by equity	10%
X_5 – ratio of loans and deposits	70%
X_7 – the ratio of assets coverage to liquid assets	30%
X_{10} – ROA	4%
X_{11} – ROE	20%
X_{12} –net interest margin	45%
X_{13} –ratio of interest expenses to interest income	80%

The training sample was made by such Ukraine's banks, that are presented in the table 2. For calculations, we choose the data period from 2017 to 2019. The values of the relevant partial indicators, selected for calculations and distributed by groups, are presented in tables 3-8.

To provide calculations, we create an electronic form using form controls. it contains, in addition to the initial data, form controls Check Boxes to include indicators in the calculations. The Combo Boxes are used to distribute the initial partial indicators between groups; choosing the method of specifying weights for partial indicators and groups (equal or user-defined); the type of baseline indicators (incentive or disincentive), as well as the method of data normalization.

Table 2. Sample of banks.

Code	Name of Bank
C_1	Public Joint-Stock Company “Commercial Bank Privatbank” (PJSC CB Privatbank);
C_2	Public Joint-Stock Company “State Savings Bank of Ukraine” (PJSC Oshcadbank);
C_3	Joint Stock Company “The State Export-Import Bank of Ukraine” (JSC Ukreximbank);
C_4	Public Joint-Stock Company “Joint Stock Bank “Ukrgazbank”;
C_5	Joint-Stock Company “Raiffeisen Bank Aval”;
C_6	Joint-Stock Company “First Ukrainian International Bank” (FUIB);
C_7	Joint-Stock Company “Alfa-Bank Ukraine”;
C_8	Joint-Stock Company “Ukrsibbank”;
C_9	Public Joint Stock Company “Joint-Stock Commercial Industrial Investment Bank” (PJSC Prominvestbank);
C_10	Public Joint-Stock Company ‘Commercial Bank ‘Financial Initiative’
C_11	Joint-Stock Company “Megabank”
C_12	Joint-Stock Company ‘Credit Agricole Bank’
C_13	Joint Stock Company OTP Bank’
C_14	Joint Stock Company ‘Procredit Bank’
C_15	Joint Stock Company ‘Tascombank’
C_16	Joint Stock Company ‘Universal Bank’
C_17	Joint Stock Company ‘ING Bank Ukraine’
C_18	Joint-Stock Company ‘Accent-Bank’ (JSC ‘A-Bank’)
C_19	Joint-Stock Company ‘Idea Bank’
C_20	Pravex Bank’ Joint-Stock Company
C_21	Public Joint-Stock Company ‘Joint Stock Commercial Bank Industrialbank’

Table 3. The first group of indicators' data for 2017.

Code	Indicators' values					
	X_1	X_2	X_3	X_4	X_5	X_6
C_1	9.9	79.5	37.4	72.0	18.3	610.4
C_2	13.5	18.7	34.4	42.7	50.2	82.7
C_3	8.5	22.6	63.3	16.0	76.3	79.7
C_4	8.1	19.2	52.0	26.3	55.4	32.2
C_5	15.1	8.5	79.6	35.4	72.3	18.9
C_6	10.5	7.1	51.8	37.5	69.1	31.8
C_7	8.0	24.6	35.6	51.8	68.2	30.0
C_8	12.0	10.9	92.4	32.0	62.1	29.1
C_9	30.6	253.4	44.2	29.7	141.3	256.0
C_10	-29.3	21.1	52.7	8.1	537.4	124.9
C_11	11.7	6.8	24.6	40.6	99.4	8.4
C_12	10.5	4.0	62.3	23.0	77.6	12.3
C_13	11.8	20.7	84.6	38.2	68.9	35.0
C_14	11.9	6.5	55.1	32.3	104.9	4.3
C_15	6.7	4.4	17.3	46.6	94.2	7.6
C_16	14.9	49.8	23.1	39.1	76.2	27.3
C_17	37.6	7.3	34.5	0.3	138.4	8.4
C_18	15.3	7.2	16.3	89.0	90.8	28.7
C_19	11.8	8.3	18.7	70.4	87.9	29.2
C_20	19.8	18.5	83.5	32.3	19.3	5.0
C_21	29.6	14.0	34.6	59.0	80.1	17.3

Table 4. The second group of indicators' data for 2017.

Code	Indicators' values						
	X_7	X_8	X_9	X_{10}	X_{11}	X_{12}	X_{13}
C_1	7.6	9.4	8.4	-9.6	-120.0	2.8	79.8
C_2	5.2	7.8	6.1	0.3	2.4	3.0	72.1
C_3	4.0	7.6	4.4	0.6	9.4	2.3	72.7
C_4	3.5	3.8	3.8	1.0	11.7	4.2	66.0
C_5	12.9	17.3	15.2	7.0	42.6	11.9	17.9
C_6	5.8	7.1	6.4	1.7	17.6	7.4	41.6
C_7	2.5	2.9	2.7	1.5	18.0	7.5	49.2
C_8	10.0	12.9	11.3	3.2	30.3	10.9	12.8
C_9	2.6	5.6	3.7	-28.1	-135.5	4.5	52.4
C_10	0.9	5.4	0.7	-11.3	-12.4	12.7	38.1
C_11	5.0	6.9	5.7	0.8	8.0	2.1	80.1
C_12	6.8	8.5	7.6	3.6	38.7	9.6	28.8
C_13	7.3	8.6	8.3	3.4	30.2	8.2	35.8
C_14	5.3	7.2	6.0	3.4	29.8	7.4	48.1
C_15	3.9	4.5	4.1	1.4	19.3	8.9	52.1
C_16	5.5	6.6	6.5	1.8	11.6	7.3	37.2
C_17	0.9	1.5	1.5	1.5	5.3	7.4	25.8
C_18	3.3	4.0	3.9	7.7	55.3	20.7	37.9
C_19	6.0	7.2	6.8	4.0	39.3	26.5	34.9
C_20	7.4	9.9	9.3	-2.3	-10.2	5.3	38.9
C_21	7.1	10.4	10.1	0.2	0.5	8.1	38.7

Table 5. The first group of indicators' data for 2018.

Code	Indicators' values					
	X_1	X_2	X_3	X_4	X_5	X_6
C_1	11.2	73.1	45.7	70.8	23.1	464.9
C_2	8.4	22.7	32.6	47.5	44.6	94.2
C_3	5.3	23.9	52.1	17.1	90.7	89.4
C_4	7.0	16.2	43.9	27.5	69.3	18.5
C_5	14.7	7.8	75.0	34.7	81.8	8.9
C_6	12.7	6.6	55.1	39.8	67.7	34.3
C_7	8.3	20.2	41.3	58.0	62.0	28.1
C_8	12.8	9.6	86.5	33.1	64.4	13.7
C_9	20.0	368.8	44.9	28.6	146.7	415.7
C_10	-123.6	38.1	53.6	8.4	269.9	297.2
C_11	10.1	6.3	25.7	43.8	95.0	9.2
C_12	12.8	3.7	66.8	23.5	81.8	10.5
C_13	16.0	18.8	84.2	40.8	83.0	25.9
C_14	14.2	5.2	53.5	30.8	136.4	3.0
C_15	12.1	6.5	17.5	51.0	94.0	7.7
C_16	10.5	37.0	32.8	57.6	71.7	40.8
C_17	37.3	6.7	50.5	0.0	153.6	8.2
C_18	20.8	6.5	16.5	78.0	98.9	13.8
C_19	17.1	7.1	18.5	71.4	102.3	45.0
C_20	44.5	21.8	80.6	51.5	51.7	4.6
C_21	25.1	17.0	56.5	43.2	60.2	16.6

Table 6. The second group of indicators' data for 2018

Code	Indicators' values						
	X_7	X_8	X_9	X_{10}	X_{11}	X_{12}	X_{13}
C_1	7.9	9.9	8.9	4.3	40.9	7.5	45.5
C_2	4.9	6.7	5.3	0.1	0.5	2.7	71.8
C_3	2.6	5.1	2.7	0.6	8.3	1.5	81.4
C_4	4.4	5.2	4.7	1.0	13.5	4.6	61.8
C_5	12.7	17.1	14.9	6.7	45.2	12.3	20.5
C_6	5.4	6.5	6.2	4.2	36.3	9.6	37.3
C_7	3.5	4.0	3.8	2.3	28.1	7.8	45.0
C_8	9.9	12.8	11.3	5.4	43.2	9.4	15.4
C_9	4.0	8.1	5.0	-20.0	-76.1	12.2	24.3
C_10	2.0	6.9	0.9	-50.2	96.8	0.7	-27.1
C_11	5.7	7.5	6.3	0.4	3.5	2.5	77.3
C_12	6.5	8.0	7.4	4.6	39.2	8.0	37.8
C_13	7.1	8.8	8.4	6.3	45.0	10.4	29.7
C_14	2.8	4.7	3.2	3.5	26.5	5.9	51.9
C_15	5.3	6.5	6.1	2.7	27.8	7.7	51.0
C_16	4.8	5.9	5.4	1.2	10.0	7.6	45.5
C_17	1.5	2.5	2.5	4.9	13.0	9.9	29.2
C_18	3.6	4.9	4.6	13.1	72.0	29.2	26.1
C_19	3.9	5.1	4.7	8.6	58.9	36.1	23.1
C_20	14.5	27.7	26.2	-2.8	-9.0	10.7	20.5
C_21	7.1	9.7	9.5	0.4	1.6	7.5	36.5

Table 7. The first group of indicators' data for 2019.

Code	Indicators' values					
	X_1	X_2	X_3	X_4	X_5	X_6
C_1	13.5	73.4	49.7	73.1	26.1	419.3
C_2	8.3	22.7	37.2	50.2	40.1	94.0
C_3	5.3	25.5	49.8	18.2	81.4	95.7
C_4	7.2	15.9	42.9	29.7	58.2	20.3
C_5	12.9	7.6	73.9	36.2	76.6	7.1
C_6	15.2	6.6	56.7	43.0	72.2	32.9
C_7	9.5	19.4	40.5	60.3	58.9	28.8
C_8	10.9	9.7	89.1	36.5	51.6	15.1
C_9	33.0	475.4	69.2	14.9	201.0	584.1
C_10	-168.3	52.0	50.5	9.4	291.4	296.3
C_11	10.4	6.8	19.4	46.5	101.9	11.7
C_12	11.1	3.2	63.2	21.3	80.2	8.6
C_13	17.8	16.8	84.1	41.7	84.6	21.4
C_14	15.4	6.6	53.7	32.4	134.1	3.4
C_15	12.6	6.4	22.5	52.2	94.4	8.2
C_16	9.4	25.8	34.2	65.0	68.9	28.2
C_17	36.9	6.8	52.8	0.0	99.5	11.4
C_18	15.7	6.1	14.2	80.9	95.3	27.5
C_19	14.4	6.0	20.5	71.6	96.6	52.2
C_20	41.1	20.3	78.1	52.2	31.8	1.7
C_21	28.2	18.9	50.0	46.8	67.2	16.7

Table 8. The second group of indicators' data for 2019.

Code	Indicators' values						
	X_7	X_8	X_9	X_{10}	X_{11}	X_{12}	X_{13}
C_1	8.3	10.6	9.5	8.4	69.9	8.7	41.2
C_2	8.0	10.2	8.7	0.1	1.0	3.5	70.4
C_3	4.9	8.8	5.1	0.7	13.7	1.5	83.9
C_4	6.8	7.7	7.3	1.0	14.3	4.3	68.8
C_5	12.2	15.6	14.0	6.1	47.7	12.0	26.5
C_6	6.8	8.4	8.1	4.5	34.6	10.5	31.8
C_7	5.7	6.6	6.3	3.2	35.3	8.4	50.2
C_8	8.3	10.4	9.3	5.9	52.7	10.5	19.0
C_9	3.5	10.7	5.3	-16.2	-48.7	12.8	29.7
C_10	0.4	1.0	0.1	-23.3	-20.4	-8.8	89.7
C_11	3.6	5.0	4.0	0.5	5.1	2.5	83.6
C_12	6.2	7.4	7.0	4.3	38.7	7.7	41.6
C_13	7.8	10.1	9.5	6.9	42.2	11.8	29.5
C_14	4.2	7.0	5.0	3.3	23.9	5.9	53.7
C_15	5.4	6.6	6.2	2.1	20.7	8.6	50.3
C_16	5.8	6.9	6.4	3.4	30.2	9.6	47.2
C_17	3.6	5.9	5.7	6.4	17.2	10.8	31.9
C_18	4.5	5.6	5.4	13.0	79.0	30.0	31.0
C_19	4.9	6.2	5.8	7.2	48.9	37.5	23.8
C_20	11.3	20.7	19.2	-3.0	-7.1	11.6	26.5
C_21	6.0	8.7	8.3	0.5	1.8	8.7	40.2

Note, that for some indicators, we have provided the recommended values as optimal for them. To select the values of weights (in case they are set by the user) we use Spin Buttons. A fragment of the electronic form for data for 2019 is shown in figure 1. In this case, we included all indicators in the calculations but took into account different ways of their normalization.

	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
1	Calculation for 2019		Indexes												
2	Initial data	Code	X1	X2	X3	X4	X5	X6	X7	X8	X9	X10	X11	X12	X13
3		C_1	13,5	73,4	49,7	73,1	26,1	419,3	8,3	10,6	9,5	8,4	69,9	8,7	41,2
4		C_2	8,3	22,7	37,2	50,2	40,1	94,0	8,0	10,2	8,7	0,1	1,0	3,5	70,4
5		C_3	5,3	25,5	49,8	18,2	81,4	95,7	4,9	8,8	5,1	0,7	13,7	1,5	83,9
6		C_4	7,2	15,9	42,9	29,7	58,2	20,3	6,8	7,7	7,3	1,0	14,3	4,3	68,8
7		C_5	12,9	7,6	73,9	36,2	76,6	7,1	12,2	15,6	14,0	6,1	47,7	12,0	26,5
8		C_6	15,2	6,6	56,7	43,0	72,2	32,9	6,8	8,4	8,1	4,5	34,6	10,5	31,8
9		C_7	9,5	19,4	40,5	60,3	58,9	28,8	5,7	6,6	6,3	3,2	35,3	8,4	50,2
10		C_8	10,9	9,7	89,1	36,5	51,6	15,1	8,3	10,4	9,3	5,9	52,7	10,5	19,0
11		C_9	33,0	475,4	69,2	14,9	201,0	584,1	3,5	10,7	5,3	-16,2	-48,7	12,8	29,7
12		C_10	-168,3	52,0	50,5	9,4	291,4	296,3	0,4	1,0	0,1	-23,3	-20,4	-8,8	89,7
13		C_11	10,4	6,8	19,4	46,5	101,9	11,7	3,6	5,0	4,0	0,5	5,1	2,5	83,6
14		C_12	11,1	3,2	63,2	21,3	80,2	8,6	6,2	7,4	7,0	4,3	38,7	7,7	41,6
15		C_13	17,8	16,8	84,1	41,7	84,6	21,4	7,8	10,1	9,5	6,9	42,2	11,8	29,5
16		C_14	15,4	6,6	53,7	32,4	134,1	3,4	4,2	7,0	5,0	3,3	23,9	5,9	53,7
17		C_15	12,6	6,4	22,5	52,2	94,4	8,2	5,4	6,6	6,2	2,1	20,7	8,6	50,3
18		C_16	9,4	25,8	34,2	65,0	68,9	28,2	5,8	6,9	6,4	3,4	30,2	9,6	47,2
19		C_17	36,9	6,8	52,8	0,0	99,5	11,4	3,6	5,9	5,7	6,4	17,2	10,8	31,9
20		C_18	15,7	6,1	14,2	80,9	95,3	27,5	4,5	5,6	5,4	13,0	79,0	30,0	31,0
21		C_19	14,4	6,0	20,5	71,6	96,6	52,2	4,9	6,2	5,8	7,2	48,9	37,5	23,8
22		C_20	41,1	20,3	78,1	52,2	31,8	1,7	11,3	20,7	19,2	-3,0	-7,1	11,6	26,5
23		C_21	28,2	18,9	50,0	46,8	67,2	16,7	6,0	8,7	8,3	0,5	1,8	8,7	40,2
24	Include index for calculation		☑	☑	☑	☑	☑	☑	☑	☑	☑	☑	☑	☑	☑
25	Select a group		1	1	1	1	1	1	2	2	2	2	2	2	2
26	Set weitght coeffitients for first group equal														
27	Set weitght coeffitients for second gro equal														
28	Change weitght coeffitients by user		1	1	1	1	1	1	1	1	1	1	1	1	1
29	Weitght coeffitients' values		0,17	0,17	0,17	0,17	0,17	0,17	0,14	0,14	0,14	0,14	0,14	0,14	0,14
30	regulatory values (if any)		10,0				70,0		30,0			4,0	20,0	45,0	80,0
31	Kind of index		incenti	incenti	incent	incenti	incenti	incenti	incenti	incenti	incenti	incenti	incenti	incenti	disince
32	Normalization Method		by reg	by scal	by scal	by scal	by reg	by scali	by reg	by scali	by scali	by reg	by reg	by scali	by reg
34	Set weitght coeffitients for groups by user		group #1	group #2											
35	Change weitght coeffitients by user		6	7											
36	Weitght coeffitients' values		0,46	0,54											

Fig. 1. The electronic form for calculation.

Let us conduct the calculation by to data normalization formulas like (1), (3), and (4) depending on the availability of the recommended values of indicators and their type. The weights for all partial indicators within the group are set the same, and for groups – in proportion to the number of indicators in the group (w_{g1}=6/13, w_{g2}=7/13). Convolution of partial indicators for each group is carried out according to formula (5). The results of the calculation of partial indicators I_1, I_2, and the financial security comprehensive index I_{COM}, also calculated by the formula (5), are shown in table 9.

The graphical configuration of banks in the space of partial indicators is presented in figures 2, 3, 4.

Analyzing the consolidated results of the calculations we can observe the following trends.

Regarding the composite index I_1, according to the data of 2017, its highest values were achieved by two banking institutions: PJSC Prominvestbank (C_9) and PJSC CB Privatbank (C_1), their index's values are 0.69 and 0.61, respectively, the remaining banks were characterized by indicators that in the vast majority were within 0.40-0.60. In 2018, the highest levels of indicators of the effectiveness of the formation of competitive potential also showed C_1 and C_9, but C_9 managed to maximize the indicator I_1 to 0.78 against the nearest value of 0.64. Regarding the dominant trend, we can observe an overall improvement in the indicators among the surveyed banks, although the range of values of the indicator has not changed. In 2019, there is a further increase in the values of this indicator with an increase in the bottom for most banks to 0.49. Thus, based on the results of calculations, it can be concluded that the efficiency of the formation of the competitive potential of banks, in general, has tended to increase.

Table 9. Values of partial indicators I_1, I_2, and the financial security comprehensive index I_{COM}.

Code	2017			2018			2019		
	I_1	I_2	I_{COM}	I_1	I_2	I_{COM}	I_1	I_2	I_{COM}
C_1	0.61	0.33	0.46	0.64	0.58	0.61	0.60	0.66	0.63
C_2	0.44	0.31	0.37	0.43	0.23	0.32	0.42	0.36	0.39
C_3	0.47	0.34	0.40	0.42	0.26	0.33	0.41	0.41	0.41
C_4	0.41	0.34	0.37	0.42	0.35	0.38	0.39	0.46	0.43
C_5	0.54	0.83	0.70	0.55	0.70	0.63	0.54	0.76	0.66
C_6	0.49	0.49	0.49	0.52	0.54	0.53	0.53	0.63	0.58
C_7	0.46	0.40	0.43	0.49	0.44	0.46	0.49	0.57	0.53
C_8	0.55	0.71	0.64	0.56	0.63	0.60	0.54	0.66	0.61
C_9	0.69	0.24	0.44	0.78	0.26	0.50	0.82	0.34	0.56
C_10	0.31	0.24	0.27	0.39	0.32	0.35	0.37	0.13	0.24
C_11	0.43	0.35	0.39	0.45	0.28	0.36	0.44	0.30	0.37
C_12	0.48	0.62	0.56	0.51	0.56	0.53	0.49	0.61	0.55
C_13	0.57	0.62	0.60	0.60	0.58	0.59	0.59	0.67	0.63
C_14	0.48	0.57	0.53	0.49	0.47	0.48	0.49	0.55	0.52
C_15	0.37	0.45	0.41	0.45	0.49	0.47	0.46	0.53	0.50
C_16	0.46	0.45	0.45	0.52	0.35	0.43	0.51	0.58	0.55
C_17	0.38	0.28	0.32	0.42	0.43	0.43	0.42	0.56	0.50
C_18	0.51	0.61	0.56	0.51	0.60	0.55	0.51	0.64	0.58
C_19	0.48	0.71	0.60	0.51	0.63	0.57	0.51	0.68	0.60
C_20	0.43	0.36	0.39	0.56	0.54	0.55	0.50	0.55	0.52
C_21	0.49	0.40	0.44	0.51	0.32	0.41	0.51	0.37	0.44

Summarizing the results of the calculation for indicator I_2, which characterizes the efficiency of the operational and financial activities of banks, we can conclude that they differ in the range of values of the main group of banking institutions, and extremal values. In particular, when most of the values obtained in 2017 are in the range of 0.31-0.62, we observe a value of 0.83 for JSC Raiffeisen Bank Aval (C_5) and 0.71 for JSC Ukrsibbank (C_8) and JSC Idea Bank (C_19), while the minimum level was recorded at 0.24 for PJSC Prominvestbank (C_9) and PJSC Commercial Bank ‘Financial Initiative’ (C_10). According to the data of 2018, there are no stable trends to change indicators, and the changes themselves are multifaceted. Thus, the value of indicator I2 for C_5 decreased to 0.70, and for JSC Pravex Bank (C_20) its value increased to 0.54. Compared to 2019, the analytical generalization of the obtained results indicates an overall improvement in the situation and an increase in the levels of I_2 for the sample banks, as a result of which the range for the predominant group of banking institutions shifted to 0.43-0.74. However, a sharp deterioration of the situation is observed for

C_10, as a result of which the level of I_2 was minimal among the whole set of values obtained and amounted to only 0.13.

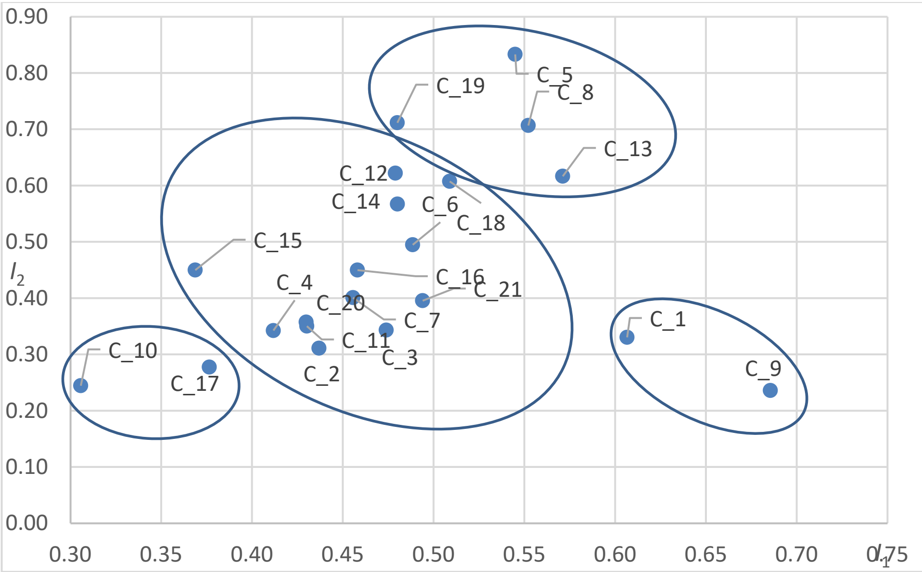


Fig. 2. Imaging banks for data of 2017.

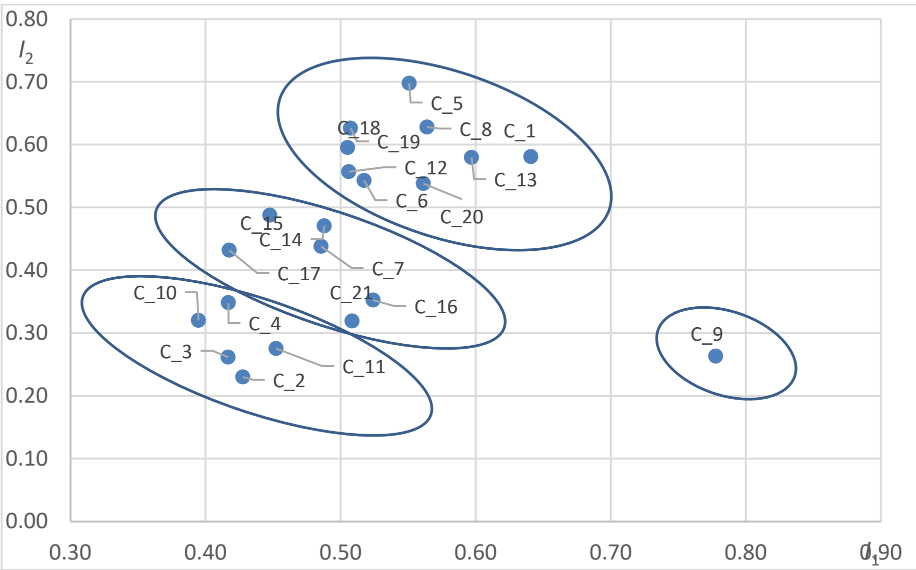


Fig. 3. Imaging banks for data of 2018.

Due to the influence of these partial indicators on the formation of the level of bank's financial security comprehensive index (I_{COM}), we note, that in 2017 the minimum level of the indicator was recorded for PJSC Commercial Bank “Financial Initiative” (C_10) and was 0.27, with a predominant overall range of 0.37 -0.56. According to the results of processing the performance of banks in 2018, we can conclude that the minimum value of the indicator increased to 0.32, and the lower limit of the main data set also increased and amounted to 0.41. As for 2019, we can note a slight improvement in most banks. An increase in the negative trend of changes in the overall level of financial security is observed only for C_10, which during 2017-2019 had the lowest values of the generalized I_{COM} indicator during the study period for this sample of banks.

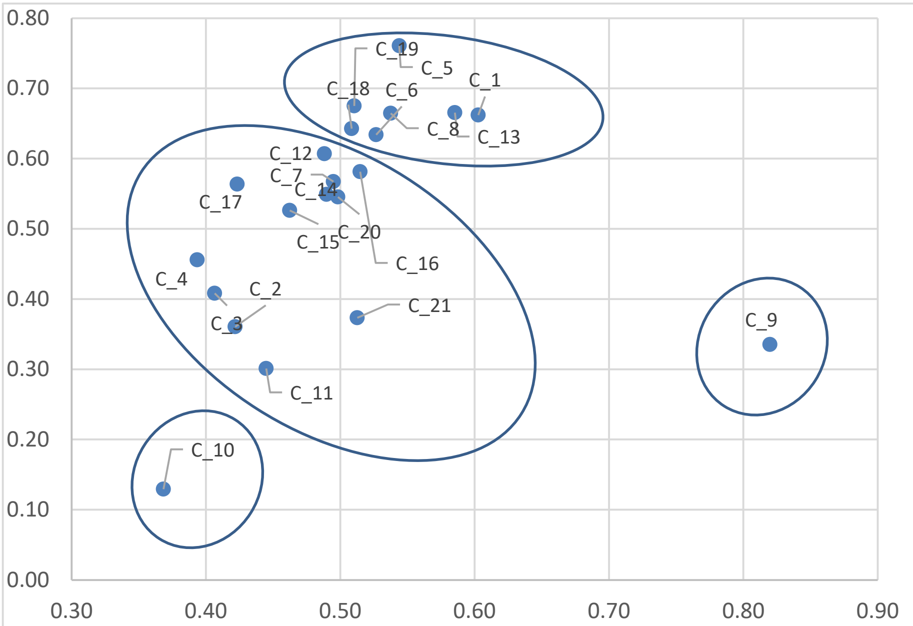


Fig. 4. Imaging banks for data of 2019.

As a result, we can state that in the realities of the existing domestic financial space, the studied stable banks show a satisfactory and sufficient level of financial security (according to the developed level’s gradation). When forming the financial security comprehensive index, the composite index I_2, which characterizes the liquidity and profitability of banks, has a more significant positive effect, while the basic components of financial security, in terms of capitalization and funding (composite index I_1) mostly need improvement.

In general, only systematic work to improve competitiveness, by providing a strong resource base and a balanced formation of assets and liabilities, and sound financial management of strategic, tactical, and operational levels will ensure the overall efficiency of operational and financial activities of the bank in a dynamic environment, with existing and potential threats to the financial security of the banking institution in the short and long term.

Analysis of graphs (fig. 2-4) shows that in general for all banks there is a fairly dense grouping of them in the space of partial composite indexes. But PJSC Prominvestbank is located as a separate object, first of all, due to high values on the scale I_1. It is one of the first banks in Ukraine, founded in 1992, it is one of the twenty banks of Ukraine in volumes of assets. However, accumulated losses affecting the level of indicators of the effectiveness of the bank's operational and financial activities are the signal of the deterioration of financial security with the established basis of competitive capacity due to the factors of capitalization, funding, and asset quality. The bank's ineffective management policy related to the ownership of its bank capital is a destabilizing factor in the bank's activities on the market.

At the same time, relatively low values on the scale of indicator I_2 are compensated by high values on the scale of indicator I_1, which allows us to position this bank quite high on the result indicator I_{COM}.

Comparison of the results of the ranking of banks on the I_{COM} scale. with the rating of their stability, presented according to the results of their comprehensive assessment on the Minfin portal [33] in general showed a coincidence of results, which indicates in favor of the proposed assessment methodology.

The next step is to determine the levels of financial security based on the values of the obtained generalized integrated indicator. Analyzing the presented in the table 9 values and graph imaging of banks in the space of partial composite indexes indicators, we can conclude that it is appropriate in this case to allocate three levels: sufficient, satisfactory, and low.

According to the results obtained in [16], to identify the boundaries of intervals for each financial security level we use the formula:

$$l_i = (\alpha_{i+1} + \beta_i)/2, \tag{10}$$

where l_i is a right bound for i-th interval, i=1..k–1, k – number of intervals; α_{i+1} – minimal value of comprehensive index, which corresponds to (i+1)-th financial security level; β_i – maximal value of comprehensive index, which corresponds to i-th financial security level. Paper [16] proposed the following boundaries for levels:

- for sufficient level: $I_{COM} \geq 0{,}51$;
- for satisfactory level: $0{,}36 \leq I_{COM} < 0{,}51$;
- for low level: $I_{COM} < 0{,}36$.

Note that the first interval, which corresponds to a sufficient level, has very wide boundaries. But in this case, using the results of calculations in table 9 and the results of grouping banks according to the levels of financial security, given above, we propose to change the low boundary for the first financial security level, leaving the ranges for other levels unchanged. Therefore, in this case, the intervals have a form:

- for sufficient level: $I_{COM} \geq 0{,}57$;
- for satisfactory level: $0{,}36 \leq I_{COM} < 0{,}51$;
- for low level: $I_{COM} < 0{,}36$.

We used the iterative procedure, given in [10], to clarify the values of the boundaries of the corresponding intervals. The results of the calculations show that for these samples there was no change in the values of the limits of the intervals of financial security levels. Therefore, the obtained generalized indicator of financial security and the corresponding scale can be used to identify the level of financial security of other banks.

4 Conclusions

An important condition for the efficient use of financial resources is the ability of the banking system to respond on time to changes in the environment and to resist external and internal threats. Therefore, it is important to assess the level of a bank's financial security, which will timely identify problem situations and develop adequate management decisions to neutralize them. The article proposes further development of the application of the technology of complex integrated assessment to determine the levels of financial security of banks. The evaluation is carried out using the procedure of the block-convolution of partial indicators. The presented evaluation tools are aimed at maximal taken into account the real state of financial institutions, are accessible and intuitive to use and provide for the data processing that has a clear quantitative dimension, are freely available on the websites of specialized institutions, or can be easily calculated from such data. The positioning of banks in the space of partial composite indexes was carried out, which allowed justifying the number of levels of financial security and their limits. The results of the calculations showed that in the realities of the existing domestic financial space the studied banks are stably functioning and show a satisfactory and sufficient level of financial security (according to the developed gradation of levels). To automate the calculations and provide the ability to model different evaluation scenarios, it is proposed to use an electronic form created in a spreadsheet using form controls. A comparison of the results of assessing the level of financial security of banks with the rating of their financial stability, presented by the Ministry of Finance, generally showed a coincidence of results. The subject of further research is the improvement of the procedure for assessing the level of bank's financial security, aimed at taking into account non-metric indicators, in particular, expert assessments; establishing and substantiating the importance of the components of the integrated indicator of financial security; improving the scale for assessing the level of financial security by identifying and justifying its high level.

References

1. Ajupov, A.A., Kurmanova, L.R., Abelguzin, N.R., Kurmanova, D.A.: A Comprehensive Assessment of the Financial Security of Banking System. International Business Management **10**, 5620–5624 (2016). doi:10.36478/ibm.2016.5620.5624
2. Alakbarov, A.: Financial Stability Assessment of the Banking Sector on the Basis of Composite Index, in Economic and Social Development. In: Ibrahimov, M., Aleksic, A.,

Dukic, D. (eds.) 37th International Scientific Conference on Economic and Social Development – "Socio Economic Problems of Sustainable Development", Baku, 2019

3. Aldasoro, I., Fender, I., Hardy, B., Tarashev, N.: Effects of Covid-19 on the banking sector: the market's assessment. BIS Bulletin 12. https://www.bis.org/publ/bisbull12.pdf (2020). Accessed 22 Aug 2020
4. Baranovskyi, O., Baryliuk, M.: Intehralna otsinka rivnia finansovoi bezpeky komertsiinoho banku (Integral Assessment of the Financial Security of Commercial Banks). Scientific Light **1**(7), 34–45 (2017)
5. Dziubliuk, O.V. Mykhailyk, R.V.: Finansova stiikist bankiv yak osnova efektyvnoho funktsionuvannia kredytnoi systemy (Financial stability of banks as a basis for the effective functioning of the credit system). Terno-graf, Ternopil (2009)
6. El-Chaarani, H.: Determinants of Bank Liquidity in the Middle East Region. International Review of Management and Marketing **9**(2), 64-75 (2019). doi:10.32479/irmm.7742
7. Financial Stability Board: COVID-19 pandemic: Financial stability implications and policy measures taken. https://www.fsb.org/wp-content/uploads/P150420.pdf (2020). Accessed 20 Aug 2020
8. Foglia, M., Angelini, E.: An explorative analysis of Italy banking financial stability. Economics Bulletin **39**(2), 1294–1308 (2019)
9. Golofaieva, I.P., Sergienko, O.A., Soldatova, Y.Y.: Managing the Financial Security of a Commercial Bank on the Basis of the Flow-Oriented Scenario Models. Business Inform **12**, 237–247 (2014)
10. Hryhoruk, P., Khrushch, N., Grygoruk, S.: An Approach to Design a Composite Index of Economic Development and Identifying the Bounds of its Levels. In: Proceedings of 9th International Conference on Advanced Computer Information Technologies ASIT'2019, Ceske Budejovice, 5-7 June 2019. IEEE (2019). doi:10.1109/ACITT.2019.8779918
11. Hryhoruk, P., Khrushch, N., Grygoruk, S.: Model for Assessment of the Financial Security Level of the Enterprise Based on the Desirability Scale. CEUR Workshop Proceedings **2422**, 169–180 (2019)
12. Ilchuk, P., Kots, O., Motoria, K.: Essence of bank financial stability and methods of its evaluation: theoretical aspects. International scientific journal "Internauka". Series: "Economic Sciences" **9** (2019). doi:10.25313/2520-2294-2019-9-5200
13. Kallaur, P.V.: The System for Ensuring the Banking Sector Sustainability of the Republic of Belarus. Money and Credit **6**, 23–29 (2008)
14. Kamran, H.W., Omran, A., Mohamed-Arshad, S.B.: Risk Management, Capital Adequacy and Audit Quality for Financial Stability: Assessment from Commercial Banks of Pakistan. Asian Economic and Financial Review **9**(6), 654–664 (2019). doi:10.18488/journal.aefr.2019.96.654.664
15. Khovrak, I.V., Petchenko, M.V.: Estimating the Level of Financial Safety in Banking Institutions. Actual Problems of Economics **2**(164), 347–354 (2015)
16. Khrushch, N., Hryhoruk, P., Prystupa, L., Vahanova, L.: Assessing Bank Financial Security Level Using the Comprehensive Index Technology. Advances in Economics, Business and Management Research **95**, 414–419 (2019). doi:10.2991/smtesm-19.2019.81
17. Klaas, J., Vagizova, V.: Tools for assessing and forecasting financial stability of the commercial bank under conditions of instability. Investment Management and Financial Innovations **11**(4), 38–44 (2014)
18. Klepkova Vodova, P.: Bank Liquidity and Profitability in the Polish Banking Sector. Scientific Journal WSFiP **1**, 19-33 (2016). doi:10.19192/wsfip.sj1.2016.2
19. Krupka, I.M.: Finansovo-ekonomichna bezpeka bankivskoi systemy Ukrainy ta perspektyvy rozvytku natsionalnoi ekonomiky (Financial and Economic Security of the

Banking System of Ukraine and Prospects of Development of National Economy). Business Inform **6**, 168-175, (2012)
20. Kulyniak, I., Prokopyshyn-Rashkevych, L., Borodavko, A.: Otsiniuvannia rivnia finansovoi bezpeky komertsiinykh bankiv Ukrainy (Evaluation of Commercial Banks Financial Security Level in Ukraine). Scientific Notes of Ostroh Academy National University, “Economics” Series **12**(40), 113–118 (2019)
21. Kvasnytska, R., Matviichuk, L., Dotsenko, I., Kovalenko, Y.: Indicative Express-Estimation of Financial Security of Banking Activity. Advances in Economics, Business and Management Research **95**, 308-313 (2019). doi:10.2991/smtesm-19.2019.60
22. Le, T., Tu, N.A.: The Determinants of Banks’ Liquidity in Vietnam. VNU Journal of Science: Policy and Management Studies **33**(2) (2017). doi:10.25073/2588-1116/vnupam.4081
23. Le, T.: The Determinants of Commercial Bank Profitability in Vietnam. SSRN Electronic Journal. https://ssrn.com/abstract=3048571 (2017). Accessed 17 Aug 2020
24. Ministerstvo ekonomichnoho rozvytku i torhivli Ukrainy. Nakaz “Pro zatverdzhennia Metodychnykh rekomedatsii shchodo rozrakhunku rivnia ekonomichnoi bezpeky Ukrainy” (Ministry of Economic Development and Trade of Ukraine Order “On approval of the Methodical Guidelines on the calculation of the level of economic security of Ukraine”). http://consultant.parus.ua/?doc=08U1H409F0 (2013). Accessed 20 Aug 2020
25. Momot, T., Rodchenko, S., Assessment of financial security of banks in conditions of macroeconomic instability. Innovative Technologies and Scientific Solutions for Industries **3**(9), 34–44 (2019). doi:10.30837/2522- 9818.2019.9.034
26. Muzychka, O.M., Zhuribida, N.R., Galko, E.O.: Metodychni pidkhody do otsiniuvannia rivnia finansovoi bezpeky bankiv (The Methodical Approaches to Evaluation of the Level of Financial Security of Banks). Business Inform **3,** 322–327 (2019)
27. National Bank of Ukraine: Financial Stability Report, June 2020. https://bank.gov.ua/en/news/all/zvit-pro-finansovu-stabilnist-cherven-2020-roku (2020). Accessed 22 Aug 2020
28. National Bank of Ukraine: Supervision. https://bank.gov.ua/en/supervision (2020). Accessed 17 Aug 2019
29. Oktaviyanti D., Purnawan, M.E.: The Behaviour of Banking Financial Stability in ASEAN-5.https://onlinelibrary.wiley.com/doi/epdf/10.1111/1759-3441.12267 - ecpa12267-note-0001 Economic Papers **38**(4), 329–344 (2019). doi:10.1111/1759-3441.12267
30. Parvin, S., Chowdhury, A.N.M.M.H., Siddiqua, A., Ferdous, J.: Effect of Liquidity and Bank Size on the Profitability of Commercial Banks in Bangladesh. Asian Business Review **9**(1), 7–10 (2019). doi:10.18034/abr.v9i1.219
31. Pestovskaya, Z.: Indicators of Banking Financial Security: Macro and Microeconomic Level. International Letters of Social and Humanistic Sciences **66**, 146–153 (2016). doi:10.18052/www.scipress.com/ILSHS.66.146
32. Poberezhnyi, S.M., Plastun, O.L., Bolgar, T.M.: Financial Security of Banking Activities. Ukrainian Academy of Banking of the National Bank of Ukraine, Sumy (2012)
33. Reityng stiikosti bankiv (Banks stability Rating). https://minfin.com.ua/banks/rating (2020). Accessed 24 Aug 2020
34. Rudhani, L., Balaj, D.: Management of Liquidity Risk and the Banking Activity. International Journal of Finance & Banking Studies **8**(2), 01-08 (2019). doi:10.20525/ijfbs.v8i2.29
35. Semerikov, S., Chukharev, S., Sakhno, S., Striuk, A., Osadchyi, V., Solovieva, V., Vakaliuk, T., Nechypurenko, P., Bondarenko, O., Danylchuk, H.: Our sustainable

coronavirus future. E3S Web of Conferences **166**, 00001 (2020). doi:10.1051/e3sconf/202016600001
36. Sergienko, O.A., Chuyko, I.M., Soldatova, Y.Y. Project Modeling of the Commercial Bank's Financial Security on the Basis of Choosing and Implementing the Vector of Strategic Priorities. Financial and credit activity: Problems of Theory and Practice **2**(19), 152–164 (2015). doi:10.18371/fcaptp.v2i19.57130
37. Sergienko, O.A., Chuyko, I.M., Soldatova, Y.Y.: Modeling Structural Elements of Commercial Bank Financial Security. Financial and credit activity: Problems of Theory and Practice **1**(18), 238–248 (2015)
38. Sergienko, O.A., Tatar, M.S., Morozova, N.L., Galych, R.V.: Banks Financial Security Management: Modeling Tools. Financial and Credit Activity: Problems of Theory and Practice **1**(22), 11–18 (2017). doi:10.18371/fcaptp.v1i22.107722
39. Sitnikova, E.V., Kolmykova, T.S., Ostapenko, E.O., Grivachev, E.A.: An Assessment of Commercial Banks' Financial Sustainability. In: Bogoviz, A., Ragulina, Y. (eds.) Industry Competitiveness: Digitalization, Management, and Integration. ISCI 2019. Lecture Notes in Networks and Systems, vol 115. Springer, Cham (2020). doi:10.1007/978-3-030-40749-0_4
40. Song, X.: Fair Value Measurement and Bank Financial Stability. Global Economy and Finance Journal **9**(2), 17–30 (2016). doi:10.21102/gefj.2016.09.92.02
41. Sukhorukov, A., Harazishvili, Y.: The Methodology of Complex Evaluation Components of Economic Security. Strategic priorities **3**(28), 5–15 (2013)
42. Voloshchuk, R.V., Stepashko, V.S.: Neliniina normalizatsiia statystychnykh pokaznykiv dlia zadachi pobudovy intehralnykh indeksiv (Nonlinear normalization of statistical indicators for the problem of constructing integral indices). Induktyvne modeliuvannia skladnykh system 6, 47–54 (2014)
43. Wojcik-Mazur, A.: Analysis of Determinants of Liquidity Risk in Polish Banking Sector. In: Jajuga K., Locarek-Junge H., Orłowski L., Staehr K. (eds) Contemporary Trends and Challenges in Finance. Springer Proceedings in Business and Economics. Springer, Cham (2019). doi:10.1007/978-3-030-15581-0_4
44. Yepifanov, A.O., Plastun, O.L., Dombrodskyi, V.S.: Financial Security of Enterprises and Banking Institutions. Ukrainian Academy of Banking of the National Bank of Ukraine, Sumy (2009)
45. Zachosova, N.: Innovative Approach in the Estimatology of Financial Institutions Economic Security: Possibilities of Usei in Management and Regulatory Activity within the Means of Provision of the State Financial Security. Baltic Journal of Economic Studies **5**(2), 45–56 (2019)
46. Zharii, Y.: Methodical Approaches to the Estimation of a Level of Financial Security of the Commercial Bank. Financial research **1**(1), 130-137 (2016)

The cryptocurrencies risk measure based on the Laplace distribution

Petro Hrytsiuk[0000-0002-3683-4766] and Tetiana Babych[0000-0001-6927-7313]

National University of Water and Environmental Engineering,
11 Soborna Str., Rivne, 33028, Ukraine
p.m.hrytsiuk@nuwm.edu.ua, t.iu.babych@nuwm.edu.ua

Abstract. Current research has led to a rejection of the hypothesis of a normal distribution of financial assets returns. Under these conditions, portfolio variance cannot serve as a good risk measure. In this paper analyzed the daily returns of the most common cryptocurrencies: Bitcoin, Ethereum, XRP, USDT, Bitcoin Cash, Litecoin. It is shown that the asset returns are not normally distributed, but with good precision follow the Cauchy distribution and Laplace distribution. The analytical expressions for risk measure were obtained using the distribution function and the VaR technique. However, the risk assessment of the return obtained on the basis of the Cauchy distribution is twice as high as the risk assessment obtained on the basis of the Laplace distribution. Therefore, the question arises: what distribution law to use to measurement the cryptocurrency risk? The paper shows that the Laplace distribution is the most adequate basis for measuring of cryptocurrencies risk.

Keywords: cryptocurrency, expected return, return distribution, risk measure, portfolio of assets.

1 Introduction

The first complete cryptographic currency appeared in 2008 thanks to the efforts of Satoshi Nakamoto. It was named Bitcoin. New varieties of digital currency appear each year due to the information technology active development and the globalization processes spread. The main advantages of cryptography are that the user controls them without any regulatory rules in the transaction. Third party costs on a transaction can be greatly reduced. This has been the main reason for the rapid development of the market for virtual currencies (crypto-currency) over the past 10 years. More than 2000 varieties of digital money have appeared on the market since the birth of Bitcoin for 5 years. Bitcoin (BTC) remains the most widespread cryptocurrency: there is the largest market capitalization among other digital currencies (about \$220 billion) [23]. The first positions of the market capitalization rating as of July 2020 are the following cryptocurrencies: ETH (Ethereum) – about \$45 billion, XRP (Ripple) – about \$12 billion, USDT (Tether) – about \$10 billion, LTC (Litecoin) and BCH (Bitcoin Cash) – \$4-5 billion each.

But investments in cryptocurrency can be quite risky as their price is very volatile [5; 12; 18; 19; 20]. Thus, during the period from July 2018 to July 2019 there were significant changes in the exchange rate. Initially, the cost of one Bitcoin was $6,600 (July 2018). There was a significant dropping in mid-December 2018 in the price – to $3,200. Then there was a sharp increasing at the end of June 2019 – to $13,000. The price of Bitcoin Cash fluctuated from $869 per unit (July 2018) to $77 per unit (mid-December 2018) to $400 per unit in June 2019. The price of the unit XRP demonstrated a sharp jump from $0.26 to $0.58 during three weeks in September 2018. Then it began to fall with slight fluctuations. The course of the ordinary currency (dollars, euros, etc.) strongly depends on inflation, politic factors and other economic conditions. Thus, its calculations can be performed fairly accurately, taking into account the influence factors changing. Instead, fluctuations in the price of cryptocurrency are very difficult to forecast. Therefore, making the correct decisions in investing and trading cryptocurrency in order to get the most return is a rather difficult task. The interaction between supply and demand, the attractiveness for investors, macroeconomic conditions and financial events are important factors in the formation of the cryptocurrency price [10]. In addition, investors rely vastly on speculation and rumors that also affect the cryptocurrency price change.

2 Literature review

Diversification is an important risk reduction tool. Creating a portfolio of financial assets is one of its instruments. In this paper, the formation of cryptocurrency investment portfolio based on the Markowitz model is investigated [17]. By changing the proportion of certain assets in a portfolio, it can be managed to maximize return or to minimize risk. The Markowitz model relies on the hypothesis of a normal distribution of returns. This hypothesis significantly simplifies the problem of choosing a portfolio for investing, since it allows you to compare alternative portfolios by just two criteria: standard deviation and mathematical expectation. However, numerous theoretical researches in the field of finance [2; 13; 15; 16; 21; 24] and the events in the financial market at the end of 2008 – early 2009 are doubted the hypothesis of a normal distribution of return.

It has been shown that the distribution of financial assets contains so-called “heavy tails”. It indicates a high likelihood of realization of very large and very small return values. The task of this work is investigating the distribution of the return of virtual currencies and using it to minimize the risk of working with portfolios of cryptocurrencies. The results of the study [11] are shown that the inclusion in the investment portfolio of several cryptocurrencies brings to investors the advantages of diversification for short term investments.

Building a portfolio solely on the basis of cryptocurrencies [8] shows that a cryptocurrencies set increases investment opportunities with a low level risk. In contrast to our research, this work does not take into account the possible deviation of the distribution of the cryptocurrency return from the normal one. In the work [1] researchers apply a portfolio diversification strategy that is based on several models of portfolio formation. So, on the basis of the modern portfolio theory, an optimal risk portfolio has been established and the effect of cryptocurrency on the usual investment portfolio of assets has been investigated. The results, obtained in [6], show that the expected return on the cryptocurrency portfolio is greater than the return of separate cryptocurrency. The risk assessment was carried out according to the quantile method, but unlike our research, the distribution of assets return does not determine.

The authors of [4] emphasize the importance of modeling nonlinearity and taking into account the behavior of tail distribution in analyzing the causal relationships between Bitcoin revenues and trading volume. For analysis the Bitcoin behavior in the study [7] taking into account heavy tails of return distribution, quantile regression is used. This made it possible to determine that Bitcoin does act as a hedge against market uncertainty. Yet, the quantile method is applied only to Bitcoin analysis without specifying the asset return distribution [4; 7]. The authors of the article [9] analyzed some statistical properties of the largest cryptocurrencies, in particular their distribution law. In the study accentuated that the return is clearly non-normal. Several types of distribution have been identified, which are subject to certain cryptocurrencies. These are the generalized hyperbolic distribution (Bitcoin and Litecoin), and the normal inverse Gaussian distribution, the generalized *t* distribution, and the Laplace distribution for smaller cryptocurrencies. The article [22] showed that the profitability of Bitcoin after risk adjustment, depending on the specific measure of risk, can be compared with the profitability of shares based on Sharpe and Sortino ratios using. In the paper [3] another approach is offered. It considers the decision-making process related to technological innovation is considered in the conditions of uncertainty and risk arising from incomplete information about the explored system. The proposed model allows describing the dynamics of multi-stage control of the technological innovation process, depending on investment resources receipt.

3 Methods

Thus, as shown by the analysis of literary sources, in present-day conditions, not only currencies and valuable metals are used for investment, but also cryptocurrency assets are added to the portfolio. Our analysis was done on the basis of historical data on prices of 6 cryptocurrency (Bitcoin, Bitcoin Cash, Litecoin, XRP, Ethereum, Tether) for the period from January 1, 2018 to June 30, 2020. This data are freely available from the www.coinmarketcap.com site – CoinMarketCap Analytical Services contains historical and actual data about cryptocurrency. The data set is divided into 6 parts, each of which refers to a specific quarter of the study period. The volume of quarterly data is 90–92 records, the total amount of data – 912 records. For comparison, we included in our analysis a study of the stock prices of such leading companies as Amazon and Google. In this case quarterly data volume is 61–64 records, the total amount of data – 628 records.

For further processing, the calculation of the corresponding normalized cryptocurrency return is performed according to following equation

$$x_{ni} = (C_{ni+1}/C_{ni} - 1) \cdot 100\%, \quad (1)$$

where x_n is the daily return of the n-th asset, C_n is the daily closing price of the n-th asset, i is the observation number.

The dynamics of cryptocurrency Bitcoin return is presented in fig. 1. The main characteristics of the investigated cryptocurrency return for the observed period are given in table 1. As well, for comparison, in table 1, we introduced the statistical characteristics of the two successful companies' stocks. The analysis of statistical characteristics, given in table 1, showed that the daily stock return of the represented companies is higher than the similar investigated cryptocurrencies return. At the same time, their risk (if we consider the risk as a standard deviation) is much lower (except for the cryptocurrency USDT). From the correlation matrix (table 2) it can be seen that the return of the cryptocurrency is sufficiently correlated with each other (except USDT).

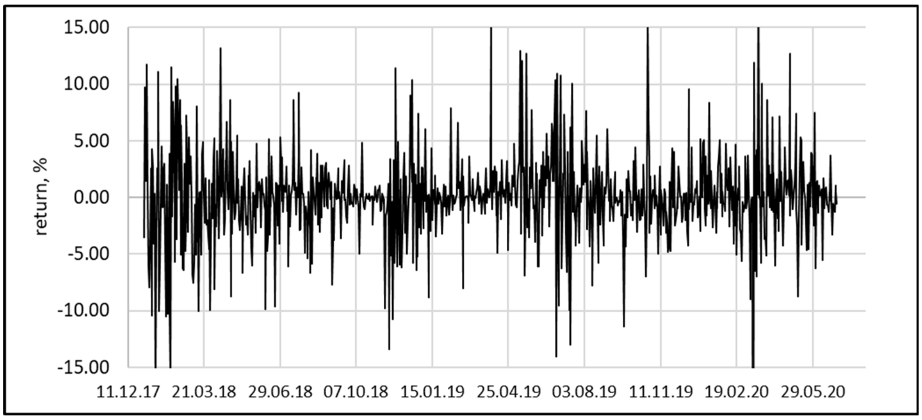


Fig. 1. Dynamics of day return cryptocurrency Bitcoin (01.01.2018 – 30.06.2020).

Table 1. Statistical characteristics of cryptocurrency return (%) for the period 01/01/2018 to 06/30/2020.

	Minimum	**Maximum**	**Average**	**Median**	**Standard deviation**	**Skewness**
Google	-11.101	10.449	0.067	0.130	1.970	-0.109
Amazon	-7.922	9.445	0.158	0.208	2.043	-0.098
BTC	-37.170	18.188	0.037	0.100	4.054	-0.662
ETH	-42.347	18.940	0.001	-0.048	5.083	-0.566
XRP	-32.899	37.989	-0.141	-0.286	5.301	0.465
USDT	-5.121	5.484	0.000	-0.009	0.540	0.511
BCH	-42.956	51.214	-0.055	-0.300	6.532	0.776
LTC	-36.177	33.722	-0.050	-0.225	5.268	0.455

Let's introduce the concept of the risk zone frontier [14]. In this capacity we will use the 5% quantile of return. To determine the risk zone frontier, it is necessary to identify the distribution of returns. Under the investor risk we understand the difference between the most expected value of cryptocurrency return and 5% quantile of return (risk zone frontier L), which is determined using the corresponding return distribution. If the distribution is normal, the most expected return value is the average value of sample $\overline{x}$. If the distribution is different from the normal one and is asymmetric, we will use the median return μ as an expected return. A significant asymmetry in the return distribution (last row of table 1) prompts as the most expected return value to choose the median sample, rather than the average value of sample.

Table 2. Correlation matrix of cryptocurrency return (%) for the period 01/01/2018 to 06/30/2020.

	BTC	**ETH**	**XRP**	**USDT**	**BCH**	**LTC**
BTC	1	0.84	0.69	-0.01	0.78	0.81
ETH	0.84	1	0.77	-0.07	0.79	0.84
XRP	0.69	0.77	1	-0.07	0.67	0.72
USDT	-0.01	-0.07	-0.07	1	-0.03	-0.04
BCH	0.78	0.79	0.67	-0.03	1	0.79
LTC	0.81	0.84	0.72	-0.04	0.79	1

Consequently, the value of the asset risk, in accordance with the above definition, can be estimated by the ratio

$$V_j = \mu_j - L_j. \tag{2}$$

For statistical research, we divided the data set into 10 time intervals, each of which corresponds to one quarter. As a result of research of the cryptocurrencies Bitcoin, Bitcoin Cash, Litecoin, XRP, Ethereum, Tether using the Pearson, Kolmogorov-Smirnov, and Shapiro-Wilk tests, in most cases the hypothesis of return normal distribution was rejected (fig. 2). Computer experiments showed that the return of the investigated cryptocurrency with good accuracy is described by both Cauchy distribution and Laplace distribution (fig. 3).

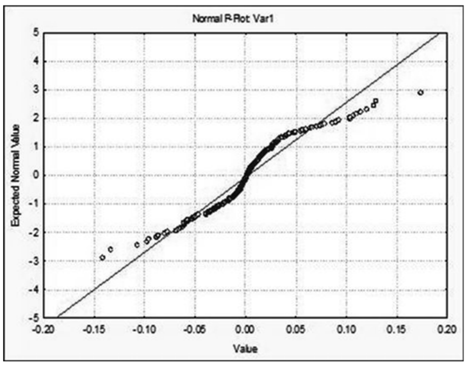


Fig. 2. Hypothesis testing on normal distribution of the Bitcoin return.

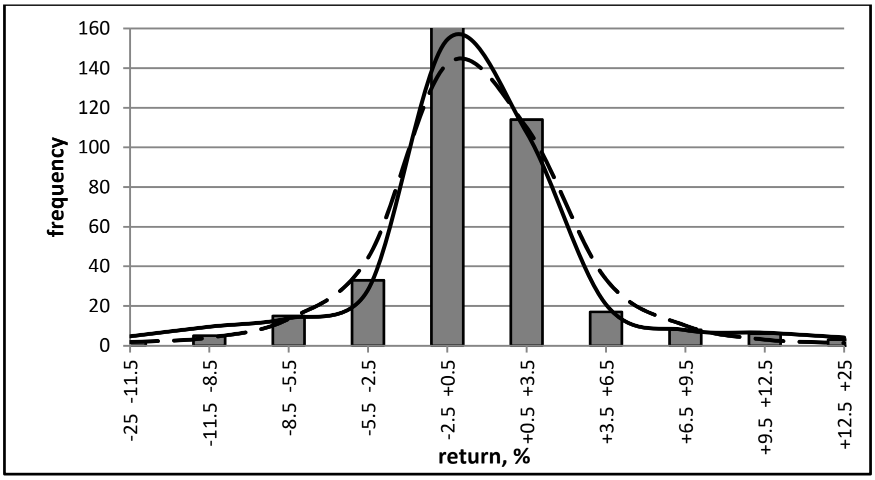


Fig. 3. Actual form of Bitcoin return distribution (gray columns), Cauchy distribution (solid line), Laplace distribution (dashed line).

To test the hypothesis of the Cauchy (Laplace) distribution of cryptocurrency returns, we used Pearson's chi-squared test (χ^2). To apply this criterion, it is necessary to calculate Pearson statistics using the formula

$$Q^2 = \sum_{i=1}^{k} \frac{(n_i - m_i)^2}{m_i}, \tag{3}$$

and compare it with tabular values $\chi^2_{crit}(\alpha, k-3)$. Here k is the number of intervals, m_i – the theoretical number of the random variable values in the i-th interval, n_i – the actual number of the random variable values in the i-th interval, $\alpha = 0.05$ – the level of significance of the test. In our case $\chi^2_{crit}(0.05, 10-3) = 14.07$. If $Q^2 \leq \chi^2_{crit}$ the hypothesis of Cauchy (Laplace) distribution is accepted, otherwise it is rejected. The results of test of hypothesis for the cryptocurrency return distribution are shown in the tables 3, 4. It is seen that for most cases the hypothesis of the corresponding distribution is accepted at the level $\alpha = 0.05$. The tables 3, 4 also show the results of testing the hypothesis of the return distribution for Google stocks and Amazon stocks. Comparing table 3 and table 4, we can conclude that the Laplace distribution more accurately describes the distribution of cryptocurrency return compared to the Cauchy distribution.

The Cauchy distribution function has the form

$$F(x) = \frac{1}{\pi} arctg\left(\frac{x-\mu}{\gamma}\right) + \frac{1}{2}. \tag{4}$$

Here μ is the mathematical expectation (median) of return, γ is the coefficient of distribution function chosen by us for each case in accordance with the least squares method.

The Laplace distribution function $F(x)$ has the form

$$F(x) = \begin{cases} \frac{1}{2}e^{\gamma(x-\mu)}, & x \le \mu \\ 1 - \frac{1}{2}e^{-\gamma(x-\mu)}, & x > \mu. \end{cases} \quad (5)$$

Here x is the return on financial assets, μ is the mathematical expectation (median) of return, γ is the coefficient of distribution function chosen by us for each case in accordance with the least squares method.

Table 3. Pearson's chi-squared test for Cauchy distribution.

	Google	Amazon	BTC	ETH	XRP	USDT	BCH	LTC
18_Q1	9.15	11.68	12.09	4.86	8.14	5.35	10.7	11.22
18_Q2	12.11	7.54	14.52	12.1	10.33	13.89	12.12	14.51
18_Q3	15.07	12.63	8.93	10.6	5.28	13.14	9.24	14.09
18_Q4	2.92	5.7	10.85	4.84	12.34	3.59	7.44	5.03
19_Q1	7.92	5.18	14.07	5.41	3.22	10.19	11.1	8.84
19_Q2	7.03	6.94	14.07	8.95	4.32	13.32	5.41	10.02
19_Q3	10.19	15.19	8.25	13.79	8.64	14.99	10.69	4.87
19_Q4	13.23	8.34	8.21	4.76	14.37	14.11	13.82	14.29
20_Q1	4.99	2.15	8.82	5.17	7.02	12.75	5.14	7.72
20_Q2	3.11	5.78	6.21	6.24	5.84	26.67	8.23	9.1
Average	*8.57*	*8.11*	*10.6*	*7.67*	*7.95*	*12.8*	*9.39*	*9.97*

Table 4. Pearson's chi-squared test for Laplace distribution.

	Google	Amazon	BTC	ETH	XRP	USDT	BCH	LTC
18_Q1	5.43	3.64	6.04	2.15	7.2	3.58	7.74	6.95
18_Q2	5.79	4.66	11.31	8.08	3.8	5.55	5.43	8.47
18_Q3	6.79	4.6	5.23	2.77	3.2	4.15	8.53	6.21
18_Q4	0.76	4.7	14.25	7.73	9.55	1.35	9.99	2.69
19_Q1	3.06	1.89	10.41	9.81	5.11	4.35	8.46	4.53
19_Q2	3.73	5.96	6.14	6.12	2.74	6.31	4.09	7.1
19_Q3	5.15	8.08	3.43	6.29	4.78	6.69	6.74	5.77
19_Q4	5.78	4.87	5.49	3.84	10.81	5.8	4.99	10.24
20_Q1	4.58	0.53	6.67	2.96	6.07	9.73	4.14	3.67
20_Q2	5.43	3.64	6.01	3.07	3.07	12.4	2.06	2.98
Average	*4.65*	*4.26*	*7.5*	*5.28*	*5.63*	*5.99*	*6.22*	*5.86*

To determine the coefficient γ an interval distribution table was constructed. The role of the minimized value was the sum of the squares of the differences between the theoretical and actual values of the frequency at different intervals (equation 3). The parameter μ (median) for the various cryptocurrencies and periods are shown in table 1.

Using the form of the Cauchy distribution function (4), we can find an analytic expression for the frontier of risk zone L_α at a given confidence level α [14]:

$$L_{\alpha} = \mu + \gamma \cdot tg\left(\pi\left(\alpha - \frac{1}{2}\right)\right). \quad (6)$$

Similarly, for the Laplace distribution, from relation (4) we determine an analytic expression for the frontier of risk zone

$$L_{\alpha} = \mu + \frac{ln(2\alpha)}{\gamma}. \quad (7)$$

Using (2), (6), (7) we calculated the risk value V at the level of 5% for each cryptocurrency at the appropriate period of time (quarter). However, the risk value calculated on the basis of the Cauchy distribution (riskC) is twice the value of the risk calculated on the basis of the Laplace distribution (riskL). For example, for Bitcoin in the 1st quarter of 2018 the value risk Cauchy $V_C = 25.89\%$, the value risk Laplace $V_L = 12.60\%$. A similar situation is observed for other cryptocurrencies and periods (table 5). For comparison, the Table 5 also shows statistics for Google stocks of and Amazon stocks. The standard deviation, which is a measure of risk in the normal distribution, almost halves the risk compared to the estimate obtained from the Laplace distribution (table 5). In this regard, the question arises: which of the two distributions described above most adequately describes the risks of cryptocurrencies: the Cauchy distribution or the Laplace distribution?

Table 5. Statistical characteristics of cryptocurrency risks, %.

Year	2018				2019				2020		Average
Quarter	Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4	Q1	Q2	
Google											
Mediane	0.38	0.05	0.09	-0.18	0	0.16	0.19	0.05	0.15	0.5	*0.14*
StDev	1.99	1.42	1.19	2.3	1.5	1.69	1.82	0.94	3.42	2.36	*1.86*
RiskC	6.87	5.84	5.3	8.49	5.47	4.54	5.09	3.71	10.55	5.69	*6.15*
RiskL	3.97	2.64	2.34	4.17	2.65	2.39	2.47	1.76	5.55	3.52	*3.15*
StDev/RiskL	0.5	0.54	0.51	0.55	0.56	0.71	0.74	0.53	0.62	0.67	*0.59*
RiskC/RiskL	1.73	2.21	2.26	2.03	2.06	1.9	2.06	2.11	1.9	1.62	*1.99*
Amazon											
Mediane	0.48	0.3	0.34	-0.38	0.29	0.1	0.01	0.02	0.09	0.74	*0.2*
StDev	1.94	1.65	1.31	3.52	1.92	1.43	1.3	1	2.88	2.14	*1.91*
RiskC	7.71	5.39	5.48	13.32	6.59	4.03	6.02	2.92	10.71	6.5	*6.87*
RiskL	4.27	2.9	2.45	6.56	3.4	2.47	2.67	1.8	5.21	3.46	*3.52*
StDev/RiskL	0.45	0.57	0.53	0.54	0.56	0.58	0.49	0.56	0.55	0.62	*0.55*
RiskC/RiskL	1.81	1.86	2.23	2.03	1.94	1.63	2.26	1.63	2.06	1.88	*1.93*
BTC											
Mediane	-0.11	0.17	0.36	-0.12	0.12	1.32	-0.18	-0.2	-0.18	0.17	*0.14*
StDev	5.97	3.7	2.81	3.87	2.21	4.59	3.9	2.98	5.61	3.16	*3.88*
RiskC	25.89	12.37	9.33	10.08	4.49	13.22	14.02	8.71	13.87	9.27	*12.13*
RiskL	12.6	6.42	4.84	6.15	3.4	6.94	6.68	4.79	7.04	5.34	*6.42*
StDev/RiskL	0.47	0.58	0.58	0.63	0.65	0.66	0.58	0.62	0.8	0.59	*0.62*
RiskC/RiskL	2.05	1.93	1.93	1.64	1.32	1.9	2.1	1.82	1.97	1.74	*1.84*
ETH											
Mediane	-0.41	0.31	-0.6	-0.18	-0.15	0.68	-0.1	-0.4	0.19	0.33	*-0.03*
StDev	6.66	5.23	4.83	5.59	4.14	4.69	4.31	3.08	6.98	4.06	*4.96*

Year	2018				2019				2020		Average
Quarter	**Q1**	**Q2**	**Q3**	**Q4**	**Q1**	**Q2**	**Q3**	**Q4**	**Q1**	**Q2**	
RiskC	28.71	21.57	16.52	16.15	10.91	15.63	15.15	9.34	17.98	12.88	*16.48*
RiskL	13.17	9.56	7.94	9.27	6.83	8.16	7.08	5.27	8.77	6.52	*8.26*
StDev/RiskL	0.51	0.55	0.61	0.6	0.61	0.58	0.61	0.58	0.8	0.62	*0.61*
RiskC/RiskL	2.18	2.26	2.08	1.74	1.6	1.92	2.14	1.77	2.05	1.98	*1.97*
XRP											
Mediane	-0.41	0.31	-0.6	-0.18	-0.15	0.68	-0.1	-0.4	0.19	0.33	*-0.03*
StDev	6.66	5.23	4.83	5.59	4.14	4.69	4.31	3.08	6.98	4.06	*4.96*
RiskC	28.71	21.57	16.52	16.15	10.91	15.63	15.15	9.34	17.98	12.88	*16.48*
RiskL	13.17	9.56	7.94	9.27	6.83	8.16	7.08	5.27	8.77	6.52	*8.26*
StDev/RiskL	0.51	0.55	0.61	0.6	0.61	0.58	0.61	0.58	0.8	0.62	*0.61*
RiskC/RiskL	2.18	2.26	2.08	1.74	1.6	1.92	2.14	1.77	2.05	1.98	*1.97*
USDT											
Mediane	-0.41	0.31	-0.6	-0.18	-0.15	0.68	-0.1	-0.4	0.19	0.33	*-0.03*
StDev	6.66	5.23	4.83	5.59	4.14	4.69	4.31	3.08	6.98	4.06	*4.96*
RiskC	28.71	21.57	16.52	16.15	10.91	15.63	15.15	9.34	17.98	12.88	*16.48*
RiskL	13.17	9.56	7.94	9.27	6.83	8.16	7.08	5.27	8.77	6.52	*8.26*
StDev/RiskL	0.51	0.55	0.61	0.6	0.61	0.58	0.61	0.58	0.8	0.62	*0.61*
RiskC/RiskL	2.18	2.26	2.08	1.74	1.6	1.92	2.14	1.77	2.05	1.98	*1.97*
BCH											
Mediane	-0.41	0.31	-0.6	-0.18	-0.15	0.68	-0.1	-0.4	0.19	0.33	*-0.03*
StDev	6.66	5.23	4.83	5.59	4.14	4.69	4.31	3.08	6.98	4.06	*4.96*
RiskC	28.71	21.57	16.52	16.15	10.91	15.63	15.15	9.34	17.98	12.88	*16.48*
RiskL	13.17	9.56	7.94	9.27	6.83	8.16	7.08	5.27	8.77	6.52	*8.26*
StDev/RiskL	0.51	0.55	0.61	0.6	0.61	0.58	0.61	0.58	0.8	0.62	*0.61*
RiskC/RiskL	2.18	2.26	2.08	1.74	1.6	1.92	2.14	1.77	2.05	1.98	*1.97*
LTC											
Mediane	-0.41	0.31	-0.6	-0.18	-0.15	0.68	-0.1	-0.4	0.19	0.33	*-0.03*
StDev	6.66	5.23	4.83	5.59	4.14	4.69	4.31	3.08	6.98	4.06	*4.96*
RiskC	28.71	21.57	16.52	16.15	10.91	15.63	15.15	9.34	17.98	12.88	*16.48*
RiskL	13.17	9.56	7.94	9.27	6.83	8.16	7.08	5.27	8.77	6.52	*8.26*
StDev/RiskL	0.51	0.55	0.61	0.6	0.61	0.58	0.61	0.58	0.8	0.62	*0.61*
RiskC/RiskL	2.18	2.26	2.08	1.74	1.6	1.92	2.14	1.77	2.05	1.98	*1.97*

4 Risk zone testing

Analysis of relations (4) and (5) showed that the Cauchy distribution has very long and heavy tails (fig. 4). In this regard, the risk zone frontier determined on the basis of the Cauchy distribution will be significantly smaller than the risk zone frontier determined on the basis of the Laplace distribution. In this case, the number of return cases that fall into the risk zone determined on the Cauchy distribution basis will be significantly less than the number of cases that fall into the risk zone determined on the Laplace distribution basis. After counting the number of cases that fall into the risk zone, we can conclude which of the two distribution laws more adequately describes the distribution of cryptocurrency returns in the negative return zone. The results of counting the number of critical cases (the case where the return falls into the risk zone) at the confidence level 5% are shown in table 6 and table 7.

