Advances in machine learning for the innovation economy: in the shadow of war

Hanna B. Danylchuk[1], Serhiy O. Semerikov[2,3,4,5,6]

[1]The Bohdan Khmelnytsky National University of Cherkasy, 81 Shevchenko Blvd., Cherkasy, 18031, 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 for Digitalisation of Education of the NAES of Ukraine, 9 M. Berlynskoho Str., Kyiv, 04060, Ukraine
[5]Zhytomyr Polytechnic State University, 103 Chudnivsyka Str., Zhytomyr, 10005, Ukraine
[6]Academy of Cognitive and Natural Sciences, 54 Gagarin Ave., Kryvyi Rih, 50086, Ukraine

Abstract

This preface introduces the selected and revised papers presented at the 10th International Conference on Monitoring, Modeling & Management of Emergent Economy (M3E2 2022), held online in Ukraine, on November 17-18, 2022. The conference aimed to bring together researchers, practitioners, and students from various fields to exchange ideas, share experiences, and discuss challenges and opportunities in applying computational intelligence and data science for the innovation economy. The innovation economy is a term that describes the emerging paradigm of economic development that is driven by knowledge, creativity, and innovation. It requires new approaches and methods for solving complex problems, discovering new opportunities, and creating value in various domains of science, business, and society. Computational intelligence and data science are two key disciplines that can provide such approaches and methods by exploiting the power of data, algorithms, models, and systems to enable intelligent decision making, learning, adaptation, optimization, and discovery. The papers in this proceedings cover a wide range of topics related to computational intelligence and data science for the innovation economy. They include theoretical foundations, novel techniques, and innovative applications. The papers were selected and revised based on the feedback from the program committee members and reviewers who ensured their high quality. We would like to thank all the authors who submitted their papers to M3E2 2022. We also appreciate the keynote speakers who shared their insights and visions on the current trends and future directions of computational intelligence and data science for the innovation economy. We acknowledge the support of our sponsors, partners, and organizers who made this conference possible despite the challenging circumstances caused by the ongoing war in Ukraine. Finally, we thank all the participants who attended the conference online and contributed to its success.

Keywords

computational intelligence, data science, innovation economy, artificial neural networks, machine learning, visualization

M3E2-MLPEED 2022: The 10th International Conference on Monitoring, Modeling & Management of Emergent Economy, November 17-18, 2022, Virtual, Online
abdanilchuk@gmail.com (H. B. Danylchuk); semerikov@acnsci.org (S. O. Semerikov)
https://kdpu.edu.ua/semerikov (S. O. Semerikov)
0000-0002-9909-2165 (H. B. Danylchuk); 0000-0003-0789-0272 (S. O. Semerikov)

1. Introduction

The **Monitoring, Modeling & Management of Emergent Economy** (M3E2, https://m3e2.ccjournals.eu/2022/) is a peer-reviewed international conference dedicated to scientific achievements in the field of complex systems, the use of information systems and technologies in the economy, interdisciplinary methods, methods of machine learning and fuzzy logic, modeling of socio-economic systems, research global transformations and challenges facing economist scientists. The M3E2 conference is a permanent scientific platform that was launched in 2008 and was formed thanks to the hard work of scientists, practicing researchers, post-graduate students who present the results of their research and have the opportunity to fruitfully discuss them.

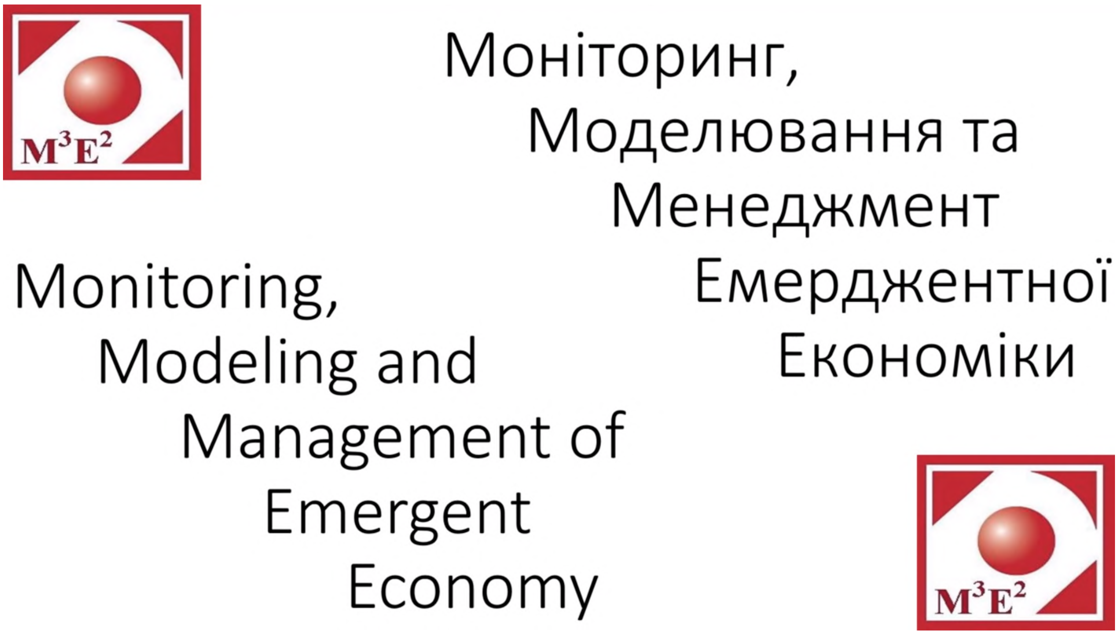


Figure 1: Conference poster.

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 [1, 2, 3, 4]:

- Complex cyberphysical systems, synergy, econophysics, economy of agents
- Dynamics of emergent markets in crisis and post-crisis period
- Economic security
- Global challenges for economic theory and practice in Europe

- Information systems and technologies in economics
- Innovation models of economic development
- Machine learning for prediction of emergent economy dynamics
- Management of the state's economic safety and economic safety of economic agents
- Methods and models of artificial intelligence in economic systems
- Modeling of hospitality sphere development
- Models of global transformations
- Monitoring, modeling and forecasting in the banking sector
- Monitoring, modeling, forecasting and preemption of crisis in socio-economic systems
- Optimal management of socio-economic processes
- Risk management models in emergent economy

This volume contains the selected and revised papers presented at the 10th International Conference on Monitoring, Modeling & Management of Emergent Economy (M3E2 2022) held on November 17-18, 2022 in Ukraine.

There were 23 submissions. Each submission was reviewed by at least 3, and on the average 3.2, program committee members. 13 papers were accepted for this volume as regular papers.

2. Program committee

2.1. M3E2 2022 program chairs

- *Serhiy Semerikov*, Kryvyi Rih State Pedagogical University, Ukraine [5]
- *Vladimir Soloviev*, Kryvyi Rih State Pedagogical University, Ukraine [6]
- *Andriy Matviychuk*, Kyiv National Economic University named after Vadym Hetman, Ukraine [7]
- *Vitaliy Kobets*, Kherson State University, Ukraine [8]
- *Liubov Kibalnyk*, The Bohdan Khmelnytsky National University of Cherkasy, Ukraine [9]
- *Hanna Danylchuk*, The Bohdan Khmelnytsky National University of Cherkasy, Ukraine [10]
- *Arnold Kiv*, Ben-Gurion University of the Negev, Israel [11]

2.2. M3E2 2022 program committee members

- *George Abuselidze*, Batumi Shota Rustaveli State University, Georgia [12]
- *Iluta Arbidane*, Rezekne Academy of Technologies, Latvia [13]
- *Vitalina Babenko*, V. N. Karazin Kharkiv National University, Ukraine [14]
- *Paul Bilokon*, Imperial College London, United Kingdom [15]
- *José Manuel Macedo Botelho*, Universidade de Évora, Portugal [16]
- *Irina Georgescu*, Bucharest University of Economics, Romania [17]
- *Lidiya Guryanova*, Simon Kuznets Kharkiv National University of Economics, Ukraine [18]

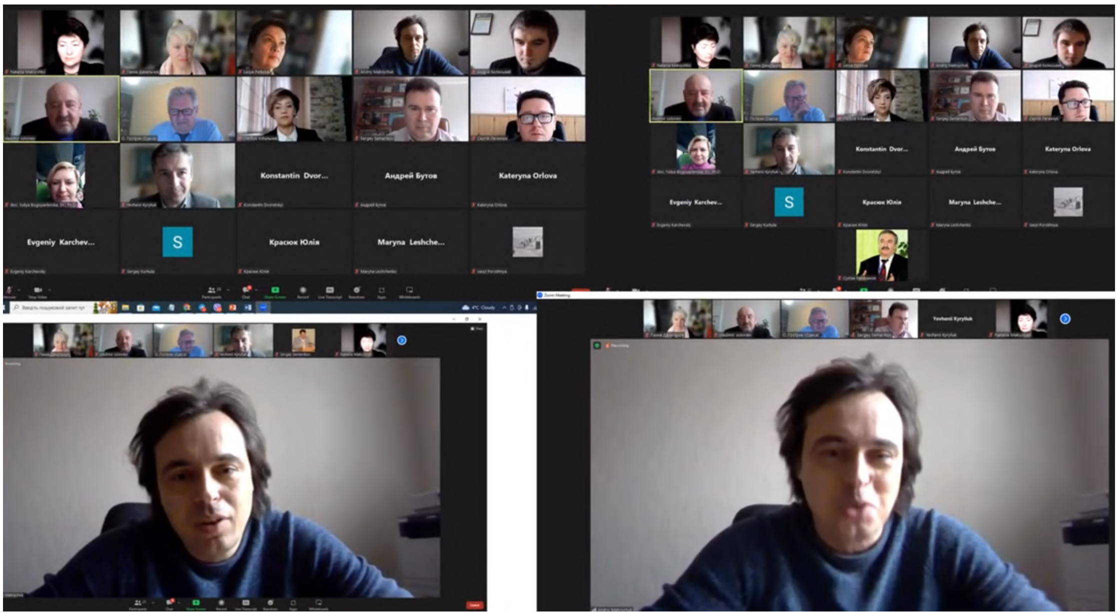


Figure 2: Conference highlights, part 1.

- *Alexey Hostryk*, Odessa National Economic University, Ukraine [19]
- *Pavlo Hryhoruk*, Khmelnytskyi National University, Ukraine [20]
- *Muhammad Jawad*, Fatima Jinnah Women University, Pakistan [21]
- *Nila Khrushch*, Khmelnytskyi National University, Ukraine [22]
- *Inesa Khvostina*, Ivano-Frankivsk National Technical University of Oil and Gas, Ukraine [23]
- *Oksana Kovtun*, University of Educational Management, Ukraine [24]
- *Serhii Lehenchuk*, Zhytomyr Polytechnic State University, Ukraine [25]
- *Nataliia Maksyshko*, Zaporizhzhia National University, Ukraine [26]
- *Abdukhakim Mamanazarov*, Center of Economic Culture Development, Uzbekistan [27]
- *Ewa Matuska*, Pomeranian University in Slupsk, Poland [28]
- *Inese Mavlutova*, BA School of Business and Finance, Latvia [29]
- *Iveta Mietule*, Rezekne Academy of Technologies, Latvia [30]
- *Dariusz Pawliszczy*, Gromadka Community, Poland [31]
- *Oleg Pursky*, Kyiv National University of Trade and Economics, Ukraine [32]
- *Michael Radin*, Rochester Institute of Technology, United States [33]
- *Sultan Ramazanov*, Kyiv National Economic University named after Vadym Hetman, Ukraine [34]
- *Kateryna Shymanska*, Prague University of Economics and Business, Czechia [35]
- *Victoria Solovieva*, State University of Economics and Technology, Ukraine [36]
- *Galyna Velykoivanenko*, Kyiv National Economic University named after Vadym Hetman, Ukraine [37]

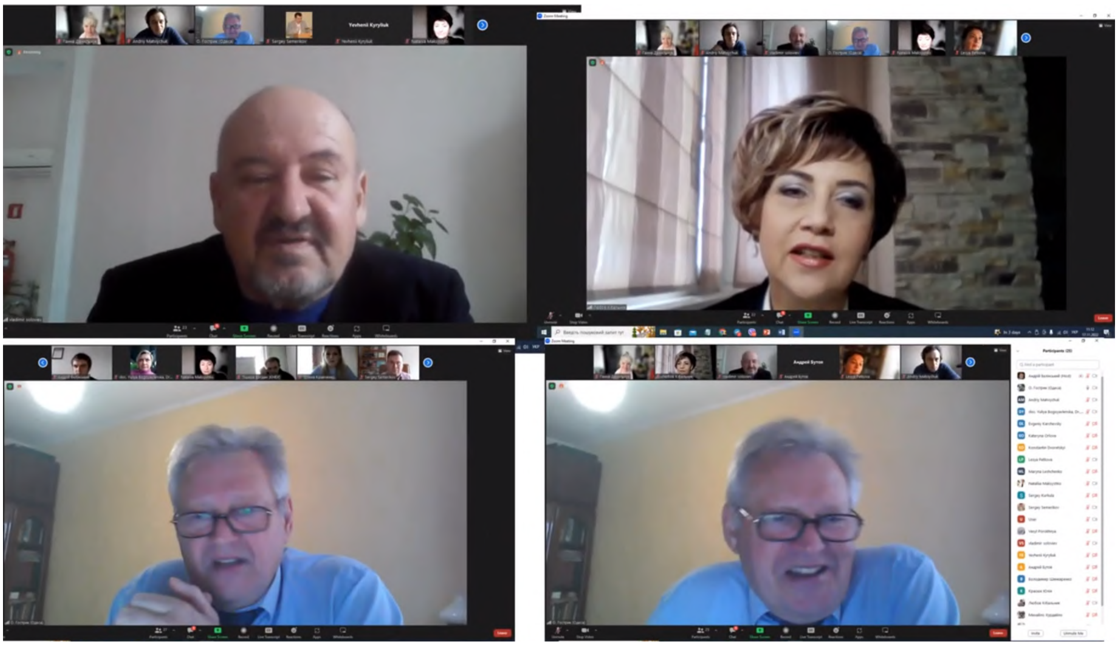


Figure 3: Conference highlights, part 2.

- *Nataliia Zachosova*, The Bohdan Khmelnytsky National University of Cherkasy, Ukraine [38]
- *Pavel Zakharchenko*, Berdyansk State Pedagogical University, Ukraine [39]

3. Articles overview

The paper titled "Assessing the educational dimension of national economy innovative development" by Olha Ilyash, Larysa Taranenko, Olena Trofymenko, Nataliia Koba, and Marzena Sobczak-Michalowska [40] examines the educational indicators that reflect the innovative development of the national economy in Ukraine. The study aims to develop a system for evaluating and enhancing the educational component of Ukraine's innovative development, which can support effective state regulation of educational processes and prevent the risks of reducing the educational security of the national economy.

The study applies a multidimensional analysis of educational indicators, using a system of complex and systemic methods, such as dynamic analysis, system generalization, statistical methods, and taxonomic analysis. The study also compares the educational indicators of Ukraine with those of other countries that have achieved educational and scientific breakthroughs.

The results of the study show that Ukraine has a low level of educational performance and potential for innovative development compared to other countries. The paper proposes some measures to improve the educational component of Ukraine's innovative development, such as increasing public investment in education, enhancing the quality and relevance of education, fostering international cooperation in education and science, and promoting a

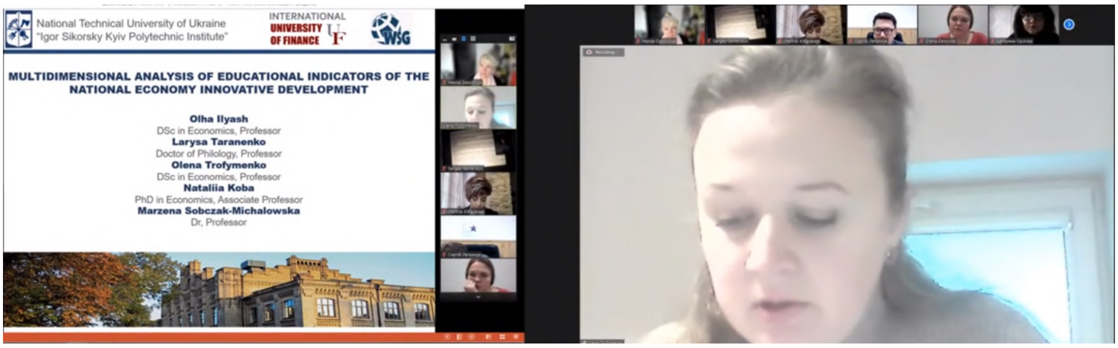


Figure 4: Presentation of paper [40].

culture of innovation among students and teachers.

Here are some of the key points of the paper:

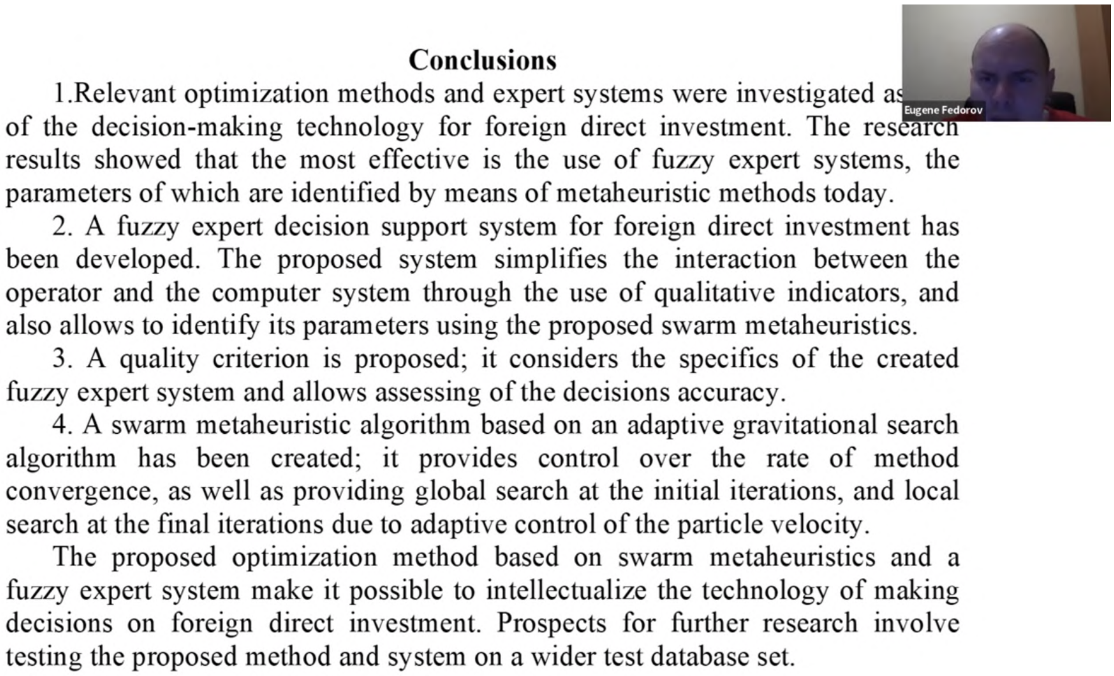
- The educational dimension of national economy innovative development is a complex and multifaceted concept that includes a wide range of factors, such as the quality of education, the level of investment in education, the international cooperation in education and science, and the culture of innovation.
- The study found that Ukraine has a low level of educational performance and potential for innovative development compared to other countries.
- The paper proposes some measures to improve the educational component of Ukraine's innovative development, such as increasing public investment in education, enhancing the quality and relevance of education, fostering international cooperation in education and science, and promoting a culture of innovation among students and teachers.

The paper "Fuzzy expert decision support system for foreign direct investment: a swarm metaheuristic approach" by Eugene E. Fedorov, Liubov O. Kibalnyk, Lesya O. Petkova, Maryna M. Leshchenko, and Vladyslav M. Pasenko [41] proposes a fuzzy expert system for foreign direct investment (FDI) decision support. The system is developed using an adaptive gravitational search algorithm (GSA) to determine the optimal parameters of the fuzzy expert system, such as the membership functions for linguistic input and output variables. The system also uses a quality criterion that considers the specificity of the fuzzy expert system and allows assessing the probability of future decisions.

The paper conducts a numerical study to test the performance of the proposed fuzzy expert system and compares it with other existing methods. The results show that the proposed fuzzy expert system has a high accuracy and robustness in FDI decision support. The paper contributes to the literature on fuzzy logic applications in economics and finance and provides a practical tool for investors to make informed decisions on FDI.

Here are some of the key points of the paper:

- The proposed fuzzy expert system is a novel approach for FDI decision support that takes into account the uncertainty and ambiguity of the decision-making process.



Conclusions

1.Relevant optimization methods and expert systems were investigated as of the decision-making technology for foreign direct investment. The research results showed that the most effective is the use of fuzzy expert systems, the parameters of which are identified by means of metaheuristic methods today.

2. A fuzzy expert decision support system for foreign direct investment has been developed. The proposed system simplifies the interaction between the operator and the computer system through the use of qualitative indicators, and also allows to identify its parameters using the proposed swarm metaheuristics.

3. A quality criterion is proposed; it considers the specifics of the created fuzzy expert system and allows assessing of the decisions accuracy.

4. A swarm metaheuristic algorithm based on an adaptive gravitational search algorithm has been created; it provides control over the rate of method convergence, as well as providing global search at the initial iterations, and local search at the final iterations due to adaptive control of the particle velocity.

The proposed optimization method based on swarm metaheuristics and a fuzzy expert system make it possible to intellectualize the technology of making decisions on foreign direct investment. Prospects for further research involve testing the proposed method and system on a wider test database set.

Figure 5: Presentation of paper [41].

- The GSA is used to optimize the parameters of the fuzzy expert system, which ensures that the system is able to make accurate and reliable decisions.
- The quality criterion is used to assess the performance of the fuzzy expert system and to compare it with other existing methods.
- The numerical study shows that the proposed fuzzy expert system has a high accuracy and robustness in FDI decision support.

The paper "The impact of intangible assets on the financial performance of Slovak ICT companies: a panel data regression analysis" by Serhii F. Lehenchuk, Tetiana A. Vakaliuk, Tetiana P. Nazarenko, Zuzana Kubaščíková, and Zuzana Juhászová [42] investigates the impact of intangible assets on the financial performance of Slovak ICT companies in the context of the knowledge economy.

The paper uses a panel data regression analysis to test the hypothesis that intangible assets have a significant positive effect on four indicators of financial performance: Return on Assets (ROA), Net Profit Margin (NPM), Assets Turnover (AT), and Return on Equity (ROE). The paper analyzes a sample of 180 Slovak ICT companies for the period 2015–2019, using eight independent variables: Research and Development Intensity (R&D), R&D Intensity Squared, Software, Intellectual Property Rights (IPR), Acquired Intangible Assets, Leverage, Size, and Dummy variable for ICT sub-sectors.

The paper applies various tests to select the appropriate estimation method and to check the adequacy of each model. The results partially confirm the hypothesis, as only R&D, R&D Intensity Squared, and Acquired Intangible Assets have a significant positive impact on some

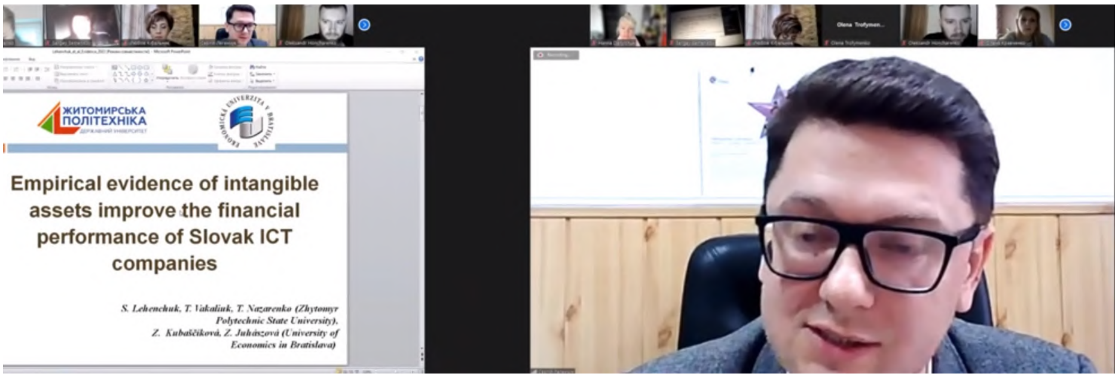


Figure 6: Presentation of paper [42].

indicators of financial performance. The paper also finds that the influence of intangible assets varies depending on the type and measure of financial performance.

The paper contributes to the literature on intangible assets and financial performance by providing empirical evidence from Slovak ICT companies. The paper also provides some implications for managers and policymakers to improve their intangible investment policy.

Here are some of the key points of the paper:

- The paper provides empirical evidence that intangible assets have a positive impact on the financial performance of Slovak ICT companies.
- The paper finds that the impact of intangible assets varies depending on the type and measure of financial performance.
- The paper suggests that managers and policymakers should focus on investing in intangible assets that have a proven positive impact on financial performance.

The paper "Maximizing customer satisfaction and business profits through Big Data technology in Society 5.0: a crisis-responsive approach for emerging markets" by Piotr Kulyk, Viktoriia Hurochkina, Bohdan Patsai, Olena Voronkova, and Oksana Hordei [43] explores the use of Big Data technology to maximize customer satisfaction and business profits in emerging markets, especially during crisis periods.

The paper begins by discussing the importance of customer satisfaction in emerging markets, where competition is often fierce and customer expectations are high. The authors argue that Big Data technology can be used to gain a deeper understanding of customer needs and preferences, which can be used to improve products and services, personalize marketing campaigns, and provide better customer service.

The paper then discusses the use of Big Data technology in loyalty programs. The authors argue that traditional discount-based loyalty programs are not effective in emerging markets, where customers are more likely to be attracted to personalized and value-added rewards. Big Data technology can be used to create personalized loyalty programs that are tailored to the specific needs and preferences of each customer.

The paper also discusses the use of Big Data technology in crisis mitigation. The authors argue that Big Data technology can be used to track customer behavior and preferences during a crisis, which can be used to adjust marketing campaigns and product offerings accordingly. Big Data technology can also be used to identify and target customers who are most likely to be affected by a crisis, and to provide them with the support they need.

The paper concludes by arguing that Big Data technology is a powerful tool that can be used to maximize customer satisfaction and business profits in emerging markets, especially during crisis periods. The paper provides a number of recommendations for businesses that are looking to use Big Data technology to improve their performance in emerging markets.

Here are some of the key points of the paper:

- Big Data technology can be used to gain a deeper understanding of customer needs and preferences.
- Big Data technology can be used to personalize marketing campaigns and customer service.
- Big Data technology can be used to create personalized loyalty programs.
- Big Data technology can be used to track customer behavior and preferences during a crisis.
- Big Data technology can be used to identify and target customers who are most likely to be affected by a crisis.

The paper "A flexible machine learning model for optimizing organizational capital development strategies and resource allocation" by Vasyl Porokhnya, Vladyslav Penev, Roman Ivanov, and Volodymyr Kravchenko [44] proposes a flexible machine learning model for optimizing organizational capital development strategies. The model is based on Q-learning, a reinforcement learning algorithm that can be used to learn optimal policies in a dynamic environment.

The model is designed to be flexible and adaptable to different organizational contexts. It can be used to optimize a variety of organizational capital development strategies, such as training and development, knowledge management, and innovation.

The model is also designed to be data-driven. It uses historical data to learn the relationship between different organizational capital development strategies and their outcomes. This allows the model to make more accurate predictions about the effectiveness of different strategies.

The paper evaluates the performance of the model using a case study of a manufacturing company. The results show that the model was able to identify the most effective organizational capital development strategies for the company. The model also helped the company to improve its resource allocation decisions.

The paper concludes that the proposed model is a valuable tool for optimizing organizational capital development strategies. The model is flexible, adaptable, and data-driven, making it suitable for a wide range of organizations.

Here are some of the key points of the paper:

- The proposed model is a flexible machine learning model that can be used to optimize organizational capital development strategies.
- The model is based on Q-learning, a reinforcement learning algorithm that can be used to learn optimal policies in a dynamic environment.
- The model is designed to be adaptable to different organizational contexts.

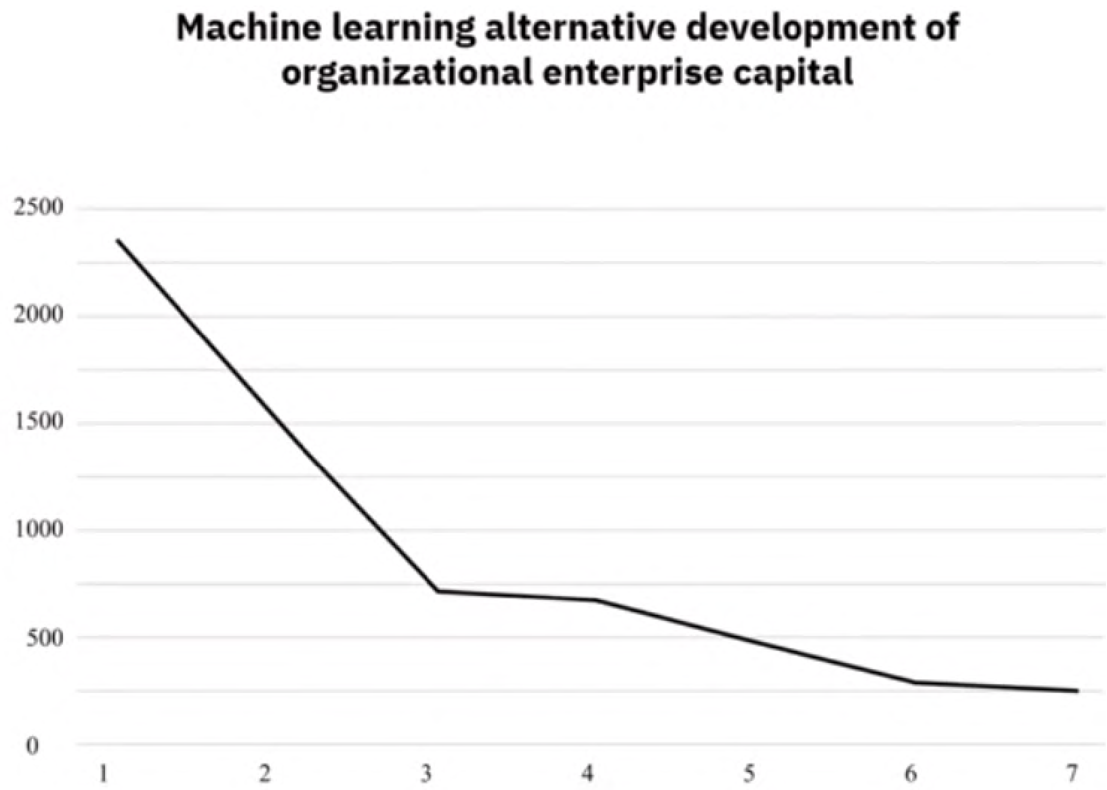


Figure 7: Presentation of paper [44].

- The model is data-driven, using historical data to learn the relationship between different organizational capital development strategies and their outcomes.
- The paper evaluates the performance of the model using a case study of a manufacturing company. The results show that the model was able to identify the most effective organizational capital development strategies for the company.
- The paper concludes that the proposed model is a valuable tool for optimizing organizational capital development strategies.

The paper "Recurrence quantification analysis of energy market crises: a nonlinear approach to risk management" by Andrii O. Bielinskyi, Vladimir N. Soloviev, Viktoria V. Solovieva, Serhiy O. Semerikov, and Michael A. Radin [45] uses recurrence quantification analysis (RQA) to analyze and construct indicators of intermittent events in energy indices.

The paper begins by discussing the importance of the energy market and the challenges posed by its unstable price dynamics. The paper then introduces RQA, a nonlinear time series analysis method that can be used to identify and quantify intermittent events.

The paper then applies RQA to daily data of Henry Hub natural gas spot prices, WTI spot prices, and Europe Brent spot prices. The results show that the recurrence measures capture the distinctive features of crashes and can be used for effective risk management strategies.

The paper concludes by discussing the implications of the findings for risk management in the energy market. The paper argues that RQA can be used to identify the early warning signs of crises and to develop strategies to mitigate their impact.

Here are some of the key points of the paper:

- The paper uses RQA to analyze and construct indicators of intermittent events in energy indices.
- The paper applies RQA to daily data of Henry Hub natural gas spot prices, WTI spot prices, and Europe Brent spot prices.
- The results show that the recurrence measures capture the distinctive features of crashes and can be used for effective risk management strategies.

The paper "High-order network analysis for financial crash identification" by Andrii O. Bielinskyi, Vladimir N. Soloviev, Serhii V. Hushko, Arnold E. Kiv, and Andriy V. Matviychuk [46] proposes to use high-order networks to study the temporal evolution of the Dow Jones Industrial Average (DJIA) index.

Figure 8: Presentation of paper [46].

The paper begins by discussing the importance of network analysis for understanding the complexity and dynamics of socio-economic systems. The paper then introduces high-order networks, which are generalized network structures that capture the higher-order dependencies that arise from the interactions of more than two nodes.

The paper then constructs high-order networks from the DJIA time series using the visibility graph method. The visibility graph method is a method for constructing networks from time series data by connecting points that are visible to each other at a given time.

The paper then measures the topological complexity of the high-order networks using various metrics. The topological complexity of a network is a measure of the richness and interconnectedness of the network.

The paper finds that the complexity of the DJIA high-order networks changes drastically during crisis events. This indicates that high-order network analysis can be used as an indicator of financial crashes.

The paper also shows that high-order network analysis and topology can provide more insights into the nonlinear and nonstationary behavior of the DJIA index than traditional tools of financial time series analysis.

Here are some of the key points of the paper:

- The paper proposes to use high-order networks to study the temporal evolution of the DJIA index.
- The paper constructs high-order networks from the DJIA time series using the visibility graph method.
- The paper measures the topological complexity of the high-order networks using various metrics.
- The paper finds that the complexity of the DJIA high-order networks changes drastically during crisis events.
- The paper shows that high-order network analysis and topology can provide more insights into the nonlinear and nonstationary behavior of the DJIA index than traditional tools of financial time series analysis.

The paper "Multidimensional statistical analysis of investment attractiveness and regional changes in the COVID-19 pandemic" by Pavlo M. Hryhoruk, Nila A. Khrushch, Svitlana S. Grygoruk, and Olena R. Ovchynnikova [47] uses multidimensional statistical analysis techniques to cluster Ukrainian regions based on their levels of investment attractiveness. The paper also examines how the regional investment attractiveness structure has changed during the COVID-19 pandemic.

The paper begins by reviewing the various approaches to assessing investment attractiveness. The authors then use the *k*-means method to cluster Ukrainian regions based on their investment attractiveness levels in 2019 and 2020. The *k*-means method is a clustering algorithm that groups data points into a predefined number of clusters.

The results of the *k*-means clustering show that the investment attractiveness of Ukrainian regions has changed during the COVID-19 pandemic. In 2019, there were four clusters of regions: high-attractiveness regions, medium-high-attractiveness regions, medium-low-attractiveness regions, and low-attractiveness regions. In 2020, the number of clusters increased to five.

The paper then uses principal component analysis (PCA) to rotate the space of selected factors. PCA is a statistical technique that reduces the dimensionality of a dataset while preserving as much information as possible. The quartimax method is a rotation method that maximizes the variance of the rotated components.

The results of the PCA and quartimax rotation show that the regional investment attractiveness structure can be explained by four main factors: economic development, infrastructure, human capital, and political stability. The authors argue that these factors are important for attracting investment, and that they have become even more important during the COVID-19 pandemic.

The paper concludes by discussing the implications of the findings for potential investors and local self-governing bodies. The authors argue that the findings can help investors to identify key investment areas, and that they can help local self-governing bodies to improve their investment attractiveness.

Here are some of the key points of the paper:

- The paper uses multidimensional statistical analysis techniques to cluster Ukrainian regions based on their levels of investment attractiveness.
- The paper examines how the regional investment attractiveness structure has changed during the COVID-19 pandemic.
- The results of the k-means clustering show that the investment attractiveness of Ukrainian regions has changed during the COVID-19 pandemic.
- The paper uses PCA and quartimax rotation to identify four main factors that explain the regional investment attractiveness structure.
- The authors argue that these factors are important for attracting investment, and that they have become even more important during the COVID-19 pandemic.

The paper "A comparative study of deep learning models for sentiment analysis of social media texts" by Vasily D. Derbentsev, Vitalii S. Bezkorovainyi, Andriy V. Matviychuk, Oksana M. Pomazun, Andrii V. Hrabariev, and Alexey M. Hostryk [48] presents a comparative study of deep learning models for sentiment analysis of social media texts.

The paper begins by discussing the challenges of sentiment analysis for social media texts. Social media texts are often informal, short, and noisy, which makes them difficult to analyze using traditional machine learning methods.

The paper then introduces three deep learning models for sentiment analysis: a convolutional neural network (CNN), a CNN with long short-term memory (LSTM) layers (CNN-LSTM), and a bidirectional LSTM with CNN layers (BiLSTM-CNN). The CNN is a deep learning model that is well-suited for processing natural language text. The LSTM is a deep learning model that is well-suited for processing sequential data.

The paper then evaluates the performance of the three models on two datasets: IMDb Movie Reviews and Twitter Sentiment 140. The IMDb Movie Reviews dataset contains 50,000 movie reviews, each of which is labeled as either positive or negative. The Twitter Sentiment 140 dataset contains 1.4 million tweets, each of which is labeled as either positive, negative, or neutral.

The results show that the CNN model achieves the best accuracy of 90.1% on the IMDb dataset, while the BiLSTM-CNN model achieves the best accuracy of 82.1% on the Sentiment 140 dataset. The proposed models are comparable to state-of-the-art models and suitable for practical use in sentiment analysis of social media texts.

Here are some of the key points of the paper:

- The paper presents a comparative study of three deep learning models for sentiment analysis of social media texts.

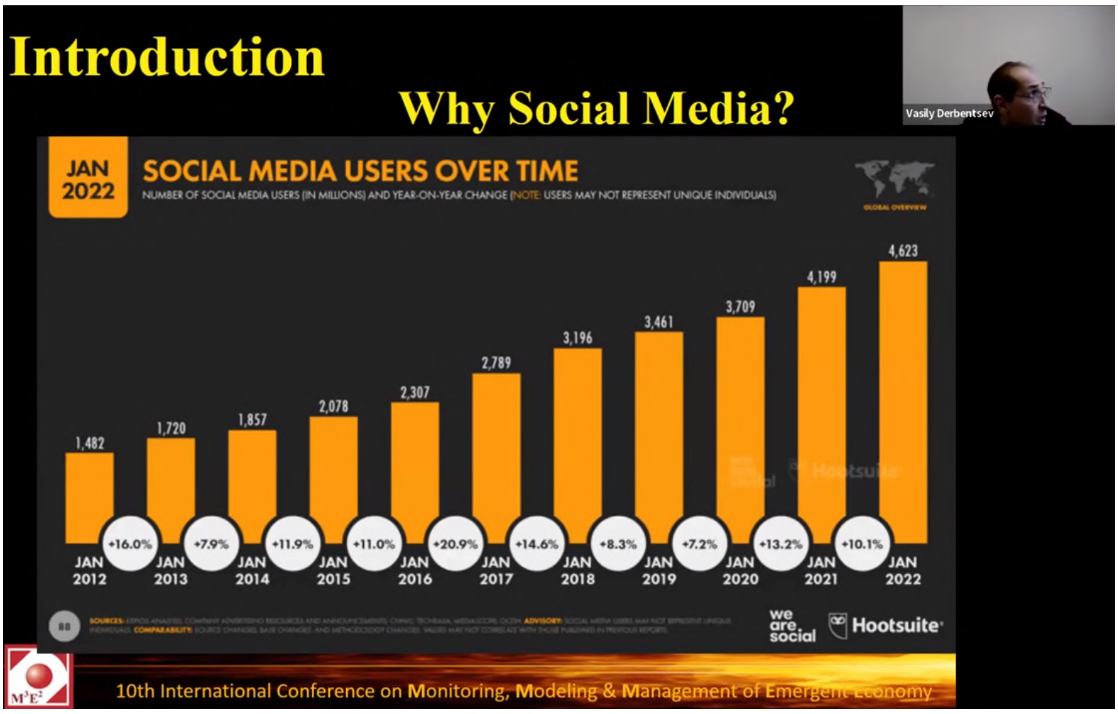


Figure 9: Presentation of paper [48].

- The models are evaluated on two datasets: IMDb Movie Reviews and Twitter Sentiment 140.
- The results show that the CNN model achieves the best accuracy of 90.1% on the IMDb dataset, while the BiLSTM-CNN model achieves the best accuracy of 82.1% on the Sentiment 140 dataset.
- The proposed models are comparable to state-of-the-art models and suitable for practical use in sentiment analysis of social media texts.

The paper "The impact of the war in Ukraine on globalization processes and world financial markets: a wavelet entropy analysis" by Hanna B. Danylchuk, Liubov O. Kibalnyk, Oksana A. Kovtun, Oleg I. Pursky, Yevhenii M. Kyryliuk, and Olena O. Kravchenko [49] uses wavelet entropy analysis to investigate the impact of the war in Ukraine on globalization processes and world financial markets.

The paper begins by discussing the importance of globalization and the world financial markets. Globalization is the process of increasing interconnectedness between countries and economies. The world financial markets are the system of institutions that facilitate the exchange of money and financial instruments between countries.

The paper then discusses the war in Ukraine and its impact on globalization and the world financial markets. The war has caused a significant disruption to global trade and investment, and has led to increased uncertainty in the financial markets.