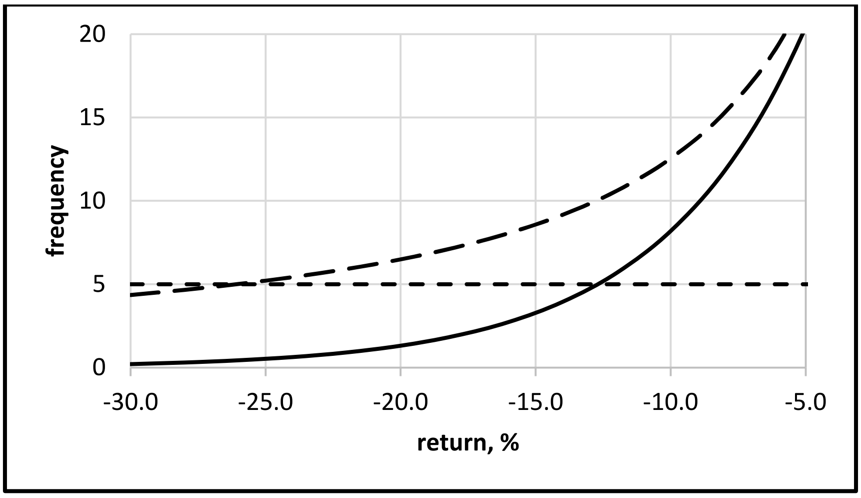


Fig. 4. The Laplace distribution (solid line), the Cauchy distribution (dashed line).

Table 6. Number of critical cases for Cauchy distribution. The frontier of the risk zone was determined at the confidence level 5%.

Year	**2018**				**2019**				**2020**		**All cases**	**All days**	**Frequency, %**
Quarter	**1**	**2**	**3**	**4**	**1**	**2**	**3**	**4**	**1**	**2**			
Google	0	0	0	0	0	2	0	0	1	1	4	628	0.64
Amazon	0	1	0	0	0	1	0	0	0	1	3	628	0.48
BTC	0	0	0	2	2	1	0	0	1	0	6	912	0.66
ETH	0	0	1	0	2	0	2	0	1	0	6	912	0.66
XRP	0	0	0	1	2	0	0	1	1	1	6	912	0.66
USDT	0	1	0	0	1	0	0	0	3	0	5	912	0.55
BCH	0	0	0	1	2	0	2	1	1	0	7	912	0.77
LTC	0	0	0	0	0	0	1	0	1	0	2	912	0.22
Stocks	0	1	0	0	0	3	0	0	1	2	7	1256	0.56
Crypto-currencies	0	1	1	4	9	1	5	2	8	1	32	5472	0.58

As can be seen from Table 6, in the case where the risk area is determined based on the Cauchy distribution at the confidence level 5%, the average probability for falling of cryptocurrency return into the risk area is 0.6% – 0.7% (except Litecoin). For the stocks return the average probability for falling into the risk zone is 0.5% – 0.6%. That is, the actual frequency of critical cases is 10 times less than theoretically predicted. Hence the conclusion about the inadequate description of the distribution of cryptocurrency (stock) returns in the negative return zone using the Cauchy distribution.

Table 7. Number of critical cases for Laplace distribution. The frontier of the risk zone was determined at the confidence level 5%.

Year	**2018**				**2019**				**2020**		**All cases**	**All days**	**Frequency, %**
Quarter	**1**	**2**	**3**	**4**	**1**	**2**	**3**	**4**	**1**	**2**			
Google	5	2	1	4	2	4	3	2	3	5	31	628	4.94
Amazon	4	4	3	1	2	3	3	0	3	3	26	628	4.14
BTC	2	5	6	5	3	5	4	2	2	4	38	912	4.17
ETH	2	5	7	6	3	3	6	6	4	3	45	912	4.93
XRP	3	4	4	4	3	1	5	4	5	4	37	912	4.06
USDT	5	4	6	5	2	1	3	2	4	3	35	912	3.84
BCH	3	2	2	5	3	4	5	6	4	3	37	912	4.06
LTC	2	3	4	4	3	1	5	6	4	3	35	912	3.84
Stocks	9	6	4	5	4	7	6	2	6	8	57	1256	4.54
Crypto-currencies	17	23	29	29	17	15	28	26	23	20	227	5472	4.16

In the case where the risk area is determined based on the Laplace distribution at the confidence level 5% (Table 7), the average probability for falling of cryptocurrency return (and stock return) into the risk area is 4% – 5%. Thus, the actual frequency of critical cases is close to the theoretically predicted. Thus, the Laplace distribution is an adequate basis for the risk measure of negative returns of cryptocurrency's and stock returns.

5 Formation of a cryptocurrency portfolio

When forming a cryptocurrencies portfolio, first of all it is necessary to take into account their return. Figure 5 shows the average return of cryptocurrencies for the two quarters of 2020. The best cryptocurrencies in terms of profitability are BTC, ETH, XRP and BCH. As can be seen from fig. 5, the average quarterly return on stocks significantly exceeds the average quarterly cryptocurrency return. This means that stocks are more attractive for long-term investments. Cryptocurrencies are a tool for speculative transactions.

Another important aspect of portfolio formation is taking into account the risks of cryptocurrencies and taking into account the correlations of their profitability. The minimum risk is typical for cryptocurrency USDT.

From the correlation matrix (table 8) it can be seen that the return of the cryptocurrency is sufficiently correlated with each other (except USDT). It is clear that the cryptocurrency USDT is the most important component of the portfolio, which will reduce its risk (fig. 6).

For the building the cryptocurrencies portfolio, let's used the technique, described in previous research [14]. Assuming that cryptocurrency returns $r_i(t)$ are poorly stationary random processes, each of which is characterized by mathematical expectations μ_i and a degree of risk V_i, then for portfolio optimization, a modified Markowitz model can be used. In this case, the mathematical description of the problem at the maximum portfolio return will have the form:

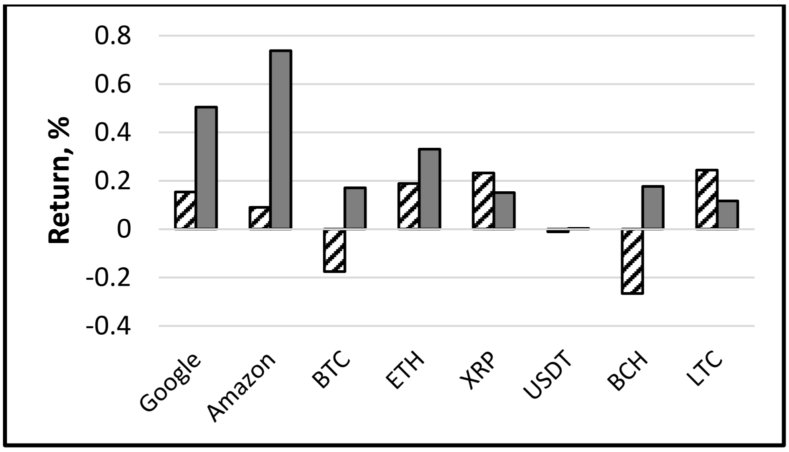


Fig. 5. Cryptocurrencies return for the I Quarter 2020 (dashed) and II Quarter 2020 (solid).

Table 8. Correlation matrix of cryptocurrency return (%) for the period 01/01/2020 to 06/30/2020.

	BTC	**ETH**	**XRP**	**USDT**	**BCH**	**LTC**
BTC	1	0.85	0.79	-0.01	0.82	0.84
ETH	0.85	1	0.86	-0.11	0.91	0.91
XRP	0.79	0.86	1	0.01	0.85	0.91
USDT	-0.01	-0.11	0.01	1	-0.02	0.00
BCH	0.82	0.91	0.85	-0.02	1	0.91
LTC	0.84	0.91	0.91	0.00	0.91	1

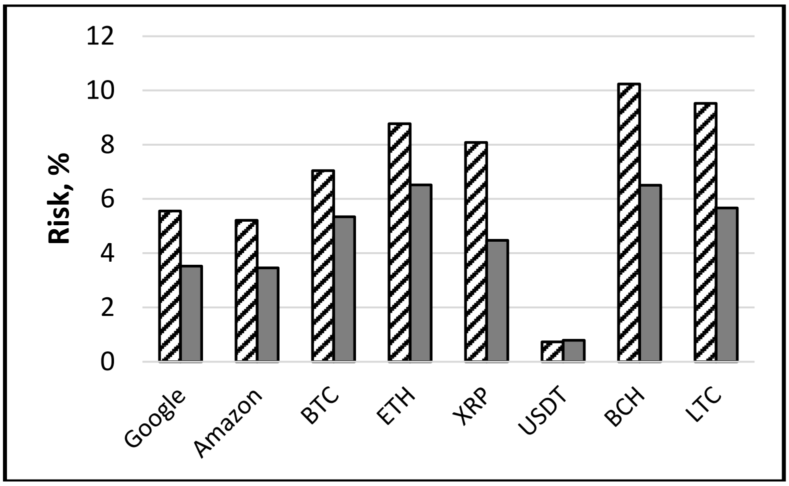


Fig. 6. Average cryptocurrencies risk for the I Quarter 2020 (dashed) and II Quarter 2020 (solid).

$$\begin{cases} R_p = w_i \times \mu_i \rightarrow max; \\ V_p = \sqrt{\sum_{i=1}^{6}\sum_{j=1}^{6}\left(w_i \times V_i \times w_j \times V_j \times \rho_{ij}\right)} \leq V_{req}; \\ w_i \geq 0;\ \sum w_i = 1. \end{cases} \tag{8}$$

To assess of portfolio risk V_p, we used an approach similar to the Markowitz approach, but for the risk measure we used definition (2), rather than the standard deviation of the cryptocurrency return.

So, using the obtained above cryptocurrency risk estimates RiskL (table 5, column 20_Q2), we constructed the set of optimal portfolios (the efficient frontier). Each such portfolio gives maximum return at the established risk level. The table 9 presents the portfolio structure for each, obtained by us, optimal solution. The analysis of the table confirms the well-known statement that a higher return level always requires a higher risk degree. As you can see, the main role in the formation of the portfolio is played by cryptocurrencies ETH and USDT. The first provides high profitability, the second guarantees low risk. Other cryptocurrencies play the role of extras and do not participate in the formation of the portfolio.

Table 9. Set of optimal portfolios (the risk measure, based on the Laplace distribution).

BTC	ETH	XRP	USDT	BCH	LTC	Risk, %	Return, %
0.000	0.027	0.000	0.973	0.000	0.000	0.768	0.012
0.000	0.123	0.000	0.877	0.000	0.000	1.000	0.043
0.000	0.221	0.000	0.779	0.000	0.000	1.500	0.075
0.000	0.305	0.000	0.695	0.000	0.000	2.000	0.103
0.000	0.385	0.000	0.615	0.000	0.000	2.500	0.129
0.000	0.463	0.000	0.537	0.000	0.000	3.000	0.155
0.000	0.540	0.000	0.460	0.000	0.000	3.500	0.180
0.000	0.617	0.000	0.383	0.000	0.000	4.000	0.205
0.000	0.694	0.000	0.306	0.000	0.000	4.500	0.230
0.000	0.770	0.000	0.230	0.000	0.000	5.000	0.255
0.000	0.846	0.000	0.154	0.000	0.000	5.500	0.280
0.000	0.922	0.000	0.078	0.000	0.000	6.000	0.305
0.000	1.000	0.000	0.000	0.000	0.000	6.518	0.331

To increase profitability, the portfolio can include shares of well-known companies. We will introduce Amazon stocks into the previous portfolio instead of the low-yield cryptocurrency LTC. Similar to the above, we obtained the set of optimal portfolios presented in table 10. The main role in the formation of the portfolio is played by stocks Amazon and cryptocurrency USDT. The stocks provide high profitability, the cryptocurrency guarantees low risk. The introduction of Amazon's stock to the portfolio halved the portfolio's risk and doubled its profitability. Thus, we conclude that optimal portfolios should be built by combining cryptocurrencies and stocks of highly profitable stable companies. The sets of optimal portfolios presented in tables 9 and 10 are illustrated in fig. 7 and fig. 8.

Table 10. Set of optimal portfolios (the risk measure, based on the Laplace distribution).

BTC	ETH	XRP	USDT	BCH	LTC	Risk, %	Return, %
0.000	0.024	0.000	0.920	0.000	0.056	0.741	0.052
0.000	0.038	0.000	0.723	0.000	0.239	1.000	0.191
0.000	0.045	0.000	0.624	0.000	0.330	1.250	0.260
0.000	0.052	0.000	0.537	0.000	0.411	1.500	0.322
0.000	0.057	0.000	0.455	0.000	0.487	1.750	0.380
0.000	0.063	0.000	0.375	0.000	0.561	2.000	0.436
0.000	0.069	0.000	0.297	0.000	0.634	2.250	0.491
0.000	0.075	0.000	0.220	0.000	0.705	2.500	0.546
0.000	0.080	0.000	0.144	0.000	0.776	2.750	0.600
0.000	0.086	0.000	0.068	0.000	0.846	3.000	0.653
0.000	0.078	0.000	0.000	0.000	0.922	3.250	0.706
0.000	0.000	0.000	0.000	0.000	1.000	3.458	0.738

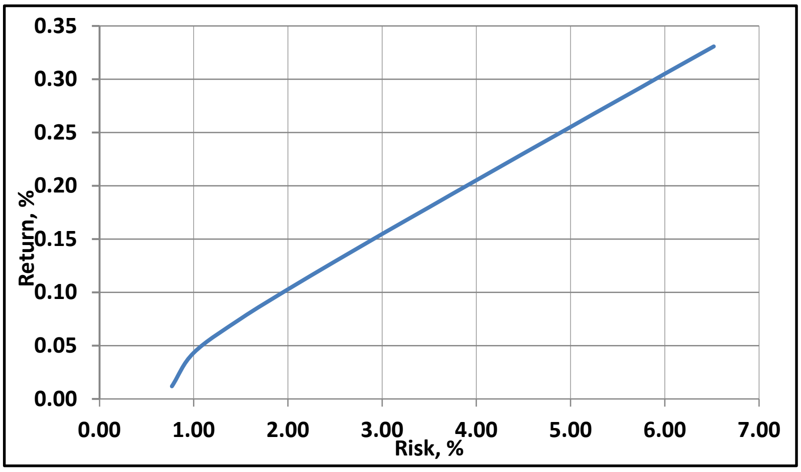


Fig. 7. Set of optimal cryptocurrency portfolios (table 9).

6 Conclusion

Due to its volatility, cryptocurrencies are an attractive tool for short-term investments. However, high volatility is a source of great risk. For assessing of cryptocurrencies risk, it is necessary to identify the return distribution. Numerous studies show that cryptocurrencies return and stocks return are not subject to normal distribution. The aim of our research is to compare the application of the Cauchy distribution and the Laplace distribution to the description of the actual distribution of cryptocurrency yields. A comparison of the actual return frequency in the critically low zone with its theoretical value was used as an evaluation criterion. Calculations performed for six cryptocurrencies over a 30-month period showed that the Cauchy distribution describes well the return distribution in the central part, but greatly overestimates the probability of marginal values of return. In our opinion, using the Laplace distribution is the most adequate approach to measuring the risk of cryptocurrencies (stocks).

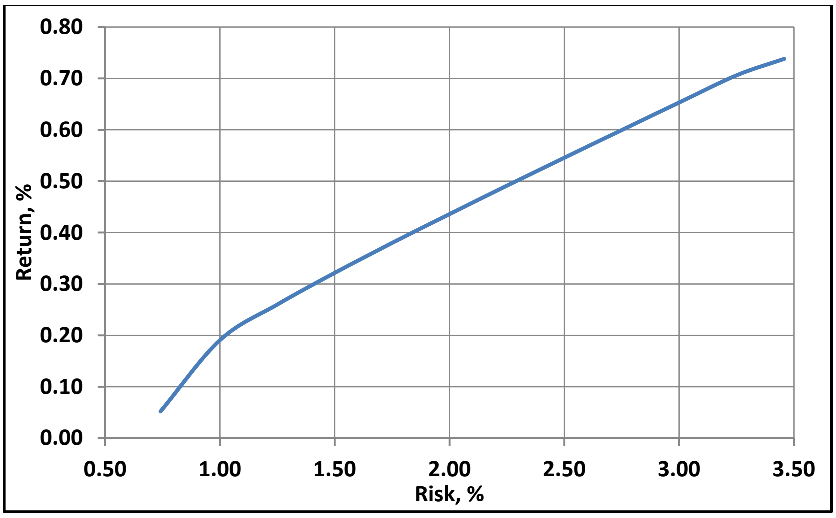


Fig. 8. Set of optimal combined portfolios (table 10).

A comparison of cryptocurrencies returns with the stocks return of leading companies showed that the average quarterly return of cryptocurrencies is low. Thus, it can be concluded that stocks are more attractive for long-term investments. Cryptocurrencies are a tool for speculative trans-actions. Inclusion of stocks of high-yield companies in the cryptocurrency's portfolio allows in-creasing portfolio profitability and reducing portfolio risk. We have shown that inclusion AMZN stocks into the cryptocurrency portfolio can double the portfolio's yield and halve its risk.

References

1. Andrianto, Y., Diputra, Y.: The Effect of Cryptocurrency on Investment Portfolio Effectiveness. J. Fin. Accoun. **5**(6), 229–238 (2017). doi:10.11648/j.jfa.20170506.14
2. Aparicio, F., Estrada, J.: Empirical Distributions of Stock Returns: Scandinavian Securities Markets, 1990-95. Eur. J. Fin. **7**, 1–21 (2001)
3. Babenko, V., Nazarenko, O., Nazarenko, I., Mandych, O.: Aspects of program control over technological innovations with consideration of risks. East.-Eur. J. Enterp. Technol. **3/4** (93), 6–14 (2018). doi:10.15587/1729-4061.2018.133603
4. Balcilar, M., Bourid, E., Guptac, R., Roubaud, D.: Can volume predict Bitcoin returns and volatility? A quantiles-based approach. Econ. Model. **64**, 74–81 (2017). doi:10.1016/j.econmod.2017.03.019

5. Bielinskyi, A.O., Soloviev, V.N.: Complex network precursors of crashes and critical events in the cryptocurrency market. CEUR Workshop Proceedings **2292**, 37–45 (2018)
6. Borri, N.: Conditional tail-risk in cryptocurrency markets. J. Empir. Fin. **50**, 1–19 (2019). doi:10.1016/j.jempfin.2018.11.002
7. Bouria, E., Guptab, R., Tiwaric, A. K., Roubaudd, D.: Does Bitcoin hedge global uncertainty? Evidence from wavelet-based quantile-in-quantile regressions. Fin. Res. Let. **23**, 87–95 (2017). doi:10.1016/j.frl.2017.02.009
8. Brauneis, A., Mestel, R.: Cryptocurrency-portfolios in a mean-variance framework. Fin. Res. Let. **28**, 259–264 (2019). doi:10.1016/j.frl.2018.05.008
9. Chan, S., Chu, J., Nadarajah, S., Osterrieder, J.: A Statistical Analysis of Cryptocurrencies. J. Risk Fin. Manag. **10**(2), 12 (2017). doi:10.3390/jrfm10020012
10. Ciaian, P., Rajcaniova, M., Kancs d'A.: The Economics of BitCoin Price Formation. Appl. Econ. **48**(19), 1799–1815 (2016). doi:10.1080/00036846.2015.1109038
11. Corbet, S., Meegan, A., Larkin, C., Lucey, B., Yarovaya, L.: Exploring the dynamic relationships between cryptocurrencies and other financial assets. Econ. Let. **165**, 28–34 (2018). doi:10.1016/j.econlet.2018.01.004
12. Derbentsev, V., Matviychuk, A., Soloviev, V.N.: Forecasting of Cryptocurrency Prices Using Machine Learning. In: Pichl, L., Eom, C., Scalas, E., Kaizoji, T. (eds.) Advanced Studies of Financial Technologies and Cryptocurrency Markets, pp. 211-231. Springer, Singapore (2020). doi:10.1007/978-981-15-4498-9_12
13. Fama, E.: The behavior of stock market prices. J. Bus. **38(**1), 34–105 (1965)
14. Hrytsiuk, P., Babych, T., Bachyshyna, L.: Cryptocurrency portfolio optimization using Value-at-Risk measure. Advances in Economics, Business and Management Research **95**, 385–389 (2019)
15. Khokhlov, V.Yu.: VaR i problema "bolshih hvostov" raspredeleniya dohodnosti (VaR and the problem of "large tails" of the profitability distribution). Risk-menedzhment v kreditnoy organizatsii **2**, 35–49 (2012)
16. Linden, M.: A model for stock returns distribution. Int. J. Fin. Econ. **6**, 159–169 (2001)
17. Markowitz, H.: Portfolio Selection. J. Fin. **7(**1), 77–91 (1952). doi:10.1111/j.1540-6261.1952.tb01525.x
18. Soloviev, V., Belinskij, A.: Methods of nonlinear dynamics and the construction of cryptocurrency crisis phenomena precursors. CEUR Workshop Proceedings **2104**, 116–127 (2018)
19. Soloviev, V.N., Belinskiy, A.: Complex Systems Theory and Crashes of Cryptocurrency Market. Communications in Computer and Information Science **1007**, 276–297 (2019). doi:10.1007/978-3-030-13929-2_14
20. Soloviev, V.N., Yevtushenko, S.P., Batareyev, V.V.: Comparative analysis of the cryptocurrency and the stock markets using the Random Matrix Theory. CEUR Workshop Proceedings **2546**, 87–100 (2019)
21. Taleb, N.: The Black Swan: The Impact of the Highly Improbable. Random House, New York (2007)
22. Van der Merwe, A.: An economic assessment of risks in Bitcoin as an alternative asset class. In: Hanke, S.H. (ed.) Studies in Applied Finance, vol. 27. J. Hopkins Ins. App. Econ., Glob. Health and St. Bus. Ent. (2019)
23. Yermack, D.: Is Bitcoin a Real Currency? An economic appraisal. In: Chuen, D.L.K. (eds.) Handbook of digital currency, pp. 31–43. Academic Press, Cambridge (2015)
24. Zabolotskyy, T.: Modeluvannya v upravlinni portfelem finansovyh aktyviv (Modeling in the management of the portfolio of financial assets). LNU, Lviv (2016)

Model for assessing and implementing resource-efficient strategy of industry

Nadiia Shmygol[1[0000-0001-5932-6580]], Francesco Schiavone[2[0000-0001-9219-6714]], Olena Trokhymets[3[0000-0001-7587-7948]], Dariusz Pawliszczy[4[0000-0003-1328-7891]], Viktor Koval[5[0000-0003-2562-4373]], Ruslan Zavgorodniy[3[0000-0002-6137-1310]] and Andrii Vorfolomeiev[6[0000-0001-5789-5149]]

[1] National University "Zaporizhzhia Polytehnic", 64 Zhukovsky Str., Zaporizhzhia, 69063, Ukraine
[2] Parthenope University of Naples, 38 Via Ammiraglio Ferdinando Acton, 80133 Napoli, Italy
[3] Classic Private University, 70B Zhukovsky Str., Zaporizhzhia, 69602, Ukraine
[4] Gromadka Commune Office, 9 General Wł. Sikorskiy Str., Gromadka, 59-706, Poland
[5] Kyiv National University of Trade and Economics, 19 Kyoto Str., Kyiv, 02156, Ukraine
[6] National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute", 37 Peremohy Ave., Kyiv, 03056, Ukraine
nadezdash@ua.fm

Abstract. The authors of the article determined that a number of scientists were involved in the development of a balanced system of indicators of the development of the oil and gas sector. Though an urgent scientific problem that needs further consideration is the development of a model of resource efficiency diagnostics in the oil and gas sector of the economy of Ukraine, taking into account the peculiarities of statistical monitoring. The scientific novelty of the paper is: this study improved the model of diagnostics of resource efficiency in oil and gas sector in the economy of Ukraine based on the additive-multiplicative compression of the formed system, which, unlike the existing ones, takes into account their variation while defining weighting coefficients which show the experts' system of preferences. It is reasonable to use the proposed model at the further economic assessment of the consequences of realization of resource-efficient strategy at enterprises of the oil and gas sector of the economy of Ukraine.

Keywords: resource-efficient strategy, oil and gas complex, model of assessment.

1 Introduction

The concept of resource efficiency in the modern and current practice of economic activity analysis has been widespread, since the efficient consumption of economic resources of any kind is associated, first of all, with intensive economic growth. That is why a lot of scientists addressed the issues of ensuring a resource-efficient economy as a necessary condition for sustainable development. According to 2018 the oil and gas sector of Ukraine provided more than 40% of the economy's needs for energy resources. At the same time oil and gas sector accounted for 42.2% of total consumption. Due to high dependency, the research of many domestic scientists is devoted to various aspects of the operation of oil and gas extraction and processing enterprises.

2 Actual scientific researches and issues analysis and the research objective

Management of any economic system is always based on its current state, the definition of which is a separate scientific task. The main tendencies of development and value of innovative technologies in the oil and gas sector were studied by foreign scientists, the following scientists are among them: Adi Karev [8], Konstantin N. Milovidov [12] and others. Various scientists were involved in the methodological bases of economic diagnostics, including the development of a balanced system of indicators for the development of the oil and gas sector, including Inesa Khvostina [9], T. F. Mantserova [10], Dani Rodrik [18], O. A. Tolpegina [26] and others.

This demonstrates that the scientific problem that needs to be solved in the framework of this research is the development of a model of resource efficiency diagnostics in the oil and gas sector of the economy of Ukraine, taking into account the existing peculiarities of statistical monitoring by the State statistical authorities.

The methods of economic analysis for estimation of resource efficiency, normalization method for bringing indicators to comparative appearance, method of additive-multiplicative convolution (compression) for generalization of results in different directions of evaluation, statistical methods of estimation of variation for substantiation of values of weight coefficients in the model of diagnostics were used.

3 Tools and models for effective development of resource efficiency in the oil and gas sector of Ukraine

In order to develop action for effective development of the resource-efficiency in the oil and gas sector of Ukraine we use the following [6; 25]:

- a system of indicators for assessing the resource efficiency of the oil and gas sector. Taking into account peculiarities of the measurement of the studied object by the State Statistics Service of Ukraine, the effective development of resource efficiency in any field of activity requires consideration in the analysis of all types of economic resources: raw materials, fixed assets, labor resources, total capital (aggregate capital);
- a model of the index of resource efficiency IRE based on the additive-multiplicative convolution (compression). Convolution of indicators is carried out by weighing their normalized o standard values on the basis of an agreed system of expert preferences. In this case, it is believed that individual indicators with equal

level of influence on the group should have the same values of the root-mean-square deviates;

- intersectoral comparative analysis of resource intensity (resource capacity) and structure of added value in the oil and gas sector. The need for this analysis is due to the fact that to diagnose the current state and efficiency of the enterprises' activity, it is appropriate to use the relevant base of comparison in economic analysis;
- scenario analysis of price equilibrium in oil and gas sector with the help of intersectoral Leontiev's model. It allows not only to perform appropriate calculations, but also to find out how these changes affect resource efficiency, including by changing the ratios of direct costs, intermediate consumption, added value and gross profit;
- assessment of the consequences of resource efficiency based on the IRE model. This allows getting recommendations on areas and mechanisms to ensure resource efficiency in the production and processing of oil and gas.

The peculiarities of the measurement of resource efficiency in oil and gas sector of economy of Ukraine include:

1. The available volume of input statistics on the basis of the State Statistics Service of Ukraine reports, with free access, significantly limits the possibilities for comprehensive assessment of the resource efficiency by all types of economic resources used in public production;
2. The change in methodology of organization of statistical observation during the recent years, and, accordingly, reporting instruments and documentation do not allow to carry out a retrospective analysis of resource efficiency indicators over a long-term period. The geopolitical changes that occurred in 2014 in the South-East of Ukraine resulted in temporary occupation of the Crimea and parts of Donetsk and Luhansk Regions, have undoubtedly had a significant impact on the oil and gas sector activity as well. That is why comparative analysis of the time periods cannot provide with objective information on the dynamics of the target indicators due to alterations in the special aggregate according to which they are calculated;
3. Some of the input data for 2016-2018 are not shown in the State Statistic Service reports due to its confidential status. First of all, it concerns assets conditions, volume of production and corresponding costs, fixed assets and a number of employees, financial results of crude oil production, natural gas extraction and production of refine products [6; 25].
4. Any diagnostics in economic analysis is possible if there is a respective base of comparisons. In scoring models of diagnostics such a turning point are the classes of indicators stability; in the models of multiplicative discriminant analysis – the intervals of stability of integral index that determine the probability of the bankruptcy of economic entities; in the express-analysis – industry standards and cross-industry comparisons; in complex analysis – dynamics and plan value of indicators, industry standards etc. As for the oil and gas sector of the economy, for the diagnostics of its resource efficiency considering available data, we will use

cross-industry comparison and analysis of time periods applying methods of statistic theory.

4 The model of resource efficiency in oil and gas sector of economy of Ukraine diagnostics

Thus, taking into account the leading experience of analysis of economic activity [1; 14; 15; 21; 22; 24; 28] and mentioned above peculiarities of information support, a model of diagnostics of resource efficiency of oil and gas sector of the economy of Ukraine has a set of indicators as its basis, which consist of the following areas of assessment: material resources, fixed assets, labor resources and aggregate capital. Let's consider them in more detail.

1. Material resources (MR_1). Technological underdevelopment (backwardness), associated with initial processing of resources, is always characterized by low added value and high material (output) ratio. That is why effective use of material resources is the priority in the development not only of oil and gas sector, but of the economy of Ukraine. This group consists of the following indicators:

- material productivity (K_{11}) – characterizes the volume of output of the inquiry period by 1 UAH of material costs. This indicator should be maximized and is calculated by the formula:

$$K_{11} = \frac{VO_1}{MC_1}, \tag{1}$$

where VO_1, MC_1 – accordingly, volume of output and material costs in the inquiry period.

- net profit (income) for 1 UAH of material costs (K_{12}), should be maximized. According to its economic essence, this indicator is the analogue of cost effectiveness (profitability), which allows to evaluate the efficiency of raw materials and supplies in the process of profit generation in the enterprises of the industry:

$$K_{12} = \frac{NP_1}{MC_1}, \tag{2}$$

where NP_1 – net profit in the inquiry period.

- coefficient of correlation of the growth rate of product output and material costs (K_{13}). Intensive economic development involves obtaining the final result not due to the greater consumption of resource productivity. That is why this coefficient should be $K_{13} > 1$.

$$K_{13} = \frac{VPO_1}{VPO_0} : \frac{MC_1}{MC_0}, \tag{3}$$

where VPO_0, MC_0 – accordingly, volume of product output and material costs in base period.

- the share of material costs in the cost of production (K_{14}). According to 2018, material costs for the economy in general were 74.3% from the cumulative costs (total costs) for production output. Accordingly, depreciation accounted for 6.7%, labor costs – 14.1%, benefits related deduction – 2.9%, and other costs – 2.0% from cumulative costs (total costs).

As we can see, the high share of material costs – is a system problem for the entire economy of Ukraine. It indicates not only the low level of social production, but also hinders increase in wages and living standards of the population. That is why this indicator should be minimized and calculated by the formula:

$$K_{14} = \frac{MC_1}{VCCP_1}, \tag{4}$$

where $VCCP_1$ – volume of cumulative costs (total costs) for production in the inquiry period.

The main production factors which are part of economic resources, are fixed assets and labor resources (human capital). In most cases they determine the production capacity of business entities and industries of the economy in general.

According to the results of 2018, the residual value of fixed assets in Ukraine was 3783.5 billion UAH, and the volume of production – 6207.7 billion UAH. Accordingly, return on assets was 1.64 UAH. The number of employed population for the same period was 16360.9 thousand persons. Thus, the annual labor productivity was 379.4 thousand UAH per employee or 31.6 thousand UAH monthly.

Thus, complex diagnostics of the resource efficiency of oil and gas sector should include comparative assessment in these areas.

2. Fixed assets (K_2). This group includes the following indicators:

- return on assets (K_{21}) – characterizes the volume of production output for the inquiry period at the rate of 1 UAH of residual value of fixed assets, and should be maximized:

$$K_{21} = \frac{VO_1}{FA_1}, \tag{5}$$

where FA_1 – the value of fixed assets in the inquiry period.

- return on assets (K_{22}) – equals the net profit on 1 UAH of residual value of fixed assets, and it should be maximized:

$$K_{22} = \frac{NP_1}{FA_1}, \tag{6}$$

- coefficient of correlation of the growth rate of product output and fixed assets costs (K_{23}). Intensive development implies an increase in aggregate production output not at the expense of additional production capacity attraction, but due to the return on assets increase. That is why this coefficient should have the inequality $K_{23} > 1$.

$$K_{23} = \frac{VO_1}{VO_0} : \frac{FA_1}{FA_0}, \tag{7}$$

where FA_0 – residual value of fixed assets in base period.

3. Labor resources (K_3). The indicators of resource efficiency of this group include:

– labor productivity (K_{31}) – characterizes the production output for the inquiry period per one employee and should be maximized:

$$K_{31} = \frac{VO_1}{AAEP_1}, \tag{8}$$

where $AAEP_1$ – average annual number of employed population in the inquiry period.

– ROI of employees (K_{32}) – equals net profit per one employee, and should be maximized:

$$K_{32} = \frac{NP_1}{AAEP_1}, \tag{9}$$

– share of labor costs in the cost of production (K_{33}). According to statistics, in most of Eurozone countries this indicator is 30-35%, which is more than 2 times ahead of the similar level of the economy of Ukraine. That is why one of the reserves for the growth of the average level of remuneration of labor is adjustment of the production cost structure, and should be maximized K_{33}:

$$K_{33} = \frac{RL_1}{VCCP_1}, \tag{10}$$

where RL_1 – amount of remuneration of labor cost in the inquiry period.

Aggregate capital is generated from both equity and borrowed sources and is allocated to fixed assets and current assets and is also an economic resource and a focus of the researches interest in terms of its effective use.

4. Aggregate capital (total capital) (K_4). In order to characterize the efficiency of capital use, in the practice of financial analysis, along with profitability indicators, indicators of turnover and duration of turnover are calculated. Let's consider them in more detail.

– aggregate capital (total capital) turnover (K_{41}) – shows how many the income of the inquiry period exceeds the corresponding amount of the raised total capital The increase in turnover shows an increase of its use:

$$K_{41} = \frac{CI_1}{CK_1}, \tag{11}$$

where CI_1 – cumulative income of the inquiry period from all types of economic activity; CK_1 – average annual amount of capital of the inquiry period, taking into account own and borrowed sources of income.

- return on aggregate capital (total capital) (K_{42}). Any borrowed capital, involved in the activity of business entities, has its price. The condition of the expediency of its use is always the excess of return on aggregate capital (total capital) over the weighted average price of the loan. Otherwise, according to financial leverage effect, economic activity will lead to a gradual decrease in equity.

$$K_{42} = \frac{BP_1}{CK_1}, \tag{12}$$

where BP_1 – balance (gross) profit of the inquiry period, excluding income tax.

- duration of circulation of aggregate (total) capital (K_{43}) – shows how many days it will take for the income received during economic activity to be equal to the amount of attracted aggregate (total) capital Speeding up the turnover means reduction of the duration of circulation and vice versa. The formula for K_{43} calculation is the following:

$$K_{43} = \frac{365}{K_{41}}, \tag{13}$$

In the numerator, in this case, there is a number of days for the inquiry period.

Fixed assets form production capacity of the economic entities and do not directly participate in the circulation. The efficiency of the use of aggregate (total) capital is directly influenced by the turnover of the operating capital according to the formula:

$$K_{44} = \frac{CI_1}{OC_1}, \tag{14}$$

where OC_1 – average annual amount of the operating capital in the inquiry period.

- duration of operation capital turnover (K_{45}) – shows how many days it will take for the received income to be equal to the amount of operating capital and is calculated by the formula:

$$K_{45} = K_{43} \times \frac{OC_1}{CK_1}, \tag{15}$$

Thus, we have formed a system of indicators for assessing the resource efficiency of oil and gas sector of the economy of Ukraine taking into account available statistics. Taking into account that all the indicators are relative indicators we will use cross-industry comparisons for diagnostics of its condition.

By direct comparison we have an opportunity to define competitive advantages or backlog of the oil and gas sector by every indicator. However, summarizing the results of such multifactor evaluation requires the corresponding compression based on the integrated index. For this reason, first of all, it is necessary to bring the value of all indicator of resource efficiency to one base of comparison, which means to normalize them. The current practice of rationing involves setting up values to the range [0, 1] using formula:

$$K' = \frac{K - K_w}{K_b - K_w}, \tag{16}$$

where K, K' – accordingly, input and normalized value of resource efficiency indicator, which belong to i group; K_w, K_b – accordingly, the worst and the best value of the indicator K, among other industries.

Since there are some indicators that should be maximized as well as minimized, then to determine the worst indicators K_w and the best indicators K_b we should follow the rule:

– if K should be maximized, then $K_b = max(K)$, $K_w = min(K)$;
– if K should be minimized, then $K_b = min(K)$, $K_w = max(K)$.

The use of formula (16), observing the rule, allows to arrange the normalized values of indicators in such a way that the best value of indicator corresponds with the normalized and vice versa.

With its help, each of the indicators (1) - (15) is reduced to a comparative form. The compression of normalized values to group and integral indexes is based on the additive-multiplicative model:

$$IPE = \sum_{i=1}^{n}(a_i \times K_i),\ K_i = \sum_{j=1}^{m_i}\left(a_{ij} \times K'_{ij}\right) \tag{17}$$

for all $i = 1...n$, where IPE – integral index of resource efficiency; K_i, a_i – accordingly, summary (consolidated) index of resource efficiency of i group and its weighing coefficient; K'_{ij}, a_{ij} – accordingly, normalized j indicator of i group and its weighing coefficient; n – a number of indicator groups; m_i – a number of indicators of i group.

There are certain limitations for weighing coefficients a_i and a_{ij}. First of all, their values should range from 0 to 1; second of all, the sum of coefficients of a certain group should equal 1.

Considering the mentioned above information, we have obtained a system of equations using numerical method for diagnostics of resource efficiency in oil and gas sector in the economy of Ukraine, taking into account equal influence of indicators, which allowed presenting a more detailed equation (18):

$$\begin{gathered} IPE = 0.328K_1 + 0.261K_2 + 0.244K_3 + 0.167K_4, \\ K_1 = 0.162K_{11} + 0.267K_{12} + 0.452K_{13} + 0.119K_{14}, \\ K_2 = 0.325K_{21} + 0.318K_{22} + 0.358K_{23}, \\ K_3 = 0.290K_{31} + 0.354K_{32} + 0.356K_{33}, \\ K_4 = 0.183K_{41} + 0.241K_{42} + 0.191K_{43} + 0.210K_{44} + 0.174K_{45}. \end{gathered} \tag{18}$$

5 Diagnosis of resource efficiency of the oil and gas sector of Ukraine taking into account the opinions of experts

If, according to the experts' preferences, individual indicators should influence differently on the group or integral index of resource efficiency, this also should be reflected in the proportions between root-mean-square deviants of such indicators considering corrective weighing coefficients.

In this case, there is a need for quantitative coordination of expert judgments, based on qualitative initial assessments. Therefore, the sequence of actions, taking into account the theory of decision-making, should be the following [24]:

1. Each of the experts, based on their personal system of preferences, organizes the sequence of evaluation of the components of the index of resource efficiency according to their importance.
2. On the basis of individual rankings of the experts, with the help of the methods of arithmetic mean ranks calculation of the generalized group ranking is carried out.
3. Verification of the consistency of the results of individual assessments of experts is performed using the variance (dispersion) coefficient of concordance.
4. If at the previous stage the verification was successfully passed, on the basis of application of procedure of pair comparison for each direction of an estimation of an index of resource efficiency, the calculation of correction factors is carried out. If the concordance coefficient indicates a high inconsistency of experts' opinions, the procedure for adjusting the parameters of equations (18) should be interrupted to find out the reasons for such inconsistency.
5. Based on the calculation of the correcting coefficients, the parameters of equations (18) are changed in order to consider the agreed and confirmed experts' opinion on the importance of areas for assessing the components of resource efficiency.

Methodical support of the planned sequence of actions should be considered in more detail. In particular, in order to decide on the correction of the parameters of equations (18), according the experts' estimation, the following are used: the method of arithmetic mean ranks, the variance (dispersion) coefficient of concordance and the method of pairwise comparison [24; 27].

To apply the arithmetic mean method, each of the experts makes individual rankings regarding the weight of the factors that affect the target coefficient. Moreover, the most important factors have the lowest rank and vice versa. Let's indicate the corresponding set of matrixes as following:

$$[r_{is}]\ _{md},\ \left(s = \overline{1, d}\ ;\ i = \overline{1, m}\right), \tag{19}$$

where m – the number of factors by which the expert assessment is conducted; d – the number of experts; r_{is} – the ranks of the i factor, which was given by the s expert.

Next, for each factor, the sum of the ranks assigned by the experts is calculated and divided by their number. Thus, the arithmetic mean simple is calculated. The weighted average can be used if the experts have different levels of competence.

Generalized group ranking $[R_i]\ _m$ is obtained on the basis of the calculated arithmetic means.

A measure of consistency of the experts' estimations is the variance (dispersion) coefficient of concordance *W*. Depending on the nature of the input data, its calculation is carried out as following:

— if individual expert assessments do not contain related ranks:

$$W = \frac{12}{d^2(m^3 - m)} \times S, \tag{20}$$

$$S = \sum_{i=1}^{m} \left(\sum_{s=1}^{d} r_{is} - \overline{r}\right)^2, \tag{21}$$

$$\overline{r} = \frac{1}{m} \sum_{i=1}^{m} \sum_{s=1}^{d} r_{is}, \tag{22}$$

— if individual expert assessments contain related ranks:

$$W = \frac{12S}{d^2(m^3 - m) - d\sum_{s=1}^{d} T_s}, \tag{23}$$

$$T_s = \sum_{k=1}^{H_s} \left(h_k^3 - h_k\right), \tag{24}$$

where T_s – indicator of the related ranks of the *s* expert; H_s – the number of groups of equal rank in the assessment of the *s* expert; h_k – the number of ranks equal to each other of the *k* group of related ranks of the s expert.

It's necessary to mention, that the formula (20) is a partial or a finite case (23). If the expert assessemnts do not contain related ranks, then we will have: $H_s = 0$; $h_k = 0$; $T_s = 0$. Accordingly, (23) is transformed into (20).

The variance (dispersion) coefficient of concordance changes within $0 \leq W \leq 1$. If *W*=1, then all individual rankings of experts are similar to each other and vice versa. The following scale is used to interpret its values: $W \in [0, 0.3]$ – the level of consistency of expert assessments is very weak; $W \in [0.3, 0.5)$ – weak; $W \in [0.5, 0.7)$ – average (moderate, medium); $W \in [0.7, 0.9)$ – high; $W \in [0.9, 1]$ – very high.

High inconsistency of experts assessments, if $W < 0.7$, it may indicate a low level of competence of individual members of the group, or a low awareness of this issue. In this case, after additional study of the problem situation, it is necessary to repeat the expert survey (questionary).

We use the method of pairwise comparison to define adjusted coefficients on the basis of a generalized group ranking $[R_i]\ _m$ in the case if $W \geq 0.7$. The elements of the matrix of pairwise comparisons s$C = \left[C_{ij}\right]_{mm}$ are determined on the basis of the rule:

$$\begin{cases} C_{ij} = 2, & if \quad R_i \succ R_j \\ C_{ij} = 1, & if \quad R_i \approx R_j, \\ C_{ij} = 0, & if \quad R_i \prec R_j \end{cases} \tag{25}$$

Then, the adjusted coefficients of the parameters of equations (25) are calculated by the formula:

$$g_i = \frac{\sum_{j=1}^{m} c_{ij}}{\sum_{i=1}^{m} \sum_{j=1}^{m} c_{ij}}, \qquad (26)$$

Accordingly, the adjusted weighting coefficients of the integrated resource efficiency index, taking into account the agreed opinion of the expert group, should satisfy the ratio: $\sigma(a_i K_i)/\sigma(a_j K_j) = g_i/g_j$, for all $i \neq j$, or $\sigma(a_{ij} K'_{ij})/\sigma(a_{iz} K'_{iz}) = g_{ij}/g_{iz}$, for all i = 1, …, n, $j \neq z$. That is, the ratio between the standard deviations of each weighted factor should be equal to the corresponding ratio between the adjusted coefficients of the generalized ranking of experts.

Within the framework of this study, the expert group consisted of three experts who expressed their opinion on the importance of the components of the resource efficiency index of the oil and gas sector of Ukraine. The difficulty of objectively assessing individual preferences for the efficient use of material resources, fixed assets, labor resources and total capital was due to the crucial role of each component in the formation of the target indicator. That is why, first of all, it was decided to perform an expert assessment based on the existing advantages and disadvantages in the resource efficiency of oil and gas sector enterprises, compared to other industries and the economy of Ukraine in general [24]. And since the enterprises of the oil and gas sector are part of both the extractive and processing industries, it is advisable to make intersectoral comparisons with them [5; 13; 16; 17].

Taking into account the developed model (18), the results of diagnostics of resource efficiency for 2015-2018 are presented in table 1.

Table 1. The results of diagnostics of the resource efficiency of oil and gas sector of the economy of Ukraine according of data of 2015-2018 years.

Industries of the economy	**Years**	**P_1**	**P_2**	**P_3**	**P_4**	**IPE**
Total	2015	0.293	0.345	0.260	0.306	0.301
	2018	0.316	0.388	0.334	0.466	0.364
Industry, including:	2015	0.281	0.360	0.233	0.540	0.333
	2018	0.297	0.435	0.327	0.661	0.401
Mining industry (primary sector) and quarrying, including:	2015	0.300	0.382	0.283	0.483	0.348
	2018	0.391	0.468	0.500	0.670	0.484
Crude oil and natural gas production*, including:	2015	0.372	0.406	0.266	0.666	0.404
	2018	0.667	0.468	0.786	0.751	0.658
Crude oil production*	2015	0.276	0.252	0.252	0.500	0.301
Extraction of natural gas*	2015	0.448	0.548	0.320	0.785	0.499
Provision of ancillary services in the field of oil and natural gas*	2015	0.306	0.252	0.345	0.577	0.347
	2018	0.300	0.735	0.395	0.861	0.530
Processing industry, including:	2015	0.272	0.387	0.213	0.609	0.344
	2018	0.285	0.542	0.300	0.748	0.433
Production of oil processing*	2015	0.255	0.503	0.223	0.733	0.392
Gas production, distribution of gaseous fuel through local pipelines*	2015	0.312	0.593	0.366	0.190	0.378
	2018	0.243	0.285	0.350	0.195	0.272

In the table 1 the asterisk symbol marks the types of economic activity which are a part of oil and gas sector. As for the separate crude oil production and natural gas extraction, and refined products production in 2018 as well, the access to the relevant statistics is limited due to their confidential character.

Special qualitative changes have taken place in the consumption of raw materials and the use of labor resources. The result of such changes was that the oil and gas sector began to outrun the extractive and all industries, as well as the average level of Ukraine's economy in terms of resource efficiency. Thus, on the one hand, we had a positive trend of increasing resource efficiency [3; 4; 11; 29]. On the other hand, it was achieved by a significant increase in product prices in recent years.

Regarding the production of oil and gas products, as well as the gas distribution system, it can be seen that in terms of the use of fixed assets and capital there is a significant lag behind other enterprises of the processing industry and the average level in the economy [2; 19; 30].

Thus, the diagnosis of resource efficiency of the oil and gas sector indicated the existing problems at refineries and significant improvements in oil and gas production [7; 20; 23]. That is why, in the formation of individual preferences, experts proceeded from the most important problems of resource efficiency in enterprises for the production of refined products and gas, table 1.

The system of preferences or advantages of each of the experts had the form:

- the first expert – $K_1 \approx K_4 \succ K_2 \approx K_3$, which means the equivalence of indicators of efficiency of use of material resources and total capital due to their importance, as refineries in the oil and gas sector have the biggest problems in these areas of assessment. Therefore, these groups of indicators are more important than the efficiency of use of fixed assets and labor resources, which are also equivalent to each other;
- the second expert – $K_1 \succ K_2 \approx K_3 \approx K_4$, that is, the problem of ensuring the efficient use of material resources, taking into account the current situation, outweighs other areas of assessment that are equivalent to each other;
- the third expert – $K_1 \succ K_4 \succ K_2 \approx K_3$. In contrast to the first system of advantages or preferences, the group of indicators K_4 is inferior to K_1 in terms of importance.

The results of the calculation of the generalized group ranking by the method of arithmetic mean ranks, taking into account the individual preferences of experts, are given in table 2. As we see, it completely coincides with the assessment of the third expert.

Table 2. The results of the calculation of generalized group ranking by the method of arithmetic mean ranks.

Group of indicators	**Individual ranking**			**Arithmetic mean ranks**	**Generalized group ranking**
	Expert I	**Expert II**	**Expert III**		
K_1	1.5	1	1	1.167	1
K_2	3.5	3	3.5	3.333	3.5
K_3	3.5	3	3.5	3.333	3.5
K_4	1.5	3	2	2.167	2

In order to use the generalized group ranking in further calculations, we will assess the consistency of experts' opinions using the variance coefficient of concordance W. Since their individual rankings had related ranks, the calculation of W was performed by the formula (23):

$$W = \frac{12S}{d^2(m^3-m)-d\sum_{s=1}^{d}T_s} = \frac{12\times 29.5}{3^2\times(4^3-4)-3\times 42} = 0.855$$

Since W ∊ [0.7, 0.9), it can be stated that we have a high level of consistency of expert assessments. Therefore, generalized ranking can be used to calculate unknown parameters. To do this, the calculation of the adjusted coefficients g_i was performed by the method of pairwise comparisons, the results of which are shown in the table 3.

Table 3. The results of the calculation of the adjusted coefficients by the method of pairwise comparisons.

Groups of indicators	K₁	K₂	K₃	K₄	Sum	The adjusted coefficients g_i
K_1	1	2	3	4	7	0.438
K_2	0	1	1	0	2	0.125
K_3	0	1	1	0	2	0.125
K_4	0	2	2	1	5	0.313
Sum	–	–	–	–	16	–

As it was noticed before, the adjusted coefficients determine the ratio between the standard deviations of each weighted factor of the IRE model as follows: $\sigma(a_iK_i)/\sigma(a_jK_j) = g_i/g_j$ for all $i \neq j$. Considering this factor, unknown parameters of the equation were obtained by numerical methods:

$$IPE = 0.555K_1 + 0.126K_2 + 0.118K_3 + 0.201K_4 \quad (27)$$

In determining the weighing coefficients in the equation (27) the variation of each indicator was: $\sigma(a_1K_1) = 0.054$, $\sigma(a_2K_2) = 0.016$, $\sigma(a_3K_3) = 0.016$, $\sigma(a_1K_1) = 0{,}039$.

6 Measures to implement a resource-efficient strategy at the enterprises of the oil and gas sector of Ukraine

In order to develop further measures t_i implement a resource-efficiency strategy at the enterprises of oil and gas sector of Ukraine let's consider the components of the IRE index in more detail. Table 4 illustrates the results of the relevant calculations according to the year 2018 data and considering price adjustments in production and processing.

Table 4. The results of the calculation of controlled indicators of resource efficiency of the oil and gas sector of Ukraine

Indicators	Oil and gas extraction	Gas production and distribution

	2018	Price adjustment	2018	Price adjustment
Resource efficiency index IRE	0.673	0.502	0.251	0.422
Material resources K_1	0.667	0.500	0.243	0.368
K_{11}	0.809	0.639	0.107	0.171
K_{12}	1.000	0.680	0.430	0.624
K_{13}	0.353	0.231	0.159	0.293
K_{14}	0.922	0.928	0.329	0.345
Fixed assets K_2	0.468	0.291	0.285	0.641
K_{21}	0.177	0.125	0.617	0.767
K_{22}	0.989	0.675	0.000	0.769
K_{23}	0.270	0.101	0.239	0.414
Labour resources K_3	0.786	0.627	0.350	0.482
K_{31}	1.000	1.000	0.156	0.254
K_{32}	1.000	0.551	0.173	0.436
K_{33}	0.399	0.398	0.684	0.713
Aggregate capital K_4	0.751	0.567	0.195	0.399
K_{41}	0.358	0.204	0.133	0.251
K_{42}	1.000	0.749	0.408	0.799
K_{43}	0.750	0.543	0.000	0.057
K_{44}	0.656	0.468	0.346	0.534
K_{45}	0.932	0.843	0.000	0.216

As can be seen from table 4, in oil and gas production, after the scenario price adjustment, almost all indicators have decreased. The exceptions were labor productivity K_{31}, as well as the share of material costs and wages in the cost of production, K_{14} and K_{33}, respectively. In this case, we can distinguish 3 main groups:

1. Indicators with a slight deterioration in their values, which remain at a competitive level relative to other industries and the economy of Ukraine in general. This group includes: K_{11}, K_{14}, K_{31}, K_{33}, K_{42} and K_{45}. Their dynamics and condition do not raise concerns about possible problems in the future. Therefore, special attention should be paid to the indicators that are part of the following two groups when developing measures to optimize the use of resources.
2. Indicators that have significantly lost their positions, however, their values still remain high. These include: K_{12}, K_{22}, K_{32}, K_{43} and K_{44}.

The first three indicators in this list are related to the reduction of profits, in relation to the volume of use of material and labor resources, as well as fixed assets. This is an objective consequence of the necessary price adjustment, which revealed the real situation with resource efficiency in oil and gas production. The recommendation, in this case, may be to optimize the number of labor resources to increase productivity K_{32}, without reducing the cost of its payment.

The last two indicators characterize the slowdown in working capital due to reduced revenues from sales. The specificity of the oil and gas sector is the high capital intensity associated with the technological features of extraction, storage and transportation. That is why the growth reserves of K_{43} and K_{44} are limited.

3. Indicators with a low level of resource efficiency, compared to other industries and the economy in general: K_{13}, K_{21}, K_{23} and K_{41}.

The coefficients K_{13} and K_{23} characterize the ratio of growth rates of output with the consumption of material resources and the volume of fixed assets. After the implementation of the proposed price adjustment, these indicators will return to the level of the last reporting period, which is positive. Reserves for further growth of K_{13} are the introduction of new technologies, which requires significant capital investment and, in the current economic stagnation, is impossible. At the same time, the increase in K_{23} is directly related to the fullest possible utilization of available production capacity. Therefore, the restoration of positive dynamics in this area of assessment is possible in conditions of economic growth.

The low return on capital K_{21} and the turnover of total capital K_{41} are associated with a high share of non-current assets (fixed assets) in their total volume. For comparison, the average for the economy in 2018 it was 42.1%; in industry – 44.2%; in the mining industry – 53.4%; in oil and gas production – 67.9%; oil – 49.1%; gas – 70.4%. For this reason, our object of study is significantly inferior to other industries and the effective use of available current assets cannot correct the situation. Therefore, the recommendations, in this case, are the decommissioning of obsolete fixed assets and those that are not used for a long time, or with a low level of load, if it is possible.

With regard to gas production and distribution companies, as a result of the proposed price adjustment, all indicators of resource efficiency included in the IRE model had a positive upward trend. The exception is a certain set of indicators, which received a positive increase, but remained low: K_{11}, K_{13}, K_{31}, K_{41}, K_{43} and K_{45}.

The high material consumption of processed products will not allow K_{11} and K_{13} to take on competitive values in the future.

The real problem that has prospects for its solution is to increase labor productivity K_{31} by reasonably optimizing the number of employees.

Problems with the turnover of working capital are caused by its high share, K_{43} and K_{45}. Thus, in gas production and distribution in 2018 it was 83.6%, and in oil refining – 73.3%. An additional financial problem of these enterprises is the negative amount of total capital due to retained losses of previous years.

7 Conclusions

Analysis of the dynamics of oil and gas production shows that during 2015-2018 these enterprises significantly improved their indicators of resource efficiency on all the areas of research, resulting in an integral index increase from 0.404 to 0.658, which is positive. Significant qualitative shifts occurred in consumption of raw materials and supplies and labor resources use. The result of such changes was that the oil and gas sector outperformed both the primary (extraction) industry and the entire industry, as well as the average level in the economy of Ukraine in terms of resource efficiency. Thus, on the one hand, we had a positive trend in resource efficiency increase. On the other hand, it was achieved by a significant increase in product prices in recent years.

Thus, the obtained model of estimating the resource efficiency index takes into account the agreed and confirmed opinion of experts on the impact of each of the factors on the performance (effective) indicator. It is reasonable to use it in further economic assessment of the consequences of the implementation of resource-efficient strategy at the enterprises of the oil and gas sector of the economy of Ukraine.

Regarding the oil and gas refining, as well as gas distribution system we can observe that according to the indicators of fixed assets and capital use there is a significant lag from other enterprises of the processing industry and average level in the economy in general.

Thus, the diagnostics of resource efficiency of oil and gas sector pointed to existing problems faced by refinery enterprises and significant improvement in oil and gas production field.

This study improved the model of diagnostics of resource efficiency in oil and gas sector in the economy of Ukraine based on the additive-multiplicative compression of the formed system, which, unlike the existing ones, takes into account their variation while defining weighting coefficients which show the experts' system of preferences.

Thus, the implementation of a resource-efficient strategy in the oil and gas sector of Ukraine should include the following practical measures:

1. Creating conditions for the redistribution of value added between extractive and processing enterprises of this sector in favor of the latter, through market pricing in a demonopolized market.
2. Measures must be taken at oil and gas production enterprises to: optimize the number of labor resources to increase labor productivity; the fullest use of existing production capacity in the current economic stagnation and the lack of significant capital investment in technological re-equipment; decommissioning of obsolete fixed assets and those that are not used, or with a low level of load.
3. At oil and gas processing enterprises it is necessary to implement resource-saving measures to: increase labor productivity by reasonably reducing the number of employees; reduction of short-term receivables to increase capital turnover, etc.

References

1. Boiko, V., Kwilinski, A., Misiuk, M., Boiko, L.: Competitive advantages of wholesale markets of agricultural products as a type of entrepreneurial activity: the experience of Ukraine and Poland. Ec. An.-XXI. 175(1-2), 68-72 (2019). doi:10.21003/ea.V175-12
2. Bringezu, S., Bleischwitz, R. (eds.): Sustainable Resource Management: Global Trends, Visions and Policies. Routledge, New York (2017)
3. De Besi, M., McCormick, K.: Towards a bioeconomy in Europe: National, regional and industrial strategies. Sustainability **7**(8), 10461–10478 (2015)
4. Domenech, T., Bahn-Walkowiak, B.: Transition towards a resource efficient circular economy in Europe: policy lessons from the EU and the member states. Ecological Economics **155**, 7–19 (2019)
5. Duflou, J.R., Sutherland, J.W., Dornfeld, D., Herrmann, C., Jeswiet, J., Kara, S., Hauschild, M., Kellens, K.: Towards energy and resource efficient manufacturing: A processes and systems approach. CIRP annals **61**(2), 587-609 (2012). doi:10.1016/j.cirp.2012.05.002
6. Fryzorenko, A.O. (ed.): Ukraine's fuel and energy resources. Statistical collection. State Statistics Service of Ukraine, Kyiv (2019)
7. Jovovic, R., Draskovic, M., Delibasic, M., Jovovic, M.: The concept of sustainable regional development – institutional aspects, policies and prospects. Journal of International Studies **10**(1), 255-266 (2017). doi:10.14254/2071-8330.2017/10- 1/18
8. Karev, A.: Top Trends in the Oil and Gas Sector. J Petrol. Tech. **65**(9), 102–106 (2015). doi:10.2118/0913-0102-JPT
9. Khvostina, I., Havadzyn, N., Horal, L., Yurchenko, N.: Emergent properties manifestation in the risk assessment of oil and gas companies. CEUR Workshop Proceedings **2422**, 157–168 (2019)
10. Mantserova, T.F., Lapchenko, D.A.: The Main Approaches to Economic Diagnostics of the Power Engineering Enterprises. ENER. Pr. CIS H. Educ. Inst. P. Eng. Assoc. **62**(4), 362–376 (2019). doi:10.21122/1029-7448-2019-62-4-362-376
11. Matopoulos, A., Barros, A.C., van der Vorst, J.G.A.J.: Resource-efficient supply chains: a research framework, literature review and research agenda. Supply Chain Management, **20**, 218–236 (2015). doi:10.1108/SCM-03-2014-0090
12. Milovidov, K.N.: Global investment in upstream oil and gas sector. Probl. Econ. Manag. Oil Gas Compl. 2(170), 46–54 (2019). doi:10.33285/1999-6942-2019-2(170)-46-54
13. Ness, D.A., Xing, K.: Toward a Resource-Efficient Built Environment: A Literature Review and Conceptual Model. Journal of Industrial Ecology **21**(3), 572–592 (2017)
14. Perevozova, I., Shmygol, N., Tereshchenko, D., Kandahura, K., Katerna, O.: Introduction of creative economy in international relations: aspects of development security. J. Sec. Sust. **9**(1), 139–154 (2019). doi:10.9770/jssi.2019.9.1(11)
15. Pinchuk, A., Tkalenko, N., Marhasova, V.: Implementation of Circular Economy Elements in the Mining Regions. E3S Web of Conferences **105**, 04048 (2019)
16. Raitu, C., Anderson, B.B., Sfakianaki, E.: Resource-efficient construction: Rethinking construction towards sustainability. World Journal of Science, Technology and Sustainable Development **12**(3), 233–242 (2015). doi:10.1108/WJSTSD-03-2015-0016
17. Ramazanov, S.: Development, Management and Forecasting in a Green Innovative Economy, Based on the Integral Dynamics Model in the Conditions of 'Industry–4.0'. ACCESS Journal: Access to science-business-innovation in digital economy **1**(1), 9–29 (2020)
18. Rodrik, D.: A Research Agenda in Economic Diagnostics. American Economic

Association, Ten Years and Beyond: Economists Answer NSF's Call for Long-Term Research Agendas. SSRN (2010). doi:10.2139/ssrn.1889365

19. Ross, J.W., Weill, P., Robertson, D.C.: Enterprise Architecture as Strategy: Creating a Foundation for Business Execution. Harvard Business School Press, Boston (2006)
20. Sandvall, A.F., Ahlgren, E.O., Ekvall, T.: Cost-efficiency of urban heating strategies– Modelling scale effects of low-energy building heat supply. Energy strategy reviews **18**, 212–223 (2017)
21. Savchenko, T., Basiurkina, N., Rodina, O., Kwilinski, A.: Improvement of the assessment methods of product competitiveness of the specialized poultry enterprises. Manag. Th. St. Rur. Bus. Infr. Dev. **41**(1), 43–61 (2019). doi:10.15544/mts.2019.05
22. Savitskaya, G.V.: Metodika kompleksnogo analiza hozyaystvennoy deyatelnosti (Methodology of a comprehensive analysis of economic activity). INFRA-M, Moscow (2007)
23. Shinkevich, A.I., Kudryavtseva, S.S., Ershova, I.G.: Modelling of Energy Efficiency Factors of Petrochemical Industry. International Journal of Energy Economics and Policy **10**(3), 465–470 (2020)
24. Shmygol, N., Łuczka, W., Trokhymets, O., Pawliszczy, D., Zavgorodniy, R.: Model of diagnostics of resource efficiency in oil and gas sector of economy of Ukraine. E3S Web of Conferences **166**, 13005 (2020). doi:10.1051/e3sconf/202016613005
25. State Statistics Service of Ukraine: Ekonomichna statystyka: diialnist pidpryiemstv. http://www.ukrstat.gov.ua/operativ/menu/menu_u/sze.htm (2019). Accessed 02 Feb 2020
26. Tolpegina, O.A.: Methodological principles of classification of types of economic diagnostics. Eff. Cr. Manag. 1–2, 64–73 (2017). doi:10.17747/2078-8886-2017-1-2-64-73
27. Tukker, A.: Product services for a resource-efficient and circular economy – a review. Journal of cleaner production **97**, 76-91 (2015)
28. Yale Center for Environmental Law & Policy: Environment Performance Index. Methodology. https://epi.envirocenter.yale.edu/2018-epi-report/methodology (2018). Accessed 02 Feb 2020
29. Zhernov, E., Nekhoda, E., Petrova, M.: Economic Transformation Impact on the Modernization of a Mineral Resource Industry Cluster. E3S Web of Conf. **174**, 04003 (2020). doi:10.1051/e3sconf/202017404003
30. Zhou, Q., Liu, R.: Strategy optimization of resource scheduling based on cluster rendering. Cluster Computing **19**(4), 2109–2117 (2016)

Fuzzy modelling of Big Data of HR in the conditions of Industry 4.0

Mykola Ivanov[1[0000-0002-1908-0763]], Sergey Ivanov[1[0000-0003-1086-0701]], Olexander Cherep[1[0000-0002-3098-0105]], Nataliia Terentieva[1[0000-0001-6930-879X]], Victoria Maltiz[1[0000-0002-3863-6456]], Iuliia Kaliuzhna[1[0000-0002-3335-6551]] and Vitaliy Lyalyuk[2[0000-0001-7258-2079]]

[1] Zaporizhzhia National University, 66 Zhukovsky Str., Zaporizhzhia, 69063, Ukraine
[2] State University of Economics and Technology,
5 Stepana Tilhy Str., Kryvyi Rih, 50006, Ukraine
nn_iva@ukr.net, flydaiver@gmail.com, cherep2508@gmail.com, terenteva_nataliya@ukr.net, ma.viktorija@ukr.net, kalyuzhnaya.ju@gmail.com, vitalij.lyalyuk@gmail.com

Abstract. In this article, a systematic methodology for analyzing and assessing the effectiveness of human resources based on fuzzy sets using big data technologies is used. Based on our research, we analyzed the big data construction method for our chosen approach using Industry 4.0. For the selected fuzzy sets, a set of sequence of procedures in the sequence of the method for assessing the effectiveness of human resources have been identified. Input and output membership functions for data mining have been developed. This article discusses process of building rules of fuzzy logic that allowed us to determine the degree of truth for each condition. The relevance achieved through the development of a methodology that includes eight procedures required for a comprehensive assessment of the economic efficiency of human resources. In this article, an approach to assessing the normative or average values of the performance of official duties by employees of an enterprise in many specialties, educational levels, levels of management, as well as taking into account the description of many positions, descriptions of compliance and interchangeability of positions, assessment of additional characteristics of employees and a description of many additional tasks and their characteristics is presented. The article presents a structural data-mining model for personnel assessment. The results of modeling the assessment of human resources is presented.

Keywords: human resources, Big Data, Industry 4.0, data-mining model, fuzzy modelling.

1 Introduction

Today business is forced to solve a whole range of complex and unique tasks. To solve such problems as the tasks of increasing and stabilizing the development of economic objects in modern conditions of economic activity, new approaches are required, which determined the emergence of a new concept by the German economist Klaus Schwab, President of the World Economic Forum in Davos [21]. According to this concept, it is argued that we live in the era of the fourth industrial revolution, when the virtual world is combined with the physical world using information technology. The fourth industrial revolution is characterized by a change in economic relations and the widespread use of intelligent technologies (cloud technologies [13], big data [12], artificial neural networks and fuzzy sets [23], data mining [7], and others). These technologies form the backbone of the digital economy.

Therefore, for the successful development of states, regions, enterprises in the era of the fourth industrial revolution (Industry 4.0), an assessment of the possible potential efficiency of human resources (HR) plays an important role, since HR can radically change and increase the development and competitiveness of business entities in modern conditions.

In addition, the global COVID-19 pandemic has forced countries to rethink their national concepts of economic development [22].

It should be said that the term Industry 4.0 is relatively new, which is applied to the concept of the fourth industrial revolution and covers a wide range of modern technologies and approaches, mainly related to the digital economy. In terms of modern technologies, Industry 4.0 is associated with the application and intelligent processing of data in the following areas: artificial neural systems, industrial Internet, cloud solutions and decentralized services, as well as processing and mining of large amounts of data. The works of Klaus-Dieter Thoben, Stefan Wiesner and Thorsten Wuest [27], Robert Lawrence Wichmann, Boris Eisenbart and Kilian Gericke [28] are devoted to these studies on the technology of processing big data in real time. According to the application of the approaches and technologies of using Industry 4.0, they can be divided into the following areas: Big Data storage, data mining and the use of artificial neural systems and fuzzy sets. As for the use of big data, first of all, one can highlight the cloud platforms Amazon Web Services (AWS), Google Cloud Platform, Microsoft's Azure and others. These decisions influenced the formation of cloud computing concepts discussed in the work of S. O. Kuyoro, F. Ibikunle and A. Oludele [14], as well as in the work on data mining and machine learning by Ian H. Witten and Elbe Frank [29] and others [11; 18].

The first direction is determined by the level of application of cloud computing. High information technologies are sweeping the world and are replacing classical methods of analysis with HR processes. Nowadays, robotic programs are being used that offer employees of the enterprise to pass an express interview or interview using expert systems, which are still in use.

The next direction is based on the application of data mining in relation to their historical layers.

This analysis and assessment of HR allows you to select employees who are capable of solving modern problems. Now new specialists under 25 are entering the labour market, who have completely different knowledge, interests and fundamentally different ideas about modern work. Young specialists are able to quickly make management decisions and promote new projects, thereby increasing the level of work at the enterprise.

Beyond these areas, it should be noted that for a long time HR management has been focused on standardization and versatility. However, today this approach is gradually becoming obsolete. This approach is being replaced by methods of personnel management focused on the maximum use of the intellectual capital of employees. This is stimulated by the simultaneous satisfaction of individual needs, desires and capabilities of employees and their synchronization with the tasks of the enterprise. Modern HR specialists are beginning to more closely monitor the development of employees in the region and within the enterprise, which allows flexible management of career growth, which can be adjusted taking into account the proposals of the employees themselves.

2 Research motivation and formal problem statement

In this section, we provide our motivation for this work and provide an analytical overview of the research questions used to develop a method for analytically assessing human resources.

2.1 Research motivation

Our main motivation for this research work is to bridge the gap between regional, production development scenarios, taking into account the assessment of the economic efficiency of human resources using big data. We are confident that this work will serve as a basis for developers of use cases for Industry 4.0, so that in the management process, more informed decisions can be made when choosing a strategy for the development of regions and enterprises.

In the course of the recent work of Mykola Ivanov, Nataliia Maksyshko, Sergey Ivanov and Nataliia Terentieva [10], we realized that there is a great need to develop a method for representing big data HR. It is also necessary to develop a methodology for analyzing and assessing the economic efficiency of personnel based on big data for the development of a region or enterprise.

Thus, this study makes it possible to carry out strategic planning for the development of a region or an enterprise, taking into account the intellectual assessment of the economic efficiency of personnel.

2.2 Literature review

Ensuring compliance with the modern requirements of Industry 4.0 and available technologies for using various technologies for processing Big Data is based on the use of previously developed directions.

The first direction is determined by the level of application of intelligent systems. High information technologies are sweeping the world and replacing the classical methods of managing HR processes. Nowadays, robotic programs are being used that offer employees of the enterprise to pass an express interview or interview using expert systems, which are still in use.

The next direction is intellectual analysis and assessment of the economic efficiency of human resources capable of learning and solving modern problems.

In addition to these areas, one should also take into account modern methods of personnel management, focused on the maximum use of the intellectual capital of employees. This is stimulated by the simultaneous satisfaction of individual needs, desires, and capabilities of employees and their synchronization with the tasks of the enterprise. Modern HR specialists are beginning to more closely monitor the development of employees within the enterprise, which allows flexible management of career growth, which can be adjusted taking into account the proposals of the employees themselves.

Personnel strategy is part of the overall strategy for the development of enterprises and long-term planning, and their business activities. An important role in these plans is played by the assessment of the degree of personnel efficiency as a factor in the renewal and increase in production efficiency in the general economic strategy of the enterprise. Building and managing modern employee data directories requires processing a lot of information. This is due to a wide range of organizational, economic, technical and technological problems solved by the personnel. Therefore, data analysis in HR process management is an urgent task.

However, the problems of rating management have not yet been resolved. The solution to the problem of rating management was the work of Iurii H. Lysenko, Volodymyr L. Petrenko, Oleh I. Bohatov and Volodymyr H. Skobeliev [16]. However, the level of personnel development and their assessment were not taken into account. This work was devoted to the solution of this problem, in which the theoretical aspects of personnel development are studied, in particular the concept, main tasks and directions of personnel development at the enterprise. V. M. Helman, Ye. V. Makazan and A. M. Buriak [6] considers the development of enterprise personnel as a change in its qualitative characteristics, in which indicators are offered in the form of a degree of activity.

Human resource management as a strategic human resource management is to-day considered as going beyond such management tasks as motivation, the level of remuneration. Instead, managers should view human resource management as a process that contributes to the success of the enterprise. Therefore, in the work of Brian E. Becker and Mark A. Huselid [1] approaches are considered in which all managers should be involved in the management process, where the role of employees is important for the competitive advantage of the enterprise. In addition, the authors, when solving these problems, considered the issues when the company develops and motivates the development of human capital. They also identified the requirements for businesses that value a skilled workforce and are more profitable than those that do not value this workforce.

The results of the work of scientists Mark A. Huselid [8], Jeffrey Pfeffer and John F. Veiga [20] show that successful enterprises have several characteristics in com-mon: stable job security, high levels of self-government and excellent wages. The most successful businesses manage people as a strategic asset and measure the performance of people in terms of their influence on the manufacturing process. Ibraiz Tarique, Dennis Briscoe and Randall S. Schuler in the work [26] writes about a situation when each employee of an enterprise effectively performs his duties and builds a highly efficient work system in which the employee bears maximum involvement and responsibility.

In modern enterprises, balancing the need to coordinate and synchronize HR across cities and around the world is an important challenge, as discussed in [25]. Achieving this balance is becoming increasingly difficult due to the level of functional diversity that states, regions and enterprises strive for in the era of Industry 4.0. Approaches to assessing the performance indicators of personnel in the context of the development of the digital economy are considered in the work of Mykola Ivanov, Sergey Ivanov, Nataliia Terentieva, Victoria Maltiz and Julia Kalyuzhnaya [9].

The use of big data in human resource management has been reviewed by Peter O'Donovan, Colm Gallagher, Kevin Leahy, Dominic T. J. O'Sullivan [19] and Alessandra Caggiano [2].

Along with these studies, not enough attention is paid to the problems of fuzzy modeling of big data of human resources in the conditions of Industry 4.0, which makes this problem very relevant today.

2.3 Formal problem statement

In the Industry 4.0 strategy, human resources are part of the overall development strategy of the state, region and enterprise, and allow them to ensure the current planning of economic activities. An important role in these plans is played by the assessment of the degree of efficiency of human resources as a factor of renewal and increase in production efficiency in the general economic strategy of the region and the enterprise. The creation of modern Big Data about the population, employees and their management requires processing a large amount of information using intelligent assessment using fuzzy logic. This is due to a wide range of tasks solved by the personnel, both organizational, economic and technological. Therefore, fuzzy modeling of big data of human resources in the conditions of Industry 4.0 is an urgent task.

2.4 Purpose of the article

The article is devoted to intellectual analysis and assessment of the effectiveness of human resources in the context of Industry 4.0. In addition, our goal was to reveal new knowledge based on the application of fuzzy set theory with the possibility of using Big Data.

3 The theory of fuzzy modeling of HR

3.1 Big Data architecture and analysis theory

Modern construction and architecture of Big Data allows not only storing, but also processing and analyzing data that is too large or too complex for traditional database management systems.

In Industry 4.0, according to research [28], they are required by using a strategic change management approach that gives them a broader approach to leverage the benefits of analytic processing.

Big Data offers tremendous opportunities to revolutionize human resource management.

Managing human resources with cloud-based solutions opens up new opportunities and solutions.

These solutions include a new level of accessibility that facilitates greater employee mobility. The ability to effectively apply data mining tools and decision-making systems. Great opportunities, flexibility and constant updating of Big Data contribute to the development of theoretical and practical developments. The Big Data architecture can be represented as follows (fig. 1).

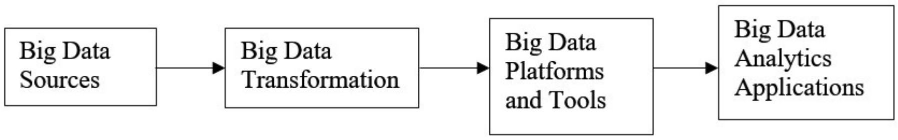


Fig. 1. Architecture Big Data analytics.

Today, the sources of Big Data for us are the Internet (social networks, websites and other applications). Big Data transformation is based on Data Warehouse principles. The Data Warehouse has a complex layered architecture called the Layered Scalable Architecture (LSA). In the Big Data system, LSA performs the logical division of data structures into several functional levels. Data is copied (saved) from level to level and transformed at the same time, in order to eventually appear in the form of consistent information. This information will be ready for further analysis.

The key components of Big Data Analytics Applications are Data Mining and OLAP (On-Line Analytical Processing) multivariate data analysis technology. OLAP is a key component of traditional data warehouse organization. OLAP systems are, in one way or another, based on a data storage and organization system.

The development of alternative methods for searching and aggregating information in sparse data hypercubes implies work in a number of directions. Among these areas, one can single out the study of the data model and the formalization of methods for assessing the density of the data hypercube, the study and development of effective methods for accessing information in the data hypercube, the development of an alternative method for aggregating the sparse data hypercube, the study of the possibilities of using various methods of visualization of data hypercubes, and others.

The standard description of a multidimensional data model is based on the following concepts: Data Hypercube, Dimension, Memders, Cells, and Measure.

A data hypercube contains one or more dimensions and is an ordered collection of components. Each component is defined by one and only one set of measurement values – labels. The component can contain a measure or be empty.

A dimension is understood as a set of marks that form one of the faces of a hypercube. An example of a time dimension is a time period: day, month, quarter, year.

An example of an economic dimension is a list of profitability indicators: working capital, cost, etc.

To gain access to the data, we need to specify one or more directions for choosing the measurement values that correspond to the necessary components. The mechanism for selecting measurement values is the fixation of marks, and the set of selected measurement values is a set of fixed marks.

The set of dimensions of a hypercube can be written in the following form

$$P = \{p_i = (op1, op2, ep1, ep2, ep3)\},\ i = \overline{1, N}\ , \tag{1}$$

where $op1$ – is a generalized indicator of job compliance, characterizing the degree of conformity of qualifications and work experience of the post, level of responsibility, as well as the quality of the performance of current work and duties,

$op2$ – is a generalized indicator of diligence, characterizing the effectiveness of the tasks (complexity, quality, timeliness),

$ep1$ – ambitiousness, a single indicator of personality characteristics,

$ep1$ – the quality of a leader, an indicator of personality characteristics,

$ep3$ – the level of attitude in the team, a single indicator of personality characteristics.

Then the set of measurement labels p_i is written

$$M_{p_i} = \{m_1,\ m_2, \dots, m_n\},\ i = \overline{1, n}\ . \tag{2}$$

And the set of fixed dimensions $D' = \{p'_1,\ p'_2, \dots, p'_n\}$ and fixed labels of the fixed dimension of the hypercube

$$M'_{p'_i} = \{m'_1,\ m'_2, \dots, m'_n\},\ i = \overline{1, n}\ . \tag{3}$$

The data hypercube will be denoted as the set of its cells $H(D, M)$, which corresponds to the sets D, M. Then the subset of the data hypercube HR, will correspond to the set of fixed values D', M' and we will denote it as $H'(D', M')$.

A single set of measurement labels $M_h \subset M$ corresponds to each component of the data hypercube HR $h \in H$. If the HR data cell is empty (does not contain data), then the set of dimensions of the hypercube $H(D, M)$ will be denoted by $V(H)$.

Consider data manipulation operations in a hypercube. We propose a method for managing HR data that includes the following steps.

Stage 1. Data Projection – Multiple Query (MDX).

A subset of the multidimensional data cube $H'(D', M')$ represents a query (Slice).

Stage 2. Building a multidimensional query.

The construction of the query is carried out in order to obtain the necessary subset of the components $H' \subset H$ and remove the values by sequentially fixing the labels. The request is usually an HR dataset.

The label $m_i \in M$ defines the hyperplane of intersection of the data hypercube corresponding to the dimension $d_i \in D$. The set of fixed labels $M' \subseteq M$, thus, defines the set of hyperplanes of sections of the data hypercube, corresponds to the set of fixed dimensions $D' \subseteq D$. The intersection of these hyperplanes determines the set of components of queries of the data hypercube $H'(D', M')$, which is needed by the management level. The essence of the process of extracting data from a hypercube is to construct a slice of the data hypercube $H'(D', M')$ by specifying the sets D', M'.

Tearing out the label $m_i \in M$, corresponding to the measurement $d_i \in D$, we determine what further interests us in the data hypercube. At each next step, the user has access to labels corresponding to the set of unfixed measurements.

Thus, by selecting a label in the dimensions $d_i \in D$ and $d_i \in D'$, we get a query to the HR data hypercube in the form of a table or surface on a MATLAB plot.

The essence of the process of adding money from the hypercube is stored in prompting for the development

Stage 3. Changing the order of presentation (visualization) of measurements.

Changing the order in which measurements are presented is called Rotate. Rotation provides the ability to visualize data in the most comfortable form for their perception. In terms of the data model under consideration, rotation means changing the sequence of fixing marks when building a slice. The result of rotation for a two-dimensional slice (table) will be replacing columns with rows, and rows with columns.

Stage 4. Convolution and detailing.

Convolution and detailing are carried out due to the presence of a hierarchical structure of dimensions. Measurement values (labels) can be grouped into hierarchies consisting of one or more levels. For example, time labels are naturally combined in a hierarchy with levels: year, quarter, month, day. The operations of convolution and detailing do not fundamentally differ from the operation of building a slice of a data hypercube, but they are distinguished to describe the work with aggregated data. The presence of a hierarchical structure of dimensions allows for data aggregation.

Stage 5. Aggregation of tributes.

The number of aggregates stored in the data hypercube along with the primary data depends on the number of labels corresponding to the levels of the hypercube dimension hierarchy, starting with $l = 1$, and can significantly exceed the amount of primary data. The total number of aggregates in the case of two dimensions will be determined by the sum of the values of the areas $A_{01},\ A_{02}, \dots, A_{22}$, which are shown in fig. 2.

Filling the HR hypercube with data with an insufficient amount of initial data leads to the formation of empty components. Data hypercubes with many empty cells are sparse.

Thus, we have built a method for creating a visual representation of a multidimensional database. This allowed us to assess the effectiveness of personnel taking into account the filling of data hypercubes and to carry out a visual search for information in the database.

3.2 Human Resources fuzzy modeling method

The task of fuzzy modeling and data mining when managing HR processes is to efficiently extract and analyze the existing data array of employees with subsequent management of personnel using cloud solutions. This will allow the rapid implementation of a new personnel management system, obtaining a new level of accessibility and increasing its mobility.

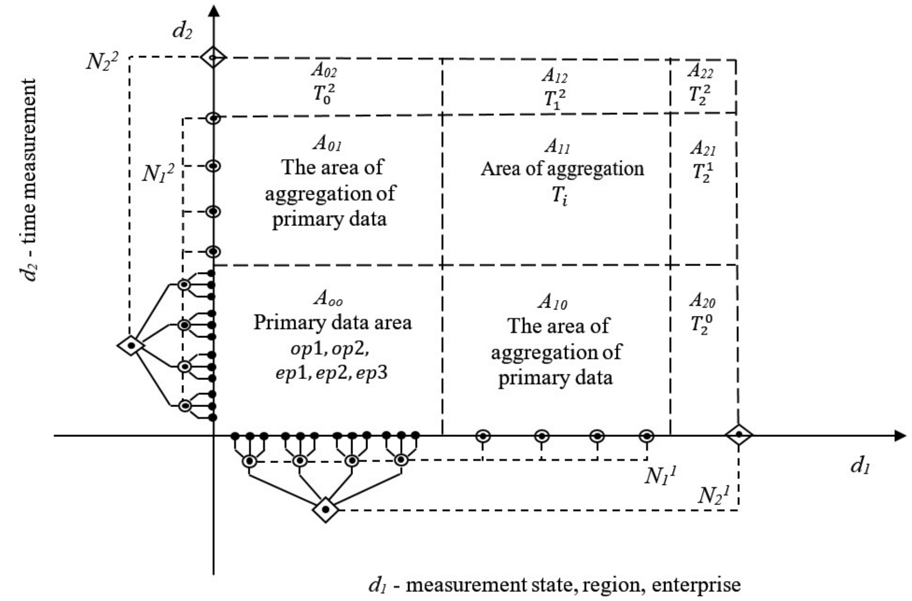


Fig. 2. Aggregation of HR data.

The resulting performance indicators of personnel at the enterprise can be represented in the form of multidimensional structures, where the corresponding indicators of the enterprise management system represent each measurement. The following method of modeling human resources is proposed, which is presented in fig. 3.

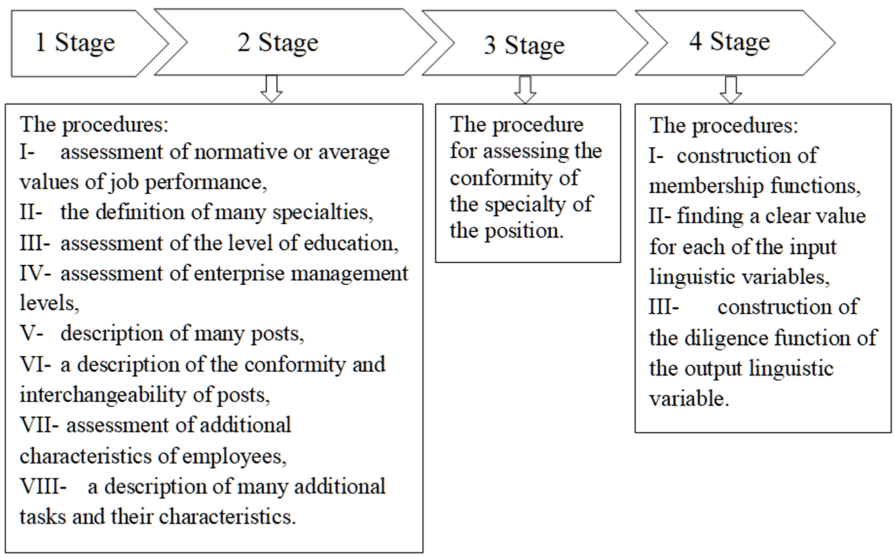


Fig. 3. The method of modelling human resources in fuzzy management.

The proposed method includes four stages.

3.3 Stage 1 Procedures

The first stage solves the problem of choosing the analyzed indicators. For this, a lot of ratings are determined (1).

3.4 Stage 2 Procedures

At the second stage, the initial information is determined, which is necessary for calculating indicators based on expert assessments, analytical indicators (for example, work experience, quality of work performed, and others).

To describe the formalized set of sets of source information, we introduce the rules, namely, if the set $P = \{pi\ = (op1, op2, op3)\}$, is defined, then to use the value of the component $op2$ of unit level 0, we will use the notation $op2$ (join operator).

At the second stage, procedures are applied that allow:

The first (I) procedure allows you to evaluate the regulatory or average value of the performance of official duties by employees – $P0$:

$$P0 = \{p0_i = (op01, op02)\},\ i = \overline{1, N}, \tag{4}$$

where $op01$ – normative or average value of the job performance of the i-th employee,

$op02$ – normative or average value of the level of assessment of the performance of tasks of the i-th employee.

The second (II) procedure is aimed at identifying many specialties (economist, programmer, builder and others) – SP:

$$SP = \{sp_r\}, r = \overline{1, L^{sp}}, \tag{5}$$

where sp_r – is the r-th specialty,

L^{sp}– is the number of specialties.

The third procedure allows you to assess the level of education (secondary, bachelor, master and others) – UO:

$$UO = \{uo_v = (name, \mu\ \}, v = \overline{1, L^{uo}}, \tag{6}$$

where uo_v – is the vector of characteristics of the v-th category,

$name$ – category name,

μ – assessment of the level of education for the category in points,

L^{uo} – is the number of categories.

The fourth (IV) procedure is aimed at assessing the level of enterprise management (higher, middle and lower level) – UD:

$$UD = \{ud_w = (name, \gamma\ )\}, w = \overline{1, L^{ud}}, \tag{7}$$

where ud_w – is the vector of characteristics of the w-th level,

$name$ – level name,

γ – is an estimate of the level in points,

L^{ud}– the number of levels, which is determined by the scale of the enterprise.

The fifth (V) procedure solves the problem of describing many positions in the enterprise – D:

$$D = \{d_j = (name, ud, uo, kl, SPD): ud \in UD,\ uo \in UO\}, j = \overline{1, L^d}, \tag{8}$$

$$SPD = \{spd_r = (sp, \beta): sp \in SP,\ 0 \le \beta \le 1\}, r = \overline{1, L_j^{spd}}, \tag{9}$$

where d_j – is the *j*-th position,

$name$ – job title,

ud – position level in the organizational and staff structure of the enterprise,

uo – the level of education required for the *j*-th position,

kl – required work experience (minimum number of years) in a given position for an optimal qualification level,

SPD – many specialties related to this position,

spd_r – is the vector of the correspondence characteristics of the *r*-th specialty of the *j*-th position β,

β – is the correspondence coefficient of the specialty sp of the *j*-th position,

L^d – is the number of posts,

L_j^{spd} – number of specialties in the *j*-th position.

The sixth (VI) procedure solves the tasks of describing correspondence and job interchangeability – SD:

$$SD = \{sd_f = (d1, d2, \alpha): d1 \in D, d2 \in D\ 0 \le \alpha \le 1,\ (d1 = d2) \Rightarrow \alpha = 1\},$$
$$f = \overline{1, L^{sd}}, \tag{10}$$

$$L^{sd} = (L^d)^2, \tag{11}$$

where sd_f – is the *f*-th vector of job matching characteristics $d1$ and $d2$,

α – is the compliance coefficient.

The seventh (VII) procedure is aimed at assessing additional characteristics of employees – A:

$$A = \{a_i = (ds, ST, OB): ds \in D,\ uo \in UO\}, i = \overline{1, N}, \tag{12}$$

$$ST = \{st_j = (d, kL): d \in D,\ d \in UO\}, j = \overline{1, L_i^{st}}, \tag{13}$$

$$OB = \{ob_w = (sp, uo, god): sp \in SP,\ uo \in UO\}, w = \overline{1, L_i^{ob}}, \tag{14}$$

where a_i – is the vector of characteristics of the *i*-th employee,

ds – the position held by the employee,

ST – many posts in which the employee previously worked and experience in them,

st_j – vector of characteristics of work experience in previous positions,

kL – length of service (number of years) in the position *d*,

OB – value, reflects the education received by the *i*-th employee,

sp – specialty,

uo – level of education,

god – year of receipt of the qualification document (certificate, certificate, diploma and others),

L_i^{st} – the number of posts previously held by the i-th employee,

L_i^{ob} – the number of specialties in which the employee was educated by the i-th employee.

The eight (VIII) procedure allows you to describe many additional tasks (determined by orders) and their characteristics in the enterprise:

$$Z = \{z_k = (t0, tk, tk', usz),\ 0 \le \beta \le 1\}, k = \overline{1, M}, \tag{15}$$

where z_k – is the vector of characteristics of the k-th task,

$t0$ and tk – the value of the beginning and end of tasks, determines the term for completing the task in units of measurement of working time (for example, working days, hours and others),

tk' – the value of time, determines the critical deadline for completing the task, after which the task is either canceled or transferred to another performer,

usz – task difficulty level,

M – the number of tasks.

The set of completing additional IZ tasks by employees can be written as follows:

$$IZ = \{iz_k = (a, z, uv^p, uv): a \in A, z \in Z, 0 \le uv^p \le 200\},\ k = \overline{1, M}, \tag{16}$$

where iz_k is the characteristic of the k-th job,

a – an employee who performs additional tasks,

z – the task

uv^p – is the percentage of the task according to the plan at the current time t ($uv^p = 0$ at time $t0$, $uv^p = 200$ at time k),

uv – is the percentage of the task at the current time t.

In case of failure to perform additional tasks, the value of $IZ = 0$.

3.5 Stage 3 Procedures

At the third stage, the procedure for assessing the conformity of the specialty of the position is performed. The function $f\beta$ returns the value of the correspondence of the specialty xsp to the position xd:

$$f\beta(xsp, xd) = \begin{cases} d_{j0}.spd_{i0}.\beta, \exists j_0, r_0: (d_{j0} = xd) \wedge (d_{j0}.spd_{r0}.sp = xsp) \\ 0, \qquad \neg\exists j_0, r_0: (d_{j0} = xd) \wedge (d_{j0}.spd_{r0}.sp = xsp) \end{cases}. \tag{17}$$

The function $f\alpha$ returns the value of the coefficient of correspondence and interchangeability of the specialty xsp and the position xd:

$$f\alpha(xds, xd) = \begin{cases} sd_g\alpha,\ \exists g_0: (sd_{g0}.d1) \wedge (sd_{g0}.d2 = xd) \\ \ 0,\ \neg\exists g_0: (sd_{g0}.d1) \wedge (sd_{g0}.d2 = xd) \end{cases}. \tag{18}$$

To determine job conformity is the level of education of the position held in conjunction with work experience in similar or related positions:

$$p_i.op1 = \delta(op11 \cdot op12) \cdot op13, \tag{19}$$

$$op11 = \sum_{w=1}^{L_i^{ob}} f\beta(a_i.ob_w.sp, a_i.ds) \cdot \frac{a_i.ob_w.god}{godT}, \tag{20}$$

$$op12 = \sum_{j=1}^{L_i^{st}} f\delta(a_i.ds.f, a_i.st.d) \cdot \frac{a_i.st_j.kL}{a_i.ds.kL}, \tag{21}$$

where $godT$ – is the value of the current year,
$op11$ – qualification level of education received,
$op12$ – qualification level, which is determined by work experience,
$op13$ – quality of job performance, determined by an expert.

When solving the problem of data mining in the management of HR processes, fuzzy logic methods are used to display the result on the interval [0; 1].

3.6 Stage 4 Procedures

Therefore, at the fourth stage, the procedure for constructing membership functions based on the theory of fuzzy sets is performed.

The following “position”, “level”, “education” can be attributed to numerical linguistic variables of employees, and “conflict”, “level of substitution” to linguistic variables. Numerical linguistic variables and their meanings serve for a qualitative description of a quantitative quantity. The values of linguistic variables are deter-mined by experts.

It should be noted that a linguistic variable, like its original term set, is associated with a specific dimensional scale on which all arithmetic operations are defined.

To assess the characteristics of employees in table 1, linguistic variables and their dimensions are proposed.

The use of the concept of stimulation and destimulation is applied taking into account the influence on the degree of personnel efficiency, namely, stimulation – the effect on the increase and destimulation – on the reduction of the factor.

Therefore, the term set $T_i^n = \{T_i^{n\prime}\}$ is associated with the set $T_i^{n\prime}$, where $T_i^{n\prime} = \langle x, \mu_{T_i^{n\prime}}(x) \,/\, x \in [x_{min}, x_{max}] \rangle$ is a fuzzy number, $i = \overline{1, m}$, m is the number of term sets, n is the number of employees.

To eliminate the influence of changes in the input variables of the metrics and, as a consequence, the correction of term sets, a transition to a normalized function is proposed. Let the previously defined term set T_i be the original one.

The normalized linguistic variable is a mapping on the interval [0; 1]:

$$D_i^n = \{D_i^{n\prime}\},\ D_i^{n\prime} = \langle z, \mu_{D_i^\prime}(z) \mid z \in [0; 1] \rangle, \tag{22}$$

where z – is a fuzzy number corresponding to the term set $D_i^\prime$ on the interval [0; 1], n – is the number of employees.

Table 1. The linguistic variables of employee characteristics.

Term set	The metric and type of exposure	x_{min}^{1}	x_{max}^{1}	The term designation
$T_1^1 = \cup T_1^j$, $j = \overline{1,3}$	Performance of duties <position>, stimulation	0	1	Not performed Partially completed Performed
$T_2^1 = \cup T_1^j$, $j = \overline{1,3}$	Job Interchangeability <Interchangeability Level>, discouragement	1	3.0	Low Average High
$T_3^1 = \cup T_1^j$, $j = \overline{1,3}$	Level of education <education>, stimulation	1	3.0	Secondary education Bachelor Master
$T_4^1 = \cup T_1^j$, $j = \overline{1,3}$	Conflict <conflict>, stimulation	0	3,0	Low Average High
$T_5^1 = \cup T_1^j$, $j = \overline{1,3}$	The importance level of the specialty <SpecialtyLevel>, discouragement	0	0,5	Low Average High

These functions allow you to display heterogeneous input variables in a single normalized interval [0; 1], which allows you to reduce errors associated with different quantities and their dimensions. This provides a convenient representation of the values, as well as their interpretation.

4 The method of constructing a model of data mining in HR process management

The structural model of data mining in HR process management is presented in fig. 4.

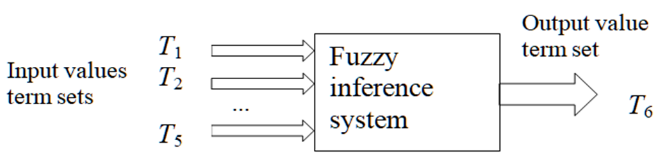


Fig. 4. The structural model of the model of data mining in the management of HR-process.

In the structural model, $T = \{T_i\}$ is a term set, where $i = \overline{1,n},$ n – is the number of sets, each of which is represented by a fuzzy variable with a domain of definition *X*.

The process of modeling fuzzy values is based on a fuzzy inference system, which allows you to convert expert estimates into fuzzy values.

In the fuzzy inference system, the procedure for finding a clear value for each of the input linguistic variables based on defuzzification is applied. Defuzzification in a fuzzy inference system is the process of finding a value for each of the output linguistic variables of the set $W = \{x_1, x_2, .. x_n\}$. The task of defuzzification is to use the results of accumulation of all output linguistic variables. It is necessary to obtain a quantitative value of each of the output variables. Output variables can be used in a fuzzy inference system relative to the input linguistic variable.

Accumulation of fuzzy inference is the process of finding the membership function for each of the output linguistic variables of the set.

The transformation of a fuzzy set into list of values of variables is named as defuzzification.

The defuzzification procedure is performed by a sequence that considers each of the output linguistic variables and the β fuzzy set $T_i = \{T_i^j\}$ related to it. The result of defuzzification for the output linguistic variable is defined as a quantitative value.

The defuzzification process is considered complete when quantitative values are determined for each of the output linguistic variables. For the fuzzy inference system, the Mamdani algorithm was applied [5; 24].

The Mamdani algorithm includes the following steps [17; 3]:

1. the formation of a rule base for fuzzy inference systems [15; 4];
2. fuzzification of input variables;
3. aggregation of conditions in fuzzy rules to find the degree of truth of the conditions of each of the rules of fuzzy logic;
4. accumulation of conclusions of fuzzy production rules;
5. defuzzification of output variables based on the center of gravity method.

An example of a rule looks like this:

```
1. If (Position is Npfd) and (InterchangeabilityLevel is low) and
(Education is secondary) and (Conflict is low) and (SpecialtyLevel is
low) then (StaffEfficiency is NotHardworking) (1)
2. If (Position is Npfd) and (InterchangeabilityLevel is low) and
(Education is bachelor) and (Conflict is low) and (SpecialtyLevel is
low) then (StaffEfficiency is NotHardworking) (1)
3. If (Position is Npfd) and (InterchangeabilityLevel is low) and
(Education is master) and (Conflict is low) and (SpecialtyLevel is low)
then (StaffEfficiency is NotHardworking) (1)
4. If (Position is Pfd) and (InterchangeabilityLevel is high) and
(Education is master) and (Conflict is low) and (SpecialtyLevel is high)
then (StaffEfficiency is Prospective) (1)
```

The values of linguistic variables are determined on an ordinal scale. It should be noted that a linguistic variable, like its original term set, is associated with a specific scale on which all arithmetic operations are defined.

Therefore, the term set $T_i = \{T_i^j\}$ is associated with the set T_i^j, where $T_i^j = \langle x, \mu_{T_i^j}(x) \,/\, x \in [x_{min}, x_{max}] \rangle$, $i = \overline{1,n}$; $j = \overline{1,m}$; n is the number of term sets, m is the number of terms.

A model that satisfies these fuzzy sets is their union:

$$\mu T_i = sup\left(\mu_{T_i^j}(x)\right),\ \ T_i = \cup\, T_i^j. \tag{23}$$

We construct membership functions for the linguistic variable characteristics of employees, presented in table. 1.

The process of converting experts' qualitative assessments into fuzzy quantities consists in mapping the elements of the original term set in the form of constructing membership functions of fuzzy quantities $T_i^j \in T_i$.

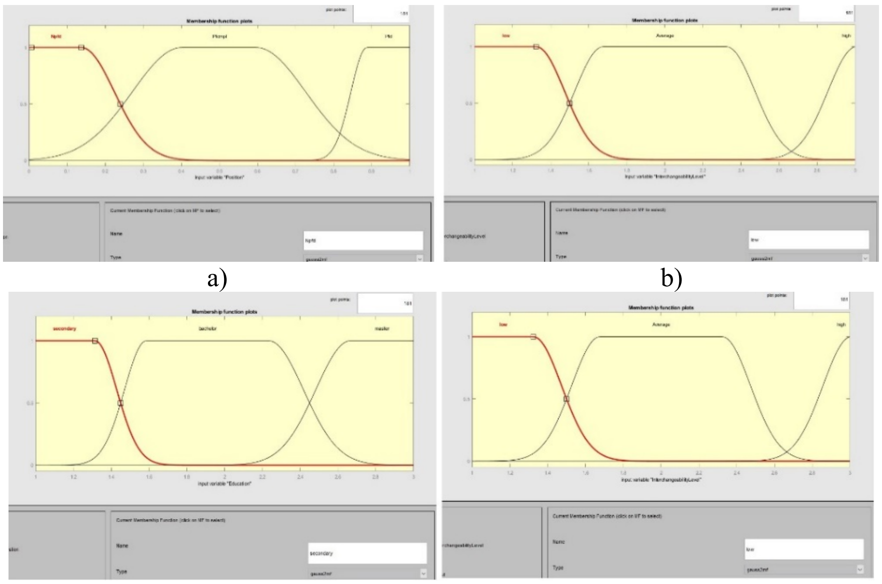
The description of linguistic variables is as follows:

```
⟨Position, {not fulfilled, partially fulfilled, fulfilled}, [0; 1]⟩,
⟨InterchangeabilityLevel, {low, medium, high}, [1; 3]⟩,
⟨Education, {secondary, bachelor, master}, [1; 3]⟩,
⟨Conflict, {low, medium, high}, [1; 3]⟩,
⟨SpecialtyLevel, {secondary, bachelor, master}, [1; 3]⟩.
```

Moreover, the values of the sets are in the range [0; 1] & [1; 3].

5 Experiments and results

The use of the Gaussian function is to use the membership function for modeling to determine fuzzy numbers. It is a form of analytical approximation using functions that include Gaussian functions.



a) b) c) d)

Fig. 5. The membership function of the input linguistic variables: a) "Position", b) "InterchangeabilityLevel", c) "Education", d) "Conflict".

The constructed membership functions of the input linguistic variables are presented in fig. 5.

In the fuzzy inference procedure for managing HR processes, it is necessary to consider the work of employees at all levels of work. The fuzzy inference procedure is implemented in the MATLAB system, which allowed obtaining the following results of assessing the degree of personnel efficiency. To perform the procedure, we built the diligence function of the output linguistic variable “Staff Efficiency”, which is presented in fig. 6.

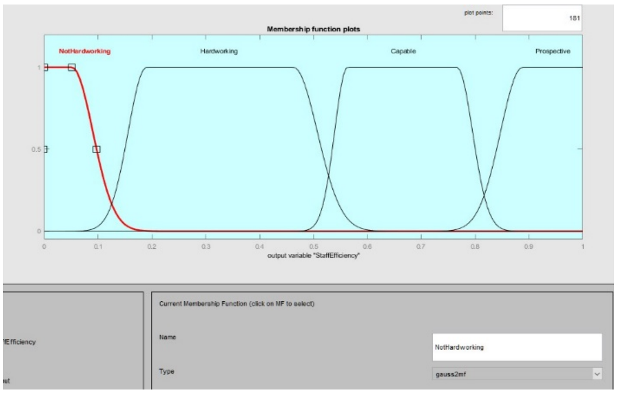


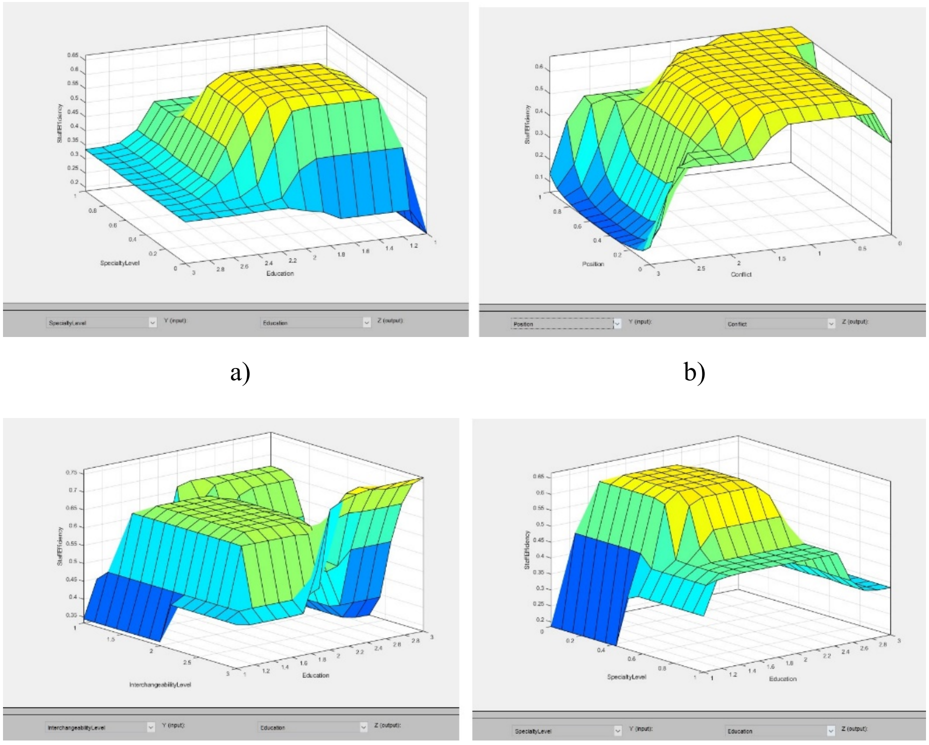
Fig. 6. The membership function of the output linguistic variable “Staff Efficiency”.

The simulation results of assessing the degree of personnel efficiency, which is presented in fig. 7.

The authors of the article propose a solution to the problem of constructing a data analysis method in human resource management and modeling the assessment of the degree of personnel efficiency based on fuzzy sets.

6 Conclusion

In this article offered the following methods for managing Big HR Data, analyzing and assessing the effectiveness of human resources. Method for analyzing and assessing the effectiveness of HR in four steps is developed. At the first stage, the problem of choosing the analyzed indicators is solved. At the second stage, eight procedures to solve the following tasks are performed. This procedure allows receive assess the standard or average values of labor productivity, determine a variety of specialties, assess the level of education, assess the levels of enterprise management, describe many job responsibilities, determine the conformity and interchangeability of work, to assess the additional characteristics of employees. At the third stage, the procedure for assessing the conformity of the specialty to the position held is carried out. At the fourth stage, the procedure for constructing membership functions based on the theory of fuzzy sets is performed. In the MATLAB system, a fuzzy inference procedure is implemented, which made it possible to assess the degree of efficiency of human resources.



a) b)

c) d)

Fig. 7. Modeling the assessment of the degree of personnel efficiency: a) the Specialty Level and Education, b) the Position and Conflict, c) the Interchangeability Level and Education, d) the Specialty Level and Conflict.

Prospects for the application of the method and procedures of fuzzy modeling in human resource management lie in the expansion of approaches and the use of models of cloud computing and big data.

References

1. Becker, B.E., Huselid, M.A.: Strategic Human Resources Management: Where Do We Go from Here? Journal of Management **32**(6), 898–925 (2006). doi:10.1177/0149206306293668
2. Caggiano, A.: Cloud-based manufacturing process monitoring for smart diagnosis services. International Journal of Computer Integrated Manufacturing **31**(7), 612–623 (2018). doi:10.1080/0951192X.2018.1425552

3. Dimitrov, V., Borisova L., Nurutdinova, I.: Development and analysis of fuzzy expert data for technological adjustment of a grain harvester header. E3S Web of Conferences **175**, 05027 (2020). doi:10.1051/e3sconf/202017505027
4. Gacto, M., Alcala, R., Herrera, F.: Interpretability of linguistic fuzzy rule-based systems: An overview of interpretability measures. Information Sciences **181**(20), 4340–4360 (2011)
5. Gottwald, S.: Universes of Fuzzy Sets and Axiomatizations of Fuzzy Set Theory. Part I: Model-Based and Axiomatic Approaches. Studia Logica **82**, 211–244 (2006). doi:10.1007/s11225-006-7197-8
6. Helman, V.M., Makazan, Ye.V., Buriak, A.M.: Personnel development as pledge of the success of enterprise. Bulletin of Zaporizhzhia National University. Economic sciences **43**(3), 112–115 (2019). doi:10.26661/2414-0287-2019-3-43-18
7. Hryvko, A., Sytnyk, O., Zhuk, Yu.: Exploring Students' Views on ICT Skills as the Components of Journalists' Professional Competence: EDM Aspect. CEUR-WS.org, online (2020, in press)
8. Huselid, M.: The Impact of Human Resource Management Practices on Turnover and Productivity. The Academy of Management Journal **38**(3), 635–672 (1995). doi:10.2307/256741
9. Ivanov, M., Ivanov, S., Terentieva, N., Maltiz, V., Kalyuzhnaya, J.: Fuzzy modeling in human resource management. E3S Web of Conferences **166**, 13010 (2020). doi:10.1051/e3sconf/202016613010
10. Ivanov, M., Maksyshko, N., Ivanov, S., Terentieva, N.: Intelligent Data Analysis in HR Process Management. CEUR Workshop Proceedings **2608**, 754–768 (2020)
11. Kholoshyn, I.V., Bondarenko, O.V., Hanchuk, O.V., Shmeltser, E.O.: Cloud ArcGIS Online as an innovative tool for developing geoinformation competence with future geography teachers. In: Kiv, A.E., Soloviev, V.N. (eds.) Proceedings of the 6th Workshop on Cloud Technologies in Education (CTE 2018), Kryvyi Rih, Ukraine, December 21, 2018. CEUR Workshop Proceedings **2433**, 403–412. http://ceur-ws.org/Vol-2433/paper27.pdf (2019). Accessed 10 Sep 2019
12. Kiv, A., Semerikov, S., Soloviev, V., Kibalnyk, L., Danylchuk, H., Matviychuk, A.: Experimental Economics and Machine Learning for Prediction of Emergent Economy Dynamics. CEUR Workshop Proceedings **2422**, 1–4 (2019)
13. Kiv, A.E., Soloviev, V.N., Semerikov, S.O.: CTE 2018 – How cloud technologies continues to transform education. CEUR Workshop Proceedings **2433**, 1–19 (2019)
14. Kuyoro, S.O., Ibikunle, F., Oludele, A.: Cloud Computing Security Issues and Challenges. International Journal of Computer Networks (IJCN). **3**, 247-255 (2011)
15. Ludermir, T., Zanchettin, C., Lorena, A.: Advances in intelligent systems. Neurocomputing **127**, 1–3 (2014). doi:10.1016/j.neucom.2013.07.040
16. Lysenko, Iu.H., Petrenko, V.L., Bohatov, O.I., Skobeliev, V.H: Reytingovoye upravleniye ekonomicheskimi sistemami. Yugo-Vostok, Donetsk (1999)
17. Mamdani, E., Assilian, S.: An experiment in linguistic synthesis with a fuzzy logic controller. International Journal of Man-Machine Studies **7**(1), 1–13 (1975)
18. Merzlykin, P.V., Popel, M.V., Shokaliuk, S.V.: Services of SageMathCloud environment and their didactic potential in learning of informatics and mathematical disciplines. In: Semerikov, S.O., Shyshkina, M.P. (eds.) Proceedings of the 5th Workshop on Cloud Technologies in Education (CTE 2017), Kryvyi Rih, Ukraine, April 28, 2017. CEUR Workshop Proceedings **2168**, 13–19. http://ceur-ws.org/Vol-2168/paper3.pdf (2018). Accessed 21 Mar 2019
19. O'Donovan, P., Gallagher, C., Leahy, K., O'Sullivan D.: A comparison of fog and cloud computing cyber-physical interfaces for Industry 4.0 real-time embedded machine learning

engineering applications. Computers in Industry **110**, 12–35 (2019). doi:10.1016/j.compind.2019.04.016
20. Pfeffer, J., Veiga, J.F.: Putting People First for Organizational Success. Academy of Management Perspectives **13**(2), 37–48 (1999). doi:10.5465/AME.1999.1899547
21. Schwab, K.: The Fourth Industrial Revolution. Currency, Geneva (2017)
22. Semerikov, S., Chukharev, S., Sakhno, S., Striuk, A., Osadchyi, V., Solovieva, V., Vakaliuk, T., Nechypurenko, P., Bondarenko, O., Danylchuk, H.: Our sustainable coronavirus future. E3S Web of Conferences **166**, 00001 (2020). doi:10.1051/e3sconf/202016600001
23. Semerikov, S.O., Teplytskyi, I.O., Yechkalo, Yu.V., Kiv, A.E.: Computer Simulation of Neural Networks Using Spreadsheets: The Dawn of the Age of Camelot. CEUR Workshop Proceedings **2257**, 122–147 (2018)
24. Siler, W., Buckley, J.: Fuzzy Expert Systems and Fuzzy Reasoning. Wiley Interscience, Birmingham (2005)
25. Stahl, G.K., Björkman, I. (eds.): International human resources management. Handbook of international management research, 2nd edn. Edward Elgar, Cheltenham (2006)
26. Tarique, I., Briscoe, D., Schuler, R.S.: International Human Resource Management: Policies and Practices for Multinational Enterprises, 5th edn. Routledge, New York (2009)
27. Thoben, K., Wiesner, S., Wuest, T.: "Industrie 4.0" and Smart Manufacturing – A Review of Research Issues and Application Examples. International Journal of Automation Technology **11**, 4–19 (2017). https://doi.org/10.20965/ijat.2017.p0004
28. Wichmann, R., Eisenbart, B., Kilian Gericke K.: The Direction of Industry: A Literature Review on Industry 4.0. Proceedings of the Design Society. International Conference on Engineering Design ICED19 **1**(1), 2129–2138 (2019). doi:10.1017/dsi.2019.219
29. Witten, I.H., Frank, E.: Data Mining: Practical Machine Learning Tools and Techniques. Elsevier, San Francisco (2005)

Using non-metric multidimensional scaling for assessment of regions' economy in the context of their sustainable development

Pavlo Hryhoruk[0000-0002-2732-5038], Svitlana Grygoruk[0000-0003-3047-2271], Nila Khrushch[0000-0002-9930-7023] and Tetiana Hovorushchenko[0000-0002-7942-1857]

Khmelnytskyi National University, 11 Instytutska Str., Khmelnytskyi, 29016, Ukraine
violete@ukr.net, grygoruk.svitlana@gmail.commailto:,
nila.ukr@gmail.com, tat_yana@ukr.net

Abstract. Solving the problems of regions' socio-economic development is strategic and most important for any country. In particular, the implementation of a new, active role of the region as a subject of sustainable development is important for the direct implementation of current regional policy. An important component of such a policy is the assessment of sustainable development of regions, which contributes to the timely detection of internal and external threats, the development of necessary stabilizing measures to prevent their negative impact, the formation of strategies aimed at sustainable regional systems. The economic system is an important subsystem of the region. The article proposes an approach to assessing the regions' economic development in the context of ensuring their sustainable development. We used the methods of multidimensional nonmetric scaling to solve this problem. The study aims to determine the structure of regions in the context of their sustainable development. Based on non-metric data reflecting the economic development of Ukraine's regions, two-dimensional space of latent scales was built based on multidimensional measures of proximity between them, and the positioning of regions in this space was carried out. The results received a semantic interpretation, which was improved by using the procedure of rotation of the scale space. The use of multidimensional non-metric scaling confirms its usefulness for the study of economic development of regions in the region and allows for their comparison and dynamics of their structure in the context of sustainable development.

Keywords: sustainable development, non-metric multidimensional scaling, region, two-dimensional space.

1 Introduction

1.1 Problem description

Regional development issues remain relevant for every country, as unreasonable and sporadic regional development policies can lead to growing disparities and exacerbation of economic, political, environmental, and social problems. In the context of European integration processes, the main goal of the state regional policy is to create conditions for dynamic, balanced development of territories, eliminate the asymmetry of development of the regions of Ukraine, intensify the involvement of human resources to ensure the competitiveness of the region's economy.

Sustainable development is a modern worldview, political and practical model of development for all countries of the world, which have started the transition from a purely economic model of development to finding the optimal balance between the three components of development - economic, social and environmental. This category is perceived around the world as a model of civilized development. Implementation of this model requires the formation of a system for managing such development.

In September 2015, during the 70th session of the UN General Assembly in New York, the UN Summit on Sustainable Development took place and adopted the 2030 Agenda for Sustainable Development. It approved new development benchmarks [31]. Summit issues covered all aspects of socio-economic development, in particular, countries' competitiveness, environmental and energy security, a global partnership for development, and were based on the principle of "Leaving no one behind". Summit Outcome Document contains 17 Sustainable Development Goals and 169 Goals. This led to the update of the Sustainable Development Strategy of Ukraine until 2030 [25]. The Strategic vision of Sustainable Development of Ukraine is focused on overcoming the imbalances that exist in the economic, social, environmental spheres and is based on the vectors defined in the Sustainable Development Strategy "Ukraine 2020" [24], one of which is the vector of development. It foresees the sustainable development of the country, carrying out structural reforms, ensuring economic growth in an environmentally sustainable way, creating favorable conditions for economic activity [25].

Thus, at the present stage of development of Ukraine's economy, the problem of transition to the sustainable development of both the country as a whole and each of its regions is urgent. The balanced region's development should be oriented towards providing conditions that will allow each region of the country to have the needed and sufficient resources to ensure decent living conditions, comprehensive development and increase the competitiveness of the economy.

On the one hand, sustainable development of the region may be seen as a positively directed process of improving the economic, social, and environmental components. On the other hand, it is considered as a process to achieve a balanced state for all of these components.

The assessment of the level and state of sustainable development of regions' economy is necessary to identify internal and external threats, which will allow devising measures to prevent their negative impact to identify scenarios of development and to develop an optimal strategy for the functioning of the country's regional economic systems.

Given this, an important role in solving these issues belongs to the analytical tools, the main tool of which is economic and mathematical modeling and modern information technology. Their application will identify trends and imbalances in the economic development of regions, carry out their structuring, and, ultimately, will contribute to the development of sound management decisions.

1.2 Literature review

At the initial stage of its development, the concept of sustainable development was associated with environmental sustainability, but soon it was also used to describe social and economic sustainability. In the modern sense, sustainable development is perceived as a socio-economic system that meets human needs, but also long-term progress towards prosperity and improving the overall quality of life.

We support the opinion of the authors of [16] that the concept of sustainable development concerns not only the well-being of people but also the world where people live, so it can be understood as a holistic philosophy that includes classical philosophical perspectives and harmonizes and integrates economic, social-political and environmental system.

The issues of assessing the level of sustainable development at both the state and regional levels remain the subject matter of the attention of many scholars. For regional systems, economic, social, and environmental components are traditionally taken into account. In our study, the main attention will be paid to assessing the sustainable development of regions by their economic component of development. In this case, it is advisable to consider sustainable development in the context of its sustainability, which according to [34] we will understand the property or quality that determines the ability of the regional system to be in a state of dynamic equilibrium in the presence of external and internal influences. It can be inherent not only in the fixed state of development of the regional system but also considered in dynamics. As the main types of such stability can be distinguished:

- the stability of development, characterized by a systematic increase in the result, which is not lower than the acceptable minimum and not higher than the objectively determined maximum;
- as permanent stability when changes, including positive ones, occur only occasionally and for a short time;
- as hyper-sustainability like a state where regions are not susceptible to development but can adapt to changes, including positive ones.

The regions' economic development is characterized by a large number of different indicators, and therefore is essentially multidimensional. This presupposes the use of multidimensional economic and statistical modeling tools to reduce the state of economic development and reduce the space dimensionality of the original characteristics.

The analysis of publications showed a variety of methodological approaches to assess sustainable development, which is determined by the identified goals to obtain evaluations. At the same time, most of the methodological approaches involve calculating a composite (integrated, comprehensive) index of sustainable development based on the use of additive or multiplicative convolution. In some cases, the author's techniques involve the analysis of the output without the convolution, in particular, the indicative method of evaluation. In our view, narrowing the assessment outcome to a single indicator makes it easier to conclude, but on the other hand, this approach causes a "compensation effect" when the low values of some indicators are offset by the high values of other ones. This disadvantage can be partially eliminated by using a weighted convolution of the initial indicators.

The paper [21] proposes a methodological approach to the assessment of sustainable development of Ukrainian regions, in which the overall assessment of its level is carried out using a composite index, based on the additive convolution of indicators of social, economic and environmental components. An integrated assessment of the sustainable development of Ukraine's regions is carried out and problems of regional development in the social, economic, and environmental fields are identified. It should be noted that the author proposes to use the financial statements, which, in our opinion, limits the application of the proposed methodology.

Paper [27] proposes to assess sustainable development on generalized indices, such as Green GDP; Human development index; Genuine Progress Indicator; Index of Sustainable Economic Welfare; Happy Planet Index; Environmental Sustainability Index; Environmental Performance Index, which is converted into a generalized index of sustainable development by weighted additive convolution. The authors provide a scale for determining the level of sustainable development, but the authors do not justify the boundaries of the levels of such a scale. Also, in our opinion, the practical use of this approach is limited by the difficulty of obtaining statistics for the selected system of baseline indicators.

The paper [9] proposes an approach to the construction of an integrated indicator, which allows assessing the level and dynamics of economic development of the region in the context of ensuring its sustainable development. A feature of the proposed approach is the simultaneous use of both metric and non-metric indicators, and the calculation of weights for components is based on the values of factor loads of the first main component, calculated for a set of metrics. Clarification of weights is carried out with the help of non-metric indicators. However, the article does not substantiate the limits for determining the levels of sustainable development.

I. V. Horiana [8] distinguishes infrastructure and innovation components in addition to the traditional part of sustainable development to design a composite index. The author defines the rules of composite index calculation based on the multiplicative convolution and formulates the conditions under which sustainable development is reached. A similar approach is implemented in the papers [7; 10]. However, these approaches leave open the issues to identify the required initial data set.

The study of O. O. Nesterenko [20] uses an approach based on the scoring model to assess the sustainable development composite index. At the same time, the author proposes to use both the statistical reporting indicators and questionnaire outcomes of the evaluation of the several components of sustainable development. The commonly proposed method uses recognized indicators of human development like the Knowledge Index, the Human Capital Development Index, and the Human Development Index. It should be noted, that some of the used indicators have a non-numerical origin and therefore their mathematical processing is not always correct.

Papers [1; 3] proposed a set of criteria for assessing the sustainable development of a region based on the use of both metric and non-metric indicators. But authors do not specify the rules for processing data and constructing the resulting metric.

Paper [14] presents the author's methodology for assessing sustainable development for the Czech Republic according to four components: Political area, Social area, Economic area and Environmental area, which uses 101 output indicators and provides for the calculation of a hierarchical integral index system, which includes 12 partial indicators, 4 partial composite indicators and one aggregate integrated index of sustainable development.

The authors of the paper [35] compared the situation of Central and Eastern Europe in terms of sustainable development based on calculating the integrated indicator. The comparison was based on indicators that emerged from the thematic areas of the EU Sustainable Development Strategy, in particular, socio-economic development, demographic change, health status, climate change, energy, and others. According to the results obtained, the countries were ranked, and the gap between the countries of Central and Eastern Europe in comparison with the average level of the EU countries was assessed.

The study [26] presents an original approach to the calculation of a comprehensive sustainable development index based on the case of Indonesia's regions. Authors propose three measures for indices: arithmetic, geometric, and entropy-based. Indices are aggregated to use for comparing regions in terms of their stability. The article also analyzes the sensitivity of the results obtained. It should be noted that despite the ranking of regions by the value of the integrated indicator of sustainable development, the authors were not offered a scale to estimate the level of sustainable development of regions.

Studies [2; 32; 11] have presented approaches for assessing the degree of achievement of the 17 Sustainable Development Goals identified in [31] in the case of Asian countries.

The conducted analysis of publications makes it possible to conclude that the presented approaches are based on the UN Sustainable Development Concept using some differences in the structure of components and the number of partial indicators. The advantages of the composite indexes for evaluation of various aspects of sustainable development include the simplicity of their calculation and the ease of results interpretation. However, the approaches don't contain criteria for identifying the level of sustainable development. A significant disadvantage of these methods is the use of an overloaded set of partial indicators, which, moreover, don't always correspond to the system of national statistics. This fact creates a multiplier effect and complicates the assessment in dynamics. There is also a methodological problem to select indicators to be included in the index and with the identification of weights of partial indicators.

Considerable attention of researchers in assessing sustainable development is paid to the application of methods of multidimensional statistical analysis. The authors of [12] considered the construction of cluster models describing the development of agritourism in the context of ensuring sustainable development of rural areas. This allowed us to identify the leading factors of the sustainable development of these areas.

The article [23] presents an approach to assessing the sustainable development of regions, which is based on a combination of methods of factor and cluster analysis. The first is used to construct integrated partial indices of sustainable development for each of its components, and the second - to group regions by the level of sustainable development. The authors propose to use the obtained cluster map of regents and municipalities for decision-makers to take action, to take measures, and determine appropriate policies to solve problems in each region.

Separately in the toolkit of multidimensional statistical analysis, it is necessary to allocate methods of multidimensional scaling (MDS) which are directed on the identification of the structure of the set of studied objects. To this end, they are reflected in some space of latent characteristics, which adequately models reality. It is built based on a matrix of measures that reflect the pairwise similarity between objects. The resulting configuration of objects allows us to conclude their set internal structure. For a long time, this area of multidimensional statistical analysis was not given due attention, primarily due to the complexity of computational procedures, especially for the group of methods of non-metric multidimensional scaling. However, the development of software tools that implement these methods has eliminated this shortcoming and opened the horizons of their application, including for economic research.

J. J. de Jongh and D. F. Meyer in [5] used MDS tools to build a multidimensional regional development index (MREDI). The authors presented the practical implementation of the developed tools for rural municipalities in the North-West province of South Africa, identified trends in the proposed indicator. We support the conclusions made in the article on the feasibility of using the MDA to measure the economic development of regions.

The application of MDA technology to study the dynamics of economic growth in the world and the impact on it of globalization, scientific and technological progress, competition is presented in the works of J. A. Tenreiro Machado and M. E. Mata [29; 30; 28]. The authors substantiated the evolution of the main indicators of economic growth, globalization, prosperity, and development of the human world economy, established periods of prosperity and crisis, growth and stagnation

Paper [33] considers the procedure for assessing the social cohesion of the counties of Lower Silesia in the period 2005-2015 based on MDS tools in combination with linear ordering and Theil decomposition. The application of MDS together with cluster analysis to assess the development of farms in Kenya and Zimbabwe is presented in the study [22]. The method was implemented in two samples using expert data processing.

Note that the analyzed studies use methods of metric multidimensional analysis, which are focused on the use of quantitatively measurable indicators. In practice, it is often necessary to deal with non-metric indicators, measured in particular in the ranking scale or obtained from the expert evaluation. In such cases, it is advisable to use methods of non-metric multidimensional scaling.

In this study, we propose a new approach for application of the method of non-metric method of multidimensional scaling (NMMDS) to determine the configuration of regions of Ukraine by indicators of economic development, which have a non-metric nature and study their structure in the context of sustainable development.

2 Research methodology

In multidimensional scaling, the source data matrix is a square symmetric matrix of objects' differences measures:

$$\Delta = \begin{bmatrix} \delta_{11} & \delta_{12} & \dots & \delta_{1m} \\ \delta_{21} & \delta_{22} & \dots & \delta_{2m} \\ \dots & & & \\ \delta_{m1} & \delta_{m2} & \dots & \delta_{mm} \end{bmatrix}, \tag{1}$$

where δ_{ij} is a measure of the difference between *i*-th and *j*-th points (corresponding to objects), what is proportional to the distance between them;

m – number of objects, to be analyzed.

The NMMDA assumes that the measure of difference is a monotonic function of distance:

$$\delta_{ij} = \mathrm{f}(d_{ij}) = f\left(\sqrt[s]{\sum_{k=1}^{p} |u_{ik} - u_{jk}|^{s}}\right), \tag{2}$$

where U is an of objects' coordinates in the new space;

p is a dimension of a new scale space;

s – Minkowski metric index for calculating the distance in space.

Neither the coordinates of the u_{ij} j objects in this space nor the dimension of the space p itself are known. The value of s is chosen in the study process based on substantive considerations. Equation (2) means that the measures of differences are calculated from the original data corresponding to the distances between objects in the new space. Therefore, the problem of metric scaling is that based on the known matrix of differences of objects (1), which is calculated by the values of the original features X_1, X_2, …, X_n we need to find the coordinates of objects in the new scale-space U_1, U_2, …, U_p under the condition that the ranking order of the estimates of distances calculated by them is as close as possible to the ranking order of the initial values.

An important component of the nonmetric scaling algorithm is the degree of correspondence, which shows how the estimates of the coordinates of the stimuli reproduce the rank order of the data δ_{ij}. Typically, it contains three sets of parameters. The first includes the estimates of the coordinates u_{ij}. The second set contains estimates of the distances d_{ij}. The third set consists of fictitious parameters called rank data images, or deviations. They are calculated to be as close as possible to the distance estimates. Deviations to determine the degree of compliance can be calculated using the stress formulas proposed by J. B. Kruskal [13]:

$$S_1 = \sqrt{\frac{\sum_{i=1}^{m}\sum_{j=1}^{m}(\delta_{ij} - d_{ij})^2}{\sum_{i=1}^{m}\sum_{j=1}^{m}(d_{ij})^2}}, \tag{3}$$

$$S_2 = \sqrt{\frac{\sum_{i=1}^{m}\sum_{j=1}^{m}(\delta_{ij} - d_{ij})^2}{\sum_{i=1}^{m}\sum_{j=1}^{m}(d_{ij} - d)^2}}, \tag{4}$$

$$d = \frac{1}{m^2}\sum_{i=1}^{m}\sum_{j=1}^{m} d_{ij}. \quad (5)$$

The algorithm of NMMDA consists of four stages:

- Determining the starting configuration;
- Standardization of distances and coordinates' estimates;
- Calculation of deviations;
- Calculation of new coordinates' estimates.

The first stage is performed only once. It sets the initial estimates of the x_{ik} stimulus coordinates in multidimensional scale space. These estimates can be calculated in different ways. One of them is the use of values obtained by the algorithm of metric scaling Torgerson [4]. A necessary condition is that the estimates obtained by any method must be standardized.

The second stage is the standardization of coordinate estimates and the standardization of distances between stimuli. If this is the first time this step has been performed, the coordinate estimates are not standardized because they are already standardized, which reduces the likelihood of obtaining a degenerate solution.

The third stage involves the calculation of deviations. It is assumed that the values of δ_i are ordered by increasing their values. This stage is often called the non-metric stage of NMMDA. It does not change estimates of stimulus coordinates and distance estimates. It only provides for a change (recalculation) of estimates of differences.

In the first step of this stage, the data of the matrix of differences *D* are written in one column in ascending order (zero values of the matrix, ie the difference between the object itself, is not taken into account). If there is an equality of values of differences for two pairs of stimuli, ie $\delta_{ij} = \delta_{sr}$, then the first indicates the value of the difference for the pair of stimuli for which the corresponding value of the distance is less.

The next step of the third stage is the implementation of a series of views on the ordered data. The purpose of each pass is to divide the deviation estimates into blocks of equal values. Before the first view, these deviations are the values of the matrix of standardized distances obtained in the previous stage. If all values are different, then each of them forms a separate block. If some adjacent values are the same, they form one block. Next, the values of adjacent blocks are compared. If the value of the block with the larger number is less than the value of the block with the smaller number, then the blocks are combined into one, and their value is equal to the arithmetic mean of the values of the blocks. The data view is completed after analyzing the values of all blocks. If blocks have been merged, the next pass is performed. That is, the passes are completed if the next step the values of the blocks are ordered in ascending order. This completes the non-metric stage of multidimensional scaling.

In the fourth stage, the coordinates of the stimuli and the distances between the stimuli are recalculated. The new coordinates are calculated by the formula:

$$x_{ij}^{k+1} = x_{ij}^{k} - \frac{1}{m}\sum_{t=1}^{m}\left(1 - \frac{\delta_{it}^{k+1}}{d_{it}^{k}}\right)\left(x_{ij}^{k} - x_{tj}^{k}\right), \quad (6)$$

If $d_{it}^{k} = 0$, then relation $\frac{\delta_{it}^{k+1}}{d_{it}^{k}}$ is arbitrarily set equal to 1. New measures of distance are standardized so that the sum of the squares of their values is equal to 1. This step is not mandatory, but to some extent can simplify the calculations.

After the completion of the fourth stage, the value of the stress formula S^{k+1} is calculated, which is then compared with the value calculated in the previous iteration S^{k}. If the deviation is less than a predetermined small value of ε> 0, the calculations are completed. Otherwise, steps 2, 3, 4 are repeated.

3 Results and discussion

Within the framework of the approach described, let us assess the economic development of Ukraine's regions in the context of their sustainable development. As mentioned above, the sustainable development of the region can be seen as a process of improving the functioning of all its subsystems, including economic. From these considerations, we will form a system of initial indicators. Recently, it is quite common to calculate and use a variety of ratings that reflect the level of regional development and are calculated based on the key performance indicators of all region's subsystems. In particular, such estimates are presented by the Ministry of Development of Communities and Territories of Ukraine [15]. They can be used as non-metric indicators of economic development. Thus, for the calculations we have chosen the following set of non-metric indicators, which are rank estimates: X_1 – Investment and innovation development and foreign economic cooperation; X_2 – Financial self-sufficiency; X_3 – Labor market efficiency.

The data source for determining indicators are materials of the Ministry of Development of Communities and Territories of Ukraine [18; 17; 19]. We will select data for the period from 2016 to 2019.

To compactly present the information, we will assign a code to each region. Relevant information is given in table 1. The values of the original data are presented in table 2.

Table 1. Relations between the name of regions and their codes.

Code	Region	Code	Region	Code	Region
r_1	Vinnytsia	r_9	Kyiv	r_17	Sumy
r_2	Volyn	r_10	Kyrovohrad	r_18	Ternopil
r_3	Dnipro	r_11	Luhansk	r_19	Kharkiv
r_4	Donetsk	r_12	Lviv	r_20	Kherson
r_5	Zhytomyr	r_13	Mykolaiv	r_21	Khmelnytskyi
r_6	Zakarpattia	r_14	Odesa	r_22	Cherkasy
r_7	Zaporizhzhia	r_15	Poltava	r_23	Chernivtsi
r_8	Ivano-Frankivsk	r_16	Rivne	r_24	Chernihiv

Table 2. Values of initial data for 2016-2019.

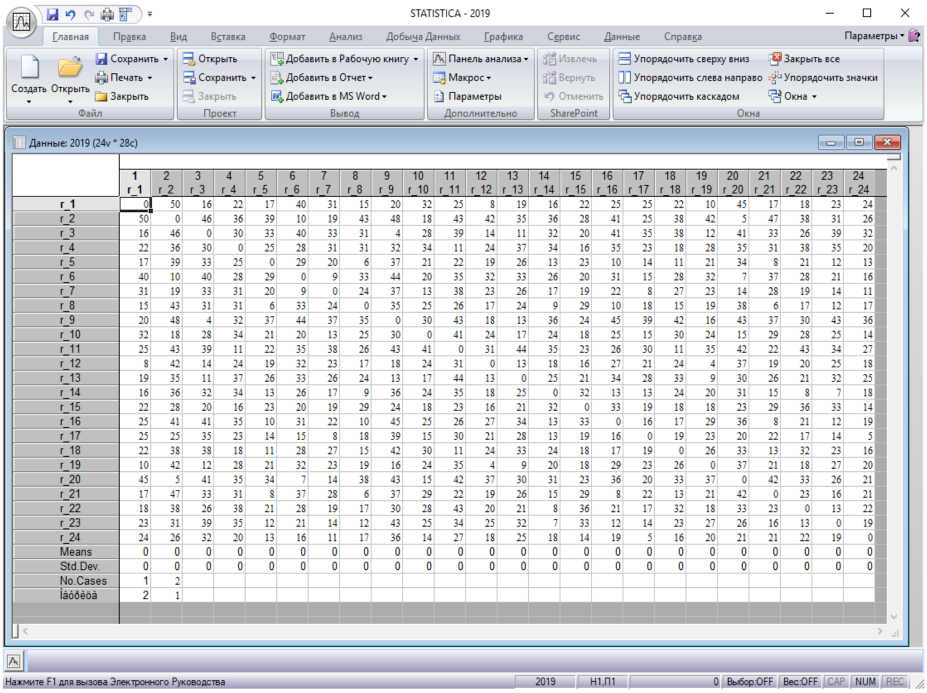
Code	Values											
	2016			2017			2018			2019		
	X_1	X_2	X_3	X_1	X_2	X_3	X_1	X_2	X_3	X_1	X_2	X_3
r_1	2	5	3	10	4	15	19	6	9	16	5	6
r_2	16	24	17	21	5	23	20	18	21	1	24	22
r_3	5	2	2	2	3	3	4	7	3	6	3	2
r_4	1	4	23	3	23	24	9	23	24	14	2	23
r_5	19	15	5	12	12	12	16	16	15	19	12	13
r_6	3	22	20	13	11	17	21	1	16	8	22	21
r_7	4	21	10	4	21	14	11	2	10	9	20	15
r_8	23	14	9	9	19	9	24	19	11	20	13	9
r_9	9	8	1	15	2	2	2	5	2	5	1	1
r_10	13	23	15	18	16	20	10	20	20	2	15	14
r_11	24	9	24	24	24	22	17	24	23	22	4	24
r_12	17	11	7	11	10	5	6	3	5	11	7	7
r_13	10	19	21	22	17	10	7	11	6	4	10	4
r_14	20	20	11	8	1	4	3	4	4	17	18	8
r_15	14	1	16	7	14	19	8	10	22	7	6	18
r_16	22	6	8	14	6	18	22	17	12	24	16	12
r_17	8	10	22	19	13	6	15	14	7	13	17	16
r_18	7	16	12	5	15	21	18	22	19	21	9	19
r_19	6	3	4	16	9	1	5	8	1	10	8	5
r_20	3	23	20	13	21	18	6	18	13	21	12	19
r_21	23	11	10	14	13	14	17	7	8	15	17	13
r_22	15	19	3	12	12	13	20	8	11	11	7	14
r_23	18	21	11	23	9	8	23	22	7	18	13	6
r_24	12	14	17	1	15	17	1	20	16	12	18	18

Given the origin of the data, as initial estimates of the objects' differences we take the measure of the Hamming's distance of (the city-block distance), calculated in the space of the original features:

$$\delta_{ij} = \sum_{k=1}^{n} |x_{ik} - u_{jk}|. \tag{7}$$

The algorithm of NMMDS is quite a time consuming from a computational point of view. An additional complication is the large dimension of the matrices that will be processed, as they will contain information about 24 regions of Ukraine. To automate the calculations, we will use special software that allows statistical data processing. One such tool is the Statistica software (Russian Localization), developed by StatSoft Inc, which has built-in tools for multidimensional scaling. The dimension of the new space is chosen to be 2. The starting configuration of the studied objects in the program is determined automatically.

The fragment of the Statistica's window with the entered values of the differences' matrix, calculated according to 2019 data, is shown in figure 1.



	1 r_1	2 r_2	3 r_3	4 r_4	5 r_5	6 r_6	7 r_7	8 r_8	9 r_9	10 r_10	11 r_11	12 r_12	13 r_13	14 r_14	15 r_15	16 r_16	17 r_17	18 r_18	19 r_19	20 r_20	21 r_21	22 r_22	23 r_23	24 r_24
r_1	0	50	16	22	17	40	31	15	20	32	25	8	19	16	22	25	25	22	10	45	17	18	23	24
r_2	50	0	46	36	39	10	19	43	48	18	43	42	35	36	28	41	25	38	42	5	47	38	31	26
r_3	16	46	0	30	33	40	33	31	4	28	39	14	11	32	20	41	35	38	12	41	33	26	39	32
r_4	22	36	30	0	25	28	31	31	32	34	11	24	37	34	16	35	23	18	28	35	31	38	35	20
r_5	17	39	33	25	0	29	20	6	37	21	22	19	26	13	23	10	14	11	21	34	8	21	12	13
r_6	40	10	40	28	29	0	9	33	44	20	35	32	33	26	20	31	15	28	32	7	37	28	21	16
r_7	31	19	33	31	20	9	0	24	37	13	38	23	26	17	19	22	8	27	23	14	28	19	14	11
r_8	15	43	31	31	6	33	24	0	35	25	26	17	24	9	29	10	18	15	19	38	6	17	12	17
r_9	20	48	4	32	37	44	37	35	0	30	43	18	13	36	24	45	39	42	16	43	37	30	43	36
r_10	32	18	28	34	21	20	13	25	30	0	41	24	17	24	18	25	15	30	24	15	29	28	25	14
r_11	25	43	39	11	22	35	38	26	43	41	0	31	44	35	23	26	30	11	35	42	22	43	34	27
r_12	8	42	14	24	19	32	23	17	18	24	31	0	13	18	16	27	21	24	4	37	19	20	25	18
r_13	19	35	11	37	26	33	26	24	13	17	44	13	0	25	21	34	28	33	9	30	26	21	32	25
r_14	16	36	32	34	13	26	17	9	36	24	35	18	25	0	32	13	13	24	20	31	15	8	7	18
r_15	22	28	20	16	23	20	19	29	24	18	23	16	21	32	0	33	19	18	18	23	29	36	33	14
r_16	25	41	41	35	10	31	22	10	45	25	26	27	34	13	33	0	16	17	29	36	8	21	12	19
r_17	25	25	35	23	14	15	8	18	39	15	30	21	28	13	19	16	0	19	23	20	22	17	14	5
r_18	22	38	38	18	11	28	27	15	42	30	11	24	33	24	18	17	19	0	26	33	13	32	23	16
r_19	10	42	12	28	21	32	23	19	16	24	35	4	9	20	18	29	23	26	0	37	21	18	27	20
r_20	45	5	41	35	34	7	14	38	43	15	42	37	30	31	23	36	20	33	37	0	42	33	26	21
r_21	17	47	33	31	8	37	28	6	37	29	22	19	26	15	29	8	22	13	21	42	0	23	16	21
r_22	18	38	26	38	21	28	19	17	30	28	43	20	21	8	36	21	17	32	18	33	23	0	13	22
r_23	23	31	39	35	12	21	14	12	43	25	34	25	32	7	33	12	14	23	27	26	16	13	0	19
r_24	24	26	32	20	13	16	11	17	36	14	27	18	25	18	14	19	5	16	20	21	21	22	19	0
Means	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Std.Dev.	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
No.Cases	1	2																						
Ìàòðèöà	2	1																						

Fig. 1. Matrix of differences for data 2019.