The paper then uses wavelet entropy analysis to study the impact of the war on the markets

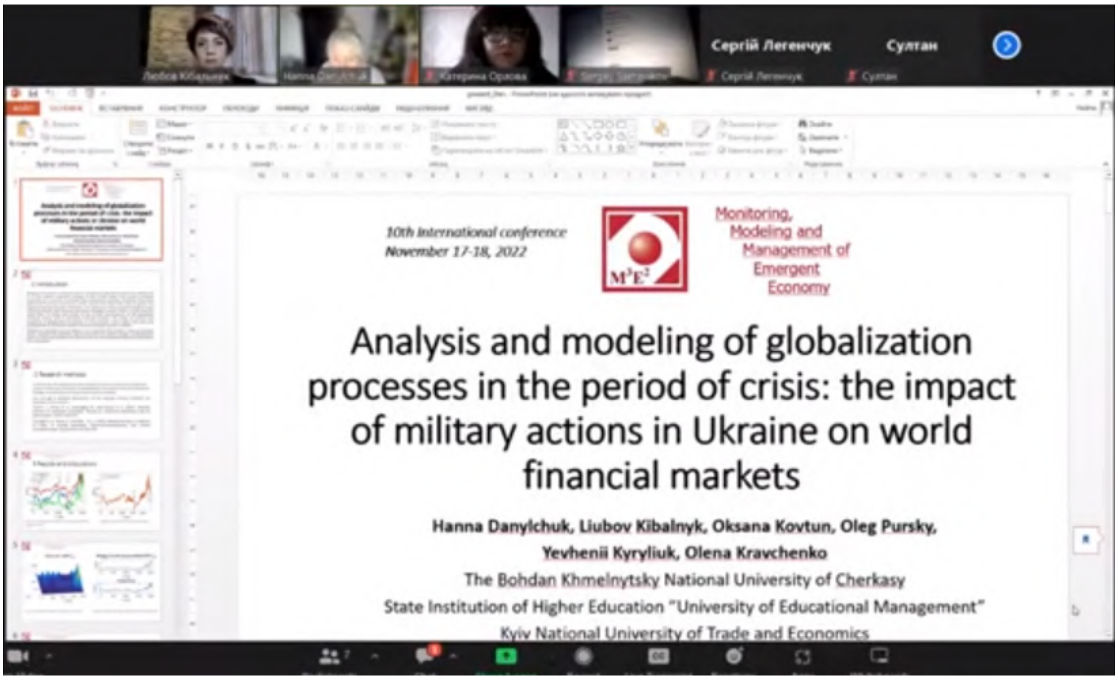


Figure 10: Presentation of paper [49].

for natural gas, oil, gasoline, and currency pairs EUR/USD and GBP/USD. Wavelet entropy is a measure of the complexity and uncertainty of a signal or system. The results show that the war has caused a significant increase in the entropy of these markets, indicating that they have become more complex and uncertain.

The paper concludes by discussing the implications of the findings for globalization and the world financial markets. The authors argue that the war has had a negative impact on globalization, and that it is likely to lead to a reconfiguration of the world economic space.

Here are some of the key points of the paper:

- The paper uses wavelet entropy analysis to study the impact of the war in Ukraine on the markets for natural gas, oil, gasoline, and currency pairs EUR/USD and GBP/USD.
- The results show that the war has caused a significant increase in the entropy of these markets, indicating that they have become more complex and uncertain.
- The authors argue that the war has had a negative impact on globalization, and that it is likely to lead to a reconfiguration of the world economic space.

The paper "Nonlinear dynamics of electric vehicle sales in China: a fractal analysis" by Serhii Kurkula, Nataliia Maksyshko, Dmytro Ocheretin, and Serhii Cheverda [50] applies three methods of nonlinear analysis to investigate the properties of the monthly sales volumes of the leading EV manufacturers in China from January 2016 to June 2022.

The paper begins by discussing the importance of electric vehicles (EVs) and the growth of the EV market in China. EVs are becoming increasingly popular due to their environmental

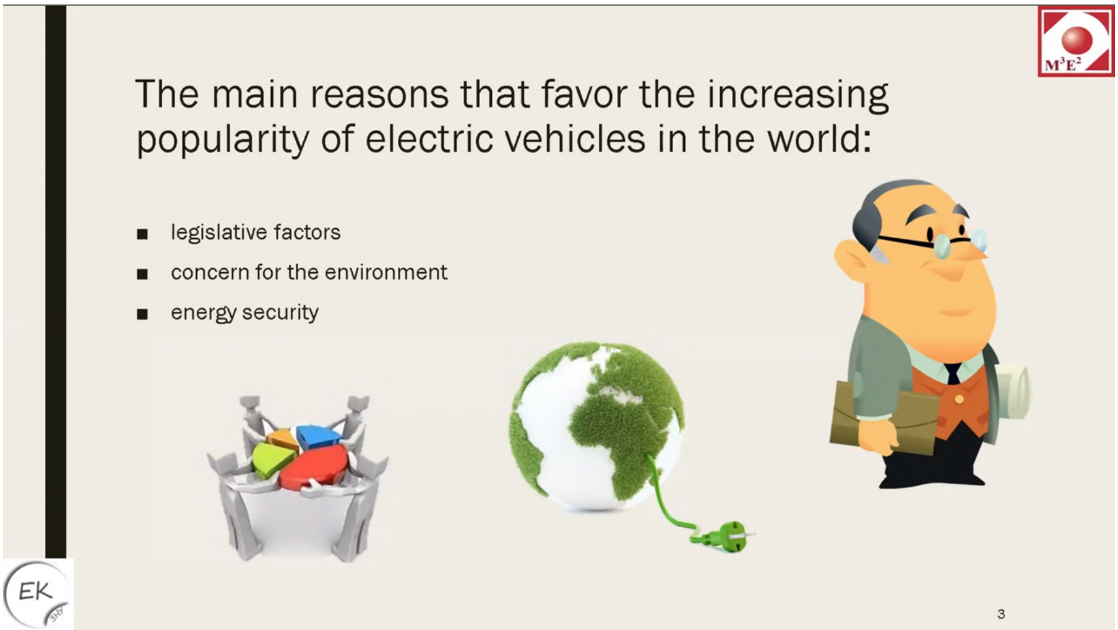


Figure 11: Presentation of paper [50].

benefits and lower running costs. China is the leading market for EVs, accounting for 45% of global sales in 2020.

The paper then discusses the challenges of forecasting EV sales. The EV market is a complex and nonlinear system, making it difficult to predict future sales.

The paper then applies three methods of nonlinear analysis to the monthly sales data: the Hurst normalized range method, phase analysis, and recurrence plots. These methods are used to identify the long-term memory, cyclicity, and determinism of the sales dynamics.

The results show that the sales dynamics exhibit fractal features, trend stability, long-term memory, cyclicity, quasi-cycles, and determinism. These findings can inform the selection of relevant forecasting methods and their parameters for the EV market in China.

Here are some of the key points of the paper:

- The paper applies three methods of nonlinear analysis to the monthly sales data of leading EV manufacturers in China.
- The results show that the sales dynamics exhibit fractal features, trend stability, long-term memory, cyclicity, quasi-cycles, and determinism.
- These findings can inform the selection of relevant forecasting methods and their parameters for the EV market in China.

The paper "A cognitive approach to modeling sustainable development of complex technogenic systems in the innovation economy" by Sultan K. Ramazanov, Bohdan O. Tishkov, Oleksandr H. Honcharenko, and Alexey M. Hostryk [51] proposes a cognitive approach to modeling sustainable development of complex technogenic production systems in the innovation economy.

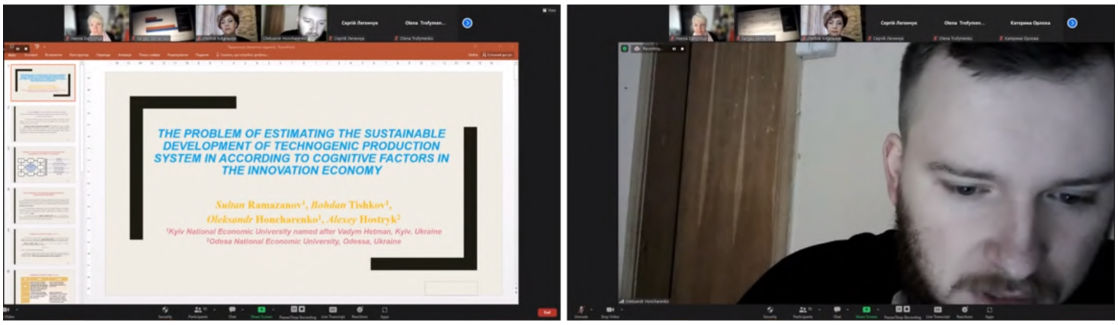


Figure 12: Presentation of paper [51].

The paper begins by discussing the challenges of sustainable development of complex technogenic systems. These systems are characterized by nonlinear dynamics and uncertainty, and they are often affected by human behavior and decision-making.

The paper then proposes a cognitive approach to modeling sustainable development. This approach takes into account the cognitive factors that affect the behavior and decision-making of the system agents. The paper also proposes a model of innovation capital dynamics for the eco-economic and socio-humanitarian system (EESHS). Innovation capital is broader than intellectual capital by its nature and content, and it is essential for sustainable development.

The paper then derives an extended integral model of nonlinear stochastic dynamics of EESHS in the innovation space. This model can be used to predict the dynamics of EESHS and to identify the factors that affect its sustainable development.

The paper concludes by discussing the implications of the findings for the modeling of sustainable development of complex technogenic systems. The paper argues that the cognitive approach is a promising approach for modeling sustainable development, and that it can be used to improve the accuracy of predictions and to identify the factors that need to be managed to achieve sustainable development.

Here are some of the key points of the paper:

- The paper proposes a cognitive approach to modeling sustainable development of complex technogenic systems.
- This approach takes into account the cognitive factors that affect the behavior and decision-making of the system agents.
- The paper also proposes a model of innovation capital dynamics for the EESHS.
- The paper derives an extended integral model of nonlinear stochastic dynamics of EESHS in the innovation space.
- The paper argues that the cognitive approach is a promising approach for modeling sustainable development, and that it can be used to improve the accuracy of predictions and to identify the factors that need to be managed to achieve sustainable development.

The paper "University competitiveness in the knowledge economy: a Kohonen map approach" by Dmytro H. Lukianenko, Andriy V. Matviychuk, Liubov I. Lukianenko, and Iryna V. Dvornyk [52] studies the factors of university competitiveness in the knowledge economy.

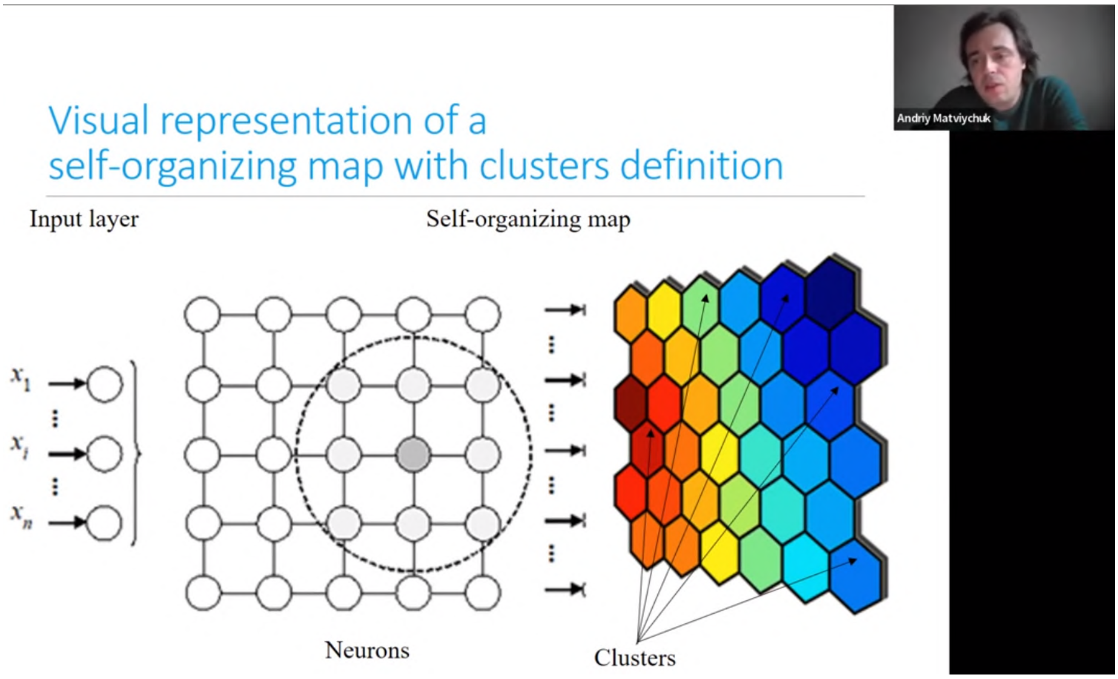


Figure 13: Presentation of paper [52].

The paper begins by discussing the importance of universities in the knowledge economy. Universities play a key role in the generation and dissemination of innovations, and they are also becoming the drivers of digital transformation in science, business, countries, and society as a whole.

The paper then proposes a clustering approach to group countries based on their university competitiveness. The clustering approach is based on the Kohonen map, which is a neural network that can be used to cluster data points into groups. The paper uses the normalized parameters of university competitiveness to cluster countries into four groups: high-performing, medium-performing, low-performing, and very low-performing.

The paper then assesses the level of significance of the normalized parameters. The results show that the most significant parameters are research productivity, internationalization, and the number of publications in high-impact journals.

The paper then proposes an organizational design for a competitive model of the university. The proposed organizational design is based on the principles of open science, education, and innovation. The paper argues that this organizational design can help universities to improve their competitiveness and become drivers of innovation and transformation.

The paper concludes by discussing the key factors of the university's success in the system of open science, education, and innovation. These factors include:

- A strong focus on research and innovation
- A commitment to open science
- A global outlook

- A focus on entrepreneurship and the transfer of knowledge to society

The findings of this study contribute to the understanding of the factors that drive university competitiveness in the knowledge economy. The proposed organizational design and key factors of success can be used by universities to improve their competitiveness and become drivers of innovation and transformation.

Here are some of the key points of the paper:

- The paper studies the factors of university competitiveness in the knowledge economy using a clustering approach.
- The paper proposes an organizational design for a competitive model of the university based on the principles of open science, education, and innovation.
- The paper discusses the key factors of the university's success in the system of open science, education, and innovation.

4. Conclusion

The vision of the M3E2 2022 is to provide 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 performed as a forum for transferring and discussing research results among researchers, students, government, private sector, or industries. Participants and presenters from several countries have attended the conference online to share their significant contributions in research related to Monitoring, Modeling, and Management of Emergent Economy.

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 innovations. We are also thankful to all the program committee members for providing continuous guidance and efforts taken by peer reviewers who contributed to improve the quality of papers. The constructive critical comments, improvements, and corrections provided to the authors are gratefully appreciated for their contribution to the success of the conference. Moreover, we would like to thank the developers and other professional staff of the *Not So Easy Science Education* platform (https://notso.easyscience.education) and the Academy of Cognitive and Natural Sciences (https://acnsci.org), who made it possible for us to use the resources of this excellent and comprehensive conference management system, from the call of papers and inviting reviewers, to handling paper submissions, communicating with the authors, etc.

The war in Ukraine has had a devastating impact on the country and its people. The scientific community in Ukraine has also been affected, with many researchers forced to flee their homes and laboratories. The M3E2 2022 conference was held in the shadow of this war, but it was also a testament to the resilience of the Ukrainian scientific community. The conference provided a platform for Ukrainian researchers to share their work and to connect with colleagues from around the world. We hope that the conference will help to rebuild the scientific community in Ukraine and to contribute to the country's recovery.

Figure 14: Hanna B. Danylchuk and Serhiy O. Semerikov, volume editors.

We also hope that the conference will contribute to the understanding of the war in Ukraine and its impact on the global economy. The papers in the proceedings address a variety of topics related to the war, including its impact on innovation, investment, and trade. We believe that these papers will be valuable resources for researchers, policymakers, and the general public.

Finally, we would like to express our solidarity with the people of Ukraine. We hope for a swift and peaceful resolution to the war.

References

[1] A. Kiv, S. Semerikov, V. N. Soloviev, L. Kibalnyk, H. Danylchuk, A. Matviychuk, Experimental Economics and Machine Learning for Prediction of Emergent Economy Dynamics, in: A. Kiv, S. Semerikov, V. N. Soloviev, L. Kibalnyk, H. Danylchuk, A. Matviychuk (Eds.), Proceedings of the Selected Papers of the 8th International Conference on Monitoring, Modeling & Management of Emergent Economy, M3E2-EEMLPEED 2019, Odessa, Ukraine, May 22-24, 2019, volume 2422 of *CEUR Workshop Proceedings*, CEUR-WS.org, 2019, pp. 1–4. URL: https://ceur-ws.org/Vol-2422/paper00.pdf.

[2] A. Kiv, P. Hryhoruk, I. Khvostina, V. Solovieva, V. N. Soloviev, S. Semerikov, Machine learning of emerging markets in pandemic times, in: A. Kiv (Ed.), 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, volume 2713 of *CEUR Workshop Proceedings*, CEUR-WS.org, 2020, pp. 1–20. URL: https://ceur-ws.org/Vol-2713/paper00.pdf.

[3] A. E. Kiv, V. N. Soloviev, S. O. Semerikov, H. B. Danylchuk, L. O. Kibalnyk, A. V. Matviychuk, A. M. Striuk, Machine learning for prediction of emergent economy dynamics III, in: A. E. Kiv, V. N. Soloviev, S. O. Semerikov (Eds.), Proceedings of the Selected and Revised Papers of 9th International Conference on Monitoring, Modeling & Management of Emergent Economy (M3E2-MLPEED 2021), Odessa, Ukraine, May 26-28, 2021, volume 3048 of *CEUR Workshop Proceedings*, CEUR-WS.org, 2021, pp. i–xxxi. URL: https://ceur-ws.org/Vol-3048/paper00.pdf.

[4] H. B. Danylchuk, S. O. Semerikov, Advances in machine learning for the innovation economy: in the shadow of war, CEUR Workshop Proceedings (2023) 1–25.

[5] S. Semerikov, Educational Technology Quarterly: in the beginning, Educational Technology Quarterly 2021 (2021) 1–50. doi:10.55056/etq.13.
[6] A. Kiv, S. Semerikov, V. Soloviev, XII International Conference on Mathematics, Science and Technology Education: conference report, Educational Technology Quarterly 2021 (2021) 140–256. doi:10.55056/etq.54.
[7] V. Derbentsev, A. Matviychuk, V. N. Soloviev, 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, Springer, Singapore, 2020, pp. 211–231. doi:10.1007/978-981-15-4498-9_12.
[8] O. Ivanov, O. Snihovyi, V. Kobets, Implementation of Robo-Advisors Tools for Different Risk Attitude Investment Decisions, in: V. Ermolayev, M. C. Suárez-Figueroa, V. Yakovyna, V. S. Kharchenko, V. Kobets, H. Kravtsov, V. S. Peschanenko, Y. Prytula, M. S. Nikitchenko, A. Spivakovsky (Eds.), Proceedings of the 14th International Conference on ICT in Education, Research and Industrial Applications. Integration, Harmonization and Knowledge Transfer. Volume II: Workshops, Kyiv, Ukraine, May 14-17, 2018, volume 2104 of *CEUR Workshop Proceedings*, CEUR-WS.org, 2018, pp. 195–206. URL: https://ceur-ws.org/Vol-2104/paper_161.pdf.
[9] H. Danylchuk, L. Kibalnyk, O. Serdiuk, Critical Phenomena Study in Economic Systems Using a Damped Oscillations Model, in: A. Kiv, S. Semerikov, V. N. Soloviev, L. Kibalnyk, H. Danylchuk, A. Matviychuk (Eds.), Proceedings of the Selected Papers of the 8th International Conference on Monitoring, Modeling & Management of Emergent Economy, M3E2-EEMLPEED 2019, Odessa, Ukraine, May 22-24, 2019, volume 2422 of *CEUR Workshop Proceedings*, CEUR-WS.org, 2019, pp. 211–225. URL: https://ceur-ws.org/Vol-2422/paper17.pdf.
[10] S. O. Semerikov, S. M. Chukharev, S. I. Sakhno, A. M. Striuk, A. V. Iatsyshin, S. V. Klimov, V. V. Osadchyi, T. A. Vakaliuk, P. P. Nechypurenko, O. V. Bondarenko, H. B. Danylchuk, 3rd International Conference on Sustainable Futures: Environmental, Technological, Social and Economic Matters, IOP Conference Series: Earth and Environmental Science 1049 (2022) 011001. doi:10.1088/1755-1315/1049/1/011001.
[11] L. Kalashnikova, I. Hrabovets, L. Chernous, V. Chorna, A. Kiv, Gamification as a trend in organizing professional education of sociologists in the context of distance learning: analysis of practices, Educational Technology Quarterly 2022 (2022) 115–128. doi:10.55056/etq.2.
[12] G. Abuselidze, G. Zoidze, The use of transferable skills in education and its impact on the economy, CTE Workshop Proceedings 10 (2023) 124–138. doi:10.55056/cte.550.
[13] V. Komarova, I. Mietule, I. Arbidane, V. Tumalavičius, D. Prakapienė, Will production in the modern world and its regions return to a slow growth regime?, Economic Annals-XXI 187 (2021) 4–14. doi:10.21003/EA.V187-01.
[14] V. O. Babenko, R. M. Yatsenko, P. D. Migunov, A.-B. M. Salem, MarkHub Cloud Online Editor as a modern web-based book creation tool, CTE Workshop Proceedings 7 (2020) 174–184. doi:10.55056/cte.342.
[15] P. Bilokon, A. Edalat, A Domain-Theoretic Approach to Brownian Motion and General Continuous Stochastic Processes, in: Proceedings of the Joint Meeting of the Twenty-Third EACSL Annual Conference on Computer Science Logic (CSL) and the Twenty-Ninth Annual ACM/IEEE Symposium on Logic in Computer Science (LICS), CSL-LICS '14,

Association for Computing Machinery, New York, NY, USA, 2014. doi:10.1145/2603088.2603102.
[16] J. M. M. Botelho, Internacionalização de empresas: contributos para a construção de um modelo de suporte à análise e à implementação de estratégias de internacionalização, Ph.D. thesis, Universidade de Évora, 2015. URL: http://hdl.handle.net/10174/17794.
[17] I. Georgescu, Arrow's Axiom and Full Rationality for Fuzzy Choice Functions, Social Choice and Welfare 28 (2007) 303–319. doi:10.1007/s00355-006-0160-9.
[18] T. S. Klebanova, L. S. Guryanova, I. K. Shevchenko, Model basis of early warning and localization of crises in economic systems of territories, Actual Problems of Economics 153 (2014) 269–278.
[19] V. N. Soloviev, V. Solovieva, A. Tuliakova, A. Hostryk, L. Pichl, Complex networks theory and precursors of financial crashes, in: A. Kiv (Ed.), 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, volume 2713 of *CEUR Workshop Proceedings*, CEUR-WS.org, 2020, pp. 53–67. URL: https://ceur-ws.org/Vol-2713/paper03.pdf.
[20] P. Hryhoruk, N. Khrushch, S. Grygoruk, Assessing the Investment Capacity of the Agricultural Sector: Case of Ukraine, in: 2020 10th International Conference on Advanced Computer Information Technologies (ACIT), 2020, pp. 183–187. doi:10.1109/ACIT49673.2020.9208927.
[21] M. Jawad, Z. Maroof, M. Naz, Development dynamics: Pre and Post Brexit analysis of United Kingdom, Quality & Quantity 53 (2019) 791–811. doi:10.1007/s11135-018-0789-3.
[22] P. Hryhoruk, S. Grygoruk, N. Khrushch, T. Hovorushchenko, Using non-metric multidimensional scaling for assessment of regions' economy in the context of their sustainable development, in: A. Kiv (Ed.), 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, volume 2713 of *CEUR Workshop Proceedings*, CEUR-WS.org, 2020, pp. 315–333. URL: https://ceur-ws.org/Vol-2713/paper35.pdf.
[23] I. Khvostina, S. Semerikov, O. Yatsiuk, N. Daliak, O. Romanko, E. Shmeltser, Casual analysis of financial and operational risks of oil and gas companies in condition of emergent economy, in: A. Kiv (Ed.), 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, volume 2713 of *CEUR Workshop Proceedings*, CEUR-WS.org, 2020, pp. 41–52. URL: https://ceur-ws.org/Vol-2713/paper02.pdf.
[24] H. Danylchuk, O. Ivanylova, L. Kibalnyk, O. Kovtun, T. Melnyk, O. Serdiuk, V. Zaselskiy, Modelling of trade relations between EU countries by the method of minimum spanning trees using different measures of similarity, in: A. Kiv (Ed.), 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, volume 2713 of *CEUR Workshop Proceedings*, CEUR-WS.org, 2020, pp. 167–186. URL: https://ceur-ws.org/Vol-2713/paper13.pdf.
[25] S. Lehenchuk, A. Raboshuk, N. Valinkevych, I. Polishchuk, V. Khodakyvskyy, Analysis of financial performance determinants: evidence from slovak agricultural companies,

Agricultural and Resource Economics: International Scientific E-Journal 8 (2022) 66–85. doi:10.51599/are.2022.08.04.03.

[26] D. S. Shepiliev, S. O. Semerikov, Y. V. Yechkalo, V. V. Tkachuk, O. M. Markova, Y. O. Modlo, I. S. Mintii, M. M. Mintii, T. V. Selivanova, N. K. Maksyshko, T. A. Vakaliuk, V. V. Osadchyi, R. O. Tarasenko, S. M. Amelina, A. E. Kiv, Development of career guidance quests using WebAR, Journal of Physics: Conference Series 1840 (2021) 012028. doi:10.1088/1742-6596/1840/1/012028.

[27] A. O. Bielinskyi, I. Khvostina, A. Mamanazarov, A. Matviychuk, S. Semerikov, O. Serdyuk, V. Solovieva, V. N. Soloviev, Predictors of oil shocks. Econophysical approach in environmental science, IOP Conference Series: Earth and Environmental Science 628 (2021) 012019. doi:10.1088/1755-1315/628/1/012019.

[28] E. M. Matuska, J. Grubicka, Employer branding and internet security, in: Brand Culture and Identity: Concepts, Methodologies, Tools, and Applications, volume 3, IGI Global, 2018, pp. 1305–1326. doi:10.4018/978-1-5225-7116-2.ch070.

[29] I. Mavlutova, J. Kuzmina, I. Uvarova, D. Atstaja, K. Lesinskis, E. Mikelsone, J. Brizga, Does car sharing contribute to urban sustainability from user-motivation perspectives?, Sustainability 13 (2021) 10588. doi:10.3390/su131910588.

[30] A. Zvaigzne, I. Mietule, I. Kotane, S. Sprudzane, V. Bartkute-Norkuniene, Digital innovations in tourism: the perceptions of stakeholders, Worldwide Hospitality and Tourism Themes (2023). doi:10.1108/WHATT-06-2023-0080.

[31] I. Gontareva, V. Babenko, N. Shmatko, D. Pawliszczy, Correlation of Income Inequality and Entrepreneurial Activity, Journal of Optimization in Industrial Engineering 14 (2021) 33–38. doi:10.22094/joie.2020.677815.

[32] H. Danylchuk, L. Kibalnyk, O. Kovtun, A. Kiv, O. Pursky, G. Berezhna, Modelling of cryptocurrency market using fractal and entropy analysis in COVID-19, in: A. Kiv (Ed.), 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, volume 2713 of *CEUR Workshop Proceedings*, CEUR-WS.org, 2020, pp. 352–371. URL: https://ceur-ws.org/Vol-2713/paper40.pdf.

[33] M. A. Radin, V. Riashchenko, Effective pedagogical management as a road to successful international teaching and learning, Forum Scientiae Oeconomia 5 (2017) 71–84. doi:10.23762/FSO_VOL5NO4_17_6.

[34] S. Ramazanov, B. Tishkov, O. Chernyak, Non-linear Forecasting of the State of a Socio-eco-oriented Innovative Economy in the Context of Systemic Crises, in: A. Kiv, S. Semerikov, V. N. Soloviev, L. Kibalnyk, H. Danylchuk, A. Matviychuk (Eds.), Proceedings of the Selected Papers of the 8th International Conference on Monitoring, Modeling & Management of Emergent Economy, M3E2-EEMLPEED 2019, Odessa, Ukraine, May 22-24, 2019, volume 2422 of *CEUR Workshop Proceedings*, CEUR-WS.org, 2019, pp. 181–193. URL: https://ceur-ws.org/Vol-2422/paper15.pdf.

[35] K. V. Shymanska, M. Kurylo, O. Karmaza, G. Timchenko, Determinants of migration motives as a precondition for the migration flows formation, Problems and Perspectives in Management 15 (2017) 352–364. doi:10.21511/ppm.15(3-2).2017.05.

[36] V. Derbentsev, S. Semerikov, O. Serdyuk, V. Solovieva, V. Soloviev, Recurrence based entropies for sustainability indices, E3S Web of Conferences 166 (2020) 13031. doi:10.

1051/e3sconf/202016613031.
[37] H. Velykoivanenko, V. Korchynskyi, Application of Clustering in the Dimensionality Reduction Algorithms for Separation of Financial Status of Commercial Banks in Ukraine, Universal Journal of Accounting and Finance 10 (2022) 148–160. doi:10.13189/ujaf.2022.100116.
[38] N. V. Zachosova, Estimating the level of economic security for assets management companies, Actual Problems of Economics (2010) 111–119.
[39] P. Zakharchenko, G. Kostenko, S. Zhvanenko, V. Mukhin, Sustainable development of environment in the tourism destination areas: tourists' perception of the issue, IOP Conference Series: Earth and Environmental Science 628 (2021) 012024. doi:10.1088/1755-1315/628/1/012024.
[40] O. Ilyash, L. Taranenko, O. Trofymenko, N. Koba, M. Sobczak-Michalowska, Assessing the educational dimension of national economy innovative development, CEUR Workshop Proceedings (2023) 26–46.
[41] E. E. Fedorov, L. O. Kibalnyk, L. O. Petkova, M. M. Leshchenko, V. M. Pasenko, Fuzzy expert decision support system for foreign direct investment: a swarm metaheuristic approach, CEUR Workshop Proceedings (2023) 47–60.
[42] S. F. Lehenchuk, T. A. Vakaliuk, T. P. Nazarenko, Z. Kubaščíková, Z. Juhászová, The impact of intangible assets on the financial performance of Slovak ICT companies: a panel data regression analysis, CEUR Workshop Proceedings (2023) 61–81.
[43] P. Kulyk, V. Hurochkina, B. Patsai, O. Voronkova, O. Hordei, Maximizing customer satisfaction and business profits through Big Data technology in Society 5.0: a crisis-responsive approach for emerging markets, CEUR Workshop Proceedings (2023) 82–94.
[44] V. Porokhnya, V. Penev, R. Ivanov, V. Kravchenko, A flexible machine learning model for optimizing organizational capital development strategies and resource allocation, CEUR Workshop Proceedings (2023) 95–109.
[45] A. O. Bielinskyi, V. N. Soloviev, V. V. Solovieva, S. O. Semerikov, M. A. Radin, Recurrence quantification analysis of energy market crises: a nonlinear approach to risk management, CEUR Workshop Proceedings (2023) 110–131.
[46] A. O. Bielinskyi, V. N. Soloviev, S. V. Hushko, A. E. Kiv, A. V. Matviychuk, High-order network analysis for financial crash identification, CEUR Workshop Proceedings (2023) 132–149.
[47] P. M. Hryhoruk, N. A. Khrushch, S. S. Grygoruk, O. R. Ovchynnikova, Multidimensional statistical analysis of investment attractiveness and regional changes in the COVID-19 pandemic, CEUR Workshop Proceedings (2023) 150–167.
[48] V. D. Derbentsev, V. S. Bezkorovainyi, A. V. Matviychuk, O. M. Pomazun, A. V. Hrabariev, A. M. Hostryk, A comparative study of deep learning models for sentiment analysis of social media texts, CEUR Workshop Proceedings (2023) 168–188.
[49] H. B. Danylchuk, L. O. Kibalnyk, O. A. Kovtun, O. I. Pursky, Y. M. Kyryliuk, O. O. Kravchenko, The impact of the war in Ukraine on globalization processes and world financial markets: a wavelet entropy analysis, CEUR Workshop Proceedings (2023) 189–205.
[50] S. Kurkula, N. Maksyshko, D. Ocheretin, S. Cheverda, Nonlinear dynamics of electric vehicle sales in China: a fractal analysis, CEUR Workshop Proceedings (2023) 206–221.

[51] S. K. Ramazanov, B. O. Tishkov, O. H. Honcharenko, A. M. Hostryk, A cognitive approach to modeling sustainable development of complex technogenic systems in the innovation economy, CEUR Workshop Proceedings (2023) 222–235.
[52] D. H. Lukianenko, A. V. Matviychuk, L. I. Lukianenko, I. V. Dvornyk, University competitiveness in the knowledge economy: a Kohonen map approach, CEUR Workshop Proceedings (2023) 236–250.

Assessing the educational dimension of national economy innovative development*

Olha Ilyash[1,2], Larysa Taranenko[1], Olena Trofymenko[1,2], Nataliia Koba[2] and Marzena Sobczak-Michalowska[3]

[1]National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute", 37 Peremohy Ave., Kyiv, Ukraine, 03056

[2]International University of Finance, 37 Peremohy Ave., Kyiv, Ukraine, 03056

[3]University of Economy in Bydgoszcz, 2 Garbary, Bydgoszcz, Poland, 85-229

Abstract

The paper examines the educational indicators that reflect the innovative development of the national economy in Ukraine. The aim of the study is to develop a system for evaluating and enhancing the educational component of Ukraine's innovative development, which can support effective state regulation of educational processes and prevent the risks of reducing the educational security of the national economy. The study applies a multidimensional analysis of educational indicators, using a system of complex and systemic methods, such as dynamic analysis, system generalization, statistical methods, and taxonomic analysis. The study also compares the educational indicators of Ukraine with those of other countries that have achieved educational and scientific breakthroughs. The results show that Ukraine has a low level of educational performance and potential for innovative development compared to other countries. The paper proposes some measures to improve the educational component of Ukraine's innovative development, such as increasing public investment in education, enhancing the quality and relevance of education, fostering international cooperation in education and science, and promoting a culture of innovation among students and teachers.

Keywords

educational indicators, multidimensional analysis, innovative development, Ukraine, international comparison

1. Introduction

In the ever-evolving landscape of globalization, the rapid strides in scientific and technological advancement, particularly stemming from the development of artificial intelligence, information technology, and transdisciplinary knowledge, have brought about profound transformations. These changes extend beyond mere economic processes, reshaping an individual's role, position,

M3E2-MLPEED 2022: The 10th International Conference on Monitoring, Modeling & Management of Emergent Economy, November 17-18, 2022, Virtual, Online

*Extended and revised version of paper [1] presented at the 10th International Conference on Monitoring, Modeling & Management of Emergent Economy.

oliai@meta.ua (O. Ilyash); larysataranenko@gmail.com (L. Taranenko); o.o.trofymenko@gmail.com (O. Trofymenko); ashatunka@gmail.com (N. Koba); marzena.sobczak@byd.pl (M. Sobczak-Michalowska)

0000-0002-7882-3942 (O. Ilyash); 0000-0002-4533-9986 (L. Taranenko); 0000-0002-2339-0377 (O. Trofymenko); 0000-0002-6444-5872 (N. Koba); 0000-0002-9533-8316 (M. Sobczak-Michalowska)

and interactions within the contemporary global economic system. The current milieu is characterized by heightened interpersonal and inter-organizational competition, which collectively sets the stage for a paradigm shift.

This phenomenon places modern society at the cusp of what can be termed the fourth industrial revolution. This revolution encompasses the gradual emergence of a knowledge-based society intertwined with the concept of Industry 4.0 [2]. Central to this shift is an all-encompassing digitalization process, coupled with the emergence of novel professional domains demanding fresh knowledge and skill sets. Consequently, this necessitates innovative approaches to education and the development of human resources.

It is evident that the progression towards Industrial Revolution 4.0 hinges upon an antecedent educational revolution. This transformative process, referred to as Education 4.0, is pivotal for orchestrating an industrial breakthrough in the era of Industry 4.0. Central to this pursuit is the realization of the fourth Sustainable Development Goal laid out by the United Nations in 2015: "Ensure inclusive and equitable quality education and promote lifelong learning opportunities for all".

2. Literature review

The processes of formation and development of the Education 4.0 system should be based, in our opinion, on the provisions of innovation theory, the notion of knowledge-based economy, the theory of lifelong learning (LLL), of knowledge management, self-study and other present-day theories and ideas.

Defining the knowledge society, Kok [3] remarks that this notion encompasses all the aspects of human activities beginning from high-tech production up to artistic professions like in media and architecture, where knowledge is provided as the basis for added value creation. In his turn, Leiber [4], while considering the knowledge society, emphasizes the crucial role of the quality of education in general and that of higher education viewed as a social, economic and environmental factor. Agrawal et al. [5] highlight a high correlation between information communication technology (ICT) and knowledge management. Valero and Van Reenen [6] prove that increases in the number of universities are positively associated with future growth of GDP per capita. Supporting this view, Benešová and Tupa [7] underline the impact of technology not only on the emergence of knowledge-intensive products and services, but also emphasize its much greater impact on people's education in general. After all, only highly qualified and highly educated specialists will be able to control these technologies. It is clear therefore the increasing role of organizational education to adapt the people to changes that occur as a result of technological and economic innovation. The paper [8] explores the role of relational power and discursive positioning in the knowledge integration process using a definite interdisciplinary project as an example and thus emphasises the necessity of carrying out more research that explicitly explores power in the knowledge integration process.

Thus, recent research [9, 10, 11, 12] has shown that human capital makes a significant contribution to economic growth and technological development primarily through education, innovation and continuous growth. At the same time, the limited development of human capital leads to the use of natural resources as the main source of income, thereby reducing the level of the countries' economic development. In addition, the relationship between human capital and innovation at the country level is based on the fact that various forms of capital can be converted into resources and other forms of economic benefit. However, it is only a properly qualified human capital that can ensure the industrial and technological development of the country as well as can serve for its economic growth. Therefore, the assessment of education through the prism of country's innovative development, in our opinion, should definitely include indicators of the level of development of the country's human capital.

Nedelko et al. [13], as a result of studying the strategies and tools for knowledge management in innovation and Industry 4.0, emphasize that the use of the notion of knowledge management in Industry 4.0 should not only be encouraged but rather necessitated. It is well known that for the emergence of new knowledge and its commercialization, which is the essence of innovation, it is necessary to ensure close ties between industry and science and education [14]. Thus, we can state that education and science today are the starting point and the driving force to ensure the innovative development of business and, consequently, of national economies.

Considering the above-mentioned tendencies, the issue of measuring the effectiveness of educational process in accordance with the dynamic global socio-economic environment has become quite acute. There are a number of scientific papers substantiating the indicators of education performance for individual countries [15, 16, 17] as well as methodologies presented by various international organizations, which include educational indicators [18, 19, 20, 21, 22, 23]. However, given the urgent need to reform education within the process of Industry 4.0 formation, there is a necessity to search for new approaches to assessing educational indicators through the prism of their impact on the innovative development of the state, which will serve the increase in the efficiency of public administration and control.

Therefore, this paper is aimed at working out the system for assessing the educational component of the national economy innovative development, which will ensure the effective state regulation of educational processes as well as will prevent the country form the risks of reducing the educational security of the national economy. Accordingly, the main issue of the study is the definition and analysis of educational indicators, and the formation of recommendations on the development of educational indicators at various levels of state policy to ensure innovative development of the economy based on the study of effective global practices. The hypothesis lies in the idea that the growth of the indicators of education quality will lead to an increase in the metric of state innovative development.

3. Methods

In the course of the study the authors employed general scientific and statistical methods, as well as the method of a taxonomic analysis. As is known, a multidimensional statistical analysis is used to determine the largest number of features that will affect the object of study. That is why to define the degree of a cumulative impact of the factorial characteristics on the level of the national economy innovative development, the authors offered to apply the taxonomic method. The necessity to opt for this method is born out of the demand to search for a single integrating indicator out of the large number of indicators that characterize innovative development, which allows increasing the efficiency of public administration and control [24].

As the data base for the use of a taxonomic analysis we chose the educational indicators of innovative development of the national economy of Ukraine as of the years 2013–2019. Such indicators include the Human Development Index (HDI), the level of expenditure on education of GDP, the Education Index, the literacy rate (i.e. expected years of schooling), and Ranking of national higher education system.

Thus, the human development index is a combined index and an indicator of the educational component of the country's innovative development [22]. The index measures the country's achievements in terms of life expectancy, access to education, actual income of the citizens, and takes into account changes in the indices of socio-economic and gender inequality and multidimensional poverty. In addition, the human development index is adjusted by political, economic, social, and environmental factors, such as: human rights and civil liberties, participation in public life, social security, the degree of territorial and social mobility of population, the level of cultural development, access to information, health, unemployment, crime, environmental protection, environmental impact and others. It should be mentioned this index comprises the following data: the acquired human capital; the expected duration of children's education at school; results of the standardised testing of schoolchildren; the percentage of adult survivors and the proportion of children without any developmental disorders [21].

The literacy level of the country's population (expected time of schooling) is set by authors as a separate indicator of educational development because it indicates the general educational level of the population.

The education expenditures is one of the key indicators of innovative development. Innovative development of the domestic economy and strengthening the social component of state economic security can be ensured only by increasing human capital expenditures. Investing in education is a vital means of increasing human capital and prospects of the country's economic growth. Therefore, the level of expenditures on education of GDP was chosen as one of the indicators for taxonomic analysis.

The Education Index, which is a sub-index of the Human Development Index, should be also included to the educational indicators.