As a result of the calculations performed according to the above non-metric scaling algorithm, implemented in Statistica, we obtain the coordinates of point-objects (regions) in the new scale space (table 3). Note that the calculations were performed separately for each year's data.

A graphical representation of the regions in the space of latent scales is shown in figures 2-5. Analysis of the charts presented allows us to conclude that there are no trends for the sustainability of economic development for Ukraine's regions during the period under study. The location of regions in the built scale-space doesn't meet certain patterns or dependencies.

Table 3. Coordinates' values of point-objects (regions) in the new scale space.

Code	Values							
	2016		2017		2018		2019	
	Axis1	Axis2	Axis1	Axis2	Axis1	Axis2	Axis1	Axis2
r_1	-0.55	0.15	-0.81	0.64	-1.00	0.58	0.57	0.84
r_2	-0.32	0.30	0.16	0.52	0.84	0.71	0.44	-0.09
r_3	-0.34	-0.36	0.14	-0.18	0.32	-0.86	-0.11	0.50
r_4	0.12	-0.19	-0.34	0.41	-0.70	0.00	0.55	0.27
r_5	-0.40	1.07	1.11	-0.89	-1.26	0.09	-1.28	-0.04
r_6	0.74	0.49	0.74	1.05	0.03	0.07	-0.01	-0.94
r_7	-0.49	0.80	1.00	0.58	-0.47	-0.95	0.28	-0.36
r_8	0.13	0.81	-0.58	-0.94	-0.29	0.70	-0.66	-0.54
r_9	0.99	-0.03	-0.46	-0.20	-0.24	0.37	-0.01	0.25
r_10	0.86	-0.44	0.27	-0.88	-0.20	-1.25	0.94	-0.58
r_11	-0.49	-0.14	0.08	-0.13	0.21	-0.37	0.58	-0.61
r_12	-0.10	-1.12	-1.18	-0.27	0.56	0.75	-0.14	0.78
r_13	-1.11	-0.39	1.05	0.10	-1.13	-0.38	-0.52	0.85
r_14	-1.23	0.15	0.66	0.19	0.89	0.05	-1.18	0.37
r_15	0.27	-0.66	-1.25	0.28	-0.22	1.10	1.15	0.25
r_16	1.09	0.27	-0.09	-1.26	-0.87	-0.82	-0.72	0.01
r_17	0.29	1.18	0.31	1.11	0.28	1.19	0.18	-0.45
r_18	0.60	-1.14	1.01	-0.39	0.78	-0.47	0.09	-1.07
r_19	0.35	0.25	0.71	-0.64	-0.30	-0.51	-0.32	1.21
r_20	1.02	-0.79	-0.57	1.08	-0.74	0.96	1.31	-0.47
r_21	1.04	0.92	-0.11	1.14	0.75	-1.05	-0.37	-0.80
r_22	-1.09	0.73	-0.99	-0.67	1.31	0.29	-1.16	0.50
r_23	-0.85	-0.72	-0.64	0.08	1.01	-0.42	-0.95	-0.72
r_24	-0.51	-1.14	-0.24	-0.74	0.47	0.21	1.35	0.84

This result can be explained by the properties of the indicators that were selected for the calculations: according to the data, there is also no clear trend in the ranking of regions of Ukraine during the study period. On the other hand, the country's economy during this period was exposed to various destabilizing internal and external influences, which also affected the disparities in regional development.

Conclusions obtained by the results of calculations coincides with the conclusion made in paper [9], where the sustainable economic development of Ukraine's regions is based on a comprehensive index according to the State Statistics Service of Ukraine [6] also has been assessed.

The presented location of regions also complicates the interpretation of axes (new scales). To solve this problem, a well-known effective procedure is the rotation of the constructed scale space, which is described by the formula:

$$U^{(1)} = U^{(0)}W, \tag{8}$$

where $U^{(1)}$ – new scale space of latent characteristics;

$U^{(0)}$ – initial scale space of latent characteristics ;

W – rotation matrix, which sets the rotation of the axes of the initial scale space by some angle φ counterclockwise.

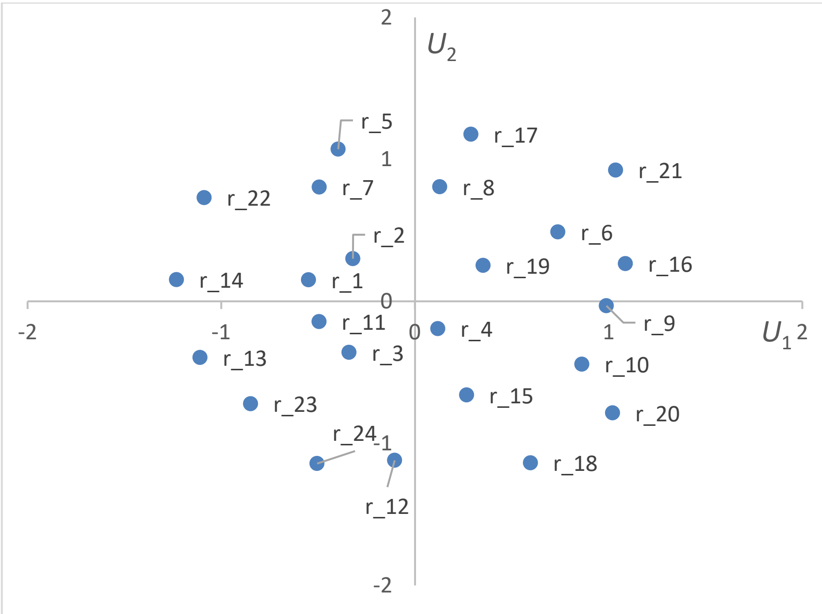


Fig. 2. Imaging regions in the new scale space for data of 2016.

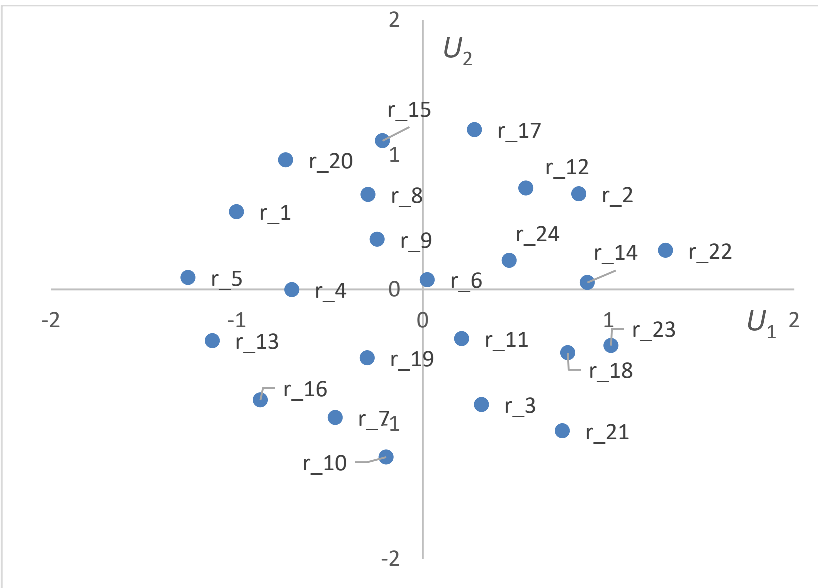


Fig. 3. Imaging regions in the new scale space for data of 2017.

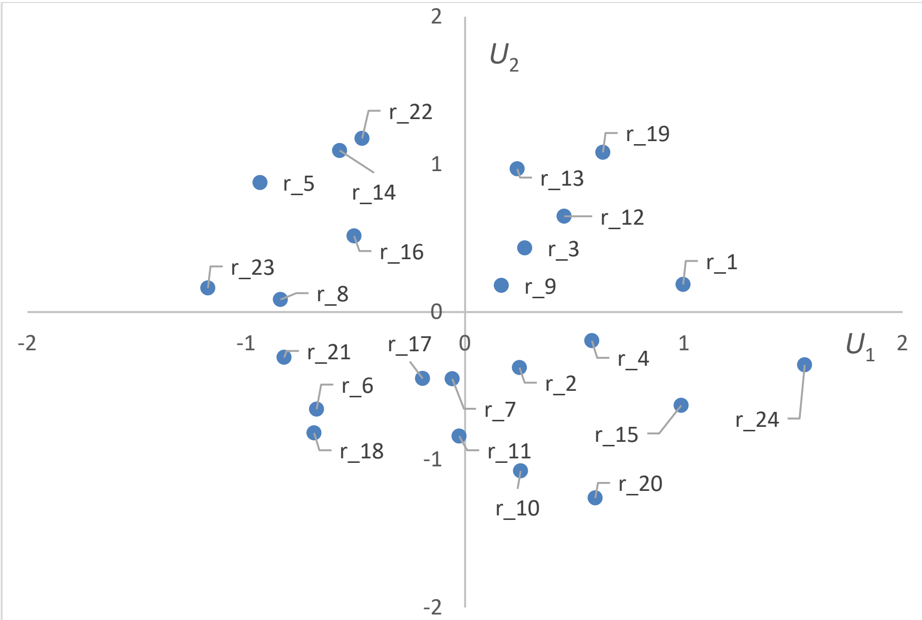


Fig. 4. Imaging regions in the new scale space for data of 2018.

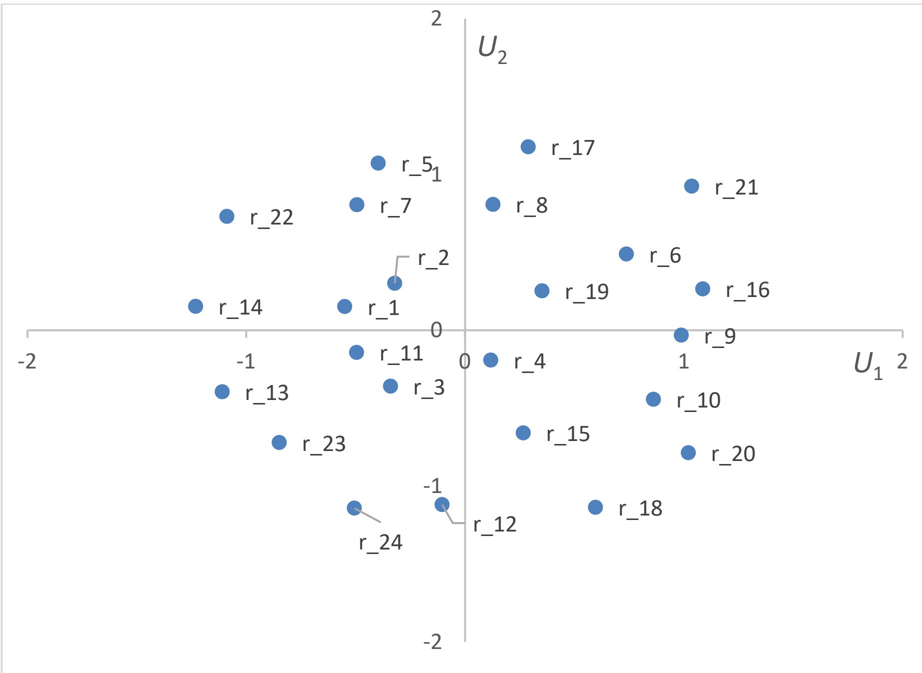


Fig. 5. Imaging regions in the new scale space for data of 2019.

The rotation matrix has a form:

$$W = \begin{bmatrix} \cos\varphi & -sin\varphi \\ sin\varphi & \cos\varphi \end{bmatrix}. \tag{9}$$

Currently, there are no formal rules that would allow analytically to justify the value of the angle of rotation φ for the best interpretation of the axes. This is usually done experimentally.

Let us illustrate the rotation of the scale-space for the graph constructed according to the data of 2019. In our study, it was found that an acceptable result is the rotation of the scale-space by the angle φ=45. The corresponding rotation matrix has the form:

$$W = \begin{bmatrix} 0.71 & -0.71 \\ 0.71 & 0.71 \end{bmatrix} \tag{10}$$

and the matrix $U_{2019}^{(1)}$, which reflects the coordinates of the points in the new space, consist of elements, presented in table 4:

Table 4. Coordinates of points in the new space for data 2019.

Code	Values	
	2019	
	Axis1	Axis2
r_1	1.00	0.19
r_2	0.25	-0.38
r_3	0.27	0.43
r_4	0.58	-0.19
r_5	-0.94	0.88
r_6	-0.68	-0.66
r_7	-0.06	-0.45
r_8	-0.84	0.08
r_9	0.17	0.18
r_10	0.25	-1.07
r_11	-0.03	-0.84
r_12	0.45	0.65
r_13	0.24	0.97
r_14	-0.57	1.09
r_15	0.99	-0.63
r_16	-0.51	0.52
r_17	-0.19	-0.45
r_18	-0.69	-0.82
r_19	0.63	1.08
r_20	0.60	-1.26
r_21	-0.83	-0.31
r_22	-0.47	1.18
r_23	-1.18	0.16
r_24	1.55	-0.36

Imaging regions in the scale space after the axis rotation for data of 2019 is shown in figure 6. Comparing the obtained configuration with the original data, we can conclude that the abscissa axis (U_1) can be interpreted as an axis that reflects the economic potential of the regions, and the U_2 axis – as the labor potential of the regions.

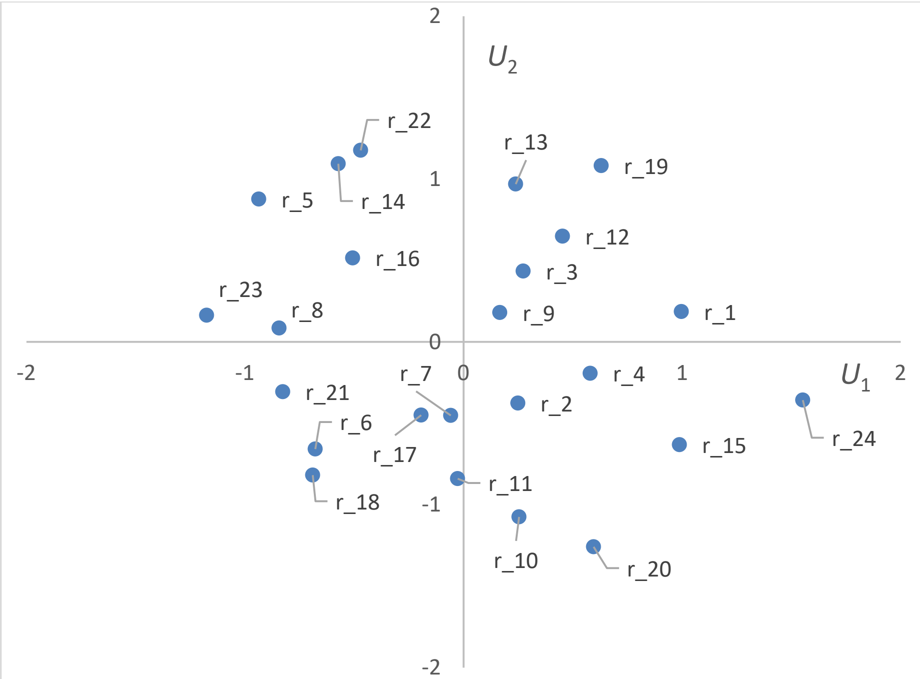


Fig. 6. Imaging regions in the scale space after the axis rotation for data of 2019.

Note that each data collection may have its rotation matrix. Moreover, there may be a situation where the rotation of space does not give the desired effect in terms of a reasonable interpretation of the axes.

A possible way out of this situation is to increase the number of initial indicators due to the condition of their close correlation.

4 Conclusions

Assessing the economies of regions in the context of ensuring their sustainable development remains an urgent problem. Studies have shown a wide range of methods for a comprehensive assessment of the economic development of regions. It is established that the multidimensionality of the description of economic development determines the use of methods of multidimensional statistics, in particular multidimensional scaling.

The scientific novelty of the proposed approach is the use of a set of non-metric indicators to assess the regions' economic development, which led to the further implementation of the methodology of multidimensional scaling in economic studies.

The approach for structuring regions in the context of sustainability of their economic development is considered in the paper. The results received a semantic interpretation, to improve which is proposed to rotate the scale space. The obtained results can be used as a basis for the formation of development strategies at both regional and national levels, as well as to assess the implementation of economic, social, and environmental aspects of sustainable development in the regions. The direction of further research is to expand the set of initial indicators for the application in multidimensional scaling, in particular, including the quantitative measured indicators into the initial data collection.

References

1. Artiushok, K.A.: Kryterii ta indykatory ekonomichnoi bezpeky i zbalansovanoho rozvytku rehionu (Criteria and Indicators for Economic Security and Balanced Development of the Region). Balanced Nature Using **3**, 58–63 (2016)
2. Bakri, B., Rustiadi, E., Fauzi, A., Adiwibowo, S.: Assessment of Regional Sustainable Development in Indonesia. International Journal of Humanities and Social Science **6**(11), 191–200. https://www.ijhssnet.com/journals/Vol_6_No_11_November_2016/22.pdf (2016). Accessed 21 Mar 2020
3. Churikanova, O.: Analiz indykatoriv staloho rozvytku (The Analysis of Sustainable Development Indicators). Ekonomika ta derzhava **2**, 56–60 (2017)
4. Davison, M.L., Sireci, S.G.: Multidimensional Scaling. In: Tinsley, H.E.A., Brown, S.D. (eds.) Handbook of Applied Multivariate Statistics and Mathematical Modeling, pp. 323-352. Academic Press, New York (2000)
5. De Jongh, J.J., Meyer, D.F.: The Multidimensional Regional Economic Development Index (MREDI) Applied in the North-West Province: A Rural Regional Application. Administratio Publica **27**(3), 162–185 (2019)
6. Derzhavna sluzhba statystyky Ukrainy (State Statistics Service of Ukraine) (2020). http://www.ukrstat.gov.ua/. Accessed 20 Aug 2020
7. Garkavaya, V.G.: Integrirovannaya ocenka ustoychivosti razvitiya regionov (Integrated Assessment of Regional Development Sustainability). http://www.rusnauka.com/CCN/Economics/13_garkavaja.doc.htm. Accessed 28 Nov 2019
8. Horiana, I.V.: Formuvannia metodyky otsiniuvannia stalls rozvytku rehioniv (Formation Evaluation Methods Sustainability of Regional Development). Ekonomichnyy Analiz **14**(1) (2013)
9. Hryhoruk, P., Khrushch, N., Grygoruk, S.: Assessment model of regions' economy in the context of their sustainable development. E3S Web of Conferences **166**, 13023 (2020). doi:10.1051/e3sconf/202016613023
10. Hryhoruk, P., Khrushch, N., Grygoruk, S.: The rating model of Ukraine's regions according to the level of economic development. Periodicals of Engineering and Natural Sciences **7**(2), 712–722 (2019). doi:10.21533/pen.v7i2.555
11. Huan, Y., Li, H., Liang, T.: A New Method for the Quantitative Assessment of Sustainable Development Goals (SDGs) and a Case Study on Central Asia. Sustainability **11**(13), 3504 (2019). doi:10.3390/su11133504
12. Kazlouski, V., Ganski, U., Platonenka, A., Vitun, S., Sabalenka, I.: Sustainable Development Modeling of Agritourism Clusters. Management Theory and Studies for

Rural Business and Infrastructure Development **42**(2), 118–127 (2020). doi:10.15544/mts.2020.12

13. Kruskal, J.B., Wish, M.: Multidimensional Scaling. Sage, Beverly Hills (1978). doi:10.4135/9781412985130
14. Mederly, P., Novacek, P., Topercer, J.: Sustainable development assessment quality and sustainability of life indicators at global, national and regional level. Foresight **5**(5), 42–49 (2003). doi:10.1108/14636680310507307
15. Ministry for Communities and Territories Development of Ukraine. https://www.minregion.gov.ua/en/ (2020). Accessed 20 Aug 2020
16. Molotokiene, E.: A Philosophical Analysis of the Concept of Sustainable Development. Regional Formation and Development Studies **31**(2), 148–156 (2020). doi:10.15181/rfds.v31i2.2106
17. Monitorynh sotsialno-ekonomichnoho rozvytku rehioniv za 2017 rik (Monitoring of socio-economic development of regions for 2017). https://www.minregion.gov.ua/wp-content/uploads/2018/05/Reytingova-otsinka-za-2017-rik-prezentatsiyni-materiali.pdf (2018). Accessed 17 Aug 2020
18. Monitorynh sotsialno-ekonomichnoho rozvytku rehioniv za 2018 rik (Monitoring of socio-economic development of regions for 2018). https://www.minregion.gov.ua/wp-content/uploads/2019/05/Reytingova-otsinka-za-2018-rik-prezentatsiyni-materiali.pdf (2019). Accessed 20 Aug 2020
19. Monitorynh sotsialno-ekonomichnoho rozvytku rehioniv za 2019 rik (Monitoring of socio-economic development of regions for 2019). https://www.minregion.gov.ua/wp-content/uploads/2020/05/rejtyngova-oczinka-za-2019-rik-prezentaczijni-materialy-2.pdf (2020). Accessed 17 Aug 2020
20. Nesterenko, O.O.: Indykatory otsinky rivnia staloho rozvytku ta yikh vplyv na pokaznyky intehrovanoi zvitnosti (Sedimentary Development Evaluation Indicators and their Effect on Integrated Reporting Indicators). Scientific Bulletin of Uzhhorod University, Series “Economics” **15**(2), 44–50 (2017)
21. Omarov, Sh.A.: Otsinka staloho rozvytku rehioniv Ukrainy (Evaluation of Sustainable Development of Regions of Ukraine). The Problems of Economy **3** (2014)
22. Pacini, G.C., Colucci, D., Baudron, F., Righi, E., Corbeels, M., Tittonell, P., Stefanini, F.M.: Combining multi-dimensional scaling and cluster analysis to describe the diversity of rural households. Experimental Agriculture **50**(3), 376–397 (2014)
23. Pravitasari, A.E., Rustiadi, E., Mulya, S.P., Fuadina, L.N.: Developing Regional Sustainability Index as a New Approach for Evaluating Sustainability Performance in Indonesia. Environment and Ecology Research **6**(3), 157–168 (2018). doi:10.13189/eer.2018.060303
24. President of Ukraine: Pro Stratehiiu staloho rozvytku “Ukraina – 2020” (On Strategy of Sustainable Development “Ukraine – 2020”). https://zakon.rada.gov.ua/laws/show/5/2015 (2015). Accessed 27 Dec 2019
25. President of Ukraine: Pro Tsili staloho rozvytku Ukrainy do 2030 roku (On the Goals of Sustainable Development of Ukraine up to 2030). https://zakon.rada.gov.ua/laws/show/722/2019 (2019). Accessed 25 Oct 2019
26. Rahma, H., Fauzi, A., Juanda, B., Widjojanto, B.: Development of a Composite Measure of Regional Sustainable Development in Indonesia**.** Sustainability **11**(20), 5861 (2019). doi:10.3390/su11205861
27. Tamenova, S., Satpayeva, Z., Sarkytova, A., Nurgissayeva, A.: Assessment of sustainable development of region. IOP Conference Series: Earth and Environmental Science **534**, 012008 (2020). doi:10.1088/1755-1315/534/1/012008

28. Tenreiro Machado, J.A., Mata, M.E.: Analysis of World Economic Variables Using Multidimensional Scaling. PLoS ONE **10**(3), e0121277 (2015). doi:10.1371/journal.pone.0121277
29. Tenreiro Machado, J.A., Mata, M.E.: Multidimensional Scaling Analysis of the Dynamics of a Country Economy. The Scientific World Journal **2013**, 594587 (2013). doi:10.1155/2013/594587
30. Tenreiro Machado, J.A., Mata, M.E.: Multidimensional scalling analysis of the world economy during the period 1972-2012. Acta Polytechnica Hungarica **12**(1), 67–82 (2015)
31. United Nations Department of Economic and Social Affairs Disability: #Envision2030: 17 goals to transform the world for persons with disabilities. https://www.un.org/development/desa/disabilities/envision2030.html (2015). Accessed 29 Nov 2019
32. Villeneuve, C., Tremblay, D., Riffon, O., Lanmafankpotin, G.Y., Bouchard, S.: A Systemic Tool and Process for Sustainability Assessment. Sustainability **9**(10), 1909 (2017). doi:10.3390/su9101909
33. Walesiak, M.: The Application of Multidimensional Scaling to Measure and Assess Changes in the Level of Social Cohesion of the Lower Silesia Region in the Period 2005-2015. Econometrics 3(57), 9–25 (2017). doi:10.15611/ekt.2017.3.01
34. Zhuchenko, A.M.: Poniattia staloho rozvytku v suchasnii ekonomitsi (The Concept of Sustainable Development in Modern Economy). Global and National Problems of Economy **13** (2016)
35. Zimny, A., Jurczak, R.: Multidimensional Analysis of Sustainable Development in Countries of Central and Eastern Europe. Journal of Agribusiness and Rural Development, **39**(1), 207–214 (2016). doi:10.17306/JARD.2016.24

Predicting the economic efficiency of the business model of an industrial enterprise using machine learning methods

Liliana Horal[1[0000-0001-6066-5619]], Inesa Khvostina[1[0000-0001-5915-749X]], Nadiia Reznik[2[0000-0001-9588-5929]], Vira Shiyko[1[0000-0002-2822-0641]], Natalia Yashcheritsyna[1[0000-0002-2926-5550]], Svitlana Korol[1[0000-0002-4804-7612]] and Vladimir Zaselskiy[3[0000-0002-7517-5433]]

[1] Ivano-Frankivsk National Technical University of Oil and Gas, 15 Karpatska Str., Ivano-Frankivsk, 76019, Ukraine
liliana.goral@gmail.com, inesa.hvostina@gmail.com, viraSh@i.ua, yashcheritsyna@gmail.com, korolSV@i.ua
[2] National University of Life and Environmental Sciences of Ukraine, 15 Heroiv Oborony Str., Kyiv, 03041, Ukraine
nadya-reznik@ukr.net
[3] State University of Economics and Technology, 5 Stepana Tilhy Str., Kryvyi Rih, 50006, Ukraine
zaselskiy52@gmail.com

Abstract. The paper considers the problem of studying the impact of key determinants on the industrial enterprise business model economic efficiency and aims to build an optimal model for predicting the industrial enterprise business model effectiveness using neural boundaries. A system of key determinants key factors has been developed. Significant factors were later used to build neural networks that characterize the studied resultant trait development vector. The procedure for constructing neural networks was performed in the STATISTICA Neural Networks environment. As input parameters, according to the previous analysis, 6 key factor indicators were selected. The initial parameter is determined by economic efficiency. According to the results of the neural network analysis, 100 neural networks were tested and the top 5 were saved. The following types of neural network architectures, multilayer perceptron, generalized regression network and linear network were used. Based on the results of the neural network modeling, 5 multilayer perceptrons of neural network architectures were proposed. According to descriptive statistics, the best model was a multilayer perceptron, with the MLP 6-10-1 architecture, which identifies a model with 6 input variables, one output variable and one hidden layer containing 10 hidden neurons. According to the analysis of the sensitivity of the network to input variables, it was determined that the network is the most sensitive to the variable the share of electricity costs in total costs. According to the results of selected neural networks standard prediction, the hypothesis of the best neural network was confirmed as Absolute res., Squared res, Std. Res for the neural network MLP 6-10-1 reached the optimal value and indicate that the selected model really has small residues, which indicates a fairly high accuracy of the forecast when using it.

Keywords: neural networks, forecasting, business model, economic efficiency, digitalization, oil transportation company.

1 Introduction

The intensification of the digital transformation of industrial enterprises in general and oil transportation in particular necessitates the urgent introduction of digitalization tools and means, which should ultimately lead to an increase in the efficiency of economic entities. Given these circumstances, the issue of assessing the economic efficiency of the enterprise business model from the introduction of modern IT is relevant. Digitalization is one of the industrial enterprises' development key determinants, it is a field for the formation of economic efficiency of the enterprise business model. Therefore, the success of companies, including their performance indicators, depends on the effective use of modern IT in the activities of business entities. In such conditions, it is important to model the enterprise business model effectiveness in the digital transformation of economic systems.

2 Background

Examining the topic of modeling the effectiveness of economic entities, we can conclude that there are many methods, techniques, tools, methods of economic and mathematical modeling of the phenomenon. The difficulty of studying this phenomenon is that for different companies, the efficiency of activities can be affected by various factors. In addition, the concept of efficiency is not unambiguous, so it can be considered through the prism of various result indicators, influencing factors and their combinations. There is a huge variety of built efficiency models for different enterprises. Therefore, consider some of the proposed models of efficiency.

In particular, we can highlight economic production quantity model with learning in production, quality, reliability and energy efficiency. The article [9] of B. Marchi, S. Zanoni and M. Y. Jaber proposes a lot sizing model for a manufacturing company that includes the relevant learning outcomes, directly and indirectly, affecting its energy efficiency. The main result of the study is to show how learning in production and energy efficiency affects each other and the optimal lot size quantity [9]. In this paper, the model of the company's success is considered through optimization derived an economic order quantity model with controllable production rate.

A. Zakharov and S.-L. Jämsä-Jounela in [16] are suggested an iterative method for optimization of the plant profit rate is proposed avoiding the control saturation and is applied to the Pulp Mill benchmark model optimization. Three different static models describing the steady state values of the manipulated variables are constructed and used in the optimization. The results of the optimization are presented and compared against the straightforward single-step optimization of the plant economic efficiency [16].

In addition Agnieszka Bezat-Jarzębowska and Włodzimierz Rembisz proposed efficiency-focused economic modeling of competitiveness in the agri-food sector [2]. Authors believe that the main source of producers' competitiveness and growth is not the increase of input factors but the efficiency of their use. The efficiency-focused modeling presented in the paper bases on the production function, more precisely on the SFA method (the Stochastic Frontier Approach), which is appropriate primarily for samples with high randomness. In the analysis Cobb-Douglas and translogarithmic models are applied [2].

Katarína Teplická in the article [15] discussed methods of pricing for the selected product and examined the impact on economic efficiency of the enterprise. Through comparing of various methods could find out reserves in area of costs decreasing by the way it could satisfy demands of client at the level of target price. She compared calculating methods at the pricing of product. In present time customers influence considerably product's price by their demands and therefore producers, businessmen must nowadays adapt prices of their products to demands of customers [15].

Mykola Havrylenko, Vira Shiyko, Liliana Horal, Inesa Khvostina, and Natalia Yashcheritsyna [4] are propose two methods for evaluating the financial efficiency of a business model of industrial enterprises. In order to evaluate the financial efficiency of the business model of an industrial enterprise, a system of single indicators for assessing the financial condition of the enterprise by such components as financial stability, liquidity and solvency, business activity and profitability was formed. Fishburne's rule weights the major components of an integral measure of an enterprise's business model financial performance. In addition, an integral measure of the financial performance of the business model is modeled using the fuzzy set method and taxonomic analysis, which will help to evaluate the financial performance level of the business model more objectively. The comparative analysis of the obtained results by different methods of calculation of integral indicators is carried out [4].

When forming a business model of an enterprise, it is necessary to take into account risks in order to minimize them. One of the methods for diversifying risks is their insurance [8]. Insurance companies take these risks into account when forming a risk portfolio. The activities of any other enterprise are aimed at increasing the country's oil and gas economy. One of the important indicators of the economic state of the country is the gross domestic product [11], which depends on the successful operation of the enterprise on the basis of a well-built business model.

3 Methodology

Entering the digital space of industrial enterprises, including oil transportation companies, opens up new opportunities for these businesses, which are to use the following technologies and approaches: artificial intelligence and advanced analytics [12]; automation and robotics; digitization of business processes; use of the Internet of Things (IoT) [5], cloud computing [7], sensors, mobile devices [10], flexible development and design thinking.

The process of assessing the impact of these IT on the efficiency of the oil transportation company has remained unexplored [6]. However, to address this issue, industrial enterprises face problems in using such technologies. These include: low level of digital culture of domestic industrial enterprises, lack of leadership, expectation of economic opportunity for change, unclear economic effect of investment in the digitalization process, lack of the digital operations benefits clear vision and better management by senior management through the digitalization process. and non-production processes.

These problems of digitalization of industrial enterprises require the transformation of production and organizational structures of economic entities for the effective implementation of ICT. The new revolution requires the formation of such a system of organization and management of activities, which will ensure, based on the analysis of performance, not only the elimination of the causes of existing discrepancies, but also help to identify and prevent their occurrence.

Thus, there is a need to assess the effectiveness and efficiency of activities “before” and “after” the introduction of ICT and to monitor them constantly. Evaluating the effectiveness of enterprises business processes digitalization will allow on the basis of the digitalized business model of the enterprise to monitor the activities in general, identify problem areas of the entity, make optimal management decisions and experiment by using a digitalized business model (digital twin of the enterprise) without intervention into the actual production process.

It should be noted that the indicators for assessing the effectiveness of digitization may be different for different processes and can be both generalized and detailed for each business process or function.

Due to the large number of the digitalization process influence area on the efficiency of the industrial enterprise digitalization effectiveness determining model will be:

$$E_d = \sum_{i=1}^{n} \sum_{\mathrm{j}=1}^{\mathrm{m}} e_{ij} \tag{1}$$

where E_d – the overall efficiency of the enterprise digitalization; e_{ij} – unit efficiency of the i-th business process (function) from the j-th IT.

The tasks performed by the procedure of evaluating of an industrial enterprise business processes digitalization effectiveness is that:

- There is a search for problem areas during the interaction of employees of different departments in the process of information support and business processes;
- The main, additional and auxiliary directions of economic entities activity for the purpose of decomposition on digital business processes are defined;
- Prerequisites are formed for the construction of an orderly electronic document management system;
- The regulation of enterprise activity is established;

— The process of conducting experiments in the activities of the enterprise is possible in order to determine the strategies of enterprise development in the current conditions of enterprises operation.

To form a system for evaluating the effectiveness of digitalization of business processes of JSC "Ukrtransnafta" define its main stages, taking into account the peculiarities of the research enterprise. Figure 1 shows the stages of evaluating the JSC "Ukrtransnafta" business processes digitalization effectiveness.

1. • Construction of the enterprise virtual business model using the analysis of the work regulation information system and determination of the digital way of performing business processes.
2. • Formation of quantitative indicators system for evaluating the effectiveness of each IT in relation to the business process on key performance indicators.
3. • Calculation of parameters of quantitative indicators of business processes digitalization efficiency and formation of efficiency matrix in a ratio IT-business / business process.
4. • Analysis of the obtained values of the business processes digitalization efficiency corfficients.
5. • Forming conclusions about the business processes digitalization effectiveness. Development of business processes digitalization efficiency increase.

Fig. 1. Stages of evaluating the JSC "Ukrtransnafta" business processes digitalization effectiveness.

Given the peculiarities of the oil transportation companies activities, the business processes digitalization impact will be carried out directly on the tariffs. Thus, the model for determining the effectiveness of digitization must be built taking into account the sequence of the researched enterprise business processes. Figure 2 defines the sequence of JSC "Ukrtransnafta" business processes.

At each of the above stages of oil transportation in the conditions of digitalization it is necessary to enter the digitization coefficient, which will eventually be reduced to an integrated indicator of activities digitization, which will reflect the synergistic impact of each stage digitization coefficients. Thus, taking into account the peculiarities of the oil transportation company business processes, the integrated indicator of digitalization, which must be taken into account when forming the main performance indicator – the tariff will take the form:

$$k_d = \prod_{i=1}^{n} k_i \tag{2}$$

where k_d – synergetic digitization factor; k_i – digitization ratio on the i-th business process; i=1, …, n – the number of oil transportation business processes.

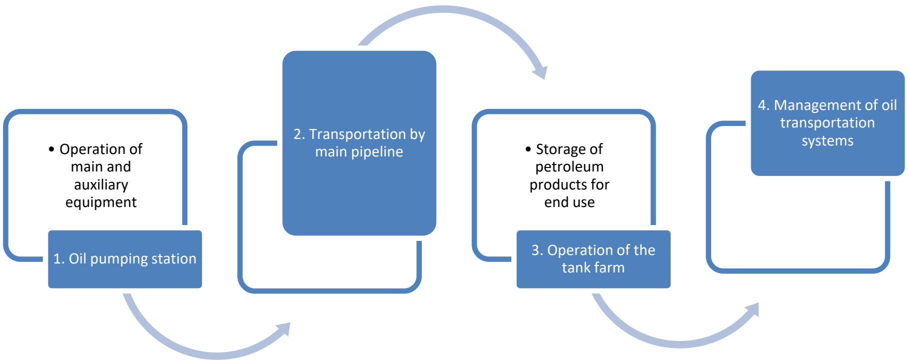


Fig. 2. The sequence of JSC "Ukrtransnafta" business processes.

It should be noted that tariffs for transportation of oil and oil products are established by the general rules of "Tariffs formation methods", which does not account for the development of alternative transportation markets, digitalization of the economy, accounting standards used time, resources, materials, regulated by the old technology, operating in the early 2000s.

Given that all of the above business processes form cost centers, it is necessary to determine the synergy of costs that arises from the digitalization of business processes. This synergetic effect of cost reduction will have a direct impact on the overall digitization ratio, which will increase economic efficiency. It is interesting for this company to form a correlation-regression model of the enterprise efficiency dependence through the introduction of ICT and reduce costs in each business process (fig. 2) of the studied enterprise. Based on the results of previous studies in the work, which indicate that for the investigated enterprise financial efficiency is at an average level and it increases at the end of the period, we consider it necessary to assess the economic efficiency of the oil company business model using computer economy methods to determine key success factors of the researched enterprise.

Therefore, the resulting indicator in the proposed correlation-regression model will be the efficiency as the ratio of net income from sales to the cost of goods sold. After all, when assessing the result of management must take into account not o nly the statement of the goal achievement, but also the optimal ways to achieve it. The choice of the regression or multifactor model type (analytical expression) depends on the type of factor features connection with the effective one. Thus, the initial form of the resulting function is reflected as a dependence:

$$Y = f(x_1; x_2; x_3 \dots, x_n), \tag{3}$$

where Y – effective feature-function of the enterprise efficiency; x_1, x_2, x_3, …, x_n – factor features.

Figure 3 shows the dynamics of the JSC "Ukrtransnafta" business model efficiency and the volume of transported oil in UAH million for 2010-2019.

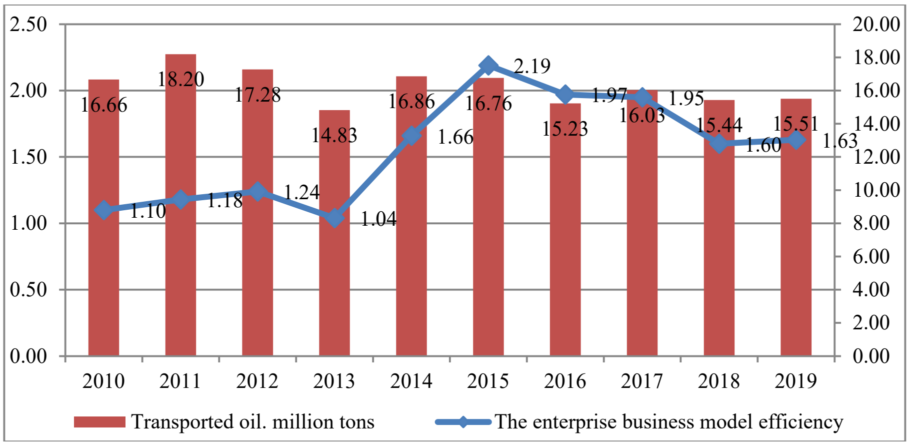


Fig. 3. Dynamics of JSC "Ukrtransnafta" business model efficiency and volume of transported oil.

Thus, from the results of calculating the enterprise business model efficiency and its relationship with the dynamics of transported oil, we can conclude that the change of the first indicator does not have a stable trend, in addition, at the end of the period there is an increase in the efficiency of the business model production efficiency (volumes of transported oil). Therefore, it is necessary to investigate and identify the factors that affect the growth of the business model economic efficiency, along with the decline in production activity of the enterprise, which is associated with both the global crisis and unstable international economic relations. To further enhance the positive impact and level the negative, provided it is impossible to turn it into a positive impact.

To determine the main factors influencing the performance indicator, it is advisable to use the method of correlation-regression analysis. The indicator of enterprise activity efficiency was chosen as the resulting indicator (Y). As mentioned above, the key determinants that affect the efficiency of the investigated enterprise are costs, the cost of petroleum products transportation, the organizational structure of the volume in table 1 we will summarize the quantitative indicators of each of the identified key evaluation determinants and include their correlation-regression model of the studied enterprise business model economic efficiency.

Summary data for correlation-regression analysis of the influence of certain factors influence on the efficiency of the oil transportation company business model are given in table 2. Given the fact that all production and non-production areas of activity, including the pipeline system will consistently move to a digital display of activities, it is advisable to introduce methods of experiments in the study of improving the efficiency of the enterprise through optimization of input, significant parameters.

Table 1. Quantitative indicators for assessing the enterprise business model effectiveness and its key determinants.

Determinants	Quantitative indicators of evaluation
The enterprise business model efficiency	1. Profitability, %; 2. Economic efficiency (Y); 3. Profit, UAH million
Costs (business processes, digitalization)	1. The level of fixed assets depreciation (x_6) 2. Share of electricity costs in the total amount of costs, % (x_8); 3. The share of costs for diagnostics and defectoscopy of equipment in the total cost, % (x_9); 4. The level of intangible assets depreciation (x_{11}); 5. The level of actual implementation of R&D (x_{12})
Cost	1. Residual value of fixed assets (x_3); 2. Volumes of oil product transit transportation, thousand tons (x_5); 3. Average tariff for transit transportation of oil products by main pipelines, UAH for 1 ton net (x_1); 4. Average tariff for transportation of oil products to Ukraine's refineries by main pipelines, UAH for 1 ton net (x_2); 5. The share of intangible assets in the total assets of the enterprise, % (x_{10}); 6. Share of capital investments in equipment repairs (x_{13}); 7. Financing of capital investments, UAH million (x_{14})
The organizational structure	1. The average number of employees, persons (x_4); 2. The share of wage costs in total costs, % (x_7)

To build an adequate regression model of the studied enterprise business model economic efficiency, it is necessary to check the selected factors-indicators for the phenomenon of multicollinearity. Using the Data Analysis package in Excel, a correlation matrix was constructed, which demonstrates the strength of the relationship between the selected factor values and the performance indicator.

The calculation of the correlation matrix allows us to conclude that there is a close relationship between the performance indicator (Y) and the factor values, except for the pair (Y and x_4, x_5, x_9, x_{10}, x_{11}, x_{12}, x_{14}). These factors will be excluded from the study when constructing the regression equation. In addition to the study of the correlation between the performance indicator and the factors, it is advisable to determine the interdependence (multicollinearity) of the factor quantities-features among themselves. The phenomenon of multicollinearity occurs when the coefficients of pairwise correlation of trait factors are equal to 0.7 or more according to the Chedok scale, which indicates a close and very close relationship between certain trait factors. In this case, those factor features between which the phenomenon of multicollinearity was established should be excluded from the correlation-regression model. Those indicators of the relationship between which remains weak, are included in the economic-mathematical model for further construction of the regression equation.

Further conducted regression analysis between the effective rate and factor variables x_1, x_2, x_3, x_6, x_7, x_8, x_{13} using the Regression tool of Data analysis package in MS Excel.

To verify the formed regression model for adequacy, significance in general and regression coefficients, in addition to identifying correlation-regression relationships between performance and factor values, the coefficient of determination, multiple regression, standard error, and Student's t-test were calculated. The multiple regression coefficient $R = 0.998$ indicates a close relationship between the performance indicator and the factors. Regarding the value of the determination coefficient R^2 of the obtained correlation-regression model $R^2 = 0.996$, the dependence of the resulting feature by 99.6% is due to the selected factor values. The remaining 0.4% are due to other factors that affect the efficiency of the enterprise, but are not included in the regression model.

Table 2. Summary data for correlation and regression analysis of the factor values impact 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}, on the JSC "Ukrtransnafta" business model economic efficiency (Y).

Period	Y – efficiency of the enterprise	x_1 – average tariff for transit transportation of oil products by main pipelines, UAH for 1 ton net	x_2 – average tariff for transportation of oil products to refineries by main pipelines, UAH for 1 ton net	x_3 – residual value of fixed assets, thousand UAH	x_4 – the average number of employees, persons	x_5 – volumes of oil transportation transit, thousand tons	x_6 – level of fixed assets depreciation	x_7 – the share of wage costs in total costs, %	x_8 – the share of electricity costs in total costs, %	x_9 – the share of costs for diagnostics and defectoscopy of equipment in the total amount of costs, %;	x_{10} – the share of intangible assets in the total assets of the enterprise,%;	x_{11} – the level of depreciation of intangible assets	x_{12} – the level of actual implementation of R&D	x_{13} – the share of capital investment for equipment repairs	x_{14} – financing of capital investments, UAH million;
2010	1.10	69.61	15.88	1849.21	5726	14567.00	0.55	11.45	5.50	2.30	0.03	0.56	0.65	1.00	48.1
2011	1.18	78.16	11.15	1885.52	5113	16789.0	0.56	16.56	4.50	4.50	0.03	0.79	0.76	1.00	90.7
2012	1.24	79.48	11.51	1886.06	4865	15670.0	0.58	15.45	6.95	2.19	0.06	0.87	0.50	1.00	86.4
2013	1.04	88.45	12.37	1799.96	4414	12980.0	0.60	12.69	5.93	3.51	0.03	0.93	0.40	1.00	99.5
2014	1.66	128.22	12.53	1583.89	4447	13456.0	0.62	24.46	11.38	8.63	0.01	0.92	0.30	1.00	70.0
2015	2.19	210.65	42.58	15376.29	4312	13640.0	0.01	27.71	12.71	0.00	0.01	0.90	0.51	0.68	233.3
2016	1.97	241.38	74.89	14977.46	4535	13822.2	0.06	27.45	10.90	0.00	0.01	0.91	0.57	0.61	285.4
2017	1.95	256.40	78.89	14750.85	4529	13937.1	0.01	25.97	9.75	0.36	0.01	0.84	0.24	0.35	143.8
2018	1.60	277.78	76.62	13947.10	3917	13334.8	0.07	26.59	7.98	4.89	0.68	0.24	0.04	0.56	598.8
2019	1.63	249.67	82.31	8428.94	3666	13126.6	0.04	32.61	11.33	2.57	0.72	0.40	9.33	0.59	1809.9

According to the results of the multiple regression coefficient calculation and determination, the relationship between the resulting indicator and the factors-features is quite natural. Analysis of the constructed model variance shows that the factors included in the regression model are significant, as the residual sum of squares, as it is indicated in table 2, is less than 1% of the total sum of squares, i.e. 99% of the resulting indicator variation is due to the studied factors and only 1% variation of the resulting indicator is caused by the action of random variables. Since F (71.37)>Significance of F (0.013) and F (0.013)<0.05, it can be stated that the results of the regression model correspond to the empirical data and the number of independent variables included in the regression model is sufficient to describe the dependent variable. The values of F and the significance of F indicate a sufficient level of the evaluation results reliability and the significance of the developed equation.

Table 3 shows the results of regression coefficients calculations and statistics of their significance. Based on the calculations, we can conclude that all factors included in the regression model are reliable, because $t_K > t_{crit}$, where the critical value of t_{crit} at the set level of significance is 0.15 and was determined using the STEERING function of the MS Excel. The data in table 3 allow obtaining the following regression model:

$$Y=0.8843-0.0047x_1-0.0013x_2+0.000000078x_3+0.9470x_6+0.0394x_7+0.0519x_8-0.8867x_{13}$$

Thus, based on the results of the correlation-regression analysis, it can be concluded that the greatest influence on the oil transportation company business model efficiency is exerted by the following factors: average tariff on the transit transportation of petroleum products by main pipelines per 1 ton net; average tariff on the transportation of petroleum products to refineries by main pipelines, UAH for 1 ton net; residual value of fixed assets, thousand UAH; level of fixed assets depreciation; proportion of wage costs in the total costs, %; proportion of electricity costs in the total amount of costs, %; proportion of capital investment in equipment repair.

Therefore, among the obtained results, the level of fixed assets depreciation has the greatest positive influence on the resulting indicator. The growth of this indicator by 1 unit will increase the resulting figure by 0.94 units. The biggest negative impact on the industrial enterprise business model efficiency has the proportion of capital investment in equipment repair, namely the growth of this factor by 1 will reduce the efficiency of the business model by 0.88 units. We calculated the elasticity for each factor by the formula:

$$E_j = b_j \frac{\overline{x_J}}{\overline{y_J}}, \tag{4}$$

where b_j – the corresponding coefficient of the regression equation; $\overline{x_J}$ – the arithmetic mean of the variable x_j; $\overline{y_J}$ – the arithmetic mean of the endogenous variable y_j.

In practice, it is often necessary to compare the separate effect on the dependent variable of different explanatory variables, when the latter ones are expressed in different units. In this case, the coefficients of elasticity are used. The coefficient of elasticity E_j shows how many percent Y will change on the average if x_j is increased by 1%. Since further calculations require additional statistical indicators for the resulting feature and factors, we will use the Statistics package of MS Excel, the obtained intermediate calculations for further analysis are summarized in table 3.

Table 3. Results of elasticity calculation for the factors of the regression model

Factor	The average value of the factor for the study period ($\overline{x_j}$)	Regression coefficient (b_j)	Coefficients of elasticity (E_j)
x_1 – average tariff for transit transportation of oil products by main pipelines, UAH for 1 ton net	167.98	-0.0047326	-0.51
x_2 – average tariff for transportation of oil products to refineries by main pipelines, UAH for 1 ton net	41.87	-0.0013603	-0.04
x_3 – fixed assets residual value, thousand UAH	7648527.70	0.0000001	0.38
x_6 – the level of fixed assets depreciation	0.31	0.9470437	0.19
x_7 – the share of wage costs in the total costs, %	22.10	0.0394198	0.56
x_8 – share of electricity costs in the total amount of costs ,%	8.69	0.0519409	0.29
x_{13} – share of capital investments in equipment repairs	0.78	-0.8867077	-0.44

Thus, based on the results of the elasticity coefficient factors calculation of the regression model, we can conclude that the growth of the average tariff on the transit transportation of petroleum products by main pipelines, UAH per 1 ton of net, by 1% will reduce the economic efficiency of the business model by 0.51%, and an increase in the average tariff on the transportation of petroleum products to refineries by main pipelines, UAH for 1 ton net, by 1% will reduce the economic efficiency by 0.04%. In this case, the management of the investigated enterprise should take measures to reduce the tariffs, especially their cost. Another factor that reduces the economic efficiency of the business model of the studied enterprise is the proportion of capital investment in equipment repair, its increase by 1% will reduce the resulting indicator by 0.44%. Of the eight factors studied, four have a positive effect on increasing the economic efficiency of the oil transportation company business model. In particular, an increase in the residual value of fixed assets by 1% will lead to an increase in the resulting indicator by 0.38%, which shows the need to update the fixed assets base. An increase in the level of fixed assets depreciation by 1% will increase the economic efficiency by 0.19%. An increase in the proportion of wage costs by 1% will increase the resulting indicator by 0.56%, and an increase in the proportion of electricity costs in total costs by 1% will increase the economic efficiency of the business model by 0.29%. Thus, the increase in the studied costs leads to an increase in the efficiency of the business model to a greater extent than the increase in the tariff on transportation, as it is proved by the developed correlation-regression model.

To determine the factors that have the greatest reserve for improving the resulting indicator, taking into account the degrees of the factors-indicators variation, we use the calculation of standardized regression β coefficients by the formula:

$$\beta_j = a_j \frac{\sigma_{xj}}{\sigma_y} \tag{5}$$

where σ_{xj} – quadratic mean deviation of the i-th factor-indicator; σ_y – quadratic mean deviation of the resulting indicator.

The delta coefficient indicates what part of the contribution the studied factor makes in the total influence of all selected factors. It is determined by the formula:

$$\Delta_j = \beta_j \frac{R_j}{R^2} \tag{6}$$

where R_j – matching correlation coefficient; R^2 – coefficient of multiple determination.

It should be noted that the increase in the number of factors-indicators, which are included in the multifactor correlation-regression model, allows establishing the additional resources of the studied resulting indicator.

The results of the obtained calculations of Δ and β coefficients are summarized in table 4.

Table 4. Δ_j and β_j coefficients of the developed correlation-regression model.

Factor-indicator	The standard deviation	Regression coefficient	β_j coefficients	Paired correlation coefficient	Δ_j coefficients
Y – the enterprise business model efficiency	0.4030771	0.8842506	-	-	-
x_1 – average tariff for transit transportation of oil products by main pipelines, UAH for 1 ton net	86.4343677	-0.0047326	-1.0148456	0.7866235	-0.8014970
x_2 – average tariff for transportation of oil products to refineries by main pipelines, UAH for 1 ton net	32.6194806	-0.0013603	-0.0005133	0.6624873	-0.0003414
x_3 – fixed assets residual value, thousand UAH	6456617.74	0.0000001	0.0154706	0.8575565	0.0133200
x_6 – the level of fixed assets depreciation	0.2879776	0.9470437	0.0000000	-0.8124769	-0.00000003
x_7 – the share of wage costs in the total costs, %	7.3646677	0.0394198	1.0081115	0.8389036	0.8490938
x_8 – share of electricity costs in the total amount of costs, %	2.8927203	0.0519409	0.0204015	0.8894378	0.0182185
x_{13} – the share of capital investment in equipment repairs, %	0.2479886	-0.8867077	-0.0760161	-0.7471013	0.0570191

Based on the obtained results of calculating Δ_j and β_j coefficients, it can be concluded that the vectors of the calculated β_j and coefficients are proved by the results of elasticity coefficients calculation in table 4. However, according to the results of the latter calculations, the greatest negative impact on the studied enterprise business model efficiency by the standard deviation is the factor of the average tariff on the transit transportation of petroleum products by main pipelines, UAH for 1 ton net. The greatest positive impact on the resulting indicator by the standard deviation is exerted by the factor the proportion of wage costs in the total amount of costs, %. Based on the results of Δ_j calculation and coefficients, it can be concluded that this statistics are confirmed by the results of β_j calculation and characteristics.

Recently, along with traditional methods of analyzing the socio-economic indicators, the use of neural networks, which belong to artificial intelligence systems, is becoming more widespread. After all, their scope is extremely large: forecasting changes in the stock exchange, making credit plans, making decisions when landing a damaged aircraft, approximating functions, solving optimization problems, managing complex processes, forecasting, etc. That is why the use of neural networks is relevant for the analysis of factors influencing the efficiency of oil transportation companies, along with traditional methods. In a series of recent works [3; 1; 14; 13], the authors have demonstrated the possibility of using the theory of complex systems and a set of developed analysis tools to calculate the corresponding measures of system complexity. These complexity measures make it possible to differentiate the systems according to the degree of their functionality, to identify and prevent critical and crisis phenomena.

Let us move on to the development of the neural network. For this purpose, use the module Neural Networks of the Statistica software package.

Table 5. Detailed results of neural network models.

Summary of active networks (Spreadsheet1_(Recovered))											
Index	Net. name	Training perf.	Test perf.	Validatio n perf.	Training error	Test error	Validatio n error	Training algorithm	Error function	Hidden activation	Output activation
1	MLP 6-10-1	0,996503	0,966398	0,999999	0,001489	,016893	0,006213	BFGS 8	SOS	Logistic	Identity
2	MLP 6-7-1	0,930660	0,970936	0,999934	0,014347	,007955	0,012498	BFGS 8	SOS	Logistic	Tanh
3	MLP 6-11-1	0,954370	0,973305	0,999893	0,010534	,012491	0,005659	BFGS 7	SOS	Logistic	Exponential
4	MLP 6-5-1	0,955839	0,967683	0,999891	0,011195	,010368	0,011513	BFGS 5	SOS	Identity	Tanh
5	MLP 6-7-1	0,958716	0,967189	0,999960	0,011143	,009940	0,011811	BFGS 5	SOS	Identity	Tanh

In the Neural Networks window, we set the following parameters:

- type of problem – “Regression”;
- as input parameters, according to the correlation analysis, we use: the average tariff on the transit transportation of petroleum products by the main pipelines for 1 t. net; average tariff on the transportation of petroleum products to refineries by main pipelines, UAH for 1 ton net; residual value of the fixed assets, thousand UAH; level of fixed assets depreciation; proportion of wage costs in the total amount of costs, %; proportion of electricity costs in the total amount of costs, %; proportion of capital investment in equipment repairs;
- the economic efficiency is taken as the input parameter (the resulting feature).

Then, with the help of the Solution Wizard, we go to the window for building Neural Networks.

In the Solution Wizard window, we set the parameters for creating neural networks:

- test 100 networks and save only the top 5;
- the types of neural network architectures, used for modeling, are multilayer perceptron, generalized regression network and linear network;
- run the analysis. We analyze the results of neural network modeling aimed at maintaining regression based on the detailed model results (table 5).

Table 5 shows that the Master created and proposed 5 multilayer perceptrons of neural network architectures. We choose the best model using the results of Descriptive Statistics, which are in table 6.

Table 6. Correlation coefficients.

	Correlation coefficients (Spreadsheet1_(Recovered))		
	Y - efficiency of the enterprises business model Train	Y - efficiency of the enterprises business model Test	Y - efficiency of the enterprises business model Validation
1.MLP 6-10-1	0,996503	0,966398	0,999999
2.MLP 6-7-1	0,930660	0,970936	0,999934
3.MLP 6-11-1	0,954370	0,973305	0,999893
4.MLP 6-5-1	0,955839	0,967683	0,999891
5.MLP 6-7-1	0,958716	0,967189	0,999960

According to the correlation indicators, shown in table 6, the best model is the first model – a multilayer perceptron, with the architecture MLP 6-10-1. The MLP 6-10-1 architecture identifies a model with 6 input variables, one output variable and one hidden layer containing 10 hidden neurons. Having investigated and analyzed the results obtained in table 6, we can make conclude that the constructed models work evenly as the correlation coefficients of the test sample are approximately at the same level. The Statistica Neural Networks program provides the ability to analyze the sensitivity of the network to input variables. This procedure allows us to make conclusions about the relative importance of the input variables for a particular neural network and, if necessary, to remove the inputs with low sensitivity. Sensitivity analysis can be used for purely informational purposes, or to remove entries.

Sensitivity analysis brings some clarity to the usefulness of the certain variables. It allows identifying the key variables, without which analysis is impossible, and identifying those that can be safely excluded from consideration.

Therefore, let us analyze the results of the input variable sensitivity to the output variable and see which variables are included in our model (table 7), and the standard prediction of the neural network, i.e. the predicted values for the resulting variables in the proposed models (table 8).

Thus, we can see that the chosen model has minor errors, and therefore it can be called reliable. The rest predicted results of the neural network models calculations are presented in tables 9-11.

Table 7. Sensitivity analysis.

Networks	Sensitivity analysis (Spreadsheet1_(Recovered)) Samples: Train, Test, Validation x7 - the share of electricity costs in total costs,%;	x6 - the level of depreciation of fixed assets;	x7 – the share of salary costs in total costs,%;	x2 – average tariff for transportation of oil products to refineries by main pipelines, UAH for 1 t. netto;	x13 – the share of capital investment in equipment repairs;	x1 – average tariff for transit transportation of oil products by main pipelines, UAH for 1 t. netto, грн. за 1 т нетто;
1.MLP 6-10-1	3,544210	2,504670	1,218176	1,826351	1,485822	1,084455
2.MLP 6-7-1	2,299395	1,350893	1,478483	1,041965	1,092716	0,991620
3.MLP 6-11-1	1,943351	1,292521	1,055444	1,128046	1,071157	0,932827
4.MLP 6-5-1	2,620613	1,509926	1,466700	1,099757	1,161099	1,028567
5.MLP 6-7-1	2,729713	1,548415	1,502538	1,080967	1,223184	1,023498
Average	2,627456	1,641285	1,344268	1,235417	1,206796	1,012194

Table 8. Predicted values of the oil transportation company efficiency based on the developed multi-architectural neural networks.

Case name	Predictions spreadsheet for Y - efficiency of the enterprises business model (Spreadsheet1_(Recovered)) Samples: Train, Test, Validation Sample	Y - efficiency of the enterprises business model Target	Y - efficiency of the enterprises business model - Output 1. MLP 6-10-1	Y - efficiency of the enterprises business model - Output 2. MLP 6-7-1	Y - efficiency of the enterprises business model - Output 3. MLP 6-11-1	Y - efficiency of the enterprises business model - Output 4. MLP 6-5-1	Y - efficiency of the enterprises business model - Output 5. MLP 6-7-1	Y - efficiency of the enterprises business model - Output Ensemble
1	Train	1,100000	1,082962	1,056674	1,158425	1,061849	1,072921	1,086566
2	Test	1,180000	1,076189	1,087971	1,159720	1,085584	1,093730	1,100639
3	Validation	1,240000	1,221474	1,226861	1,180044	1,234785	1,243894	1,221411
4	Test	1,040000	1,105908	1,098549	1,161789	1,108089	1,117473	1,118362
5	Train	1,660000	1,666791	1,661751	1,620909	1,648881	1,650108	1,649688
6	Train	2,190000	2,083504	1,861664	1,937052	1,896471	1,894925	1,934723
7	Validation	1,970000	1,832867	1,777675	1,840524	1,786291	1,781861	1,803844
8	Validation	1,950000	1,815363	1,755506	1,833396	1,762113	1,761710	1,785618
9	Train	1,600000	1,584646	1,671356	1,723884	1,642662	1,635257	1,651561
10	Test	1,630000	1,923665	1,819298	1,874339	1,850579	1,844931	1,862562

Table 9. Absolute results of the predicted values of the oil transportation enterprise activity efficiency based on the various developed architectural neural networks.

Case name	Predictions spreadsheet for Y - efficiency of the enterprises business model (Spreadsheet1_(Recovered)) Samples: Train, Test, Validation Sample	Y - efficiency of the enterprises business model - Abs. Res. 1. MLP 6-10-1	Y - efficiency of the enterprises business model - Abs. Res. 2. MLP 6-7-1	Y - efficiency of the enterprises business model - Abs. Res. 3. MLP 6-11-1	Y - efficiency of the enterprises business model - Abs. Res. 4. MLP 6-5-1	Y - efficiency of the enterprises business model - Abs. Res. 5. MLP 6-7-1	Y - efficiency of the enterprises business model - Abs. Res. Ensemble
1	Train	0,017038	0,043326	0,058425	0,038151	0,027079	0,013434
2	Test	0,103811	0,092029	0,020280	0,094416	0,086270	0,079361
3	Validation	0,018526	0,013139	0,059956	0,005215	0,003894	0,018589
4	Test	0,065908	0,058549	0,121789	0,068089	0,077473	0,078362
5	Train	0,006791	0,001751	0,039091	0,011119	0,009892	0,010312
6	Train	0,106496	0,328336	0,252948	0,293529	0,295075	0,255277
7	Validation	0,137133	0,192325	0,129476	0,183709	0,188139	0,166156
8	Validation	0,134637	0,194494	0,116604	0,187887	0,188290	0,164382
9	Train	0,015354	0,071356	0,123884	0,042662	0,035257	0,051561
10	Test	0,293665	0,189298	0,244339	0,220579	0,214931	0,232562

Tables 9-11 show that the model MLP 6-10-1 does have small residues, which indicates a fairly high accuracy of the forecast. The graphs of actual and predicted values of the business model efficiency of the oil transportation enterprise in the period from 2010 to 2019 are presented in figure 8.

Table 10. Squared results of the predicted values of the oil transportation enterprise activity efficiency based on the various developed architectural neural networks.

Case name	Sample	Y - efficiency of the enterprises business model - Squared Res. 1. MLP 6-10-1	Y - efficiency of the enterprises business model - Squared Res. 2. MLP 6-7-1	Y - efficiency of the enterprises business model - Squared Res. 3. MLP 6-11-1	Y - efficiency of the enterprises business model - Squared Res. 4. MLP 6-5-1	Y - efficiency of the enterprises business model - Squared Res. 5. MLP 6-7-1	Y - efficiency of the enterprises business model - Squared Res. Ensemble
	Predictions spreadsheet for Y - efficiency of the enterprises business model (Spreadsheet1_(Recovered)) Samples: Train, Test, Validation						
1	Train	0,000290	0,001877	0,003414	0,001455	0,000733	0,000180
2	Test	0,010777	0,008469	0,000411	0,008914	0,007442	0,006298
3	Validation	0,000343	0,000173	0,003595	0,000027	0,000015	0,000346
4	Test	0,004344	0,003428	0,014833	0,004636	0,006002	0,006141
5	Train	0,000046	0,000003	0,001528	0,000124	0,000098	0,000106
6	Train	0,011341	0,107805	0,063983	0,086159	0,087069	0,065166
7	Validation	0,018805	0,036989	0,016764	0,033749	0,035396	0,027608
8	Validation	0,018127	0,037828	0,013596	0,035302	0,035453	0,027022
9	Train	0,000236	0,005092	0,015347	0,001820	0,001243	0,002659
10	Test	0,086239	0,035834	0,059702	0,048655	0,046195	0,054085

Table 11. Studied results of the predicted values of the oil transportation enterprise activity efficiency based on the various developed architectural neural networks.

Case name	Sample	Y - efficiency of the enterprises business model - Std. Res. 1. MLP 6-10-1	Y - efficiency of the enterprises business model - Std. Res. 2. MLP 6-7-1	Y - efficiency of the enterprises business model - Std. Res. 3. MLP 6-11-1	Y - efficiency of the enterprises business model - Std. Res. 4. MLP 6-5-1	Y - efficiency of the enterprises business model - Std. Res. 5. MLP 6-7-1	Y - efficiency of the enterprises business model - Std. Res. Ensemble
	Predictions spreadsheet for Y - efficiency of the enterprises business model (Spreadsheet1_(Recovered)) Samples: Train, Test, Validation						
1	Train	0,44151	0,36172	-0,56925	0,36058	0,25653	0,13611
2	Test	2,69010	0,76832	0,19759	0,89236	0,81726	0,80407
3	Validation	0,48007	0,10970	0,58417	0,04929	-0,03688	0,18834
4	Test	-1,70790	-0,48881	-1,18662	-0,64353	-0,73392	-0,79394
5	Train	-0,17598	-0,01462	0,38088	0,10509	0,09371	0,10448
6	Train	2,75967	2,74118	2,46454	2,77423	2,79533	2,58640
7	Validation	3,55357	1,60567	1,26151	1,73629	1,78230	1,68346
8	Validation	3,48889	1,62377	1,13610	1,77578	1,78373	1,66548
9	Train	0,39789	-0,59573	-1,20703	-0,40321	-0,33400	-0,52240
10	Test	-7,60985	-1,58039	-2,38066	-2,08476	-2,03610	-2,35627

4 Results

Based on the investigation of JSC "Ukrtransnafta" business model economic efficiency it is possible to draw conclusions that the accompanying process of digitalization of production for the given enterprise is characterized by a number of problems which arise both at initial stages and in adaptation process. In addition, it has been proposed to use a synergetic coefficient of digitalization efficiency, which takes into account the sequence of the enterprise business processes and partial indicators of digitalization efficiency of each individual business process. However, it cannot be calculated for the enterprise under study, as the processes at JSC "Ukrtransnafta" production digitalization are currently underway. Taking into account these circumstances, a system of the previously determined key determinants factors-features of the impact on the company business model economic efficiency of the oil transportation enterprise and the resulting features has been developed. According to the analysis of the impact levels of the selected factors, the main ones have been determined, which have been used to develop the neural network models of the enterprise's business model economic efficiency. The main factors influencing the business model efficiency of the studied enterprise are the average tariff on the transit transportation of petroleum products by main pipelines, UAH for 1 t. net; average tariff on the transportation of petroleum products to refineries by main pipelines, UAH for 1 t. net; level of fixed assets depreciation; proportion of salary costs in total costs, %; proportion of electricity costs in total costs, %; proportion of capital investment in equipment repair. The regression neural network analysis has proved six factors that can be used to predict the economy of the oil transportation company business model economic efficiency. The performed analysis of the neural networks sensitivity, which indicates the direction and strength of the influence of factors, will allow the surveyed enterprise to manage the economic efficiency, and the state, in its turn, to find directions for industry development.

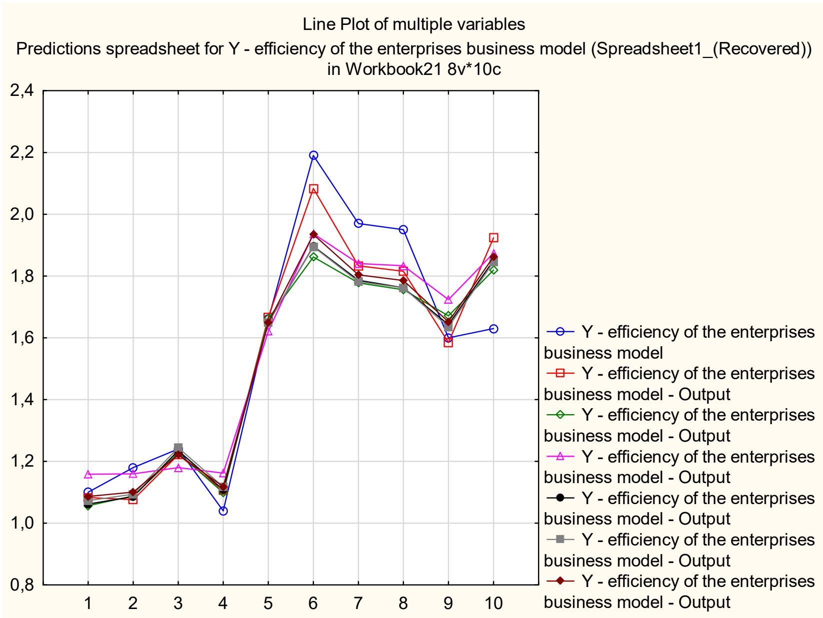


Fig. 4. Graphs of actual and predicted values of the business model of the industrial enterprise efficiency for different neural network models.

5 Conclusion

Thus, comparing the results of this research work and the previous work [4] we can draw unambiguous conclusions that according to [4], the financial efficiency of the studied enterprise is at the average level, it increases at the end of the analyzed period mainly due to the profitability component, which is usually a positive trend. Based on the results of regression neural network analysis, the key factors influencing the economic efficiency have been identified and ranked, and the optimal neural network model for forecasting the resulting indicator has been developed, which can be used for predictions and will allow the company's management to rationally manage the economic efficiency at relatively low costs.