It is the economic development and competitiveness of the country that serve as the indicators of the country's economic security and largely depend on the number of educated and competent professionals, as well as technologies that increase their productivity. The higher education sector contributes significantly to realisation of these needs. In addition, in the modern world of alterglobalism, those high-quality higher education systems, which have broad links at the international level and contribute to the country's global development through the exchange of students, researchers, projects and ideas, demonstrate a high level of national economy. Therefore, one of the indicators of the educational component of the state innovative development at the global level is the ranking of national higher education systems (U21 Ranking of National Higher Education Systems) which enables assessing the overall higher education system of different countries at various stages of economic and social development [25].

To conduct a taxonomic analysis of the educational component of innovative development, it is rational to perform a sequence of the following methodological steps [26]:

- to form a matrix with the initial data necessary for the study of educational indicators of

innovative development;
- to standardize the values of the indicators matrix;
- to form a reference vector representing the growth of the innovative development component under study;
- to calculate the distance between individual variables and the reference vector; item to define the taxonomic indicator of innovative development.

In accordance with the outlined algorithmic steps, it is expedient to form an observation matrix based on the input data. It should be mentioned that in our study the units (I_i) are represented by the innovative development educational indicators. Within the scope of these indicated we single out the educational component of innovative development (E), while the years sand for characteristic features.

The construction of the matrix respresenting the initial data by components comprises the following steps:

The first step presupposes the use of $I^{(E)}$ for the matrix in order to reveal the educational component of innovative development (size 5×7).

At the second stage, since the indicators of innovative development have different measurement units, it is necessary to form a matrix of standardized values. This procedure is performed by replacing the criteria values with the coefficients standardized indicators [24] according to the following formula (1):

$$Z_i = \frac{I_i}{\bar{I}} \tag{1}$$

where:

I_i is the value of the i^{th} indicator;

$\bar{I}$ is the average value of the i^{th} indicator.

After indicators' standardization, to carry out a further taxonomic analysis, the features of the observation matrix are to be divided into those of stimulators and destimulators that determine the direction of the impact on the national economy innovative development. In this case, stimulatory factos have a positive effect on the development level, while destimulatory factors restrain.

Differentiation of the studied factors into stimulating and destimulating ones is given in the table 1.

Table 1
Educational indicators of innovative development (stimulator / destimulator).

Symbol	Indicator	Stimulator or destimulator
E	Educational component	
I_1	Human Development Index	Stimulator
I_2	Level of expenditures on education of GDP, %	Stimulator
I_3	Education index	Stimulator
I_4	Literacy rate of the country's population (expected years of schooling)	Stimulator
I_5	Ranking of the national higher education system	Stimulator

The division of the indicators into stimulators and destimulators can serve as the basis for finding out the ideal reference vector and forming the values of the indicators [24] in the following way:

$$\begin{cases} I_{oi} = \max I_{ij} \quad (stimulator) \\ I_{oi} = \min I_{ij} \quad (destimulator) \end{cases} \tag{2}$$

After that, we receive a vector-standard of the innovative development level within the framework of educational component. To calculate the integrated taxonomic index, it is necessary to find the average distance ($\overline{C_0}$), the mean value of the square root of the average square of the difference of values of characteristics (S_0), deviation of the distance between the point-unit and the upper pole point from the value of characteristics distance (d_i) for the educational component of innovation development according to the following formulas (3-6):

1) average distance:

$$\overline{C_0} = \frac{1}{m} \sum_{i=1}^{m} C_{i0} \tag{3}$$

where:

m – the number of periods;

C_{i0} – the distance between the point-unit and the point E_{M4};

2) the mean value of the square root of the average square of the difference between the values of characteristics:

$$S_0 = \sqrt{\frac{1}{m} \sum_{i=1}^{m} (C_{i0} - \overline{C_0})^2} \tag{4}$$

$\overline{C_0}$ – the average distance;

C_{i0} – the distance between the point-unit and the point E_{M4};

3) deviation of the distance between the point-unit and the point the reference vector from the value of characteristics distance:

$$C_0 = \overline{C_0} + 2S_0 \tag{5}$$

$$d_i = \frac{C_{i0}}{C_0} \tag{6}$$

where:

S_0 – the mean value of the square root of the average square of the difference of values of characteristics;

C_{i0} – the distance between the point-unit and the point E_{M4};

C_0 – the distance.

On the basis of the obtained results, we can define the taxonomic indicator of the level of the system economic security by the following formula (7):

$$K = 1 - d \tag{7}$$

where:

d – deviation of the distance between the point-unit and the point E_{M4} from the value of characteristics distance.

Thus, the obtained indicator can acquire higher values with the higher values of stimulants, and, consequently, lower ones with low values of stimulants. By calculating this indicator, we will be able to analyze the directions and scales of changes that occur in the system under study, in particular, in the innovation system of the national economy on the basis of one synthetic feature of the educational component.

4. Results

For the innovative development of the national economy, the development of human capital assets undoubtedly remains the driving force. That is why the study of educational and scientific direction of innovative economic development, which is the basis for ensuring the capital development, is of an urgent need. At present, the model of education aimed at training highly qualified industry personnel is almost completely lost, while industrial enterprises cannot function efficiently with a shortage of specialists having an up-to-date training. In order to preserve the industrial potential, the structure of innovative education should form a symbiosis of higher education institutions, research institutes, production facilities of industrial enterprises and public authorities. However, the majority of organizations cannot make the use of human capital resources since they are limited by the approaches aimed at performing specific tasks rather than being focused on research and development. At the same time, industrial enterprises agree that managing human capital development is one of the priorities of innovation progress.

Unfortunately, in Ukraine there is a significant gap between the knowledge and competencies of the students who graduate almost without any practical experience, forcing employers to spend time preparing them for a particular job [27, 28]. To build innovative human capital, the educational system should include more practical skills, in particular through the integration of business into the educational process, and, thus, provide the generation-to-come with the up-to-date theoretical and innovative practical tools.

The research results of the different countries' experience on the development of education are given below.

In particular, such countries as Singapore, China, India, South Korea, USA, Japan, Finland, Brazil, and others demonstrated the educational breakthrough achieved through high quality STEM education, due to increase in expenditures on education, by means of supporting fundamental research and a number of programs aimed at developing the level of general public digital competence as well as thanks to the growth in patent productivity in priority sectors of the economy.

Singapore, for instance, has introduced programs to set up technical schools and international corporations to train unskilled workers in the spheres of information technology, petrochemistry and electronics. The strategy of involving Singapore's multinational organizations in training its workforce has contributed in the long run to the country's economic prosperity. As a result of the education system reform taken place in Singapore, minimum compulsory educational standards have been introduced for all schools, English has become an obligatory discipline for all types of schools and a number of other subjects are tought in English. The government is constantly investing in the education of Singaporean students in the best universities world, while creating at the same time leading research and educational centers in Singapore. Distinguished results in education have been achieved due to implementation of the Plan of Research and Innovation Enterprises completed by 2020 [29].

China's economic reform program was based on lowering government norms on prices in education sphere and increasing investment in education of personnel [30].

India initiated a policy of promoting the quality of the workforce. The introduction of the language law made English a second national language, which contributed to the growth of the country's technological development and signing international agreements with the leading companies in the IT sector.

South Korea, the USA, Japan, and China have the policy of increasing the share of GDP in research and development (R&D), which results in the growth of intensity and quality of research and development [31].

In Finland, there are programs to expand cooperation with foreign experts since due to the aging of its population in some industries, there is a shortage of workforce. There are also programs aimed at financial support of innovations in R&D. One euro invested in innovation for research brings about 10-20 euros net profit, which corresponds to 70% of investments (compare, for example, in Russia they spend 10% of investments, while in France 90%).

In Brazil, there is the Bolsa Familia Income Transfer Programme (Family Assistance), which provides monetary benefits to families who send their children to school. The government has also introduced the policy of increasing investments in education to strengthen human resources.

It seems inevitable that automation, digitalization and other forms of technology will put an end to millions of jobs and will create new opportunities for the workforce. At the same time, it is of vital importance to prepare the next generation of workers to participate in the development of Industry 4.0. It is education that should be the driver of future skills by means of going beyond the traditional teaching, including entrepreneurship, soft leadership, technology and workforce readiness. Accordingly, it proves the need to provide such the condition for the development of human resources required to meet the changing demands of the twenty-first century as the improvement of quality of primary schools education.

Since the adoption of the Universal Declaration of Human Rights in 1948, countries have been making efforts to universalize primary education. However, the quality of education is going down due to the low quality of primary education. Thus, in 2019, Ukraine demonstrates the lowest level of quality of primary education as compared to the world's leading countries. The percentage of students with the highest results in at least one field (reading, mathematics, and natural sciences) is only 7.5%, which is from three to five times lower than in China (49.3%), Singapore (43.3%), South Korea (26.6%) and Canada (24.1%), while it accounts for 50% of the quality of primary education in Denmark (15.8%). Some of the advantages of high quality primary education in Singapore and South Korea comprise effective leadership, quality teacher advanced training, high salaries for teachers, teachers' professional development strategy, a high percentage of modern equipment supply, the ability to work with innovative interactive technologies as well as social security of teachers [23]. In the USA, the Regional Councils for Economic Education and the State Federal Reserve offer the teachers an annual weekly summer training program based on the model "Key to Financial Success". A comparison of the main indicators of the educational and scientific breakthrough of Ukraine with countries of progressive development in 2019 is presented in the table 2.

Table 2
Indicators of educational breakthrough in Ukraine and the countries of progressive development in 2019 (based on [18, 22, 21, 23]).

Indicators	Ukraine	Finland	Germany	Denmark	USA	Canada	Australia	South Korea	China	Singapore	Japan
Quality of primary education (% of students with the highest results in at least one of the disciplines: (reading, mathematics, natural sciences, level 5 or 6)	7,50	21,00	19,10	15,80	17,10	24,10	18,90	26,60	49,30	43,30	23,30
The literacy level of the country's population (the average number of years spent studying) and the expected duration of education	10,40	17,10	12,40	12,60	13,40	13,80	12,40	12,10	7,80	11,90	12,80
Index of digital competencies of economically active population	4,50	5,10	5,80	5,40	5,30	5,10	5,00	5,00	4,70	5,60	4,40
Expenditures on education, % of GDP according to IMD data	6,00	5,70	4,10	6,50	6,00	4,40	5,00	5,00	3,60	2,70	3,20

Over the last two centuries, alongside the increase in the number of students acquiring primary education, there has been a steady increase in the level of literacy of the world population. However, in some African countries, the literacy rate among young people is still below 50.0%. According to the level of population literacy, Ukraine ranks 51st in the world ranking of competitiveness. Thus, the average number of years spent on studying and the expected duration of education in Ukraine is 10,4 years, which sets it ahead of China (7,8 years) by 2,6 years and below Germany (17,1 years), Canada (13,8 years) and the United States (13,4 years). Among the possible threats of insufficient literacy in Ukraine's population, to name but a few, are the state's non-recognition of other than official forms, formats and methods of training, lack of the culture of dual education within the framework of labor relations, lack of employers' interest in financing employee training, the employees' insufficient practical and soft skills [32].

As for the education in Japan, it is almost 32% funded by private sources. The education in Denmark is entirely funded by Danish government which guarantees free education for all [31]. In the Netherlands and other countries of Northern Europe, the government has developed programmes to provide opportunities to participate in formal and / or informal education for adults and the unemployed (64% and 57% of adults and the unemployed have already participated in these programmes).

A feature of the national education system is a relative high level of funding, i.e. the maximum amount of expenditures is allocated on education. Thus, the indicators of education financing in Ukraine in 2019 exceed the average indicators of the OECD countries (by 6% of GDP) [33], in particular Denmark (6,5% of GDP), the United States (6%), Finland (5,7%) and are 2,5 times higher than in Singapore (2,7%), Japan (3,2%), and China (3,6%). A common feature of the leading countries in terms of financing higher education with the insignificant level of public sector spending is a high share of of funding provided by the private sources, in particular, in the United States – 26%, Australia – 23%.

Digital competencies of the population play quite an important role in the educational and scientific breakthrough [34, 35]. However, as of 2017, according to the Digital Skills Index of the European Digital Economy and Society Index (DESI), almost half (44,0%) of the EU population does not have the necessary skills to use digital technologies [36]. To ensure an educational breakthrough in Ukraine, the Digital Agenda of Ukraine – 2020 was adopted with an aime to use digital technologies, create a digital society and ensure the competitiveness of the country and its citizens. Thanks to digitalization, Ukraine is able to reduce the gap in international indicators of competitiveness [37], since in 2019 it ranked 56th with the index 4,5 economic units (EU) and as compared to 2016 it increased the level by only 8 points. The leading countries are Finland (5,8), Singapore (5,6), Denmark (5,4) and the United States (5,3). With the support of the National Library Council and the Singapore Cybersecurity Agency, curricula have been updated with a view to providing better cybersecurity awareness and acquiring skills to detect fake news and protect onself against it. Besides, the government introduced a route map of teaching technology intensity, which is a three-year plan to help small and medium-sized enterprises (SMEs) accelerate technology implementation and help the population expand their digital capabilities (e.g., a Memorandum on cooperation between SkillsFuture Singapore and Microsoft has been signed aimed at helping make profit for 5000 people and 100 small- and medium-sized enterprises). Denmark has developed the Digital Literacy Manifesto, which has inspired politicians to reflect on digital skills and transformation.

We shall remark here that the process of monitoring the educational component of innovation development should be carried out within the system of indicators: correlation between the level of education expenditures to GDP, human development index, education level index, literacy level and the rating of the higher education system. The system of the mentioned indicators is presented in the table 3.

Table 3
The rate of change of the educational component indicators signifying the innovative development of Ukraine within the period of 2013–2019 (based on [18, 21, 22]).

No	Indicators	Years							Growth rate, %		
		2013	2014	2015	2016	2017	2018	2019	$\frac{2019}{2018}$	$\frac{2019}{2015}$	$\frac{2019}{2013}$
	Educational indicators										
1.	Human development index	0,74	0,75	0,74	0,75	0,75	0,75	0,77	102,67	104,05	104,0
2.	Level of education expenditures to GDP, %	5,90	6,20	6,70	6,70	6,00	5,90	5,00	84,75	74,63	84,75
3.	Education level index	0,791	0,8	0,791	0,792	0,794	0,792	0,799	100,88	101,01	101,01
4.	Literacy level of the country's population (expected time of schooling)	14,90	15,00	14,90	15,10	15,10	15,10	15,10	100,00	101,34	101,34
5.	Ranking of the national higher education system	49,9	43,9	44,0	42,1	47,7	47,4	45,1	95,15	102,50	90,38

According to the report submitted by UNDP in 2019, Ukraine ranked 74th in the human development index among 183 countries surveyed. Given the significant part of the population with higher education (82%), Ukraine has an average human development index among European countries, which is 0,779 points. At the same time, the share of people engaged in scientific activities is only 0,34% out of the total employed population.

It should be added that 66 countries of the world belong to the category of high level of human development, among which: Switzerland (2nd place, 0,955 points), Germany (6th place, 0,947 points), Great Britain (13th place, 0,931 points), Canada (16th place, 0,929 points), Estonia (29th place, 0,892 points), Lithuania (34th place, 0,882 points), Poland (35th place, 0,880 points), Romania (49th place, 0,828 points), Kazakhstan (51st place, 0,825 points), Russia (52nd place, 0,824 points), Belarus (53rd place, 0,823 points), Turkey (54th place, 0,820 points), Bulgaria (56th place, 0,816 points), Georgia (61st place, 0,812 points), Serbia (64th place, 0,806 points).

The World Bank estimates the human capital index in Ukraine at 0,63 points, so Ukraine ranks 50th among 157 countries. Thus, a child born in Ukraine can rely upon acquiring only 63% of the potential level of human capital, being possible only under condition of receiving complete education and having a good health. In terms of figures, Singapore (88%), Hong Kong (81%) and Japan (80%) have been ranked first for several years in a row. The top ten countries also include South Korea, Canada, Finland, Macau and Sweden (about 88%), Ireland and the Netherlands (79%). The lowest positions in ranking are occupied by the CAR (29%), Chad (30%) and South Sudan (31%).

The comparison of the index of human capital with GDP per capita in the world allows us to trace the correlation between the level of the country's economic development and the level of education received by human capital (figure 1). Thus, in general, the countries with the highest human capital index have higher GDP per capita: Singapore (0,88 and 652333), Japan (0,80 and 40246), Canada (0,80 and 46194), Finland (0,80 and 48782), Switzerland (0,80 and 51615), Ireland (0,79 and 78660,96). The lowest positions are occupied by Congo (0,37 and 553), Yemen (0,37 and 774), Rwanda (0,38 and 820), Ethiopia (0,38 and 855), Burundi (0,39 and 261). It should be added that in 2019 the value of the human development index in Ukraine was 0,63, while GDP per capita amounted to 3656 [21].

The total amount of education funding (from public, local and private sources) varies from 5,0% to 6,7% of GDP and is characterized by declining dynamics. Although the Law of Ukraine "On Education" states that education funding should comprise at least 7% of GDP, in 2019 the amount of financial support in this area was only 5,0% and the share of expenditures on education in the budget of Ukraine comprised 17,1% [38]. Examining the dynamics of this indicator changing throughout 2014–2019, we can single out two periods: the 1st period of 2014–2016 is marked by a decrease of 3,6 percentage points, while the 2nd period (2016–2019) is characterised by an increase of 1,6 percentage points.

We should also point out that starting from the year 2015, the consolidated budget expenditures on education in GDP have also been decreasing. Thus, despite its unstable dynamics, expenditures on education as a percentage ratio of GDP amounted to 6,1% in 2019, while in 2016 it was 5,4%, which is 0,9% lower than in 2014. At the same time, it should be noted that in comparison with the EU countries, Ukraine spends much more on education. Thus, the total expenditure on education from GDP in Poland is 4,6%, in Latvia – 4,7%, in Italy – 3,8%, Germany – 4,8%, Estonia – 5,2%, Switzerland – 5,1%, and Romania – 3,0%. The high level of

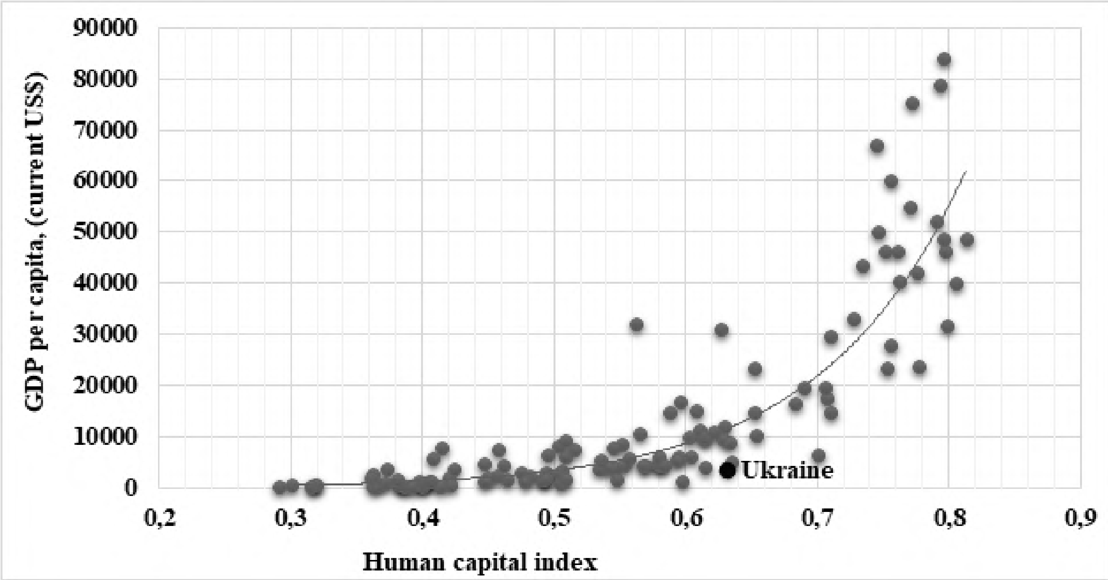


Figure 1: The ratio of the human capital index and GDP per capita of Ukraine and the world in 2019 (based on [21]).

expenditures on education in Ukraine is explained by the fact that the majority of Ukrainian higher education institutions are financed from the state budget (72%), while in other countries a significant share is made up of private educational institutions (43%).

Another key indicator of innovative development in the context of ensuring the economic security of the country is the Education Index, which is a sub-index of the Human Development Index [22]. The optimal value of the indicator for the developed countries is no less than 0,8 points. That is why in 2019 Ukraine occupied the 46th rank (0,797 points) which testifies a significant achievement of the country's population in education in terms of adult literacy and the total number of students receiving education. Figure 2 shows the general dynamics of the education index among other indicators comprising the educational component of innovative development considered within the analysed period.

The education index also allows estimating the average number of, as well as the expected duration of education of the population, which in Ukraine correspond to 11,4 and 15,1 years respectively. Leading positions in the world are occupied by Australia (22,0 years), Belgium (19,8 years), Sweden (19,5 years), Finland (19,4 years), Iceland (19.1 years), Denmark (18.9 years), New Zealand (18,8 years), Ireland (18,7 years), the Netherlands (18,5 years), and Norway (18,1 years). We will emphasise that creation of conditions for lifelong learning ensures the adaptation of labor capital to rapid technological changes, and, consequently, accelerates economic development and serves for the growth of national economy competitiveness.

For instance, in 2020, Ukraine ranked 36th. If to evaluate this indicator in terms of its individual constituents, we can see that according to the degree of investments from both the private and public sectors Ukraine occupies the 27th place (52,6 points); as to the public policy and regulation as well as the possibility of acquiring education, Ukraine takes 39th place (70,6 points); according to the level of international cooperation, which demonstrates the degree of

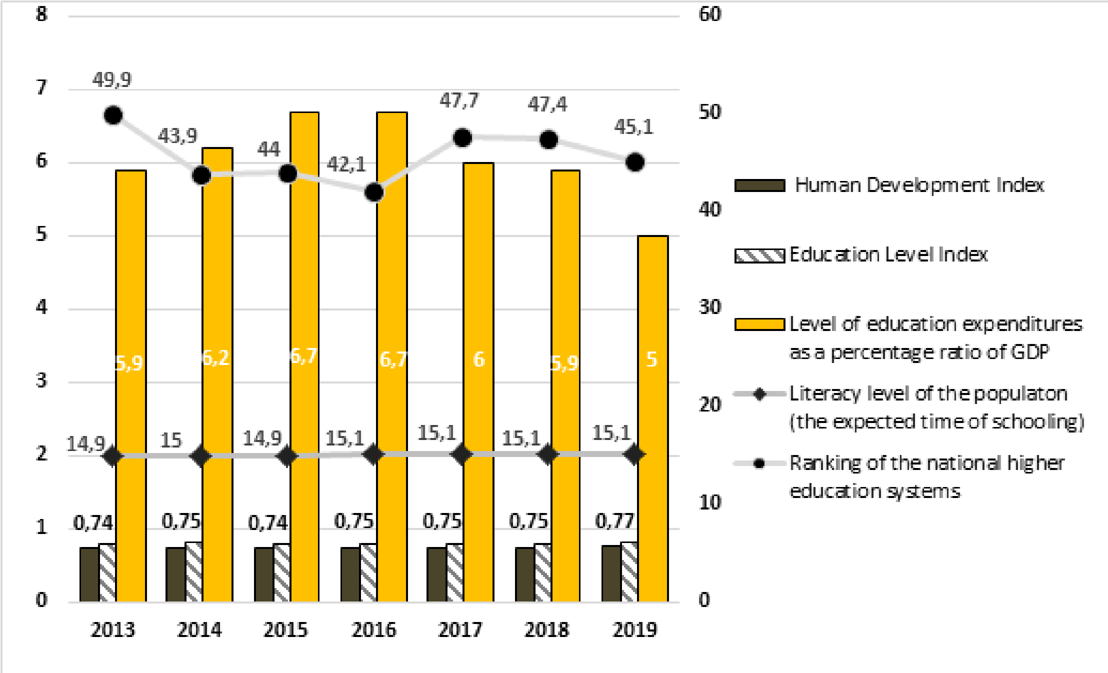


Figure 2: Changes in the education index within the system of indicators comprising the educational component of Ukrainian innovative development from 2013 to 2019 (based on [25, 21, 22]).

openness of the higher education system, Ukraine is at the 38th place (40,4 points); considering the quality of scientific research, scientific publications, compliance of higher education with the demands of the national labor market, including further employment of educational institutions' graduates, Ukraine occupies 42nd place with a score of 28,7 points. At the same time, the highest ranking include the USA, Switzerland, Denmark, Singapore, Sweden, Great Britain, Canada, Finland, Australia, the Netherlands, and Norway, where the overall indicator value is no less than 80 points.

It is worth noting that recently there has been a tendency to reduce the number of higher education institutions. Thus, in 2019, the Ministry of Education and Science of Ukraine granted the right to carry out activities in the field of higher education only to 281 educational institutions, which is 25% less than in 2010. A similar tendency is typical of the indicator "the number of students per 10000 of population", which during the last 9 years dropped 0,6 times from 476 to 302 students. The reasons for this negative tendency is the decrease in the birth rate, which, in its turn, led to a reduction in the number of university applicants and consequently in a number of students in 2019 by 63% as compared to 2010 (figure 3). Regarding the academic and teaching staff in the field of education, it should be mentioned that there are 14,25 students per teacher in secondary schools, and 10,75 students per one teacher in higher educational institution, which is 1,5 times lower comparing with the average indicator in the majority of economically developed countries.

As for education expenses, there has been a gradual increase in private sector expenditures on R&D (by UAH 177 billion or 66,6% within the period of 2013–2019), the following indices

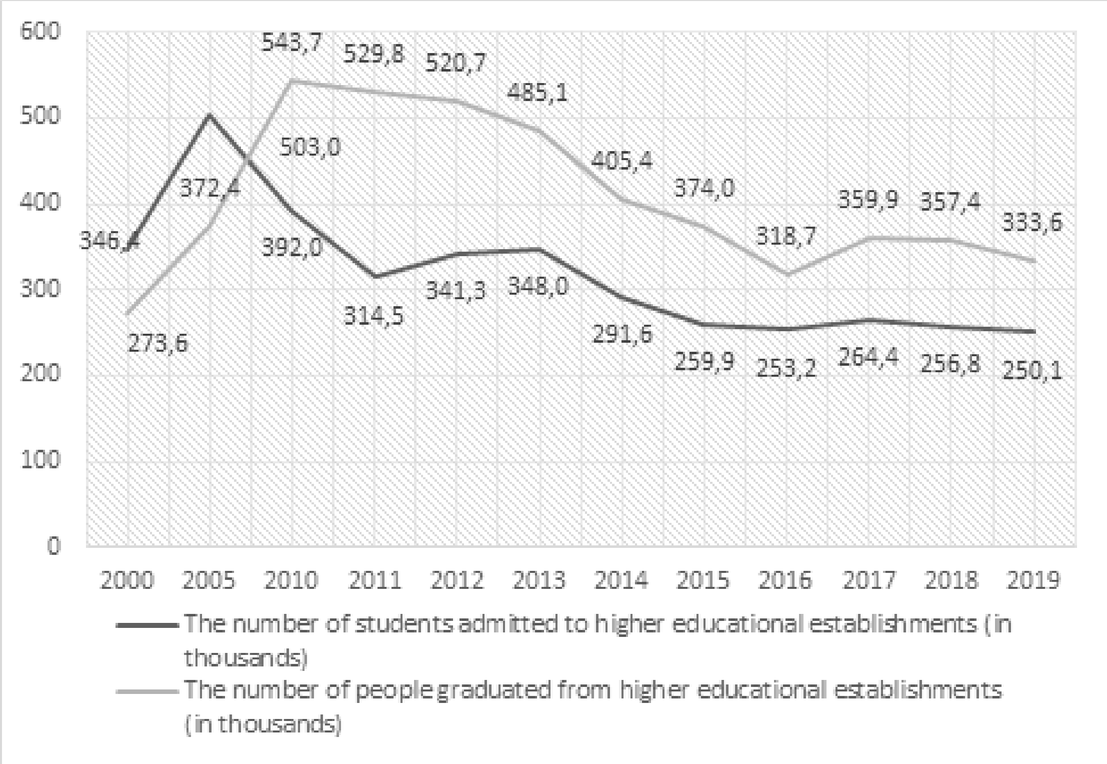


Figure 3: Dynamics of the indicators of higher education in Ukraine in 2000, 2005 and during the period from 2010 to 2019 (based on [21, 22, 25]).

remain low: the levels of expenditure on education (5,0% of GDP in 2019 and a decrease by 0,9 percentage points by 2013), implementation of scientific and technical work (0,43% and 0,27 percentage points, respectively), ranking of the national system of higher education (45,1% and a decrease by 4,8 percentage points by 2013), the amount of scientific and technical work being realized (0,3% to GDP and a decrease by 0,17 percentage points by 2013) and specialists who perform scientific and technical work (0,48% out of the total number of employees and a decrease by 0,32 percentage points by 2013). It is also observed declining in the level of patent productivity, reduction in the number of specialists involved in R&D implementation, failing to maintain leading positions in public funding of scientific and technical work, budget funding including [38].

To overcome the negative phenomena in the field of education, it is necessary, first of all, to build the relevant skills and competencies in specialists-to-be through a STEM-oriented approach to teaching and learning [39, 40]. Thus, in the UK, to meet the demands for specialists in the STEM sphere it is necessary to have more than 100 thousand people graduate by 2020, while in Germany there is a shortage of 210 thousand workers in natural sciences, mathematics, technology and computer science. However, to ensure the dynamic development of the national economy, it is essential to either significantly increase the intellectual potential of STEM specialists, or there should be a transition of the economic development to a new phase with the use of technologies of a new generation [41]. For the development of digitalization processes it is important also taking into account such world experience as Digital Competence Program, the Danish "Digital Literacy Manifesto" [37].

Thus, modernisation of the education system must be carried out taking into account the directions of the international economy development as well as defining the role and place of the country in the global dimension. With such an approach to the education system formation in a short period of time Ukraine will have had the trained personnel of necessary specializations.

After studying the above mentioned factors influencing education development in Ukraine as well as in other countries, we have applied the taxonomic analysis. First, we compiled a matrix of input data I^{M4}:

$$\begin{bmatrix} 0{,}74 & 0{,}75 & 0{,}74 & 0{,}75 & 0{,}75 & 0{,}75 & 0{,}77 \\ 5{,}90 & 6{,}20 & 6{,}70 & 6{,}70 & 6{,}00 & 5{,}90 & 5{,}00 \\ 0{,}79 & 0{,}80 & 0{,}79 & 0{,}79 & 0{,}79 & 0{,}79 & 0{,}79 \\ 14{,}9 & 15{,}0 & 14{,}9 & 15{,}1 & 15{,}1 & 15{,}1 & 15{,}1 \\ 49{,}9 & 43{,}9 & 44{,}0 & 42{,}1 & 47{,}7 & 47{,}4 & 45{,}1 \end{bmatrix}$$

The next step was the formation of a matrix of standardized values of Z^{M4}:

$$\begin{bmatrix} 0{,}99 & 1{,}00 & 0{,}99 & 1{,}00 & 1{,}00 & 1{,}00 & 1{,}03 \\ 0{,}95 & 1{,}00 & 1{,}08 & 1{,}08 & 0{,}97 & 0{,}95 & 0{,}91 \\ 1{,}00 & 1{,}01 & 1{,}00 & 1{,}00 & 1{,}00 & 1{,}00 & 1{,}01 \\ 0{,}99 & 1{,}00 & 0{,}99 & 1{,}00 & 1{,}00 & 1{,}00 & 1{,}00 \\ 1{,}10 & 0{,}97 & 0{,}97 & 0{,}93 & 1{,}05 & 1{,}04 & 0{,}99 \end{bmatrix}$$

According to the results of the chosen methodology application, we received the vector of the educational component being the standard of the level of innovative development within the framework of the educational component:

$$E^{M4} = (1.03; 1.11; 1.01; 1.00; 1.10)$$

At the next step of our research we calculated the taxonomic indicators of the educational component of the country's innovative development.

The closer the value of the taxonomic coefficient to one, the greater is the impact of a particular educational indicator on the national economy innovative development. The calculations of taxonomic indicators of the educational component of the national economy innovative development within the period from 2013 to 2019 are given in the table 4.

The taxonomic indicator's dynamics in the context of the educational component of the country's innovative development is shown in figure 4.

Despite the average value of the human development index of European countries, the taxonomic indicator of the educational component as a part of innovative development is characterized by a low share (0,008).

Thus, the taxonomic analysis of the educational component of the national economy innovative development proves its negative dynamics, which signifies the importance of taking into account the changes in educational indicators as well as timely response to these changes for the effective state regulation of the human capital development, which serves as the basis for

Table 4
Values of taxonomic indicators of the educational component in the system of the country's economic security within the period from 2013 to 2019.

Title	Indices	Years							Values
		2013	2014	2015	2016	2017	2018	2019	
Educational component (M4)	$C_{i0}^{(4)}$	0,139	0,157	0,136	0,173	0,128	0,146	0,300	$\overline{C_0} = 0,179$
	$d_i^{(4)}$	0,499	0,564	0,488	0,620	0,459	0,522	0,992	$S_0 = 0,055$
	$K_i^{(4)}$	0,501	0,436	0,512	0,380	0,541	0,478	0,008	$C_0 = 0,279$

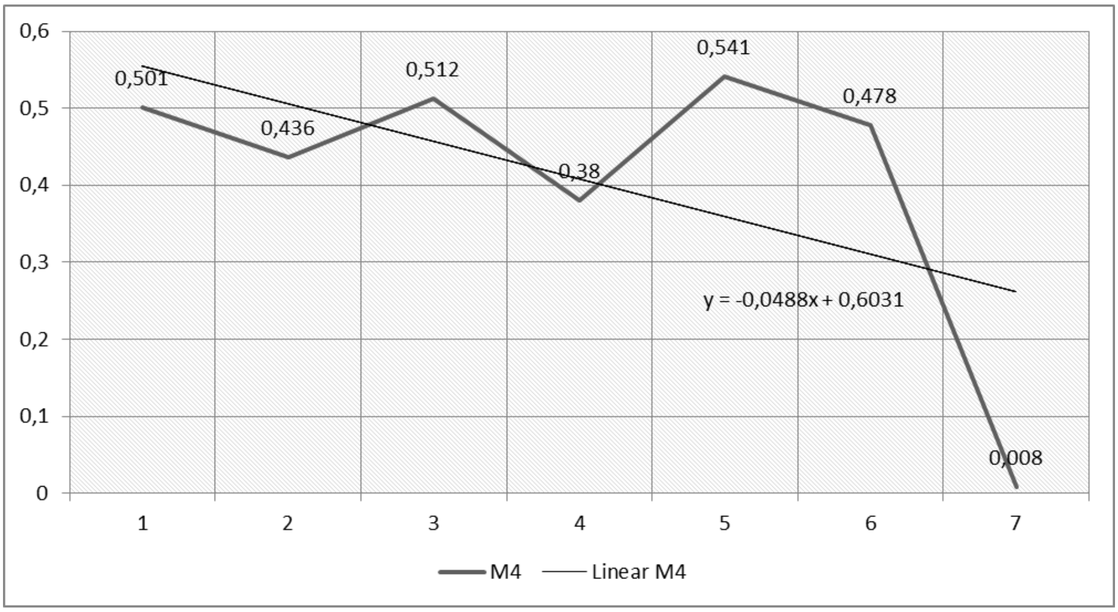


Figure 4: Dynamics of the taxonomic indicator of the country's innovative development (from a perspective of an educational component) within the period from 2013 to 2019.

the development of innovation and prevention from the risks of reducing educational security of the national economy.

5. Discussion and conclusions

The developed in the paper advanced system for assessing the educational constituent of the national economy innovative development, performed on the basis of taxonomic analysis allows taking into account the main indicators of the educational development in the country, thus simplifying the analysis of their impact on the national innovation system and, thus, serves as an effective tool for finding optimal solutions in the state regulation of educational processes as the basis for innovative activities of all its participants.

As a result of the performed research, it was found out that the potential of the educational and scientific components of the national economy of Ukraine innovative development is not realized to its full. These results elucidate the reasons for Ukraine's weak position in the state innovative economic development, at present being one of the most important factors of economic security. Therefore, judging from the research results, in order to guarantee the national economy innovative development, the primary task is to ensure the growth of taxonomic indicators within the educational component.

In particular, the lack of systematic managerial decisions by state authorities in solving problems of educational policy development requires scientific research to put forward proposals and set primary functions of corresponding ministries and agencies in order to ensure the innovative development of the national economy. According to the results of the present study the primary functions of the state policy to ensure growing of the educational component of the national economy innovative development have been defined. In particular, at the level of the Ministry of Education and Science of Ukraine it is expedient to introduce the Strategy of accelerated formation of the teachers' educational potential. This will contribute to the spread of innovation in the education system of Ukraine in the context of global digitalization. Based on the high indicators of education in Japan, Denmark and the Netherlands obtained in the process of the conducted research and taking into account the directions of their development,it is advisable to introduce the programs of learning English as a first foreign language into primary school; to develop, in cooperation with private educational institutions, educational programs aimed at introducing up-to-date teaching methods and programs to provide the opportunities for adults and the unemployed to take part in formal and / or informal education. A number of programs should be initiated to start technical schools and fee-paying international corporations to train unskilled workers in the field of information technology, petrochemistry and electronics (those who could not get industrial jobs, the government could increase the number of labor-intensive retail services such as in the sphere of tourism and transport). Taking into account the analysis of the development of human capital, it is advisable to take the experience of Singapore as a basis,and to recommend to develop a Strategy for Involving Multinational Organizations in Labor Training in Ukraine, and to work out a Strategy for uniting the country's largest technical universities with the purpose to develop advanced clusters for the research of future technologies and ensure the presence of the association members in public, political and economic circles to warrant a high level of training of future personnel.

At the level of the Department of Scientific and Technical Development and the Ministry of Economy of Ukraine it is worthwhile to introduce a New Plan for the Development of Research Innovative Enterprises until 2025, that allows to take into account the successful USA experience.

At the level of the Ministry of Finance of Ukraine, alongside the Ministry of Economy of Ukraine, it is important to assist small and medium enterprises (SMEs) in accelerating technology implementation and helping the population expand its digital potential.

At the level of the Cabinet of Ministers of Ukraine it is expedient to introduce Programs of funding educational institutions regardless of the form of ownership, to develop an effective formula of mixed financing in various proportions, which implies a gradual reduction of state funding while increasing the share of private funding (Singapore experience), to stimulate government to invest in the education of Ukrainian students in the best universities in the world, while creating leading research and educational centers at home.

At the level of the Ministry of Digital Transformation of Ukraine, it is recommended to introduce an online resource with a view to increasing the digital competence of citizens, following the example of Digital Competence. The introduction of the analogue of such a program will give impetus to digital initiatives, the increase in digital competencies of Ukrainian citizens as well as the use of the Danish "Digital Literacy Manifesto", which will stimulate the development of digital skills and critical thinking.

Implementation of the suggested measures within the framework of achieving educational and scientific breakthrough will allow to increase the overall indicator of the educational component in order to ensure the national economy innovative development.

The limitations of the study are impossibility to take into account the state of education in Ukraine during the war. In our opinion, the deterioration of the state of the educational component of the state's innovative development is expected and obvious. Among the positive expected educational trends is the growth of digital literacy of the population and the development of new technologies. Therefore, the study of the educational component and the development of mechanisms to ensure its strengthening in the post-war period will be an important and priority area of the future research.

Acknowledgments

This study was carried out in the framework of cooperation with the Ministry of Economy of Ukraine (was named the Ministry of Economic Development, Trade and Agriculture of Ukraine at the moment of the project) under the project "Synergy of knowledge, experience and creativity for the future".

References

[1] O. Ilyash, L. Taranenko, O. Trofymenko, N. Koba, M. Sobczak-Michalowska, Multidimensional Analysis of Educational Indicators of the National Economy Innovative Development, in: S. Semerikov, V. Soloviev, A. Matviychuk, V. Kobets, L. Kibalnyk, H. Danylchuk, A. Kiv (Eds.), Proceedings of 10th International Conference on Monitoring, Modeling & Management of Emergent Economy - M3E2, INSTICC, SciTePress, 2023, pp. 23–37. doi:10.5220/0011930800003432.

[2] V. Kobets, V. Yatsenko, Influence of the Fourth industrial revolution on divergence and convergence of economic inequality for various countries, Neiro-Nechitki Tekhnolohii Modelyuvannya v Ekonomitsi 2019 (2019) 124–146. doi:10.33111/nfmte.2019.124.

[3] W. Kok, Facing the challenge: the Lisbon strategy for growth and employment, Office for Oficial Publications of the European Communities, Luxembourg, 2004. URL: https://www.bmuv.de/fileadmin/bmu-import/files/pdfs/allgemein/application/pdf/facing_the_challenge.pdf.

[4] T. Leiber, Impact evaluation of quality management in higher education: a contribution to sustainable quality development in knowledge societies, European Journal of Higher Education 8 (2018) 235–248. doi:10.1080/21568235.2018.1474775.

[5] A. Agrawal, C. Kumar, S. K. Mukti, Role of Information and Communication Technology (ICT) to Enhance the Success of Knowledge Management (KM): a Study in a Steel Plant, Journal of the Knowledge Economy 12 (2021) 1760–1786. doi:10.1007/s13132-020-00694-6.

[6] A. Valero, J. Van Reenen, The economic impact of universities: Evidence from across the globe, Economics of Education Review 68 (2019) 53–67. doi:10.1016/j.econedurev.2018.09.001.
[7] A. Benešová, J. Tupa, Requirements for Education and Qualification of People in Industry 4.0, Procedia Manufacturing 11 (2017) 2195–2202. doi:10.1016/j.promfg.2017.07.366.
[8] Å.-K. Engstrand, C. Enberg, The power in positionings: A Foucauldian approach to knowledge integration processes, Management Learning 51 (2020) 336–352. doi:10.1177/1350507620904307.
[9] P. Thakur, R. Arora, Exploring the Relationship Between Satisfaction and Intention to Stay Among Millennial Employees: The Moderating Role of Managerial Support, International Journal of Human Capital and Information Technology Professionals (IJHCITP) 13 (2022) 1–17. doi:10.4018/ijhcitp.293229.
[10] F. Mane, D. Miravet, Using the job requirements approach and matched employer-employee data to investigate the content of individuals' human capital, Journal for Labour Market Research 49 (2016) 133–155. doi:10.1007/s12651-016-0203-3.
[11] D. M. Marchiori, R. G. Rodrigues, S. Popadiuk, E. W. Mainardes, The relationship between human capital, information technology capability, innovativeness and organizational performance: An integrated approach, Technological Forecasting and Social Change 177 (2022) 121526. doi:10.1016/j.techfore.2022.121526.
[12] H. Kucherova, Y. Honcharenko, D. Ocheretin, O. Bilska, Fuzzy logic model of usability of websites of higher education institutions in the context of digitalization of educational services, Neiro-Nechitki Tekhnolohii Modelyuvannya v Ekonomitsi 2021 (2021) 119–135. doi:10.33111/nfmte.2021.119.
[13] Z. Nedelko, S. Jevšenak, et al., Strategies and tools for knowledge management in innovation and the future industry, in: The role of knowledge transfer in open innovation, IGI Global, 2019, pp. 179–202. doi:10.4018/978-1-5225-5849-1.ch009.
[14] O. Trofymenko, O. Shevchuk, N. Koba, Y. Tashcheiev, T. Pavlenco, Knowledge and innovation management for transforming the field of renewable energy, in: International Conference on Artificial Intelligence and Sustainable Computing, Springer, 2021, pp. 73–87. doi:10.1007/978-3-030-82322-1_6.
[15] A. Braunstein, V. Deutscher, J. Seifried, E. Winther, A. Rausch, A taxonomy of social embedding – a systematic review of virtual learning simulations in vocational and professional learning, Studies in Educational Evaluation 72 (2022) 101098. doi:10.1016/j.stueduc.2021.101098.
[16] K. A. Riley, D. L. Nuttall (Eds.), Measuring Quality: Education Indicators: United Kingdom and International Perspectives , Routledge, 1994. URL: https://eric.ed.gov/?id=ED374559.
[17] G. Csomós, Introducing recalibrated academic performance indicators in the evaluation of individuals' research performance: A case study from Eastern Europe, Journal of Informetrics 14 (2020) 101073. doi:10.1016/j.joi.2020.101073.
[18] IMD - International Institute for Management Development, World Competitiveness Center Rankings, 2022. URL: https://www.imd.org/centers/world-competitiveness-center/rankings/.
[19] S. Dutta, B. Lanvin, S. Wunsch-Vincent (Eds.), Global Innovation Index 2020: Who Will Finance Innovation?, 13 ed., Ithaca, Fontainebleau, and Geneva, 2020. URL: https://www.

wipo.int/edocs/pubdocs/en/wipo_pub_gii_2020.pdf.
[20] A. Schleicher, PISA 2018: Insights and Interpretations, OECD Publishing, 2019. URL: https://www.oecd.org/pisa/PISA%202018%20Insights%20and%20Interpretations%20FINAL%20PDF.pdf.
[21] UNDP, Human Development Report 2019. Beyond income, beyond averages, beyond today: Inequalities in human development in the 21st century, 2019. doi:10.18356/838f78fd-en.
[22] WorldBank, The Human Capital Index 2020 Update: Human Capital in the Time of COVID-19, 2021. doi:10.1596/978-1-4648-1552-2.
[23] K. Schwab (Ed.), The Global Competitiveness Report 2018, World Economic Forum, Cologny/Geneva, 2018. URL: https://www3.weforum.org/docs/GCR2018/05FullReport/TheGlobalCompetitivenessReport2018.pdf.
[24] O. I. Ilyash, V. I. Stefaniak, Zastosuvannia taksonomichnoho analizu yak instrumentu derzhavnoho rehuliuvannia ta otsinky rivnia sotsialnoi bezpeky Ukrainy [Applying of taxonomic analyses as a tool of state regulation and estimation of the social safety's level of Ukraine], Ekonomichnyi prostir (2012) 80–91.
[25] R. Williams, A. Leahy, U21 Ranking of National Higher Education Systems 2020, 2020. doi:10.6017/ihe.2016.84.9107.
[26] O. Ilyash, O. Trofymenko, I. Dzhadan, T. Tsarova, Ecological and economic effects of industrial and technological development, IOP Conference Series: Earth and Environmental Science 915 (2021) 012004. doi:10.1088/1755-1315/915/1/012004.
[27] S. Kahkonen, Why Ukraine's Education System is Not Sustainable, 2018. URL: https://www.worldbank.org/en/news/opinion/2018/09/12/why-ukraines-education-system-is-not-sustainable.
[28] A. Bhattacharya (Ed.), Strategic human capital development and management in emerging economies, IGI Global, 2017. doi:10.4018/978-1-5225-1974-4.
[29] Research, Innovation and Enterprise Secretariat, Research Innovation and Enterprise 2020 Plan: Winning the Future through Science and Technology, 2016. URL: https://www.mti.gov.sg/-/media/MTI/Resources/Publications/Research-Innovation-and-Enterprise-RIE-2020/RIE2020.pdf.
[30] A. M. Santacreu, H. Zhu, Has China Overtaken the U.S. in Terms of Innovation?, 2018. URL: https://www.stlouisfed.org/on-the-economy/2018/march/china-overtaken-us-terms-innovation.
[31] World Bank, World Development Report 2019 : The Changing Nature of Work, World Bank, Washington, DC, 2019. doi:10.1596/978-1-4648-1328-3.
[32] M. M. Karpenko, Osvita protiahom zhyttia yak chynnyk liudskoho rozvytku. Analitychna zapyska [Lifelong learning as a factor of human development. Analytical note], 2015. URL: http://old2.niss.gov.ua/articles/1865/.
[33] O. Zhylinska, O. Melnychuk, L. Antoniuk, O. Humenna, A. Radchuk, Y. Stoliarchuk, S. Taruta, H. Kharlamova, N. Chala, O. Shnyrkov, Ukraina — 2030: Doktryna zbalansovanoho rozvytku [Ukraine – 2030: The doctrine of balanced development], 2 ed., Kalvariia, Lviv, 2017. URL: http://ukraine2030.org/img/_book/E-Book-Doctrine-2030.pdf.
[34] O. Bondarchuk, V. Balakhtar, O. Gorova, N. Lytvynenko, N. Pinchuk, O. Shmanko, A. Kiv, V. Oleksiuk, Features of responsibility of future specialists of the socionomic professions as an indicator of their digital competence, Educational Technology Quarterly 2022 (2022)

35–55. doi:10.55056/etq.12.
[35] O. V. Prokhorov, V. O. Lisovichenko, M. S. Mazorchuk, O. H. Kuzminska, Implementation of digital technology for student involvement based on a 3D quest game for career guidance and assessing students' digital competences, Educational Technology Quarterly 2022 (2022) 366–387. doi:10.55056/etq.430.
[36] V. Russo, Digital Economy and Society Index (DESI). European guidelines and empirical applications on the territory, in: Qualitative and Quantitative Models in Socio-Economic Systems and Social Work, Springer, 2020, pp. 427–442. doi:10.1007/978-3-030-18593-0_31.
[37] V. Kuybida, O. Petroye, L. Fedulova, G. Androshchuk, Digital competences as a condition to the development of quality of human capital, Zbirnyk naukovykh prats Natsionalnoi akademii derzhavnoho upravlinnia pry Prezydentovi Ukrainy (2019) 118–133. URL: http://nbuv.gov.ua/UJRN/znpnadu_2019_1_16.
[38] State Statistics Service of Ukraine, Statistics, 2021. URL: http://www.ukrstat.gov.ua/.
[39] L. T. B. Le, T. T. Tran, N. H. Tran, Challenges to STEM education in Vietnamese high school contexts, Heliyon (2021) e08649. doi:10.1016/j.heliyon.2021.e08649.
[40] L. Hrynevych, N. Morze, V. Vember, M. Boiko, Use of digital tools as a component of STEM education ecosystem, Educational Technology Quarterly 2021 (2021) 118–139. doi:10.55056/etq.24.
[41] T. R. Kelley, J. G. Knowles, A conceptual framework for integrated STEM education, International Journal of STEM education 3 (2016) 1–11. doi:10.1186/s40594-016-0046-z.

Fuzzy expert decision support system for foreign direct investment: a swarm metaheuristic approach*

Eugene E. Fedorov[1], Liubov O. Kibalnyk[2], Lesya O. Petkova[1], Maryna M. Leshchenko[1] and Vladyslav M. Pasenko[1]

[1]Cherkasy State Technological University, 460 Shevchenko Blvd., Cherkasy, 18006, Ukraine

[2]The Bohdan Khmelnytsky National University of Cherkasy, 81 Shevchenko Blvd., Cherkasy, 18031, Ukraine

Abstract

This paper develops and optimizes a fuzzy expert system for foreign direct investment decision support. The paper aims to provide a reliable and flexible tool for investors to evaluate the attractiveness of different countries for foreign direct investment. The paper uses an adaptive gravitational search algorithm to determine the optimal parameters of the fuzzy expert system, such as the membership functions for linguistic input and output variables. The paper also uses a quality criterion that considers the specificity of the fuzzy expert system and allows assessing the probability of future decisions. The paper conducts a numerical study to test the performance of the proposed fuzzy expert system and compares it with other existing methods. The results show that the proposed fuzzy expert system has a high accuracy and robustness in foreign direct investment decision support. The paper contributes to the literature on fuzzy logic applications in economics and finance and provides a practical tool for investors to make informed decisions on foreign direct investment.

Keywords

fuzzy expert system, foreign direct investment, adaptive gravitational search algorithm, quality criterion, decision support

1. Introduction

In contemporary contexts, decision-making systems for foreign direct investment have gained significant prominence. Machine learning techniques [2, 3, 4], such as regression [5] and auto-regressive methods [6], are commonly employed to construct these systems. However, these approaches often result in linear models, limiting their scope. Expert systems, utilizing a knowledge base typically represented as production rules, are another avenue for building decision-making systems for foreign direct investment [7]. Yet, these systems are criticized for their sole reliance on quantitative assessments, which can pose challenges when operators prefer qualitative estimates.

M3E2-MLPEED 2022: The 10th International Conference on Monitoring, Modeling & Management of Emergent Economy, November 17-18, 2022, Virtual, Online

*Extended and revised version of paper [1] presented at the 10th International Conference on Monitoring, Modeling & Management of Emergent Economy.

fedorovee75@ukr.net (E. E. Fedorov); liubovkibalnyk@gmail.com (L. O. Kibalnyk); l{-}petkova@ukr.net (L. O. Petkova); mari.leshchenko@gmail.com (M. M. Leshchenko); pasenko{-}vlad@ukr.net (V. M. Pasenko)

0000-0003-3841-7373 (E. E. Fedorov); 0000-0001-7659-5627 (L. O. Kibalnyk); 0000-0003-4519-3726 (L. O. Petkova); 0000-0002-0210-9582 (M. M. Leshchenko); 0000-0002-7411-2625 (V. M. Pasenko)

Fuzzy expert systems have emerged as a means to simplify human-computer interactions. These systems leverage fuzzy inference mechanisms such as Larsen, Mamdani, Tsukamoto, and Sugeno [8, 9]. Nonetheless, a key shortcoming is the lack of automation in determining their parameters [10, 11]. Addressing these limitations calls for the utilization of optimization methods to fine-tune the parameters of fuzzy expert systems.

However, contemporary optimization methods are not without their own set of challenges:

- Many possess high computational complexity.
- Several are prone to converging into local extrema.
- Some lack convergence guarantees.

In this regard, there is an actual problem of optimization methods' insufficient efficiency.

Consequently, the quest for more efficient optimization methods is pertinent. This has led to the adoption of metaheuristics, a class of modern heuristics aimed at hastening the discovery of quasi-optimal solutions to optimization problems and reducing the likelihood of converging into local optima [12, 13].

Yet, current metaheuristics have their own limitations:

- Certain methods exhibit inadequate accuracy [14].
- Others provide only abstract descriptions or are tailored to specific problems [15].
- The process of parameter determination remains non-automated [16].
- The influence of iteration count on solution search is often disregarded [17].
- Some methods lack the capability to address conditional optimization problems [18].
- Incompatibility with non-binary potential solutions exists [19].
- Convergence guarantees may be absent [20].

Hence, the challenge of constructing efficient metaheuristic optimization methods arises [21, 22]. One noteworthy example of such a metaheuristic is the gravitational search algorithm, a member of the swarm metaheuristics family [23].

This research is driven by the need to develop adept fuzzy expert systems using parametric identification for adaptation and refinement in the realm of foreign direct investment decisions.

Objective: This study aims to enhance the effectiveness of foreign direct investment decisions by constructing a fuzzy expert system trained through the utilization of metaheuristics.

To accomplish this overarching goal, the following tasks have been undertaken:

1. Design a fuzzy expert decision support system for foreign direct investment.
2. Select an appropriate quality criterion for the proposed fuzzy expert system.
3. Develop a metaheuristic approach based on an adaptive gravitational search algorithm for parameter determination of the proposed fuzzy expert system.
4. Conduct extensive numerical investigations.

2. The fuzzy expert decision support system for foreign direct investment

The foreign direct investment analysis is based on the data of the GDP per capita volume, inflation rates, goods and services exports volume, and labor force indicators. To make decisions on foreign direct investment, a fuzzy expert system is proposed. It involves the following steps:

1) linguistic variables formation;
2) fuzzy knowledge base formation;
3) Mamdani fuzzy inference mechanism formation:
    - fuzzification;
    - sub-conditions aggregation;
    - conclusions activation;
    - aggregation of conclusions;
    - defuzzification.
4) identification of parameters based on metaheuristics.

2.1. Linguistic variables formation

The following input variables were chosen:

- the volume of gross domestic product (GDP) per capita (per year, US dollars), x_1;
- the inflation indicator (according to the consumer price index, which reflects the annual percentage change in the cost for the average consumer of purchasing a goods and services basket, per year, %), x_2;
- the volume of goods and services export indicator (total volume, per year, USD), x_3;
- the labor force indicator (labor force is people aged 15 and over who provide labor for the production of goods and services, per year, number of people), x_4.

The following indicators were chosen as linguistic input variables. They are qualitative indicators:

- the GDP volume $\tilde{x}_1$ with values $\tilde{\alpha}_{11}$ = *little*, $\tilde{\alpha}_{12}$ = *medium*, $\tilde{\alpha}_{13}$ = *much*, where the ranges are fuzzy sets $\tilde{A}_{11} = \{(x_1, \mu_{\tilde{A}_{11}}(x_1))\}$, $\tilde{A}_{12} = \{(x_1, \mu_{\tilde{A}_{12}}(x_1))\}$, $\tilde{A}_{13} = \{(x_1, \mu_{\tilde{A}_{13}}(x_1))\}$;
- the inflation indicator $\tilde{x}_2$ with values $\tilde{\alpha}_{21}$ = *little*, $\tilde{\alpha}_{22}$ = *medium*, $\tilde{\alpha}_{23}$ = *much*, where the ranges are fuzzy sets $\tilde{A}_{21} = \{(x_2, \mu_{\tilde{A}_{21}}(x_2))\}$, $\tilde{A}_{22} = \{(x_2, \mu_{\tilde{A}_{22}}(x_2))\}$, $\tilde{A}_{23} = \{(x_2, \mu_{\tilde{A}_{23}}(x_2))\}$;
- the volume of goods and services export indicator $\tilde{x}_3$ with values $\tilde{\alpha}_{31}$ = *little*, $\tilde{\alpha}_{32}$ = *medium*, $\tilde{\alpha}_{33}$ = *much*, where the ranges are fuzzy sets $\tilde{A}_{31} = \{(x_3, \mu_{\tilde{A}_{31}}(x_3))\}$, $\tilde{A}_{32} = \{(x_3, \mu_{\tilde{A}_{32}}(x_3))\}$, $\tilde{A}_{33} = \{(x_3, \mu_{\tilde{A}_{33}}(x_3))\}$;
- the labor force indicator $\tilde{x}_4$ with values $\tilde{\alpha}_{41}$ = *little*, $\tilde{\alpha}_{42}$ = *medium*, $\tilde{\alpha}_{43}$ = *much*, where the ranges are fuzzy sets $\tilde{A}_{41} = \{(x_4, \mu_{\tilde{A}_{41}}(x_4))\}$, $\tilde{A}_{42} = \{(x_4, \mu_{\tilde{A}_{42}}(x_4))\}$, $\tilde{A}_{43} = \{(x_4, \mu_{\tilde{A}_{43}}(x_4))\}$.

The volume of foreign direct investment (net flows for the year, USD) was chosen as a clear output variable $\tilde{y}$. It is a qualitative indicator.

The volume of foreign direct investment was chosen $\tilde{y}$ with its values $\tilde{\beta}_1$ = *little*, $\tilde{\beta}_2$ = *medium*, $\tilde{\beta}_3$ = *much*, where the ranges are fuzzy sets $\tilde{B}_1 = \{(y, \mu_{\tilde{B}_1}(y))\}$, $\tilde{B}_2 = \{(y, \mu_{\tilde{B}_{42}}(y))\}$, $\tilde{B}_3 = \{(y, \mu_{\tilde{B}_3}(y))\}$;

2.2. Fuzzy knowledge base formation

Fuzzy knowledge is represented as the following fuzzy rules that contain a linguistic output variable R^n : IF $\tilde{x}_1$ is $\tilde{a}_{1i}$ AND $\tilde{x}_2$ is $\tilde{a}_{2j}$ AND $\tilde{x}_3$ is $\tilde{a}_{3k}$ AND $\tilde{x}_4$ is $\tilde{a}_{4p}$ then $\tilde{y}$ is $\tilde{B}_m$

In the case of linguistic variables specific values, fuzzy knowledge is presented in relational form in table 1.

Table 1
Relational form of fuzzy knowledge representation.

The rule	$\tilde{x}_1$	$\tilde{x}_2$	$\tilde{x}_3$	$\tilde{x}_4$	$\tilde{y}$
R^1	$\tilde{\alpha}_{11}$	$\tilde{\alpha}_{21}$	$\tilde{\alpha}_{31}$	$\tilde{\alpha}_{41}$	$\tilde{\alpha}_1$
R^2	$\tilde{\alpha}_{12}$	$\tilde{\alpha}_{21}$	$\tilde{\alpha}_{31}$	$\tilde{\alpha}_{41}$	$\tilde{\alpha}_1$
R^3	$\tilde{\alpha}_{13}$	$\tilde{\alpha}_{21}$	$\tilde{\alpha}_{31}$	$\tilde{\alpha}_{41}$	$\tilde{\alpha}_2$
R^4	$\tilde{\alpha}_{11}$	$\tilde{\alpha}_{22}$	$\tilde{\alpha}_{31}$	$\tilde{\alpha}_{41}$	$\tilde{\alpha}_2$
...	...	...	...	...	...
R^{81}	$\tilde{\alpha}_{13}$	$\tilde{\alpha}_{23}$	$\tilde{\alpha}_{33}$	$\tilde{\alpha}_{43}$	$\tilde{\alpha}_3$

2.3. Mamdani fuzzy inference mechanism formation

2.3.1. Fuzzification

We will determine the truth degree of each sub-condition of each rule, using the membership function $\mu_{\tilde{A}_{ij}}(x_i)$.

As membership functions of sub-conditions, we chose:

- piecewise linear Z-shaped function, i.e.

$$\mu_{\tilde{A}_{i1}}(x_i) = \begin{cases} 1, & x_i \le a_i \\ \frac{b_i - x_i}{b_i - a_i}, & a_i < x_i < b_i \\ 0, & x_i \ge b_i \end{cases} ,i \in \overline{1,4}$$

- piecewise linear Π-shaped function, i.e.

$$\mu_{\tilde{A}_{i2}}(x_i) = \begin{cases} 0, & x_i \le a_i \\ \frac{x_i - a_i}{b_i - a_i}, & a_i \le x_i \le b_i \\ 1, & b_i \le x_i \le c_i \\ \frac{d_i - x_i}{d_i - c_i}, & c_i \le x_i \le d_i \\ 0, & x_i \ge d_i \end{cases} ,i \in \overline{1,4}$$

- piecewise linear S-shaped function, i.e.

$$\mu_{\tilde{A}_{i3}}(x_i) = \begin{cases} 0, & x_i \leq c_i \\ \frac{x_i - c_i}{d_i - c_i}, & c_i < x_i < d_i \\ 1, & x_i \geq d_i \end{cases}, i \in \overline{1,4},$$

where a_i, b_i, c_i, d_i - membership function parameters.

2.3.2. Sub-condition aggregation

The condition membership functions for each rule R^n are determined based on the minimum value method:

$$\mu_{\bigcup_{i=1}^{4} \tilde{A}_{i,f(n,i)}}(x_1, x_2, x_3, x_4) = \min_{i \in \overline{1,4}} \left\{ \mu_{\tilde{A}_{i,f(n,i)}}(x_i) \right\},$$

where f – a function that returns the value number of the i-th linguistic input variable of the n-th rule and is determined on the basis of table 1. For example, if the linguistic input variable $\tilde{x}_1$ rules R^{81} matters $\tilde{\alpha}_{13}$, then $f(81, 1) = 3$.

2.3.3. Activation of conclusions

The membership functions of the conclusion for each rule R^n are determined based on the minimum value method (based on the Mamdani rule):

$$\mu_{\tilde{B}_{g(n)}}(y) = \min \{ \mu_{U_{i=1}^{4} \tilde{A}_{i,f(n,i)}}(x_1, x_2, x_3, x_4), \mu_{\tilde{B}_{g(n)}}(y) \},$$

where g – a function that returns the value number of the linguistic output variable of n-th rule and determined on the basis of table 1.

For example, if the linguistic output variable $\tilde{y}$ of the rule R^{81} is $\tilde{\beta}_3$, then $g(81) = 3$.

A piecewise linear triangular function was chosen as the membership functions of the conclusions, i.e.

$$\mu_{\tilde{B}_m}(y) = \begin{cases} 0, & y \leq e_m \\ \frac{y - e_m}{u_m - e_m}, & e_m \leq y \leq u_m \\ \frac{v_m - y}{v_m - u_m}, & u_m \leq y \leq v_m \\ 0, & y \geq v_m \end{cases}, m \in \overline{1,3},$$

where e_m, u_m, v_m – membership function parameters.

In the case of such a membership function, the kernel of each fuzzy set $\tilde{B}_m$ is:

$$\ker \tilde{B}_m = \{ y \in Y | \mu_{\tilde{B}_m}(y) = 1 \} = \{ u_m \}.$$

2.3.4. Aggregation of conclusions

The membership functions of the final conclusion are defined, which contains a linguistic output variable based on the maximum value method:

$$\mu_{\tilde{B}_m}(Y) = \max_{n \in 1,81} \{ \mu_{\tilde{B}_g(n)}(y) \}$$

2.3.5. Defuzzification

The volumes of foreign direct investment are determined basedon the centroid method:

$$y^* = \frac{\sum_{y\in Y} \mu_{\tilde{B}}(y)y}{\sum_{y\in Y} \mu_{\tilde{B}}(y)}, Y = \{1, 2, 3\}$$

3. Quality criterion for the proposed fuzzy expert system

The objective function is chosen as a quality criterion, representing the accuracy as probability of correct foreign direct investment

$$\begin{gathered} F = \frac{1}{P}\sum_{p=1}^{P}, [y_p = d_p] \to \max_{\theta}, \\ [p = q] = \begin{cases} 1, & p = q \\ 0, & p \neq q \end{cases}, \end{gathered} \tag{1}$$

where d_p – test foreign direct investment,

y_p – foreign direct investment received as a result of fuzzy inference,

P – number of test implementations,

$\theta = (a_1, b_1, c_1, d_1, ..., a_4, b_4, c_4, d_4, e_1, u_1, v_1, ..., e_3, u_3, v_3)$ – parameter vector of membership functions.

4. Metaheuristic method based on an adaptive gravitational search algorithm for determining the parameters of the proposed fuzzy expert system

The particle velocity (not the gravitational constant) depends on the iteration number in this method, which provides control over the convergence rate of the method, as well as providing a global search at the initial iterations, and a local search at the final iterations. The parameter vector of membership functions corresponds to the position vector of one particle x. The quality criterion is used as the goal function (1).

1. Initialization.
   1.1. Setting the gravitational constant G, the maximum number of iterations N, the population size K, the length of the particle position vector M (it corresponds to the length of the parameter vector of membership functions and is equal to 25), the minimum and maximum values for the position vector $x_j^{\min}$, $x_j^{\max}$, $j \in \overline{1, M}$, the minimum and maximum values for the velocity vector $v_j^{\min}$, $v_j^{\max}$, $j \in \overline{1, M}$.
   1.2. The best position vector randomly generating $x^* = (x_1^*, ..., x_M^*)$,
   $x_j^* = x_j^{\min} + (x_j^{\max} - x_j^{\min})U(0, 1)$,
   where $U(0, 1)$ – a function that returns a uniformly distributed random number in a range $[0, 1]$.

1.3. The initial population creation

1.3.1. Particle number $k = 1$, $P = \varnothing$.

1.3.2. A position vector at random x_k generating $x_k = (x_{k1}, ..., x_{kM})$, $x_{kj} = x_j^{\min} + (x_j^{\max} - x_j^{\min})U(0,1)$.

1.3.3. Random velocity vector v_k generating $v_k = (v_{k1}, ..., v_{kM})$, $v_{ij} = v_j^{\min} + (v_j^{\max} - v_j^{\min})U(0,1)$.

1.3.4. If $(x_k, v_k) \notin P$, then $P = P \cup \{(x_k, v_k)\}$, $k = k + 1$.

1.3.5. If $k \leq K$, then go to step 1.3.2.

2. Iteration number $n = 1$.
3. The computation of the best and worst particle of a population from a target function

$$l = arg\min_k F(x_k), x^{best} = x_l,$$

$$l = arg\max_k F(x_k), x^{worst} = x_l.$$

4. The computation of all particles masses.
5. The computation of the gravitational force acting between all pairs of particles

5.1. $m_k = G\frac{F(x_k) - F(x^{worst})}{F(x^{best}) - F(x^{worst})}, k \in \overline{1, K}$.

5.2. $M_k = \frac{m_k}{\sum_{s=1}^{K} m_s}, k \in \overline{1, K}$.

6. The computation of the gravitational force acting between all pairs of particles

$$f_{kl} = G\frac{M_k M_l}{d(x_k, x_l) + \varepsilon}(x_l - x_k), k, l \in \overline{1, K},$$

where $d(x_k, x_l)$ – distance between particles k and l (e.g. Euclid distance).

7. The computation of the resulting force acting on all particles

$$r_{kl} = U(0,1), k, l \in \overline{1, K}$$

$$f_k = \sum_{\substack{l=1 \\ l \neq k}}^{K} r_{kl} f_{kl}, k \in \overline{1, K}$$

8. Modification of the acceleration of all particles

$$a_k = \frac{f_k}{M_k}, k \in \overline{1, K}$$

9. Speed modification of all particles

$$r_k = U(0,1), k \in \overline{1, K}$$

$$v_k = r_k v_k + a_k, k \in \overline{1, K}$$

10. Modification all of the particles' position, taking into account the iteration number

10.1. $x_k = x_k + v_k\left(1 - \frac{n}{N}\right), k \in \overline{1, K}$

10.2. $x_{kj} = \max\{x_j^{\min}, x_{kj}\}, x_{kj} = \min\{x_j^{\max}, x_{kj}\}, j \in \overline{1, M}, k \in \overline{1, K}$

11. If $n < N$, then $n = n + 1$, go to step 3

The result is x^*.

5. Numerical research

Numerical research was carried out using the Keras submodule of the TensorFlow module. The Pandas module was used to fill in missing values through linear interpolation, as well as for tabular data I/O operations. The Scikit-fuzzy module was used to create a fuzzy expert system.

The fuzzy expert system was researched using the World Bank economic indicators database (https://databank.worldbank.org/home.aspx). The economic indicators of 145 countries for 10 years were used. The size of the original sample was 1450.

For the proposed adaptive gravity search algorithm, the gravity constant G was 100, the maximum number of iterations was 1000, and the population size was 50.

The comparison results of the proposed fuzzy expert system with the operator are presented in table 2.

Table 2
Comparison results of the proposed fuzzy expert system with an operator.

Accuracy	
fuzzy expert system	operator
0.98	0.8

The comparison results of the proposed fuzzy expert system with the proposed meta-heuristic adaptive gravitational search algorithm (AGSA) and the traditional meta-heuristic adaptive gravitational search algorithm (AGSA) operator are presented in table 3.

Table 3
Comparison results of the proposed fuzzy expert system of the proposed meta-heuristic and the traditional meta-heuristic.

Accuracy	
GSA	AGSA
0.93	0.98

Table 4
Comparison results of the proposed fuzzy expert system based on the back-propagation method and proposed meta-heuristic.

Accuracy	
BP	AGSA
0.90	0.98

Figure 1 shows the accuracy for the proposed fuzzy expert system trained based on the proposed meta-heuristic adaptive gravitational search algorithm (AGSA) and on the proposed meta-heuristic gravitational search algorithm (GSA).

The comparison results of the proposed fuzzy expert system trained on the basis of back-propagation (BP) and the proposed meta-heuristic adaptive gravitational search algorithm (AGSA) are presented in table 4.

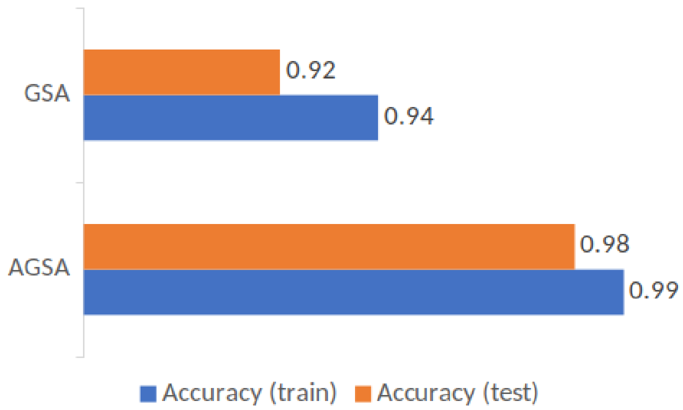


Figure 1: Accuracy of the proposed fuzzy expert system with GSA and AGSA.

Figure 2 shows the accuracy for the proposed fuzzy expert system trained on the basis of back-propagation (BP) and the proposed meta-heuristic adaptive gravitational search algorithm (AGSA).

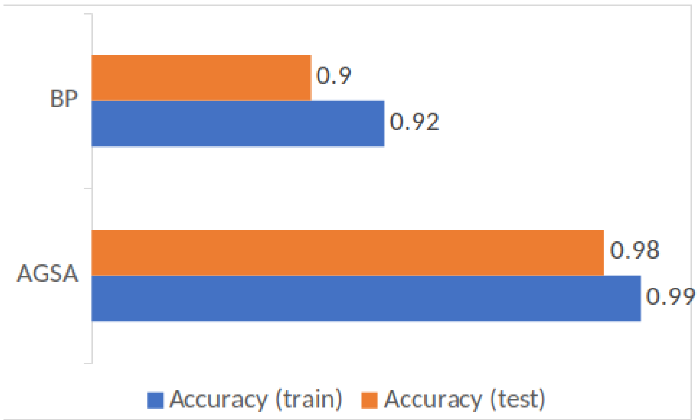


Figure 2: Accuracy of the proposed fuzzy expert system with BP and AGSA.

Figures 3-7 shows the membership functions for the values of linguistic variables $\tilde{x}_1$, $\tilde{x}_2$, $\tilde{x}_3$, $\tilde{x}_4$ and y.

6. Discussion

The traditional non-automatic approach to assessing the foreign direct investment effectiveness reduces the accuracy of a correct assessment (table 2). The proposed method eliminates this disadvantage.

The traditional method of the gravitational search algorithm ignores the iteration number during the particle position calculating; this reduces the accuracy of finding a solution (table 3);

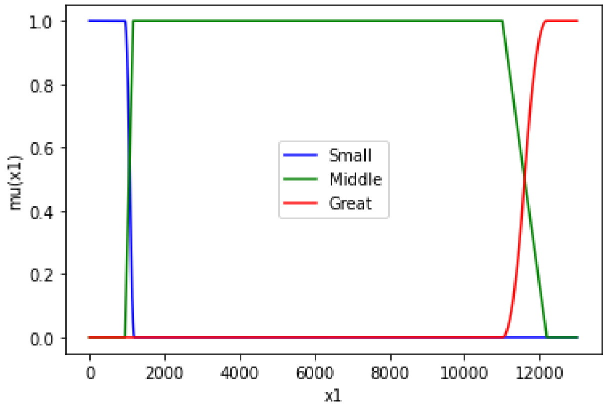


Figure 3: Membership functions for linguistic variable values $\tilde{x}_1$.

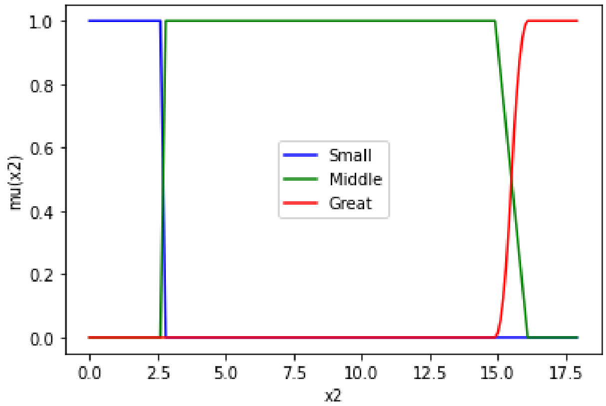


Figure 4: Membership functions for linguistic variable values $\tilde{x}_2$.

requires a large number of parameters associated with the gravitational constant calculating. The proposed method eliminates these shortcomings.

The traditional approach to training a fuzzy expert system based on back propagation reduces the probability of correct estimation (table 4). The proposed method eliminates this disadvantage.

7. Conclusions

This section presents the key outcomes and contributions of the research, highlighting the insights gained and the novel methodologies developed:

1. **Exploration of Relevant Methods**: The study delved into the landscape of optimization methods and expert systems within the realm of foreign direct investment decision-making. The findings underscored the potency of employing fuzzy expert systems, parameterized through contemporary metaheuristic techniques.
2. **Development of Fuzzy Expert System**: A novel fuzzy expert decision support system for foreign direct investment has been conceived. This innovative system streamlines

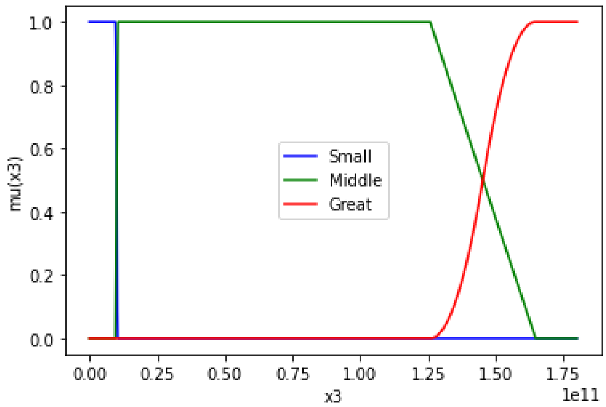


Figure 5: Membership functions for linguistic variable values $\tilde{x}_3$.

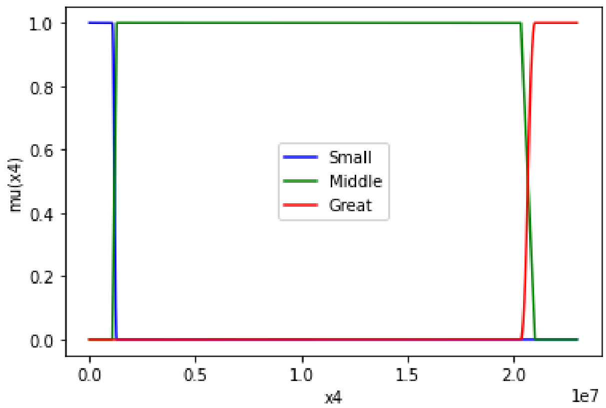


Figure 6: Membership functions for linguistic variable values $\tilde{x}_4$.

operator-computer interactions by integrating qualitative indicators, facilitating parameter identification through the proposed swarm metaheuristics.

3. **Introduction of Quality Criterion**: A bespoke quality criterion has been introduced, tailored to the nuances of the newly devised fuzzy expert system. This criterion enables a comprehensive assessment of decision accuracy.
4. **Creation of Adaptive Swarm Metaheuristic Algorithm**: An adaptive swarm metaheuristic algorithm, founded on the principles of the gravitational search algorithm, has been crafted. This algorithm boasts the capability to regulate convergence rate, execute global exploration in initial iterations, and transition to local exploration in later iterations via adaptive particle velocity control.
5. **Empowering Decision Technology**: The amalgamation of the swarm metaheuristic optimization method with the fuzzy expert system furnishes a means to infuse sophistication into foreign direct investment decision-making technology. This intellectualized approach holds significant promise in enhancing decision accuracy and effectiveness.
6. **Future Prospects**: The envisioned trajectory involves subjecting the proposed method

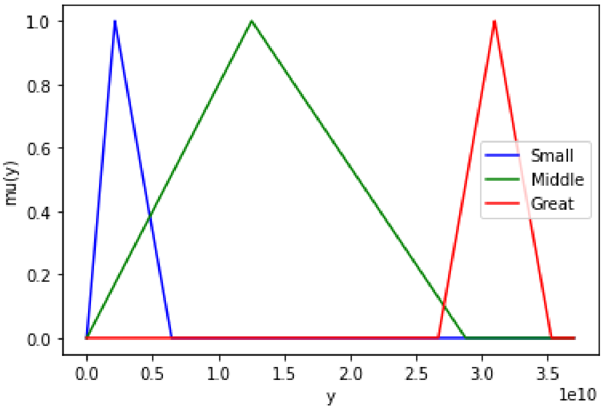


Figure 7: Membership functions for linguistic variable values $\tilde{y}$.

and system to more extensive testing using a broader array of test databases. This step is pivotal in validating the method's robustness and efficacy in diverse scenarios.

References

[1] E. Fedorov, L. Kibalnyk, L. Petkova, M. Leshchenko, V. Pasenko, Fuzzy Expert System of the Decision Making Support on Foreign Direct Investment, in: S. Semerikov, V. Soloviev, A. Matviychuk, V. Kobets, L. Kibalnyk, H. Danylchuk, A. Kiv (Eds.), Proceedings of 10th International Conference on Monitoring, Modeling & Management of Emergent Economy - M3E2, INSTICC, SciTePress, 2023, pp. 15–22. doi:10.5220/0011930700003432.

[2] A. Kiv, S. Semerikov, V. N. Soloviev, L. Kibalnyk, H. Danylchuk, A. Matviychuk, Experimental Economics and Machine Learning for Prediction of Emergent Economy Dynamics, in: A. Kiv, S. Semerikov, V. N. Soloviev, L. Kibalnyk, H. Danylchuk, A. Matviychuk (Eds.), Proceedings of the Selected Papers of the 8th International Conference on Monitoring, Modeling & Management of Emergent Economy, M3E2-EEMLPEED 2019, Odessa, Ukraine, May 22-24, 2019, volume 2422 of *CEUR Workshop Proceedings*, CEUR-WS.org, 2019, pp. 1–4. URL: https://ceur-ws.org/Vol-2422/paper00.pdf.

[3] V. Derbentsev, A. Matviychuk, V. N. Soloviev, 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, Springer, Singapore, 2020, pp. 211–231. doi:10.1007/978-981-15-4498-9_12.

[4] P. V. Zahorodko, S. O. Semerikov, V. N. Soloviev, A. M. Striuk, M. I. Striuk, H. M. Shalatska, Comparisons of performance between quantum-enhanced and classical machine learning algorithms on the IBM Quantum Experience, Journal of Physics: Conference Series 1840 (2021) 012021. doi:10.1088/1742-6596/1840/1/012021.

[5] D. Milovanović, N. Marković, Strategic decision making and influence of economic freedoms on foreign direct investment (FDI) in Bosnia and Herzegovina, Strategic Management 27 (2022) 44–56. URL: https://www.smjournal.rs/index.php/home/article/view/239.

[6] P. Kurecic, F. Kokotovic, The Relevance of Political Stability on FDI: A VAR Analysis

and ARDL Models for Selected Small, Developed, and Instability Threatened Economies, Economies 5 (2017) 22. doi:10.3390/economies5030022.

[7] M. Šamanović, M. Ćukusić, M. Jadrić, Rule based approach to determining the influence of various business regulations on the amount of foreign direct investment in a country, in: Proceedings of the ITI 2010, 32nd International Conference on Information Technology Interfaces, Cavtat, 2010.

[8] D. Ruan (Ed.), Intelligent Hybrid Systems: Fuzzy Logic, Neural Networks, and Genetic Algorithm, Kluwer Academic Publishers, 1997. doi:10.1007/978-1-4615-6191-0.

[9] L. H. Tsoukalas, R. E. Uhrig, Fuzzy and Neural Approaches in Engineering, John Wiley & Sons, Inc, New York, 1997.

[10] S. Abe, Neural Networks and Fuzzy Systems: Theory and Application, Kluwer Academic Publishers, Boston, 1997. doi:10.1007/978-1-4615-6253-5.

[11] A. Rotshtein, S. Shtovba, I. Mostav, Fuzzy rule based innovation projects estimation, in: Proceedings Joint 9th IFSA World Congress and 20th NAFIPS International Conference (Cat. No. 01TH8569), volume 1, 2001, pp. 122–126. doi:10.1109/NAFIPS.2001.944238.

[12] E.-G. Talbi, Metaheuristics: From Design to Implementation, Wiley & Sons, Hoboken, New Jersey, 2009.