References

1. Belinskyi, A., Soloviev, V., Semerikov, S., Solovieva, V.: Detecting stock crashes using Levy distribution. CEUR Workshop Proceedings **2422**, 420–433 (2019)
2. Bezat-Jarzębowska, A., Rembisz, W.: Efficiency-focused Economic Modeling of Competitiveness in the Agri-Food Sector. Proc. - Soc. and Behav. Sci. **81**, 359–365 (2013). doi:10.1515/foli-2016-0005
3. Derbentsev, V., Semerikov, S., Serdyuk, O., Solovieva, V., Soloviev, V.: Recurrence based entropies for sustainability indices. E3S Web Conf. **166**, 13031 (2020). doi:10.1051/e3sconf/202016613031
4. Havrylenko, M., Shiyko, V., Horal, L., Khvostina, I., Yashcheritsyna, N: Economic and mathematical modeling of industrial enterprise business model financial efficiency estimation. E3S Web Conf. **166**, 13025 (2020). doi:10.1051/e3sconf/202016613025
5. Herts, A., Tsidylo, I., Herts, N., Barna, L., Mazur, S.-I.: PhotosynQ – cloud platform powered by IoT devices. E3S Web of Conferences **166**, 05001 (2020). doi:10.1051/e3sconf/202016605001
6. Khvostina, I., Havadzyn, N., Horal, L., Yurchenko, N.: Emergent properties manifestation in the risk assessment of oil and gas companies. CEUR Workshop Proceedings **2422**, 157–168 (2019)
7. Kiv, A.E., Soloviev, V.N., Semerikov, S.O.: CTE 2018 – How cloud technologies continues to transform education. CEUR Workshop Proceedings **2433**, 1–19 (2019)
8. Kozmenko, O., Oliynyk V.: Statistical model of risk assessment of insurance company's functioning. Inv. Man. and Fin. Innov. **12**(2), 189–194 (2015)
9. Marchi, B., Zanoni, S., Jaber, M.Y.: Economic production quantity model with learning in production, quality, reliability and energy efficiency. Eng. Comp. & Industr. Eng. **129**, 502–511 (2019). doi:10.1016/j.cie.2019.02.009
10. Modlo, Ye.O., Semerikov, S.O., Shajda, R.P., Tolmachev, S.T., Markova, O.M., Nechypurenko, P.P., Selivanova, T.V.: Methods of using mobile Internet devices in the formation of the general professional component of bachelor in electromechanics competency in modeling of technical objects. CEUR Workshop Proceedings **2643**, 500–534 (2020)
11. Oliinyk, V., Kozmenko S.: Forecasting and management of gross domestic product. J. of Int. Stud. **12**(4), 214–228 (2019). doi:10.14254/2071-8330.2019/12-4/14
12. Semerikov, S.O., Teplytskyi, I.O., Yechkalo, Yu.V., Kiv, A.E.: Computer Simulation of Neural Networks Using Spreadsheets: The Dawn of the Age of Camelot. CEUR Workshop Proceedings **2257**, 122–147 (2018)
13. Soloviev, V., Bielinskyi, A., Solovieva, V.: Entropy Analysis of Crisis Phenomena for DJIA Index. CEUR Workshop Proceedings **2393**, 434–449 (2019)
14. Soloviev, V.N., Belinskiy, A.: Complex Systems Theory and Crashes of Cryptocurrency Market. Communications in Computer and Information Science **1007**, 276–297 (2019)
15. Teplická, K.: Comparison of Methods for Pricing of the Product and its Impact on Economic Efficiency of Enterprise. Proc. Econ. and Fin. **34**, 149–155 (2015). doi:10.1016/S2212-5671(15)01613-5
16. Zakharov, A., Jämsä-Jounela, S.-L.: Iterative optimization of the economic efficiency of an industrial process within the validity area of the static plant model and its application to a Pulp Mill. Computers & Chemical Engineering **35**(2), 245–254 (2011). doi:10.1016/j.compchemeng.2010.10.010

Modelling of cryptocurrency market using fractal and entropy analysis in COVID-19

Hanna Danylchuk[1][0000-0002-9909-2165], Liubov Kibalnyk[1][0000-0001-7659-5627], Oksana Kovtun[2][0000-0002-0159-730X], Arnold Kiv[3][0000-0002-0991-2343.], Oleg Pursky[4][0000-0002-1230-0305] and Galina Berezhna[2][0000-0002-0596-8534]

[1] The Bohdan Khmelnytsky National University of Cherkasy, 81 Shevchenka Blvd., Cherkasy, 18031, Ukraine
abdanilchuk@gmail.com, liubovkibalnyk@gmail.com
[2] University of Educational Management, 52A Sichovyh Striltsiv Str., Kyiv, 04053, Ukraine
kovtun.oa71@gmail.com, galina52@yahoo.com
[3] Ben-Gurion University of the Negev, P.O.B. 653, Beer Sheva, 8410501, Israel
kiv.arnold20@gmail.com
[4] Kyiv National University of Trade and Economics, Kyiv, 19 Kyoto Str., 02156, Ukraine
pursky_o@ukr.net

Abstract. In this article, we present the results of simulation for cryptocurrency market based on fractal and entropy analysis using six cryptocurrencies in the first 20 of the capitalization rating. The application of the selected research methods is based on an analysis of existing methodologies and tools of economic and mathematical modeling of financial markets. It has been shown that individual methods are not relevant because they do not provide an adequate assessment of the given market, so an integrated approach is the most appropriate. Daily values of cryptocurrency pairs from August 2016 to August 2020 selected by the monitoring and modelling database. The application of fractal analysis led to the conclusion that the time series of selected cryptocurrencies were persistent. And the use of the window procedure for calculating the local Hurst coefficient allowed to detail and isolate the persistant and antipersistant gaps. Interdisciplinary methods, namely Tsallis entropy and wavelet entropy, are proposed to complement the results. The results of the research show that Tsallis entropy reveals special (crisis) conditions in the cryptocurrency market, despite the nature of the crises' origin. Wavelet entropy is a warning indicator of crisis phenomena. It provides additional information on a small scale.

Keywords: cryptocurrency market, fractal analysis, wavelet entropy, Tsallis entropy, crisis, COVID-19.

1 Introduction

In the context of digitalization and virtualization of the financial sphere, the problem of forecasting the cryptocurrency market dynamically developing in recent years, is of particular importance. The improvement of tools for modelling and forecasting of the cryptocurrency market and risk management are necessary in the context of public distrust of cryptocurrency as an innovative financial asset, and the presence of a number of risks inherent in the cryptocurrency transactions (asset security, hedging complexity, high price volatility, lack of guarantees, legal barriers, the limited use as a means of payment).

The COVID-19 pandemic, in which the governments of the world's major economies have been forced to adopt unprecedented security measures, has highlighted this problem. As a result, social, economic, psychological and behavioural crises can occur in the world economy and social spheres [18; 33]. Cash flows have had to be diverted to the treatment of patients, the elimination of the effects of the pandemic, and the development and testing of influenza drugs instead of investing in financial and stock markets. In the research [17] the author draws a parallel between natural disasters, terrorist events, crises, epidemics, pandemics and proves that the effects of the pandemic are particularly pronounced in the financial and banking sectors, in the stock markets and in insurance. Stock market reactions at COVID-19 are the subject of the research [4] in which the researcher identifies the dependency and volatility of stock markets on flu morbidity and mortality. Stock markets have been shown to respond quickly to COVID-19, but the intensity of the reaction has depended more on the number of diseases than on the number of deaths. According to the author, the fluctuating cycle of the stock market's decline and increase is 40-60 days since the onset of the pandemic in the country.

The impact of COVID-19 on stock markets, which is developing, is even more threatening. This leads to instability in the financial, fiscal, and other spheres. Researching the 26 stock markets that are developing, the author in publication [36] has empirically proven that at the onset of the pandemic, with limited economic activity, mobility of the population, cash transfers, trade, commercial service activities, there was a free fall of indices in all the markets studied. Using a Driscoll-Kraay estimate, the researcher concluded that coronavirus had a negative and statistically significant impact on stock markets until mid-April, and then in countries where compensation mechanisms were used, this effect gradually diminished.

Authors of [7] point out that the formation and development of the cryptocurrency market are associated with Bitcoin release and blockchain technology that have some spheres of application; and cryptocurrencies are innovative financial asset that attracts investors.

According to the authors of [22], in assessing stability during a pandemic, there is a higher level of risk to the cryptocurrency market than the stock market. That is, digital currencies exhibit instability, increased irregularity and volatility, indicating a decline in their investment attractiveness during this period.

An important segment of research is the study of the significant impact of the cryptocurrency market in a context of limited communication and commercial activity, where Internet commerce and services are becoming increasingly important. The current development of the economy is characterized by the rapid growth of the pace of informatization of economic processes, the expansion of the scale of e-business, and, in particular, of retail e-trade on a large scale [32]. E-commerce has grown rapidly, spreading comprehensively and offering an increasingly a diverse assortment of goods and services, e-commerce becomes an instrument for the integration of individuals, enterprises, industries, state institutions, and states into a united community, within which the interaction of partners is effectively and unhindered by means of information and telecommunication technologies [6]. The basis of e-commerce is the new information technologies for commercial operations and the management of production processes. E-commerce, compared to traditional business, has substantial advantages. In particular, the use of new electronic communication channels and new cryptocurrencies significantly reduce costs related to organization and support e-business, and the possibilities of e-commerce allow re-designing business strategy at any moment [30].

Thus, in the context of COVID-19, there is a demand among investors and businessmen for up-to-date information on the state and dynamics of the cryptocurrency market, as well as the possibility of forecasting it. The use of modelling tools for research and forecasting of the cryptocurrency markets using an integrated approach and interdisciplinary methods is therefore a topical scientific and applied challenge.

National and international researches pay great attention to this problem. Different approaches and methods are used for modelling and forecasting the cryptocurrency market and forming the investment portfolio. As this market becomes more and more difficult, the potential investors need tools allowing them to form a highly profitable investment portfolio that may include such an asset as cryptocurrency.

To calculate the cryptocurrency market efficiency, paper [3] offers to use Factor Augmented VAR (TVP-FAVAR)-model considering the impact of a large number of variable factors on a dependent variable and allowing to study the dynamics of more than 45 cryptocurrencies. This approach helps to conclude that the periods of high/low uncertainty in the market correspond to strong/weak link. The authors explain the trend by the increased degree of the market uncertainty associated with the process of cryptocurrency price fluctuations. In this situation, they propose to form a dual investment portfolio, the structure of which can be varied by the dynamic hedge ratio and the dynamic weights of the portfolio. Using the ARFIMA-FIGARCH model with two distributions and the modified logarithmic periodogram method, authors [27] studied the stability of eight biggest cryptocurrency markets and made a conclusion that they were unstable, volatile, had the limited trading horizons and time gaps, complicating the process of attracting investors.

Paper [10] proposes to assess the dynamics of the cryptocurrency profitability and stability by the multivariate stochastic model, which allows to calculate the average currency volatility. Its practical application shows the significant impact of such factors as the volatility of the cryptocurrency market on the one hand, and the rapid growth in demand for it on the other.

In paper [16], researchers use a Lotka-Volterra model with variable intervals to model the number of transactions for Bitcoin, Litecoin and Ripple using two- and three-dimensional models that allow them to get high-precision forecast for Bitcoin and satisfactory level for Litecoin and Ripple.

The investigation of cryptocurrency time series using econometric models of stochastic volatility is shown in papers [9; 29]. Alternative one-dimensional dynamic linear and multivariate vector autoregressive models are compared in paper [9]. This approach improves the accuracy of the forecast.

The paper of scientists [8] shows that the use of standard GARCH models can make incorrect VAR and ES forecasts, and thus, lead to inefficient risk management and optimization of the investment portfolio. Therefore, the authors propose to use the method of model building (model complex) of VAR and ES recheck based on a confidence model (MCS) to minimize risks and financial losses.

The traditional method of the cryptocurrency market analysis is technical analysis, but in the conditions of high market volatility, it is reasonable to combine the methods of technical analysis with the methods of economic and mathematical modelling. Thus, authors [1] propose a non-parametric model based on technical analysis as an alternative method of assessment and forecasting of the cryptocurrency market. This approach calculates the forecast values of Bitcoin's profitability through a neural network and indicates the speculative nature of the market.

Traditional methods and modelling do not allow to make accurate forecasts and calculations for the development of the cryptocurrency market being a nonlinear complex economic system and to detect the occurrence of crisis phenomena.

Therefore, it is necessary to use the methods being traditionally inherent in other fields of science. Econophysics has rather powerful methodological apparatus for modelling complex socio-economic systems in modern economic science.

Thus, paper [2] points out that it is necessary to apply comprehensive approach for the forecast considering the nonlinear dynamics and the inherent chaos and fractality of the digital currencies. The authors propose the hybrid model for the forecast based on the neural network of long-term memory (LSTM) and empirical wavelet decomposition (EWT), along with the cuckoo search algorithm (CS) for digital currency time series to obtain more accurate forecast values.

Authors [11] use monofractal analysis to investigate the price volatility in the cryptocurrency market and multifractal fluctuation analysis to test the model for stability.

A strong impulse effect in the Bitcoin and Ethereum markets and a reversal effect for Ripple and EOS at high fluctuations were found as a result of the model implementation. The application of this model will help to form effective alternative strategies for the allocation of assets in the investment portfolio.

Article [26] contains a multifractal analysis of the cryptocurrency market's performance level before and after the COVID-19 pandemic. The authors' calculations show that, overall, the cryptocurrency market became more efficient during the pandemic. The research [25] analyzed asymmetrical multifractal oscillations without a trend for Bitcoin and Ethereum, which revealed structural breaks (shifting), long memory, and asymmetric multifractures. Bitcoin has been shown to exhibit low performance on the general upward and downward trends, with the gap between the latter two being small when the temporal scale is reduced and rather, growing as it grows.

Entropy and wavelet analysis are an effective tool for modelling and research of the cryptocurrency market. The authors of article [20] propose to use a multifractional decomposition approach on a time scale for analysis of prices and trading volumes at different sampling frequencies in high-frequency cryptocurrency markets. The analysis of a large amount of data with the combination of statistical testing showed a heterogeneous multifractal dynamic at certain intervals, so the application of this methodology allowed the authors to obtain a more precise definition of the singularity spectrum. This will allow potential investors to develop an efficient and profitable strategy for high-frequency cryptocurrency trading.

With the help of Renyi entropy, the authors [21] analyze the processes of information exchange and information chaos in the stock markets, cryptocurrency markets, oil and gas markets, precious metals markets before and after a pandemic in order to identify investors' expectations. Empirical research has concluded that investment portfolios containing Bitcoin and precious metals (gold, silver), Bitcoin and Brent have high risk and need to diversify into balanced portfolios, For example, between gas and silver, gold and silver, Brent and silver, Bitcoin and gold.

The application of interdisciplinary and integrated tools for the modelling, forecasting and monitoring of financial markets, in particular stock, currency and cryptocurrency, is also the subject of the work of domestic scientists.

Ukrainian scientists [12] propose the procedure for determining the normalized economic coordinates, economic mass and heterogeneous economic time, based on the basic concepts of general theory of relativity and relativistic quantum mechanics. They are based on the analysis of time series describing socio-economic phenomena and economic interpretation of uncertainty by Heisenberg. The authors confirm that the economic mass of the time series can be an indicator of crisis phenomena.

In the research [14] the authors modelled and monitored cryptocurrency pairs with different capitalization degrees by the recursive and *R/S*-analysis. The usage of the R/S-method led them to conclude that the time series of selected cryptocurrencies were persistent, indicating clear market trends. The highest level of trends' persistence is the currency pair XRP/USD. And with the help of recurrence analysis, crisis periods in the cryptocurrency market were identified.

Therefore, a combination use of different methods of economic and mathematical modelling of the cryptocurrency market allows to increase the accuracy of forecasting, to identify the indicators of crisis phenomena, to analyze the degree of volatility and risk of both individual cryptocurrencies and their pairs.

2 Research methodology

2.1 Fractal analysis

R/S-analysis was first proposed by B. B. Mandelbrot and J. R. Wallis [23] and is based on H. Hurst hydrological studies. According to Hurst, the essence of fractal analysis for time series is to process the structure of a series that reflects a certain process and demonstrates a quantitative degree of chance. In paper [19], this approach was proposed to quantify the nature of self-affine structures.

The method of *R/S*-analysis allows calculating the parameter of self-similarity H, which measures the intensity of long-term dependencies in a time series. This metric is based on the analysis of the parameter range and the standard deviation.

Here is an algorithm for R/S-analysis. Suppose that the time series $y = \{y_i\}$, $I = 1, 2, \ldots, n$ is given, which must be led to "logarithmic returns". The resulting sequence is divided into initial segments $y = y_1, y_2, \ldots, y_N$.

(i) The average value and standard deviation of S_n are calculated for each segment.

(ii) The cumulative deviation is calculated:

$$x_{t,N} = \sum_{i=1}^{t}(y_i - M_N), \tag{1}$$

where y_i – increase in a year i, M_N – average, y_i for N periods.

(iii) Range is determined:

$$R(N) = \max(x_{t,N}) - \min(x_{t,N}), \tag{2}$$

where $R(N)$ – deviation range $x_{t,N}$, $\max(x_{t,N})$ – maximum value for $x_{t,N}$, $\min(x_{t,N})$ – minimum value for $x_{t,N}$.

(iv) From the ratio

$$R/S = (aN)^H, \tag{3}$$

where R/S – normalized range, N – observation number, a – constant, the Hurst exponent H is derived.

By the value of the Hurst coefficient, we can identify time series and draw conclusions about the minimum forecast of these series behaviour:

(i) $H = 0.5$ – the sequence is white noise; the time series is random. The future values of this series are independent of the previous ones. System status is random walk;

(ii) $0 < H < 0.5$ – anti-persistent or ergodic time series (pink noise), i.e. a series characterized by the so-called "return to average": if there is a decline in the indicator in the studied system over a period of time, an increase should be expected in the next period. The closer H is to 0, the more variable the series is. These changes happen quickly, often, but they are small. Note that such processes are few in reality. System status – flute;

(iii) $0.5 < H < 1$ – persistent or trend resistant time series (black noise, Brownian motion). The time series is characterized by the effect of long-term memory. If the series is growing, it is likely to continue to grow. If it falls today, a decline should be expected tomorrow. The closer the value of H is to 1, the more trend resistance is. The closer the value of N is to 0.5, the noisier the number will be. System status is a trend.

The method R/S-analysis has a significant limitation as it shows adequate results only for stationary series. However, most market processes are not static. The Method of Detrended Fluctuation Analysis (DFA), proposed in [28], is devoid of this shortcoming. It involves removing local trends on all segments. The local trend is understood as approximating a polynomial of some fixed order m. Approximation involves linear, quadratic, cubic or higher order polynomials (commonly called DFA1, DFA2, DFA3, etc.). Note that each value of the order of polynomial m corresponds to a separate scaling index $H_{[\mathrm{DFA}\,(m)]}$.

The advantages of DFA over other methods are that it exhibits long-term correlations of non-stationary time series, as well as ignores obvious random correlations, are a consequence of non-stationality.

Compared to *R*/*S*-analysis, the DFA method gives great possibilities to interpret the scaling index $H_{[\mathrm{DFA}\,(m)]}$ and allows to provide such a classification of series:

(i) $H_{[\mathrm{DFA}(m)]} = 0.5$ – random series;

(ii) if only short-term correlations are present $H_{[\mathrm{DFA}(m)]}$ may differ from 0.5, but tends to follow up to 0.5 when the size of the split segments increases;

(iii) $0.5 < H_{[\mathrm{DFA}(m)]} \leq 1$ – shows persistent long-term correlations;

(iv) $0 < H_{[\mathrm{DFA}(m)]} < 0.5$ – antipersistent series;

(v) $H_{[\mathrm{DFA}(m)]} = 1$ – existing flicker noise $1/f$;

(vi) $H_{[\mathrm{DFA}(m)]} > 1$, correlations exist but cease to reflect power relationships [35].

Application of different orders DFA allows to reveal hidden patterns in the input row.

The time sequences have both persistent and antipersistent gaps.

The local DFA method is based on the use of a mobile window of constant size. The window size shall be sufficient to use the standard DFA method by which the value of the local coefficient is determined. The movable window moves in a sequence with a certain step and the calculations are repeated. The window continues to move until it is fully depleted. Analysis of the local Hurst coefficient allows to identify and interpret significant changes in the dynamics of the system [35].

2.2 Tsallis and wavelet entropy

In economic systems, there is generally a strong interaction between its individual elements. As a result, we have broken the hypothesis of total chaos, and because of this, we have lost the relevance of its characteristics.

The idea of C. Tsallis [38] is to generalize the classical Shannon formula for entropy in the case of complex non-additive statistical systems by defining

$$S_q = \frac{k}{q-1}\left(1 - \sum_{i=1}^{n} p_i^q\right), \tag{4}$$

where k – is the Boltzmann constant, q – is the real number ($q \in \mathrm{R}$), n – is the total number of configurations, p_i – is the probability corresponding to the microscopic configurations.

Tsallis' entropy is a generalization of the non-additive measure and is calculated by the formula

$$S_q = \frac{1 - \int f(x)^q dx}{q-1}. \tag{5}$$

For anomalous systems with long memory and (or) long-range interactions, the coefficient q provides insight into the behavior of complex systems: at $q < 1$ in the system there is a predominance of rare phenomena, at $q > 1$ – the advantage of recurring phenomena if $q \to 1$ entropy of Tsallis is reduced to Shannon's entropy. The high values of this coefficient can be considered as a long-term memory parameter because they correspond to long-term relationships between the states of the system [37; 41]. It should be pointed out that for economic series $q < 1$ has not been observed yet.

As already noted, Tsallis' entropy is non-additive: if A and B are two independent states for which it is true that $p\ (A\ B) = p\ (A)\ p\ (B)$

$$S_q(A+B) = S_q(A)+S_q(B) - (1-q)S_q(A)S_q(B). \tag{6}$$

The parameter q is calculated from the cumulative distribution function

$$P_{q,k}(X \geq x) = \left(1 - \frac{(1-q)x}{k}\right)^{\frac{1}{1-q}}, \tag{7}$$

where k – defines limitations. Note that the usage of numerous and mathematical methods of parameter calculation leads to a system of nonlinear equations, the solution of which has a very large bias. Therefore, the research [34] proposes a maximum likelihood method with the previous reduction of the distribution function to the Pareto distribution to estimate the value of the parameter in order to minimize the calculation bias.

An important accent is that Tsallis entropy is calculated for a time series based on the probability distribution of these values, which is substantially non-Gaussian. Such distributions have profitable financial and economic series.

The wavelet transforms provide two-dimensional expansion of the investigated one-dimensional signal, that is, we have the possibility simultaneously to analyze the signal in both physical and frequency spaces. Wavelet transforms [5] are also used to describe the dynamics of complex non-linear processes where different perturbations in the spatial and temporal frequency ranges are possible. It is in the amplitude-time form that the information about the signal is written.

The essence of the wavelet-analysis method is that an appropriate basis and characteristics of the time series are introduced using the amplitude distribution (series count) in the introduced basis. Any function studied can only be decomposed into a basis if it is required to be orthogonal. This arrangement allows the restoration of the series [15].

Two types of wavelet transform are used in the studies: continuous and discrete.

Continuous wavelet transform (CWT) is directed at quantifying functions at a specific frequency and at a certain point in time. With CWT it is possible to decompose a function into elementary components, which is interesting when studying functions with complex structure. Note that CWT is better suited to studying functions than economic time series [24].

The discrete wavelet transform (DWT) can be used to derive a time series with values in the form of coefficients in the wavelet sequence. These coefficients provide complete information on a number of factors and make it possible to obtain local energy of its constituents at different scales. Note that information can be organized as a hierarchical schema of nested subspaces (multiscale analysis).

Let us give theoretical data of the bases of wavelet entropy and mathematical calculations [31; 40].

Consider time series $X = x_i$, $i = 1, \ldots, M$, the elements of which are represented on the uniform grid with the sampling time t_s (frequency ω_s). Apply the set of scales $1, \ldots, N$ and we will get

$$X(t) = \sum_{j=1}^{N} \sum C_j(k)\Psi_{j,k}(t) = \sum_{j=1}^{N} r_j(t). \quad (8)$$

Here $r_j(t)$ – a value containing information on a series X at frequencies $2^{j-1}\omega_s \leq |\omega_s| \leq 2^j\omega_s$, Ψ – the analytic function, called the parent wavelet, is the base for the wavelet transform. The spectral coefficients (wavelet coefficients) are defined as $C_j(k) = \langle S, \Psi_{j,k} \rangle$ and allow the energy at each scale j = 1, ..., N to be represented as the energy of the series elements [13].

The sum of the energies of each level gives us the total wavelet energy of the spectrum:

$$E_{tot} = \|X\|^2 = \sum_{j=1}^{N} \sum_k \left|C_j(k)\right|^2 = \sum_{j=1}^{N} E_j, \quad (9)$$

where N – maximum expansion level.

The paper [15] shows that the wavelet conversion of the original series is based on wavelet entropy, which is a quantitative measure of the ordered signal and is calculated by the formula:

$$WE = -\sum_{j=1}^{N} p_j ln(p_j), \quad (10)$$

where $p_j = \frac{E_j}{E_{tot}}$ – the relative wavelet energy over which the wavelet energy distribution is determined by the level of the arrangement. Values p_j at different scales form an energy probability distribution:

$$\sum_{j=1}^{n} p_j = 1. \quad (11)$$

Time scale density can be used to determine the characteristics of the time series, its characteristics both in time and in frequency space $\{p_j\}$.

Shannon entropy, which is a measure of the information contained in the distribution, is considered a criterion for analysing and comparing probability distributions. For the determination of normalized total wavelet entropy (*NTWE*) we have [40]:

$$E_{WT} = \frac{-\sum_{j=1}^{N} p_j \cdot ln p_j}{X_{max}}, \quad (12)$$

where $X_{max} = lnN$ the normalizing constant. Depending on the values that will be normalized wavelet entropy, it is possible to characterize the processes occurring in the time series. Normalized total wavelet entropy can be seen as a measure of the degree of chaotic time series, which can provide information about hidden dynamic processes occurring in the system [15].

In order to improve the algorithms for calculating the characteristics of wavelet entropy, the usage of window procedure is meaningful. The wavelet transform is automatically embedded in a movable frequency-time window. For this purpose, the windows of a certain length L are selected, without overlapping, forming i intervals $i = 1, \dots, N_T$, де $N_T = \frac{M}{L}$. The corresponding values of the time series of each interval shall be compared to the centre of the window. Therefore, at least one wavelet coefficient at each scale will be the smallest window length [13].

The moving window is equally good for both low- and high-frequency signal characteristics, which is an advantage of this method. By formula find wavelet energy at scale *j* for the time window *i*

$$E_j^{(i)} = \sum_{k=(i-1)\cdot L+1}^{i\cdot L} \left|C_j(k)\right|^2, i = 1, \dots, N_T. \tag{13}$$

So, we can find shared power in this time window

$$E_{tot}^{(i)} = \sum_{j=-N}^{-1} E_j^{(i)}. \tag{14}$$

The following formulas can be used to calculate the relative and normalized wavelet entropy reflecting changes over time:

$$p_j^{(i)} = \frac{E_j^{(i)}}{E_{tot}^{(i)}}, E_{WT}^{(i)} = -\sum_{j=-N}^{-1} p_j^{(i)} \cdot \frac{lnp_j^{(i)}}{X_{max}}. \tag{15}$$

In order to avoid differences in the number of wavelet coefficients at all scales, when the wavelet transform was performed, copies of the wavelet coefficient were added to the time output row on the right and left, so when the wavelet coefficients of C_{ij} are indicated, there are two indexes – scale number and point number respectively.

To calculate wavelet entropy, the Shannon entropy formula was used, which provides a simple method for calculating the probability density of time series energy.

In the paper [31] the indicators of wavelet entropy – scale and point entropy derived from the energy of wavelet coefficients are defined. These types of indicators are based on the Shannon formula approach: scale for entropy, hour for point for scale.

As a result of these calculations, find the wavelet coefficient energy field

$$E_{ij} = C_{ij}^2, \tag{16}$$

which is normalized by the standard average deviation of the output time series

$$\tilde{E}_{ij} = \frac{E_{ij}}{\sigma_t}. \tag{17}$$

The formula for calculating the scale of entropy is:

$$E_{WS} = -\sum_i \left(\left(\sum_j p_{ij}\right) \cdot \frac{log_2 \sum_j p_{ij}}{log_2 N} \right), \tag{18}$$

where $p_{ij} = \frac{\tilde{E}_{ij}}{\tilde{E}_{tot}}$ – entropy probability distribution, $E_{tot} = \sum_i \sum_j \tilde{E}_{ij}$, $log_2 N$ – normalizing multiplier (constant), *N* – number of time series elements.

The formula for calculating point entropy is:

$$E_{WP} = -\sum_i \left(\left(\sum_j p_{ij}\right) \cdot \frac{log_2 \sum_j p_{ij}}{log_2 N} \right). \tag{19}$$

To calculate point entropy, the energy values at each scale would be determined $\tilde{E}_i = \sum_i \tilde{E}_{ij}$ and is the distribution of the probability density of the energies by dividing the energies of the wavelet coefficients by the total energy of the corresponding scale $p_{ij} = \frac{\tilde{E}_{ij}}{\tilde{E}_i}$.

Thus, the technology of using wavelets allows to identify "spikes", sharp changes in the values of quantitative indicators at different periods of time. Wavelet analysis can identify not only obvious anomalies in the investigated system, but also critical values that may be hidden behind relatively small absolute values of the elements.

3 Research findings

Selected six cryptocurrencies in the first 20 capitalization rankings: Bitcoin (BTC), Ethereum (ETH), Ripple (XRP), EOS (EOS), Tether (USDT), USDCoin (USDC). In addition, cryptocurrencies were selected for the following species: those mined (BTC, ETH) – model Proof-of-Work, and those not mined (all others) – model Proof-of-Stake. Cryptocurrencies that those not mined, are also classified as cryptocurrencies (XRP, EOS) and stablecoins (USDT, USDC).

Here are the estimated values of the Hurst indicator, scaling index $H_{[\mathrm{DFA}(m)]}$ for cryptocurrency pairs (table 1). The study covers the time span from November 2019 to August 2020. The calculations were made in the MATLAB based data of [39].

Table 1. Results of the Hurst coefficient and scaling index $H_{[\mathrm{DFA}(m)]}$ calculations.

Currency pair	Hurst coefficient value (H), which have been calculated by R/S-analysis	Scaling index $H_{[\mathrm{DFA}(m)]}$, which have been calculated by DFA
BTC/USD	0.7122	0.54955
XRP/USD	0.7089	0.58829
ETH/USD	0.70511	0.58659
EOS/USD	0.70016	0.58658
USDT/USD	0.58997	0.21414
USDC/USD	0.571	0.15673

It is obvious that the values of the Hurst coefficient (table 1) are exaggerated. Consequently, an integrated approach is needed to evaluate cryptocurrency pairs. For this purpose, the scaling index $H_{[\mathrm{DFA}(m)]}$ is calculated, which will allow to clarify the conclusions. The value of the Hurst coefficient and the scaling index range from 0.5 to 1 for the cryptocurrency pairs BTC/USD, XRP/USD, ETH/USD, EOS/USD, indicates the persistence of the time series of the cryptocurrencies, i.e., the markets show clear trends. The XRP/USD currency pair has the highest meaning of the coefficient. In our view, the legal recognition of this cryptocurrency at the state level is an influential factor. It is one of the most popular cryptocurrencies today. It was created to accelerate transactions and currency conversion; it is also one of the cryptocurrencies used to support the technology of the future – Internet of Things.

Regarding the cryptocurrency pairs USDT/USD, USDC/USD with respect to these two indices, it is difficult to draw unambiguous conclusions. As we can see, the values of the indicators show opposite properties. However, as already noted, the Hurst coefficient is overestimated. Therefore, taking into account the obtained values of the scaling index $H_{[DFA(m)]}$, we can state the antipersistence of said rows. This behavior may be caused by the fact that the stablecoins under investigation are pegged to the value of fiat money – a dollar. Since the pandemic, which is in our view a form of crisis, has caused destabilizing developments in the world economy, fluctuations in the currency market, so do dollar-backed stablecoins.

Figure 1 shows the results of local DFA calculation for cryptocurrency pairs from August 2016 to August 2020 (BTC/USD, XRP/USD, ETH/USD, USDT/USD), from October 2018 to August 2020 (USDC/USD) and from July 2017 to August 2020 (EOS/USD). Calculations were performed in MATLAB with the following parameters: window width 100 points, step – 10 points based on data [39].

Thanks to the moving window procedure, we can observe the dynamics of the local Hurst coefficient and draw conclusions about changes in the dynamics of cryptocurrency pairs. The most stable cryptocurrency pair is XRP/USD (figure 2d). The local Hurst coefficient is sensitive to changes in the cryptocurrency market. The characteristic feature of the indicator is a significant rapid decline at the onset of the crisis. In the vicinity of point 25 (figure 1a, b, d), which corresponds to June 2017, the coefficient values are falling rapidly. There was a crisis in the cryptocurrency market – selloff. We also see the reaction to the next crisis of the cryptocurrency market – the "Great Chinese Cold" in the vicinity of Point 35 and the peak of this crisis in December 2017 (the vicinity of Point 70). In March 2020, the World Health Organization announced a pandemic, and at the end of March, partial or complete quarantine was introduced with the support of national Governments. The cryptocurrency market reacted by rapidly dropping local Hurst coefficients in the vicinity of point 120 (figure 1a, b, d). According to the figures. 1e for a pair of Stablecoin USDC/USD observes an extremely unstable dynamic of the local Hurst coefficient, whose value is predominantly less than 0.5, confirming the preliminary conclusion of the antipersistence of the series. For the USDT/USD pair (figure 1.f), the values of the coefficient are between 0.5 and 1, and at most intervals this series exhibits the persistent properties. According to the behavior of the local Hurst coefficient, we can observe the reaction of this cryptocurrency pair to the declaration of a pandemic. Despite the findings, it is worth noting that the cryptocurrency market is speculative. This is confirmed by the dynamics of the local Hurst coefficient for each cryptocurrency pair. As for the stablecoins we should admit that since USDT/USD and USDC/USD are actually provided with the dollar 1:1, there is a perceived artificial correction, and this may be a negative signal to investors, leading to additional risks.

Consider the results of the Tsallis entropy calculations (figure 2) and wavelet entropy (figure 3) for the cryptocurrency pairs studied.

Tsallis entropy was calculated in the MATLAB using a mobile window procedure and parameters: width of the window – 750 points, step – 5 points.

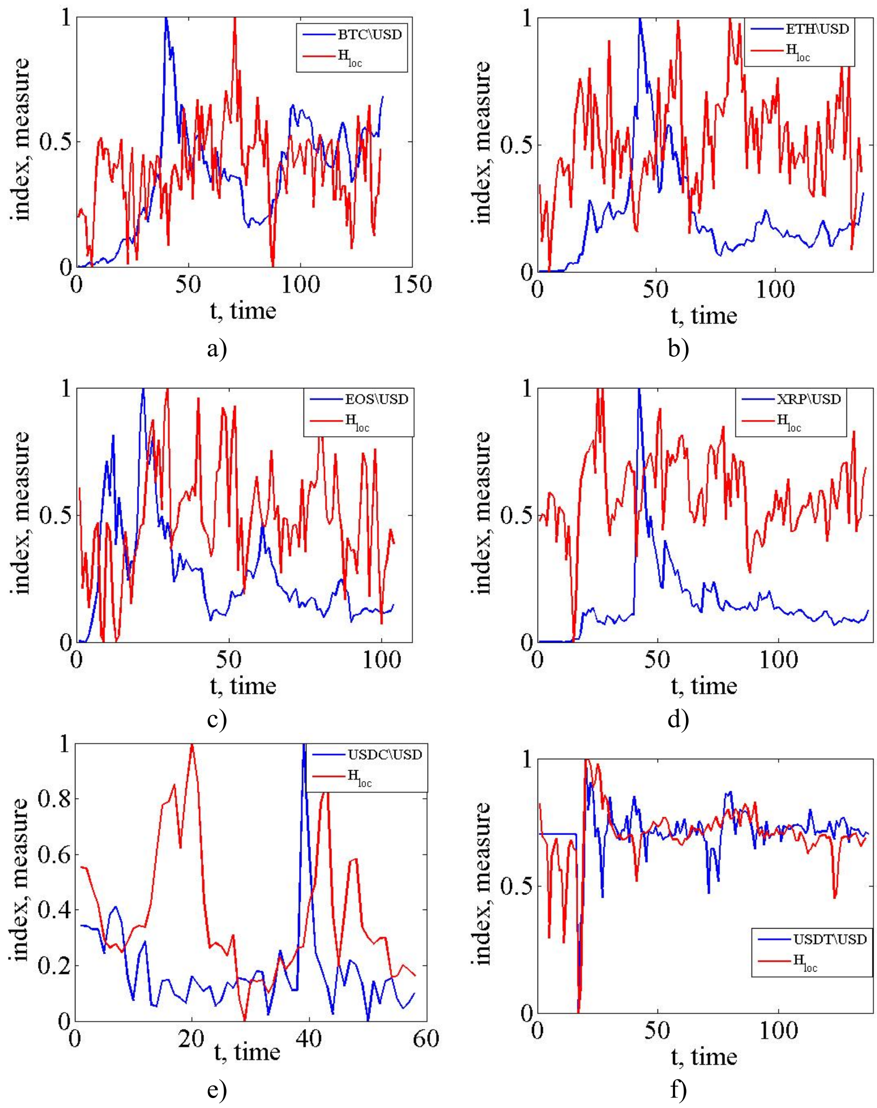
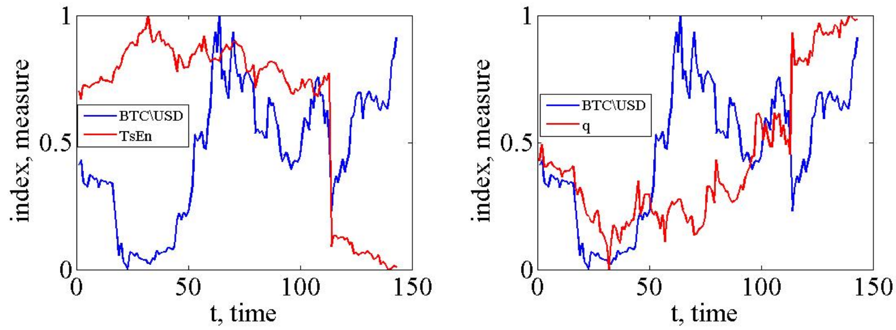


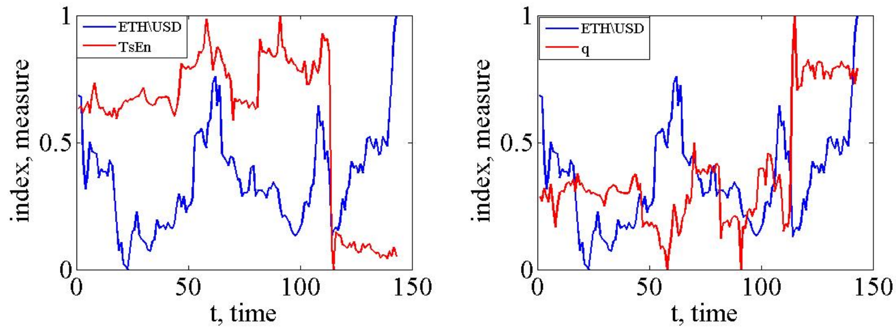
Fig. 1. Relative dynamics of the original series and the local Hurst coefficient for cryptocurrency pairs: a) BTC/USD, b) ETH/USD, c) EOS/USD, d) XRP/USD, e) USDT/USD, f) USDC/USD.

Tsallis entropy (TsEn) and q are indicators of critical phenomena in complex systems. The characteristics of these indicators are the rapid decline (TsEn) and the rapid growth (q) of values at the time of the crisis.

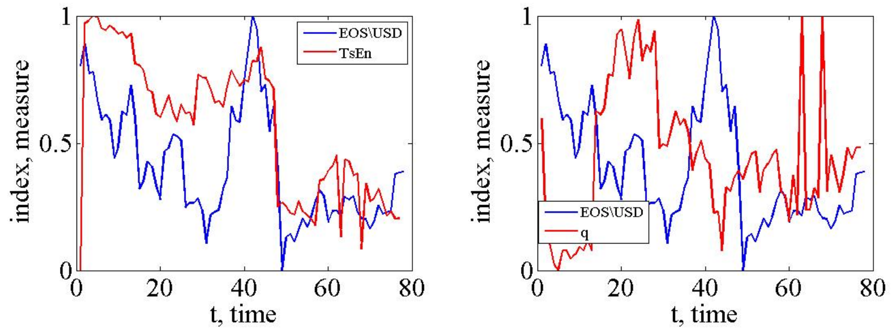
The following conclusions can be drawn from this analysis. For cryptocurrency pairs BTC/USD, ETH/USD in the vicinity of point 120 (March 2020), the characteristic behavior of the Tsallis entropy index and the coefficient q is observed. These values have not returned to the pre-pandemic values. Although the value of these cryptocurrencies has reached the pre-crisis period, the entropy rate for them is affected by the crisis period. For the cryptocurrency pair XRP/USD, the rapid change in entropy occurred earlier, and in this case, it worked as a warning indicator of a crisis event.



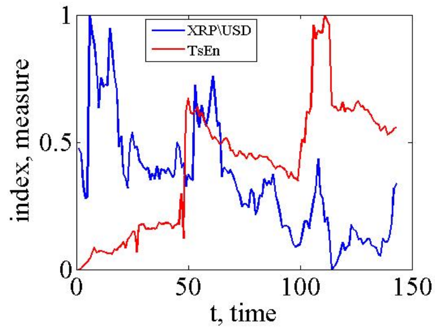
a)



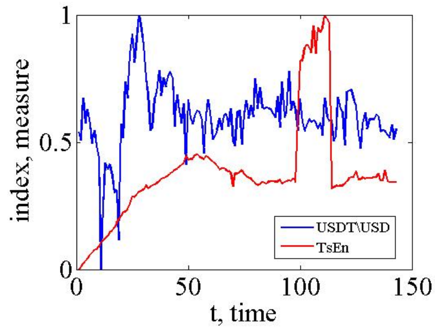
b)



c)



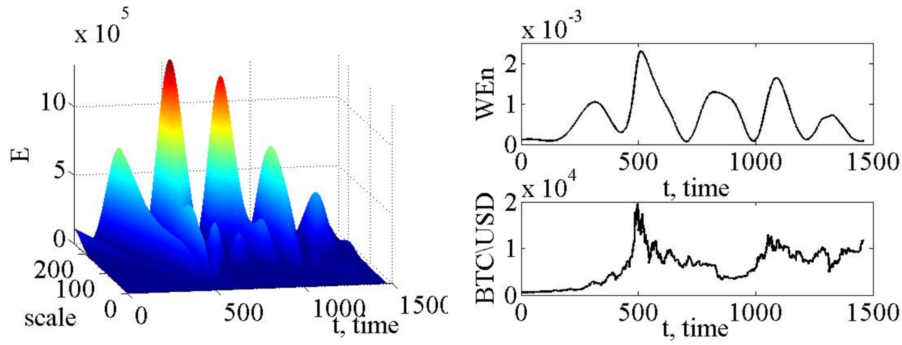
d)



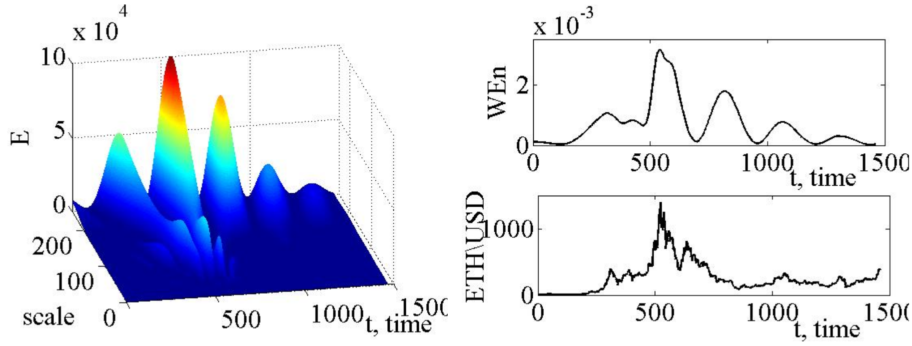
e)

Fig. 2. Relative dynamics of Tsallis entropy (left) and *q* coefficient (right) and cryptocurrency pairs: a) BTC/USD, b) ETH/USD, c) EOS/USD, d) XRP/USD, e) USDT/USD.

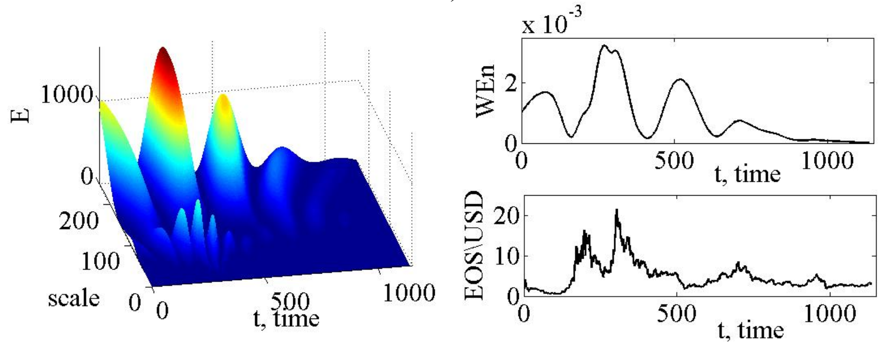
Tsallis' entropy behavior and the coefficient *q* for the cryptocurrency pair USDT / USD has also changed rapidly in the vicinity of point 110, but the return of these values to the pre-crisis values in a very short period of time confirms the artificial corrections of the stablecoins, provided by the dollar.



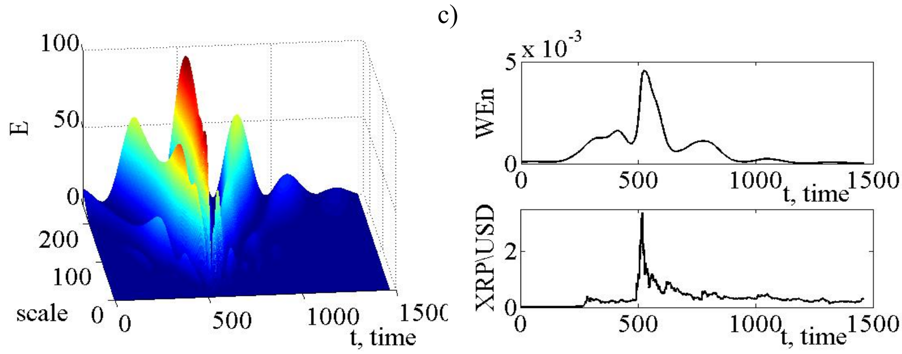
a)


x 10^4
E
scale
t, time
x 10^-3
WEn
t, time
ETH\USD
t, time

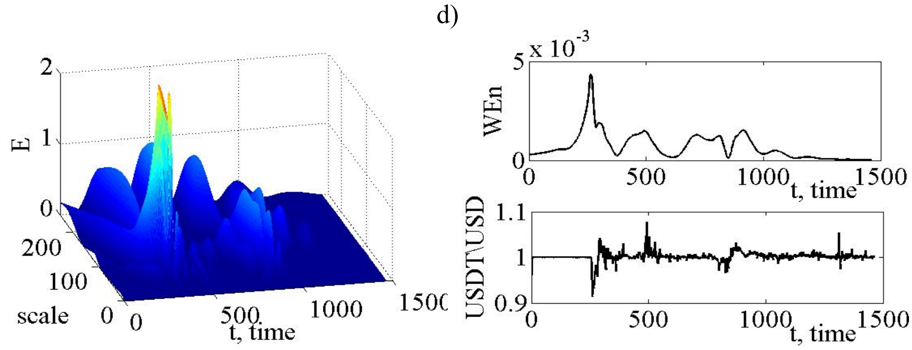
b)


E
scale
t, time
x 10^-3
WEn
t, time
EOS\USD
t, time

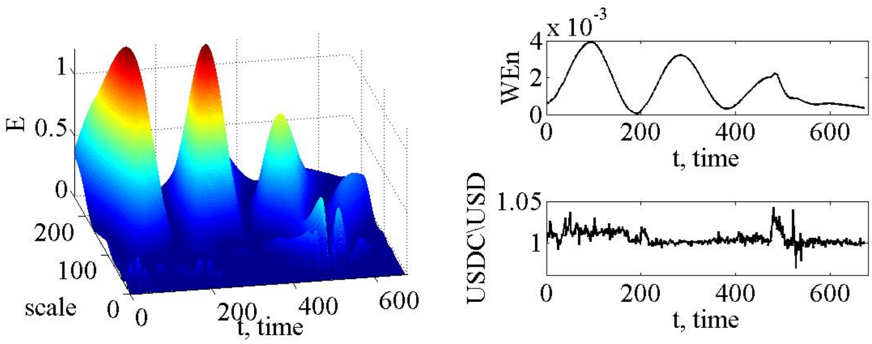
c)


E
scale
t, time
x 10^-3
WEn
t, time
XRP\USD
t, time

d)


E
scale
t, time
x 10^-3
WEn
t, time
USDT\USD
t, time

e)



f)

Fig. 3. Wavelet coefficient energy (left) and point wavelet entropy dynamics (right) for cryptocurrency pairs a) BTC/USD, b) ETH/USD, c) EOS/USD, d) XRP/USD, e) USDT/USD, f) USDC/USD.

Wavelet entropy is a reliable indicator of crisis phenomena in social and economic systems. As with the previous indicator, wavelet entropy has characteristics of behaviour before and after the crisis. Empirically proved the response of "law of three waves" in the pre-crisis period: on the energy surface (as on the point wave entropy graph) characteristic increasing oscillations are formed, with the maximum of each subsequent wave being greater than the previous one. The signal of a possible crisis is the exceedance of the wavelet entropy of the third wave maximum of the previous wave. In the post-crisis period fluctuations are markedly attenuated. Wavelet entropy (WEn) was calculated with the following parameters: width of the window – 750 points, step – 5 points, scale – 300 points.

Figure 3 shows that for one cryptocurrency there are no "three waves" observed, either at the beginning of the pandemic or at other periods. The results of the wavelet entropy calculation allow to refine the preliminary conclusions of the cryptocurrency market according to other indicators. In our opinion, the cryptocurrency market has been shocked by the COVID-19 outbreak and unprecedented measures. However, the ability to explore the time series relative to scale provides additional information about hidden processes in this market. In fact, disturbances for all cryptocurrency pairs are present on a small scale. This confirms once again the high activity within the market and it is likely to be speculative.

4 Conclusions

We can conclude based on the results of the cryptocurrency market research that cryptocurrencies have long-term potential and prospects in today's globalized economy being subject to crisis.

The analysis of the cryptocurrency market conducted in the article on the basis of fractal analysis leads to the conclusion that the cryptocurrency pair XRP/USD has the highest level of trends' persistence. Stable antipersistence detected for cryptocurrency pair USDC / USD. The results of calculating the local Hurst coefficient at the time of the official declaration of a pandemic by the World Health Organization, for virtually all the cryptocurrency pairs under investigation, justify anti-persistence. However, in the future, the cryptocurrency market is witnessing a return of pre-pandemic trends.

Interdisciplinary methods of financial market research, namely Tsallis entropy and wavelet entropy, are proposed to compare and refine the results of the calculations. The results of the calculations show that the cryptocurrency market has indeed reacted to a crisis such as COVID-19. It should be noted, however, that no significant impact has been discovered and the cryptocurrency market is now considered to be more susceptible to speculation.

The proposed methodology for monitoring and modelling the cryptocurrency market is of practical importance as it will allow potential investors to form a profitable portfolio with a high level of reliability and stability over time.

In addition, the results of the research can be useful to all scientists who are working on the effects of COVID-19 on this market.

References

1. Adcock, R., Gradojevic, N.: Non-fundamental, non-parametric Bitcoin forecasting. Physica A: Statistical Mechanics and its Applications **531**, 121727 (2019). doi:10.1016/j.physa.2019.121727
2. Altan, A., Karasu, S., Bekiros, S.: Digital currency forecasting with chaotic meta-heuristic bio-inspired signal processing techniques. Chaos, Solitons & Fractals **126**, 325–336 (2019). doi:10.1016/j.chaos.2019.07.011
3. Antonakakis, N., Chatziantoniou, I., Gabauer, D.: Cryptocurrency market contagion: Market uncertainty, market complexity, and dynamic portfolios. Journal of International Financial Markets, Institutions and Money **61**, 37–51 (2019). doi:10.1016/j.intfin.2019.02.003
4. Ashraf, B.N.: Stock markets' reaction to COVID-19: Cases or fatalities? Research in International Business and Finance **54**, 101249 (2020). doi:10.1016/j.ribaf.2020.101249
5. Astafeva, N.M.: Veyvlet-analiz: osnovyi teorii i primeryi primeneniya (Wavelet Analysis: Theory Basics and Application Examples). Uspehi fizicheskih nauk, 166(11), 1145–1170 (1996)
6. Belew, S., Elad, J.: Starting an Online Business All-in-One for Dummies, 5th edn. For Dummies, New Jersey (2017)
7. Berentsen, A., Schar, F.: A Short Introduction to the World of Cryptocurrencies. Federal Reserve Bank of St. Louis Review **100**(1), 1–16 (2018). doi:10.20955/r.2018.1-16
8. Caporale, G.M., Zekokh, T.: Modelling volatility of cryptocurrencies using Markov-Switching GARCH models. Research in International Business and Finance **48**, 143–155 (2019). doi:10.1016/j.ribaf.2018.12.009
9. Catania, L., Grassi, S., Ravazzolo, F.: Forecasting cryptocurrencies under model and parameter instability. International Journal of Forecasting **35**(2), 485–501 (2019). doi:10.1016/j.ijforecast.2018.09.005
10. Chaim, P., Laurini, M.P.: Nonlinear dependence in cryptocurrency markets. The North American Journal of Economics and Finance **48**, 32–47 (2019). doi:10.1016/j.najef.2019.01.015

11. Cheng, Q., Liu, X., Zhu, X.: Cryptocurrency momentum effect: DFA and MF-DFA analysis. Physica A: Statistical Mechanics and its Applications **526**, 120847 (2019). doi:10.1016/j.physa.2019.04.083
12. Danilchuk, A., Solovev, V.: Ispolzovanie princzipa neopredelennosti Gejzenberga dlya modelirovaniya krizisnykh yavlenij na rynke kriptovalyut (Use of the Heisenberg uncertainty principle to model crisis phenomena in the cryptocurrency market). In: Ekonomicheskie i finansovye mekhanizmy innovaczionnogo razvitiya czifrovoj ekonomiki, vol. 2, pp. 172–177. Institut biznesa BGU, Minsk (2019)
13. Danylchuk H.B., Soloviov V.M.: Vykorystannia entropiinykh pokaznykiv dlia vymiriuvannia skladnosti ekonomichnykh system (Use of entropy indicators to measure the complexity of economic systems). Visnyk Kryvorizkoho ekonomichnoho instytutu KNEU 2(14), 61–69 (2008)
14. Danylchuk, H., Kovtun, O., Kibalnyk, L., Sysoiev, O.: Monitoring and modelling of cryptocurrency trend resistance by recurrent and R/S-analysis. E3S Web of Conferences **166**, 13030 (2020). doi:10.1051/e3sconf/202016613030
15. Danylchuk, H.B., Serdiuk, O.A.: Vykorystannia riznykh metodiv otsinky entropii dlia analizu ta modeliuvannia ekonomichnykh system (Using different methods of estimation of entropy for analysis and modeling of economic systems). In: Modeliuvannia ta informatsiini tekhnolohii v ekonomitsi, pp. 111–139. Brama-Ukraina, Cherkasy (2014)
16. Gatabazi, P., Mba, J.C., Pindza, E.: Modeling cryptocurrencies transaction counts using variable-order Fractional Grey Lotka-Volterra dynamical system. Chaos, Solitons & Fractals **127**, 283–290 (2019). doi:10.1016/j.chaos.2019.07.003
17. Goodell, J.W.: COVID-19 and finance: agendas for future research. Finance Research Letters **35**, 101512 (2020). doi:10.1016/j.frl.2020.101512
18. Hamaniuk, V., Semerikov, S., Shramko, Y.: ICHTML 2020 – How learning technology wins coronavirus. SHS Web of Conferences **75**, 00001 (2020). doi:10.1051/shsconf/20207500001
19. Hurst, H.E.: Long-Term Storage Capacity of Reservoirs. Transactions of the American Society of Civil Engineers 116(1), 770–799 (1951)
20. Lahmiri S., Bekiros S.: Big data analytics using multi-fractal wavelet leaders in high-frequency Bitcoin markets. Chaos, Solitons & Fractals **131**, 109472 (2020). doi:10.1016/j.chaos.2019.109472
21. Lahmiri S., Bekiros S.: Renyi entropy and mutual information measurement of market expectations and investor fear during the COVID-19 pandemic. Chaos, Solitons & Fractals **139**, 110084 (2020). doi:10.1016/j.chaos.2020.110084
22. Lahmiri, S., Bekiros, S.: The impact of COVID-19 pandemic upon stability and sequential irregularity of equity and cryptocurrency markets. Chaos, Solitons & Fractals **138**, 109936 (2020). doi:10.1016/j.chaos.2020.109936
23. Mandelbrot, B.B., Wallis, J.R.: Robustness of the rescaled range R/S in the measurement of noncyclic long run statistical dependence. Water Resources Research **5**(5), 967–988 (1969). doi:10.1029/WR005i005p00967
24. Masset, P.: Analysis of Financial Time-Series using Fourier and Wavelet Methods. SSRN (2008). doi:10.2139/ssrn.1289420
25. Mensi, W, Lee, Y-J., Al-Yahyaee, K.H., Sensoy, A., Yoon, S-M.: Intraday downward/upward multifractality and long memory in Bitcoin and Ethereum markets: An asymmetric multifractal detrended fluctuation analysis. Finance Research Letters **31**, 19–25 (2019). doi:10.1016/j.frl.2019.03.029

26. Mnif, E., Jarboui, A., Mouakhar K.: How the cryptocurrency market has performed during COVID 19? A multifractal analysis. Finance Research Letters **36**, 101647 (2020). doi:10.1016/j.frl.2020.101647
27. Omane-Adjepong, M., Alagidede, P., Kwame Akosah, N.: Wavelet time-scale persistence analysis of cryptocurrency market returns and volatility. Physica A: Statistical Mechanics and its Applications **514**, 105–120 (2019). doi:10.1016/j.physa.2018.09.013
28. Peng, C.-K., Buldyrev, S.V., Havlin, S., Simons, M., Stanley, H.E., Goldberger, A.L.: Mosaic organization of DNA nucleotides. Physical Review E **49**, 1685 (1993). doi:10.1103/PhysRevE.49.1685
29. Phillip, A., Chan, J., Peiris, S.: On generalized bivariate student-t Gegenbauer long memory stochastic volatility models with leverage: Bayesian forecasting of cryptocurrencies with a focus on Bitcoin. Econometrics and Statistics **16**, 69–90 (2018). doi:10.1016/j.ecosta.2018.10.003
30. Pursky, O.I., Grynyuk, B.V., Shestopal, D.A.: Planning of advertising costs and vendor number at e-trade market. Actual Problems of Economics 177(3), 407–413 (2016)
31. Quian Quiroga, R., Rosso, O.A., Başar, E.: Wavelet entropy: a measure of order in evoked potentials. Electroencephalography and clinical neurophysiology Supplement 49, 299–303 (1999)
32. Schneider, G.: Electronic Commerce, 12th edn. Cengage Learning, Boston (2016)
33. Semerikov, S., Chukharev, S., Sakhno, S., Striuk, A., Osadchyi, V., Solovieva, V., Vakaliuk, T., Nechypurenko, P., Bondarenko, O., Danylchuk, H.: Our sustainable coronavirus future. E3S Web of Conferences **166**, 00001 (2020). doi:10.1051/e3sconf/202016600001
34. Shalizi, C.R.: Maximum likelihood estimation for q-exponential (Tsallis) distributions. arXiv:math/0701854v2 [math.ST] (2007). Accessed 14 May 2020
35. Soloviov, V.M., Derbentsev, V.D., Serdiuk, O.A., Sharapov, O.D.: Synerhetychni ta ekonofizychni metody doslidzhennia dynamichnykh ta strukturnykh kharakterystyk ekonomichnykh system (Synergetic and econophysical methods for the study of dynamic and structural characteristics of economic systems). Brama–Ukraina, Cherkasy (2010)
36. Topcu, M., Gulal, O.S.: The impact of COVID-19 on emerging stock markets. Finance Research Letters **36**, 101691 (2020). doi:10.1016/j.frl.2020.101691
37. Tsallis, C.: Introduction to Nonextensive Statistical Mechanics: Approaching a Complex Word. Springer, New York (2009)
38. Tsallis, C.: Possible generalization of Boltzmann-Gibbs statistics. Journal of Statistical Physics **52**(1/2), 479–487 (1988). doi:10.1007/BF01016429
39. Verizon Media: Yahoo Finance - Stock Market Live, Quotes, Business & Finance News. https://finance.yahoo.com (2020). Accessed 17 Aug 2020
40. Zunino, L., Pérez, D.G., Garavaglia, M., Rosso, O.A.: Wavelet entropy of stochastic processes. Physica A: Statistical Mechanics and its Applications **379**(2), 503–512 (2007). doi:10.1016/j.physa.2006.12.057
41. Zunino, L., Pérez, D.G., Kowalski, A., Martín, M.T., Garavaglia, M., Plastino, A., Rosso O.A.: Fractional Brownian motion, fractional Gaussian noise, and Tsallis permutation entropy. Physica A: Statistical Mechanics and its Applications **387**(24), 6057–6088 (2008)

Econophysics of sustainability indices

Andrii Bielinskyi[1[0000-0002-2821-2895]], Serhiy Semerikov[1,2,3[0000-0003-0789-0272]], Oleksandr Serdyuk[4[0000-0002-1230-0305]], Victoria Solovieva[5[0000-0002-8090-9569]], Vladimir Soloviev[1,4[0000-0002-4945-202X]] and Lukáš Pichl[6]

[1] Kryvyi Rih State Pedagogical University, 54 Gagarin Ave., Kryvyi Rih, 50086, Ukraine
[2] Kryvyi Rih National University, 11 Vitalii Matusevych Str., Kryvyi Rih, 50027, Ukraine
[3] Institute of Information Technologies and Learning Tools of NAES of Ukraine, 9 M. Berlynskoho Str., Kyiv, 04060, Ukraine
[4] The Bohdan Khmelnytsky National University of Cherkasy, 81 Shevchenka Blvd., 18031, Cherkasy, Ukraine
[5] State University of Economics and Technology, 16 Medychna Str., Kryvyi Rih, 50005, Ukraine
[6] International Christian University, 3-10-2 Osawa, Mitaka-shi, Tokyo, 181-8585, Japan

krivogame@gmail.com, semerikov@gmail.com, serdyuk@ukr.net, vikasolovieva2027@gmail.com, vnsoloviev2016@gmail.com, lukas@icu.ac.jp

Abstract. In this paper, the possibility of using some econophysical methods for quantitative assessment of complexity measures: entropy (Shannon, Approximate and Permutation entropies), fractal (Multifractal detrended fluctuation analysis – MF-DFA), and quantum (Heisenberg uncertainty principle) is investigated. Comparing the capability of both entropies, it is obtained that both measures are presented to be computationally efficient, robust, and useful. Each of them detects patterns that are general for crisis states. The similar results are for other measures. MF-DFA approach gives evidence that Dow Jones Sustainability Index is multifractal, and the degree of it changes significantly at different periods. Moreover, we demonstrate that the quantum apparatus of econophysics has reliable models for the identification of instability periods. We conclude that these measures make it possible to establish that the socially responsive exhibits characteristic patterns of complexity, and the proposed measures of complexity allow us to build indicators-precursors of critical and crisis phenomena.

Keywords: Dow Jones Sustainability Index, measures of complexity, precursors of stock market crashes.

1 Introduction

Current economic trends have convincingly demonstrated that green development is a necessary condition for sustainable development, which is essential for a better life in the future [40]. Economists have described climate change as a global market failure estimating that without action, the rising overall costs of climate could result in losing at least 5% of global GDP each year. A growing number of financial institutions are joining in a constructive dialogue on the relationship between economic development, environmental protection, and sustainable development. Financial institutions, including banks, insurers, and investors, work with the United Nations Environment Programme – Finance Initiative to better understand environmental, social, and governance challenges, why they matter to finance, and how to take steps to address them [30].

The availability of stock indexes based on sustainability screening makes increasingly viable for institutional investors the transition to a portfolio based on a Socially Responsible Investment (SRI) benchmark at a relatively low cost.

The 2008 subprime crisis and increased social awareness have led to a growing interest in topics related to Socially Responsible Investment. SRI is a long-term investment that integrates environmental, social, and corporate governance criteria (ESG). According to the Global Sustainable Investment Alliance (GSIA), SRI reached 24 trillion euro's in 2016, registering a growth of 25.2% between 2014 and 2016. So, green and sustainable finance is more important nowadays than ever before [14].

This increased social interest coincides with international initiatives aimed at developing environmental and social policies on sustainable finance issues, such as the Action Plan on sustainable finance adopted by the European Commission in March 2018. This plan has three main objectives:

(i) to redirect capital flows towards sustainable investment to achieve sustainable and inclusive growth,

(ii) to manage financial risks stemming from climate change, environmental degradation, and social issues, and

(iii) to foster transparency and long-termism in financial and economic activity. Therefore, the main purpose is to enhance the role of finance and to build an economy that enables the goals of the Paris Agreement (2015) and the EU for sustainable development to be reached [15].

The Dow Jones Sustainability Index (DJSI) comprises global sustainability leaders as identified by SAM. It represents the top 10% of the largest 2,500 companies in the S&P Global BMI based on long-term economic, environmental, and social criteria [59]. Founded in 1995, RobecoSAM is an investment specialist focused exclusively on Sustainability Investing [38].

The S&P Global Broad Market Index (BMI) is the only global index suite with a transparent, modular structure that has been fully float-adjusted since 1989. This comprehensive, rules-based index series employs a transparent and consistent methodology across all countries and includes more than 11,000 stocks from 25 developed and 25 emerging markets [39]. The SAM Corporate Sustainability Assessment (CSA), established by RobecoSAM, is now issued by S&P Global. RobecoSAM, an asset manager focused entirely on sustainable investing, established the CSA in 1999. The CSA has become the basis for numerous S&P ESG Indices over the last two decades attracting billions of USD in assets. Besides, S&P Global acquired RobecoSAM's ESG Ratings and Benchmarking businesses which operate out of S&P Global Switzerland. SAM is a registered trademark of S&P Global. ESG is a generic term used in capital markets and used by investors to evaluate corporate behavior and to determine the future financial performance of companies. In the conditions of a wide variety of sustainable development indices, investors need to have a comparative characteristic of traditional indices with sustainable development indices obtained by *quantitative methods*. At the same time, the set of tools of modern financial analysis took shape in a separate rapidly growing applied science – *fintech*. Financial technology ('fintech') is emerging as a core disruptor of every aspect of today's financial system. Fintech covers everything from mobile payment platforms to high-frequency trading, and from crowdfunding and virtual currencies to blockchain. In combination, such forceful innovations will threaten the viability of today's financial sector business models, and indeed the effectiveness of current policies, regulations, and norms that have shaped modern finance.

The use of financial technology innovations is of course not new – but a step change is now expected with the novel application of several technologies in combination, notably involving blockchain, the 'Internet of things', and artificial intelligence [6]. The widespread introduction of fintech makes it possible to talk about *green finance* as a strategy for the financial sector and broader sustainable development that is relevant around the world [1; 7; 24; 33]. Green economy, green finance, and green development are the peculiar coordinates of the phase space in which today it is generally accepted to evaluate the sustainable development of world civilization.

Financial systems are complex systems and consist of a plurality of interacting agents possessing the ability to generate new qualities at the level of macroscopic collective behavior, the manifestation of which is the spontaneous formation of noticeable temporal, spatial, or functional structures [54]. For many years financial markets have been attracting the attention of many scientists like engineers, mathematicians, physicists, and others for the last two decades. Such vast interest transformed into a branch of statistical mechanics – *econophysics* [25]. Physics, economics, finance, sociology, mathematics, engineering, and computer science are fields of science which, as a result of cross-fertilization, have created the multi-, cross-, and interdisciplinary areas of science and research such as econophysics and sociophysics, thriving in the last two and a half decades. These mixed research fields use knowledge, methodologies, methods, and tools of physics for modeling, explaining, and forecasting economic and social phenomena and processes. Accordingly, econophysics is an interdisciplinary research field, applying theories and methods originally developed by physicists to solve problems in economics, usually those including uncertainty or stochastic processes, nonlinear dynamics, and evolutionary games. Obviously, quantitative econophysical methods for studying financial markets are an interesting and promising area of fintech.

Our research structured as follows. Section 2 contains a brief description of socially responsive indexes and an analysis of previous work on a comparative quantitative analysis of this variety of indices. Section 3 describes algorithms for constructing econophysical measures of complexity based on the informational, (multi-)fractal and quantum physical properties of a time series. These measures are calculated based on the DJSI index. Section 4 summarizes the results obtained and indicates the direction of subsequent studies.

2 Review of previous research

In the last 20-25 years, a huge number of social responsibility or sustainability indices have been created and their number continues to grow [13; 53]. Briefly consider the most commonly used.

The Dow Jones Sustainability Indices are a family of best-in-class benchmarks for investors who have recognized that sustainable business practices are critical to generating long-term shareholder value and who wish to reflect their sustainability convictions in their investment portfolios (http://www.sustainability-indices.com/). The family was launched in 1999 as the first global sustainability benchmark and tracks the stock performance of the world's leading companies in terms of economic, environmental, and social criteria. Dow Jones Sustainability World Index, the most important global stock market valuation index of corporate social responsibility.

FTSE4Good was created by the FTSE Group to facilitate investments in companies that meet globally recognized corporate responsibility standards and constitutes an important reference point for the establishment of benchmarks and ethical portfolios. Companies in the FTSE4Good Index have met stringent environmental, social, and governance criteria, and are therefore potentially better positioned to capitalize on the benefits of responsible business practice (http://www.ftse.com/).

MSCI is a leading provider of investment decision support tools to investors globally, including asset managers, banks, hedge funds, and pension funds. MSCI Global Sustainability Indexes include companies with high ESG ratings relative to their sector peers (http://www.msci.com/).

CDP (formerly the “Carbon Disclosure Project”) is one of the world’s leading not-for-profit climate change organizations, assessing transparency in the disclosure of information on climate change and greenhouse gas emissions, as well as in the management of water resources (http://www.cdp.net/).

United Nations Global Compact 100 (“GC 100”), a global stock index developed and released by the UN Global Compact in partnership with the research firm Sustainalytics (https://www.unglobalcompact.org/). The index lists the 100 companies which globally outstand for executive leadership commitment and consistent baseline profitability, as well as their adherence to the Global Compact’s ten principles, on human rights, labor, environment, and anti-corruption issues.

STOXX Global ESG Leaders Indices, a group of indices based on a fully transparent selection process of the performance, in terms of sustainability, of 1,800 companies worldwide (http://www.stoxx.com/). The ratings are calculated for three sub-areas – environmental, social, and governance – and are then combined to form the overall index. The indices are managed by STOXX, the owner of some of the most important international stock indices, such as the STOXX50.