[13] A. P. Engelbrecht, Computational Intelligence: An Introduction, 2 ed., Wiley & Sons, Chichester, West Sussex, 2007.

[14] E. Alba, A. Nakib, P. Siarry (Eds.), Metaheuristics for Dynamic Optimization, volume 433 of *Studies in Computational Intelligence*, Springer-Verlag, Berlin, 2013.

[15] K. F. Doerner, M. Gendreau, P. Greistorfer, W. Gutjahr, R. F. Hartl, M. Reimann (Eds.), Metaheuristics: Progress in Complex Systems Optimization, volume 39 of *Operations Research/Computer Science Interfaces Series*, Springer, New York, 2007. doi:10.1007/978-0-387-71921-4.

[16] O. O. Grygor, E. E. Fedorov, T. Y. Utkina, A. G. Lukashenko, K. S. Rudakov, D. A. Harder, V. M. Lukashenko, Optimization method based on the synthesis of clonal selection and annealing simulation algorithms, Radio Electronics, Computer Science, Control (2019) 90–99. doi:10.15588/1607-3274-2019-2-10.

[17] O. Bozorg-Haddad, Meta-heuristic and Evolutionary Algorithms for Engineering Optimization, Wiley & Sons, Hoboken, New Jersey, 2017.

[18] E. Fedorov, V. Lukashenko, T. Utkina, A. Lukashenko, K. Rudakov, Method for Parametric Identification of Gaussian Mixture Model Based on Clonal Selection Algorithm, in: D. Luengo, S. Subbotin, P. Arras, Y. V. Bodyanskiy, K. Henke, I. Izonin, V. G. Levashenko, V. Lytvynenko, A. Parkhomenko, A. Pester, N. Shakhovska, A. Sharpanskykh, G. Tabunshchyk, C. Wolff, H. Wuttke, E. Zaitseva (Eds.), Proceedings of the Second International Workshop on Computer Modeling and Intelligent Systems (CMIS-2019), Zaporizhzhia, Ukraine, April 15-19, 2019, volume 2353 of *CEUR Workshop Proceedings*, CEUR-WS.org, 2019, pp. 41–55. URL: http://ceur-ws.org/Vol-2353/paper4.pdf.

[19] J. Radosavljević, Metaheuristic Optimization in Power Engineering, The Institution of Engineering and Technology, New York, 2018. doi:10.1049/PBPO131E.

[20] B. Chopard, M. Tomassini, An Introduction to Metaheuristics for Optimization, Natural Computing Series, Springer, Cham, 2018. doi:10.1007/978-3-319-93073-2.

[21] K.-L. Du, M. N. S. Swamy, Search and Optimization by Metaheuristics: Techniques and

Algorithms Inspired by Nature, Springer, Cham, 2016. doi:10.1007/978-3-319-41192-7.
[22] J. Brownlee, Clever Algorithms: Nature-Inspired Programming Recipes, Melbourne, 2011. URL: https://github.com/clever-algorithms/CleverAlgorithms.
[23] E. Rashedi, H. Nezamabadi-pour, S. Saryazdi, Gsa: A gravitational search algorithm, Information Sciences 179 (2009) 2232–2248. doi:10.1016/j.ins.2009.03.004.

The impact of intangible assets on the financial performance of Slovak ICT companies: a panel data regression analysis*

Serhii F. Lehenchuk[1], Tetiana A. Vakaliuk[1,2,3,4], Tetiana P. Nazarenko[1], Zuzana Kubaščíková[5] and Zuzana Juhászová[5]

[1]*Zhytomyr Polytechnic State University, 103 Chudnivsyka Str., Zhytomyr, 10005, Ukraine*

[2]*Institute for Digitalisation of Education of the National Academy of Educational Sciences of Ukraine, 9 M. Berlynskoho Str., Kyiv, 04060, Ukraine*

[3]*Kryvyi Rih State Pedagogical University, 54 Gagarin Ave., Kryvyi Rih, 50086, Ukraine*

[4]*Academy of Cognitive and Natural Sciences, 54 Gagarin Ave., Kryvyi Rih, 50086, Ukraine*

[5]*University of Economics in Bratislava, Dolnozemská cesta 1, 852 35 Petržalka, Slovakia*

Abstract

This paper investigates the impact of intangible assets on the financial performance of Slovak ICT companies in the context of the knowledge economy. The paper uses a panel data regression analysis to test the hypothesis that intangible assets have a significant positive effect on four indicators of financial performance: Return on Assets, Net Profit Margin, Assets Turnover, and Return on Equity. The paper analyzes a sample of 180 Slovak ICT companies for the period 2015–2019, using eight independent variables: Research and Development Intensity, Research and Development Intensity Squared, Software, Intellectual Property Rights, Acquired Intangible Assets, Leverage, Size, and Dummy variable for ICT sub-sectors. The paper applies various tests to select the appropriate estimation method and to check the adequacy of each model. The results partially confirm the hypothesis, as only Research and Development Intensity, Research and Development Intensity Squared, and Acquired Intangible Assets have a significant positive impact on some indicators of financial performance. The paper also finds that the influence of intangible assets varies depending on the type and measure of financial performance. The paper contributes to the literature on intangible assets and financial performance by providing empirical evidence from Slovak ICT companies. The paper also provides some implications for managers and policymakers to improve their intangible investment policy.

Keywords

intangible assets, financial performance, panel data regression, Slovak ICT companies, intangible investment policy

M3E2-MLPEED 2022: The 10th International Conference on Monitoring, Modeling & Management of Emergent Economy, November 17-18, 2022, Virtual, Online

*Extended and revised version of paper [1] presented at the 10th International Conference on Monitoring, Modeling & Management of Emergent Economy.

legenchyk2014@gmail.com (S. F. Lehenchuk); tetianavakaliuk@gmail.com (T. A. Vakaliuk); tatyana.nazarenko12@gmail.com (T. P. Nazarenko); kubascikova.zuzana@gmail.com (Z. Kubaščíková); zuzana.juhaszova@euba.sk (Z. Juhászová)

0000-0002-3975-1210 (S. F. Lehenchuk); 0000-0001-6825-4697 (T. A. Vakaliuk); 0000-0001-7702-8122 (T. P. Nazarenko); 0000-0001-6739-1278 (Z. Kubaščíková); 0000-0001-8592-0137 (Z. Juhászová)

1. Introduction

Over the past few decades, extensive discourse among researchers has centered on the evolving role of various types of capital in shaping the economic value and sustainability of enterprises, paving the way for their enduring success. Particularly noteworthy is the recognition of intellectual capital's pivotal role in this metamorphosis, influenced by shifts in production focus and management strategies. This shift is marked by a transition from the conventional emphasis on physical capital and labor to a spotlight on intellectual capital and the exchange of ideas, buoyed by intellectual property rights, particularly patents for their integration with technology [2, 3, 4, 5]. In this transformative landscape, the valuation and profitability of enterprises are increasingly tethered to their adeptness in harnessing their innovative potential and capitalizing on their intangible assets. Against the backdrop of overcoming the COVID-19 pandemic's repercussions [6, 7] and the global implementation of proactive sanctions strategies, which have led to reduced trade in conventional goods and services, the role of unique intellectual technologies in anchoring stable enterprise value gains prominence.

This new paradigm necessitates a growing reliance on national intellectual capital to fortify the economies of developed nations. Moreover, for numerous enterprises, financial performance hinges on the judicious crafting of policies that foster both the creation of novel intangible assets and the efficient utilization of existing ones. This involves their seamless integration into enterprise operations, fostering effective partnerships, governance, and control. In this context, enterprises frequently encounter network effects and an elevated exposure to market and technological risks. Consequently, a recalibration of business strategies becomes imperative, encompassing initiatives that leverage intangible assets to their strategic advantage. This is especially pertinent for high-tech entities, characterized by their substantial reliance on intangible assets to develop innovative technological products and services.

In light of these economic transformations, the quest for fresh theories and strategies is imperative to underpin informed decisions and management conduct in enterprises rich in intangible assets. A pertinent avenue of investigation is the impact of intangible assets on companies' financial performance, specifically focusing on the context of Slovak enterprises operating in the information and communications technology (ICT) sector. These companies, encompassing those engaged in information processing, storage, transfer, production of computing and telecommunication devices, and related services, epitomize high-tech enterprises. Their value creation processes are intrinsically linked to the efficient utilization of intangible assets. Investments in high-tech intangible assets within the ICT sector hold the promise of enhancing their financial metrics. However, as Huňady et al. [8] have indicated, Slovakia still exhibits a limited presence of business Research and Development (R&D) in the ICT sector. This suggests a cautious approach to intangible investments among Slovak ICT companies, possibly due to perceived risks and uncertainties surrounding their potential returns. Consequently, to mitigate such risks and uncertainties and to formulate an effective intangible investment policy, it becomes paramount to ascertain the intricate relationships between diverse intangible asset categories and various financial performance metrics.

Over the last decade, the Slovak Republic has witnessed vigorous development in its ICT sector. The number of individuals employed in the information and communication technology services sector surged from 28,905 in 2009 to 53,676 in 2019 [8], underscoring a near doubling of the workforce over the span of a decade. Reflecting this expansion, the ICT sector's contribution to the country's Gross Domestic Product (GDP) reached 4.2%, exerting significant influence on related industries as well. Endowed with numerous advantages, including high adaptability to enterprise activities, elevated value addition, well-established educational infrastructure, robust institutional networks, diversification in the telecommunications segment, strategic geographical positioning, extensive data and network coverage, and attractive investment incentives, the ICT sector garners substantial investor interest [9]. The Slovak government's active support for the sector, as evidenced by the provision of incentives such as tax reliefs, cash grants, job creation contributions, discounted real estate transactions, and a favorable R&D tax regime, further amplifies its allure [9]. Hence, examining the impact of intangible assets on the financial performance of Slovak ICT firms, within this conducive environment, assumes particular significance. Such an investigation can illuminate avenues for developing and fine-tuning intangible investment policies to enhance the financial performance of these companies.

Given the pivotal role of intangible assets in shaping the efficiency of high-tech enterprises, a research hypothesis has been formulated. This study hypothesizes a significant positive correlation between intangible assets and the financial performance of ICT companies. Recognizing that the strength of this influence may also vary based on company size, levels of borrowed capital, and sub-sector categorization within the ICT industry, the analysis of the impact of intangible assets on financial performance takes into account these factors. The resulting insights are expected to inform recommendations for Slovak ICT companies' intangible asset investment decisions.

2. Theoretical background

Problems of influence of intangible assets in their broad (economic) understanding on financial performance of high-tech companies are paid considerable attention of academicians. First of all, this is conditioned by the decisive role of intellectual capital for such enterprises in the context of the development of knowledge economy, which is based on ideas, R&D, innovations and technological progress. Scientists analyze of the impact of different intangible values on financial performance: intangible assets (the concept of IAS 38 [10]), intellectual capital (as a combination of human, organizational, and client capital), or separate components of two data. These studies cover different types of enterprises from different countries of the world, which represent different sectors of the economy. Since intellectual capital includes, to the most extent, all intangible assets that are the result of human intellectual activity, this article also analyzes the impact of intellectual capital and its components on the financial performance of ICT companies. In addition, a number of researchers are conducting studies of the impact of intangible assets both on individual components of financial performance, in particular, on profitability, and on broader categories, in particular, on total performance of the company or companies value.

Table 1 lists the number of articles and their quotations, which reveal the relationship between "Intangible assets" / "Intellectual capital" and "Financial performance" in science-based databases of Scopus, Web of Science and Google Scholar.

Table 1
Number of scientific articles by direction of researches and their quotations in academic literature for the period 2018–2022 as of July 01, 2022 (via Scopus, Web of Science and Google Scholar databases).

	Results found			Sum of the times cited		
Searching phrases	Scopus	Web of Science	Google Scholar	Scopus	Web of Science	Google Scholar
"Intangible assets" and "Financial performance"	161	576	21	894	16100	38
"Intellectual capital" and "Financial performance"	329	970	494	2530	13904	2235

The results of analysis of scientific databases are obtained (table 1) testify to the existence of a considerable number of publications in this direction of researches, as well as their influence on scientific works of other authors, which is confirmed by a considerable number of references to data of other authors and their constant growth from year to year. The cluster analysis of the key words of the articles from the databases of the Scopus and Web of Science on the basis of the use of VOSviewer allowed to confirm this conclusion. There was also a large number of publications that examined the impact of structural elements of intangible assets or intellectual capital (research and development, intangible resources, customer capital, structural capital, human capital, social capital, relational capital) on financial performance (figure 1). In addition, publications have been identified that investigate the impact of intangible assets or intellectual capital on other types of indicators that characterize the performance of the enterprise – firm performance, business performance, corporate performance, firm value, effectiveness, efficiency, profitability, ROA, competitive advantage etc. (figure 2).

Little attention is paid directly to the issue of impact of intangible values on financial performance of ICT companies, although the presence of significant positive relationships between with two variables is confirmed in the vast majority of results. Gan and Saleh [11] studied the connection between intellectual capital components of corporate performance among high-tech companies listed on Bursa Malaysia, in particular, profitability, and productivity. Based on the use of regression analysis, it was found that companies with larger intellectual capital as a rule have better profitability (ROA) and more efficient productivity (ATO).

Li and Wang [12] investigated the impact of different intangible assets (R&D expenditure, employee benefit, sales training) on profitability indicators (ROA) of Hong Kong Listed IT companies using regression analysis. They found a positive relationship between intangible assets and ROA.

Dženopoljac et al. [13] examined the role of intellectual capital and its key components in provision for financial performance (ROA, ROE, ROIC, ATO) of Serbian ICT sector companies during 2009–2013. They used Value-added intellectual coefficient (VAIC) as a measure of the IC contribution to value creation. The results obtained by the authors revealed that only one component of VAIC – CEE (capital-employed efficiency) had a significant impact on financial performance indicators, except for the indicator ROIC. Khan [14] also used VAIC as firms intangibility measure when analyzed the impact of intellectual capital on the financial performance of the 51 Indian IT companies for the period 2006–2016. He found a significant positive association

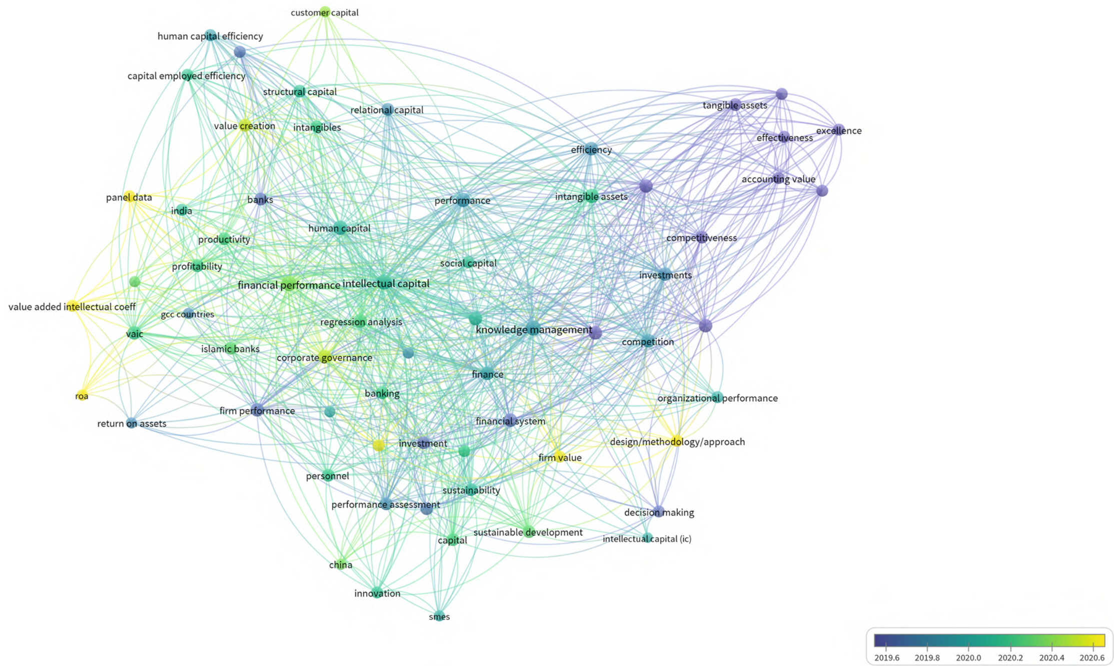


Figure 1: Bibliometric map of publications' keywords on the query "Intangible assets", "Intellectual capital" and "Financial performance" according to Scopus database in 2018–2022.

of VAIC with profitability, and an insignificant relationship with productivity, and significant positive association of CEE with profitability and productivity of Indian IT companies.

Zhang [15] analysed the relationship between degree of intangible assets and profitability for 17 Chinese listed telecommunication firms' for the period from 2014 to 2016. He found a positive and significant effect of Intangible assets ratio on ROA. Also, he emphasized the possibility of the inaccuracy of the obtained results due to the conservative nature of Chinese accounting standards rules in measuring intangible assets.

Huňady et al. [8] examined the role of innovations in performance of ICT sector companies from 24 countries during the years 2008–2016. Using regression analysis for macro-level data, they found positive effect of R&D expenditure on apparent labour productivity and value added in ICT sector.

Qureshi and Siddiqui [16] analyzed an effect of intangible assets on financial performance (ROA, ROE, ROIC, ATO and NPM) of the 80 global technology firms for the period from 2015 to 2018. They confirmed a significant negative effect of intangible assets on ROE, ROIC, ATO, and insignificant positive impact on companies' profitability. Moreover, the force of this influence considerably varies depending on the country's innovative development.

Lopes and Ferreira [17] also investigated the impact of intangibles on the performance indicators of major world technological firms (Turnover, ROA, ROE, ROS, EPS), have received evidence of existence of negative correlation between all intangible variables, control variables (Size, Leverage) with ROA. These conclusions are also confirmed by Sundaresan et al. [18], who investigated the impact of intangible assets on financial performance of 38 Taiwanese listed

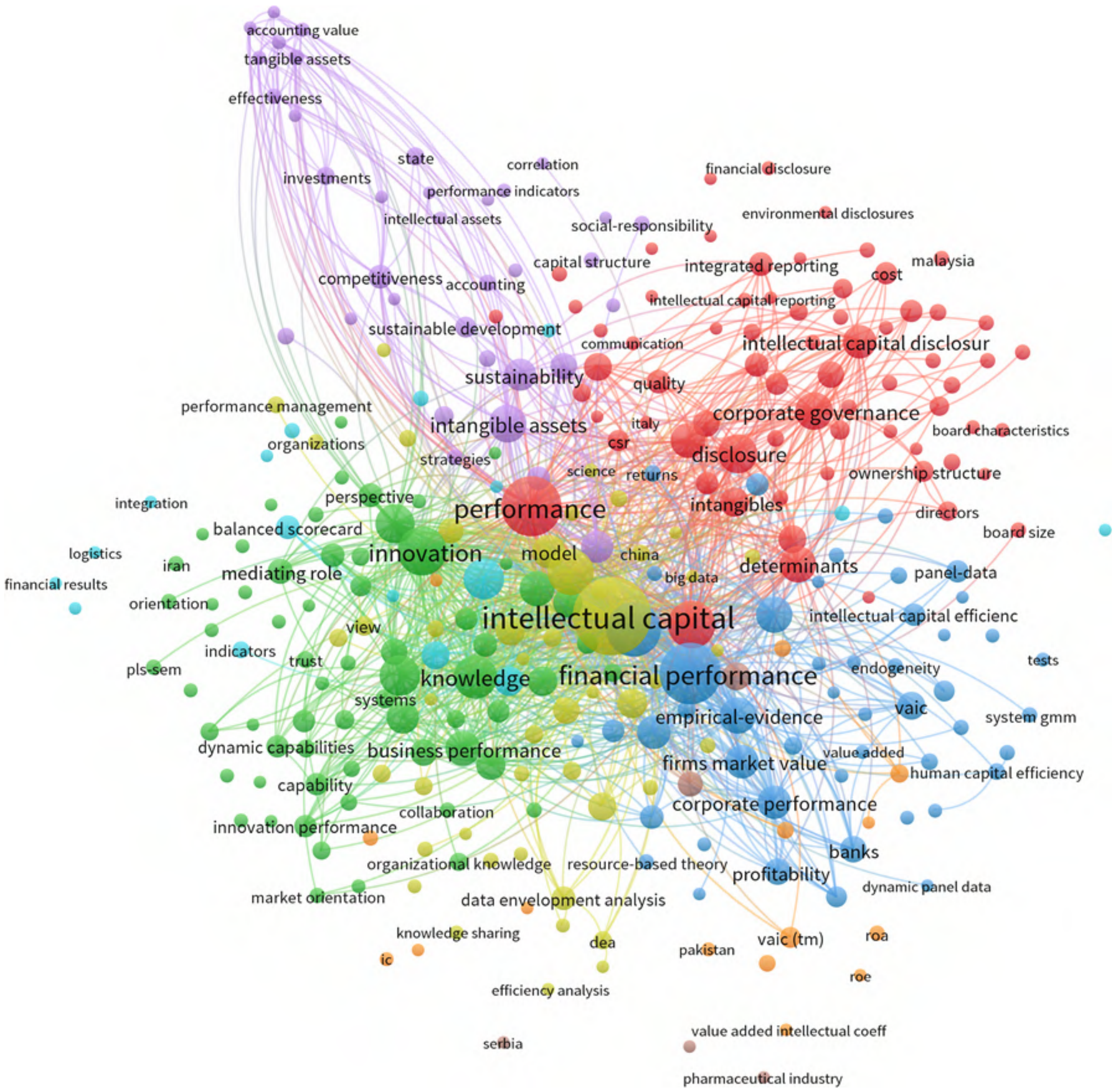


Figure 2: Bibliometric map of publications' keywords on the query "Intellectual capital" and "Financial performance" according to Web of Science database in 2018–2022.

technology firms for the period 2015–2019. The authors also revealed the existence of a lack of a significant relationship between intangible assets and ROA, but found significant influence of size on ROA. At the same time, they confirmed significant impact of intangibles on ROE. The results of the ROA received by Lopes and Ferreira [17], Sundaresan et al. [18] are in direct contradiction with most of the conclusions obtained by the authors who studied impact of intangibles on performance of ICT companies.

Radonić et al. [19] studied the role of intellectual capital components (human, relational, structural and innovation capital) in ensuring the achievement of financial performance indicators (ROA, ROE, Net Profit, etc.) of South-East Europe IT industry companies. In their study, as a theoretical background they used a resource-based view on intellectual capital, which involves analyzing the impact of its individual components on financial performance indicators. In particular, the authors established that innovation capital has the strongest impact and human capital has an indirect impact on the financial performance of IT companies. A similar resource-based approach was also used by Serpeninova et al. [20], who as a result of a study of the impact of intellectual capital on the profitability of Slovak software development companies (ROA, NPM, GPM, EBITM) found an absence of a significant relationship between them. The authors considered the main reason for this to be the imperfection of the current accounting standards, for instance, IAS 38, in terms of criteria for recognizing and evaluating the intellectual capital of enterprises.

The analysis of studies on the issues of the research made it possible to establish the existence of mutually contradictory evidence regarding the impact of intangible assets on the financial performance. In general, this does not allow the management of enterprises to effectively control intangible values aimed at creating internal value, and for investors – to receive clear signals for making effective investments. Considering the above, the following objectives were formulated: to measure the relationship between intangible assets and the financial performance of Slovak ICT companies; to investigate which components of intangible assets have the most significant or insignificant impact on the financial performance of Slovak ICT companies; to form recommendations for improving the investment policy of ICT companies, based on the level of significance of the elements of intangible assets from the point of view of increasing financial results.

3. Data and methodology

3.1. Sample selection

To determine whether intangible assets stimulate financial performance, was analyzed sample of 180 Slovak ICT companies for the period 2015–2019. In particular, the panel data information from financial statements of such enterprises, available in the open access, as well as the information from database "FinStat" was used to form panel data. Only those companies, for which the necessary information for the 5-year period was available, were included in the sample. The selected 180 companies provide a valid and complete set of data in order to carry out relevant statistical analysis.

Investigated enterprises proceeding from EU Economic Activity Classification and from the SK NACE 2 classification belongs to group 26 "Manufacture of computer, electronic and optical products", includes direct production of computers, computer peripheral equipment (input device, output device, input/output device), communication equipment (public switching equipment, transmission equipment, customer premises equipment), measuring, medical, navigation, radio, optical and other electronic equipment, as well as production of various types of accessories for such products (electrical boards, magnetic and optical media, etc.). In order to take into account the influence sub-sectors affiliation on financial performance of ICT companies two groups were allocated in their composition. The first group included enterprises dealing with the production of different types of electronics and components, and the second group involved enterprises producing communication equipment and components.

Based on the form of ownership, most of the companies investigated – 160, companies with limited liability, 16 – is a joint-stock company, 2 – production cooperative, 1 – limited partnership, 1 – general partnership. By type of ownership, the companies investigated are divided as follows: private domestic – 64%; foreign – 21%; international with a predominant private sector – 13%; cooperative – 1%; state – 1%.

3.2. Variables

In the research for characteristics of financial performance of ICT companies were used four dependent variables – Return on Assets, Net Profit Margin, Return on Equity, Assets Turnover, and used in their work by researchers for simiral empirical analysis of the relationship between intangibles values and company financial performance [11, 13, 16, 18, 19, 20]. For explanation of a relation between intangible assets and financial performance of ICT companies used intangible assets variables – Research and Development Intensity, Research and Development Intensity Squared, Software, Intellectual Property Rights, Acquired Intangible Assets. The election of such independent variable is justified by the financial statements of Slovak ICT companies in the disclosure of information about intangible assets. As it was revealed by Huňady et al. [8], the firm's ICT sector account for significant share of total business R&D expenditure in economy in most countries. Therefore, in the analysis impact of intangible assets on financial performance of ICT sector an important role should be assigned to R&D indicators. As a result, the study does not use the indicator of R&D costs but uses two calculation ratios that characterize the R&D of the companies. In addition, based on previous studies [21, 22, 20] in our study used three control variables – Leverage, Size and Dummy variable for ICT sub-sectors. Use of these variables will allow to control for a significant effects of company size, level of borrowing capital, and unseen role of ICT sub-sectors affiliation.

Types, calculation procedures, and abbreviations used in the Variables study are shown in table 2.

The dynamics of four indicators, that characterize financial performance of Slovak ICT companies (ROA, NPM, ROE, ATO) for the period 2015–2019 showed in figure 3.

Figure 3 displays the change in time of financial performance indicators for the 2015–2019 period. It allows to identify a number of common trends: Simultaneous growth in all indicators for 2017–2018 years; decrease in ATO, ROA and NPM indicators for 2015–2016 years, their growth in 2016–2018 years, as well as their simultaneous decrease in 2018–2019; during 2018–2019 years only growth of ROE indicator occurs. In general, common behavior was found for ATO, ROA and NPM, as well as almost completely different behavior of ROE compared to these indicators.

4. Research models

To understand the relationship between intangible assets and financial performance indicators, this study examined four following models:

Model 1: $\mathrm{ROA}_{it} = \alpha + \beta_1 \cdot \mathrm{RDI}_{it} + \beta_2 \cdot \mathrm{RDI2}_{it} + \beta_3 \cdot \mathrm{SOFT}_{it} + \beta_4 \cdot \mathrm{IPR}_{it} + \beta_5 \cdot \mathrm{AIA}_{it} + \beta_6 \cdot \mathrm{LEV}_{it} + \beta_7 \cdot \mathrm{l\_SIZE}_{it} + \beta_8 \cdot \mathrm{DVICTSS}_{it} + \epsilon_{it}$

Table 2
Variable definitions and abbreviations.

Variable	Calculation (Source)	Abbreviation
Dependent Variables		
Return on Assets	Net turnover / Total Assets	ROA
Net Profit Margin	Net profit / Total Sales	NPM
Assets Turnover	Total Sales / Total Assets	ATO
Return on Equity	Net profit / Total Equity	ROE
Independent Variables		
Intangible Assets Variables		
Research and Development Intensity	Capitalized R&D Costs / Total Sales	RDI
Research and Development Intensity Squared	Squared function of RDI	RDI2
Software	Software (Intangible Asset)	SOFT
Intellectual Property Rights	Valuable Intellectual Property Rights	IPR
Acquired Intangible Assets	Acquired long-term intangible assets are charged until the time of their use	AIA
Control Variables		
Leverage	Total liabilities / Total Assets	LEV
Size	Logarithm of Total Assets	l_SIZE
Dummy variable for ICT sub-sectors	1 for electronic producers, 0 for communication producers	DVICTSS

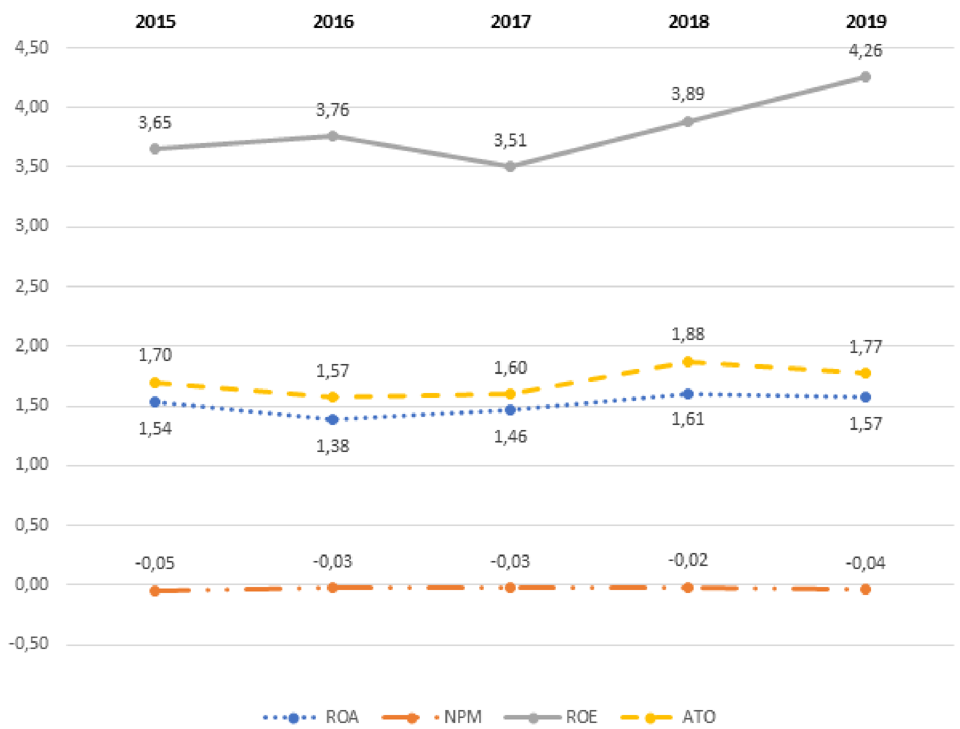


Figure 3: Dynamics of financial performance indicators of Slovak ICT companies for the 2015-2019 period.

Model 2: $\text{NPM}_{it} = \alpha + \beta_1 \cdot \text{RDI}_{it} + \beta_2 \cdot \text{RDI2}_{it} + \beta_3 \cdot \text{SOFT}_{it} + \beta_4 \cdot \text{IPR}_{it} + \beta_5 \cdot \text{AIA}_{it} + \beta_6 \cdot \text{LEV}_{it} + \beta_7 \cdot \text{l\_SIZE}_{it} + \beta_8 \cdot \text{DVICTSS}_{it} + \epsilon_{it}$

Model 3: $\text{ATO}_{it} = \alpha + \beta_1 \cdot \text{RDI}_{it} + \beta_2 \cdot \text{RDI2}_{it} + \beta_3 \cdot \text{SOFT}_{it} + \beta_4 \cdot \text{IPR}_{it} + \beta_5 \cdot \text{AIA}_{it} + \beta_6 \cdot \text{LEV}_{it} + \beta_7 \cdot \text{l\_SIZE}_{it} + \beta_8 \cdot \text{DVICTSS}_{it} + \epsilon_{it}$

Model 4: $\text{ROE}_{it} = \alpha + \beta_1 \cdot \text{RDI}_{it} + \beta_2 \cdot \text{RDI2}_{it} + \beta_3 \cdot \text{SOFT}_{it} + \beta_4 \cdot \text{IPR}_{it} + \beta_5 \cdot \text{AIA}_{it} + \beta_6 \cdot \text{LEV}_{it} + \beta_7 \cdot \text{l\_SIZE}_{it} + \beta_8 \cdot \text{DVICTSS}_{it} + \epsilon_{it}$

where: ROA, NPM, ATO, ROE – dependent variables, where i is entity and t is time;

α – Identifier;

μ – Variance introduced by the unit-specific effect for unit i;

β – Regression coefficient;

RDI, RDI2, SOFT, IPR, AIA – independent intangible variables, LEV, l_SIZE, DVICTSS – independent control variables;

ϵ_{it} – error term.

Figure 4 shows the conceptual framework of the study.

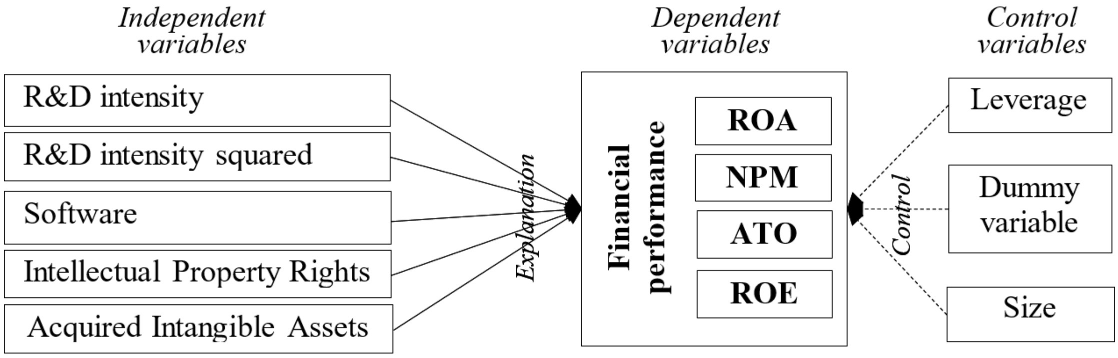


Figure 4: Conceptual framework of the study.

5. Results

5.1. Descriptive statistics and correlations

The descriptive statistics (observation, mean, median, standard deviation, minimum, maximum) of a full sample are presented in table 3.

From table 3 it can be observed that the full sample is measured with 180 units. The largest deviations in variables are related to SOFT ($5{,}95 \cdot 10^4$), IPR ($2{,}89 \cdot 10^4$), AIA ($1{,}61 \cdot 10^5$) and ROE (4,30). Large differences between the minimum and the maximum values of ROA, ATO, and ROE show that the financial performance levels of ICT companies are quite distinct. For some variables (ATO, LEV, IPR, AIA, l_SIZE) the mean value is greater than the standard deviation value, as a result, the data in these variables have a small distribution. ROA, NPM, and ROE have a higher standard deviation than their mean. This indicates a relatively large set of ratios that will characterize the normal distribution curve and will not be outliers. The closeness of the mean (13,5) and median (13,3) values for l_SIZE indicates a high level of symmetry in the

Table 3
Descriptive statistics of variables (based on observations 1:1 – 180:5).

Variables	Observation	Mean	Median	St. Dev.	Minimum	Maximum
ROA	180	1,51	1,27	1,52	2,75e-005	24,9
NPM	180	-0,0325	0,00798	0,501	-6,80	2,97
ATO	180	1,70	1,39	1,61	6,88e-005	24,6
ROE	180	3,81	2,33	4,30	0,000120	38,2
LEV	180	0,438	0,429	0,265	0,000	0,988
RDI	180	0,129	0,000	0,652	-0,0346	9,91
RDI2	180	0,442	0,000	4,96	0,000	98,2
SOFT	180	$2{,}00 \cdot 10^4$	0,000	$5{,}95 \cdot 10^4$	0,000	$5{,}18 \cdot 10^5$
IPR	180	$8{,}27 \cdot 10^3$	0,000	$2{,}89 \cdot 10^4$	$-2{,}57 \cdot 10^4$	$2{,}67 \cdot 10^5$
AIA	180	$2{,}00 \cdot 10^4$	0,000	$1{,}61 \cdot 10^5$	0,000	$3{,}20 \cdot 10^6$
l_SIZE	180	13,5	13,3	2,00	8,35	18,9

distribution of range values, that is, the size of the studied enterprises. The mean value of the LEV ratio is 0,438, and this means that approximately 44% of the total assets of ICT companies are financed through borrowed resources.

In general, correlation matrix of variables used in Models 1-4 (figure 5), testifies to absence multicollinearity problem, since in most cases, the correlation coefficient is less than 0,5 (−0,5). The only exception is the high correlation coefficient between variables RDI and RDI2 (0,9), which is understandable given that RDI2 is a squared function of RDI. However, as Özkan [23] notes, the practice of applying such mutually-correcting indicators is normal in the regression analysis performed to check the effect of interrelated variables on financial performance indicators. In particular, simultaneous use in regression models of variables RDI and RDI2 allows to detect presence U-inverted relation between R&D and financial performance of a company.

5.2. Selection of estimate panel data parameter

The choice of estimate panel data parameter for each of the selected models plays an important role in the regression analysis of panel data. This parameter should be adequately correlated with the data used in the corresponding model. Proceeding from F-statistics test for Model 1 F(179; 712) = 1,17767 with p-value 0,0766456, which is more than 0,05 and confirms null hypothesis in relation to pooled OLS model. The need for such a choice estimate parameter for Model 1 also confirmed the application Breusch-Pagan test, according to which chi-square (1) > 2,04561 p-value = 0,152645, which is larger than 0.05 and confirms zero hypotheses. The use of F-statistics test and Breusch-Pagan test also confirmed the need for use pooled OLS model as a quality estimate parameter for Model 2. For Model 3 after application F-statistics test it was received F(179; 712) = 1,23387 with p-value 0,0331413, that is less than 0,05 and testifies to the adequacy of application Fixed effects method (FEM). However, this conclusion is refuted as a result Breusch-Pagan test, according to chi-square (1) > 3,58479 p-value = 0,0583107, which is larger than 0,05 and confirms zero hypothesis of adequacy pooled OLS model. Considering the results Hausman test (p-value = prob(chi-square (8) > 4,34179) = 0,825045), according to which more appropriate is the application of Random effects method (REM) than FEM, for

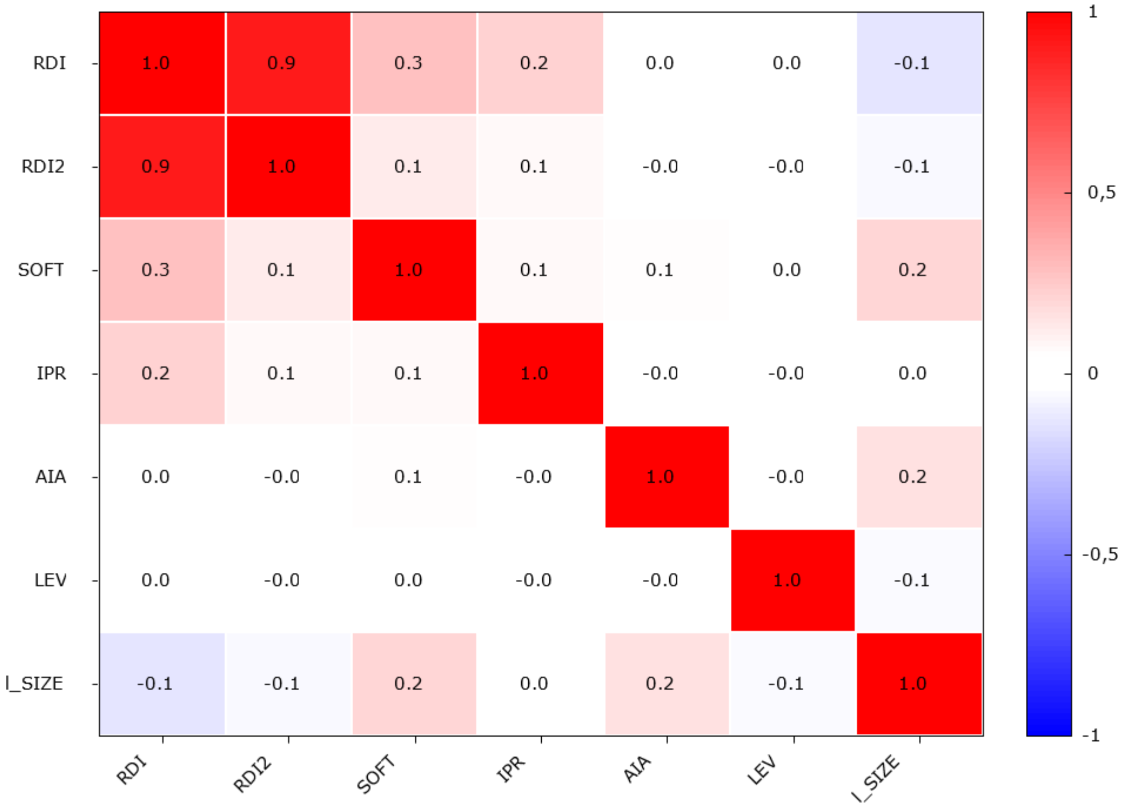


Figure 5: Correlation matrix of variables used in Models 1-4 (calculated via GRETL software package).

Model 3 more appropriate also consider the application of pooled OLS model. For Model 4 after application of F-statistics test F(179; 712) = 1,32394 of p-value 0,00693691, which is less than 0,05 and shows the adequacy of application of FEM. This is the test followed by the p-value = P(chi-square (1) > 6,04321) = 0,0139599.

5.3. Assumption test results

To verify the adequacy of the Panel data for Models 1-4 that is collected about ICT companies, it should be diagnosed using Normality test, Autocorrelation test and Heteroscedasticity test. Normality test for all Models 1-4 allowed to detect abnormal distribution of the error. For example, for Model 1 for chi-square (2) = 4119,75 p-value = 0, which is less than 0,05, and does not confirm zero hypotheses about the normal distribution of balances. Review null hypothesis about no first-order autocorrelation based on usage Wooldridge test for autocorrelation allowed to confirm it for all four models. In particular, for all Models 1-4 p-value it is more than 0,05 (0,73367; 0,923389; 0,193049; 0,227822), confirming null hypothesis. White test was used to check the heteroscedasticity of a models 1–3. Since the obtained p-value for each of the three models (0,284134; 0,999935; 0,421088) is more than the critical value, the zero hypothesis about the absence of heteroscedasticity is forgiven. For Model 4 with estimate parameter FEM was applied non-parametric Walk test, which also was established the presence of heteroscedasticity. In particular, chi-square(180) = 78593,1 p-value = 0 was received. Since p-value is less than 0,05, there is an inhomogeneous observation and a different variance of a Model 4 random error, which confirms the existence of heteroscodesticity.

To solve the problem of inadequacy of all Models 1–4 used by this data due to the problem of improper distribution of the error and heteroscedasticity, the use of robust estimators is proposed. They help minimize or eliminate impact of outliers in a Models 1-4, improving the results of panel data regression analysis. Practice of use robust standard errors in regression analysis was also used in research of scientists who study the impact of intangible assets and their components on the performance of enterprises [23, 20].

5.4. Panel data regression results

Model 1 (ROA). Tables 4–5 show the results of regression analysis performed using pooled OLS model. They show how the independent variable will affect the dependent variable, which of the regressions have significant influence, force and direction of such influence.

Table 4
Model 1 (ROA). Pooled OLS model (Robust standard errors), using the observations 1–900.

Variable	Coefficient	Standard error	z	P-value	Significance by t-statistics
const	1,83324	0,632512	2,898	0,0038	***
RDI	−1,16410	0,157720	−7,381	<0,0001	***
RDI2	0,110937	0,0175566	6,319	<0,0001	***
SOFT	$1{,}65440 \cdot 10^{-6}$	$5{,}38521 \cdot 10^{-7}$	3,072	0,0021	***
IPR	$1{,}34184 \cdot 10^{-6}$	$9{,}24827 \cdot 10^{-7}$	1,451	0,1468	
AIA	$-4{,}98766 \cdot 10^{-7}$	$1{,}38889 \cdot 10^{-6}$	−3,591	0,0003	***
LEV	−0,137738	0,214780	−0,6413	0,5213	
l_SIZE	−0,0379800	0,0454674	−0,8353	0,4035	
DVICTSS	0,168307	0,105500	1,595	0,1106	

Note: *** Significant at the 1% level.

Table 5
Model 1 (ROA). Pooled OLS model (Robust standard errors), using the observations 1–900.

Indicator	Value	Indicator	Value
Mean dependent var.	1,511304	S.D. dependent var.	1,524745
Sum squared resid.	1991,445	S.E. of regression	1,495014
R-squared	0,047173	Adjusted R-squared	0,038618
F(8, 179)	21,20706	P-value (F)	$1{,}86 \cdot 10^{-22}$

Model 1 can be interpreted through the following equation:

$\hat{y} = 1,83324 - 1,16410 \cdot 10^{-6} x_1 + 0,110937 x_2 + 1,65440 \cdot 10^{-6} x_3 + 1,34184 \cdot 10^{-6} x_4 - 4,98766 \cdot 10^{-7} x_5 - 0,137738 x_6 - 0,0379800 x_7 + 0,168307 x_8$ where: $\hat{y}$ – ROA; x_1 – RDI; x_5 – AIA; x_2 – RDI2; x_6 – LEV; x_3 – SOFT; x_7 – l_SIZE; x_4 – IPR; x_8 – DVICTSS.

Based on the results of the regression analysis, const, RDI, RDI2, SOFT and AIA are statistically significant (there are stars in the last column of table 4), having the highest level of significance at the 1% level. Accordingly, these indicators have the highest impact on ROA. In addition to RDI and AIA, other significant indicators have a direct impact on ROA and RDI and AIA are rotating. The presence of a different direction of influence in RDI and RDI2 indicates the presence of U-inverted relationship between R&D and ROA [24]. Similar U-inverted behavior is common to most of the costs of non-material nature, in particular, social and environmental costs [25]. The results also show that there is no significant influence of control variables (Lev, l_SIZE, DVICTSS) on ROA.

The overall content of the regression coefficient of Model 1 is that with an increase of 1 directly influencing the ROA, the last increase in the ratio will be increased. For example, if SOFT is increased by 1, the ROA will increase by 1,65440 $\cdot 10^{06}$. And for indicators that have a positive impact on ROA, their increase by 1 for ICT enterprises will result in corresponding decrease of ROA (depending on the coefficient of regression).

Table 5 indicates that the coefficient of determination (R-squared) of Model 1 is 0,047173. This means only that 4,7% of the variation of ROA can be explained by the variation of the independent variables (const, RDI, RDI2, SOFT, IPR, AIA, LEV, l_SIZE, DVICTSS).

Model 2 (NPM). Model 2 can be interpreted through the following equation:

$\hat{y} = -0,274718 + 0,0626252x_1 - 0,00669466x_2 - 5,13111 \cdot 10^{-8}x_3 + 2,00654 \cdot 10^{-7}x_4 - 5,88982 \cdot 10^{-8}x_5 - 0,0630010x_6 + 0,0269143x_7 - 0,0517929x_8$

where: $\hat{y}$ – NPM; $x_1 - x_8$ – the same as in Model 1.

Based on table 6, the most significant effect on NPM is changed to l_SIZE. Accordingly, with the growth of the enterprise volume by 1 increases the value of the NPM indicator by 0,0269143. Significant at the 5% level in NPM explanation have regressors const, RDI, RDI2 and AIA. Also significant at the 10% level is the DVICTSS regression, which has an indirect effect. Indirect effects on NPM are also affected by the RDI2 and AIA indicators. This means that, as investments in such types of intangible assets increase, the corresponding (depending on the regression coefficient) reduction of the dependent variable will occur. By comparing the coefficient of Model 2 with RDI and RDI2, it is possible to note the existence of the upper limit of investments in R&D of Slovak ICT companies, after which their negative impact on NPM will already be observed.

Table 7 indicates that the R-squared of Model 2 is 0,01, a very low value and does not allow to speak about the significant role of intangible assets in NPM provision. This means that 1,33% of the variation of the NPM can be explained by the variation of regressors.

Model 3 (ATO). Model 3 can be interpreted through the following equation:

$\hat{y} = 2,80330 - 1,42622x_1 - 0,134772x_2 + 2,76920 \cdot 10^{-6}x_3 + 4,61116 \cdot 10^{-6}x_4 - 4,38715 \cdot 10^{-7}x_5 - 0,139781x_6 - 0,0951285x_7 + 0,150857x_8$

where: $\hat{y}$ – ATO; $x_1 - x_8$ – the same as in Model 1.

For dependent variable ATO except for LEV and DVICTSS, all other regressions are significant. In particular, l_SIZE significant at the 5% level, and all other regressions (const, RDI, RDI2, SOFT, IPR and AIA) significant at the 1% level. Direct effects on ATO from the regression data are RDI2, SOFT and IPR, while others are affected. In particular, as in Model 1 for ROA, making a small amount of investments in R&D of Slovak ICT companies has a negative impact on ATO. Only their implementation from a certain volume, in particular, in the volume of RDI2, ensures

Table 6
Model 2 (NPM). Pooled OLS model (Robust standard errors), using the observations 1–900.

Variable	Coefficient	Standard error	z	P-value	Significance by t-statistics
const	−0,274718	0,116547	−2,357	0,0184	**
RDI	0,0626252	0,0295138	2,122	0,0338	**
RDI2	−0,00669466	0,00314309	−2,130	0,0332	**
SOFT	$-5{,}13111 \cdot 10^{-8}$	$1{,}17552 \cdot 10^{-7}$	−0,4365	0,6625	
IPR	$2{,}00654 \cdot 10^{-7}$	$1{,}99115 \cdot 10^{-7}$	1,008	0,3136	
AIA	$-5{,}88982 \cdot 10^{-8}$	$2{,}90523 \cdot 10^{-8}$	−2,027	0,0426	**
LEV	−0,0630010	0,0813673	−0,7743	0,4388	
l_SIZE	0,0269143	0,00838749	3,209	0,0013	***
DVICTSS	−0,0517929	0,0272405	−1,901	0,0573	*

Note:
* Significant at the 10% level;
** Significant at the 5% level;
*** Significant at the 1% level.

Table 7
Model 2 (NPM). Pooled OLS model (Robust standard errors), using the observations 1–900.

Indicator	Value	Indicator	Value
Mean dependent var.	-0,032511	S.D. dependent var.	0,501315
Sum squared resid.	222,9308	S.E. of regression	0,500203
R-squared	0,013291	Adjusted R-squared	0,004432
F(8, 179)	2,238141	P-value (F)	0,026686

Table 8
Model 3 (ATO). Pooled OLS model (Robust standard errors), using the observations 1–900.

Variable	Coefficient	Standard error	z	P-value	Significance by t-statistics
const	2,80330	0,637648	4,396	<0,0001	***
RDI	−1,42622	0,174982	−8,151	<0,0001	***
RDI2	0,134772	0,0192228	7,011	<0,0001	***
SOFT	$2{,}76920 \cdot 10^{-6}$	$6{,}38619 \cdot 10^{-7}$	4,336	<0,0001	***
IPR	$4{,}61116 \cdot 10^{-6}$	$1{,}59970 \cdot 10^{-6}$	2,883	0,0039	***
AIA	$-4{,}38715 \cdot 10^{-7}$	$1{,}36505 \cdot 10^{-7}$	−3,214	0,0013	***
LEV	−0,139781	0,233510	−0,5986	0,5494	
l_SIZE	−0,0951285	0,0438796	−2,168	0,0302	**
DVICTSS	0,150857	0,124085	1,216	0,2241	

Note:
** Significant at the 5% level;
*** Significant at the 1% level.

the growth of ATO. Based on an equal to 1,3 RDI2 growth by 1 increases the NPM value by 0,0269143. Table 9 indicates that the R-squared of Model 3 is 0,056. This means that 5,61% of the variation of the ATO can be explained by the variation of regressors.

Model 4 (ROE). Model 4 can be interpreted through the following equation:

Table 9
Model 3 (ATO). Pooled OLS model (Robust standard errors), using the observations 1–900.

Indicator	Value	Indicator	Value
Mean dependent var.	1,703817	S.D. dependent var.	1,614880
Sum squared resid.	2212,759	S.E. of regression	1,575898
R-squared	0,056170	Adjusted R-squared	0,047696
F(8, 179)	15,95424	P-value (F)	$1{,}15 \cdot 10^{-17}$

$\hat{y} = -1,06067 - 2,79903x_1 + 0,272431x_2 + 5,89712 \cdot 10^{-6}x_3 + 8,97938 \cdot 10^{-7}x_4 - 1,27997 \cdot 10^{-6}x_5 + 8,94081x_6 + 0,0265371x_7 + 0,392670x_8$

where: $\hat{y}$ – ROE; $x_1 - x_8$ – the same as in Model 1.

Model 4 has five statistically significant regressors – RDI, RDI2, SOFT, AIA and LEV (table 10). All of them have the highest level of significance – 1%, therefore they have the greatest influence on the dependent variable (ROE). The equation of Model 4 shows that most of the independent variables (RDI2, SOFT, IPR, LEV, l_SIZE and DVICTSS) have a direct influence, and only two variables (const, RDI and AIA) have a rotational influence on the ROE. As in Models 1 and 3, Model 4 has a U-inverted relationship between R&D and ROA, characterized by the need to increase investment in R&D of Slovakia ICT companies to ensure their positive impact on ROE.

Table 10
Model 4 (ROE). FEM (Robust standard errors), using the observations 1–900.

Variable	Coefficient	Standard error	z	P-value	Significance by t-statistics
const	−1,06067	1,27812	−0,8299	0,4066	
RDI	−2,79903	0,466001	−6,006	<0,0001	***
RDI2	0,272431	0,0565526	4,817	<0,0001	***
SOFT	$5{,}89712 \cdot 10^{-6}$	$2{,}18175 \cdot 10^{-6}$	2,703	0,0069	***
IPR	$8{,}97938 \cdot 10^{-7}$	$3{,}30604 \cdot 10^{-6}$	0,2716	0,7859	
AIA	$-1{,}27997 \cdot 10^{-6}$	$3{,}62115 \cdot 10^{-7}$	−3,535	0,0004	***
LEV	8,94081	0,614350	14,55	<0,0001	***
l_SIZE	0,0265371	0,0848603	0,3127	0,7545	
DVICTSS	0,392670	0,344744	1,139	0,2547	

Note: *** Significant at the 1% level.

Table 11 indicates that the LSDV R-squared of Model 4 is 0,51. This is quite a high value compared to the 1–3 models, but not enough to speak about the significant role of intangible assets in providing of financial performance of ICT companies. This means that 51,61% of the variation of the ROE can be explained by the variation of the regressors.

Table 11
Model 4 (ROE). FEM (Robust standard errors), using the observations 1–900.

Indicator	Value	Indicator	Value
Mean dependent var.	3,812005	S.D. dependent var.	4,304137
Sum squared resid.	8058,382	S.E. of regression	3,364216
LSDV R-squared	0,516144	Within R-squared	0,348421

6. Discussion

The results obtained in the article partially confirm the conclusions of the analyzed works on the role of intangible assets in the promotion of financial performance of high-tech companies. As for some regressions, they are in conflict with such conclusions. The existence of a positive and significant relationship between intangible assets and some financial performance measures was confirmed, which is also set in the works of Li and Wang [12], Dženopoljac et al. [13], Zhang [15]. The presence was also established of negative and significant impact of AIA on all financial performance indicators, this confirms the results of the research [16, 17]. At the same time, the direction and influence of different types of regressions used in the study are not the same in all formed models, but depends on a particular kind of financial performance indicator. One of the reasons for this is that the relationship between intangible assets on financial performance may depend on macroeconomic factors, in particular, on the level of science capacity in the industry and on the level of innovation in the country, which is noted by Qureshi and Siddiqui [16]. Another reason for such results may be incomplete information about intangible assets disclosed in the financial statements of Slovak ICT companies. In turn, this is a consequence of the conservatism of the current methodology of recognizing and evaluating intangible assets, which Zhang [15] also points out, Radonić et al. [19]. Therefore, the findings of this study confirm the proposal of Serpeninova et al. [20] regarding the necessity of expanding the criteria for recognizing and the structure of financial reporting for high-tech companies regarding intangible assets.

The results of the survey refutes the conclusions of Gan and Saleh [11] on the positive impact of the company's size on the improvement of financial performance (ROA), but such an impact was found with respect to NPM. The above confirms the hypothesis of Del Monte and Papagni [26] that to increase the returns from intangible investments should be provided with their proper quality level, not quantitative imitations. Therefore, an intangible investment policy of ICT companies should be based not only on quantitative parameters, that is, not on the basis of total investment in the company, but on the individual role of certain types of intangible assets in improving of financial performance.

The study has some limitations, which should be taken into account by other scientists when evaluating the results of a study. Firstly, given the sufficient breadth of the term "financial performance", a list of dependent variables used in the study can be specified. Second, the list of independent variables used in a study can be expanded by uncapitalized intangible assets that also affect the financial performance of Slovakia ICT companies. However, it is necessary to separate from the composition of different types of expenses of ICT companies those expenses connected with creation of intangible assets (client, ecological, social, etc.), as such data are not in financial statements of companies. Third, to determine the role of intangible assets in improvement of financial performance, research can be carried out not only on the examples of companies of ICT industry, but also on the example of other branches of economy. This will allow to carry out an interindustry comparison and establish in which areas of management of enterprises should pay the most attention to development of an intangible investment policy.

7. Conclusion

This research was undertaken with the objective of comprehending the ramifications of intangible assets on the financial performance of high-tech enterprises. The focus of this study was the analysis of 180 Slovak ICT companies over the period spanning 2015 to 2019. This inquiry gains particular pertinence against the backdrop of the pivotal role that the ICT sector plays in propelling the development of the Slovak economy. The Slovak Government has proactively established conducive institutional conditions to facilitate the growth of ICT companies and has initiated specialized programs to incentivize investments in this sector.

Panel data regression analysis served as the foundational methodology for this investigation. The financial performance was characterized through four dependent variables: Return on Assets, Net Profit Margin, Assets Turnover, and Return on Equity. For each of these indicators, a distinct model was constructed, incorporating eight independent variables. The intangible asset variables encompassed Research and Development Intensity, Research and Development Intensity Squared, Software, Intellectual Property Rights, and Acquired Intangible Assets. Additionally, three control variables were integrated: Leverage, Size, and a Dummy variable denoting ICT sub-sectors, within the temporal scope of 2015 to 2019. The selection of the optimal panel data parameter for each model was grounded in statistical tests, such as the F-statistics test, Breusch-Pagan test, and Hausman test (Models 1 to 3 – pooled OLS model, Model 4 – Fixed Effects Method). To assess the models' compatibility with the generated data, the Normality test, Autocorrelation test (Wooldridge test for autocorrelation), and Heteroscedasticity test (White test, Walk test) were employed, substantiating the application of robust standard errors due to partial model adequacy.

The study's hypothesis was affirmed to a certain extent through the outcomes of the panel regression analysis. The results unveiled that not all categories of intangible assets wield a significant positive influence on the financial performance of Slovak ICT companies. Specifically, Research and Development Intensity (RDI), Research and Development Intensity Squared (RDI2), and Acquired Intangible Assets (AIA) exhibited substantial impact across various degrees on the four distinct financial performance indicators. This emphasizes the rationale for management to channel investments into these specific categories of intangible assets for Slovak ICT companies. Furthermore, the contrasting directions of influence of RDI and RDI2 on financial performance indicators underscore the existence of a U-inverted relationship between R&D investment and the financial performance metrics. Depending on the model, RDI either operates beyond the threshold of returns on R&D investments or within it, while RDI2 exhibits the inverse relationship. These findings offer managerial insights into optimizing R&D investments based on the desired financial performance outcomes. Remarkably, across all models, Acquired Intangible Assets (AIA) demonstrated substantial significance but negatively affected the financial performance of Slovak ICT enterprises. This indicates the need for more expeditious integration of these long-term intangible assets into the operational fabric of the companies.

Furthermore, the research underscores the need for a comprehensive system to plan the assimilation of intangible assets, tailoring them to the company's exigencies as a core component of its intangible investment strategy. The analysis of control variables, including Leverage, Size, and Dummy variable for ICT sub-sectors, demonstrated selective impact on financial performance indicators. Notably, only the variable l_SIZE exhibited a significant influence on Net Profit Margin (NPM) and Assets Turnover (ATO), with implications for managerial considerations in optimizing these metrics. The variable DVICTSS exhibited limited significance on NPM, while LEV affected Return on Equity (ROE). These findings underscore the nuanced and selective nature of control variables' influence on various financial performance indicators, with no discernible effect on Return on Assets (ROA).

In the broader context, this study serves to illuminate the intricate relationships between intangible assets and financial performance in the ICT sector. The insights garnered herein have the potential to inform prudent decision-making within the domain of intangible asset investments and their subsequent impact on financial performance for Slovak ICT companies. Furthermore, this research offers valuable insights for policy and strategic adjustments aimed at fostering the incorporation of long-term intangible assets into business operations, thus enhancing financial performance. As the ICT sector continues to evolve and shape the Slovak economy, these findings can serve as a compass guiding effective strategies for leveraging intangible assets in this dynamic landscape.

As the global business environment continually evolves, future research endeavors could delve deeper into the interplay between various intangible asset categories and financial performance across different industries and geographical contexts. Such investigations would not only enrich the existing body of knowledge but also provide actionable insights for enterprises seeking to optimize their financial performance through targeted intangible asset investments.

Acknowledgments

This article is an output of the project of the Scientific Grant Agency of the Ministry of Culture of the Slovak Republic and Slovak Academy of Sciences (VEGA) no. 1/0517/20 (2020–2022) "Virtual Cryptochains as a Relevant Tool to Eliminate Economic Crime"

References

[1] S. Lehenchuk, T. Vakaliuk, T. Nazarenko, Z. Kubaščíková, Z. Juhászová, Empirical Evidence of Intangible Assets Improve the Financial Performance of Slovak ICT Companies, in: S. Semerikov, V. Soloviev, A. Matviychuk, V. Kobets, L. Kibalnyk, H. Danylchuk, A. Kiv (Eds.), Proceedings of 10th International Conference on Monitoring, Modeling & Management of Emergent Economy - M3E2, INSTICC, SciTePress, 2023, pp. 38–52. doi:10.5220/0011931000003432.

[2] J. H. Daum, Intangible Assets and Value Creation, John Wiley & Sons Ltd., Chichester, 2002.

[3] I. Abeysekera, Intellectual Capital Accounting: Practices in a developing country, Routledge, London, 2008. doi:10.4324/9780203937617.

[4] M. D. Moberly, Safeguarding Intangible Assets, Butterworth-Heinemann, 2014. doi:10.1016/C2013-0-15617-6.

[5] E. Ullberg, L. Edvinsson, C. Yeh-Yun Lin, Intangible Asset Gap in Global Competitiveness: Mapping and Responding to the New Economy, SpringerBriefs in Business, Springer, Cham, 2021. doi:10.1007/978-3-030-55666-2.

[6] T. Vakaliuk, O. Spirin, O. Korotun, D. Antoniuk, M. Medvedieva, I. Novitska, The current level of competence of schoolteachers on how to use cloud technologies in the educational process during COVID-19, Educational Technology Quarterly 2022 (2022) 232–250. doi:10.55056/etq.32.

[7] S. S. Iyer, L. Gernal, R. Subramanian, A. Mehrotra, Impact of digital disruption influencing business continuity in UAE higher education, Educational Technology Quarterly 2023 (2023) 18–57. doi:10.55056/etq.29.

[8] J. Huňady, P. Pisar, I. Durcekova, Business R&D Expenditure in the ICT Sector: Effects on Business Performance Indicators, in: Proceedings of the ENTRENOVA - ENTerprise REsearch InNOVAtion Conference, volume 5, Rovinj, Croatia, 2019, pp. 519–530. URL: https://www.econstor.eu/handle/10419/207714.

[9] SARIO, Information & Communications Technology Sector in Slovakia, SARIO. Slovak Investment and Trade Development Agency, Bratislava, 2021. URL: https://sario.sk/sites/default/files/sario-ict-2021-02-05.pdf.

[10] Deloitte Touche Tohmatsu Limited, IAS 38 — Intangible Assets, 2022. URL: https://www.iasplus.com/en/standards/ias/ias38.

[11] K. Gan, Z. Saleh, Intellectual Capital and Corporate Performance of Technology-Intensive Companies: Malaysia Evidence, Asian Journal of Business and Accounting 1 (2008) 113–130. URL: https://ajba.um.edu.my/article/view/2197.

[12] H. Li, W. Wang, Impact of Intangible Assets on Profitability of Hong Kong Listed Information Technology Companies, Business and Economic Research 4 (2014) 98–113. doi:10.5296/ber.v4i2.6009.

[13] V. Dženopoljac, S. Janoševic, N. Bontis, Intellectual capital and financial performance in the Serbian ICT industry, Journal of Intellectual Capital 17 (2016) 373–396. doi:10.1108/JIC-07-2015-0068.

[14] A. M. Khan, An Empirical Study of the Impact of Intellectual Capital on the Financial Performance of the Indian IT Sector, Journal of Corporate Finance Research 15 (2018) 7–19. doi:10.17323/j.jcfr.2073-0438.12.1.2018.7-19.

[15] N. Zhang, Relationship between intangible assets and financial performance of listed telecommunication firms in China, based on empirical analysis, African Journal of Business Management 11 (2017) 751–757. doi:10.5897/AJBM2017.8429.

[16] M. J. Qureshi, D. A. Siddiqui, The Effect of Intangible Assets on Financial Performance, Financial Policies, and Market Value of Technology Firms: A Global Comparative Analysis, Asian Journal of Finance & Accounting 12 (2020) 26–57. doi:10.5296/ajfa.v12i1.16655.

[17] I. T. Lopes, C. F. P. Ferreira, Intangibles as innovative drivers for competitive businesses, International Journal of Business Innovation and Research 24 (2021) 238–260. doi:10.1504/ijbir.2020.10024988.

[18] M. Sundaresan, T. P. N. Linh, M. Rey, The Effects of Intangible Assets on Financial Performance and Financial Policies of Listed Technology Firms in Thailand, Apheit International Journal 10 (2021) 1–17. URL: https://www.journals.apheit.org/jounal/Inter-vol10no1/INT-01.pdf.

[19] M. Radonić, M. Milosavljević, S. Knežević, Intangible Assets as Financial Performance Drivers of IT Industry: Evidence from an Emerging Market, E&M Economics and Management 24 (2021) 119–135. doi:10.15240/tul/001/2021-2-008.

[20] Y. Serpeninova, S. Lehenchuk, M. Mateášová, T. Ostapchuk, I. Polishchuk, Impact of intellectual capital on profitability: Evidence from software development companies in the Slovak Republic, Problems and Perspectives in Management 20 (2022) 411–425. doi:10.21511/ppm.20(2).2022.34.
[21] V. Ievdokymov, T. Ostapchuk, S. Lehenchuk, D. Grytsyshen, G. Marchuk, Analysis of the impact of intangible assets on the companies' market value, Naukovyi Visnyk Natsionalnoho Hirnychoho Universytetu 3 (2020) 164–170. doi:10.33271/nvngu/2020-3/164.
[22] T. Zavalii, O. Vikarchuk, C. Constantinou, Do marketing-related intangible assets affect the company's net income?, Public Policy and Accounting 1 (2022) 3–14. doi:10.26642/ppa-2022-1(4)-3-14.
[23] N. Özkan, R&D spending and financial performance: an investigation in an emerging market, International Journal of Management Economics and Business 18 (2022) 38–58. doi:10.17130/ijmeb.964849.
[24] S. Lehenchuk, M. Tumpach, I. Vyhivska, V. Makarovych, S. Laichuk, The Impact of Innovation on the Profitability of Slovak Pharmaceutical Companies, Marketing and Management of Innovations 2 (2022) 184–296. doi:10.21272/mmi.2022.2-25.
[25] O. Sokil, D. Zvezdov, V. Zhuk, S. Kucherkova, L. Sakhno, Social and environmental costs: The impact of accounting and analytical support on enterprises' sustainable development in Germany and Ukraine, Economic Annals-XXI 181 (2020) 124–136. URL: http://dx.doi.org/10.21003/ea.V181-11.
[26] A. Del Monte, E. Papagni, R&D and the growth of firms: Empirical analysis of a panel of Italian firms, Research Policy 32 (2003) 1003–1014. doi:10.1016/S0048-7333(02)00107-5.

Maximizing customer satisfaction and business profits through Big Data technology in Society 5.0: a crisis-responsive approach for emerging markets*

Piotr Kulyk[1], Viktoriia Hurochkina[1,2], Bohdan Patsai[2], Olena Voronkova[2] and Oksana Hordei[2]

[1]*University of Zielona Góra, 9 Licealna, Zielona Góra, 65-417, Poland*
[2]*State Tax University, 31 Universitetska Str., Irpin, 08200, Ukraine*
[3]*Taras Shevchenko National University of Kyiv, 60 Volodymyrska Str., Kyiv, 01033, Ukraine*

Abstract

Big Data technology is a powerful tool for businesses in emerging markets, especially in times of crisis. It can help them understand their customers better, improve their products and services, and increase their profits. This study aims to use Big Data technology to design personalized loyalty programs that can enhance customer satisfaction and loyalty in the context of Society 5.0, a vision of a human-centered society that integrates physical and digital realms. The study builds on utility theory, firm theory, welfare economics, and Big Data theory to develop a conceptual framework for information-centric loyalty programs. These programs can use Big Data to analyze customer behavior, preferences, income, and mobility in real time and offer customized incentives and discounts. The study also explores the benefits and challenges of using Big Data technology for businesses in emerging markets, as well as its role in crisis management and recovery. The study suggests that Big Data technology is essential for businesses to gain a competitive edge and survive in the dynamic and uncertain environment of Society 5.0.

Keywords

customer satisfaction, business profits, emerging markets, crisis periods emerging markets, welfare theory, needs theory, Big Data technology, loyalty programs, predictive analysis

1. Introduction

In the aftermath of the Coronavirus disease (COVID-19) pandemic [2] and the repercussions of the Russian-Ukrainian war [3], harnessing competitive advantages in the realm of goods and services assumes paramount importance. Informed by foundational principles of economic theory, a pivotal question arises: can the doctrine of utility be effectively employed? This query finds resonance in the proposition put forth by Jules Dupuit [4], who posited that

M3E2-MLPEED 2022: The 10th International Conference on Monitoring, Modeling & Management of Emergent Economy, November 17-18, 2022, Virtual, Online

*Extended and revised version of paper [1] presented at the 10th International Conference on Monitoring, Modeling & Management of Emergent Economy.

p.kulyk@wez.uz.zgora.pl (P. Kulyk); v.hurochkina@g.elearn.uz.zgora.pl (V. Hurochkina); b.pacaj@gmail.com (B. Patsai); Voronkova303@gmail.com (O. Voronkova); ohordei@gmail.com (O. Hordei)
0000-0003-2786-4020 (P. Kulyk); 0000-0001-8869-0189 (V. Hurochkina); 0000-0001-5636-9219 (B. Patsai); 0000-0002-7956-7768 (O. Voronkova); 0000-0001-6938-0548 (O. Hordei)

the same commodity could be sold to distinct customers at disparate prices, irrespective of cost differentials. Furthermore, Dupuit's insights unveil a prerequisite condition for such a strategy – a monopolistic market position that empowers the seller with pricing authority. This monopolistic structure permits discernment between consumer groups of varying affluence, allowing for calibrated pricing based on divergent proclivities. Here, the underpinning of Dupuit's utility theory predominantly rests on the consumer's vantage.

Contemporaneously, the quest for optimizing individual needs and interests was also explored by Dionysius Lardner, a British economist and engineer. In his analysis [5], Lardner delved into maximizing income from the prism of firm theory. A significant facet of Lardner's contributions pertains to the strategic utilization of price competition to heighten profits. His scrutiny of railway tariffs furnished insights into their differential structuring based on distance and cargo characteristics, attributed to variations in elasticity and consumer demand heterogeneity. Notably, Lardner's work underscores the pivotal role of demand elasticity in satiating consumer needs.

The economics of welfare, as elucidated by Pigou [6], furnishes a pivotal framework for comprehending price competition and its multifaceted types. Pigou postulates a nexus between the feasibility of price competition and a set of general conditions, contingent on the non-interdependency of demand prices for distinct units of commodities. To this end, it necessitates a scenario where units of goods transacted in one market cannot be seamlessly transposed to another market, coupled with the restriction that units of demand in one market remain exclusively tethered to their native domain. However, achieving such equilibrium is intricate, demanding comprehensive information encompassing consumer benefits and purchasing potential across domestic and global markets. This daunting task is further compounded by the complexities introduced by analogous products. As the post-COVID-19 and post-conflict landscape unfolds, the global race for end consumers is poised to escalate.

A salient determinant of success resides in positive emergent properties, which pivot on innovation germination and progression, particularly in the realm of consumer engagement and burgeoning economies [7, 8, 9]. The catalysts of such positive emergent properties coalesce around the zenith of human capital development. Pertinently, in the context of developing economies, intricate examinations are undertaken to scrutinize innovative human capital development facets across various domains [10, 11]. The contours of human capital realization within developing economies are further compounded by the far-reaching impacts of COVID-19 and the Russian-Ukrainian war, leading to demographic shifts and internally displaced populations.

2. Literature review

Industry 4.0 will contribute to the emergence of a new Society 5.0. Innovative technologies of Industry 4.0 will contribute to the rapid recovery and overcoming the consequences of the Russian-Ukrainian war.

Fundamental provisions of formation Society 5.0 and implementation of innovative Industry 4.0 technologies are considered in a number of work. Kitsuregawa [12] highlight the questions of how Japan is launching Society 5.0 and the vision for a future smarter society. The work of Aquilani et al. [13] is devoted to the advanced manufacturing solutions, augmented reality, the cloud, and big data in the emergence of a new level of social development. Rahmanto et al. [14] note the potential of huge advantages of big data technology in the emergence of a new level of social development and a breakthrough revolution in people's lives thanks to the use of technologies taking into account the humanitarian aspect.

Foresti et al. [15], Hayashi and Nagahara [16] highlight the role of artificial intelligence in the functioning of automated planning and data analysis with the help of smart programs, smart infrastructure, smart systems, and smart networks.

Ellitan [17] focuses on the lack of HR (human resources) skills and the existing problem of security of communication technologies, and the inability of stakeholders to change, while in society 5.0 there is a clear priority due to the reliable and stable operation of production machines, which in turn leads to the negative consequences of worker losses places through automation. for the rapid adaptation of human capital for the benefit of improving public and business services, achieving a high level of literacy in working with data and its data analysis is an important condition. Simatupang [18] noted that the slow progress of Society 5.0 can be achieved through the development of integrated information technologies in universities and education. De Felice et al. [19] noted that in order to achieve Society 5.0 it is important to manage the transition and identify the enabling factors that integrate Industry 4.0. According to Önday [20], digital transformation creates new values and becomes a pillar of the industrial policy of many countries. Therefore, in Society 5.0, the basis of quality functioning is the achievement of convergence between physical and cyberspace. But it should be noted that the key drivers of the implementation of Industry 4.0 in Society 5.0 will contribute to rapid recovery in the post-war period, new economies will emerge, the only question will be the transfer of technologies for recovery and adaptation at the fastest pace.

3. Methodology

If you determine the level of demand in various market segments and in the markets of various countries, you can set an individual price for each unit of a homogeneous product, which will be equal to the price of its demand. This price is called the reserved price of the buyer. In its pure form, such a pricing policy is difficult to implement. The company does not know the reserved price of each buyer, but also cannot know its level from the buyer, since it is in his interests to reduce its value. It is the lack of information that does not allow the full introduction of perfect price competition and the largest financial effect.

The options (based on the collected data) for setting different prices for certain consignments of goods in accordance with the same demand function are used today. In practice, it often takes the form of various kinds of discounts (depending on the size of purchases, prepaid periods, etc.). In this case, the monopolist increases the volume of sales, and the consumer can achieve certain economies of purchase volume.

Differentiation of buyers into groups with different demand functions and subsequent pricing for each such group occurs separately during market segmentation. Segmentation is usually carried out by gender, age, income level, social status. There is the practice of setting different prices for students, senior citizens, people with disabilities and people of working age. Segmentation of end consumers is being made considering price and non-price ways of increase

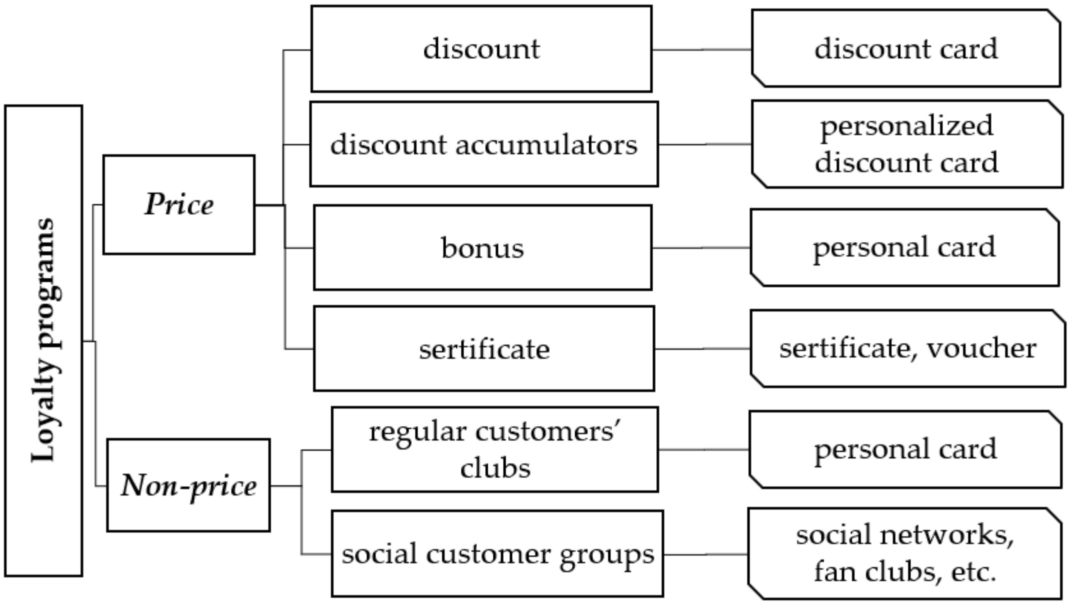


Figure 1: Classification of loyalty programs.

influence on sales (figure 1), which are reflected in loyalty programs.

However, the discount loyalty programs have some disadvantages:

- the ability to saturation and, consequently, decrease the efficiency of use;
- the complexity of how to form a group of supporters as well as the completion of the closure of the current program;
- the remoteness of non-regular customers and the usual price overpricing.

Nowadays discount accumulators and bonus cards are mostly used. Among the reasons that led to a change in the accounting policies of many enterprises there is a possibility of:

- the creation of various offers for various groups of clients;
- provision of discounts in the form of a certificate is an incentive for the client to return to the purchase of well-known goods and services;
- tracking the movement of regular customers and changing their preferences.

Introduction of such loyalty programs became possible thanks to the rapid development of information technologies that are capable to solve new problems. In addition, these cards can significantly reduce the turnover of small bills. But the main feature of these changes is the personalization of discount programs.

Personalization of seller-buyer relationships, using data mining (OLAP technology), allows you to analyze the dependencies of any values contained in the database and respond to the situation quickly. Important information for the seller is not only attracting new customers, but also controlling relationships with regulars. Firstly, the sales increase may be a consequence of a successful advertising company and, secondly, sales decrease for personalized discount cards is a consequence of low level of service, which will lead to a sharp decrease in sales in the medium and long term.

Currently, in order to increase the effectiveness of consumer segmentation the enterprise is trying to group them according to the level of the product value perception. In this case consumers are allocated:

- price-sensitive and thus easily change suppliers;
- sensitive to the quality of goods and services;
- are focused on creating long-term relationships and, as a result, strive to establish long-term partnerships to improve the quality of goods and services.

Internet trade has the greatest relevance during the lockdown. It is devoid of such shortcomings that are characteristic of the real sector of the economy:

- is not strictly connected with the territory of the physical existence of the consumer;
- can be carried out without any territorial restrictions;
- the rapid development of the information society and information growth gave impetus to the development of new methods of its implementation.

In particular the Big Data theory is rapidly developing [21, 22]. The term "Big Data" usually refers to a series of approaches, tools and methods for processing of structured and unstructured large volumes and the different nature data to obtain a consumer acceptable result. The introduction of the term "Big Data" is associated with Clifford Lynch [23] who was an editor of Nature magazine and prepared a series of topical works. Quite often the "triple V" criterion is used to describe "Big Data": volume, velocity, variety. Some leading manufacturers of business intelligence software, such as SAS [24], additionally use two more: variability and complexity. In addition to growing speeds and data varieties, data flows can also be characterized by periodic peaks. Such peak data loads can be difficult to manage. It is worth to note the complexity factor as the most important factor when you are working with Big Data. While increasing the amount of data to variable n, the number of links between them grows in proportion to $n!$ (n factorial). So the problem is not limited only to the processing of large amounts of data but also requires an additional solution to the problem of analyzing connections' $n!$.

To identify a consumer on the Internet data for analysis is needed. The profile of the network is formed not only with the registration data on particular Internet resources but also activity in social networks, forums, blogs and the like. Thus, data reflecting the user is unstructured.

4. Results

Leading corporations have developed platforms for big data business analytics [21]. In particular IBM, creating a full profile from social network data in the Big Data Analytical System, uses all the data that is more or less related to a specific consumer (table 1). At the first stage analysis of the texts takes place, at the second the linking of attributes takes place, at the third formation of statistical models and at the fourth formation of business logic take place.

Table 1
The data structure that is used to form a complete social user profile.

Full social customer profile	Personal characteristics	Identifiers
		Interests
		Social status
	Relationships	Personal
		Business
	Chronological activity	Purchase intention
		Current location
		Feedback on products and services
		Incident
		Loyalty facts
	Goods and interests	Personal relation to goods
		Shopping history
		Recommendations
	Politics	Attitude to power
		Political views
		Perception of reform
	Life events	Personal
		Reactions to events

Economic-mathematical modeling of the socio-economic system based on online Big Data algorithms makes it possible to predict consumer behavior based on the identification of business logic and to form a consumer profile in the decision-making system. This method is traditional, but the selection of characteristic functional features for forecasting efficiency and optimization of Slick-Through-Rate forecasting processes is special in view of machine learning as a tool for economic and mathematical modeling of the management decision-making system.

Taking into account the presented data structure of the full profile of a social network user and the model of Big Data online algorithms, we have the possibility of flexible targeting of the target audience, adaptation of advertising content in accordance with user interests, the possibility of forecasting the effectiveness of advertising and its impact on consumer behavior. In addition, when building a model of Big Data algorithms, it is worth taking into account traffic segmentation and the Real-Time Bidding Exchange RTB auction (corresponding to the business logic of the consumer).

The use of Big Data in e-commerce provides such competitive advantages:

1) customer service: Big Data helps to give the consumer a sense of self-worth because his needs are maximally met by creating a certain connection between him and the brand. This cultivates consumers' loyalty and influence on their emotional level;
2) dynamic and point pricing: analysis of market data allows you to set an attractive price for each specific consumer;
3) personalization: in the process of analyzing consumers' information, personalized solutions are offered that become a competitive advantage for the client;
4) predictive analysis: Big Data allows you to carry out medium-term forecasting in the market and respond accordingly to possible changes in the market environment.