In our previous work [11], we performed a comparative analysis of the index DJSI [62] with its classic and traditional counterpart – the index Dow Jones Industrial Average (DJIA) [61].

In a comparative analysis of structural and dynamic properties of traditional stock market indices and social responsibility indices, descriptive statistics methods are used in most works [2; 31; 34; 43].

Descriptive statistics (mean, maximum, minimum, and standard deviation) of the financial information required to apply the Ohlson [34] valuation model reviewed in [31]. They were examining whether sustainability leadership – proxied by the membership of the Dow Jones Sustainability Index Europe – is value relevant for investors on the 10 major European stock markets over the 2001–2013 period. These results reveal that there exist significant differences across markets.

The article [43] analyzes rate-of-return and risk related to investments in socially responsible and conventional country indices. The socially responsible indices are the DJSI Korea, DJSI US, and Respect Index, and the corresponding conventional country indices are the Korea Stock Exchange Composite KOSPI, DJIA, and WIG20. Shown, that conclude that investing in the analyzed SRI indices do not yield systematically better results than investing in the respective conventional indices, both in terms of neoclassical risk and return rate.

The authors [2] examined sustainable investment returns predictability based on the US DJSI and a wide set of uncertainty and financial distress indicators for the period January 2002 to December 2014. They employ a novel nonparametric causality-in-quantile approach that captures nonlinearity in returns distribution. The authors conclude that the aggregate Economic Policy Uncertainty indicator and some components have predictive ability for real returns of the US sustainable investments index. Paper [55] explores the relationship between sustainability performance and financial performance by looking at the impact of sustainability index changes on the market value of a company. The author has studied the price effects of changes in the DJSI and FTSE4Good Index. He failed to observe statistically significant positive abnormal returns for companies being added to a sustainability index. On the opposite, he finds negative abnormal returns for companies being deleted from the FTSE, however not in the case of the DJSI. This can be explained by studying the volume effects and the behavior of investment managers.

However, the first works appeared using more modern methods of analysis, using the achievements of nonlinear dynamical systems and complexity theory [16; 26; 29; 32; 58]. The authors [58] constructed a sustainable regional green economy development index system from five aspects - economic, social, technological, resources, and environmental - using DPSIR (drivers, pressures, state, impact, response model) and entropy-TOPSIS (a technique for order preference by similarity to an ideal solution) coupling coordination to horizontally and vertically quantitatively analyze the sustainable green economy development. The model was verified by the actual situation of green economy development in Shandong Province from 2010 to 2016, which confirmed the feasibility of the method.

A sustainable development capacity measure model for Sichuan Province was established by applying the information entropy calculation principle and the Brusselator principle [26]. Each subsystem and entropy change in a calendar year in Sichuan Province were analyzed to evaluate Sichuan Province's sustainable development capacity. It was found that the established model could effectively show actual changes in sustainable development levels through the entropy change reaction system, at the same time this model could clearly demonstrate how those forty-six indicators from the three subsystems impact on the regional sustainable development, which could make up for the lack of sustainable development research.

A similar approach is implemented to measure the tourist attractiveness of the region [16]. And in work [29] information and entropy theory used for the sustainability of coupled human and natural systems.

Authors of [32] used R/S analysis to calculate the Hurst exponent as a measure of persistence (efficiency of traditional stock market indices and social responsibility stock market indices). The presence of persistence was evidence in favor of less efficiency. According to empirical results, SRI has lower efficiency, in particular the Dow Jones Sustainability Index. Lower efficiency was also observed in the emerging markets with a responsible investment segment, compared to the traditional stock market indices.

In paper [27] authors suggest three new indicators based on an engineering approach of irreversibility. They allow evaluating both the technological level and the environmental impact of the production processes and the socio-economic conditions of the countries. Indeed, they are based on the energy analysis and on the irreversible thermodynamic approach, in order to evaluate the inefficiency both of the process and of the production systems, and the related consequences. Three applications are summarized in order to highlight the possible interest from different scientists and researchers in engineering, economy, etc. [19; 23], in order to develop sustainable approaches and policies for decision-makers.

All mentioned measures can capture nonlinearity and complexity that peculiar even for sustainability indices. Analysis of previous papers [47; 48; 49; 51; 52] shows that indicators have theoretical perspectives and, in accordance with other studies, such approaches are presented to be robust and computationally efficient. In some aspects, the results of the multifractal analysis are presented to be better, but the computational costs leave a lot to be desired. Therefore, due to computationally efficiency and ability to monitor, and prevent crisis events in advance, the entropy measures present to be the most attractive. However, empirical results of quantum and multifractal measures present to be optimal that motivate further research work.

3 Econophysical measures of DJSI complexity and precursors of crisis states

In a series of recent works [4; 44; 45; 50], we have demonstrated the possibility of using the theory of complex systems and a set of developed analysis tools to calculate the corresponding measures of system complexity. These complexity measures make it possible to differentiate systems according to the degree of their functionality, to identify and prevent critical and crisis phenomena.

Since the DJSI index is used as a calculation base, we will provide more detailed information for it. DJSI measures the performance of companies selected for economic, environmental, and social criteria that weighted by market capitalization using a best-in-class approach. In assessing sustainability, the key factor in selecting components for any DJSI index is the overall company sustainability rating (TSS). The first CSA was undertaken in 1999 with the launch of the original DJSI family of indexes. The annual CSA process begins in March of each year and is published with new estimates in September. The index is calculated using the divisor methodology that is used for all Dow Jones Index stock indices. Indices are calculated daily throughout the calendar year. The exception is those days when all exchanges on which the index constituents are quoted are officially closed or if the WM Reuters exchange rate services are not published.

The table 1 and figure 1 provide information on the key companies in the index basket and the weight of the respective economic sectors to which they belong.

Table 1. Top 10 components by index weight.

No	Constituent	Symbol	Sector
1	Microsoft Corp	MSFT	Information
2	Technology Alphabet Inc C	GOOG	Communication Services
3	Nestle SA Reg	NESN	Consumer
4	Staples United health Group Inc	UNH	Health Care
5	Taiwan Semiconductor Manufacturing Co Ltd	2330	Information
6	Technology Roche Hldgs AG Ptg Genus	ROG	Health Care
7	Adobe Inc.	ADBE	Information
8	Technology Novartis AG Reg	NOVN	Health Care
9	Cisco Systems Inc	CSCO	Information
10	Technology Bank of America Corp	BAC	Financials

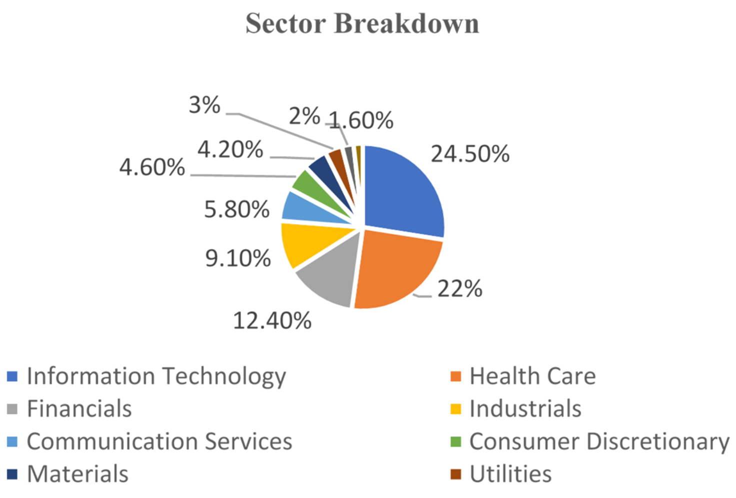


Fig. 1. Weights for each sector of the index, %.

For the daily DJSI time series $\{x(t)|t = 1, \ldots, N\}$ we will carry out calculations of the corresponding measures of complexity within the framework of the moving window algorithm. For this purpose, the part of the time series (window), for which there were calculated measures of complexity, was selected, then the window was displaced along with the time series in a predefined value, and the procedure repeated until all the studied series had exhausted. Further, comparing the dynamics of the actual time series and the corresponding measures of complexity, we can judge the characteristic changes in the dynamics of the behavior of complexity with changes in the cryptocurrency. If this or that measure of complexity behaves in a definite way for all periods of crashes, for example, decreases or increases during the pre-crashes or pre-critical period, then it can serve as their indicator or precursor.

The returns over some time scale Δt is defined as the forward changes in the logarithm of the corresponding time series: $G(t) \equiv ln\, x\,(t+\Delta t)/\, ln\, x\,(t)$. We will determine standardized returns $g(t) \equiv [G(t) - \langle G \rangle]/\sigma$, where $\sigma \equiv \sqrt{\langle G^2 \rangle - \langle G \rangle^2}$ is the standard deviation of G, and $\langle \dots \rangle$ denotes the average over the time period under study.

In our previous paper [11] we devoted to a comparative analysis complexity of traditional stock market indices and social responsible indices in the example Dow Jones Sustainability Indices and Dow Jones Industrial Average. As measures of complexity, the entropies of various recurrence indicators are chosen – the entropy of the diagonal lines of the recurrence diagram, recurrence probability density entropy and recurrence entropy. It is shown that these measures make it possible to establish that the socially responsive Dow Jones index is more complex. In this paper, we will continue to use econophysical measures of complexity, considering other than recurrent entropy measures, as well as fractal and quantum measures of complexity in relation to the index DJSI.

4 Entropy complexity measures for an index DJSI

The most important quantity that allows us to parameterize complexity in deterministic or random processes is entropy. Originally, it was introduced by Clausius [8], in the context of classical thermodynamics, where according to his definition, entropy tends to increase within an isolated system, forming the generalized second law of thermodynamics. Then, the definition of entropy was extended by Boltzmann and Gibbs [5; 18], linking it to molecular disorder and chaos to make it suitable for statistical mechanics, where they combined the notion of entropy and probability.

After the fundamental paper of Shannon [42] in the context of information theory, where entropy denoted the average amount of information contained in the message, its notion was significantly redefined. After this, it has been evolved along with different ways and successful enough used for the research of economic systems [57].

A huge amount of different methods, as an example, from the theory of complexity, the purpose of which is to quantify the degree of complexity of systems obtained from various sources of nature, can be applied in our study. Such applications have been studied intensively for an economic behavior system.

The existence of patterns within the series is the core in the definition of randomness, so it is appropriate to establish such methods that will be based on the different patterns and their repetition [9]. In this regard, Pincus described the methodology *Approximate* *entropy* (ApEn) [37] to gain more detail analysis of relatively short and noisy time series, particularly, of clinical and psychological. Pincus and Kalman [36], considering both empirical data and models, including composite indices, individual stock prices, the random-walk hypothesis, Black-Sholes, and fractional Brownian motion models to demonstrate the benefits of ApEn applied to the classical econometric modeling apparatus. This research the usefulness of ApEn on the example of three major events of the stock market crash in the US, Japan, and India. During the major crashes, there is significant evidence of a decline of ApEn during and pre-crash periods. Based on the presented results, their research concludes that ApEn can serve as a base for a good trading system. Duan and Stanley [12] showed that it is possible to effectively distinguish the real-world financial time series from random-walk processes by examining changing patterns of volatility, ApEn, and the Hurst exponent. The empirical results prove that financial time series are predictable to some extent and ApEn is a good indicator to characterize the predictable degree of financial time series. Alfonso Delgado-Bonal [10] gives evidence of the usefulness of ApEn. The researcher quantifies the existence of patterns in evolving data series. In general, his results present that degree of predictability increases in times of crisis.

Permutation entropy (PEn), according to the previous approach, is a complexity measure that is related to the original *Shannon entropy* (ShEn) that applied to the distribution of ordinal patterns in time series. Such a tool was proposed by Bandt and Pompe [3], which is characterized by its simplicity, computational speed that does not require some prior knowledge about the system, strongly describes nonlinear chaotic regimes. As an example, Henry and Judge [20] applied PEn to the Dow Jones Industrial Average to extract information from this complex economic system. The result demonstrates the ability of the PEn method to detect the degree of disorder and uncertainty for the specific time that is explored.

4.1 Approximate entropy

When ApEn is calculated, for a given time series $\{x(i)|i = 1, \ldots, N\}$, non-negative embedding parameter d_E, with $d_E \leq N$, and a filter r we construct subsequences $\vec{X}(i)$=$[x(i), x(i+1), \ldots, x(i+d_E-1)]$ and $\vec{X}(j)$=$[x(j), x(j+1), \ldots, x(j+d_E-1)]$. The relative neighborhoods in phase space are measures by L_∞ norm between all pairs of $\vec{X}(i)$ and $\vec{X}(j)$. Then, for each $i = 1, \ldots, N-d_E+1$ we count the number of $j = 1, \ldots, N-d_E+1$ that lie within a suitable distance r and define it as $N_i{}^{d_E}(r)$. For further estimations, we need to define the probability of finding such patters of the length d_E that will be similar to the given pattern:

$$C_i{}^{d_E}(r) = \frac{N_i{}^{d_E}(r)}{(N-d_E+1)},$$

or it can be presented in an equivalent form

$$C_i{}^{d_E}(r) = \frac{1}{N-d_E+1}\sum_{j=1}^{N-d_E+1} \Theta(r - d[\vec{X}(i),\ \vec{X}(j)]),$$

where $\Theta(\cdot)$ is the Heaviside function which counts the instances where the distance d

is below the threshold r.

Next, we define the logarithmic average over all the vectors of the ${C_i}^{d_E}(r)$ probability as

$$F^{d_E}(r) = \frac{1}{(N-d_E+1)} \sum_{i=1}^{N-d_E+1} log({C_i}^{d_E}(r))$$

and ApEn of a corresponding time series is defined as an increment of the absolute entropy $F^{d_E}(r)$ during the transition from a sequence of patterns of length d_E to a sequence of length $d_E + 1$ according to the following formula:

$$ApEn(d_E, r) = F^{d_E}(r) - F^{d_E+1}(r), \tag{1}$$

i.e., equation (1) measures the logarithmic likelihood that sequences of patterns that are close for d_E observations will remain close after further comparisons. Therefore, if the patterns in the sequence remain close to each other (high regularity), the ApEn becomes small, and hence, the time series data has a lower degree of randomness. High values of ApEn indicate randomness and unpredictability. But it should be considered that ApEn results are not always consistent, thus it depends on the value of r and the length of the data series. However, it remains insensitive to noise of magnitude if the values of r and d_E are sufficiently good, and it is robust to artefacts and outliers. Although ApEn remains usable without any models, it also fits naturally into a classical probability and statistics frameworks, and, generally, despite its shortcomings, it is still the applicable indicator of system stability, which significantly increased values may prognosticate the upcoming changes in the dynamics of the data.

The empirical results for the corresponding measure of entropy of DJSI are presented in figure 2:

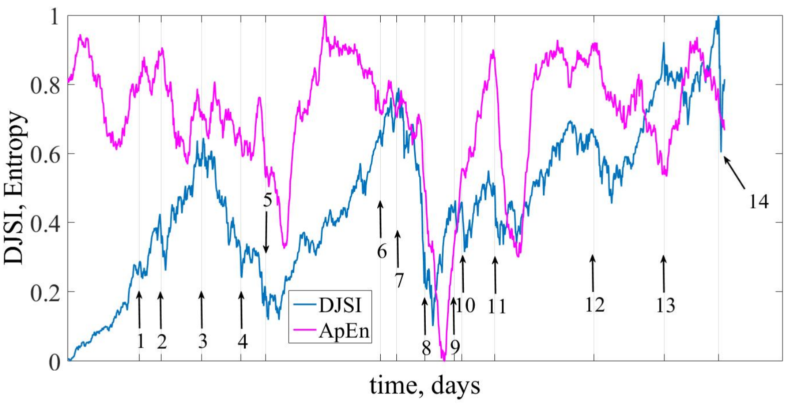


Fig. 2. ApEn dynamics of the entire time series of DJSI.

Long before the crisis, the value of this type of entropy begins to decrease, the complexity of the system decreases. This measure, in our opinion, is one of the earliest precursors of the crisis.

4.2 Permutation entropy

According to this method, we need to consider "ordinal patterns" that consider the order among time series and relative amplitude of values instead of individual values. For evaluating PEn, at first, we need to consider a time series $\{x(i)|i = 1, \dots, n\}$ which relevant details can be "revealed" in d_E-dimensional vector:

$$\vec{X}(i) = [x(i),\ x(i+\tau),\ .\ .\ .\ ,\ x(i+(d_E-1)\tau)],$$

where $i = 1, 2, \dots, N-(d_E-1)\tau$, and τ is an embedding delay of our time delayed vector. By the ordinal pattern, related to the time i, we consider the permutation $\pi_l(i) = (k_0, k_1, \dots, k_{d_E-1})$ of $[0, 1, \dots, d_E-1]$ where $1 \le l \le d_E!$. Then each of the subvectors is arranged in ascending order:

$$x(i+k_0\tau) \le x(i+k_1\tau) \le \cdots \le x(i+k_{d_E-1}\tau).$$

We will use ordinal pattern probability distribution as a basis for entropy estimation. Further, the relative frequencies of permutations in the time series are defined as

$$p(\pi_l) = \frac{the\ number\ of\ patterns\ thathas\ type\ \pi_l}{N-(d_E-1)\tau},$$

where the ordinal pattern probability distribution is given by $P = \{p_l(\pi_l)|l = 1, \dots, d_E!\}$. The Permutation entropy (denoted by $S[P]$) of the corresponding time series presented in the following form:

$$S[P] = -\sum_{l=1}^{d_E!} p_l\, log\, p_l\,.$$

Then, to take more convenient values, we calculate *Normalized permutation entropy* as

$$E_s[P] = \frac{S[P]}{S_{max}}$$

whose $S\ log\ d_{Emax}$ represents the maximum value of $E_s[P]$ (a normalization constant), and normalized entropy restricted between 0 and 1. Here, the maximal entropy value is realized when all $d_E!$ possible permutations are uniformly distributed (more noise and random data). With the much lower entropy value, we get a more predictable and regular sequence of the data. Therefore, the PEn gives a measure of the departure of the time series from a complete noise and stochastic time series.

In figure 3 we can observe the empirical results for permutation entropy, where it serves as indicator-precursor of the possible crashes and critical events.

Information measures of complexity due to their initial validity and transparency, ease of implementation and interpretation of the results occupy a prominent place among the tools for the quantitative analysis of complex systems.

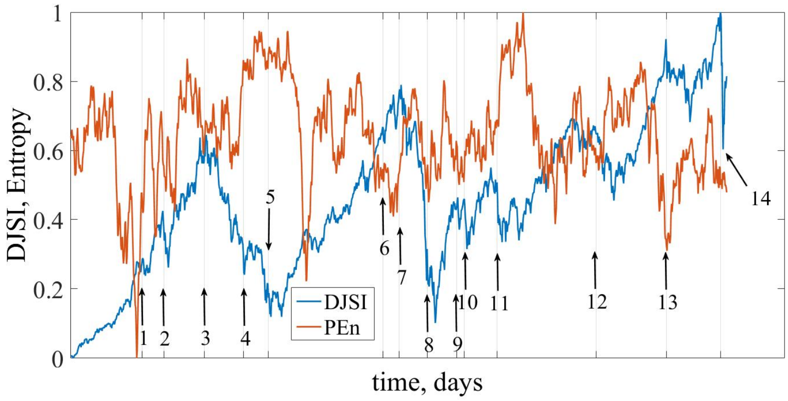


Fig. 3. PEn dynamics of the entire time series of DJSI.

5 Fractal and multifractral measures of complexity

The economic phenomena that cannot be explained by the traditional efficient market hypothesis can be explained by the fractal theory proposed by Mandelbrot [28]. Before, fractal studies focus on the Rescaled Range (R/S) analysis were proposed by Hurst [21] in the field of hydrology. Peng et al. [35] proposed a widely used Detrended Fluctuation Analysis (DFA) that uses a long-range power-law to avoid significant long-range autocorrelation false detection. As a multifractal extension (MF) of the DFA approach, Kantelhardtet et al. [22] introduced the MF-DFA method that for a long time has been successfully applied for a variety of financial markets. An especially interesting application of multifractal analysis is measuring the degree of multifractality of time series, which can be related to the degree of efficiency of financial markets [56].

Similarly to our article [17] where we applied the MF-DFA method to Ukrainian and Russian stock markets, we use it here to explore the multifractal property of DJSI and construct a reliable indicator for it.

As an extension to the original DFA, the multifractal approach estimates the Hurst exponent of a time series at different scales. Based on a given time series $\{x(i)|i = 1, \dots, N\}$, the MF-DFA is described as follows:

1. Construct the profile $Y(i)$ (accumulation) according to the equation below

$$Y(i) = \sum_{j=1}^{i}(g(j) - \langle g \rangle),$$

where $\langle g \rangle$ denotes the average of returns.

2. Then we need to divide the profile $\{Y(i)\}$ into $N_s \equiv int(N/s)$ non-overlapping segments of equal length s, and the local trend Y_v^{fit} for each segment is calculated by the least-square fit. Since time scale s is not always a multiple of the length of the time series, a short period at the end of the profile, which is less than the window size, may be removed. For taking into account the rejected part and, therefore, to use

all the elements of the sequence, the above procedure is repeated starting from the end of the opposite side. Therefore, the total $2N_s$ segments are obtained together, and the variance is computed as

$$F^2(v,\ s) = \frac{1}{s}\sum_{i=1}^{s}[Y((v-1)s+i) - Y_v^{fit}(i)]^2, \quad \text{for } v = 1,\ \dots\ ,\ N_s$$

and

$$F^2(v,\ s) = \frac{1}{s}\sum_{i=1}^{s}[Y(N-(v-N_s)s+i] - Y_v^{fit}(i)]^2, \quad \text{for } v = N_s+1, \dots, 2N_s.$$

Various types of MF-DFA such as linear, quadratic, or higher order polynomials can be used for eliminating local trend in segment v.

3. Considering the variability of time series and the possible multiple scaling properties, we obtain the q-th order fluctuation function by averaging over all segments:

$$F_q(s) = \left[\frac{1}{2N_s}\sum_{v=1}^{2N_s}[F^2(v,\ s)]^{\frac{q}{2}}\right]^{\frac{1}{q}}.$$

The index q can take any non-zero value. For $q = 0$, $F_q(s)$ is divergent and can be replaced by an exponential of a logarithmic sum

$$F_0(s) = exp\left[\frac{1}{4N_S}\sum_{s=1}^{2N_S} ln(\,F^2(v,\ s))\right].$$

4. At least, we determine the scaling behavior of the fluctuation function by analyzing $log\, F_q\,(s)$ vs $log\, s$ graphs for each value of q. Here, $F_q(s)$ is expected to reveal power-law scaling

$$F_q(s) \sim s^{h(q)}.$$

The scaling exponent $h(q)$ can be considered as generalized Hurst exponent. With $q = 2$ MF-DFA transforms into standard DFA, and $h(2)$ is the well-known Hurst exponent.

5. Another way of characterizing multifractality of a time series is in terms of the multifractal scaling exponent $\tau(q)$ which is related to the generalized Hurst exponent $h(q)$ from the standard multifractal formalism and given by

$$\tau(q) = qh(q) - 1. \tag{2}$$

Equation (2) reflects temporal structure of the time series as a function of moments q i.e., it represents the scaling dependence of small fluctuations for negative values of q and large fluctuations for positives values. If (2) represents linear dependence of q, the time series is said to be monofractal. Otherwise, if it has a non-linear dependence on q, then the series is multifractal.

6. The different scalings are better described by the singularity spectrum $f(\alpha)$ which can be defined as:

$$\alpha = \frac{d\tau(q)}{dq} = h(q) + q\frac{dh(q)}{dq},$$
$$f(\alpha) = q[\alpha - h(q)] + 1,$$

with α is the Hölder exponent or singularity strength. Following the methods described above, we present results that reflect multifractal behavior of the DJSI time series.

Fig. 4(a) presents $F_q(s)$ in the log-log plot. The slope changes dependently on q, which indicates the multifractal property of a time series. As it was pointed out, multifractality emerges not only because of temporal correlation, but also because DJSI returns distribution turns out to be broad (fat-tailed), and this distribution could contribute to the multifractality of the time series. The same dependence can be observed in the remaining plots. The scaling exponent $\tau(q)$ remains nonlinear, as well as generalized Hurst exponents that can serve as evidence that Bitcoin exhibit multifractal property.

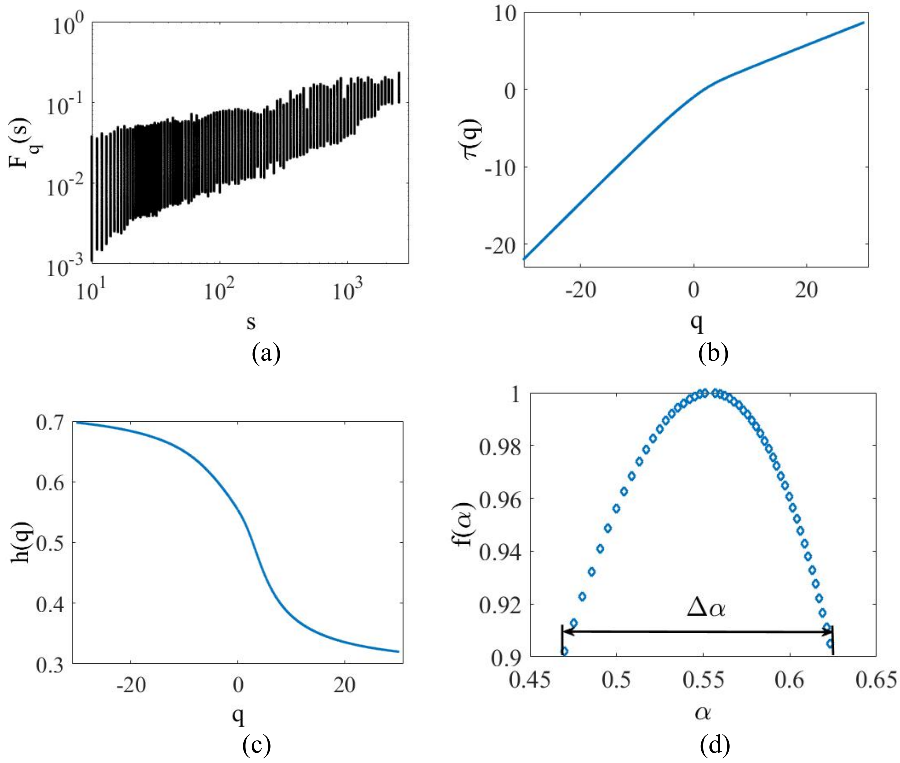


Fig. 4. The fluctuation function $F_q(s)$ (a), multifractal scaling exponent $\tau(q)$ (b), $h(q)$ versus q (c), and singularity spectrum $f(\alpha)$ (d) of the DJSI return time series obtained from MF-DFA.

In the case of multifractals, the shape of the singularity spectrum typically resembles an inverted parabola (see Fig. 4(d)); furthermore, the degree of complexity is straightforwardly quantified by the width of $f(\alpha)$, simply defined as

$\Delta\alpha = \alpha_{max} - \alpha_{min}$, where α_{min} and α_{max} correspond to the opposite ends of the α α values as projected out by different q-moments.

In the figure below we present the width of the spectrum of multifractality that changes over time accordingly to the sliding window approach. The whole figure consists of both a three-dimensional plot (singularity spectrum) and two-dimensional representation of its surface (fig. 5).

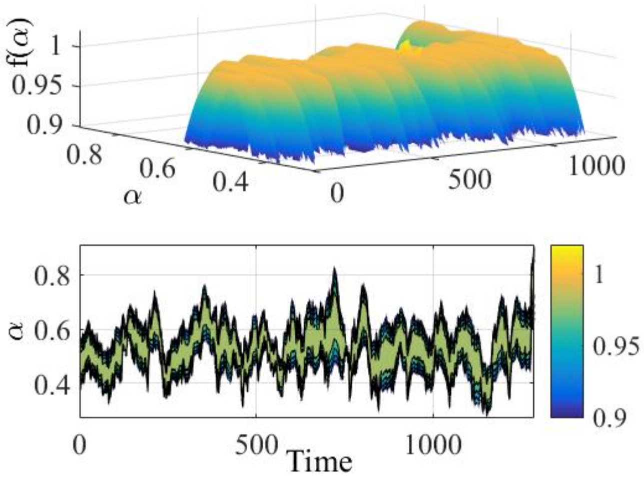


Fig. 5. Changes in the spectrum of multifractality in time.

If the series exhibited a simple monofractal scaling behavior, the value of singularity spectrum $f(\alpha)$ would be a constant. As can be observed, here our series exhibits a simple multifractal scaling behavior, as the value of singularity spectrum $f(\alpha)$ changes dependently on α, i.e., it exhibits different scalings at different scales. Moreover, with the sliding window of the corresponding length, we understand that at different time periods DJSI becomes more or less complex. The value of $\Delta\alpha$ gives a shred of additional evidence on it (fig. 6).

As we can see from the presented results, the width of the singularity spectrum after the crisis starts to increase, which tells us that more violent price fluctuations are usually expected. With the decreasing width of the singularity spectrum, the series is expected to hold the trend. As the rule, it reaches its minimum before the crash of DJSI value.

6 Heisenberg uncertainty principle and economic "mass" as a quantum measure of complexity

In this section, we will demonstrate the possibilities of quantum econophysics on the example of the application of the Heisenberg uncertainty principle [46]. In our paper [41], we have suggested a new paradigm of complex systems modeling based on the ideas of quantum as well as relativistic mechanics. It has been revealed that the use of quantum-mechanical analogies (such as the uncertainty principle, the notion of the operator, and quantum measurement interpretation) can be applied to describing socio-economic processes. Methodological and philosophical analysis of fundamental physical notions and constants, such as time, space, and spatial coordinates, mass, Planck's constant, light velocity from modern theoretical physics provides an opportunity to search for adequate and useful analogs in socio-economic phenomena and processes.

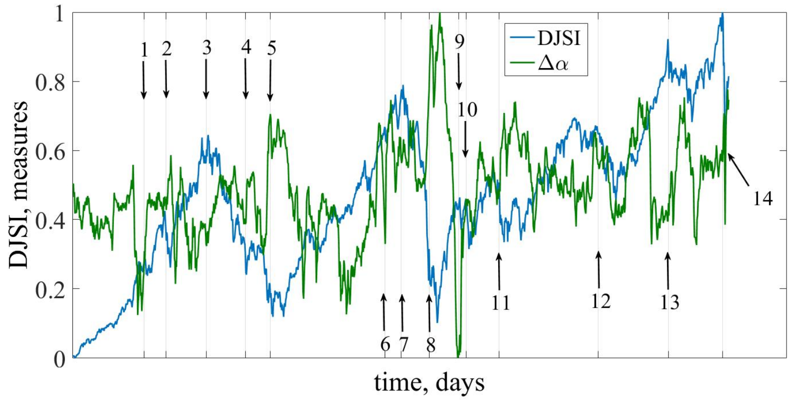


Fig. 6. The comparison of the DJSI time series with the width of the multifractality spectrum measure.

To demonstrate it, let us use the known Heisenberg's uncertainty ratio which is the fundamental consequence of non-relativistic quantum mechanics axioms and appears to be

$$\Delta x \cdot \Delta v \geq \frac{\hbar}{2m_0}, \tag{3}$$

where Δx and Δv are mean square deviations of x coordinate and velocity v corresponding to the particle with (rest) mass m_0, $\hbar$ – Planck's constant. Considering values Δx and Δv to be measurable when their product reaches their minimum, according to equation (3) we derive:

$$m_0 = \frac{\hbar}{2 \cdot \Delta x \cdot \Delta v},$$

i.e., the mass of the particle is conveyed via uncertainties of its coordinate and velocity – time derivative of the same coordinate.

Economic measurements are fundamentally relative, local in time, space and other socio-economic coordinates, and can be carried out via consequent and/or parallel comparisons "here and now," "here and there," "yesterday and today," "a year ago and now," etc.

Due to these reasons constant monitoring, analysis, and time series prediction (time series imply data derived from the dynamics of stock indices, exchange rates, cryptocurrencies prices, spot prices, and other socio-economic indicators) become relevant for the evaluation of the state, tendencies, and perspectives of global, regional, and national economies.

Suppose there is a set of K time series, each of N samples, that correspond to the single distance T, with an equally minimal time step Δt_{min}:

$$X_i(t_n), t_n = \Delta t_{min} n, \text{ for } n = 0, 1, 2, \dots, N-1, \text{ for } i = 1, 2, \dots, K.$$

To bring all series to the unified and non-dimensional representation, accurate to the additive constant, we normalize them, have taken a natural logarithm of each term of the series. Then, consider that every new series $X_i(t_n)$ is a one-dimensional trajectory of a certain fictitious or abstract particle numbered i, while its coordinate is registered after every time span Δt_{min}, and evaluate mean square deviations of its coordinate and speed in some time window $\Delta T = \Delta N \cdot \Delta t_{min} = \Delta N$, $1 << \Delta N << N$. The "immediate" speed of i particle at the moment t_n is defined by the ratio:

$$v_i(t_n) = \frac{x_i(t_{n+1}) - x_i(t_n)}{\Delta t_{min}} = \frac{1}{\Delta t_{min}} ln \frac{X_i(t_{n+1})}{X_i(t_n)},$$

with variance D_{v_i} and mean square deviation Δv_i.

After some transformations, we can write an uncertainty ratio for this trajectory:

$$\frac{1}{\Delta t_{min}} \left(< ln^2 \frac{X_i(t_{n+1})}{X_i(t_n)} >_{n,\Delta N} - \left(< ln \frac{X_i(t_{n+1})}{X_i(t_n)} >_{n,\Delta N} \right)^2 \right) \sim \frac{h}{m_i},$$

where m_i – economic "mass" of an X_i series, h – value which comes as an economic Planck's constant.

Since the analogy with physical particle trajectory is merely formal, h value, unlike the physical Planck's constant $\hbar$, can, generally speaking, depend on the historical period, for which the series are taken, and the length of the averaging interval (e.g., economical processes are different in the time of crisis and recession), on the series number i etc. Whether this analogy is correct or not depends on the particular series' properties.

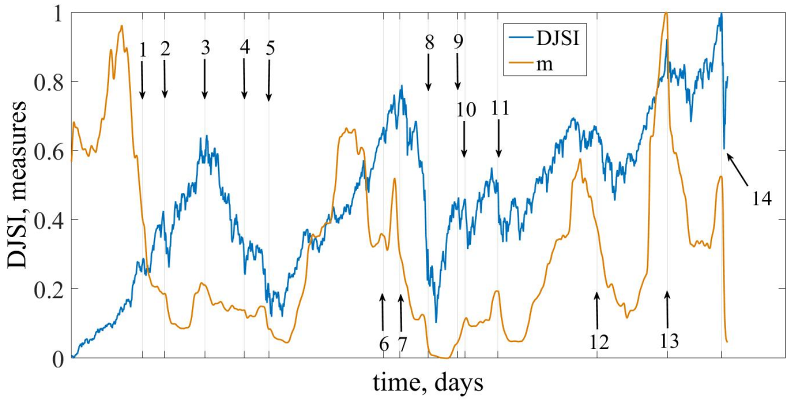


Fig. 7. Dynamics of measure m, and its dynamics with the window size of 250 days and step of 5 days.

Obviously, there is a dynamic characteristic values *m* depending on the internal dynamics of the market. In times of crashes and critical events marked by arrows, mass *m* is significantly reduced in the pre-crash and pre-critical periods (fig. 7). Obviously, *m* remains a good indicator-precursor even in this case. Value *m* is considerably reduced before a special market condition. The market becomes more volatile and prone to changes.

7 Conclusions

In this paper, for the first time, econophysical measures of complexity based on the analysis of entropy, multifractal, and quantum properties of time series are used for the analysis of sustainable development indices. Using the DJSI index as an example, it is shown that, firstly, all econophysical measures are complex measures and, secondly, they respond to critical and crisis conditions of the stock market.

In the future, a similar study for a set of other indices would be of interest, as well as a comparison with the results of using other quantitative measures of complexity.

References

1. Accelerating Green Finance: A report to Government by the Green Finance Taskforce. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/703816/green-finance-taskforce-accelerating-green-finance-report.pdf (2018). Accessed 25 Oct 2018
2. Antonakakis, N., Babalos, V., Kyei, C.K.: Predictability of sustainable investments and the role of uncertainty: Evidence from a non-parametric causality-in-quantiles test. Applied Economics **48**(48), 4655–4665 (2016). doi:10.1080/00036846.2016.1161724
3. Bandt, C., Pompe, B.: Permutation Entropy: A Natural Complexity Measure for Time Series. Phys. Rev. Lett. **88**, 174102 (2002). doi:10.1103/PhysRevLett.88.174102
4. Bielinskyi, A., Soloviev, V., Semerikov, S.: Detecting Stock Crashes Using Levy Distribution. CEUR Workshop Proceedings **2422**, 420–433 (2019)
5. Boltzmann, L.: Weitere Studien über das Wärmegleichgewicht unter Gasmolekülen. Sitzungsberichte Akademie der Wissenschaften **66**, 275–370 (1872). doi:10.1142/9781848161337_0015
6. Castilla-Rubio, J.C., Zadek, S., Robins, N.: Fintech and sustainable development: Assessing the implications. International Environment House, Geneva. http://unepinquiry.org/wp-content/uploads/2016/12/Fintech_and_Sustainable_Development_Assessing_the_Implications.pdf (2016). Accessed 21 Mar 2017
7. Cen, T., He, R.: Fintech, Green Finance and Sustainable Development. Advances in Social Science, Education and Humanities Research **291**, 222–225 (2018). doi:10.2991/meeah-18.2018.40
8. Clausius, R., Hirst, T.A.: The Mechanical Theory of Heat: With Its Applications to the Steam-engine and to the Physical Properties of Bodies. John van Voorst, London (1867)
9. Delgado-Bonal, A., Marshak, A.: Approximate Entropy and Sample Entropy: A Comprehensive Tutorial. Entropy **21**(6), 541 (2019). doi:10.3390/e21060541
10. Delgado-Bonal, A.: Quantifying the randomness of the stock markets. Scientific Reports **9**, 12761 (2019). doi:10.1038/s41598-019-49320-9

11. Derbentsev, V., Semerikov, S., Serdyuk, O., Solovieva, V., Soloviev, V.: Recurrence based entropies for sustainability indices. E3S Web of Conferences **166**, 13031 (2020). doi:10.1051/e3sconf/202016613031
12. Duan, W.-Q., Stanley, H.: Volatility, irregularity, and predictable degree of accumulative return series. Phys. Rev. E 81, 066116 (2010). doi: 10.1103/PhysRevE.81.066116
13. Durand, R., Paugam, L., Stolowy, H.: Do investors actually value sustainability indices? Replication, development, and new evidence on CSR visibility. Strategic Management Journal **40**(9), 1471–1490 (2019). doi:10.1002/smj.3035
14. Escrig-Olmedo, E., Fernández-Izquierdo, M.A., Ferrero-Ferrero, I., Rivera-Lirio, J.M., Muñoz-Torres, M.J.: Rating the Raters: Evaluating how ESG Rating Agencies Integrate Sustainability Principles. Sustainability **11**(3), 915 (2019). doi:10.3390/su11030915
15. Fabregat-Aibar, L., Barberà-Mariné, M.G., Terceño, A., Pié, L.: A Bibliometric and Visualization Analysis of Socially Responsible Funds. Sustainability **11**(9), 2526 (2019). doi:10.3390/su11092526
16. Feng, H., Chen, X., Heck, P., Miao, H.: An Entropy-Perspective Study on the Sustainable Development Potential of Tourism Destination Ecosystem in Dunhuang, China. Sustainability **6**(12), 8980–9006 (2014). doi:10.3390/su6128980
17. Ganchuk, A., Derbentsev, V., Soloviev, V. N. Multifractal Properties of the Ukraine Stock Market. arXiv:physics/0608009v1 [physics.data-an] (2006). Accessed 17 Aug 2020
18. Gibbs, J.W.: Elementary Principles in Statistical Mechanics: Developed with Especial Reference to the Rational Foundation of Thermodynamics. C. Scribner's Sons, New York (1902). doi:10.5962/bhl.title.32624
19. Havrylenko, M., Shiyko, V., Horal, L., Khvostina, I., Yashcheritsyna, N.: Economic and mathematical modeling of industrial enterprise business model financial efficiency estimation. E3S Web of Conference **166**, 13025 (2020). doi:10.1051/e3sconf/202016613025
20. Henri, M., Judge, G.: Permutation Entropy and Information Recovery in Nonlinear Dynamic Economic Time Series. Econometrics **7**(1), 10 (2019). doi:10.3390/econometrics7010010
21. Hurst, H.E.: A Suggested Statistical Model of some Time Series which occur in Nature. Nature **180**, 494 (1957). doi:10.1038/180494a0
22. Kantelhardt, J.W., Zschiegner, S.A., Koscielny-Bunde, E., Havlin, S., Bunde, A., Stanley, H.E.: Multifractal detrended fluctuation analysis of nonstationary time series. Physica A: Statistical Mechanics and its Applications 316(1–4), 87–114 (2002). doi:10.1016/S0378-4371(02)01383-3
23. Khvostina, I., Havadzyn, N., Horal, L., Yurchenko, N.: Emergent Properties Manifestation in the Risk Assessment of Oil and Gas Companies. CEUR Workshop Proceedings **2422**, 157–168 (2019)
24. Kung, O.Y.: "Green Finance for a Sustainable World" - Keynote Speech by Mr Ong Ye Kung, Minister for Education, Singapore and Board Member, Monetary Authority of Singapore, at SFF x SWITCH 2019 on 11 November 2019. https://www.mas.gov.sg/news/speeches/2019/green-finance-for-a-sustainable-world (2019). Accessed 28 Nov 2019
25. Kutner, R., Ausloos, M., Grech, D., Matteo, T.Di., Schinckus, C., Stanley, H.E.: Econophysics and sociophysics: Their milestones & challenges. Physica A: Statistical Mechanics and its Applications **516**, 240–253 (2019). doi:10.1016/j.physa.2018.10.019
26. Liang, X., Si, D., Zhang, X.: Regional Sustainable Development Analysis Based on Information Entropy-Sichuan Province as an Example. International Journal of Environmental Research and Public Health **14**(10), 1219 (2017).

doi:10.3390/ijerph14101219
27. Lucia, U., Grisolia, G: Exergy inefficiency: An indicator for sustainable development analysis. Energy Reports **5**, 62–69 (2019). doi:10.1016/j.egyr.2018.12.001
28. Mandelbrot, B.B.: The Fractal Geometry of Nature. W. H. Freeman and Company, New York (1982)
29. Mayer, A.L., Donovan, R.P., Pawlowski, C.W.: Information and entropy theory for the sustainability of coupled human and natural systems. Ecology and Society **19**(3), 11 (2014). doi:10.5751/ES-06626-190311
30. Mills, M., Wardle, M.: The Global Green Finance Index 4 (2019). doi:10.13140/RG.2.2.28337.33124
31. Miralles-Quiros, M.M., Miralles-Quiros, J.L., Arraiano, I.G.: Sustainable Development, Sustainability Leadership and Firm Valuation: Differences across Europe. Business Strategy and the Environment **26**(7), 1014–1028 (2017). doi:10.1002/bse.1964
32. Mynhardt, H., Makarenko, I., Plastun, A.: Market efficiency of traditional stock market indices and social responsible indices: the role of sustainability reporting. Investment Management and Financial Innovations **14**(2), 94-106 (2017). doi:10.21511/imfi.14(2).2017.09
33. Nassiry, D.: The Role of Fintech in Unlocking Green Finance: Policy Insights for Developing Countries. ADBI Working Paper Series **883**. https://www.adb.org/sites/default/files/publication/464821/adbi-wp883.pdf (2018). Accessed 25 Oct 2019
34. Ohlson, J.A.: Earnings, Book Values, and Dividends in Equity Valuation: An Empirical Perspective. Contemporary Accounting Research **18**(1), 107–120 (2001). doi:10.1506/7TPJ-RXQN-TQC7-FFAE
35. Peng, C.-K., Buldyrev, S.V., Havlin, S., Simons, M., Stanley, H.E., Goldberger, A.L.: Mosaic organization of DNA nucleotides. Physical Review E **49**, 1685 (1993). doi:10.1103/PhysRevE.49.1685
36. Pincus, S., Kalman, R.E.: Irregularity, volatility, risk, and financial market time series. PNAS 101(38), 13709–13714 (2004). doi: 10.1073/pnas.0405168101
37. Pincus, S.M.: Approximate entropy as a measure of system complexity. PNAS **88**(6), 2297–2301 (1991). doi:10.1073/pnas.88.6.2297
38. RobecoSAM: About RobecoSAM. https://www.robecosam.com/en/about-us/about-robecosam.html (2020). Accessed 17 Aug 2020
39. S&P Dow Jones Indices: S&P Global BMI. https://www.spglobal.com/spdji/en/indices/equity/sp-global-bmi (2020). Accessed 17 Aug 2020
40. S&P Global Switzerland: The Sustainability Yearbook 2020. https://www.spglobal.com/esg/csa/yearbook/ (2020). Accessed 17 Aug 2020
41. Saptsin, V., Soloviev, V.: Relativistic quantum econophysics – new paradigms in complex systems modelling. arXiv:0907.1142v1 [physics.soc-ph] (2009). Accessed 17 Aug 2020
42. Shannon, C.E.: A mathematical theory of communication. The Bell System Technical Journal **27**(3), 379–423 (1948). doi:10.1002/j.1538-7305.1948.tb01338.x
43. Śliwiński, P., Łobza, M.: Financial Performance of Socially Responsible Indices. International Journal of Management and Economics **53**(1), 25–46 (2017)
44. Soloviev, V., Belinskij, A.: Methods of nonlinear dynamics and the construction of cryptocurrency crisis phenomena precursors. CEUR Workshop Proceedings **2104**, 116–127 (2018)
45. Soloviev, V., Bielinskyi, A., Solovieva, V.: Entropy Analysis of Crisis Phenomena for DJIA Index. CEUR Workshop Proceedings **2393**, 434–449 (2019)

46. Soloviev, V., Saptsin, V.: Heisenberg uncertainty principle and economic analogues of basic physical quantities. arXiv:1111.5289v1 [physics.gen-ph] (2011). Accessed 17 Aug 2020
47. Soloviev, V., Semerikov, S., Solovieva, V.: Lempel-Ziv Complexity and Crises of Cryptocurrency Market. Advances in Economics, Business and Management Research **129**, 299–306 (2020). doi:10.2991/aebmr.k.200318.037
48. Soloviev, V., Serdiuk, O., Semerikov, S., Kohut-Ferens, O.: Recurrence entropy and financial crashes. Advances in Economics, Business and Management Research **99**, 385–388 (2019). doi:10.2991/mdsmes-19.2019.73
49. Soloviev, V., Solovieva, V., Tuliakova, A., Ivanova, M.: Construction of crisis precursors in multiplex networks. Advances in Economics, Business and Management Research **99**, 361–366 (2019). doi:10.2991/mdsmes-19.2019.68
50. Soloviev, V.N., Belinskiy, A..: Complex Systems Theory and Crashes of Cryptocurrency Market. In: Ermolayev V., Suárez-Figueroa M., Yakovyna V., Mayr H., Nikitchenko M., Spivakovsky A. (eds) Information and Communication Technologies in Education, Research, and Industrial Applications. ICTERI 2018. Communications in Computer and Information Science, vol 1007, pp. 276–297. Springer, Cham (2019)
51. Soloviev, V.N., Bielinskyi, A., Serdyuk, O., Solovieva, V., Semerikov, S.: Lyapunov Exponents as Indicators of the Stock Market Crashes. CEUR Workshop Proceedings (2020, in press)
52. Soloviev, V.N., Yevtushenko, S.P., Batareyev, V.V.: Comparative analysis of the cryptocurrency and the stock markets using the Random Matrix Theory. CEUR Workshop Proceedings **2546**, 87–100 (2019)
53. Sustainalytics: Index Research Services. https://www.sustainalytics.com/index-research-services/ (2020). Accessed 10 Apr 2020
54. Thurner, S., Hanel, R., Klimek, P.: Introduction to the Theory of Complex Systems. Oxford University Press, Oxford (2018)
55. Tillmann, J.: The link between sustainability performance and financial performance – An event study on the impact of sustainability index changes on the market value of a company. Master Thesis, University of Tilburg (2012)
56. Tiwari, A.K., Albulescu, C.T., Yoon, S.-M.: A multifractal detrended fluctuation analysis of financial market efficiency: Comparison using Dow Jones sector ETF indices. Physica A: Statistical Mechanics and its Applications **483**, 182–192 (2017)
57. Tsallis, C.: Introduction to Nonextensive Statistical Mechanics: Approaching a Complex World. Springer, New York (2009)
58. Wang, W., Zhao, X., Gong, Q., Ji, Z.: Measurement of Regional Green Economy sustainable development ability based on Entropy Weight-Topsis-Coupling Coordination Degree – A case study in Shandong Province, China. Sustainability **11**(1), 280 (2019)
59. Wikipedia: Dow Jones Sustainability Indices. https://en.wikipedia.org/w/index.php?title=Dow_Jones_Sustainability_Indices&oldid=974037263 (2020). Accessed 20 Aug 2020
60. Wikipedia: List of stock market crashes and bear markets. https://en.wikipedia.org/w/index.php?title=List_of_stock_market_crashes_and_bear_markets&oldid=984366080 (2020). Accessed 19 Oct 2020
61. Yahoo Finance: Dow Jones Industrial Average (^DJI). https://finance.yahoo.com/quote/%5EDJI?p=^DJI (2020). Accessed 17 Aug 2020
62. Yahoo Finance: Dow Jones Sustainability World (^W1SGI) Charts, Data & News. https://finance.yahoo.com/quote/%5EW1SGI/ (2020). Accessed 17 Aug 2020

Asymptotic methods in optimization of multi-item inventory management model

Lidiia Horoshkova[1,2][0000-0002-7142-4308], Ievgen Khlobystov[3][0000-0002-9983-9062], Volodymyr Volkov[1][0000-0002-1270-895X], Olha Holovan[1][0000-0002-9410-3830], Svitlana Markova[1][0000-0003-0675-0235], Alexander Golub[3][0000-0003-1823-2523] and Oleksandr Oliynyk[1][0000-0003-0511-7681]

[1] Zaporizhzhia National University, 66 Zhukovskogo Str., Zaporizhzhia, 69600, Ukraine
volkovvp49@gmail.com, oaholov@gmail.com, masvvi@outlook.com, a.n.oleynick@gmail.com

[2] Bila Tserkva Institute of Continuous Education of University of Educational Management, 52 Levanevskogo Str., Bila Tserkva, 09108, Ukraine
goroshkova69@gmail.com

[3] National University of “Kyiv-Mohyla academy”, 2 Skovorody Str., Kyiv, 04070, Ukraine
ievgen.khlobystov@ukr.net, agolub@ukma.edu.ua

Abstract. The study proposes asymptotic methods for optimizing multi-item inventory model. To achieve the objective of the study, formulas of the optimal value of multi-item delivery frequency based on the asymptotic approach under conditions of minor changes in the input parameters have been obtained. The discrete increase in the execution costs and inventory holding costs which depend on the “small parameter” as well as a gradual increase in periodic fluctuations in demand for products have been taken as variable parameters of the system. Easy-to-use analytical formulas for determining optimal order interval when ordering and inventory holding costs, as well as demand meet insufficient changes have been obtained. Testing of the proposed approach to the multi-item inventory model has been carried out on the example of HoReCa regional market segment. The proposed formulas allow to apply the obtained results for optimization and forecasting of decision-making in the system of procurement logistics of a company amid variation of input parameters describing changes of external and internal business environment.

Keywords: multi-item order, optimal order, delivery interval, small parameter, asymptotic methods.

1 Introducation

Sharpening of international and local competition leads to higher requirements for business processes’ competitiveness and efficiency. These could be achieved through the optimization of product, pricing, sales, innovation policy, as well as promotion policy. Therefore, forecasting and rational planning of all the subsystems of business process management when minimizing operating and sales costs, strengthening cooperation and coordination between company's structural units become relevant.

Solving these problems is not possible without the reduction of company's inventory management costs. Excess stocks are frozen current assets, which do not bring additional benefits, but also require additional costs for their maintenance, holding, ensuring their appropriate quality.

The modern inventory management system is characterized by the introduction of the integrated approach to the inventory management within the internal logistics system, which is subject to the business strategy and provides the ability to determine the optimal inventory based on the demand forecast and changes in external and internal environment.

There are inventory management models and methods in the logistics system of an enterprise and corresponding software products that improve inventory management quality and efficiency in the market. Current decision-making support systems fulfill mainly the accounting functions, which significantly limits their effectiveness for companies' management. To further improve these software products and ensure the possibility of their implementation to forecast various business activities, it is necessary to develop analytical tools for the inventory management system.

Asymptotic methods and approaches that allow to obtain analytical solutions of the applied problems, to assess sensitivity of these solutions to variations of system input parameters aimed at forecasting business activities are promising for the improvement of the analytical tools.

2 Literature review

Studies [1; 13] contain the main ideas of the asymptotic approach and perturbation techniques; they put their outlined features, advantages and limits of their application in various fields into layman's terms.

[7] is devoted to the technique of the asymptotic developments' application in a series by small parameter degrees. This parameter is selected either artificially or emerges naturally in the model. The proposed technique assumes that the solution is sought in the form of the asymptotic sequence, which is taken as power function of a small parameter ε. Perturbation techniques are widely used in various natural science problems: solid mechanics and mathematics, in particular when solving differential equations. These methods are chosen as they allow to obtain an easy-to-use analytical formula to analyze system sensitivity to variable input parameters.

The scope of perturbation techniques can be extended to solving economic problems, modelling of relevant processes, managerial decision-making in general and one of the topical issues – inventory management issue in particular.

[10; 15] studies analyze features of the procurement logistics' deterministic models building, which help managers in the decision-making process. The EOQ-model of the procurement logistics is made in [8] taking into account the shortage due to the possible defective products. The proposed model is limited by the minor discrete changes in execution costs. Implementation of the asymptotic methods to determine the order quantity of a single-product inventory management problem amid insignificant discrete changes in execution costs is proposed in [6]. Analytical tools improvement using the asymptotic approach and taking into account the periodicity in the demand for products is implemented in [14]. However, the authors did not take into account the variability of storage costs. Further research referring to the application of asymptotic methods in the inventory management problems has been made in [2]. The authors proposed the model that considers execution and holding costs' small gradual changes, as well as small periodic fluctuations in demand for products. However, the asymptotic approach was applied to the single-product inventory management problem.

In [22], a dynamic model of inventory management with the shift of the deficit of demand, which changes linearly, is built. The proposed model is aimed at determining the main efficiency indicators of the order management system with constant input parameters.

[12] studies the system of inventory planning for the industries producing nondurable goods. For small and medium-sized businesses in the food industry, the main issue is the compliance with the expiration dates of the proposed products. Therefore, the issue of reducing the amount of products, which past the expiry date, determining the optimal amount of products and time of multi-product order is especially relevant. Nevertheless, the authors use fixed parameters to determine the optimal order quantity and order period. Change in demand and inventory damage due to the sub-optimal inventory location or improper storage conditions are studied in [23]. Besides, the authors of the study consider the case when the inventory damage rate is distributed by the Weibull function, and inventory holding costs are discrete. Purchases of perishable goods are modelled in [3; 11] under the conditions of possible payment delays and inflation, but these models do not take into account possible fluctuations in demand for products. The study [24] concerned the integrated EOQ model building for the supplier`s and buyer`s monopoly who operate in the monopsony market diminishing ordering costs, but keeping inventory costs unchanged.

Researchers are trying to adapt the proposed models to the situations faced by companies who implement real production procedures in the logistics management system. For instance: fluctuations of demand, supply, component prices, inflation, consumer expectations, etc. [19] proposed the inventory management model in retail, which allows to maximize profit in the reverse logistics system, considering the level of supply and the term of delivery. Unlike most studies, [17] takes into account transport costs to obtain the optimal order, but the researchers propose the iterative approach, which is difficult to apply. Study [21] is another adaptation of the inventory management model considering spatial and temperature constraints on storage equipment. However, the results do not take into account changes in the execution and inventory costs.

The EOQ problem was studied in [26] for the case of the stochastic nature of input parameters and calculation of probability distribution using the geometric programming model. The stochastic problem of finding the optimal order quantity on a time interval is solved in [9]. However, the studies are mainly theoretical ones; there is lack of analytical formulas convenient for application and further analysis.

In [5] the multi-item problem of the inventory management amid insignificant changes of execution costs based on asymptotic approach was solved. Nonetheless, specification of the optimal order did not take into consideration the variability of inventory costs and demand.

The impact of the demand sensitive to marketing incentives on the solution of the multi-item EOQ-model problem for the two-level supply chain under the conditions of payment delay of outgoing inventory spending was considered in [4]. In [25] the multi-item inventory management model within limited warehouse space and application of the quantitative discount system was studied. The authors determined the optimal order applying Lagrange multiplier technique and the dynamic programming method, which is close to the Wagner-Within algorithm. These methods were compared to find the best solution to determine the number of orders. However, the authors use already known methods and algorithms to solve the applied problem.

In [20] genetic algorithm based on total inventory cost minimization was used to determine the quantity of raw materials' multi-product batch. Nonetheless, inventory holding costs were based on the warehouse space during a period and unit`s space dimension.

The study [16] was devoted to the EOQ-model building, which is not without limitations. In particular, there are assumptions like deteriorating inventory quality, shortages, inventory space availability and the modesty of overall budget for purchasing of goods. [18] study proposed deficit-free multi-product EOQ model with inaccurate limitations for fragile goods. However, the proposed models do not take into account the variability of input parameters.

The analysis of well-known researchers' works in the field of inventory management revealed that the proposed models allow to determine key inventory management system parameters, namely: order quantity, time interval between orders, etc. Still, there are difficulties in their application due to the complex mathematical apparatus and the impossibility of obtaining analytical formulas, convenient for the model behavior prediction with changing input parameters.

3 Research tasks

The study objective is optimization of inventory management multi-item model amid insignificant changes of input parameters using asymptotic methods. To achieve this objective, the following tasks were set:

1. To obtain the asymptotic formula of the multi-item inventory management model with a slight discrete spending spree on order execution and inventory holding;
2. To obtain the asymptotic formula of the multi-item inventory management model with variable order execution costs and small fluctuations in the amplitude of demand.

4 Results

4.1 Asymptotic expansion of the multi-item inventory management problem solution with a slight discrete spending spree on order execution and inventory holding

Inventory management model`s application allows company's management to reduce fixed and variable production costs, order and sales costs. All costs related to the multi-item resources or goods ordering from one supplier are presented in the form of two components:

C_0 – Costs formed when transporting,

C_i – Costs depending on the activities when a specific order is being formed. Thus, order execution costs k of product units y of one supplier are presented as (1).

$$C_{order} = C_0 + \sum_{i=1}^{k} C_i = \sum_{i=0}^{k} C_i \quad (1)$$

In case of simultaneous delivery of k product groups, total costs are presented as (2) [10].

$$C_{\Sigma} = \frac{D}{T}\sum_{i=0}^{k} C_i + \frac{T}{2D}\sum_{i=1}^{k} S_i Cx_i \to min, \quad (2)$$

where S_i – total consumption of i product during period, Cx_i – inventory costs for one unit produced, T – delivery frequency, D – duration.

Under the conditions of model parameters with fixed values (2), the optimal value of multi-item delivery T_{opt} frequency is determined as follows [10]:

$$T_{opt} = D\sqrt{\frac{2\sum_{i=0}^{k} C_i}{\sum_{i=1}^{k} S_i C_{x_i}}} \quad (3)$$

The assumption of model's fixed parameters reduces possibility and effectiveness of its practical application. The asymptotic methods allow, without violating this condition, to perturb model's parameters. Order execution costs are one of the following parameters fixed in the inventory management optimization model. In practice, this parameter may increase with certain repetitions caused by inflation, consumer and producer expectations, the "ratchet" effect, etc. To build the model it was assumed that during the specified time order execution costs, namely their transport component, rise repeatedly by l% and in n periods become $C_0 \cdot \left(1 + \frac{l\%}{100\%}\right)^n$. By introducing a small perturbation parameter, this correlation was presented as follows: $C_0 \cdot (1 + \varepsilon)^n$, when $\varepsilon \ll 1$.

Practically, order execution costs and inventory holding costs rise due to the price boost for electricity and utilities. It was assumed in the study, that inventory holding costs surge every period by j%. Analogously, taking value $\beta = \frac{j\%}{100\%}$ ($\beta \ll 1$) as a small parameter, it was obtained dependence of inventory as $C_{xi}(1 + \beta)^m$.

To improve the efficiency of the proposed model for manufacturing, it is advisable to take into account different combinations of parameters n and m, ε and β. The multiplicity *m*, which characterizes the frequency of inventory holding costs changes, is lower than the multiplicity *n*, which characterizes order execution cost changes. In the proposed model, the conditions were taken into account through definition of dependence's general form, for example, taking $m=[n/2]$, $m=[n/3]$, etc., where [ ] – the quotient.

$T^*{}_{opt}$ was presented as an asymptotic expansion of two small parameters ε and β:

$$T^*{}_{opt} = T_0 + T_1\varepsilon + T_2\beta + T_3\varepsilon^2 + T_4\varepsilon\beta + T_5\beta^2 + \ldots \tag{4}$$

where ε and β – perturbation parameters.

Substituting the perturbation values of execution and inventory holding costs' transport component, expansion of $T^*{}_{opt}$ by small parameters degrees ε and β in formula (3) and neglect terms ε^3, β^3, $\varepsilon^2\beta$, $\varepsilon\beta^2$ and above, allowed to obtain:

$$(T_0 + T_1\varepsilon + T_2\beta + T_3\varepsilon^2 + T_4\varepsilon\beta + T_5\beta^2)^2 = D^2\frac{\left(2C_0(1+\varepsilon)^n + 2\sum_{i=1}^{k} C_i\right)}{(1+\beta)^m \sum_{i=1}^{k} S_i C_{xi}}. \tag{5}$$

Function expansions $(1+\varepsilon)^n$ and $(1+\beta)^{-m}$ in the Taylor series, neglect of the higher order terms and raising square of both parts of equation (5) led to:

$$T_0^2 + 2T_0T_1\varepsilon + 2T_0T_2\beta + (T_1^2 + 2T_0T_3)\varepsilon^2 + (2T_0T_4 + 2T_1T_2)\varepsilon\beta +$$
$$+\left({T_2}^2 + 2T_0T_5\right)\beta^2 = D^2\frac{\left(2C_0\left(1+n\varepsilon+\frac{n(n-1)}{2}\varepsilon^2\right)+2\sum_{i=1}^{k} C_i\right)\cdot\left(1-m\beta+\frac{m(m+1)}{2}\beta^2\right)}{\sum_{i=1}^{k} S_i C_{xi}} \tag{6}$$

Comparison of the coefficients with the same parameter's degrees ε and β, allowed to obtain:

$$T_0 = D\sqrt{\frac{2\sum_{i=0}^{k} C_i}{\sum_{i=1}^{k} S_i C_{x_i}}},\ T_1 = D\sqrt{\frac{2\sum_{i=0}^{k} C_i}{\sum_{i=1}^{k} S_i C_{x_i}}}\cdot\frac{C_0 n}{2\sum_{i=0}^{k} C_i},\ T_2 = D\sqrt{\frac{2\sum_{i=0}^{k} C_i}{\sum_{i=1}^{k} S_i C_{x_i}}}\cdot\left(\frac{-m}{2}\right),$$
$$T_3 = D\sqrt{\frac{2\sum_{i=0}^{k} C_i}{\sum_{i=1}^{k} S_i C_{x_i}}}\cdot\left(\frac{C_0 n(n-1)}{4\sum_{i=0}^{k} C_i} - \frac{C_0^2 n^2}{8\left(\sum_{i=0}^{k} C_i\right)^2}\right),\ T_4 = D\sqrt{\frac{2\sum_{i=0}^{k} C_i}{\sum_{i=1}^{k} S_i C_{x_i}}}\cdot\frac{C_0(-mn)}{4\sum_{i=0}^{k} C_i},$$
$$T_5 = D\sqrt{\frac{2\sum_{i=0}^{k} C_i}{\sum_{i=1}^{k} S_i C_{x_i}}}\cdot\frac{m(m+2)}{8}. \tag{7}$$

Asymptotic representation of the optimum value of multiitem delivery cycle T^*_{opt} took the form (8) and (9):

$$T^*_{opt} = D\sqrt{\frac{2\sum_{i=0}^{k} C_i}{\sum_{i=1}^{k} S_i C_{x_i}}}\cdot$$
$$\cdot\left(1 + \frac{C_0 n}{2\sum_{i=0}^{k} C_i}\varepsilon - \frac{m}{2}\beta + \left(\frac{C_0 n(n-1)}{4\sum_{i=0}^{k} C_i} - \frac{C_0^2 n^2}{8\left(\sum_{i=0}^{k} C_i\right)^2}\right)\cdot\varepsilon^2 - \frac{C_0 nm}{2\sum_{i=0}^{k} C_i}\varepsilon\beta + \frac{m(m+2)}{8}\beta^2\right), \tag{8}$$

$$T_{opt}^{*} = T_{opt} \cdot$$
$$\cdot \left(1 + \frac{C_0 n}{2\sum_{i=0}^{k} C_i}\varepsilon - \frac{m}{2}\beta + \left(\frac{C_0 n(n-1)}{4\sum_{i=0}^{k} C_i} - \frac{C_0^2 n^2}{8\left(\sum_{i=0}^{k} C_i\right)^2}\right) \cdot \varepsilon^2 - \frac{C_0 nm}{4\sum_{i=0}^{k} C_i}\varepsilon\beta + \frac{m(m+2)}{8}\beta^2\right). \quad (9)$$

If ε=0 and β=0 one can obtain transformation to “not perturbed” solution in the form (3).

The built multi-item model of inventory management optimization under “perturbed” parameters was tested on the example of an enterprise operating in HoReCa market of Zaporizhzhia. Output data and calculation of “not perturbed” system parameters are given in table 1.

Table 1. Output data and calculation results of a multi-item delivery of HoReCa market (Zaporizhzhia).

Type of products	Annual demand S_i, units	Inventory holding costs, Cx_i, MU.	Execution costs, MU.		Optimal value of delivery frequency T_{opt}, full days	Order quantity, $q_i = T_{opt}\frac{S_i}{D}$, units
			C_0	C_i		
Tea (10 sachets)	900	1,5	70	2	84	207
Whole bean coffee (1 kg)	270	5,5	70	3		62
Sugar (800 sticks)	95	1,0	70	3		22

“Perturbed” values of a multi-item delivery T_{opt}^{*}, corresponding output data of table 1, as well as its relation with “not perturbed” parameter between orders ($T^{*}{}_{opt}/T_{opt}$) for different values of ε, n and β are presented in tables 2 and 3. There is comparative analysis of optimal cycle’s values of a multi-item delivery of HoReCa (Zaporizhzhia) when $m = \left[\frac{n}{6}\right]$ in table 2.

Data analysis presented in table 2 shows that, for example, with a gradual ramp up of order execution costs’ transport component by 1% (ε = 0,01) during the first 6 periods (n = 5, m = 0) and fixed inventory holding costs, the order interval rises by 2,3%. A one-time increase in holding costs (n = 6 and m = 1) leads to decrease of the studied parameter by 2%. Further period extension induces the recurrence of the percentage change in order intervals. Growth of order execution costs’ transport component by 2% (ε = 0,02) during the first 6 periods (n = 5, m = 0) and fixed inventory holding costs come amid order interval boosted by 4,6%. A one-time increase in holding costs (n = 6, m = 1) leads to a decrease in the studied indicator by 1,7% compared to the previous value.

Fig. 1 shows correlation between the dynamics of order quantity change for the type of product – tea with 5% (β = 0,05) rate of holding costs change and different rates of order execution costs change.

Table 2. Comparative analysis of optimal cycle's values of a multi-item delivery, β=0,05, $m = \left[\frac{n}{6}\right]$.

Period		ε = 0,01		ε = 0,015		ε = 0,02	
n	m	$T^*{}_{opt}/T_{opt}$	$T^*{}_{opt}$, full days	$T^*{}_{opt}/T_{opt}$	$T^*{}_{opt}$, full days	$T^*{}_{opt}/T_{opt}$	$T^*{}_{opt}$, full days
0	0	1,0000	84,00	1,0000	84,00	1,0000	84,00
1	0	1,0045	84,38	1,0067	84,56	1,0089	84,75
2	0	1,0090	84,75	1,0135	85,13	1,0180	85,51
3	0	1,0135	85,13	1,0203	85,70	1,0271	86,28
4	0	1,0181	85,52	1,0272	86,28	1,0363	87,05
5	0	1,0226	85,90	1,0341	86,86	1,0457	87,84
6	1	1,0018	84,15	1,0150	85,26	1,0283	86,38
7	1	1,0062	84,52	1,0217	85,82	1,0374	87,14
8	1	1,0107	84,89	1,0285	86,39	1,0466	87,91
9	1	1,0151	85,27	1,0353	86,96	1,0559	88,69
10	1	1,0196	85,64	1,0422	87,54	1,0652	89,48
11	1	1,0241	86,02	1,0491	88,12	1,0747	90,28
12	2	1,0025	84,21	1,0286	86,40	1,0555	88,66
13	2	1,0068	84,57	1,0353	86,96	1,0647	89,43
14	2	1,0111	84,94	1,0421	87,53	1,0740	90,22
15	2	1,0155	85,30	1,0489	88,11	1,0834	91,01
16	2	1,0199	85,67	1,0557	88,68	1,0930	91,81
17	2	1,0243	86,05	1,0627	89,26	1,1026	92,62
18	3	1,0019	84,16	1,0408	87,43	1,0814	90,84
19	3	1,0062	84,52	1,0475	87,99	1,0908	91,63
20	3	1,0105	84,88	1,0542	88,56	1,1002	92,42
21	3	1,0148	85,24	1,0610	89,13	1,1098	93,22
22	3	1,0191	85,60	1,0679	89,70	1,1195	94,04
23	3	1,0234	85,97	1,0748	90,29	1,1292	94,86
24	4	1,0002	84,02	1,0515	88,33	1,1062	92,92

Thus, with a gradual ramp up of order execution costs by 1,5% (ε=0,015) and 2% (ε=0,02) there is "perturbed" order quantity growth compared to the optimal (table 1) by 3,4% and 4,8% respectively (when n =5); by 4,8% and 7,7% (when n =11) and by 6,3% and 10,1% (when n =17). Periods which come amid holding costs shift (n = 6, m = 1), (n = 12, m = 2), etc. and gradual order execution costs surge by 2% (ε=0,02) face lower order quantity compared to the previous value by 1,9% respectively when n =6; by 1,8% when n = 12 and when n = 18.

Optimal periods of a multi-item delivery of HoReCa segment in Zaporizhzhia when β=0,05 and $m = \left[\frac{n}{12}\right]$ are calculated in table 3.

Calculations given in table 3 reveal that at ε = 0.02 gradual execution costs change of a multi-item delivery leads to higher order interval by 0.9% at average at all intervals n, except for those periods when there is holding costs change. The change in inventory holding costs at n = 12, m = 1 and n = 24, m = 2 leads to lower order interval by 1.8% and 2%, respectively, relative to the previous value.

Table 3. Calculation of optimal cycle`s values of a multi-item delivery of HoReCa (Zaporizhzhia), β=0,05, $m = \left[\frac{n}{12}\right]$.