An example of this approach can be an application developed for the clothing brand Free People which provided sales growth of 38 percent [25]. The application allows users to discuss the latest collections, share their photos on Pinterest and Instagram social resources and vote for the best photos. This interaction is an example of the monetization of accumulated data by retailers using social platforms.

Point discounts of Internet commerce can be divided by analogy with traditional commerce into two types depending on the technology that is used. The first type is personalized which provides for mandatory registration on a web resource, the second is not personalized (does not require registration). The first option of a point discount is for a price offer based on customer data, a history of web surfing (viewing products on a store page) and purchase history. Retailers often use social media accounts to register. It simplifies the registration procedure and gains access to user data. This significantly increases the amount of data to be analyzed.

Based on the data (table 1) on using Big Data, a consumer profile is formed and its segment affiliation is determined. In the future the client is offered an individual price offer. The price that is offered is minimal in order for the fact of purchase. In addition, goods are offered in accordance with the target audience. In other words, an individual approach to proposals is formed based on the analytical processing of unstructured data.

For convenience we have built EPC diagram [26], which is often used to describe the workflow in ArisExpress environment (figure 2). If the visitor is not a consumer of goods and services, HTTP-cookie analysis of the web page is carried out that allow carrying out authentication, storage of personal user preferences and settings, session state tracking of user access, maintain user statistics.

It is also possible when there is not enough data to determine the profile of the visitor. This may be due to both the low activity of the Internet user and his conscious reluctance to "external tracking". One such way is to use an anonymous session. In this case the basic offers are determined by the system.

For machine learning target audience targeting [27], we use the Datch approach, taking into account the social network user profile, to build a model of online Big Data algorithms. The Datch approach is based on two-level testing of Big Data algorithms: training dataset and test dataset. The condition of the model is the constancy of the data of the decision-making system over time. At the same time, the dynamism of the system and the resonance of news on the website can become an emergent property of the socio-economic system, which will contribute to a further change in the trend.

The model of Big Data algorithms for the task of predicting CTR is based on the systematization of the modeling process by stages and on a certain set of parameters of the data structure of the complete profile of a social network user.

$$\begin{aligned} W_{sc_{p+1}} = argmin \sum_{i=1}^{t-1} v(w_p, w_r, w_{ch}, w_{gi}, w_{pol}\ , w_l) \\ + R(w_p,\ w_r, w_{ch},\ w_{gi},\ w_{pol}\ , w_l) \end{aligned} \quad (1)$$

where:

$W_{sc_{p+1}}$ – function social customer profile;

Figure 2: Structurally Logical Pricing Scheme in an EPC Chart.

$v(w_p)$ – loss function for optimization Personal characteristics (Identifiers, Interests, Social status);

$v(w_r)$ – loss function for optimization Relationships (Personal, Business);

$v(w_{ch})$ – loss function for optimization Chronological activity (Purchase intention, Current location, Feedback on products and services, Incident, Loyalty Facts);

$v(w_{gi})$ – loss function for optimization Goods and interests (Personal relation to goods, Shopping history, Recommendations);

$v(w_{pol})$ – loss function for optimization Politics (Attitude to power, Political views, Perception of reform);

$v(w_l)$ – loss function for optimization Life events (Personal, Reactions to events).

$R(w_p)$ – regularization function Personal characteristics (Identifiers, Interests, Social status);

$R(w_r)$ – regularization function Relationships (Personal, Business);

$R(w_{ch})$ – regularization function Chronological activity (Purchase intention, Current location, Feedback on products and services, Incident, Loyalty Facts);

$R(w_{gi})$ – regularization function Goods and interests (Personal relation to goods, Shopping history, Recommendations);

$R(w_{pol})$ – regularization function Politics (Attitude to power, Political views, Perception of reform);

$R(w_l)$ – regularization function Life events (Personal, Reactions to events).

The loss function for optimizing the profile characteristics of a social network user will have the following form:

$$v_t\left(w_p,\ w_r, w_{ch},\ w_{gi},\ w_{pol}\ , w_l\right)\ = \|w - x_t\|\ ^2 \tag{2}$$

Under the conditions of a linear loss function in order to optimize the characteristics of the social network user profile, the formula will have the following form:

$$v_t(w_p,\ w_r, w_{ch},\ w_{gi},\ w_{pol}\ , w_l)\ = \langle w, x_t\rangle \tag{3}$$

Under conditions of activation of emergent properties in the socio-economic system, such as dynamic system changes or trend changes under the influence of high-profile news on the site, which contribute to the manifestation of binary dependence at the bifurcation point, the function will have the following form:

$$v_t(w_p, w_r, w_{ch}, w_{gi}, w_{pol}, w_l) = \left(\sigma\left((w_p,\ w_r, w_{ch},\ w_{gi},\ w_{pol}\ , w_l)\, x_t\right) - y_t\right)x_t \tag{4}$$

σ – sigmoidal function:

$$\sigma(\alpha) = \frac{1}{1+e^a} \tag{5}$$

With the activation of emergent properties in the socio-economic system, the regularization function will have the following form:

$$R(w_p,\ w_r, w_{ch},\ w_{gi},\ w_{pol}\ , w_l)\ = \frac{1}{2_n}\|w\|\ ^2 \tag{6}$$

Under the conditions of if $\eta > 0$, then the iteration of the machine learning algorithm will include a stepwise gradient descent algorithm and will look like:

$$w_{sc_{p+1}} = -\eta \sum_{i=1}^{p} z_i = Wsc_p - \eta z_i = Wsc_p - \nabla v_t(w_p, w_r, w_{ch}, w_{gi}, w_{pol}, w_l) \tag{7}$$

The resulting formula for optimizing management decisions, taking into account the parameters of the data structure of the full profile of a social network user, will look like this:

$$w_{p,i} = \begin{cases} 0 & |x_i| \le \varepsilon_1 \\ -\left(\frac{\beta+\sqrt{n_i}}{\alpha}\right)(x_i - sign(x_i)\varepsilon_1) & |x_i| > \varepsilon_1 \end{cases} \tag{8}$$

where *x* and *n* iteration parameters, $\varepsilon_1,\ \varepsilon_2$ are regularization intensity parameters according to the selected type and α, β – are input parameters characterizing the learning rate.

Since, based on the above, in order to achieve the optimum at each step of the algorithm execution, the optimal decision is made and the previous ones are not foreseen, then this model belongs to the Greedy algorithm. A characteristic feature of these algorithms is relative simplicity and speed of execution.

This technique of point discount has been actively developing over the past three years. One of the first companies that offered this service was Freshplum whose founder was Sam Odai. Later Freshplum joined the TellApart company [24], which operates in the market of services for online stores. Moreover, the algorithm for potential customers' selection of this company uses a number of "non-standard" indicators such as: place of residence (city center or outskirts), weather, etc. This allows you to increase the likelihood of making a purchase up to 36 percent [28].

For the first time the analysis of differential pricing in online stores was conducted by the The Wall Street Journal. The editors conducted a study [29] of pricing in 200 online stores.

The economic situation in the world is extremely dependent on the geopolitical risks that can now be observed (for example the corona virus pandemic and the consequences of the Russian-Ukrainian war). Therefore, the widespread use of Big Data concept may increase the profitability of enterprises. The use of Big Data methods will become an additional source of budget revenues after taxation. This will maximally satisfy the needs of consumers whose incomes have recently been declining due to devaluation and inflationary processes. In order to increase competitiveness of European goods and services markets the use of big data is a mandatory requirement of our time.

5. Conclusions

In an era characterized by global geopolitical uncertainties, underscored by events such as the COVID-19 pandemic and the repercussions of the Russian-Ukrainian war, the economic trajectory of nations is intricately interwoven with the contours of geopolitical risks. Amidst this backdrop, the strategic assimilation of foundational principles of Big Data emerges as an indispensable tool for enhancing enterprise profitability. The application of Big Data methodologies assumes the mantle of an auxiliary revenue stream, akin to taxation, endowed with the potential to assuage the financial strains borne by consumers grappling with income diminutions catalyzed by devaluation and inflationary forces. Additionally, as European markets for commodities and services necessitate a heightened competitive edge, the incorporation of Big Data not only becomes a compelling imperative but a quintessential mandate of our epoch.

At the nexus of this pursuit lies the economic-mathematical model, a potent tool that drives optimization at each algorithmic juncture. This model, emulating the Greedy algorithm's attributes of simplicity and swift execution, engenders optimal decision-making. With Big Data as its cornerstone, this methodology fosters an environment where the quintessence of profitability and competitiveness converge to navigate the complexities of today's economic landscape.

Acknowledgments

The work was carried out within the framework of the program to support scientists from Ukraine during the implementation of the project “Mechanism to strengthen the social responsibility of refugees and people fleeing the Russian armed conflict on the territory of Ukraine” on funding from the Polish Academy of Sciences and the National Academy of Sciences of the United States and University of Zielona Góra.

References

[1] P. Kulyk, V. Hurochkina, B. Patsai, O. Voronkova, O. Hordei, Intensifying Use of Big Data for Emerging Markets in Society 5.0, in: S. Semerikov, V. Soloviev, A. Matviychuk, V. Kobets, L. Kibalnyk, H. Danylchuk, A. Kiv (Eds.), Proceedings of 10th International Conference on Monitoring, Modeling & Management of Emergent Economy - M3E2, INSTICC, SciTePress, 2023, pp. 63–70. doi:10.5220/0011931300003432.

[2] S. S. Iyer, L. Gernal, R. Subramanian, A. Mehrotra, Impact of digital disruption influencing business continuity in UAE higher education, Educational Technology Quarterly 2023 (2023) 18–57. doi:10.55056/etq.29.

[3] V. I. Kovalchuk, S. V. Maslich, L. H. Movchan, Digitalization of vocational education under crisis conditions, Educational Technology Quarterly 2023 (2023) 1–17. doi:10.55056/etq.49.

[4] G. Numa, The monetary economics of Jules Dupuit, The European Journal of the History of Economic Thought 23 (2016) 453–477. doi:10.1080/09672567.2014.951673.

[5] D. L. Hooks, Monopoly Price Discrimination in 1850: Dionysius Lardner, History of Political Economy 3 (1971) 208–223. URL: https://EconPapers.repec.org/RePEc:hop:hopeec:v:3:y:1971:i:1:p:208-223.

[6] A. C. Pigou, The Economics of Welfare, Macmillan & Co., London, 1920. URL: https://oll.libertyfund.org/title/pigou-the-economics-of-welfare.

[7] V. Dzhedzhula, V. Hurochkina, I. Yepifanova, A. Telnov, Fuzzy Technologies for Modeling Social Capital in the Emergent Economy, WSEAS Transactions on Business and Economics 19 (2022) 915–923. doi:10.37394/23207.2022.19.80.

[8] V. Hurochkina, S. Reshmidilova, L. Bohatchyk, A. Telnov, L. Skorobogata, N. Riabinina, Modeling Effectiveness of Financial Support for the Social Capital Development in Economic Emergence, WSEAS Transactions on Environment and Development 17 (2021) 262–270. doi:10.37394/232015.2021.17.27.

[9] O. Hordei, B. Patsai, V. Hurochkina, O. Ovdiienko, R. Mishchenko, Optimization of the Investment Portfolio in the Environment of Table Processor MS Excel, Studies of Applied Economics 39 (2021). doi:10.25115/eea.v39i5.4983.

[10] B. Czyżewski, A. Sapa, P. Kułyk, Human Capital and Eco-Contractual Governance in Small Farms in Poland: Simultaneous Confirmatory Factor Analysis with Ordinal Variables, Agriculture 11 (2021) 46. doi:10.3390/agriculture11010046.

[11] B. Czyżewski, A. Poczta-Wajda, P. Kułyk, J. Drozdz, Small farm as sustainable nexus of contracts: understanding the role of human capital and policy based on evidence

from Poland, Environment, Development and Sustainability (2022). doi:10.1007/s10668-022-02485-2.
[12] M. Kitsuregawa, Transformational Role of Big Data in Society 5.0, in: 2018 IEEE International Conference on Big Data (Big Data), 2018, pp. 3–3. doi:10.1109/BigData.2018.8621989.
[13] B. Aquilani, M. Piccarozzi, T. Abbate, A. Codini, The Role of Open Innovation and Value Co-creation in the Challenging Transition from Industry 4.0 to Society 5.0: Toward a Theoretical Framework, Sustainability 12 (2020) 8943. doi:10.3390/su12218943.
[14] F. Rahmanto, U. Pribadi, A. Priyanto, Big Data: What are the Implications for Public Sector Policy in Society 5.0 Era?, IOP Conference Series: Earth and Environmental Science 717 (2021) 012009. doi:10.1088/1755-1315/717/1/012009.
[15] R. Foresti, S. Rossi, M. Magnani, C. Guarino Lo Bianco, N. Delmonte, Smart Society and Artificial Intelligence: Big Data Scheduling and the Global Standard Method Applied to Smart Maintenance, Engineering 6 (2020) 835–846. doi:10.1016/j.eng.2019.11.014.
[16] N. Hayashi, M. Nagahara, Distributed Sparse Modeling for Society 5.0: Big Data Analysis over Multiagent Networks, IEICE ESS Fundamentals Review 13 (2019) 95–107. doi:10.1587/essfr.13.2_95.
[17] L. Ellitan, Competing in the Era of Industrial Revolution 4.0 and Society 5.0, Jurnal Maksipreneur: Manajemen, Koperasi, dan Entrepreneurship 10 (2020) 1–12. doi:10.30588/jmp.v10i1.657.
[18] A. Simatupang, Digitalisasi dan Internasionalisasi Pendidikan Tinggi Dalam Pembentukan Society 5.0 dan Industri 5.0, in: L. Sintha, T. Guswantoro, F. Tobing, A. A. Purnamasari, S. Putra (Eds.), Digitalisasi dan Internasionalisasi Menuju APT Unggul dan UKI Hebat, UKI Press, 2020, p. 217–218. URL: http://surl.li/dzmxm.
[19] F. De Felice, M. Travaglioni, A. Petrillo, Innovation Trajectories for a Society 5.0, Data 6 (2021) 115. doi:10.3390/data6110115.
[20] Ö. Önday, Japan's Society 5.0: Going Beyond Industry 4.0, Business and Economics Journal 10 (2019) 1000389. URL: https://www.academia.edu/39149435/Japan_s_Society_5_0_Going_Beyond_Industry_4_0.
[21] Market Research Future, Healthcare Big Data Analytics Market Size, Growth | Industry Trends, 2030, 2022. URL: http://surl.li/dzmwt.
[22] Z. Seidametova, Z. Abduramanov, G. Seydametov, Hackathons in computer science education: monitoring and evaluation of programming projects, Educational Technology Quarterly 2022 (2022) 20–34. URL: https://acnsci.org/journal/index.php/etq/article/view/5. doi:10.55056/etq.5.
[23] C. Lynch, How do your data grow?, Nature 455 (2008) 28–29. doi:10.1038/455028a.
[24] SAS Institute Inc, Big Data Analytics: What it is and why it matters, 2022. URL: https://www.sas.com/en_us/insights/analytics/big-data-analytics.html.
[25] L. Dishman, How Free People is Using Big Data and Social Commerce for Bigger Sales, 2013. URL: http://surl.li/dzmwk.
[26] Software AG, Event-driven process chain (EPC), 2022. URL: https://www.ariscommunity.com/event-driven-process-chain.
[27] V. Derbentsev, A. Matviychuk, V. N. Soloviev, 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, Springer, Singapore, 2020, pp. 211–231. doi:10.1007/978-981-15-4498-9_12.

[28] A. Tanner, Different Customers, Different Prices, Thanks To Big Data, 2014. URL: https://cutt.ly/q018Ugn.

[29] J. Valentino-DeVries, J. Singer-Vine, A. Soltani, Websites Vary Prices, Deals Based on Users' Information, The Wall Street Journal (2012). URL: https://www.wsj.com/articles/SB10001424127887323777204578189391813881534.

A flexible machine learning model for optimizing organizational capital development strategies and resource allocation*

Vasyl Porokhnya[1], Vladyslav Penev[1], Roman Ivanov[2] and Volodymyr Kravchenko[1]

[1]*Classical private university, 70B Zhukovsky Str., Zaporizhzhia, 69061, Ukraine*
[2]*Oles Honchar Dnipropetrovsk National University, 72 Gagarin Ave., Dnipro, 49000, Ukraine*

Abstract

In this paper, we present a comprehensive approach to optimizing organizational capital development strategies through the application of a flexible machine learning model. We advance from initial conceptualization to the implementation stage, employing Q-leaning to enhance the selection process of the most effective organizational capital development strategies within the framework of intellectual capital. Our model aims to improve decision-making reliability by employing data-driven techniques. In the final phase of our study, we simulate various alternative strategies for organizational capital development using machine learning techniques. This simulation framework streamlines the process of exploring different strategic options, enabling more informed management decisions. To enhance the machine learning process, we introduce coefficients that influence decision-making, resulting in more accurate and effective outcomes. Our findings emphasize that innovative information potential is a key facet of successful organizational capital development strategies. Furthermore, our approach demonstrates the potency of integrating intellectual capital management mechanisms with other capital types.

Keywords

machine learning, organizational capital, strategy optimization, Q-leaning, intellectual capital management

1. Introduction

In the contemporary economy, the significance of intellectual capital as a potent driver of effectiveness is well substantiated. The notion of intellectual capital surpasses the confines of intellectual property and intangible assets, while closely aligning with the concept of intangible capital, a term explored in economic theory and econometrics since the 1970s [2].

The work by Daum [3] provided a definition of intangible capital rooted in the interconnectedness of structured knowledge and competencies, which bear the potential to foster development and value creation.

Leontiev [4] conceptualized intellectual capital as encompassing the value of an enterprise's

M3E2-MLPEED 2022: The 10th International Conference on Monitoring, Modeling & Management of Emergent Economy, November 17-18, 2022, Virtual, Online

*Extended and revised version of paper [1] presented at the 10th International Conference on Monitoring, Modeling & Management of Emergent Economy.

vprhnp76@gmail.com (V. Porokhnya); penev.vladislav@gmail.com (V. Penev)
0000-0003-0820-8749 (V. Porokhnya); 0000-0003-2086-5004 (R. Ivanov)

collective intellectual assets, encompassing intellectual property, innate and acquired cognitive abilities of its personnel, as well as amassed knowledge repositories and beneficial collaborations with other entities.

In the perspective of Roos et al. [5], intellectual capital comprises all non-monetary, intangible resources contributing to an organization's value generation, which it possesses full or partial control over.

Intellectual capital, by nature, poses challenges in terms of assessment and quantification due to the intricacies of delineating its constituent elements. Yet, it can be deconstructed into various capitals: human capital, organizational capital, and customer or consumer capital.

Each facet of intellectual capital can be further delineated as follows:

1. Human capital, embodying the value contributed by employees through their skill sets, expertise, and knowledge. This form of capital resides within individuals and can be attributed to an organization.
2. Organizational capital comprises: technological capital; branding capital; business culture capital; economic value added (EVA) capital; information potential innovation strategy capital. The evaluation criteria encompass manufacturability, productivity, innovativeness, cooperativeness, adaptability, and efficiency.
3. Customer equity, encompassing elements such as customer relationships, supplier relationships, trademarks and trade names (whose value stems solely from customer relationships), licenses, and franchises.

This multi-faceted nature of intellectual capital engenders complexity in its assessment. Methods for measuring intellectual capital encompass four primary categories, as posited by Sveiby [6]: Direct Intellectual Capital Methods; Market Capitalization Methods; Return on Assets Methods; Scorecard Methods. However, each category has its limitations that necessitate integration with machine learning techniques. By adopting a unified machine learning algorithm, a comprehensive mathematical model can be formulated, enhancing the precision of estimates across all constituent elements of intellectual capital.

2. Results

If we consider the structure of Organizational Capital (OC) as a set of its qualities and properties, their ratios, which directly affect labor productivity, which increases the income for personnel, the company as a whole, society, and the nation, then there is an opportunity to cover all possible options for its evaluation.

1. Assessment of the level of manufacturability

Let's move on to the assessment of the properties of the components of manufacturability capital. We will use its structure, which consists in determining the share a_t^k of the t-th property in the formation of the k-th type of components of manufacturability capital (k_k^t), which allows us to establish the probable level of the k-th type of manufacturability capital:

$$KT_{t^k} = \sum_{k=1}^{n_{ip}} k_{k^t} * a_{k^t}, \tag{1}$$

where

KT_{t^k} – technological capital;

k_{k^t} – exploitation and repair manufacturability of the structure to k item for t-th indicator (materials, energy, labor, compatibility, etc.);

a_{k^t} – volatility of the injection of the t-th indicator for manufacturability of k item.

2. Capital assessment of business culture

$$CC_{k^t} = \sum_{k=1}^{n_{ip}} c_{k^t} * b_{k^t} \tag{2}$$

where

CC_{k^t} is the capital of business culture;

c_{k^t} – organizational and corporate culture of a certain business model of doing business according to the t-th indicator (liberty and democracy, monoactivity of the business culture type; polyactivity of the business culture type; reactivity of the business culture type, etc.);

b_{k^t} – the importance of the impact of the t-th indicator on the cultural capital of the k-th business model of doing business.

3. The efficiency capital of added economic value

The productivity of the production process has a significant range of properties, the characteristic features of which are formed and reflected by a significant network of indicators that have branched relationships of quantitative and qualitative capital assessment of performance. Among the important features of performance, the following should be noted:

- Activation of human heuristic abilities and structuring of discovered knowledge and verification according to the criterion of objectivity;
- Orderliness of the communication process for the exchange of information flows, emotions, social and individual values, economic interests;
- Formation and growth of the fundamental and market value of the enterprise as a criterion of performance.
- Identification and elimination of dysfunctions in enterprise management, which arise due to a malfunction.

Capital assessment of efficiency of added economic value. Performance is assessed as the level of intellectual leverage (LIL) and is calculated according to the formula:

$$LIL = \frac{\triangle EVA\%}{\triangle NOPLAT\%} \tag{3}$$

where:

$\triangle EVA\%$ is the rate of profit growth;

$\triangle NOPLAT\%$ is the growth rate of economic added value.

LIL – the degree of sensitivity of profit to changes in economic added value.

The level of intellectual leverage shows: how many times the growth rate of economic added value exceeds the growth rate of profit. This excess is provided with the help of the effect of intellectual leverage, one of the components of which is its differential (the ratio of the involved intellectual capital to its own).

4. The capital of the strategy of attracting innovations of the information potential

The information capital of the strategy or the capital of the strategy of attracting innovations of the information potential determines the trajectory of intellectual capital and the direction of the implementation of the proposed strategy within the framework of the implementation of innovations of the information potential, which is aimed at increasing the value of capital and depends on the speed of updating this strategy. Informational capital and its potential act as investment capital to maximize the value of intellectual capital:

$$\left(\frac{\sum_{i=1}^{k} EVA}{ROI_{opt} - WACC} - CAPITAL\right) \rightarrow max \tag{4}$$

where

ROI_{opt} is the economic profitability of intellectual capital;

$WACC$ – weighted average interest rate of the involved intellectual capital;

$CAPITAL$ – the capital of the strategy of attracting innovations of the information potential.

5. Capital of turning knowledge into a result

The capital of the transformation of knowledge into a result declares the path of transformations from an idea to the formalization of knowledge in official documents and its structuring for communicative use [7]. Therefore, its components are the following indicators that reflect the characteristic properties of transformations: an idea as a creative and spiritual message, and the level of formalization of knowledge in official documents.

An idea has its own depth of penetration into the macro or micro world [5]. Based on Einstein's thesis that the development of society requires the improvement of everyday thinking, it is appropriate to consider an idea-concept as a complex of properties and relationships that determine the characteristics of the image of the object of research. we can establish a connection between intellectual capital (figuratively speaking, the mass of intellectual substance that is at rest or in motion, that is, in its use) and the strategy of interaction of processes in an economic object and its results. The question arises, does the strategy have energy? It is known that the strategy has different value, that is, weight. Suppose that, like any economic potential, it has potential energy, and when the process of its realization takes place, it also has kinetic energy. That is, strategy is the energy of capital that goes to the realization of an idea-concept. Therefore, it can have its own dimension. Strategy, like any energy, consists of the energy of rest and the momentum of intellectual capital. As the speed of this impulse, we will take the speed of the generation of an idea-concept in the direction predetermined by the strategy. To measure images-properties, that is, the amount of intellectual substance, a unit is introduced, – image.

Any image of intellectual substance contains the same number of images-properties that reflect the properties of the object of the real world. For example, the number of images-properties that characterize a person is a constant value, a number that can be established experimentally, as Avogadro's number was established at one time (the principle of equivalence in nature). But each person has a different number of images-relationships characterizing his intellectual capital. This value of images-relationships, corresponding to intellectual capital, will be assigned the unit of measurement – intel. Intel measures the level (mass) of intellectual capital of a person, enterprise, state.

The definition of images-properties is a consequence of the same type of process properties during the realization of an idea-concept in time, which contain a certain number of these images in one unit. We denote the number of images-properties by N_{img}:

$$N_{img} = \frac{100}{image} - const \tag{5}$$

From here we can determine the amount of the level (mass) of the intellectual capital of the economic system, which corresponds to the capital of transforming knowledge into a result:

$$ic = \frac{N}{N_{img}} M_{ic} \tag{6}$$

where

N – the number of images-properties, respectively, ideas-concepts,

M_{ic} - the intellectual mass of image-properties per image-property for a specific phenomenon, intel / image.

The level of an idea-concept can be represented in four quantitative measurements with the introduction of a unit of measurement – *id*, which contains a certain integral number of images-objects that characterize the properties of this very idea-concept using established criteria:

- Elementary level (household, cognitive, which does not require the formation of new knowledge), where $id = 1$.
- The technological level associated with the emergence of new technologies, etc., where $id = 1000 = 1K$.
- Conceptual level containing new knowledge and discoveries, where $id = 1000000 = 1M = 1000K$.
- The planetary level is determined by the depth of penetration of human activity into the macro and micro world, where $id = 1000000000 = 1G = 1000M = 1000000K$.

$InfConvert_k^t$ – informativeness as a measure of usefulness. The level of structuring of knowledge of special and general scientific terms and its verification according to the criterion of objectivity of the k-th type of the indicator of capital transformations according to the t-th component of this indicator.

$InfCap_k^t = InfConvert_k^t / TotalExp$ – the level of orderliness of the communication process for the exchange of information flows, emotions, social and individual values, economic interests of the k-th type of the indicator of capital transformations according to the t-th component of this indicator.

Evaluation of the capital of the transformation of knowledge into a result

$$CP_k^t = \sum_{k=1}^{n_{ip}} (ic_k^t + InfConvert_k^t + InfCap_k^t) d_k^t \tag{7}$$

where

CP_k^t – the capital of transforming knowledge into a result;

ic_k^t – the capital level of the transformation of knowledge into the result of the k-th type of the indicator of capital transformations according to the t-th component of this indicator;

d_k^t – the weight of the influence of the k-th indicator of transformations on the capital of the transformation of knowledge into a result according to the t-th component of this indicator of transformations.

For a preliminary analysis of the capital criteria, their importance, influence on the choice of the best alternative for the development of the properties of organizational capital, we will use the method of hierarchical comparisons when evaluating the level of priorities of alternatives, the results of which are shown in the table 1.

Table 1
Influence of criteria on a choice of alternatives (properties) of improvement of the level of capital.

Criteria	Properties				
	Intellectual	Communicative	Strategic	Cognitive	Innovative
Branding capital	0.14	0.12	0.14	0.1	0.09
Technology capital	0.12	0.14	0.13	0.14	0.1
Capital efficiency of added economic value	0.11	0.12	0.11	0.13	0.12
Capital of business culture	0.11	0.12	0.13	0.11	0.12
The capital of the strategy of attracting innovations of the information potential	0.1	0.12	0.13	0.12	0.1
General approach	0.11	0.12	0.13	0.115	0.11

The structure of OK is primarily related to branding capital, which is the main relative indicator of the company's attractiveness on the market and to some extent attests to the fate of the firm's market capital, which is adjusted to its organizational, i.e., intellectual capital.

The relevance of the use of machine learning in the field of economics [8, 9, 10, 11] allows us to consider many aspects of the strategy for the development of organizational capital and ways to optimize the cost of resources for its development in different ways. Learning to find the most optimal and less resource-intensive way of developing organizational capital can be presented as a continuous cycle that will end only after the specified conditions are reached (figure 1).

In the reinforcement learning algorithm, the agent's actions are directed to the steps to achieve success with a reward estimate. After $\triangle t$ steps into the next step, the human capital will decide some next step. The weight for this step is calculated as $\gamma^{\triangle t}$, where γ is the discount factor, which can take a value from 0 and 1 ($0 \leq \gamma \leq 1$) and has the effect of evaluating actions that are aimed at achieving the human capital goal. γ can be called the level of success in achieving the desired state by human capital, when the investment data changes at the $\triangle t$ step.

Thus, we can conclude that a function is required that will determine the quality of combinations of the state of human capital and the action aimed at it:

$$Q \div S \times A \rightarrow R. \tag{8}$$

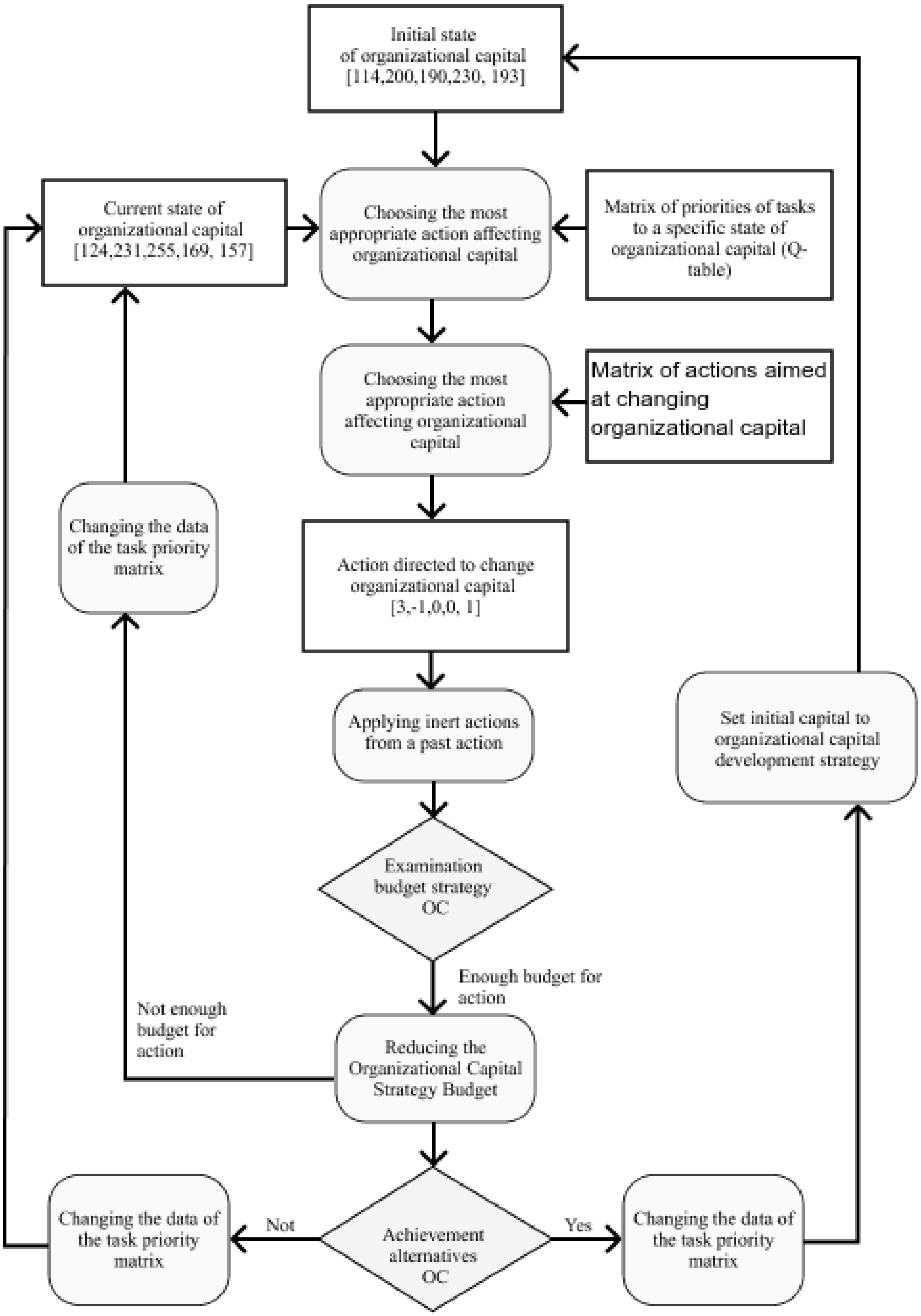


Figure 1: Machine learning of alternative development of human capital of the enterprise.

At the beginning of training, Q is initialized, possibly with an arbitrary fixed value – 0. After initialization, at each moment of time t, the agent selects an action, observes a reward, enters a new state (that may depend on both the previous state and the selected action), and Q is updated. The core of the algorithm is a Bellman [12] equation as a simple value iteration update, using the weighted average of the old value and the new information [13]:

$$Q^{new}(s_t, a_t) \leftarrow Q(s_t, a_t) + \alpha \times (r_t + \gamma \times maxQ(s_{t+1}, a) - Q(s_t, a_t)), \tag{9}$$

where r_t is the reward received when moving from the state S_t to the state S_{t+1}, and $0 < \alpha \leq 1$;

Note that $S^{new}(s_t, \alpha_t)$ is the sum of three factors:

$(1 - \alpha)Q(s_t, \alpha_t)$: the current value weighted by the learning rate. Values of the learning rate near to 1 made faster the changes in Q;

αr_t: the reward $r_t = r(s_t, a_t)$ to obtain if action a_t is taken when in state s_t (weighted by learning rate);

$\alpha\gamma \max Q(s_{t+1}, \alpha)$: the maximum reward that can be obtained from state s_{t+1} (weighted by learning rate and discount factor).

Each action has its own parameters, and system changes can be limited by parameters that can be correlated with the required resource costs to apply the action chosen by machine learning. Thus, each iteration of training implies two possible effects:

1. Changes in the coefficient of effectiveness of the action, depending on the state that the system acquires as a result of the application of the action.
2. Return of the iteration to the initial state due to non-compliance with the specified restrictions for machine learning.

For the application of Q-Learning, the following parameters were selected:

- Impact on the Intellectual Capital criteria
- Time spent in days
- Resource costs equivalent to monetary units
- The coefficient of the complexity of the action
- Risk ratio of failure to take action
- Inert influence on the system
- Coefficient of possibility of inert influence on the system

Each action parameter is used in the calculation of the effectiveness of the action taken at each training step. Applied properties of actions can be represented as a table of actions, which is presented in figure 2.

Thus, at each iteration, the system calculates a promising system that has already been acted upon and recalculates the result of intellectual capital with new parameters.

Thus, we can say that the calculation of the effectiveness of the action is carried out according to the following formula:

$$AE = IK_{t+n}, \tag{10}$$

where

AE – action efficiency;

	IK									PIK					
Action	Branding Capital	Technology Capital	Value Added Efficiency Capital	Business Culture Capital	Implementation capital innovation information capacity	T	RE	WI	RoD	Branding Capital	Technology Capital	Value Added Efficiency Capital	Business Culture Capital	Implementation capital innovation information capacity	PP
1	1	3	0	1	2	24	40	0.3	0.3	0	0	0	0	1	0.3
2	0	0	1	2	0	12	120	0.1	0.4	0	1	1	1	0	0.1
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
23	T	T	T	T	T	40	240	0.3	0.12	1	1	1	1	0	0.5

Figure 2: Action properties used in machine learning with resource cost parameters.

IK – the cost of intellectual capital;
IK_{t+n} — the cost of intellectual capital after applying the action.
So the value of AE will be rewards for moving to the next machine learning state.

$$Q^{new}(s_t, a_t) \leftarrow Q(s_t, a_t) + \alpha \times (AE + \gamma \times maxQ(s_{t+1}, a) - Q(s_t, a_t)), \tag{11}$$

However, each action additionally has a time cost parameter for performing this action, which can optionally be included in the formula. For greater accuracy of calculations, you can use hours, days, months or quarters. In this case, integer values of days were used.

Thus, the new formula for calculating efficiency can be represented as follows:

$$AE = IK_{t+n} * T, \tag{12}$$

where AE – action efficiency, T is the time spent on applying the action

Also, an optional parameter can be resource costs, which are presented in monetary terms. To simplify the loads and quick calculations, all action parameters can be divided by a certain coefficient Mk. In this case, $Mk = 1000$.

Thus, if Action 1 has a resource cost (FE) of 1300000, then the resources spent can be represented as RE and calculated by the formula:

$$RE = \frac{FE}{Mk} * T. \tag{13}$$

Taking into account resource costs, the action efficiency formula will look like this:

$$AE = \frac{IK_{t+n}}{RE} * T. \tag{14}$$

The calculation of resource costs can also include the coefficient of complexity of performing an action (WI), which can be represented by a value in the range from 0.1 to 1.0. Thus, now the resource costs can be represented as:

$$RE = \frac{FE}{Mk} * T * WI. \tag{15}$$

Also, given the individuality of the systems to which actions can be applied, it is worth considering the risks of not performing an action (*RoD*) or its success in execution.

The risk of investing in organizational capital is the possibility that the accumulated organizational capital will not bring the expected return, will not be in demand in the market, or will not bring the expected return. This value can be represented as a range from 0 to 1. A low value of this coefficient means a low level of success of the action and its high risks. Given the risk ratio, the formula for the effectiveness of action can be represented as follows:

$$RE = RoD\frac{IK_{t+n}}{RE}. \tag{16}$$

The relationship of all parameters of intellectual capital does not exclude the influence of the development of some parameters on the possibility of developing other parameters as a result of these actions.

Thus, each action has the values of the inert development of intellectual capital and the coefficient of the possibility of this development.

Given these parameters, the formula for the effectiveness of actions can be represented as follows:

$$RE = ROD\frac{IK_{t+n}}{RE} + PIK * PP * RoD, \tag{17}$$

where PIK is the value of the possible inert development of intellectual capital, PP is the probability coefficient of the development of intellectual capital.

Thus, each iteration of training affects the value of intellectual capital by changing the values of its parameters. However, it is the efficiency values of the action that are written to the state table, not the cost of capital. Having an unlimited resource of investments, achieving the desired value of the cost of intellectual capital had a large set of action algorithms, but given the parameters of each of the actions, machine learning will find the most optimal algorithm for this system.

The development of Intellectual capital occurs with the choice of an alternative to which the capital must approach as a result of learning.

For more effective training and achievement of the most favorable conditions for achieving the desired alternative, development alternatives were introduced. Development alternatives are coefficients for each of the parameters of actions that affect the state of capital. Using the hierarchy analysis method, the following coefficients were introduced (table 2).

Table 2
Alternatives of the development method for managing the choice of effective action.

	Accelerated	Safe	Risky	Budgetary	Effective
IK	0.21	0.26	0.31	0.2	0.32
T	0.23	0.12	0.2	0.08	0.13
FE	0.12	0.12	0.12	0.09	0.12
WI	0.12	0.16	0.13	0.32	0.1
RoD	0.1	0.07	0.08	0.09	0.13
PIK	0.1	0.19	0.1	0.08	0.08
PP	0.12	0.07	0.06	0.14	0.12

For this study, a risky alternative of the method of developing capital for machine learning was chosen.

Thus, each iteration of learning and applying actions to the system will affect the state table and calculate its new values according to the following formula:

$$AE = (RoD\ a_5)\frac{IK_{t+n}\ a_1}{(FE\ a_3)\ (T\ a_2)} + (RoD\ a_5)\ (PIK\ a_6)\ (PP\ a_7) \tag{18}$$

After carrying out the calculations with the initial data, the results describing the strategy for investing in organizational capital shown in table 3.

Table 3
Factor of importance of action properties for learning.

IK	T	FE	WI	RoD	PIK	PP
a_1	a_2	a_3	a_4	a_5	a_6	a_7
0.31	0.2	0.12	0.13	0.08	0.1	0.06

It should be noted that the coefficients of capital alternatives and development alternatives affect value preferences and spending.

The first stages of training provide impressive indicators of cost optimization for investment in organizational capital. With an increase in training cycles, obtaining a better result becomes more rare.

The data in the table 4 and in the figure 3 show optimization costs of developing organizational capital to achieve the cost of organizational capital, taking into account the chosen alternative. It can be concluded that in order to achieve the best results, it is necessary to conduct a sufficient number of training cycles.

Thus, after each stage of learning new indicators, alternatives should be identified and calculations should be made that determine subsequent investments in human capital. It should also be borne in mind that each the alternative has its own characteristic features and characteristics, behavioral connections and influence on the choice of options capital investment.

Table 4
Initialized data affecting machine learning training in the search for optimal investments in organizational capital.