Period		ε = 0,01		ε = 0,015		ε = 0,02	
n	m	T^*_{opt}/T_{opt}	T^*_{opt}, full days	T^*_{opt}/T_{opt}	T^*_{opt}, full days	T^*_{opt}/T_{opt}	T^*_{opt}, full days
0	0	1,0000	84,00	1,0000	84,00	1,0000	84,00
1	0	1,0045	84,38	1,0067	84,56	1,0089	84,75
2	0	1,0090	84,75	1,0135	85,13	1,0180	85,51
3	0	1,0135	85,13	1,0203	85,70	1,0271	86,28
4	0	1,0181	85,52	1,0272	86,28	1,0363	87,05
5	0	1,0226	85,90	1,0341	86,86	1,0457	87,84
6	0	1,0272	86,29	1,0411	87,45	1,0551	88,63
7	0	1,0319	86,68	1,0481	88,04	1,0646	89,43
8	0	1,0365	87,07	1,0552	88,64	1,0742	90,24
9	0	1,0412	87,46	1,0624	89,24	1,0840	91,05
10	0	1,0459	87,85	1,0696	89,85	1,0938	91,88
11	0	1,0506	88,25	1,0769	90,46	1,1037	92,71
12	1	1,0286	86,40	1,0561	88,71	1,0843	91,08
13	1	1,0332	86,78	1,0631	89,30	1,0940	91,89
14	1	1,0377	87,17	1,0702	89,90	1,1037	92,71
15	1	1,0423	87,56	1,0774	90,50	1,1136	93,54
16	1	1,0469	87,94	1,0846	91,10	1,1236	94,38
17	1	1,0516	88,33	1,0918	91,71	1,1336	95,23
18	1	1,0563	88,73	1,0991	92,33	1,1438	96,08
19	1	1,0610	89,12	1,1065	92,95	1,1541	96,94
20	1	1,0657	89,52	1,1139	93,57	1,1644	97,81
21	1	1,0704	89,92	1,1214	94,20	1,1749	98,69
22	1	1,0752	90,32	1,1290	94,83	1,1855	99,58
23	1	1,0800	90,72	1,1366	95,47	1,1961	100,48
24	2	1,0560	88,70	1,1127	93,47	1,1727	98,51

Comparing the results presented in table 2 and table 3, it can be concluded that the increasing multiplicity of holding costs changes for products (m) in a multi-item delivery causes order interval contraction relative to the optimal. For example, higher holding costs (n = 24, m = 2) at ε = 0,02, β = 0,05 and when (n = 24, m = 4) come amid the contraction of the order interval from 98,51 to 92,92 full days or by 6%, which is quite significant.

Fig. 2 illustrates correlation between tea order quantity and parameters values ε and n.

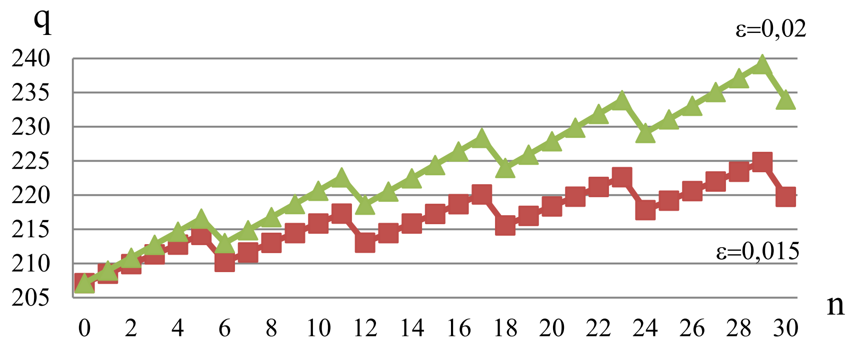


Fig. 1. Correlation between tea order quantity and parameters values *ε* and *n* when *β*=0,05, $m = \left[\frac{n}{6}\right]$.

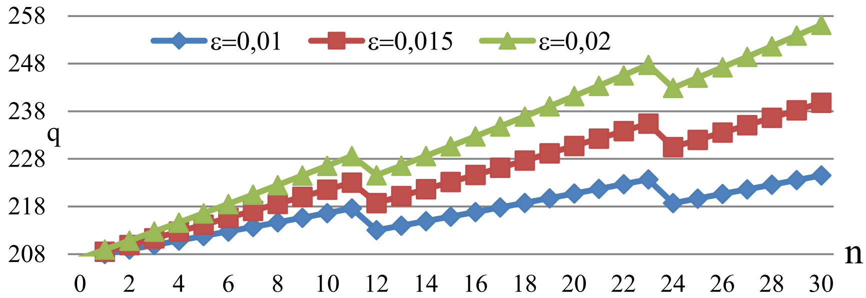


Fig. 2. Correlation between tea order quantity and parameters values *ε* and *n* at *β*=0,05, $m = \left[\frac{n}{12}\right]$.

The calculation revealed that n growth, which characterizes the frequency of order execution costs boost, to the value $n = 11$ at $\varepsilon = 0.02$ causes the increase in the order quantity to 10.4% relative to the optimal value. Shift of m ($m = 1$) leads to the decrease in order quantity by 1.8% compared to the previous value at $\varepsilon = 0.02$, and at ($m = 2$) leads to lower order quantity by 2.3%.

Figure 3 illustrates correlation of company's total costs, operating in HoReCa segment in Zaporizhzhia from multiplicity of order and holding costs growth (n and $m = \left[\frac{n}{6}\right]$).

One can see that higher order execution costs growth rate (fig. 3) causes total cost boost of a multi-item delivery. For example, parameter ε growth from 0,01 to 0,02 or by 1% leads to total costs boost by 3% at ($n = 6$, $m = 1$). Thenceforward, total costs growth takes place. Moreover, parameter's ε impact becomes more significant. For example, at $n = 12$, $m = 2$ the difference between total costs values at $\varepsilon = 0{,}01$ and $\varepsilon = 0{,}02$ is 5,55%.

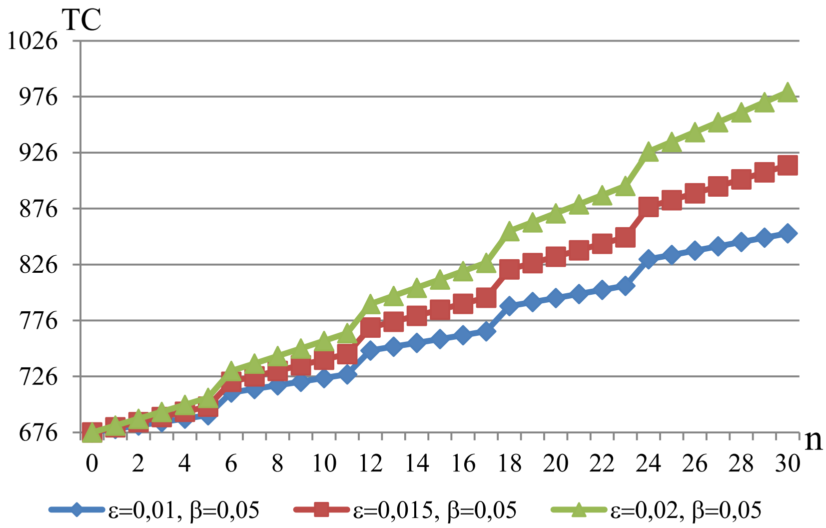


Fig. 3. Correlation of customer's total costs change to parameters values ε and β at $m = \left[\frac{n}{6}\right]$.

Fig. 4 illustrates shift of company's total costs from execution costs in case $m = \left[\frac{n}{12}\right]$, i.e. inventory holding costs alter once in 12 periods.

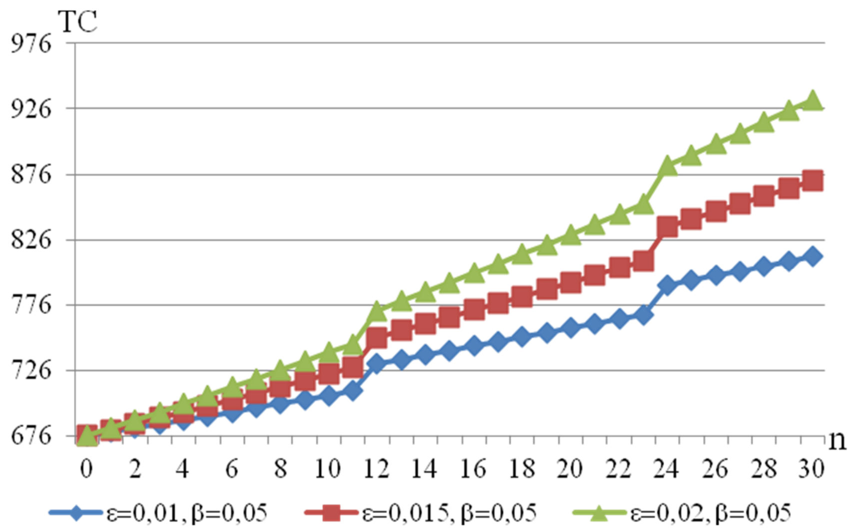


Fig. 4. Correlation of company's total costs' change to parameters' values ε and β at $m = \left[\frac{n}{12}\right]$.

Comparing the results demonstrated in fig. 3 and fig. 4, one can conclude that the contraction of multiplicity of holding costs changes m causes total costs reduction. For example, at $\varepsilon = 0{,}01$ and $n = 6$, $m = 1$ (fig. 3) total costs are UAH 711,64, and at ε = 0,01 and $n = 6$, $m = 0$ (fig. 4) they are UAH 694,49, which is 2,4% less. At $\varepsilon = 0.02$ and $n = 6$, $m = 1$ (fig. 3) total costs are UAH 730,95, and at $\varepsilon = 0{,}02$ and $n = 6$, $m = 0$ (fig. 4) they are UAH 713,33, which is 2,5% less. Thenceforward, this discrepancy is growing.

Thus, one can conclude that the higher periods' multiplicity of ordering and holding costs of products (n and m) is, then period`s, order quantities of a multi-item delivery and execution costs' cyclical fluctuations take place.

4.2 Solution of the multi-item inventory management problem with variable costs of order execution and small fluctuations in the amplitude of demand

The practical application of the studied model assumes that not only order execution costs change, but also the seasonal demand for some goods. It was assumed in the study that the demand for products gradually increases with seasonal fluctuations. Demand dependence was chosen in the form $S_0\left(1-\alpha\sin\frac{\pi m}{2}\right)e^{\gamma m}$, where the parameter m affects the period of the demand change.

Selection of parameters' γ and α values is determined by the nature of the shift in demand for products. Thus, the parameter γ sets the average growth rate of demand for a certain period (for example, the average annual growth rate). The parameter α determines the amplitude of seasonal fluctuations in demand for products.

This study provides analytical formulas that can be applied by companies' management in HoReCa sector to determine quantity and period of a multi-item delivery for the categories of products, such as coffee and tea, characterized by both seasonal demand and the growing one in Ukraine and around the world.

Fig. 5 demonstrates the chosen form of demand dependence on the average growth rates γ and the amplitude of seasonal fluctuations α. In fig. 5 one can see the nature of the demand amplitude shift for seasonal products. The type of function can be explained by the decline in demand for tea and coffee in the spring and the summer, and its gradual increase in the autumn and the winter. In addition, this dependence makes it possible to take into account the trend of constant demand growth by choosing different values of the parameter γ. Variation of the parameter α allows to analyze model's behaviour amid changing amplitude of demand fluctuations.

The desired optimal interval between orders $T^*{}_{opt}$ a for multi-item delivery is presented as an asymptotic expansion of two small parameters ε and γ:

$$T^*{}_{opt} = T_0 + T_1\varepsilon + T_2\gamma + T_3\varepsilon^2 + T_4\varepsilon\gamma + T_5\gamma^2 + \ldots, \qquad (10)$$

where ε, γ – perturbation parameters.

Substituting perturbed values of order execution costs $C_0 \cdot (1+\varepsilon)^n$, demand for products $S_i\left(1-\alpha\,sin\frac{\pi m}{2}\right)e^{\gamma m}$, and expansion T^*_{opt} (11) by small parameter degrees ε and γ into formula (3) and neglecting terms ε^3, γ^3, $\varepsilon^2\gamma$, $\varepsilon\gamma^2$ and above, it was obtained:

$$(T_0 + T_1\varepsilon + T_2\gamma + T_3\varepsilon^2 + T_4\varepsilon\gamma + T_5\gamma^2)^2 = D^2 \frac{\left(2C_0(1+\varepsilon)^n + 2\sum_{i=1}^{k} C_i\right)}{\left(1-\alpha \sin\frac{\pi m}{2}\right)\cdot e^{\gamma m} \sum_{i=1}^{k} S_i C_{xi}}. \quad (11)$$

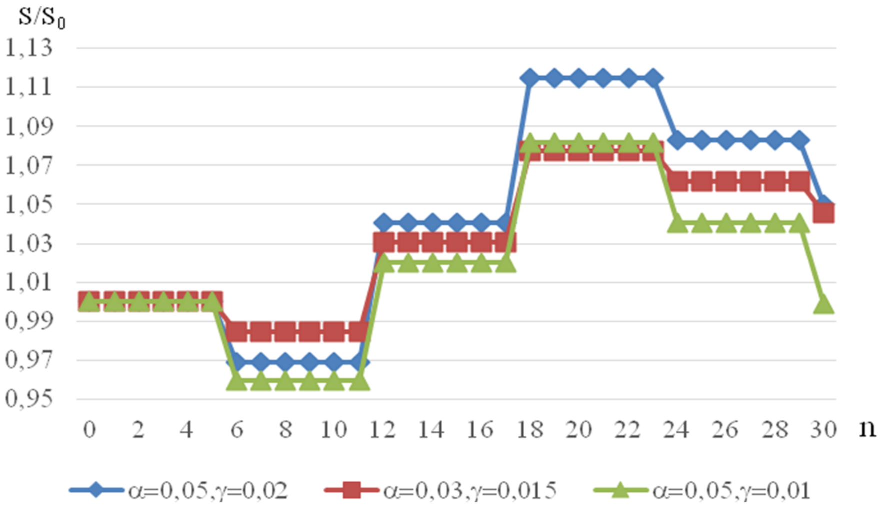


Fig. 5. Type of the demand dependence on the growth γ amplitude and seasonal fluctuations α at $m = \left[\frac{n}{6}\right]$.

After expanding functions $(1+\varepsilon)^n$ and $e^{-\gamma m}$ in the Taylor series and neglecting high-order terms, after raising square of both parts of the equation (12), it was obtained:

$$T_0^2 + 2T_0T_1\varepsilon + 2T_0T_2\gamma + (T_1^2 + 2T_0T_3)\varepsilon^2 + (2T_0T_4 + 2T_1T_2)\varepsilon\gamma +$$
$$+\left({T_2}^2 + 2T_0T_5\right)\gamma^2 = D^2 \frac{\left(2C_0\left(1+n\varepsilon+\frac{n(n-1)}{2}\varepsilon^2\right)+2\sum_{i=1}^{k} C_i\right)\cdot\left(1-m\gamma+\frac{m^2\gamma^2}{2}\right)}{\left(1-\alpha \sin\frac{\pi m}{2}\right)\sum_{i=1}^{k} S_i C_{xi}}. \quad (12)$$

Comparison of the coefficients having the same parameters degrees ε i γ allowed to obtain:

$$T_0 = D\sqrt{\frac{2\sum_{i=0}^{k} C_i}{\left(1-\alpha \sin\frac{\pi m}{2}\right)\cdot\sum_{i=1}^{k} S_i C_{xi}}},\quad T_1 = D\sqrt{\frac{2\sum_{i=0}^{k} C_i}{\left(1-\alpha \sin\frac{\pi m}{2}\right)\cdot\sum_{i=1}^{k} S_i C_{xi}}}\cdot\frac{C_0 n}{2\sum_{i=0}^{k} C_i},$$

$$T_2 = D\sqrt{\frac{2\sum_{i=0}^{k} C_i}{\left(1-\alpha \sin\frac{\pi m}{2}\right)\cdot\sum_{i=1}^{k} S_i C_{xi}}}\cdot\left(\frac{-m}{2}\right),$$

$$T_3 = D\sqrt{\frac{2\sum_{i=0}^{k} C_i}{\left(1-\alpha \sin\frac{\pi m}{2}\right)\cdot\sum_{i=1}^{k} S_i C_{xi}}}\cdot\left(\frac{C_0 n(n-1)}{4\sum_{i=0}^{k} C_i} - \frac{C_0^2 n^2}{8\left(\sum_{i=0}^{k} C_i\right)^2}\right) \quad (13)$$

$$T_4 = D\sqrt{\frac{2\sum_{i=0}^{k} C_i}{\left(1-\alpha \sin\frac{\pi m}{2}\right)\cdot\sum_{i=1}^{k} S_i C_{xi}}}\cdot\frac{C_0(-mn)}{4\sum_{i=0}^{k} C_i},\quad T_5 = D\sqrt{\frac{2\sum_{i=0}^{k} C_i}{\left(1-\alpha \sin\frac{\pi m}{2}\right)\cdot\sum_{i=1}^{k} S_i C_{xi}}}\cdot\frac{m^2}{8}.$$

“Perturbed” optimal cycle`s values of a multi-item delivery T^*_{opt} is (14)-(15):

$$T^*_{opt} = D\sqrt{\frac{2\sum_{i=0}^{k} C_i}{\left(1-\alpha\, sin\frac{\pi m}{2}\right)\cdot\sum_{i=1}^{k} S_i C_{x_i}}}\cdot$$

$$\cdot\left(1+\frac{C_0 n}{2\sum_{i=0}^{k} C_i}\varepsilon-\frac{m}{2}\gamma+\left(\frac{C_0 n(n-1)}{4\sum_{i=0}^{k} C_i}-\frac{C_0^2 n^2}{8\left(\sum_{i=0}^{k} C_i\right)^2}\right)\cdot\varepsilon^2-\frac{C_0 nm}{2\sum_{i=0}^{k} C_i}\varepsilon\gamma+\frac{m^2}{8}\gamma^2\right), \quad (14)$$

$$T^*_{opt} = \frac{T_{opt}}{\sqrt{\left(1-\alpha\, sin\frac{\pi m}{2}\right)}}\cdot$$

$$\cdot\left(1+\frac{C_0 n}{2\sum_{i=0}^{k} C_i}\varepsilon-\frac{m}{2}\gamma+\left(\frac{C_0 n(n-1)}{4\sum_{i=0}^{k} C_i}-\frac{C_0^2 n^2}{8\left(\sum_{i=0}^{k} C_i\right)^2}\right)\cdot\varepsilon^2-\frac{C_0 nm}{4\sum_{i=0}^{k} C_i}\varepsilon\gamma+\frac{m^2}{8}\gamma^2\right) \quad (15)$$

To study interval sensitivity between orders of the multi-item inventory management model to changes in input parameters, namely to executive costs and fluctuations in demand, data of tables were used. Calculations were made for α = 0,05, γ = 0,01, with values of the interval of a multi-item delivery T^*_{opt} and ${T^*}_{opt}/T_{opt}$ for different values ε, n and α = 0,05 given in table 4. It provides comparative analysis of the optimal periodicity of a multi-item delivery for business in HoReCa sector (Zaporizhzhia) at $m = \left[\frac{n}{6}\right]$.

Data presented in table 4 reveal that fixed holding costs and demand, as well as order execution costs growth by 1,5% (ε = 0,015) during the first 12 periods (n = 11, m = 1) is characterized by the rising order interval from 84 up to 92 days or by 9,5%. When demand changes (n = 12, m = 2), the order period contracts by 2,2%, which can be explained by the type of function selected to approximate demand. Similar changes in the order interval are observed at ε = 0,02 and ε = 0,01: during the first 12 periods, the order interval goes up with its subsequent reduction. In the future, the trend of changing periods between orders is kept.

Correlation between coffee order quantity changes and parameters values ε and n at α = 0,05, $m = \left[\frac{n}{6}\right]$ are presented in fig. 6. Fig. 6 shows that during the first 12 periods (n = 11, m = 1) when order execution costs change by 1% (ε = 0,01), 1,5% (ε = 0,015) and 2% (ε = 0,02), coffee orders increase by 8,1%, 9,7% and 12,9% compared to the initial period, respectively. Further change in demand (n = 12, m = 2) is characterized by the contraction of coffee orders by 2,9%, 1,5% and 1,4% compared to the previous value, respectively.

Dependence of the coffee order change on parameters values ε and n at α = 0,05, $m = \left[\frac{n}{12}\right]$ is presented in fig. 7.

Comparing fig. 6 and fig. 7, one can conclude that extension of the period of change in demand for products (from $m = \left[\frac{n}{6}\right]$ to $m = \left[\frac{n}{12}\right]$) causes the decrease in optimal order’s number of amplitude fluctuations from 4 to 2. True value of coffee order quantity, corresponding to the “perturbed” interval, at n = 23, m = 1 exceeds values corresponding to n = 23, m = 3 by 6,1%, 5,7%, 5,4%, respectively, for ε = 0,01, ε = 0,015 and ε = 0,02.

Table 4. Comparative analysis of optimal periodicity of a multi-item delivery, α=0,05, $m = \left[\frac{n}{6}\right]$, γ=0,01.

Period		ε = 0,01		ε = 0,015		ε = 0,02	
n	m	T^*_{opt}/T_{opt}	T^*_{opt}, full days	T^*_{opt}/T_{opt}	T^*_{opt}, full days	T^*_{opt}/T_{opt}	T^*_{opt}, full days
0	0	1,0000	84,00	1,0000	84,00	1,0000	84,00
1	0	1,0045	84,38	1,0067	84,56	1,0089	84,75
2	0	1,0090	84,75	1,0135	85,13	1,0180	85,51
3	0	1,0135	85,13	1,0203	85,70	1,0271	86,28
4	0	1,0181	85,52	1,0272	86,28	1,0363	87,05
5	0	1,0226	85,90	1,0341	86,86	1,0457	87,84
6	1	1,0487	88,09	1,0628	89,28	1,0771	90,48
7	1	1,0534	88,48	1,0700	89,88	1,0868	91,29
8	1	1,0581	88,88	1,0772	90,49	1,0967	92,12
9	1	1,0629	89,28	1,0845	91,10	1,1066	92,95
10	1	1,0677	89,69	1,0919	91,72	1,1166	93,80
11	1	1,0725	90,09	1,0993	92,34	1,1268	94,65
12	2	1,0449	87,77	1,0734	90,17	1,1027	92,63
13	2	1,0496	88,17	1,0807	90,78	1,1127	93,47
14	2	1,0544	88,57	1,0881	91,40	1,1229	94,32
15	2	1,0591	88,97	1,0955	92,02	1,1331	95,18
16	2	1,0639	89,37	1,1030	92,65	1,1434	96,05
17	2	1,0688	89,78	1,1105	93,28	1,1539	96,92
18	3	1,0425	87,57	1,0857	91,20	1,1307	94,98
19	3	1,0473	87,97	1,0932	91,83	1,1410	95,85
20	3	1,0520	88,37	1,1007	92,45	1,1515	96,72
21	3	1,0568	88,77	1,1082	93,09	1,1620	97,61
22	3	1,0616	89,18	1,1158	93,72	1,1726	98,50
23	3	1,0665	89,58	1,1234	94,37	1,1833	99,40
24	4	1,0923	91,76	1,1533	96,88	1,2176	102,28

Let us analyze sensitivity of the optimized formula proposed to determine the order interval of a multi-item delivery to changes in input parameters. Let us visualize the obtained results. Surface demonstrating dependence of order interval of a multi-item delivery on the period n and the growth rate of order execution costs ε at α=0,05, $m = \left[\frac{n}{6}\right]$, γ=0,01 is illustrated in fig. 8. As one can see in fig. 8, parameter`s ε growth induces interval extension between orders relative to the optimal (T_{opt} = 84). Taking this fact into account one can optimize company's logistics costs and provide additional competitive advantages.

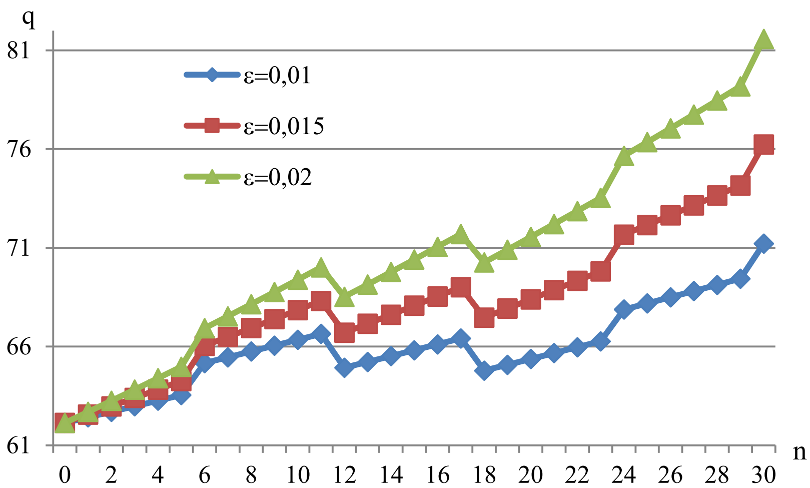


Fig. 6. Correlation between coffee order quantity changes and parameters values ε and n at α=0,05, $m = \left[\frac{n}{6}\right]$, γ=0,01.

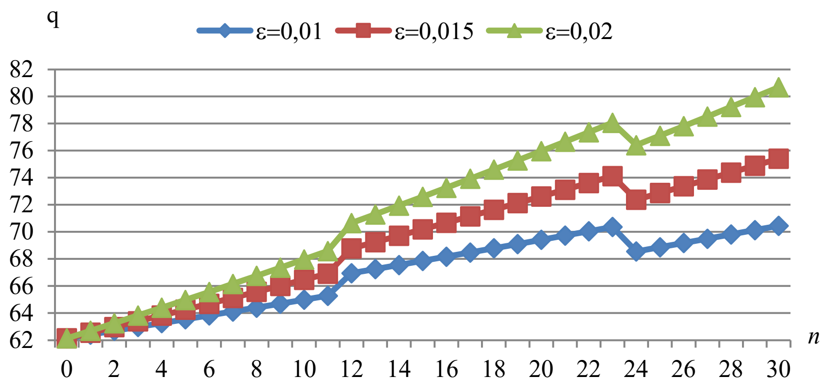


Fig. 7. Dependence of the coffee order change on parameters values ε and n at α=0,05, $m = \left[\frac{n}{12}\right]$, γ=0,01.

Visualization of the dependence of order interval of a multi-item delivery on the period n and the growth rate of demand for products γ, whilst execution costs are fixed at 2% (ε = 0,02) and the amplitude of demand change is α = 0,05, and $m = \left[\frac{n}{6}\right]$ is presented in fig. 9. Fig. 9 demonstrates parameter γ growth leading to gradual increase in the interval between orders relative to the optimal one (T_{opt}=84). However, if to compare the order intervals for different values of the parameter γ, one can see the tendency to the decreased interval between orders with growing value of γ.

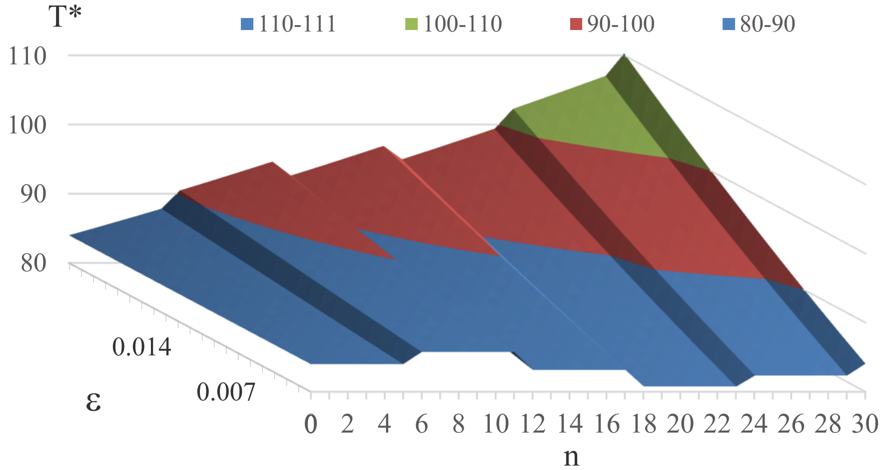


Fig. 8. Dependence of order interval of a multi-item delivery on the period n and the growth rate of order execution costs ε at α=0,05, $m = \left[\frac{n}{6}\right]$, γ=0,01.

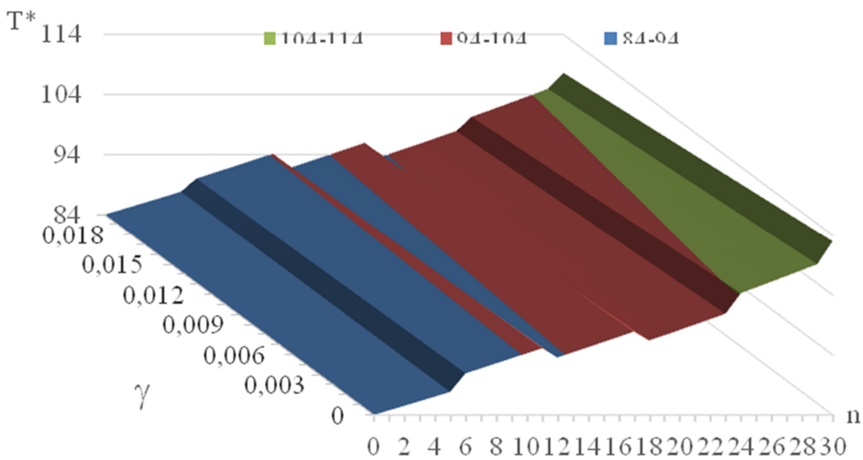


Fig. 9. Dependence of order interval of a multi-item delivery on the period n and the growth rate of demand for products γ at α=0,05, $m = \left[\frac{n}{6}\right]$, ε =0,02.

Let us analyze total costs of a multi-item delivery taking into consideration execution costs changes and gradual growth of demand for classical and perturbed order intervals (α=0,05, ε=0,02, γ=0,01). Their comparison for $m = \left[\frac{n}{6}\right]$ and $m = \left[\frac{n}{12}\right]$ respectively is presented in fig. 10 and fig. 11.

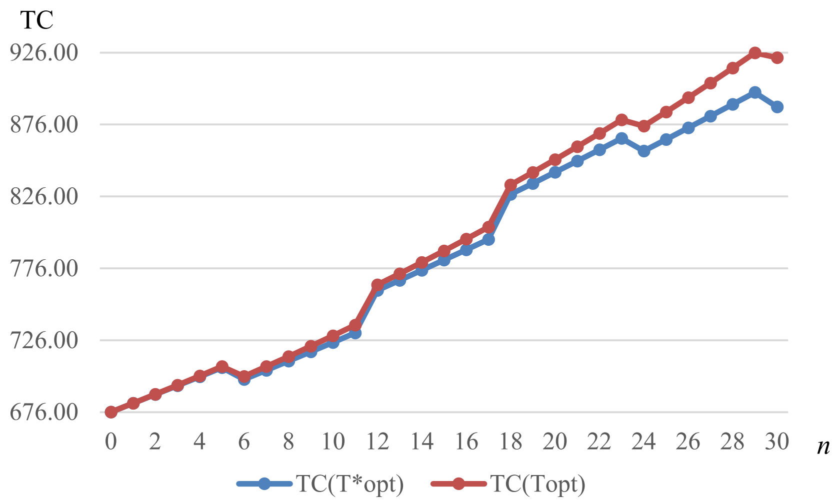


Fig. 10. Comparative analysis of total costs of a multi-item delivery for classical and perturbed order intervals α=0,05, ε=0,02, γ=0,01, $m = \left[\frac{n}{6}\right]$.

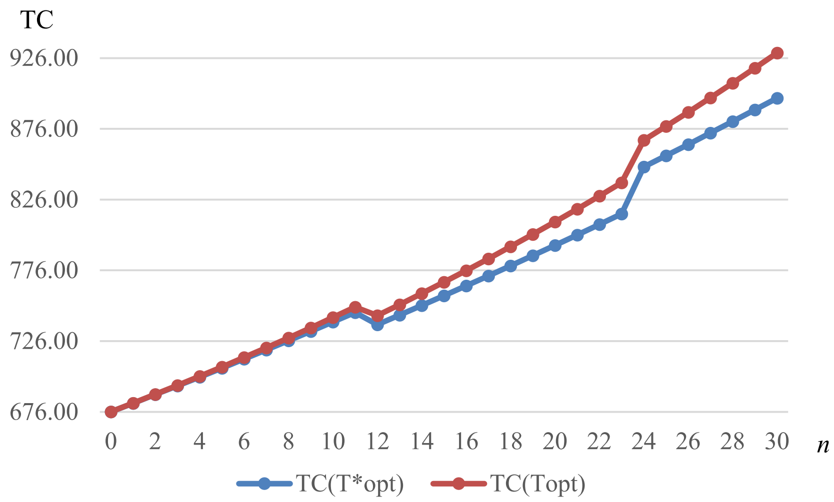


Fig. 11. Comparative analysis of total costs of a multi-item delivery for classical and perturbed order intervals α=0,05, ε=0,02, γ=0,01, $m = \left[\frac{n}{12}\right]$.

As one can see in fig. 10 and fig. 11, the obtained "perturbed" formula (15) to determine the order interval of a multi-item delivery allows company's management to optimize its total costs. Costs downtrend does not depend on correlation between parameters n and m, which is one of business' market competitiveness determinants.

5 Discussion of the results of multi-item inventory model's optimization based on the asymptotic methods

The asymptotic approach proposed for solving the multi-item inventory management model, in contrast to the classical approach, allows to vary system input parameters, which significantly expands the scope of this model`s application. In the study, the degree of system parameters variation is insignificant as percentage of the initial values. Input parameters such as order execution and inventory holding costs (by introduction of small values rate of change by periods) and demand for products (by introduction of amplitude shift and growth rate parameters by periods) were varied.

Execution costs growth rate is characterized by a small parameter ε. Interval from 0 to 0,02 (i.e. 0–2,0%) was considered as the parameter's range of changes. Inventory holding costs growth rate is determined by the parameter β, which range of variation was limited by the interval from 0 to 0,05 (i.e. 0-5,0%). These parameters values can be explained by their specificity. For instance, parameter β had bigger range of change than ε, because the percentage utility costs growth is less frequent, but is more substantial.

The gradual growth rate of demand for products is characterized by the parameter γ. The interval from 0 to 0,02 (i.e. 0–2,0%) was taken as the range of change of this parameter in the study. Parameter α was chosen as the parameter characterizing the amplitude of seasonal demand fluctuations, which range was set at 0,05 (i.e. 5%). However, to illustrate the change in the nature of the demand function, this parameter was also set at 3%. The choice of parameters α and γ is determined by the specifics of HoReCa market segment and the supplied products (coffee, tea, sugar).

The asymptotic formulas obtained for the multi-item supply model also contain the parameters n and m. These parameters specify periods of execution costs change (n) and holding costs change (m). In addition, parameter m typifies periods of seasonal demand change and growth. So, as utility costs change occurs not so often as the change in order execution costs (namely, its transport component), the parameter m was chosen as a mathematical function of quotient $m = [n / 6]$ and $m = [n / 12]$, which corresponds to the parameter change once every six months or once a year, respectively.

Solution of the multi-item supply problem amid small discrete execution and inventory holding costs growth was obtained in the form of a two-parameter asymptotic formula (9). Assessment of the developed model's sensitivity to changes in the input parameters revealed that the relative deviation of the time interval between orders (tables 2, 3) varies from 0% to +19% depending on the period. Calculation of order quantity for different values of the parameter ε (fig. 2) shows the growing tendency of order quantity with this parameter's boost.

The asymptotic formula for determining the optimal order period of a multi-item supply under the condition of changing order costs and periodic demand growth for products was obtained in the form (15). The study of the obtained "perturbed" formula's sensitivity to changes in input parameters (table 4) found out that the time interval between orders depends on the periods of input parameters change, as well as of their percentage change. Calculation of the time interval deviation between orders according to the formula (15) (table 4) at different values of the input parameters shows an upward trend. Thus, the deviation can range from 0% to +22% for the respective periods. Calculation of the interval between orders for different values of the parameter ε demonstrates its growth with respect to the optimal. Considering this fact, one can optimize company's logistics costs and provide additional advantages to the competitiveness.

Among the limitations of the study one should note application of the selected forms to approximate functions that characterize order execution and holding costs change, as well as seasonal demand function for products.

The resulting asymptotic solutions of the multi-item inventory management model are of practical significance, as the resulting "perturbed" formulas are convenient for companies' management to forecast possible changes in company's logistics system amid demand, order and holding costs variation.

The proposed asymptotic method for two parameters is the development of analytical tools for the procurement and inventory management in contrast to [7; 10]. In particular, the proposed formulas allow to apply the obtained results for optimization and forecasting of decision-making in the system of procurement logistics of a company amid variation of input parameters describing changes of external and internal business environment. An easy-to-use model that takes into account demand and costs shift makes it possible to optimize the process of business procurement organization, to forecast company's total costs in order to ensure its market competitive position.

Prospects for further research are associated with building of asymptotic solutions of inventory management models with minor changes in input parameters under scarce resources, including warehouse space, vehicle capacity, available current assets, etc.

6 Conclusions

Optimization streams of the multi-item inventory management model under the condition of insignificant changes of input parameters based on asymptotic methods were proposed.

1. Asymptotic formula to determine the order period of the multi-item inventory management model with a slight discrete order execution and inventory holding costs growth was obtained. Formula contains two small parameters that characterize order execution and inventory holding costs growth rate in accordance with the period. It makes it possible to specify the order execution period and the order quantity of product categories included into a multi-item delivery. Thus, the higher periods' multiplicity of ordering and holding costs of products (n and m) is, then period's cyclical fluctuations and the order quantity with the growing tendency take place. The improved asymptotic formula is convenient for business processes' planning and forecasting.
2. Asymptotic formula of the multi-item inventory management model with variable order execution costs and insignificant fluctuations in the amplitude of growing demand was made. Order execution costs growth rate and the demand growth rate, which depend on the period and the functions selected for approximation, were chosen as small parameters. The study of the "perturbed" interval`s deviation nature

between orders under the different conditions of gradual order execution costs and demand growth found out that it basically goes up even with minor changes in the model's input parameters. The parameter growth, which characterizes the demand growth rate, causes gradual interval value boost between orders relative to the optimal, used for forecasting and logistics decision-making by the company's management.

References

1. Andrianov, I.V., Manevitch, L.I.: Asymptology: Ideas, Methods, and Applications. Springer US, New York (2002). doi:10.1007/978-1-4419-9162-1
2. Bikulov, D., Holovan, O., Oliynyk, O., Shupchynska K., Markova, S., Chkan, A., Makazan, E., Sukhareva, K., Kryvenko, O.: Optimization of inventory management models with variable input parameters by perturbation methods. Eastern-European Journal of Enterprise Technologies 3(105), 6–15 (2020). doi:10.15587/1729-4061.2020.204231
3. Brodetskii, G.L.: Influence of order payment delays on the efficiency of multiitem reserve control models. Automation and Remote Control **11**, 94–104 (2017). doi:10.1134/S0005117917110078
4. Cárdenas-Barrón, L.E., Sana, S.S.: Multi-item EOQ inventory model in a two-layer supply chain while demand varies with promotional effort. Applied Mathematical Modelling **39**(21), 6725-6737 (2015). doi:10.1016/j.apm.2015.02.004
5. Golovan, O.O., Oliynyk O., Kovalenko, N.V.: Adaptation of logistics management systems using asymptotic methods. Actual problems of economics 5, 395–401 (2016)
6. Golovan, O.O., Oliynyk O., Shyshkin, V.O.: Logistic business processes modelling using asymptotic methods. Actual problems of economics 9, 428–433 (2015)
7. Gristchak, V.Z.: A Hybrid Asymptotic Methods and Technique of Application. Zaporizhzhya National University, Zaporozhye, (2009)
8. Jaggi, C.K., Goel, S.K., Mittal, M.: Credit Financing in Economic Ordering Policies for Defective Items with Allowable Shortages. Applied Mathematics and Computation **219**(10), 5268–5282 (2013). doi:10.1016/j.amc.2012.11.027
9. Kuzmin, O. V., Tungalag, N., E`nkhbat, R., E`rde`ne`bat, M.: Optimization approach to the stochastic problem of the stocks control. Modern technologies. System analysis. Modeling 3(55), 106–110 (2017). doi: 10.26731/1813-9108.2017.3(55).106-110
10. Lukinskiy, V.S, Lukinskiy, V.V., Pletneva, N. G.: Logistics and supply chain management. Yurayt, Moscow (2016)
11. Mittal, M., Khanna, A., Jaggi, C.K.: Retailer's ordering policy for deteriorating imperfect quality items when demand and price are time-dependent under inflationary conditions and permissible delay in payments. International Journal of Procurement Management **10**(4), 461–494 (2017). doi:10.1504/IJPM.2017.085037
12. Nasution, A.A., Rizkya, I., Ksyahpturib, Mariana, R.R.: Iventory policy for multi item products by short expiration period. In: 2nd Talenta Conference on Engineering, Science and Technology, TALENTA-CEST 2019; Grand Aston Hotel Medan; Indonesia; 17 October 2019 Volume 801, Issue 1, 2 June 2020 (2020). doi:10.1088/1757-899X/801/1/012110
13. Nayfeh, A.H.: Introduction to Perturbation Techniques. Wiley, New York (2011)
14. Oliynyk, O., Golovan, O.O., Makazan, E.V.: Directions of analytical tools improvement in management of machine-building enterprise's logistic subsystem. Actual problems of economics 9, 383–390 (2016)

15. Pentico, D.W., Drake, M.J.: A Survey of Deterministic Models for the EOQ and EPQ with Partial Backordering. European Journal of Operational Research **214**(2), 179–198 (2011). doi:10.1016/j.ejor.2011.01.048
16. Rabbania, M., Rezaeia, H., Lashgarib, M., Farrokhi-Aslb. H.: Vendor managed inventory control system for deteriorating items using metaheuristic algorithms. Decision Science Letters **7**(1), 25–38 (2018). doi:10.5267/j.dsl.2017.4.006
17. Rasay, H., Golmohammadi, A.M.: Modeling and Analyzing Incremental Quantity Discounts in Transportation Costs for a Joint Economic Lot Sizing Problem. Iranian Journal of Management Studies **13**(1), 23–49 (2020). doi: 10.22059/IJMS.2019.253476.673494
18. Saha, A., Roy, A., Kar, S., Maiti, M.: Inventory models for breakable items with stock dependent demand and imprecise constraints. Mathematical and Computer Modelling **52**(9–10), 1771–1782 (2010). doi:10.1016/j.mcm.2010.07.004
19. Sanni, S., Jovanoski, Z., Sidhu, H.S.: An economic order quantity model with reverse logistics program. Operations Research Perspectives **7**, 100133 (2020). doi:10.1016/j.orp.2019.100133
20. Saraswati, D., Sari, D.K., Johan, V.: Using genetic algorithm to determine the optimal order quantities for multi-item multi-period under warehouse capacity constraints in kitchenware manufacturing. IOP Conference Series: Materials Science and Engineering **273**, 012020 (2017). doi:10.1088/1757-899X/245/1/012020
21. Satiti, D., Rusdiansyah, A., Dewi, R.S.: Modified EOQ Model for Refrigerated Display's Shelf-Space Allocation Problem. IOP Conference Series: Materials Science and Engineering **722**, 012014 (2020). doi:10.1088/1757-899X/722/1/012014
22. Tripathi, R.P., Singh, D., Mishra, T.: Economic Order Quantity with Linearly Time Dependent Demand Rate and Shortages. Journal of Mathematics and Statistics **11**(1), 21–28 (2015). doi:10.3844/jmssp.2015.21.28
23. Tyagi, A.P.: An Optimization of an Inventory Model of Decaying-Lot Depleted by Declining Market Demand and Extended with Discretely Variable Holding Costs. International Journal of Industrial Engineering Computations **5**(1), 71–86 (2014). doi:10.5267/j.ijiec.2013.09.005
24. Vijayashree, M., Uthayakumar, R.: A single-vendor and a single-buyer integrated inventory model with ordering cost reduction dependent on lead time. Journal of Industrial Engineering International volume **13**, 393–416 (2017). doi:10.1007/s40092-017-0193-y
25. Yassa, R.I., Ikatrinasari, Z.F.: Determination of multi-item inventory model with limitations of warehouse capacity and unit discount in leading garment industry in Indonesia. International Journal of Mechanical and Production Engineering Research and Development **9**(2), 161–170 (2019)
26. Yousefli, A., Ghazanfari, M.: A Stochastic Decision Support System for Economic Order Quantity Problem. Advances in Fuzzy Systems **2012**, 650419 (2012). doi:10.1155/2012/650419

Modeling and prediction of the gas pipelines reliability indicators in the context of energy security of Ukraine

Halyna Zelinska1[0000-0003-2175-4883], Irina Fedorovych1[0000-0002-7139-6327], Uliana Andrusiv1[0000-0003-1793-0936], Oksana Chernova1[0000-0002-6424-7569] and Halyna Kupalova2[0000-0003-4486-8349]

[1] Ivano-Frankivsk National Technical University of Oil and Gas, 15 Karpatska Str., Ivano-Frankivsk, 76019, Ukraine
[2] Taras Shevchenko National University of Kyiv, 60 Volodymyrska Str., Kyiv, 01033, Ukraine
zelinska_haluna@i.ua

Abstract. Many years of experience in operation of the gas transportation system shows that the largest accidents with severe consequences arise due to untimely detection and elimination of gas leaks in underground gas pipelines. The decrease in the reliability of the gas transportation system functioning can be considered from the following two perspectives: the first perspective is the economic one – it leads to an increase in the economic expenses of an enterprise; and the second perspective is the social and environmental one – it results in emergence of a threat to public health, as well as loss of human and natural resources. Hence, the issue of modeling and prediction of the reliability indicators of natural gas transportation via gas pipelines becomes especially urgent because of the requirements for reliable operation of the system. It has been proven that the main problem leading to a decrease in the reliability of the gas transportation process is the significant deterioration of fixed assets, which requires investment of considerable financial resources in the gas transportation system of Ukraine (GTS). The article substantiates that it is possible to increase the reliability of operation of the line section of the main gas pipelines (LSMGP) through a high-quality system of repairs and equipment modernization. The main factor allowing to reduce the number of accidents is considered to be timely detection of damages on gas pipelines and their prediction. It has been determined that the failure rate depends on the diameter and number of lines of a gas pipeline. The authors propose to conduct a comprehensive diagnosis of the process of reliability of gas pipelines together with of their technical and economic indicators, based on the development of a system of measures to improve the safety of gas pipelines in Ukraine. A system of measures has been developed to improve the reliability of gas pipelines operation in Ukraine.

Keywords: reliability, expenses, modeling, accidents, economic benefits, failures, prediction, safety.

1 Introduction

The most important task of Ukraine's energy security system is to provide the reliability of the inner gas transmission system. To maintain its elements and prevent their premature deterioration is technically difficult and expensive. The core reasons of the significant repair expenses in the gas pipeline system and their sharp increase at last years are: (1) increasing the average operating age of the gas transmission lines, (2) the construction of a large number of them in areas with high soil aggressiveness, high wetlands. That's why these expanses require the deeper economic justification to reduce financial costs and obtain economic benefits. These circumstances underline the great relevance and importance of the study of the economic problems of the gas transmission lines repairing. These circumstances emphasize the great relevance and importance of studying not only the technical aspect of the problem, but also the economic problems associated with the maintenance of main gas pipelines. Lack of financial resources of an enterprise leads to the search for alternative problem solutions. One of such options consists in modeling and prediction of the gas pipelines operation reliability indicators.

2 Literature review

The increase in the level of dependence of the socio-economic development on supply of energy resources shows a clear need for improvement of the scientific foundations of the country's energy security. This issue is relatively new, so many scientists and practitioners pay more and more attention to the study of the issues associated with prediction of the gas pipelines reliability indicators in the context of the country's energy security [22]. Victoria Dergachova and Nadiia Pysar [7] are proposed differential equations that take into account price factors for the fuel and energy resources, exports, imports, as well as the mutual influence of the production volumes of certain types of energy resources on others. The reliability of supply of energy resources has become one of the most urgent political issues for the countries of Central and Eastern Europe in recent years [5]. At the same time, Mabroor Hassan, Manzoor Khan Afridi and Muhammad Irfan Khan [10] pay attention to the relationship of the policy of reliable energy supply with environmental safety and sustainable development. Shahrouz Abolhosseini, Almas Heshmati and Masoomeh Rashidghalam [1] argue in their work that energy security is the dominant factor in international stability. The researchers have established that Iran and the Caucasus are reliable energy suppliers for Europe. Hence, the scientists have proposed some alternative solutions on how to compete with competing countries in order to improve energy security. Jack D. Sharples [24] concludes in the study that the absence of market mechanisms affects negatively the country's energy security and argues that Ukraine's integration into the European gas market and reduction of the bilateral Russian-Ukrainian dependence neutralizes all concerns about regional energy security in Eastern Europe. Ganna Kharlamova and Andriy Stavytskyy prove in the [14] that the energy security of Ukraine is unsatisfactory and, using the statistical analysis, carry out a detailed analysis of the natural gas supply to Ukraine. The predictions for the country's production, import, and transit of gas for 2018-2025 are calculated based on economic and mathematical approaches. The scientists conclude that the government should carefully and transparently approach the negotiation processes with other countries interested in joint projects in gas production and transportation. Also the methodology for assessing the reliability of gas supply to the natural gas pipeline system was developed and three aspects of uncertainty and the hydraulic characteristics of the natural gas pipeline system was considered [27]. In the article [8], P. Eser, N. Chokani and R. Abhari found that a projected reduction in domestic European gas production would lead to a 12% deficit in EU gas demand by 2030. And they offered two different strategies to overcome the shortage: (1) to increase liquefied natural gas imports from various global sources against increasing supply of the Russian gas pipeline through the Nord Stream, (2) a new model of the gas system that is capturing the market as well. The potentials of studying the effect of failures and recovery coefficients on integral reliability of gas distribution systems are also considered by Nikolay I. Ilkevich, Tatyana V. Dzyubina and Zhanna V. Kalinina [12].

Luo Zheng-shan, Xi Yi-chen and Wang Hong-chao [16] have developed a comprehensive assessment of the risks of gas pipelines operation, which makes it possible to predict accidents on gas pipelines and their possible consequences. Z. Y. Han and W. G. Weng [9], in their turn, proposed a complex method for quantitative risk analysis for the gas transportation system, which consists of the incident probability assessment, consequence analysis, and risk assessment.

L. Manian and A. Hodgdon [17] assessed the integrity of gas pipelines and their management, where they proved that the result of their functioning depends on efficient management. O. Ivanov, O. Avdeuk, K. Bushmeleva, I. Ivanov and S. Uvaysov [13] develop a model for calculating the reliability of the wireless sensor telecommunication system for monitoring the state of the gas transmission network, which allows to fulfill the required level of probability of no-failure operation when a certain number of the wireless sensor modules fail. In [4] discusses the elements and structure of the distributed wireless monitoring system for detecting gas leaks in real time that ensures safe and reliable operation of different objects in the gas transmission network. Chiara Belvederesi, Megan S. Thompson and Petr E. Komers [3] carry out a comparative analysis of the safety of the US and Canadian pipelines and determine that the Pipeline and Hazardous Materials Safety Administration (PHMSA) governs 76% of pipelines in the United States, while NEB controls only 9% of pipelines in Canada and offers Canadian federal agencies to improve the accuracy and consistency in the recording of past accidents and collection of pipeline data in order to prevent and minimize future pipeline failures. Nadiia Shmygol, Władysława Łuczka, Olena Trokhymets, Dariusz Pawliszczy and Ruslan Zavgorodniy [25] improved the model for diagnosing the efficiency of resources use in the Ukrainian economy gas sector on the basis of the additive-and-multiplicative multiplier, which, unlike the existing one, takes into account their changes when determining the weighting coefficients. Mykhaylo Voynarenko, Mariia V. Dykha, Oksana Mykoliuk, Ludmyla Yemchuk and Anastasiia Danilkova [26] proposed a mathematical model of the hierarchy of factors from the point of view of their influence on the energy security of an enterprise using the graph theory. The developed model of the hierarchy of factors, which is based on the applied scientific-and-methodological approach to determining their impact on the energy security of an enterprise, makes it possible to obtain a detailed understanding of the interaction of factors, as well as the relationships and impact on the energy security of an enterprise, which ultimately leads to development of complex optimal/coordinated management solutions in the context of development and implementation of an enterprise energy security system. Ensuring the reliability of the gas pipeline system, it is necessary to find the optimal distribution of funds aimed at modernization and repair of equipment, say the authors [20]. For the optimal distribution of available resources, they propose to use a mathematical apparatus based on portfolio theory. As a criterion of optimality, the parameter of minimizing the financial costs of the enterprise is proposed. Inesa Khvostina, Nataliia Havadzyn, Liliana Horal and Nataliia Yurchenko in [15] proposed an approach to risk assessment taking into account the manifestation of emergent properties and using the method of taxonomy and factor analysis, which involves building economic and mathematical models that take into account qualimetric and structural components of the production process.

However, solution of the issues on increasing the reliability level of gas pipelines operation in the context of ensuring the country's own energy security requires further studying. Therefore, practical modeling and prediction of the reliability indicators of gas pipelines at the gas transmission enterprises require careful study and analysis.

3 Results and discussion

As an applied field of knowledge, the science of reliability is based on fundamental mathematical and natural sciences. It studies the patterns of change in the quality of technical devices and systems and provide its trouble-free exploitation with the minimum time and resources expense.

The reliability of gas transmission lines, like any technical object, is defined as the ability to perform the specified functions, while maintaining the specified performance over time. However, being a complex feature, depending on the purpose and conditions of operation, reliability may include failure-free, durability, maintainability or certain combination of these features.

The specific features of reliability are: (1) the time factor as the change in initial parameters during the operation of the equipment is estimated, (2) an object behavior prediction that maintains its original parameters (quality indicators).

The reliability of the gas supply system, its subsystems and facilities depends on many factors, among which are the following:

- the level of reliability of the elements of the equipment included in the system;
- the operation and management level of the system;
- the composition of the elements included in the system and the structure of the relations between them;
- the amount and structure of gas reserves;
- management efficiency.

The reliability and technological characteristics of the elements of these systems (average time of emergency and scheduled repairs, await repair time, the elements performance) largely depend on the quality of the equipment and the level of operation of the system. The values of these parameters are limited by the level of scientific and technological progress achieved and the economic feasibility of additional costs for the technology improvement. These factors can change both through the rational use and allocation of costs for the system's creation and development, as well as the costs for reserves, and also by increasing of these costs. So, reducing the reliability level of the gas pipeline results both in the weakening of the economic security of the state and in the fatal consequences for humans.

When planning repairs, upgrades and reconstruction of gas transmission lines it is important to prevent gas pipeline failures and crashes. Therefore, it is necessary to pay the greatest attention to the amount of equipment depreciation of the linear part of the pipeline. However, the service life can be extended as a result of inspections and diagnostics of the technical condition of the gas pipelines. That's why it is necessary to take into account the volume of natural gas transportation planned; safety of gas transmission lines; gas pipelines demolition.

The main characteristics of the gas transmission lines reliability: trouble-free, durability and maintainability.

We have determined that gas transmission lines may be in one of the following states: (1) loaded, (2) scheduled preventive repair, (3) forced (emergency) idle time. As a result of the influence of various external connections of a random nature, the gas transmission lines in the process of exploitation passes from one state to another. The transition is made at random times. Only sometimes it is possible to predict the exact residence time of gas transmission lines. But even so, there is uncertainty about the onset of the moment of change of state.

The stability of the gas transportation system is determined by the material base, which includes the gas-pumping machinery and the linear part of the main gas pipelines. Only highly qualified personnel can satisfy the needs of its high efficiency and reliability. Predictions of gas pipeline failures and rapid elimination of the consequences of accidents are particularly relevant at modern, powerful gas pipelines.

Determination of economic losses caused by accidents is important for the gas transmission company. Economic risks are not subject to mandatory assessment by supervisory authorities. Therefore, the permitted levels of these risks are not regulated. The economic risks assessment model based on economic feasibility and efficiency of gas pipeline further operation was offered. We find the economic losses caused with the failures and accidents determination to be particularly important for gas transmission companies.

We offer to determine the enterprise's economic losses by the following steps.

1. The main losses from failures of a gas pipeline are the cost of direct gas losses *Ld*, losses during gas pipeline downtime *Ldt* and the repair cost *Lr*. It also includes losses from possible damage to technical facilities and communications, crossing the gas pipeline at the accident site *Lt* and the social costs caused by the possible people displacement or the restoration of buildings *Ls*.

2. Gas leakage losses from gas leak in monetary terms is:

$$Ld = (V1 + V2) \cdot Pg \tag{1}$$

where $V1$ – the volume of gas exiting the pipeline by the time the taps are closed;
$V2$ – the volume of gas exiting the pipeline after the taps have been closed until the pipeline is fully released;
Pg – the price of 1 cubic meter of gas.

3. Losses during gas pipeline downtime *Ldt*:

$$Ldt = Kdn \times Vd \times Pg + Pp \tag{2}$$

where *Kdn* – downtime, days;
Vd – daily volume of gas transportation;
Pp – the penalties according to the contract.
The repair cost depends on the type of failure, the amount of equipment involved, the availability of the facility and the repair type.

4. Losses from possible damage to technical facilities and communications include the amount of compensation for repair or the losses incurred by the pipeline owner if these facilities are on his balance sheet.
5. Social loss is associated with the possible displacement of people or the restoration of social facilities that have been affected by the failure. Economic social losses are estimated if a threat to social objects that border or are in the natural areas of the main gas pipelines affected. The magnitude of these losses depends on the type of failure and the distance to the social object. In the case of insurance policies, the total loss may be reduced by the amount of insurance claims.

Therefore, by determining the economic losses by individual components, gas transportation companies will be able to determine the order of precautionary feasibility measures to minimize costs.

The analysis of statistical data for 2016-2018 showed that the current state of the GTS enterprises of Ukraine is a reflection of the general economic crisis phenomena. More than 50% of gas pipelines have been operated for over 30 years.

The structure of gas pipelines in terms of service includes: up to 10 years – 3%; 10–15 years – 5%; 16–20 years – 9%; 21–30 years – 32%; over 30 years – 51%. Powerful gas pipelines such as “Soyuz”, “Progress” and “Urengoy-Pomary-Uzhgorod” were built and used for transit only. Most of the compressor stations of these gas pipelines have imported high-tech equipment, but a large part of it already requires to be changed or renovated. More than 700 gas pumping units installed at compressor stations of gas pipelines of Ukraine. Almost 30% of them have already passed the final date of its exploitation period according to the documents. The consequence of this situation is a rather low effectiveness of functioning of the units, their efficiency coefficient of 24–26%, an overconsumption of fuel gas, and, accordingly, a decrease in the productivity and reliability of the main gas pipelines operation, which, in general, will have a negative impact on the energy security over the years.

Accidence is a contrast to a reliable energy transportation process. The largest number of accidents on gas transmission lines occurs due to defects in pipe metal, as well as to violations of the rules of gas transmission systems operation, arising as a result of poor-quality welding of patches when cutting openings for the installation of rubber balls. Many accidents are caused by temperature deformation of the gas pipeline.

The failure rate of gas pipelines shows an indicator $\lambda(t)$, which can also show the intensity of failures [17].

$$\lambda(t) = \frac{n}{\Delta t n(t) L}, \tag{3}$$

where n – the number of failures during Δt over the entire length of the pipeline;

$n(t)$ – the number of non-failing elements by the time t;

L – the length of the pipeline.

For main gas pipelines, this value is given depending on the diameter as the ratio of the number of days of emergency status to the entire period of operation (year). For main gas pipelines, this value is given depending on the diameter as the ratio of the number of days of emergency status to the entire period of operation (year). The failure rate also depends on the number of lines and the diameter of the pipeline (table 1).

Table 1. The dependence of the failure rate of the gas transmission lines on the number of pipes and their diameter (based on [12]).

Diameter, mm	**The failure rate (at 1/hour*10^{-3}) for different lines number *N***			
	N=1	*N*=2	*N*=3	*N*=4
1020	17,8	27,4	35,9	43,6
1220	24,6	38,1	49,9	60,4
1420	32,6	50,2	65,5	79,5

For example, here are sections consisting of several lines of gas pipelines in the territory of Western Ukraine, namely: gas pipelines “Belchevolitsa-Dolina” and “Ivacevichi-Dolina-III”; gas pipelines “Bogorodchany-Dolina – “Torzhok-Dolina”; gas pipelines “DUD-I”, “DUD-II” and “Progress”; gas pipelines “Pasichna-Dolyna”, “Bogorodchany-Dolyna”, “UPU” and “Soyuz”; gas pipelines “Uhersko-Ivano-Frankivsk”, “Uhersko-Ivano-Frankivsk-Chernivtsi”, “Pukenichi-Dolyna” and “KZU-II” and others.

It is possible to approximate the failure rates for pipelines of other length using the following coefficients (table 2).

Accidents and failures on gas transmission lines are discrete quantities that are independent of each other, and this allows us to predict these figures with the help of a statistical and mathematical apparatus. Thus, we have determined that the statistics on the failure of the linear part of gas pipelines are quite consistent with the exponential probability distribution function.

Table 2. Coefficients for determining the failure rate of gas pipelines of different lengths (based on [12]).

Length, km	**The number of pipes *N*, diameter, mm**		
	***N*=1 *d*=1020**	***N*=2 *d*=1220**	***N*=3 *d*=1420**
1000	1,0	1,0	1,0
2000	1,53	1,58	1,60
3000	2,00	2,08	1,15

The additional verification of the exponential distribution law compliance with the actual data for the calculations fully confirmed the assumption formulated in the paper about the probability distribution law and a number of the following important properties:

– time between gas pipeline failures is described by exponential law:

$$F_1(t) = 1 - e^{-\lambda t} \tag{4}$$

with a distribution density:

$$f_1\,(t) = \lambda e^{-\lambda t} \tag{5}$$

– the probability of the number of failures in a gas pipeline of length L in a single gas line at time T is described by Poisson's law:

$$p\{n\} = \frac{(\lambda LT)^n}{n!} e^{-\lambda LT} \tag{6}$$

where λ is a constant positive value.

– the failure rate of the commissioned gas pipelines decreases monotonically over time, and the failure rate increases over time for gas pipelines operated for more than 20 years.

Our additional studies have shown that in most cases it is possible to use an exponential representation of the law of distribution (fig. 1).

Recovery time distribution function is represented as:

$$F_{2\,(t)=1-e^{-\lambda t}} \tag{7}$$

and can be used to build methodological bases for reliability assessment.

The failure rate increases with the length of the pipeline, its corrosion etc. Our processing of a large number of statistics shows the existence of a linear relationship between the specific failure rate and the diameter of the pipeline:

$$\lambda = a_1 d + b_1 \tag{8}$$

The coefficients b here were determined based on the built-in functions of the trend curves of the Excel and are relevant $a_1 = 0{,}89\cdot10^{-10}$; $b_1 = 0{,}987\cdot10^{-8}$.

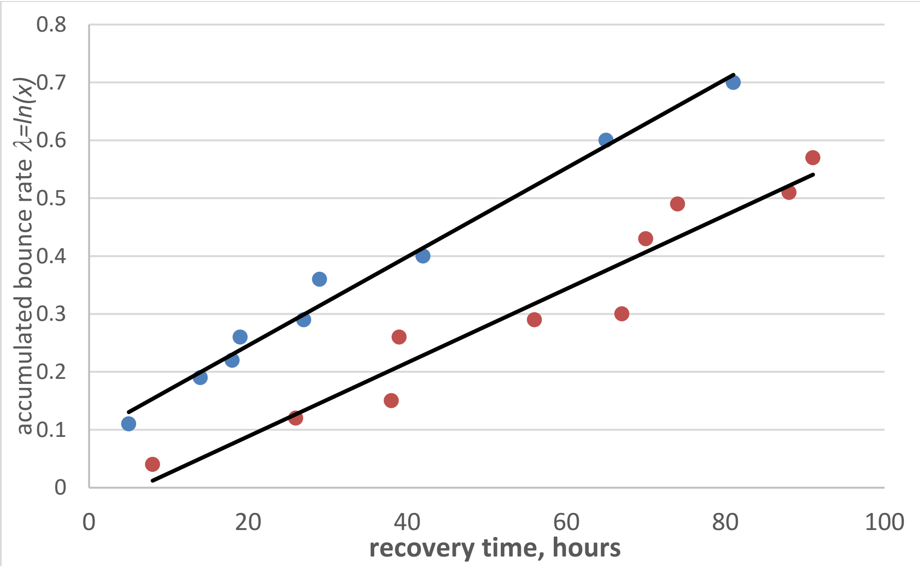


Fig. 1. Direct correspondence of experimental data to the exponential law of distribution of the operating time of the linear part of gas pipelines: 1 – diameter 1220 mm; 2 – diameter 1420 mm [30].

Fig. 2, 3 (based on the commercial statistics of Western Ukrainian Gas Pipelines Department for 2017-2018) shows the results of the study: a graphical interpretation of the dependence of (a) the failure rate and (b) the average recovery time from the diameter of the pipeline for the actual data and their trends. As we can see, we have a high accuracy of getting theoretical dependencies on the actual data. As we can see, the actual data are highly correspondent to theoretical trends.

To estimate the intensity of failures $\lambda(t)$ different methods can be used. For the linear sections of pipeline the approaches used in [21] allow, on the basis of information on the displacement for a certain set of surface points of the studied body, to determine the law of motion of each point of the investigated pipeline in the form:

$$\vec{r}\,(s,\varphi,r,t) = \vec{r_l}(s,\varphi,r,t) - R\,\vec{n_l} +$$
$$+\rho(s,\varphi,r,t)\big[\cos\omega(s,\varphi,r,t)\cdot\vec{b_1} + \sin\omega(s,\varphi,r,t)\,\vec{n_l}\big] + \psi(s,\varphi,r,t)\vec{L_l} \tag{9}$$

where $\rho(s,\varphi,r,t)$ $\omega(s,\varphi,r,t)$ $\psi(s,\varphi,r,t)$ – functions that characterize the points movement of the body studied in radial, transverse and longitudinal directions;

variables s,φ,r – are related to a curvilinear coordinate system:

s – coordinate along the axis of the body, $0 \le s \ \le L$;

φ – coordinate in the polar angle, $0 \le \varphi \le 2\pi$;

r – coordinate for the radius of the pipeline: $R_{int} \le r \le R_{out}$, $R_{ви}$ – internal and $R_з$ – the outer radius of the pipe,

$\vec{n_l}.\vec{b_1}\,,\vec{L_l}$ – components corresponding to normal, binormal and tangent at the studied point of the body.

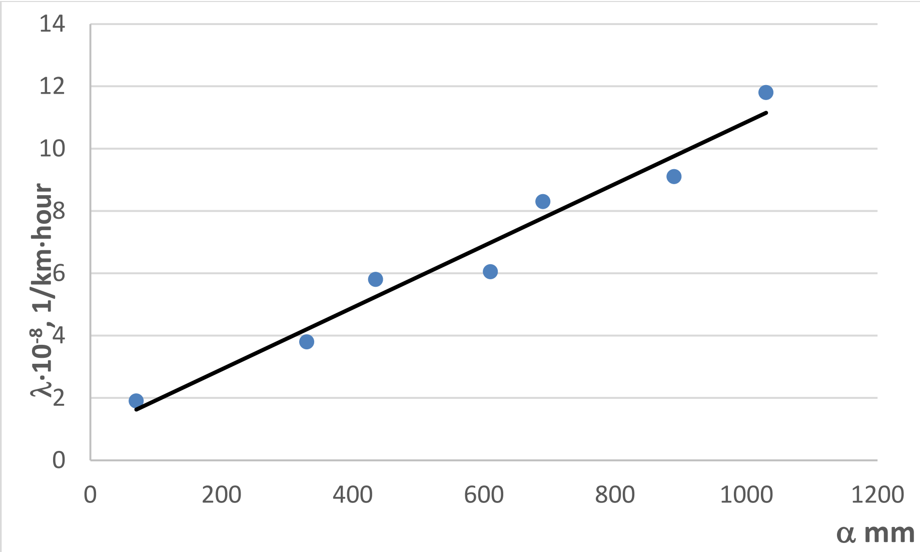


Fig. 2. The dependence of the specific failure rate on the diameter of the pipeline.

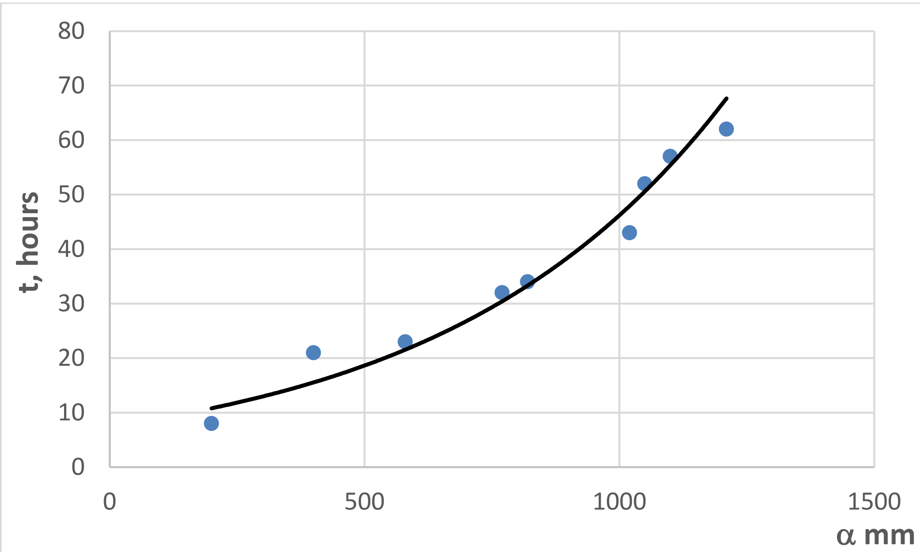


Fig. 3. The dependence of the average recovery time on the diameter of the pipeline.

The formula (3) is valid for a quasi-linear section of objects. In [21; 30] the ideas are given for conical and spherical sections of pipelines used in various industrial systems. The presentation of the form (1) allows us to calculate the change of the stress-strained state of the studied objects within the model of a stress-strained isotropic or anisotropic body using the formulas [21]:

— for deformation tensor components:

$$\varepsilon_{ij} = \frac{1}{2}(\nabla_i w_j + \nabla_j w_i) \tag{10}$$

where w_i – components of the displacement vector calculated by (1);

∇_i – covariant differentiation operator in the corresponding coordinate system (Cartesian, cylindrical, conical, orthogonal) [30];

— for stress tensor component (isotropic model):

$$\sigma^{ij} = \lambda I_1(\varepsilon_{ij})g^{ij} + 2\mu\varepsilon^{ij} \tag{11}$$

where $\sigma^{ij}, \varepsilon^{ij}$ – contravariants components of the strain and stress tensor;

s,φ, r – pseudopolar coordinates;

$I_1(\varepsilon_{ij})$ – the first invariant of the strain tensor:

$$I_1(\varepsilon_{ij}) = \sum_{ij=1}^{3} \varepsilon_{ij} g^{ij} \tag{12}$$

$$\mu = \frac{E}{2(\sigma+1)} \text{ and } \lambda = \frac{E\sigma}{(1+\sigma)(1-2\sigma)} \tag{13}$$

— for stress tensor component (isotropic model):

$$\sigma_{ij} = \sum_{k,l=1}^{3} C_{ijkl}\varepsilon_{kl} \tag{14}$$

where C_{ijkl} – components of the elastic module tensor, there is a relationship between the covariant and contravariant components of the tensor:

$$\varepsilon^{kl} = \sum_{i,j=1}^{3} \varepsilon_{ij}\, g^{ik} g^{jl} \tag{15}$$

On the fig. 4 it is shown that zones off potential section's failure can be defined as zones of great stresses change – greater than critical value 400 MPa.

It is possible to realize the technology (3)-(9) to calculate the stresses changing in different moments of time. If *N* – the total number of points, in which the stresses changing is defined, *M* – the total number of points, in which the value of tresses changing is greater that valid values, the value $\lambda(t)$ can be estimated as.

$$\lambda(t) = {}^{M}\!/_{N} \tag{16}$$

The analysis of the results showed that when the diameter increases, the increase in λ can be explained by the following reasons:

— when the weight of the pipes increases, loading and unloading operations become more complicated, raising the likelihood of damage to the pipes during such operations;
— joining of pipes during welding when the diameter increases is complicated even if the specified technical conditions are observed. The radius of the pipeline sag bend increases with an increase in the diameter, which sometimes causes difficulties

when laying pipelines in a trench at the turns of the route and results in appearance of increased stress;
- the temperature regime of large-diameter gas pipelines during operation is more severe than the temperature regime of small-diameter pipelines, which can lead to thermal deformations.

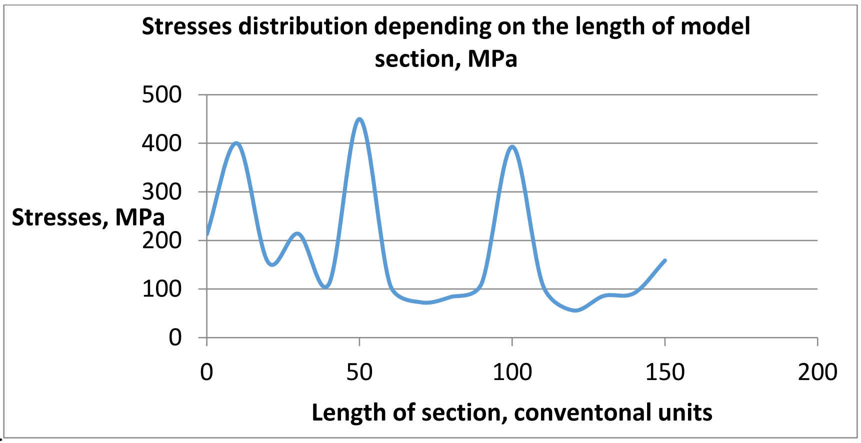


Fig. 4. Stresses distribution depending on the length of model section, MPa (based on [19]).

The next important indicator of the gas pipeline reliable operation is the mean time to recovery, which is determined by formula (15), or the recovery rate of μ, which is the inverse value of the mean time to recovery.

$$t_{av} = \frac{1}{n}\sum_{i=1}^{n} t_i \tag{17}$$

where t_i is the time of liquidation of the i-th accident;
n is the total number of accidents.

The recovery time depends on the nature of the accident, time of the year, conditions of the gas pipeline route, distance between the accident scene and emergency repair station, equipment of the emergency team with transport, machines, and mechanisms, as well as qualification of the personnel involved in the reproduction process. The time for repair of the same diameter gas pipeline varies widely. However, the dependence of the mean time to recovery on the diameter can be approximated by the following dependence:

$$t_{av} = a_2 d^2 + b_2 \tag{18}$$

The values of the coefficients b (18) are determined by prediction. We obtained the following results: $a_2 = 3.07 \cdot 10^{-5}$; $a_2 = 8.97$.

The graph (fig. 3) shows the curve of dependence (18), as well as the actual values of t_{av} obtained as a result of processing the statistical data. When the diameter increases, the growth of t_{av} is explained by the following reasons:

- the length of the pipeline rupture increases with an increase in the diameter leading to a growth in the volume of repair works;
- an increase in the diameter results in a growth in the volume of excavation and welding works;
- joining of pipes during repair works on large-diameter gas pipelines becomes sharply more complicated and takes a large share of the total time spent for accident elimination.

Financial losses from accidents are considered to be one of the main characteristics that determine the strategy for financing the LSMGP repair works. Prediction of the amount of losses in the gas pipeline section in the event of accidents determines the priority in financing the repair works of the section. Taking into account the fact that the cost of gas is growing, it is necessary to determine the financial losses of an enterprise from the volume of the lost gas, as well as the main factors that led to occurrence of accidents at the facilities under study. The complexity of the algorithm for assessing the losses from accidents at the LSMGP primarily consists in the fact that its value depends on many different factors (diameter of the gas pipeline, its length, laying conditions, age of the gas pipeline, etc.). Such circumstances induce to consider losses as a random variable.

Failure prevention is possible as a result of the high-quality repair service (works on replacement or repair of the gas pipeline, the parameter value of which approached the limit), which requires financial investments and highly educated personnel. The equipment, which is not only physically, but also morally obsolete (it results in frequent stoppages in work and emergency situations; the equipment is often operated in uneconomical modes), is now used in the system of transportation and storage of natural gas in Ukraine. This leads to an increase in consumption of the fuel and energy resources (FER), as well as to an increase in their prime cost and processing losses.

In order to predict further development of the gas pipeline defective areas by the time an emergency situation occurs, it is necessary to consider the amount of losses when it actually occurs.

We carried out statistical processing of the data on defects in gas pipelines based on in-line inspection (INI). It was found that the relative failure rate and overhaul work scope do not have a clear trend, which is determined by the crisis in the economy, decrease in the volume of natural gas transportation, and change in the pricing policy for the services of Naftogaz of Ukraine, NJSC.

The most important indicator characterizing the efficiency of operation of the main gas pipelines in the context of the country's energy security is the availability (reliability) coefficient.

In order to determine the tendency of this indicator to change, we carried out some prediction based on an expert survey of the following three groups of specialists: scientists, practitioners, and gas pipeline personnel.

In order to determine the numerical score of the complex indicator of the gas pipeline reliability as one of the main indicators affecting the energy security of the country, the methodology developed by Dow Chemical Co. and based on the relative index of gas pipeline reliability (relative index of pipeline safety – RIPS) was taken as a basis [18]. This criterion is determined using 5 indices. Four of them (F_1, ..., F_4) characterize the most typical causes of the line section failures, which include anthropogenic influences, corrosion, design errors, and operational control errors. The fifth α characterizes the severity of the consequences in emergency situations.

We improved and detailed this methodological approach for reliability assessment, as well as filled table 3 with the indicators that describe the most common causes of failures on the gas pipelines in the Western region of Ukraine.

Thus, based on the developed-by-us assessment of the reliability of the line section of the main gas pipelines, we will get the complex indicator of the LSMGP availability characterizing the reliability and calculated using formula (19):

$$K_q = \sum_{i=1}^{11} \beta_i \cdot F_i \tag{19}$$

where β is the probability of emergency situation occurrence at the i-th factor influence.

After applying the study results with the help of the expert assessments, we will obtain the following dependence that is typical for the enterprises in the Western region (Dolyna LPDMGP is taken as an example):

$$K_q = 0{,}11F_1 + 0{,}1F_2 + 0{,}1F_3 + 0{,}1F_4 + 0{,}07F_5 + 0{,}09F_6 + 0{,}1F_7 + 0{,}09F_8 + \\ +0{,}09F_9 + 0{,}08F_{10} + 0{,}07F_{11} \qquad (20)$$

Table 3 shows the procedure for calculating the indicators, with the help of which it is possible to quantitatively represent the factors that affect the occurrence of failures and accidents on the line section of the main gas pipelines reducing their reliability. As can be seen from the above Table 3, one of the indicators that affect the complex indicator of the gas pipelines reliability is the coefficient of serviceability. Let's predict its value for the period of 2020-2035. Let's consider the scenario approach that provides for realistic and optimistic predictions (tables 4 and 5). The first one involves predicting the coefficient of serviceability based on the statistical data from different gas transmission enterprises using linear regression. The second one consists in making investments into the GTS, its modernization, and human development. The investments into human development should be considered by an enterprise as one of the priority tasks of its strategy.

Therefore, according to the provided predicted data based on the second scenario (table 5), the gas transmission enterprises of the Western region will be able to significantly improve the indicator of the gas pipeline operation reliability that will reach the value of 0.68 in 2035, which is 0.1 higher than the value according to the predicted data based on the first scenario.

Since the processes of gas transportation and storage are considered to be energy-intensive (the share of expenses for the FER in the total prime cost of gas transportation is 60-80%) and has a direct impact on the energy security of our country, the priority for an enterprise is to carry out technological production re-equipment, i.e. improvement of technologies, which is not possible without highly professional personnel. At present, the cost of labor is determined by the following quality characteristics: intelligence, education, and professionalism [2; 23; 29].

Table 3. Factors and indicators of their measurement affecting gas pipelines operation reliability.

Factor	Designation	Indicator calculation algorithm
Quality of the works on the gas pipeline construction	F_1	Ratio of the volume of the works on the gas pipeline construction carried out in compliance with the construction instructions and standards to the total volume of the works on the gas pipeline construction
Quality of the gas pipelines repair service	F_2	Ratio of the volume of the repair works carried out in compliance with the requirements and standards to the total volume of the gas pipelines repair works
State of the gas pipeline insulation coating	F_3	Ratio of the length of the gas pipelines covered with the suitable insulation coating to their total length
Level of the gas pipeline corrosion failure	F_4	Ratio of the number of the defective corrosion damages on the pipe body to the length of gas pipelines
Natural-and-geographical location and laying environment of gas pipelines	F_5	Depth, at which the gas pipeline is located; Atmospheric conditions (temperature, humidity); Possibility of ground distortion; Possibility of uneven soil compaction; Possibility of erosion of the gas pipeline due to flooding or changes in the river bed
Level and quality of the gas pipeline inspection and cleaning	F_6	Ratio of the length of the gas pipelines that have passed the in-line inspection and cleaning to the total length of gas pipelines
Qualification level of the repairmen, engineers, and other technical workers	F_7	Ratio of the number of the engineers and other technical workers that correspond to the position held according to the certification results to the number of the employees who have passed the certification
Coefficient of serviceability of the main gas pipelines, fraction unit	F_8	Ratio of the weighted service life of gas pipelines to the average standard service life of gas pipelines
Level of sophistication in making management decisions on operation and restoration of main gas pipelines	F_9	Ratio of the number of correct and timely made management decisions to the total number of the management decisions related to the process of operation and restoration of gas pipelines
Level of pipe defects	F_{10}	Ratio of the volume of the used pipe with some factory and mechanical defects to the total volume of the used pipe
Level of protection of gas pipelines	F_{11}	Ratio of the length of the gas pipelines protected against corrosion with the help of electrochemical protection to the total length of gas pipelines

Table 4. Prediction of the coefficient of serviceability based on the first scenario.

Predicted values				Years				
2020	2025	2030	2035	2013	2014	2015	2016	2017
0.58	0.54	0.56	0.58	0.68	0.66	0.640	0.526	0.618

Table 5. Prediction of the coefficient of serviceability based on the second scenario.

Predicted values				Years				
2020	2025	2030	2035	2013	2014	2015	2016	2017
0.65	0.65	0.66	0.68	0.68	0.66	0.640	0.526	0.618

In order to improve the reliability indicators of the Ukrainian GTS functioning in the context of energy security, we propose to introduce the following [28]:

A) system of technical and technological measures:

- Modernization and replacement of pipelines.
- Improvement of the anti-corrosion protection.

These measures will allow to prevent emergency situations on the line section of the gas pipelines, increase their resistance to adverse natural and climatic conditions, as well as reduce gas pollution during transportation (prevention of deterioration of quality parameters) and gas leaks. All of this will lead to optimization of the system as a whole, as well as to reduction of the FER losses, and, hence, total expenses of an enterprise.

- Introduction of new energy-efficient engines. The largest portion of the engines that drive compressor stations (hereinafter referred to as CS) is made by gas turbine engines, the efficiency coefficient (efficiency) of which is very low and does not even reach 25%, therefore, it is necessary to replace them with more energy-efficient engines (the efficiency of which is higher). A sufficient niche of manufacturers of gas turbine engines (GTE) with the best technical quality parameters and greater efficiency formed on the domestic market.
- Reduction of energy losses associated with the change in the load of gas pipelines and achievable by increasing the level of automatic control and regulation of CS operation, as well as by introducing the automated systems for enabling and disabling a gas-pumping unit (GPU).

B) The necessary prerequisite for implementation of the above measures is the fulfillment of a number of socio-economic tasks, the results of which will make the basis for their realization. In particular, in order to control the deviations of the FER consumption rates, it is necessary to develop and establish them first, therefore, the first of the socio-economic measures should be the following [6; 11]:

1. Development of economically justified norms of the FER specific consumption rate. In order to do this, it is suggested to conduct an in-depth analysis of available equipment, study technical characteristics, determine optimal operating and

loading modes, as well as develop economically justified norms of the FER consumption rate on the basis of the comprehensive knowledge obtained.
2. Introduction of the system of stimulation and personal responsibility of employees. Material stimulation for an efficient use of energy resources, carrying out of the work to improve the efficiency of using the FER, introduction of different energy-saving technologies, etc. by awarding the employees with bonus payments within the established share of the cost of the saved FER will lead to the material interest of each employee, since the economic result of the whole enterprise will depend on their work.
3. The efficient and flexible management system capable of making operational management decisions on effective business activities in various conditions should become the quintessence of the work of a gas transmission enterprise in the market conditions.

4 Conclusions

The statistical models for calculation of the gas pipelines reliability indicators were obtained on the basis of actual data on the gas pipelines operation in the context of the country's energy security. Utilization of these models made it possible to carry out a prediction based on the partial indicators and make adjustments to the system of gas pipelines maintenance, which will increase their operational reliability and energy security of Ukraine.

Justification of effective management decisions and reforms requires accuracy not only of the relative quantitative assessment of the reliability level of gas supply and impact of individual threats, but also of the absolute and qualitative assessment, which will allow modeling individual scenarios of the measures for the reform implementation. Constant monitoring and use of the predicted parameters of the gas pipelines state will provide a possibility to reduce the accident rate on main pipelines, save significant financial resources, and obtain an economic effect due to the system of technical, technological, economic, social, and environmental measures. Since the indicators of the gas pipelines operation reliability are influenced by a significant number of different factors, it is very difficult to assess the influence of each of them individually and in a complex, therefore the methods for statistical modeling and prediction are suggested to be utilized. When using them, it is possible to take into account the influence of all the factors on the reliability indicators, on which the functioning of the gas transportation system, as well as the energy security of Ukraine, will largely depend.

References

1. Abolhosseini, S., Heshmati, A., Rashidghalam, M.: Energy security and competition over energy resources in Iran and Caucasus region. AIMS Energy **5**(2), 224–238 (2017). doi:10.3934/energy.2017.2.224
2. Andrusiv, U., Kinash, I., Cherchata, A., Polyanska, A., Dzoba, O., Tarasova, T., Lysak,

H.: Experience and prospects of innovation development venture capital financing. Management Science Letters **10**(4), 781–788 (2020). doi:10.5267/j.msl.2019.10.019

3. Belvederesi, C., Thompson, M., Komers, P.: Canada's federal database is inadequate for the assessment of environmental consequences of oil and gas pipeline failures. Environmental Reviews **25**(4), 415–422 (2017). doi:10.1139/er-2017-0003
4. Bushmeleva, K.I., Plyusnin, I.I., Bushmelev, P.E., Uvaisov, S.U.: Distributed wireless system for monitoring the technical state of objects in a gas-transport network. Measurement Techniques **56**(3), 226–231 (2013). doi:10.1007/s11018-013-0184-3
5. Butler, E.: Conclusion: Central and eastern European energy security - more than Russia. In: Ostrowski, W., Butler, E. (eds.) Understanding Energy Security in Central and Eastern Europe: Russia, transition and national interest, pp. 222–231. Routledge, London (2018). doi:10.4324/9781315651774
6. Cherchata, A., Popovychenko, I., Andrusiv, U., Simkiv, L., Kliukha, O., Horai, O.: A methodology for analysis and assessment of business processes of Ukrainian enterprises. Management Science Letters **10**(3), 631–640 (2020). doi:10.5267/j.msl.2019.9.016
7. Dergachova, V., Pysar, N.: Implementation of the market approach to the processes of management of the energy sector of Ukrainian economy under conditions of European integration. Eastern-European Journal of Enterprise Technologies 3(3-93), 40–49 (2018). doi:10.15587/1729-4061.2018.133437
8. Eser, P., Chokani, N.: Abhari, R.: Impact of Nord Stream 2 and LNG on gas trade and security of supply in the European gas network of 2030. Applied Energy **238**, 816–830 (2019). doi:10.1016/j.apenergy.2019.01.068
9. Han, Z.Y., Weng, W.G.: An integrated quantitative risk analysis method for natural gas pipeline network. Journal of Loss Prevention in the Process Industries **23**(3), 428–436 (2010). doi:10.1016/j.jlp.2010.02.003
10. Hassan, M., Afridi, M.K., Khan, M.I.: Energy policies and environmental security: a multi-criteria analysis of energy policies of Pakistan. International Journal of Green Energy **16**(7), 510–519 (2019). doi:10.1080/15435075.2019.1593177
11. Ievdokymov, V., Lehenchuk, S., Zakharov, D., Andrusiv, U., Usatenko, O., Kovalenko, L.: Social capital measurement based on "The value explorer" method. Management Science Letters **10**(6), 1161–1168 (2020). doi:10.5267/j.msl.2019.12.002
12. Ilkevich, N.I., Dzyubina, T.V., Kalinina, Z.V.: Development of analysis methods for ensuring the reliability of gas distribution systems in the problems of complex energy supply. E3S Web of Conferences **102** 02005 (2019). doi:10.1051/e3sconf/201910202005
13. Ivanov, O., Avdeuk, O., Bushmeleva, K., Ivanov, I., Uvaysov, S.: Model for calculating the reliability of a wireless sensor telecommunication system for monitoring the gas transmission network state. Paper presented at the Moscow Workshop on Electronic and Networking Technologies, MWENT 2018 - Proceedings, 1-5 March 2018. doi:10.1109/MWENT.2018.8337229
14. Kharlamova, G., Stavytskyy, A.: Economic Modeling of Energy Security: Simulation of Economic Processes (Case of Ukraine Gas system). CEUR Workshop Proceedings **2104**, 128–143 (2018)
15. Khvostina, I., Havadzyn, N., Horal, L., Yurchenko, N.: Emergent Properties Manifestation in the Risk Assessment of Oil and Gas Companies. CEUR Workshop Proceedings **2422**, 157–168 (2019)
16. Luo, Z., Xi, Y., Wang, H.: Research on comprehensive quantitative risk-assessment of urban nature gas pipeline. Paper presented at the International Conference on Management Science and Engineering - Annual Conference Proceedings, 2196-2202

(2013). doi:10.1109/ICMSE.2013.6586568
17. Manian, L., Hodgdon, A.: Pipeline integrity assessment and management. Materials Performance **44**(2), 18–22 (2005)
18. Oliinyk, A., Chernova, O.: Estimation of gas losses based on the characteristics of the state of wells of Dashava storage. Eastern-European Journal of Enterprise Technologies 6(8-90), 25 – 32 (2017). doi:10.15587/1729-4061.2017.116806
19. Oliinyk, A., Robur, L., Yarovoj, K.: Determination of stress-deformation state of vibrating constructions by fiber laser Doppler anemometer. III International Conference of Vibration Measurements by Laser Techniques: Advances and Applications, Ancona, Italy, Proc. SPIE **3411**, 404–407 (1998)
20. Oliinyk, V., Kozmenko, O.: Optimization of investment portfolio management. Serbian Journal of Management **14**(2), 373–387 (2019). doi:10.5937/sjm14-16806
21. Oliynyk, A.: Mathematical models for the process of quasi-stationary deformation of pipeline and industrial systems when changing their spatial configuration. IFNTUOG, Ivano-Frankivsk (2010)
22. On the National Security Strategy of Ukraine. https://zakon.rada.gov.ua/laws/show/n0008525-15 (2015). Accessed 17 Aug 2015
23. Popadynets, I., Andrusiv, U., Shtohryn, M., Galtsova, O.: The effect of cooperation between universities and stakeholders: Evidence from Ukraine. International Journal of Data and Network Science **4**(2), 199-212 (2020). doi:10.5267/j.ijdns.2020.1.001
24. Sharples, J.D.: The International Political Economy of Eastern European Energy Security: Russia, Ukraine, and the European Union. In: Davidson, D.J., Gross, M. (eds.) Oxford Handbook of Energy and Society, pp. 137–164. Oxford University Press, Oxford (2018). doi:10.1093/oxfordhb/9780190633851.013.7
25. Shmygol, N., Łuczka, W., Trokhymets, O., Pawliszczy, D., Zavgorodniy, R.: Model of diagnostics of resource efficiency in oil and gas sector of economy of Ukraine. E3S Web of Conferences **166**, 13005 (2020). doi:10.1051/e3sconf/202016613005
26. Voynarenko, M., Dykha, M. V., Mykoliuk, O., Yemchuk, L., Danilkova, A.: Assessment of an enterprise's energy security based on multi-criteria tasks modeling. Problems and Perspectives in Management **16**(4), 102–116 (2018). doi:10.21511/ppm.16(4).2018.10
27. Yu, W., Gong, J., Song, S., Huang, W., Li, Y., Zhang, J., Hong, B., Zhang, Y., Wen, K., Duan, X.: Gas supply reliability analysis of a natural gas pipeline system considering the effects of underground gas storages. Applied Energy **252**, 113418 (2019). doi: 10.1016/j.apenergy.2019.113418
28. Zablodska, I., Akhromkin, Y., Akhromkin, A., Bielousova, L., Litvinova, I.: World experience in public administration of the transformation of energy-dependent regions in the context of their sustainable development. Problemy Ekorozwoju **15**(2), 235–244 (2020)
29. Zelinska, H., Andrusiv, U., Simkiv, L.: Knowledge economy: Trends in the world and analysis of Ukraine. Journal of Eastern European and Central Asian Research **7**(1), 104–113 (2020). doi:10.15549/jeecar.v7i1.325
30. Zelinska, H., Fedorovych, I., Andrusiv, U., Yurchenko, N.: Modeling of the Gas Transmission Reliability as a Component of Economical Security of Ukrainian Gas Transmission System. Advances in Economics, Business and Management Research **99**, 127–132 (2019). doi:10.2991/mdsmes-19.2019.25

Machine learning approaches for financial time series forecasting

Vasily Derbentsev[1[0000-0002-8988-2526]], Andriy Matviychuk[1[0000-0002-8911-5677]], Nataliia Datsenko[1[0000-0002-8239-5303]], Vitalii Bezkorovainyi[1[0000-0002-4998-8385]] and Albert Azaryan[2[0000-0003-0892-8332]]

[1] Kyiv National Economic University named after Vadym Hetman, 54/1 Peremohy Ave., Kyiv, 03057, Ukraine
derbv@kneu.edu.ua, editor@nfmte.com, d_tashakneu@ukr.net, retal.vs@gmail.com
[2] Kryvyi Rih National University, 11 Vitalii Matusevych Str., Kryvyi Rih, 50027, Ukraine
azaryan325@gmail.com

Abstract. This paper is discusses the problems of the short-term forecasting of financial time series using supervised machine learning (ML) approach. For this goal, we applied several the most powerful methods including Support Vector Machine (SVM), Multilayer Perceptron (MLP), Random Forests (RF) and Stochastic Gradient Boosting Machine (SGBM). As dataset were selected the daily close prices of two stock index: SP 500 and NASDAQ, two the most capitalized cryptocurrencies: Bitcoin (BTC), Ethereum (ETH), and exchange rate of EUR-USD. As features we used only the past price information. To check the efficiency of these models we made out-of-sample forecast for selected time series by using one step ahead technique. The accuracy rates of the forecasted prices by using ML models were calculated. The results verify the applicability of the ML approach for the forecasting of financial time series. The best out of sample accuracy of short-term prediction daily close prices for selected time series obtained by SGBM and MLP in terms of Mean Absolute Percentage Error (MAPE) was within 0.46-3.71 %. Our results are comparable with accuracy obtained by Deep learning approaches.

Keywords: financial time series, short-term forecasting, machine learning, support vector machine, random forest, gradient boosting, multilayer perceptron.

1 Introduction

Forecasting financial tine series have been in focus of researchers for a long time. This topic continues to be relevant from both theoretical and applied points of view. Brokers, financial analysts and traders make daily decisions about buying and selling various financial assets, including currency, stocks, bonds and others. To reduce the risk of such transactions and to obtain the expected return on their investments, each of them must analyze a number of factors that affect market conditions and generate upward or downward trends.

In this regard, the problem of developing adequate forecasting approaches is relevant to the scientific community as well as to financial analysts, investors and traders.

There are two main approaches to solving the problem of forecasting financial assets. The first one is to construct a casual model that describes the relationship between the asset's value and other macroeconomic factors. This approach was implemented within the framework of fundamental analysis and based on different mathematical tools, such as econometric modeling and systems of differential equations [13; 17; 28].

Another approach is based on the analysis of past observations selected asset and used variety of technical indicators and oscillators that help predict market trends. This approach has been realized in technical analysis which is actively used now in addition to time series analyses [7; 13; 28]. Within time series framework has been developed manifold class of linear and nonlinear approaches, such as ARIMA-GARCH models [6; 27].

Recent time the methods and algorithms of Machine Learning (ML) which have developed within Data Science paradigm [14; 36] ML have been also applied to forecasting financial and economic time series [2; 12], and various automated trading systems (bots) built on these algorithms began to be used for trading. Results of numerous empirical studies have shown that ML approaches outperform time series models in forecasting different financial assets [10; 18; 22; 26; 31; 38].

The main advantage of ML is that the algorithms themselves interpret the data, so we don't need to perform their initial decomposition. Depending on the purpose of the analysis, these algorithms themselves build the logic of modeling on the basis of available data.

This avoids the complex and lengthy pre-model stage of statistical testing of various hypotheses about studied process. The main hypothesis, in particular, in terms of the purpose of our study, is only the thesis of the ability of ML methods to effectively analyze the financial time series, to identify hidden patterns and time correlations, which are the basis for making qualitative short-term forecasts.

The main goal of our paper is to compare the predictive properties of the most efficient ML algorithms: Artificial Neural Network (ANN), Support Vector Machine (SVM), Random Forest (RF) and Gradient Boosting Machine (GBM) for short-term forecast financial time series (stock indices, currencies and cryptocurrencies). At the same time, as predictors (features) we used only the past values of the studied time series. Our main assumption is that ML methods be able to extract latent patterns from the data, which allows us to make more efficient predictions.

This paper is organized as follows: in Section 2 is represented brief literature review devoted using ML approaches in the field of financial time series forecasting. Section 3 is described the main concept of applied methods. Data description and empirical results are given in Section 4. Concluding remarks and future perspectives are given in Section 5.

2 Brief review recent studies

It should be noted that financial time series forecasting have been studied for a long time. Since, ML approaches proved their efficiency in many areas and became popular; they have been widely used for research financial time series. Numerous articles in scientific journals, reviews, conferences and internet resources are devoted to this topic.

Last five years researchers basically have been focused attention on novel network ML approaches Deep Learning (DL), which includes asset of powerful methods, such as Recurrent Neural Network (RNN), Long-Short Time Memory (LSTM), and Convolutional Neural Network (CNN) and so on [22; 24; 32; 34; 38]. Recently was published detailed overview devoted to using DL approaches in the field of financial forecasting [24]. The main finding this survey is that generally DL framework outperforms time series models and often shows higher accuracy than traditional ML algorithms.

The key advantage of DL models is very powerful in feature learning and selection of input data using a general-purpose learning procedure. But DL models have such disadvantage that it takes much more time to train them, besides, it is a nontrivial problem is tuning hyperparameters. At the same time, traditional ML models often show comparable accuracy in time series forecasting.

As for using ML algorithms in financial forecasting, the most common are Neural Networks (ANNs) of various architecture [1; 8; 10; 20; 21; 33; 38], Support Vector Machines (SVM) [23; 25; 29; 30; 35], and Fuzzy Logic (FL) [25; 39].

The application of these approaches for forecasting task has shown their efficiency for both traditional financial assets [1; 8; 18; 20; 21; 23; 24; 29; 30; 35] and cryptocurrencies [8; 26; 31; 38].

Several studies [1; 8; 20; 21] presented the results that ANNs have better predictive properties then other ML approaches for forecasting financial time series. At the same time, there are a number of research papers (see, for example, Okasha, [29]; Sapankevych and Sankar, [30]; Hitam and Ismail, [19]), which presented results that SVMs have also been proven to outperform other non-linear techniques including neural-network based non-linear prediction techniques such as multi-layer perceptron (MLP).

It should be noted, that much less attention has been paid to another powerful class of ML approaches of designing ensembles Classification and Regression Trees (C&RT): Random Forest (RF) [4; 5] and Gradient Boosting Machine (GBM) [15; 16], which used bagging (RF) and boosting (GBM) technique. Both RF and GBM are powerful methods that can efficient capture complex nonlinear patterns in data.

Thus, Varghade and Patel [35] tested RF and SVM to forecasting stock market index S&P CNX NIFTY. They noted that the Decision Trees model outperforms the SVR, although RF at times is found to overfit the data.

Kumar and Thenmozh [23] explored set of classification models for predicting direction of index S&P CNX NIFTY. Their empirical results suggest that both the SVM and RF outperforms the other classification methods (NN, Linear Discriminant Analysis, Logit), in terms of predicting the direction of the stock market movement, but at the same time SVM it turned out to be more accurate.

Recently there have been appeared several papers devoted to applying ensembles approaches for forecasting cryptocurrency prices [3; 9; 11]. Borges and Neves [3] tested four ML algorithms for prediction price trend: LR, RF, SVM and GBM. All learning algorithms outperform the Buy and Hold investment strategy in cryptomarket. The best result was obtained by ensembles voting (accuracy 59.3%).

Chen et al. [9] applied a set of learning models including RF, XGBoost, Quadratic Discriminant Analysis, SVM and LSTM for Bitcoin 5-minute interval and daily prices. Authors used wide dataset including as features technological, market and trading, socio-media and fundamental factors. Somewhat unexpected was that for daily prices better results were obtained by using statistical methods (average accuracy 65%) unlike ML methods (average accuracy 55.3%). Among the best ML the SVM was the best, with an accuracy of 65.3%.

3 Methodology

In this paper we have been applied supervised ML technique for forecasting financial time series. Consider a sample of pairs of features $\boldsymbol{x} = (x_1, x_2, \dots, x_p, \dots, x_n)$ and the labels y: $(\boldsymbol{x}_i, y_i)_{i=1,2,\dots,n}$ length n. In our case labels (or target) are values of selected financial assets, and features are only lagged daily values these assets $y_{i-1}, y_{i-2}, \dots, y_{i-p}, i > p$.

Our main goal is to predict future value of target variable on the next time period (next day since we used daily quotes) by using several ML approaches (SVM, ANN, RF and GBM) and compare their forecasting performance.

Thus, our task is to construct some functional (regression) or rule-based (decision tree based) dependence of the form

$$\hat{y}_{n+1} = f(\boldsymbol{x}_j, \boldsymbol{w}_j), \tag{1}$$

where $\boldsymbol{x}_j = (x_i)_{i=1}^n, j = 1,2, \dots, k$ are vectors of features; $\boldsymbol{w}_j$ – weights of the features, n – total number of samples in dataset; k – number of features.

3.1 Support Vector Machine (SVM)

The support vector machine (SVM) is an extension of the support vector classifier that results from enlarging the feature space in a specific way by using kernels function. The main idea of the SVM method is to map the original vectors into a space of a higher dimension and search for a separating hyperplane with a maximum margin in this space. Two parallel hyperplanes are constructed on both sides of the hyperplane separating the classes. The separating hyperplane will be the hyperplane that maximizes the distance to two parallel hyperplanes. The algorithm works under the assumption that the lager difference or distance between these parallel hyperplanes (margine) provides the smaller average error of the classifier.

Support Vector Regression (SVR) is the regression process performed by SVM which tries to identify the hyperplane that maximizes the margin between two classes and minimize the total error. In order for an efficient SVM to be constructed, a penalty of complexity is also introduced, balancing forecasting accuracy and computational performance.

Unlike classic regression problem SVR seeks coefficients that minimize a different type of loss, where only residuals larger in absolute value than some positive constant contribute to the loss function. This is an extension of the margin used in support vector classifiers to the regression setting.

The mathematical formalization of SVR is reduced to the following. Let's regression equation is written in the form

$$a(\boldsymbol{x}) = \langle \boldsymbol{x}, \boldsymbol{w} \rangle - w_0, \tag{2}$$

where $\langle \cdot,\cdot \rangle$ – is operator of inner product; w_0 is a constant.

Then the problem is reduced to minimizing functional:

$$\langle \boldsymbol{w}^T, \boldsymbol{w} \rangle + \frac{1}{2C} \sum_{i=1}^{l} (|\langle \boldsymbol{w}, x_i \rangle - w_0 - y_i| - \varepsilon) \rightarrow \min_{w, w_0}, i = 1,2, \dots, l, \tag{3}$$

where C is the regularization parameter or penalty coefficient for incorrectly estimating the output associated with input vectors, which also controls the relationship between a smooth boundary; l is the number of samples in training set ($l < n$, as a rule $l \approx 0.7 \div 0.8n$); ε is the margin value.

After changing variables and some algebraic transformations loss function for SVR can be presented in such form:

$$\frac{1}{2} \langle \boldsymbol{w}^T, \boldsymbol{w} \rangle + C \sum_{i=1}^{l} (\xi_i^+ + \xi_i^-) \rightarrow \min_{w, w_0,, \xi_i^+,, \xi_i^-}, i = 1,2, \dots, l, \tag{4}$$

where $\xi_i^- = (-a(x_i) + y_i - \varepsilon)$, $\xi_i^+ = (a(x_i) - y_i - \varepsilon)$ are slack variables, that allow individual observations to be on the wrong side of the margin or the hyperplane; $K(x_i, x_j)$ is the kernel function. The most commonly used kernel functions are Linear, Polynomial, Gaussian, Radial Based Function (RBF) and so on.

Loss function (4) is minimized under condition

$$\begin{cases} y_i - \varepsilon - \xi_i^- \leq \boldsymbol{w}^T K(x_i, x_j) - w_0 \leq y_i + \varepsilon + \xi_i^+ \\ \xi_i^+, \xi_i^- \geq 0, i = 1,2, \dots, l \end{cases}. \tag{5}$$

The Lagrangian of this problem can be expressed in terms of the dual variables λ_i^+, λ_i^-, thus the regression equation on support vectors can be written in the such form:

$$a(\boldsymbol{x}) = \sum_{i=1}^{l} (\lambda_i^+ - \lambda_i^-) K(x_i, x_j) - w_0. \tag{6}$$

3.2 Artificial Neural Network (ANN)

ANNs are the most popular methods of ML. Numerous empirical studies show the efficiency of ANNs in the different fields both for classification and regression problem: pattern recognition, image and voice analysis, machine translation and so on. Several last decides they are widely used for analysis and forecasting financial time series. In [12; 22; 31] it was shown that ANNs have better predictive properties than time series models and other ML algorithms for financial time series forecasting problem.

In this paper we have been used network model of the most common architecture: Multi-Layer Perceptron (MLP) with three layers: input layer, one hidden layer and output layer with one neuron that represent target variable (predicted value). It should be noted that despite simple structure MLP be able to take into account complex patterns in data due using different nonlinear activation functions.

The network output depends on its configuration, weights and activation functions of neurons on the hidden and output layers:

$$\hat{y}_{n+1} = g\left(\sum_{i=1}^{k} w_i f\left(\sum_{j=1}^{p} \omega_{ji} y_{n-j+1} + b_i\right) + b_0\right), \tag{7}$$

where $f(\cdot), g(\cdot)$ – activation functions of neurons of the hidden and input layer, respectively; w_i – the weight of the connections between the i-th neuron of the hidden layer and the output of the network; ω_{ji} – the weight of the connections between the j-th neuron of the input and the i-th neuron of the hidden layers; b_0, b_i – bias neurons of the output and hidden layers.

Network learning consists in finding and setting the neurons weights (synaptic weights) which minimized difference between the target variable and the network output. The search of minimum of the loss function was performed by the gradient descent method, embodied in the back-propagation algorithm.

3.3 Gradient Boosting Machine (GBM)

Boosting is a procedure for sequentially building a composition of machine learning algorithms, when each of them seeks to compensate for the shortcomings of the composition of all previous algorithms. In contrast to bagging, boosting does not use simple voting but a weighted one. The major attractions of boosting are that it is easy to design computationally efficient weak classifiers (as a rule used shallow decision trees). Boosting over decision trees is considered one of the most efficient methods in terms of classification quality.

Gradient Boosting Machine method (GBM) was proposed Friedman [15; 16]. Commonly the basic steps of GBM are the next.

The final classifier $a_N(\boldsymbol{x})$ is constructed as a weighted sum of N basic algorithms $h_i(\boldsymbol{x}, \theta)$ (Decision Trees):

$$a_N(\boldsymbol{x}) = \frac{1}{N} \gamma_N \sum_{i=0}^{N} h_i(\boldsymbol{x}, \theta), \tag{8}$$

where θ is vector of adjusted parameters, γ_N is the weight coefficient.

Let's we choose the initial classifier $h_0(\boldsymbol{x}, \theta)$, for example, it may be the median or mean of the time series (target variable).

If we on the N–1 step have already built new classifier $a_{N-1}(\boldsymbol{x}_i, \theta)$, then we select the next basic algorithm $h_N(\boldsymbol{x})$ that reduced the error given by previous classifier as much, as possible:

$$\sum_{i=1}^{l}\left[L\left(y_i, a_{N-1}(\mathbf{x}_i,\theta)+\gamma_N h_N(\mathbf{x}_i,\theta)\right)\right] \to \min_{\gamma_N, h_N}, \tag{9}$$

where $L(\cdot)$ – is loss function.

We can select $h_N(\boldsymbol{x}_i.\theta)$ that minimized sum of squares deviations for all samples in training set

$$h_N(\mathbf{x},\theta) = \underset{h(\mathbf{x},\theta)}{\operatorname{argmin}} \sum_{i=1}^{l}(h(x_i,\theta) - s_i)^2, \tag{10}$$

where s_i is deviations that equal to the anti-gradient of the loss function $L(\cdot)$.

In this way we perform predictions for samples in the training set by using gradient descent in the *l*-dimensional space.

If a new basic algorithm has been found, it is possible to select its coefficient γ_N by analogy with the gradient descent:

$$\gamma_N = \underset{\gamma}{\operatorname{argmin}} \sum_{i=1}^{l} L\left(y_i, a_{N-1}(\mathbf{x}_i,\theta)+\gamma_N h_N(\mathbf{x}_i,\theta)\right), \tag{11}$$

It should be note, that boosting usually does not result the overfitting problem because shallow decision trees are used. These trees have a large bias, but are not inclined to overfitting.

The effective ways to solve this problem is to reduce the step: instead of moving to the optimal direction of the anti-gradient, a shortened step can be taken by

$$a_N(\boldsymbol{x},\theta) = a_{N-1}(\boldsymbol{x},\theta) + \lambda\left(\gamma_N h_N(\boldsymbol{x})\right), \tag{12}$$

where $\lambda \in [0,1]$ is the learning rate.

3.4 Random Forest (RF)

The main concept of the RF is that a composition of weak classifiers can give good results for both classification and regression problems. Proposed by Breiman [4; 5] in 1996 the RF is based on bagging technique (bootstrap aggregation) over decision trees. Bagging reduces the variance of the base algorithms if they are weakly correlated. In RF the correlation between trees is reduced by randomization in two directions.

Firstly, each tree is trained on a bootstrapped subset. Secondly, the feature by which splitting is performed in each node is not selected from all possible features, but only from their random subset of size m. The main distinction between bagging and RF is the choice of these features subset. RF works well when all of the features are at least marginally relevant, since the number of features selected for any given tree is small. Using a small value of m will typically be helpful when we have a large number of correlated predictors.

The RF algorithm generates each of the N trees independently, which makes it very easy to parallelize. For each tree, it constructs a full binary tree of maximum depth. The main concept is that classifiers (trees) do not correct each other's mistakes, but compensate for them when voting. Basic classifiers should be independent and they can be based on different groups of methods or trained on independent datasets. Bagging allows us to reduce prediction error in the case when the variance of the error base method is high.

Thereby efficiency of RF performance is achieved even though some trees will query on useless features and make random predictions. But some of the trees will happen to query on good features and will make good predictions (because the leaves are estimated based on the training data).

If we have enough trees, the random ones will wash out as noise, and only the “good” trees will have an effect on the final result (classification or prediction).

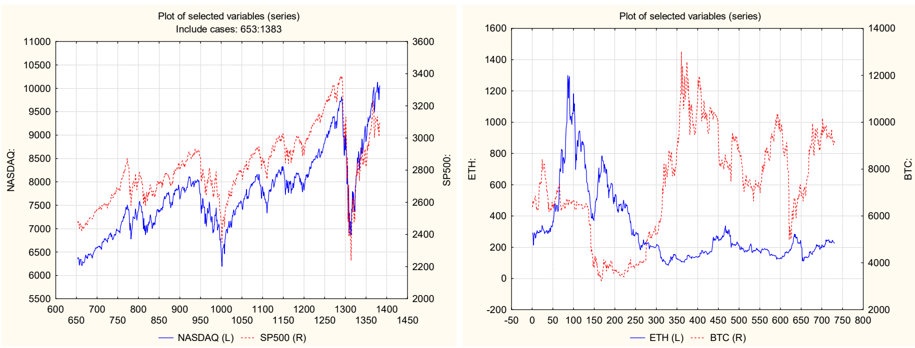
4 Empirical results

4.1 Dataset

To reduce our analysis to the most popular financial assets, we were used daily close prices two stock index: Nasdaq and SP&500, two most capitalized cryptocurrencies: Bitcoin (BTC), Ethereum (ETH), and exchange rate EUR USD. Our initial dataset covers the period from 01/01/2015 to 30/06/2020 for all series (for ETH from 06/08/2020) according to the Yahoo Finance [37].

So, our dataset includes 1384 observations for Nasdaq, 1383 for SP&500, 1434 for exchange rate EUR USD, 2008 for BTC and 1278 for ETH.

It should be noted that selected time series during this period had different type of dynamics due to we can better estimate forecasting performance for ML approaches (see fig. 1).



(a) (b)

Fig. 1. Dynamics of traditional assets (a) and cryptocurrencies (b) from 30/06/2018 to 30/06/2020.

On purpose of training models, fitting and tuning their parameters dataset was divided into the training and test subsets in the ratio of 80% and 20%. Moreover, the last 100 observations (from 22/03/2020 to 30/06/2020) were reserved for validation which was performed by out-of-sample one-step ahead forecast.

Since we focus on ML approach of forecasting financial time series data, the main purpose of our paper is to get the most accurate one-step ahead forecast of daily prices, based on only their past value.

According to some empirical studies devoted forecasting financial time series, there is a seasonal lag which is a multiple of 5 if we use daily observations and a multiple of 7 for cryptocurrencies because the fact that cryptocurrencies are traded 24/7.

For stabilization variance all features were taken in to natural logarithm. This is special case of Box-Cox transform.

4.2 Hyper-parameters tuning

It should be noted that hyper-parameters tuning is an important and sophisticated step of the model design. First of all, it is necessary to choose the functional form of the loss function. In the point of main purposes of our study the quadratic loss, which generally used for solving the regression problem, was selected.

According to our hypothesis regarding lag length as MLP models, we tested the following architectures:

- 7 inputs and from 5 to 14 hidden layer neurons for cryptocurrencies;
- 5 inputs and from 5 to 10 hidden layer neurons.

The most common functions such as logistic, hyperbolic tan, exponential and ReLu were tested as activation functions. Training MLP for each time series and different lag values (number of input neurons) was conducted over 100 epochs, of which the best 5 architectures were selected for each case (in terms of minimum PE error on the test sample and matching the model residuals to normal distribution).

The final prediction for each asset was obtained as the prediction of the ensemble of networks, that is, average of the best 5 corresponding MLP models.

For SVM models we have chosen RBF as a kernel which is the best for regression problem. Regularization parameter was estimated by the greed search in the range from 1 to 15 and it was selected C=10.

Both of tree-based methods (RF, and GBM) based on partitions the data into training and testing sets by randomly selecting cases. We applied in this study stochastic modification of GBM (SGBM) which based on such partition. The training sample is used to fitting models by adding simple trees to ensembles. Testing set is used to validate their performance. For regression tasks validation is usually measured as the average error. We select 30% of the dataset as test cases for both approaches.

Since the RF is not inclined to overfitting, one can choose a large number of trees for the ensemble. We designed RF model with 500 trees. At the same time, in order for the model to be able to describe complex nonlinear patterns in data, it is necessary to use complex trees. So, we have been chosen 15 the maximum number of levels.

Other important parameter for RF is the number of features to consider at each split. As noted in the Section 3.2 it is recommended to choose this value as $m \approx \frac{M}{3}$ (where M is the total number of features) for regression task. We tested different RF models with value m within 8 to 12.

As stop condition for number of trees in SGBM (boosting steps) we took the number of trees at which the error on the test stops decreasing. This is necessary in order to avoid the overfitting. For boosting, unlike the RF, the simple trees are usually used. That's why we fitted maximum number of levels in trees and number of terminal nodes by the criteria of lowest average squared error on both training and test samples.

For GBM an important parameter is a learning rate (shrinkage). Regularization by shrinkage consists in modifying the update rule (12) by tuning λ. We selected this value on the grid search according to minimum prediction error on the test set. The final values of hyper-parameters setting are reported in table 1.

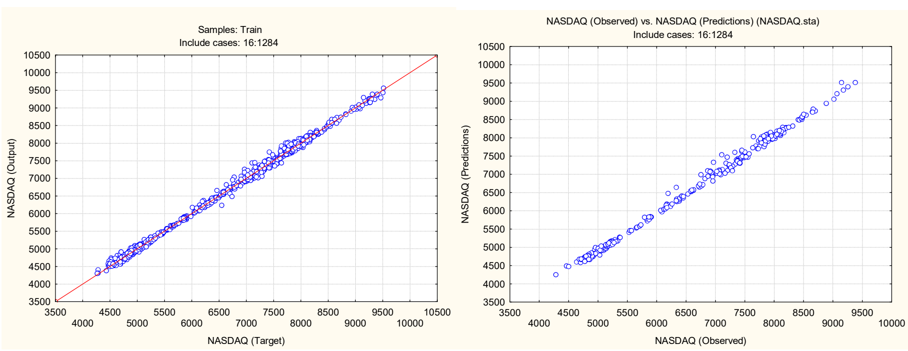
Table 1. Final hyper-parameters setting for RF and SGBM.

Parameters	**RF**	**GBM**
Loss-function	quadratic	quadratic
Training / test subsamples proportion, %	70/30	70/30
Random subsample rate	0.7	0.7
Maximum number of trees in ensemble	500	400
Maximum number of levels in trees	10	5
Maximum number of features to consider at each split	12	-
Maximum number of terminal nodes in trees	150	15
Minimum samples in child nodes	5	-
Learning rate (shrinkage)		0.1

4.3 Forecasting performance

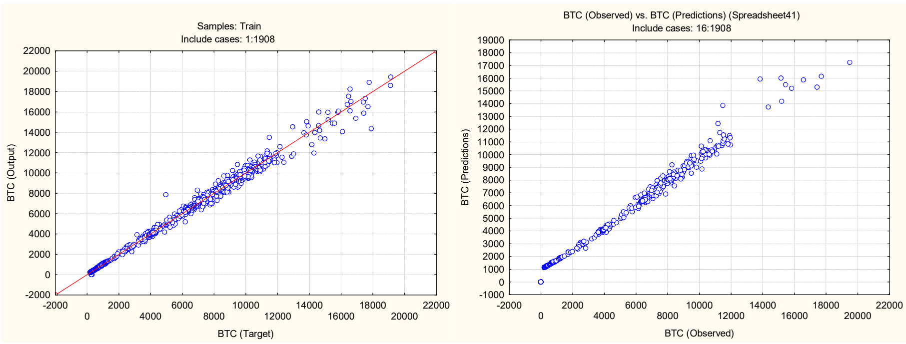
The short-term forecasts for selected time series were made for absolute values of prices (log prices). The target variable is the prediction the value of close prices for each series in the next time period (day) although we used daily observation. All models were trained with the same set of features.

On figures 2-3 were shown quality models fitting for BTC, and NASDAQ obtained by MLP and SVM. figures 4-5 presented results for RF, and SGBM respectively.



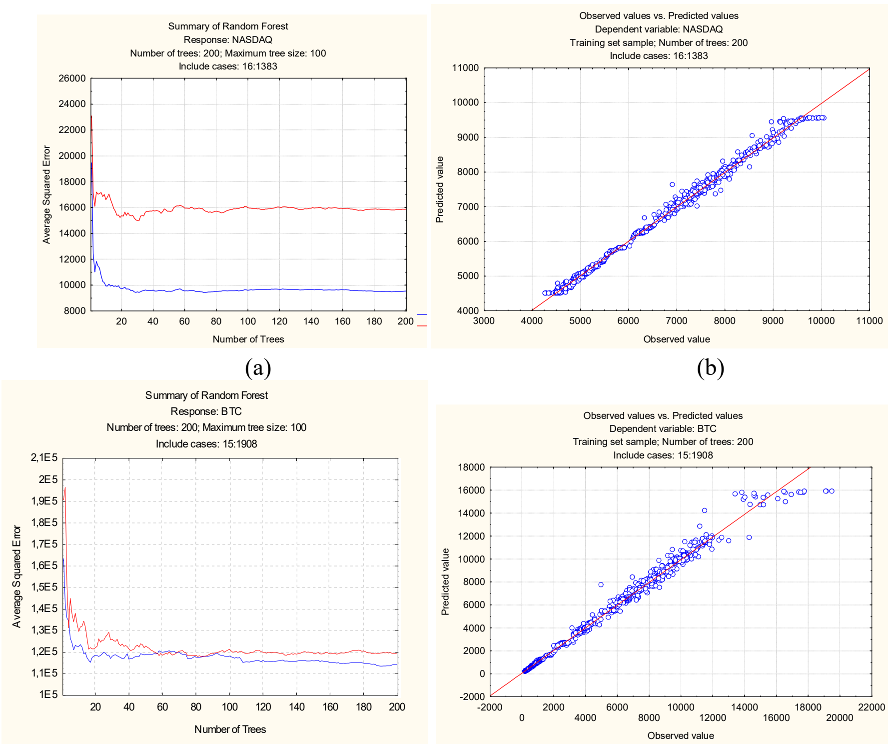
(a) (b)

Fig. 2. Fitting accuracy on the training and test subsets for NASDAQ: (a) MLP, (b) SVM.



(a) (b)

Fig. 3. Fitting accuracy on the training and test subsets for BTC: (a) MLP, (b) SVM.



(a) (b)

(c) (d)

Fig. 4. Fitting accuracy on the training and test subsets for RF: (a, b) NASDAQ, (c, d) BTC.

These graphs characterize the dependence of the predicted values (vertical axis) on the actual data (horizontal axis) on the test set and allow us to visually determine the quality of the fitting.

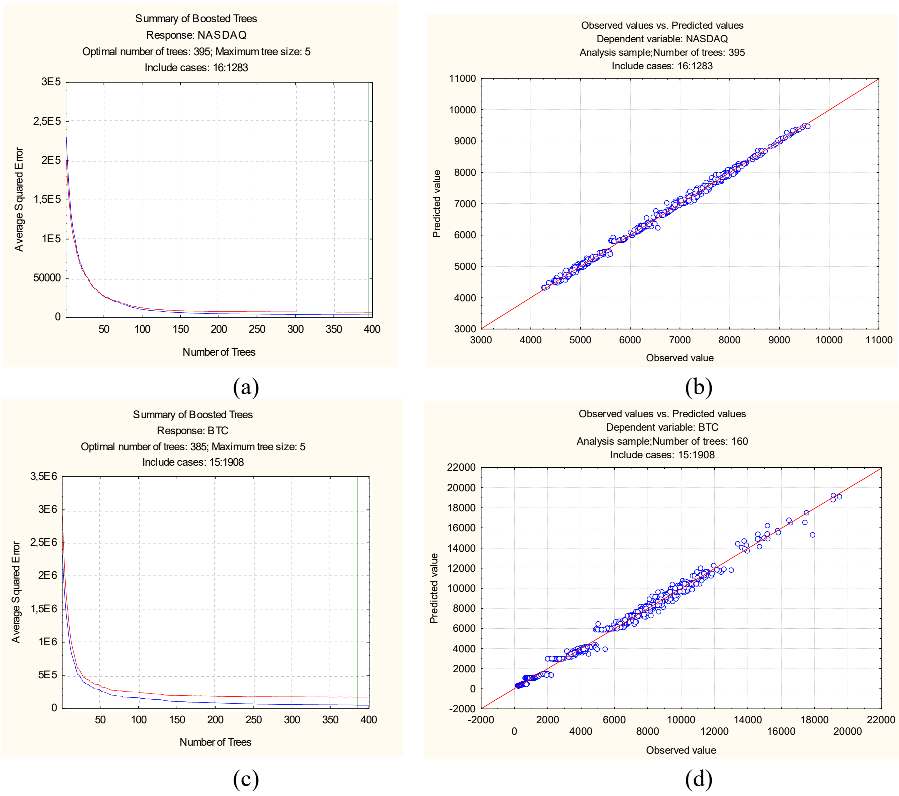


Fig. 5. Fitting accuracy on the training and test subsets for SGBM: (a, b) NASDAQ, (c, d) BTC.

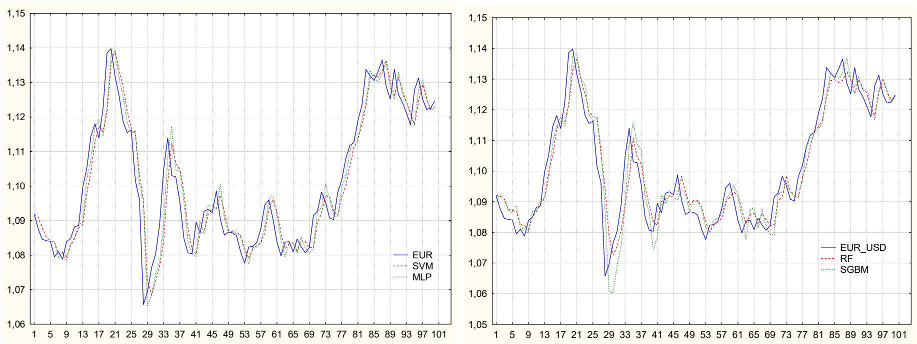
Figures 4-5 shows both the dependence of the predicted values (vertical axis) on the actual data (horizontal axis) (graphs (b, d)) and dependence of models fitting quality on number of trees in ensemble for RF (fig. 4) and SGBM (fig. 5) on the training and test subsets (graphs (a, c)) for selected assets (NASDAQ, BTC).

Forecasting results for last 100 observations (hold-out dataset) for all models and selected time series are shown on fig. 6-9. On fig. 6-10 (a) represented results obtained by SVM and MLP, on fig. 6-10 (b) results for RF and SGBM.

Analysis of the graphs allows us to conclude that SGBM and MLP well approximate time series dynamics, but one can see a certain delay in the model graphs in comparison to real data. RF and SVM showed good approximation not for all time series.

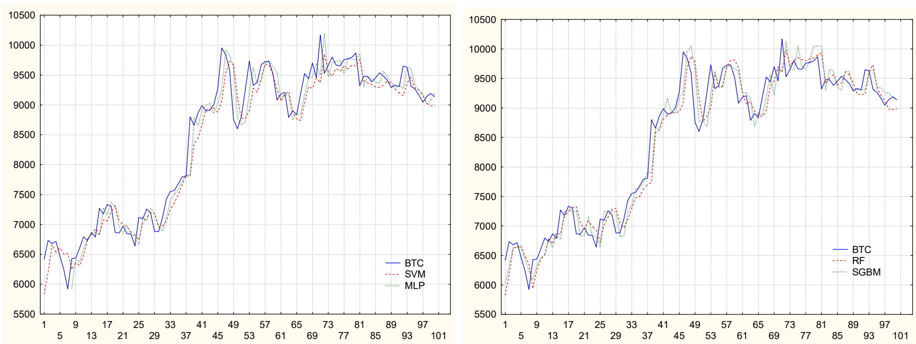
Summary accuracy results in terms of MAPE and RMSE metrics are shown in table 2.

Thus, we can conclude MLP, SVM and SGBM methods have the same order of accuracy for the out-of-sample dataset prediction, although boosting also was somewhat more accurate. The best prediction performance is produced by SGBM for EUR-USD – 0.46 % (MAPE), and the best result for NASDAQ also provided SGBM – 2.38%. For BTC better performance shown SVM but MLP outperformed other models for SP&500.



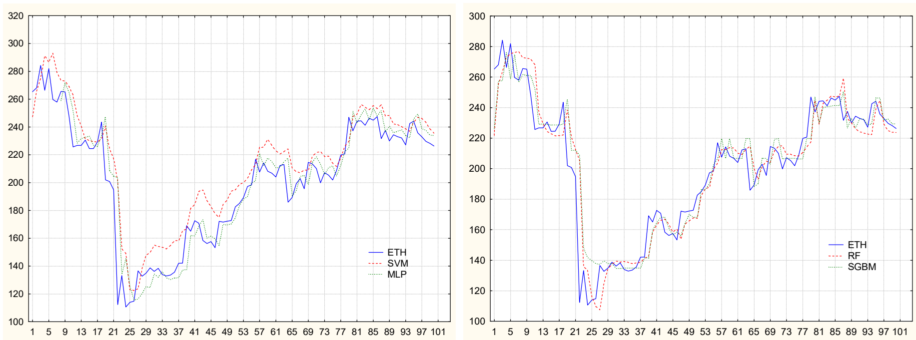
(a) (b)

Fig. 6. Out of sample prediction EUR/USD.



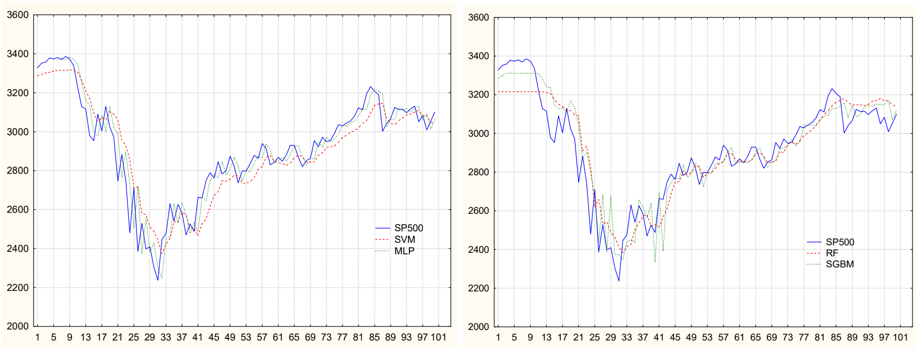
(a) (b)

Fig. 7. Out of sample prediction BTC /USD.



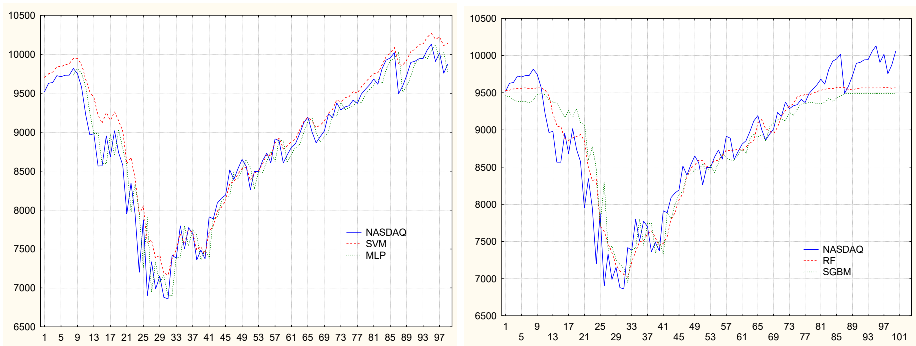
(a) (b)

Fig. 8. Out of sample prediction ETH/USD.



(a) (b)

Fig. 9. Out of sample prediction S&P 500.



(a) (b)

Fig. 10. Out of sample prediction NASDAQ.

Table 2. Out-of-sample accuracy forecasting performance results.

	SGBM		RF		SVM		MLP	
	MAPE, %	RMSE	MAPE, %	RMSE	MAPE, %	RMSE	MAPE, %	RMSE
EUR/USD	0.46	0.0656	0.47	0.0067	**0.40**	0.0065	0.45	0.0067
BTC/USD	2.44	283.6	2.65	321.8	**1.03**	106.5	2.25	274.5
ETH/USD	**5.09**	15.6	5.17	15.89	8.36	260.4	5.25	14.83
S&P 500	2.54	97.99	2.85	108.9	2.91	106.5	**2.35**	91.2
NASDAQ	**2.38**	289.3	2.51	289.1	2.77	340.3	2.23	257.6

It should be noted that our results are comparable with accuracy obtained by Deep learning approaches [24]. Therefore, using both tree-based ensembles, ANNs and SVM are powerful enough forecasting tools for financial time series.

5 Conclusion and discussion

Our research has shown efficiency of using ML approaches to predicting financial time series. According to our results, the out of sample accuracy of short-term forecasting daily quotes obtained by SGBM has the same rate as MLP and SVM. In terms of MAPE for selected time series it was within 0.46-5.9 %. Moreover, for NASDAQ and ETH SGBM outperformed other approaches. The worse results were obtained by RF which was used as a baseline.

At the same time all models showed the worst results for ETH, accuracy rate (MAPE) turned out in the range from 5.9 (SGBM) to 8.38 % (RF).

By designing models, we explored different sets of features: from 5 to 15 lags of target variable (from 7 to 14 for cryptocoins). Our final dataset contained only past values of target variable with 14 and 15 lag depth. In this case larger dataset provided better training for all models and given more efficient results.

It should be noted that we used a minimal dataset - only lag values of the studied series (closing prices). In our opinion, forecasting accuracy can be improved by including additional features, for example, open, max, min and average prices, fundamental variables, different indicators and oscillators, such as, Price rate-of-change, Relative strength index, and so on.

Future research should extend by investigating of the prediction power of described ML approaches by using additional features. In the conclusion, we note that the proposed methodology by the development of combined ensemble of C&RT with other powerful ML models, such as NN and SVM is a promising approach to forecasting financial time series. Moreover, it seems to us promising to use DL approaches for features selection and making prediction.

References

1. Bahrammirzaee, A.: A comparative survey of artificial intelligence applications in finance: artificial neural networks, expert system and hybrid intelligent systems. Neural Computing and Applications **19**(8), 1165–1195 (2010). doi:10.1007/s00521-010-0362-z
2. Bontempi, G., Taieb S., Borgne Y.: Machine Learning Strategies for Time Series Forecasting. Business Intelligence. Lecture Notes in Business Information Processing **138**, 62–77 (2013). doi:10.1007/978-3-642-36318-4_3
3. Borges, T.A., Neves, R.N.: Ensemble of Machine Learning Algorithms for Cryptocurrency Investment with Different Data Resampling Methods. Applied Soft Computing **90**, 106187 (2020). doi:10.1016/j.asoc.2020.106187
4. Breiman, L., Friedman, H., Olshen, R.A., Stone, C.J.: Classification and Regression Trees. Chapman & Hall/CRC, Boca Raton (1984)
5. Breiman, L.: Random Forests. Machine Learning **45**, 5–32 (2001). doi:10.1023/A:1010933404324
6. Brockwell, P.J., Davis, R.A.: Introduction to Time Series and Forecasting, 3rd edn. Springer International Publishing, New York (2016). doi:10.1007/978-3-319-29854-2
7. Brooks, A.: Trading Price Action Trends: Technical Analysis of Price Charts Bar by Bar for the Serious Trader. John Wiley & Sons, New Jersey (2012)

8. Caporale, G.M., Plastun, A., Oliinyk, V.: Bitcoin fluctuations and the frequency of price overreactions. Financial Markets and Portfolio Management **33**, 109–131 (2019). doi:10.1007/s11408-019-00332-5
9. Chen, Z., Li, C., Sun, W.: Bitcoin price prediction using machine learning: An approach to sample dimension engineering. Journal of Computational and Applied Mathematics **365**, 112395 (2020). doi:10.1016/j.cam.2019.112395
10. Derbentsev, V., Datsenko, N., Stepanenko, O., Bezkorovainyi, V.: Forecasting Cryptocurrency Prices Time Series Using Machine Learning. CEUR Workshop Proceedings **2422**, 320–334 (2019)
11. Derbentsev, V., Matviychuk, A., Soloviev, V.N.: Forecasting of Cryptocurrency Prices Using Machine Learning. In: Pichl, L., Eom, C., Scalas, E., Kaizoji, T. (eds.) Advanced Studies of Financial Technologies and Cryptocurrency Markets, pp. 211–231. Springer, Singapore (2020). doi:10.1007/978-981-15-4498-9_12
12. Di Persio, L., Honchar O.: Multitask Machine Learning for Financial Forecasting. International Journal of Circuits, Systems and Signal Processing **12**, 444–451 (2018)
13. Eiamkanitchat, N., Moontuy, T., Ramingwong, S.: Fundamental analysis and technical analysis integrated system for stock filtration. Cluster Computing **20**, 883–894 (2017). doi:10.1007/s10586-016-0694-2
14. Flach, P.: Machine Learning: The Art and Science of Algorithms that Make Sense of Data. Cambridge University Press, Cambridge (2012)
15. Friedman, J.H.: Greedy Function Approximation: A Gradient Boosting Machine. The Annals of Statistics **29**(5), 1189–1232 (2001)
16. Friedman, J.H.: Stochastic Gradient Boosting. Computational Statistics & Data Analysis **38**(4), 367–378 (2002). doi:10.1016/S0167-9473(01)00065-2
17. Glabadanidis, P.: Market Timing and Moving Averages: An Empirical Analysis of Performance in Asset Allocation. Palgrave Macmillan, New York (2015). doi:10.1057/9781137359834
18. Hamid, S.A., Habib, A.: Financial Forecasting with Neural Networks. Academy of Accounting and Financial Studies Journal **18**(4), 37–55 (2014)
19. Hitam, N.A., Ismail, A.R.: Comparative Performance of Machine Learning Algorithms for Cryptocurrency Forecasting. Indonesian journal of electrical engineering and computer science **11**(3), 1121–1128 (2018). doi:10.11591/ijeecs.v11.i3.pp1121-1128
20. Kara, Y., Boyacioglu, M., Baykan, Ö.K.: Predicting direction of stock price index movement using artificial neural networks and support vector machines: The sample of the Istanbul Stock Exchange. Expert Systems with Applications **38**(5), 5311–5319 (2011). doi:10.1016/j.eswa.2010.10.027
21. Kourentzes, N., Barrow, D.K., Crone, S.F.: Neural network ensemble operators for time series forecasting. Expert Systems with Applications **41**(9), 4235-4244 (2014). doi:10.1016/j.eswa.2013.12.011
22. Kumar, D., Rath, S.K.: Predicting the Trends of Price for Ethereum Using Deep Learning Technique. In: Dash, S., Lakshmi, C., Das, S., Panigrahi, B. (eds.) Artificial Intelligence and Evolutionary Computations in Engineering Systems. Advances in Intelligent Systems and Computing, vol. 1056, pp. 103–114. Springer, Singapore (2020). doi:10.1007/978-981-15-0199-9_9
23. Kumar, M., Thenmozhi, M.: Forecasting Stock Index Movement: A Comparison of Support Vector Machines and Random Forest. Paper presented at Indian Institute of Capital Markets 9th Capital Markets Conference (2006). doi:10.2139/ssrn.876544

24. Liu, W., Wang, Z., Liu, X., Zeng, N., Liu, Y., Alsaadi, F.E.: A survey of deep neural network architectures and their applications. Neurocomputing **234**, 11-26 (2017). doi:10.1016/j.neucom.2016.12.038
25. Matviychuk, A.: Fuzzy logic approach to identification and forecasting of financial time series using Elliott wave theory. Fuzzy Economic Review XI(2), 51–68 (2006). doi:10.25102/fer.2006.02.04
26. McNally, S., Roche, J., Caton, S.: Predicting the price of Bitcoin using Machine Learning. In: 2018 26th Euromicro International Conference on Parallel, Distributed and Network-based Processing (PDP), 21-23 March 2018, Cambridge, UK(2018). doi:10.1109/PDP2018.2018.00060
27. Mills, T.C.: Forecasting Financial Time Series. In: Clements, M.P., Hendry, D.F. (eds.) The Oxford Handbook of Economic Forecasting. Oxford University Press, Oxford (2011). doi:10.1093/oxfordhb/9780195398649.013.0019
28. Nti, I.K., Adekoya, A.F., Weyori, B.A.: A systematic review of fundamental and technical analysis of stock market predictions. Artificial Intelligence Review **53**, 3007–3057 (2020). doi:10.1007/s10462-019-09754-z
29. Okasha, M.K.: Using Support Vector Machines in Financial Time Series Forecasting. International Journal of Statistics and Applications **4**(1), 28–39 (2014). doi:10.5923/j.statistics.20140401.03
30. Sapankevych, N.I., Sankar, R.: Time Series Prediction Using Support Vector Machines: A Survey. IEEE Computational Intelligence Magazine **4**(2), 24–38 (2009). doi:10.1109/MCI.2009.932254
31. Saxena, A., Sukumar, T.R.: Predicting bitcoin price using lstm And Compare its predictability with ARIMA model. International Journal of Pure and Applied Mathematics **119**(17), 2591–2600 (2018)
32. Semerikov, S.O., Teplytskyi, I.O., Yechkalo, Yu.V., Kiv, A.E.: Computer Simulation of Neural Networks Using Spreadsheets: The Dawn of the Age of Camelot. CEUR Workshop Proceedings **2257**, 122–147 (2018)
33. Sezer, O.B., Gudelek, M.U., Ozbayoglu, A.M.: Financial time series forecasting with deep learning : A systematic literature review: 2005–2019. Applied Soft Computing **90**, 106181 (2020). doi:10.1016/j.asoc.2020.106181
34. Tarasenko, A.O., Yakimov, Y.V., Soloviev, V.N.: Convolutional neural networks for image classification. CEUR Workshop Proceedings **2546**, 101–114 (2019)
35. Varghade, P., Patel, R.: Comparison of SVR and Decision Trees for Financial Series Prediction. IJACTE **1**(1), 101–105 (2012)
36. Volkova, N.P., Rizun, N.O., Nehrey, M.V.: Data science: opportunities to transform education. CEUR Workshop Proceedings **2433**, 48–73 (2019)
37. Yahoo Finance: Stock Market Live, Quotes, Business & Finance News. https://finance.yahoo.com (2020). Accessed 30 Jun 2020
38. Yao, Y., Yi, J., Zhai, S., Lin, Y., Kim, T., Zhang, G., Lee, L.Y.: Predictive Analysis of Cryptocurrency Price Using Deep Learning. International Journal of Engineering & Technology **7**(3), 258–264 (2018). doi:10.14419/ijet.v7i3.27.17889
39. Zarandi, M.H.F., Rezaee, B., Turksen, I.B., Neshat, E.: A type-2 fuzzy rule-based experts system model for stock price analysis. Expert Systems with Applications **36**(1), 139–154 (2009). doi:10.1016/j.eswa.2007.09.034

Author Index