Action number	Impact values on organizational capital					
	Salary	Branding Capital	Technology Capital	Value Added Efficiency Capital	Business Culture Capital	Implementation capital innovation information capacity
16637	2219	52	286	313	45	333
25352	1431	155	121	95	128	318
74521	725	157	151	54	138	340
168348	684	184	115	74	123	286
2236341	485	133	87	33	142	118
14330450	336	197	90	51	165	114
17735547	294	127	44	201	153	134

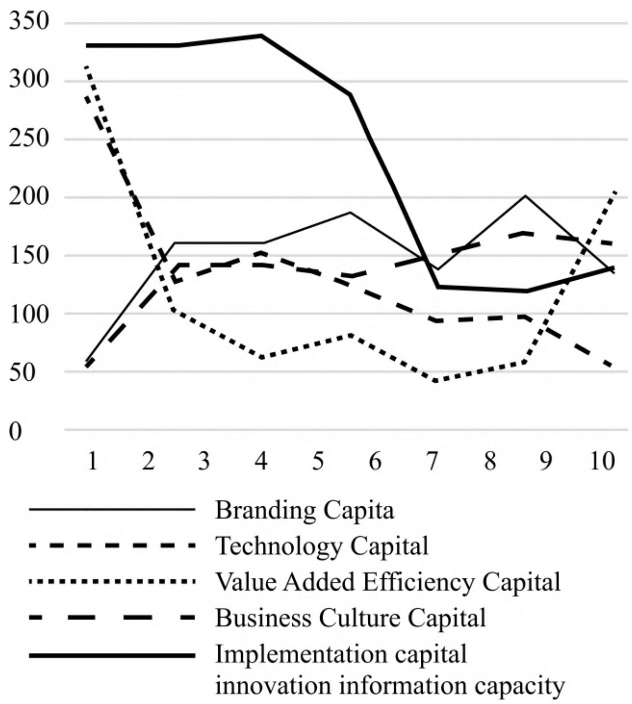


Figure 3: Machine learning of alternative development of organizational capital of the enterprise.

Taking into account the dynamics of changes in results, it can be concluded that subsequent training cycles can bring more optimized costs. Figure 4 shows the optimization of the costs of organizational capital development, taking into account the same level of organizational capital development.

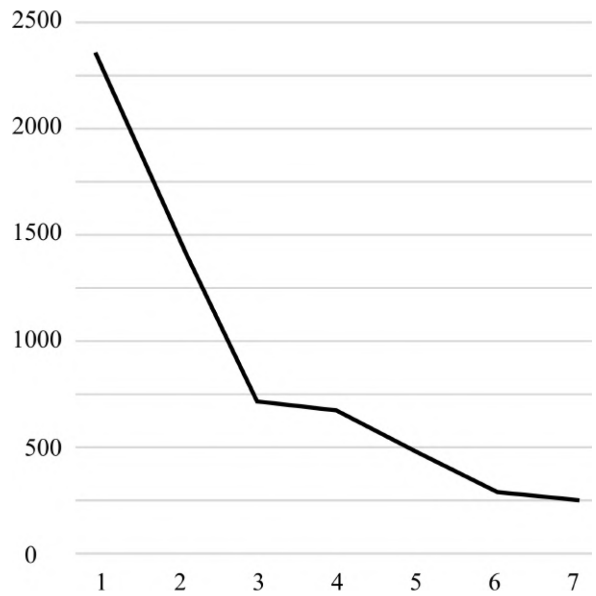


Figure 4: Machine learning of alternative development of organizational capital of the enterprise.

It is also worth noting that when the input data changes, machine learning will be able to rebuild and generate calculations and optimize the result better and faster than a person.

3. Conclusions

The study presents a conceptual framework for the application of Q-leaning to ascertain the most effective developmental strategy for organizational capital within the context of intellectual capital. This approach aims to bolster the reliability of the outcomes achieved.

As a result, the strategy's capital for fostering information potential innovations and the capital of alternatives independently undertake pivotal roles in shaping and implementing mechanisms for managing intellectual capital, both in conjunction with and separately from other capital types.

The crux of this approach lies in the judicious selection of significance indicators (returns) for contributions to various organizational capital facets, driving iterative learning cycles. Such an approach streamlines the exploration and formulation of organizational capital development strategies, opening pathways to genuine alternatives and simplifying decision-making processes.

Notably, tuning training by altering parameters such as reward magnitude, data optimization value, and training constraints can yield superior outcomes by accelerating training processes and furnishing a more proficient AI capable of delivering enhanced results.

Leveraging machine learning to optimize costs linked with organizational capital development stands as the most effective method. The advantages of swiftness, objectivity, and adaptability to external shifts distinguish this approach from human-centric alternatives.

To bolster outcomes, fine-tuning of these actions and precise selection of alternatives for action-based choices are deemed essential.

References

[1] V. Porokhnya, V. Penev, R. Ivanov, V. Kravchenko, Flexible Evolutionary Model of Machine Learning of Organizational Capital Development Strategies with Optimization of Spent Resources, in: S. Semerikov, V. Soloviev, A. Matviychuk, V. Kobets, L. Kibalnyk, H. Danylchuk, A. Kiv (Eds.), Proceedings of 10th International Conference on Monitoring, Modeling & Management of Emergent Economy - M3E2, INSTICC, SciTePress, 2023, pp. 71–79. doi:10.5220/0011931400003432.

[2] A. Brooking, Intellectual Capital: Core Assets for the Third Millenium, International Thompson Business Press, 1996.

[3] J. H. Daum, Intangible Assets, Wiley, 2002, pp. 152–154.

[4] B. B. Leontiev, Tcena intellekta. Intellektualnyi kapital v rossiiskom biznese [The price of intelligence. Intellectual capital in Russian business], Moscow, 2002.

[5] G. Roos, S. Pike, L. Fernstrom, Managing Intellectual Capital in Practice, Butterworth-Heinemann, 2005. URL: https://www.researchgate.net/publication/274392498.

[6] K.-E. Sveiby, Methods for Measuring Intangible Assets, 2010. URL: https://www.sveiby.com/article/Methods-for-Measuring-Intangible-Assets.

[7] V. Porokhnya, The general theory of relativity of economic growth, State and regions. Economy and entrepreneurship series (2009) 153–160.

[8] V. Kobets, O. Novak, EU countries clustering for the state of food security using machine learning techniques, Neiro-Nechitki Tekhnolohii Modelyuvannya v Ekonomitsi 2021 (2021) 86–118. doi:10.33111/nfmte.2021.086.

[9] V. Derbentsev, A. Matviychuk, V. N. Soloviev, 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, Springer, Singapore, 2020, pp. 211–231. doi:10.1007/978-981-15-4498-9_12.

[10] A. Kiv, S. Semerikov, V. N. Soloviev, L. Kibalnyk, H. Danylchuk, A. Matviychuk, Experimental Economics and Machine Learning for Prediction of Emergent Economy Dynamics, in: A. Kiv, S. Semerikov, V. N. Soloviev, L. Kibalnyk, H. Danylchuk, A. Matviychuk (Eds.), Proceedings of the Selected Papers of the 8th International Conference on Monitoring, Modeling & Management of Emergent Economy, M3E2-EEMLPEED 2019, Odessa, Ukraine, May 22-24, 2019, volume 2422 of *CEUR Workshop Proceedings*, CEUR-WS.org, 2019, pp. 1–4. URL: https://ceur-ws.org/Vol-2422/paper00.pdf.

[11] A. Kiv, P. Hryhoruk, I. Khvostina, V. Solovieva, V. N. Soloviev, S. Semerikov, Machine learning of emerging markets in pandemic times, in: A. Kiv (Ed.), 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, volume 2713 of *CEUR Workshop Proceedings*, CEUR-WS.org, 2020, pp. 1–20. URL: https://ceur-ws.org/Vol-2713/paper00.pdf.
[12] R. Bellman, Dynamic Programming, Princeton University Press, New York, 1957. URL: https://www.gwern.net/docs/statistics/decision/1957-bellman-dynamicprogramming.pdf.
[13] C. J. C. H. Watkins, P. Dayan, Q-learning, Machine Learning 8 (1992) 279–292. doi:10.1007/BF00992698.

Recurrence quantification analysis of energy market crises: a nonlinear approach to risk management*

Andrii O. Bielinskyi[1,2], Vladimir N. Soloviev[1,3], Viktoria V. Solovieva[2], Serhiy O. Semerikov[1,4,5,6,7] and Michael A. Radin[1]

[1]Kryvyi Rih State Pedagogical University, 54 Gagarin Ave., Kryvyi Rih, 50086, Ukraine
[2]State University of Economics and Technology, 16 Medychna Str., Kryvyi Rih, 50005, Ukraine
[3]Kyiv National Economic University named after Vadym Hetman, 54/1 Peremogy Ave., Kyiv, 03680, Ukraine
[4]Kryvyi Rih National University, 11 Vitalii Matusevych Str., Kryvyi Rih, 50027, Ukraine
[5]Institute for Digitalisation of Education of the National Academy of Educational Sciences of Ukraine, 9 M. Berlynskoho Str., Kyiv, 04060, Ukraine
[6]Zhytomyr Polytechnic State University, 103 Chudnivsyka Str., Zhytomyr, 10005, Ukraine
[7]Academy of Cognitive and Natural Sciences, 54 Gagarin Ave., Kryvyi Rih, 50086, Ukraine
[8]Rochester Institute of Technology, 1 Lomb Memorial Dr, Rochester, NY 14623, United States

Abstract

The energy market is characterized by unstable price dynamics, which challenge the quantitative models of pricing processes and result in abnormal shocks and crashes. We use recurrence quantification analysis (RQA) to analyze and construct indicators of intermittent events in energy indices, where regular patterns are interrupted by chaotic fluctuations, which could signal the onset of crisis events. We apply RQA to daily data of Henry Hub natural gas spot prices, WTI spot prices, and Europe Brent spot prices. Our empirical results show that the recurrence measures capture the distinctive features of crashes and can be used for effective risk management strategies.

Keywords

energy market, recurrence quantification analysis, crash detection, risk management, price dynamics, instability, abnormal shocks

1. Introduction

Crude oil stands as a linchpin in the stability of global economic and financial systems, rendering it a strategic asset for national economic progress [2, 3]. Analyzing the multifaceted determinants impacting crude oil prices becomes paramount for investors, governmental bodies, and stakeholders. These price fluctuations stem from diverse sources, including fundamental

M3E2-MLPEED 2022: The 10th International Conference on Monitoring, Modeling & Management of Emergent Economy, November 17-18, 2022, Virtual, Online
*Extended and revised version of paper [1] presented at the 10th International Conference on Monitoring, Modeling & Management of Emergent Economy.
krivogame@gmail.com (A. O. Bielinskyi); vnsoloviev2016@gmail.com (V. N. Soloviev); vikasolovieva2027@gmail.com (V. V. Solovieva); semerikov@gmail.com (S. O. Semerikov); michael.a.radin@gmail.com (M. A. Radin)
0000-0002-2821-2895 (A. O. Bielinskyi); 0000-0002-4945-202X (V. N. Soloviev); 0000-0002-8090-9569 (V. V. Solovieva); 0000-0003-0789-0272 (S. O. Semerikov); 0000-0001-9951-7955 (M. A. Radin)

factors such as crude oil supply and demand [4], as well as non-fundamental factors like investor sentiment and speculations [5]. Notably, the interplay of the global economic landscape, geopolitical stability among oil-producing nations, and economic policy uncertainty significantly shape crude oil prices.

Given the pivotal role of crude oil in economic advancement, the market's volatile nature has led to substantial economic repercussions, especially for oil-import-dependent countries. Consequently, numerous studies have delved into the drivers of crude oil price volatility, fuelling debates within academia regarding the mechanics of the crude oil market [6, 7, 8]. Amidst this discourse, the intricate risks posed by crude oil price fluctuations, driven by their stochastic and complex nature, have come to the forefront [9, 10, 11, 12].

Two key benchmarks, namely WTI and Brent contracts, typically set oil prices. These benchmarks are favored by hedge funds and traders, resulting in considerable interest in the WTI-Brent pricing structure encompassing futures curve shapes, benchmark price disparities, and market integration levels. Financial institutions actively engage in these markets, amplifying their influence on jet fuel, diesel, heating oil, and gasoline prices. Furthermore, the WTI-Brent spread serves as a foundation for derivative financial products like swaps and options.

While microeconomic theory attributes crude oil prices to supply and demand dynamics, the financialization of oil in the past decade has heightened speculative influences, complicating price determination [13, 14].

The natural gas industry has flourished due to substantial market demand, abundant cost-effective supply, and thriving global trade. Predicting natural gas prices holds critical importance in trading, electric power planning, and regulatory decision-making. Henry Hub (U.S.), NBP (U.K.), and LNG (Japan) now serve as vital international natural gas trading hubs. Among these, Henry Hub boasts the highest liquidity, widest impact, and strongest reflection of supply-demand dynamics. Beyond fundamental factors, natural gas prices are influenced by elements such as extreme weather, geopolitical conflicts, and international relations [15].

Given the nonlinear and non-stationary traits of crude oil and natural gas markets under intricate influences, enhancing the early detection accuracy of market crises remains a pivotal research goal. This paper introduces recurrence analysis-based indicators (indicator-precursors) for this purpose.

2. Methodology of recurrence analysis

In 1890 Poincaré introduced *Poincaré recurrence theorem* [16], which states that certain systems return to their arbitrarily close, or exactly the same initial states after a sufficiently long but finite time. Such property in the case of deterministic behavior of the system allows us to make conclusions regarding its future development.

2.1. Time delay method

The state of the system can be described by the set of variables. Its observational state can be expressed through a d-dimensional vector or matrix, where each of its components refers to a single variable that represents a property of the system. After a while, the variables change, resulting in different system states.

Usually, not all relevant variables can be captured from our observations. Often, only a single variable may be observed. *Thakens' theorem* [17] that was mentioned in previous sections ensures that it‘s possible to reconstruct the topological structure of the trajectory formed by the state vectors, as the data collected for this single variable contains information about the dynamics of the whole system.

For an approximate reconstruction of the original dynamics of the observed system, we project the time series onto a Reconstructed Phase Space [18, 19, 20] with the commonly used time delay method [19] which relied on the *embedding dimension* and *time delay*.

The embedding dimension is being the dimensionality of the reconstructed system (corresponds to the number of relevant variables that may differ from one system to another. The time delay parameter specifies the temporal components of the vector components.

2.2. Recurrence plot

Recurrence plot (RP) have been introduced to study dynamics and recurrence states of complex systems. When we create RP, at first, from recorded time series we reconstruct phase-space trajectory. Then, according to Eckmann et al. [21], we consider a trajectory $\vec{\mathrm{X}}(i)$ on the reconstructed trajectory. The recurrence plot is an array of dots in a $N \times N$ matrix, where dot is placed at (i, j) whenever $\vec{\mathrm{X}}(j)$ is sufficiently close to $\vec{\mathrm{X}}(i)$, and both axes are time axes which mathematically can be expressed as

$$\begin{gathered} R_{ij} = \Theta(\epsilon - \left\| \vec{\mathrm{X}}(i) - \vec{\mathrm{X}}(j) \right\|), \\ \text{for } i, j = 1, \dots, N, \end{gathered} \tag{1}$$

where $\| \; \|$ is a norm (representing the spatial distance between the states at times i and j); ϵ is a predefined recurrence threshold, and $\Theta(\cdot)$ is the Heaviside function. As a result, the matrix captures a total of N^2 binary similarity values.

Typically, L_p-norm is applied to determine the pairwise similarity between two vectors. According to Webber and Zbilut [22], the L_1-norm (Taxicab metric), the L_2-norm (Euclidean metric), and the L_∞-norm (Chebyshev metric) can serve as candidates for measuring distance between trajectories in phase space.

Also, as it can be seen from equation (1), the similarity between vectors is determined by a threshold ϵ. The choice of $\epsilon > 0$ ensures that all vectors that lie within this radius are similar to each other, and that dissimilarity up to a certain error is permitted [16].

The fixed radius for recurrent states is the commonly used condition, which leads to equally sized ϵ-neighborhoods. The shape in which neighborhoods lie is determined by the distance metric. Applying the fixed threshold with the distance metric, we define recurrence matrices that are symmetric along the middle diagonal. The self-similarity of the multi-dimensional vectors reflects in the middle diagonal, which is commonly referred to as the line of identity (LOI). In contrast, it is not guaranteed that a recurrence matrix is symmetric if the condition of the fixed number of nearest neighbors is applied. For specific purposes (e.g., quantification of recurrences), it can be useful to exclude the LOI from the RP, as the trivial recurrence of a state with itself might not be of interest [23].

The main purpose of RP is the visualization of trajectories and hidden patterns of the systems [24, 23].

The dots within RP, representing the time evolution of the trajectories, exhibit characteristic large-scale and small-scale patterns. Large-scale patterns of RP can be classified as

- *homogeneous* – autonomous and stationary systems, which consist of many recurrence points that are homogeneously distributed (relaxation times are short);
- *periodic* – long, uninterrupted, and diagonally oriented structures that represent which indicate periodic behavior. These lines are usually distributed regularly;
- *drift* – systems with patterns paling or darkening from the LOI to the outer corners of RP;
- *disrupted* – systems with drastic changes as well as extreme events in the system dynamics.

The small-scale clusters can represent a combination of *isolated dots* (abrupt events). Similar evolution at different periods in time or in reverse temporal order will present *diagonal lines* (deterministic structures) as well as *vertical/horizontal lines* to inscribe laminar states (intermittency) or systems that paused at singularities. For the quantitative description of the system, such small-scale clusters serve the base of the *recurrence quantification analysis* (RQA).

2.3. Recurrence quantification analysis

The graphic representation of the system suits perfectly for a qualitative description. However, the main disadvantage of graphical representation is that it forces users to subjectively intuit and interpret patterns and structures presented within the recurrence plot. Also, with the increasing size of RP, they can be hardly depicted on graphical display as a whole. As a result, we need to work with separated parts of the original plot. Analysis in such a way may create new defects, which should distort objectivity of the observed patterns and lead to incorrect interpretations. To overcome such limitation and spread an objective assessment among observers, in the early 1990s by Webber and Zbilut [25, 26] were introduced definitions and procedures to quantify RP's complexity, and later, it has been extended by Marwan et al. [27].

The first known measure of the RQA is *recurrence rate*, which measures the probability that the studied process will recur (RR):

$$RR = \frac{1}{N^2} \sum_{i,j=1}^{N} R_{i,j}. \tag{2}$$

Another measure is based on frequency distribution of line structures in the RP. First, we consider the histogram of the length of the diagonal structures in the RP

$$\begin{aligned} P(l) = \sum_{i,j=1}^{N} & (1 - R_{i-1,j-1}) \\ & \times (1 - R_{i+l,j+l}) \prod_{k=0}^{l-1} R_{i+k,j+k}. \end{aligned} \tag{3}$$

The percentage of recurrence points that form diagonal segments of minimal length l_{min} parallel to the main diagonal is the measure of *determinism* (*DET*):

$$DET = \sum_{l=l_{min}}^{N} lP(l) \Bigg/ \sum_{l=1}^{N} lP(l). \tag{4}$$

Systems that are characterized by long diagonal lines are presented to be periodic. From chaotic signals, we would expect short diagonal lines, and stochastic processes would not present any diagonal lines. Performing the RQA, typically, we rely on the lines with minimal length, which excludes the shorter lines, which may be spurious for characterizing deterministic processes. In our case, $l_{min} = 2$ is considered. In case when $l_{min} = 1$, DET and RR are identical.

Considering diagonal line segments, we can emphasize the longest one – L_{max}. This indicator measures the maximum time that two trajectories remain close to each other and can be interpreted as the maximum prediction time:

$$L_{max} = \max\left(\{l_i \,|\, i = 1, \ldots, N_l\}\right), \tag{5}$$

where $N_l = \sum_{l \geqslant l_{min}} P(l)$ is the total number of diagonal lines.

Divergence (DIV) is the inverse of L_{max} characterizes the exponential divergence of the phase space trajectory [28, 29]:

$$DIV = 1 \Big/ L_{max}. \tag{6}$$

For longer diagonal lines system is more deterministic and, therefore, the measure of divergence is also lower. The smaller L_{max}, the more divergent are trajectories and more chaotic the studied system. According to Eckmann et al. [21], *DIV* can be used to estimate the largest positive Lyapunov exponent.

Another measure which is related to the diagonal line segments is the *average diagonal line length* (L_{mean}):

$$L_{mean} = \sum\nolimits_{l=l_{min}}^{N} lP(l) \Big/ \sum\nolimits_{l=l_{min}}^{N} P(l) \tag{7}$$

It can be interpreted as the mean prediction horizon of the system, and it measures average time that two trajectories remain close to each other.

Using the classic Shannon entropy, we can measure the hidden complexity of recurrence structures in the RP. In accordance with this study, the entropy of diagonal line histogram (*DLEn*) is of the greatest interest. It can be defined as:

$$DLEn = -\sum_{l=l_{min}}^{N} p(l) \ln p(l) \tag{8}$$

and

$$p(l) = P(l) \Bigg/ \sum_{l=l_{min}}^{N} P(l), \tag{9}$$

where $p(l)$ captures the probability that a diagonal line has exactly length l, and *DLEn* reflects the complexity of deterministic structure in the system. The more uniform is the frequency distribution of diagonal lines, the higher the value of *DLEn*. If there is predominant deterministic behavior with a particular period l, then *DLEn* becomes lower.

As it was mentioned, the RP structure consists of vertical (horizontal lines). For them Marwan and Webber [30] proposed additional recurrence measures. The first of them is the *laminarity* (*LAM*) Analogously to the equation (4), which measures the percentage of diagonal lines with minimal length l_{min} in the RP, we can calculate the fraction of recurrence points forming vertical structures of minimal length v_{min}:

$$LAM = \sum_{v=v_{min}}^{N} vP(v) \Big/ \sum_{v=1}^{N} vP(v) \tag{10}$$

with

$$\begin{aligned} P(v) = \sum_{i,j=1}^{N} & \left(1 - R_{i,j-1}\right) \\ & \times \left(1 - R_{i,j+v}\right) \prod_{k=0}^{v-1} R_{i,j+k} \end{aligned} \tag{11}$$

as the histogram of lengths of vertical lines.

Since it measures the overall amount of vertical lines, it characterizes the percentage of laminar states within the system. If *LAM* increases, then there are more vertical or diagonal structures than isolated recurrent points.

Similarly to L_{max}, we can define the measure which will indicate the maximum time that a system holds an unchangeable pattern – the *maximal vertical lines length* (V_{max}):

$$V_{max} = \max\left(\{v_i \,|\, i = 1, \ldots, N_v\}\right), \tag{12}$$

where $N_v = \sum_{v \geqslant v_{min}} P(v)$ is the total number of vertical lines.

Vertical line divergence (*VDIV*) is the analogous to (6), which can be related to the rate of divergence from laminar state:

$$VDIV = 1 \Big/ V_{max}. \tag{13}$$

Consequently, we can define the average time that two trajectories remain at a specific state – *trapping time* (*TT*):

$$TT = \sum\nolimits_{v=v_{min}}^{N} vP(v) \Big/ \sum\nolimits_{v=v_{min}}^{N} P(v). \tag{14}$$

For high *TT* values we would expect the system to consist of more laminar states, whereas low *TT* values would indicate abrupt changes in the system's dynamics.

The variability of laminar states with different duration time can be measured in the same way as for diagonal lines – using Shannon entropy. The complexity of vertical lines can be measures according to the following equation:

$$VLEn = -\sum_{v=v_{min}}^{N} p(v) \ln p(v) \tag{15}$$

with

$$p(v) = P(v) \Bigg/ \sum_{v=v_{min}}^{N} P(v) \tag{16}$$

indicating the probability of a vertical line to have length $v \geqslant v_{min}$.

In the same manner, we can quantify the variation (complexity) of abrupt changes during the studied periods in the energy markets. Regarding equation (7), we can quantify the average time of divergence when two trajectories in the phase-space remain out of recurrence threshold ϵ. This measure can be called as *average white vertical line length* (WVL_{mean}):

$$WVL_{mean} = \sum_{w=w_{min}}^{N} wP(w) \Bigg/ \sum_{w=w_{min}}^{N} P(w), \tag{17}$$

where $P(w)$ is the frequency of white vertical lines in the RP. This measure can be interpreted as the mean horizon of unpredictability of the system.

This kind of complexity is associated with the white vertical lines in the RP and can be quantified in the following way:

$$WVLEn = -\sum_{w=w_{min}}^{N} p(w) \ln p(w) \tag{18}$$

with

$$p(w) = P(w) \Bigg/ \sum_{w=w_{min}}^{N} P(w) \tag{19}$$

indicating the probability of a white vertical line to have length $w \geqslant w_{min}$.

The further measure is based on the ration between DET and RR, and known as *ratio* (DET/RR):

$$DET/RR = N^2 \sum_{l=l_{min}} P(l) \Bigg/ \left(\sum_{l=1}^{N} lP(l)\right)^2 \tag{20}$$

In the same manner, we can define another measure which is based on the ratio between LAM and DET:

$$\begin{aligned} LAM/DET = &\sum_{v=v_{min}}^{N} vP(v) \cdot \sum_{l=1}^{N} lP(l) \\ &\Bigg/ \sum_{v=1}^{N} vP(v) \cdot \sum_{l=l_{min}}^{N} lP(l). \end{aligned} \tag{21}$$

This measures can be used to uncover hidden transitions in the dynamics of the system [25].

3. Results and analysis

Regarding previous studies, we present additional analysis on co-movement between 3 energy-related indices and construct indicators or indicators-precursors based on the using recurrence analysis.

The presented work uses daily data of Henry Hub natural gas spot prices (US$/MMBTU) ranged from 7 February 1997 to 18 October 2022; Cushing, OK WTI spot prices FOB (US$/BBL) ranged from 20 May 1987 to 17 October 2022; Europe Brent spot prices FOB (US$/BBL) ranged from 20 May 1987 to 17 October 2022 [31, 32].

In figure 1 are presented:

- the dynamics of the initial time series;
- standardized returns, where returns can be calculated as $G(t) = [x(t + \Delta t) - x(t)]/x(t)$ and their standardized version as $g(t) = [G(t) - \langle G \rangle]/\sigma$;
- probability density function of the standardized returns.

We can see that most periods in energy markets are defined by events that exceed $\pm 3\sigma$. Both WTI and Brent returns are characterized by much more extensive crashes. Previous studies pointed out that such events are located in fat-tails of the probability distribution. Such crashes are the main source of high complexity and non-linearity in the studied systems.

Most of our results are based on the sliding window approach. The idea here is to take a sub-window of a predefined length w. For that sub-window, we perform recurrence quantification analysis, get necessary indicators that are appended to the array. Then, the window is shifted by a predefined time step h, and the procedure is repeated until the time series is completely exhausted.

We have performed RQA under sliding window procedure for standardized returns and standardized initial time series [33, 34, 35, 36, 37, 38]. We have found that standardized initial time series better expresses internal complexity and recurrent properties of the energy market indices.

RQA was performed for the following parameters:

- embedding dimension $d_E = 1$;
- time delay $\tau = 1$;
- recurrence threshold $\epsilon = 0.3$;
- L_2-norm as a candidate for measuring distance between trajectories in phase space;
- minimum diagonal line length $l_{min=2}$;
- minimum vertical line length $v_{min} = 2$;
- minimum white vertical line length $w_{min} = 2$;
- sliding window length $w = 500$ days;
- sliding window time step $h = 1$ day.

Worth to mention that the experiments were performed for sliding window lengths of 250 days and 500 days. We have chosen the second option since it represents a more reliable and smoother dynamics of all the presented indicators. All described measures result into highly volatile variation with the sliding window of 250 days that difficult to interpret.

In figure 2 are presented RPs for the studied series.

Recurrence plots in figure 2 represent that the studied energy markets are highly inhomogeneous. As it was expected, nonlinear structure of WTI and Brent is presented to be very similar, comparing to Henry Hub. Recurrence structure of all indices varies across time. They do not

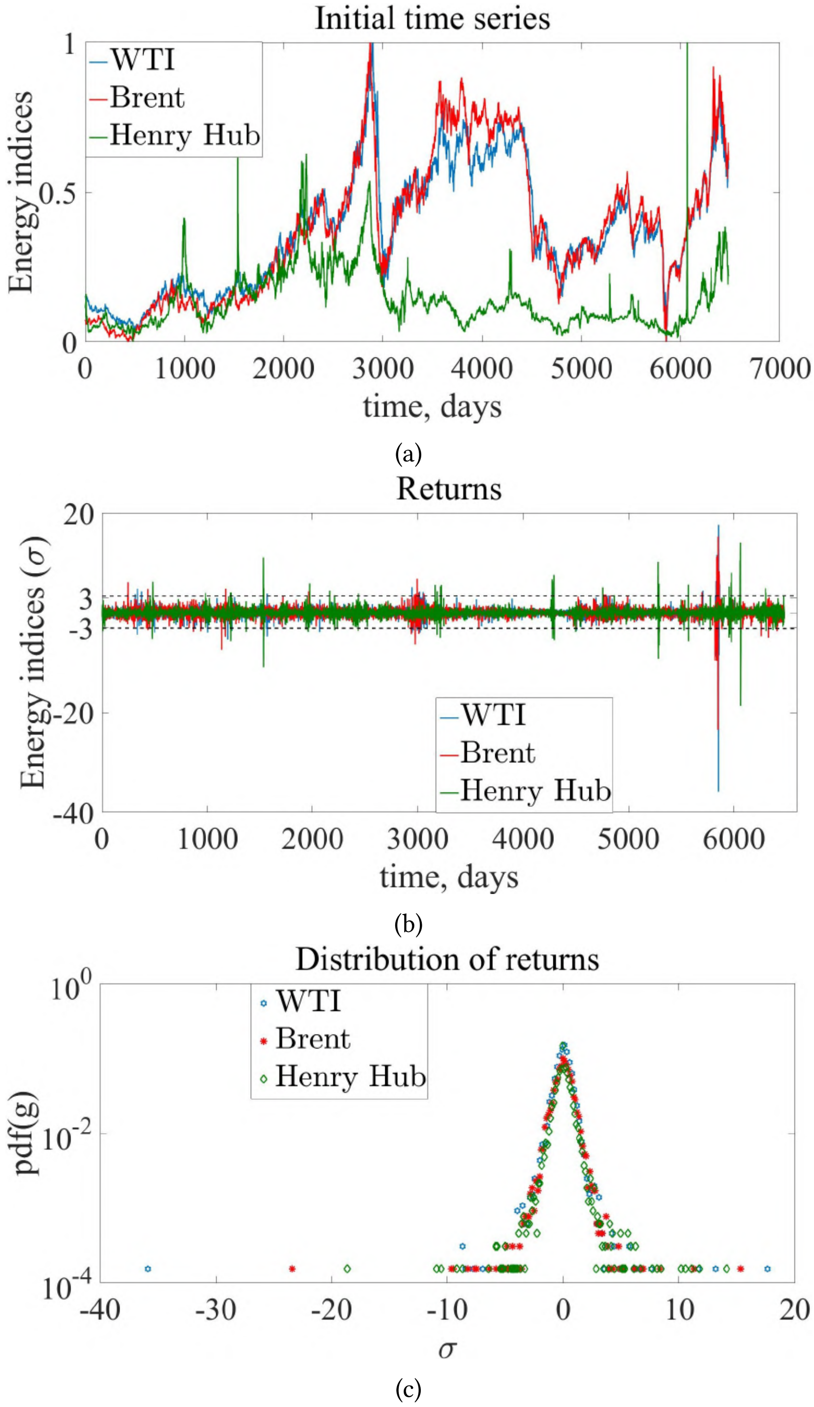
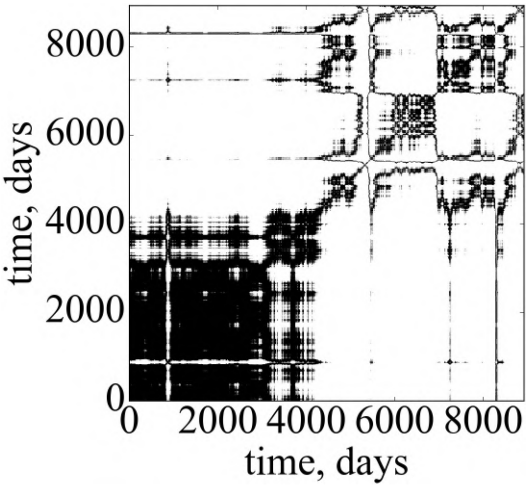
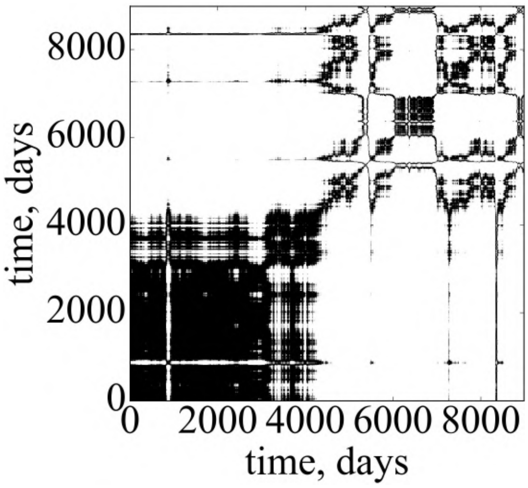


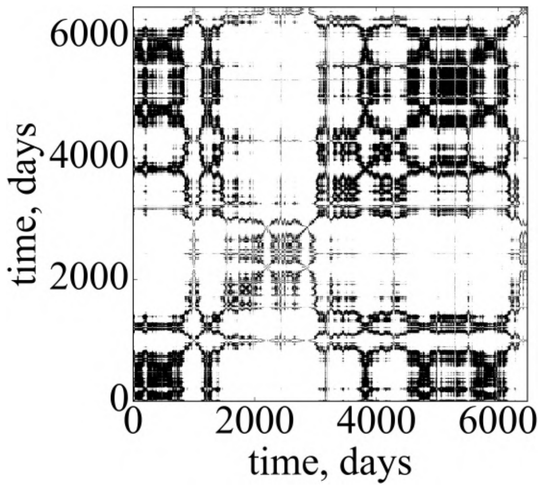
Figure 1: Initial time series (a), standardized returns (b), and pdf of standardized returns of WTI spot prices (WTI), Europe Brent spot prices (Brent), and Henry Hub natural gas spot prices (Henry Hub).



(a)



(b)



(c)

Figure 2: Recurrence plots calculated for WTI (a), Brent (b), and Henry Hub (c) standardized time series.

follow a certain pattern, presented to be non-periodic, and there are differences in the patterns that concern the frequency of their appearance, shape, and size. It should be noticed that for the oil markets first 4000 days are presented to be highly recurrent, while the remaining days seem to be more volatile, which is indicated by high proportion of white regions. The recurrence structure of Henry Hub index is presented to be more uniformly distributed. The variations of recurrence patterns should be more noticeable during crashes. Recurrence quantitative indicators should give a more accurate representation of the complex, chaotic structure of the studied markets.

Figure 3 represents recurrence measures of determinism (*DET*) and laminarity (*LAM*).

In figure 3 we see that *DET* and *LAM* increase during crisis events of all markets. We may conclude that those critical states are characterized by high degree of laminarity and

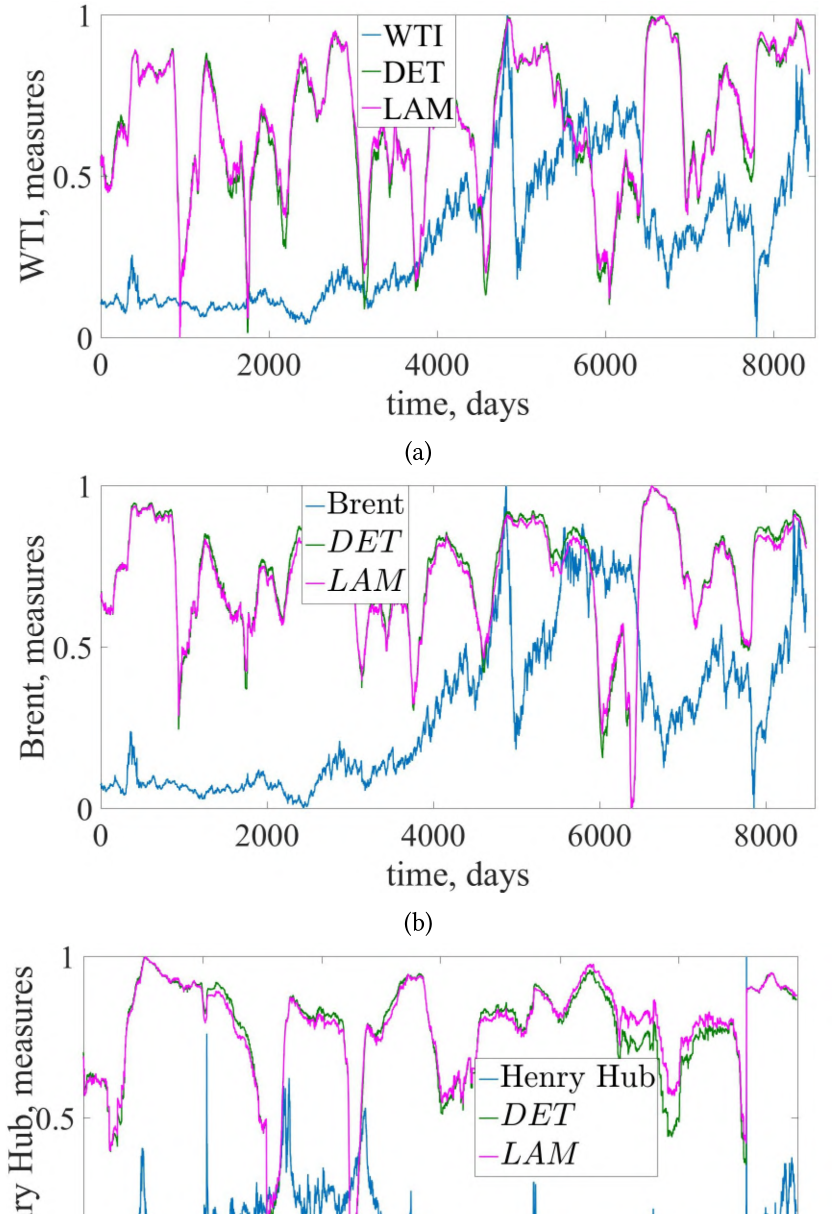


Figure 3: Recurrence measures of determinism (*DET*) and laminarity (*LAM*) calculated for WTI (a), Europe Brent (b), and Henry Hub (c) indices.

determinism. Crashes are presented to be highly complex and deterministic. Their degree of predictability becomes higher, and corresponding recurrence measures seem to be indicators or even indicators-precursors of such changes.

Figure 4 represents recurrence measures of ratios DET/RR and LAM/DET.

From figure 4 we can see that both measures decrease during crisis events of energy indices. For ratio DET/RR we may say that the overall percentage of recurrence points in RP becomes higher than the percentage of only diagonal structures in RP. For ratio LAM/DET we see precisely the same behavior during crashes, i.e., it starts to decline during crisis or even in advance. Thus, it can be seen that the overall determinism of the system during crashes is much higher than the degree of laminarity.

Figure 5 shows recurrence measures of diagonal (DIV) and vertical line ($VDIV$) divergences.

Figure 5 demonstrates that the divergence of deterministic and laminar structure of energy-related markets becomes lower during critical states. Since both measures are inverse quantities to maximum diagonal and vertical line length (L_{max} and V_{max}), such behavior has to be obvious. Previous measures have made it clear to us that the crisis phenomena of energy indices are characterized by a high degree of determinism and laminarity. In this case, the lengths of diagonal and vertical lines should also increase, which indicate an increase in the horizon of predictability and immutability.

Figure 6 represents recurrence measures of recurrence rate (RR), average diagonal line length (L_{mean}), and trapping time (TT).

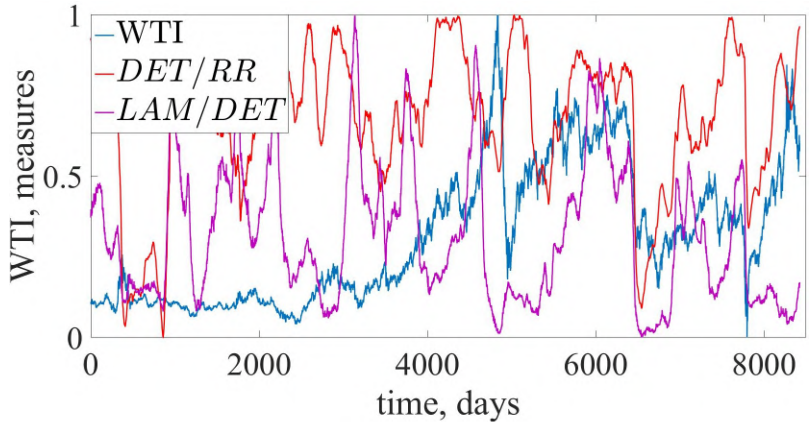
In figure 6 we see that recurrence rate increases during crisis phenomena. This means that the total number of trajectories in the phase space that are close enough to each other becomes larger on the eve of a crisis or at the moment of its onset. Thus, the probability of recurrence state increases during crash. Regarding previous measures, RR and L_{mean}, we see that the average degree of predictability during crisis increases. The same can be seen for trapping time: average degree of changeability increases during crashes. Based on this indicator, we may conclude that the system is 'trapped' in a state of crisis.

Figure 7 presents recurrence measures of average white vertical line length (WVL_{mean}), and diagonal, vertical and white vertical line entropies ($DLEn$, $VLEn$, and $WVLEn$).

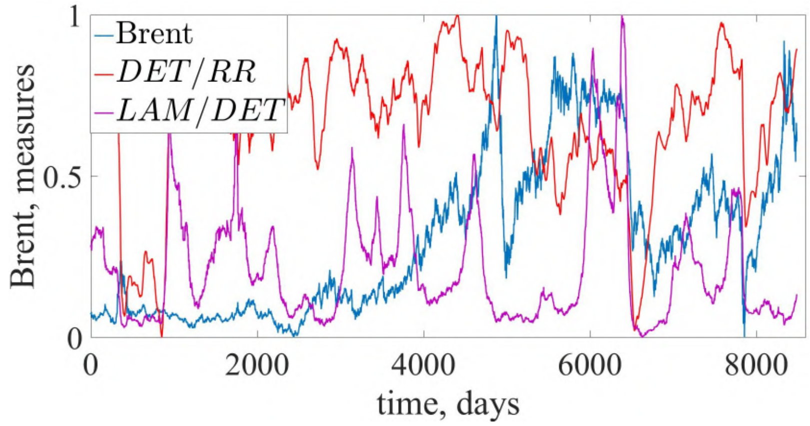
From figure 7 we can see that all the presented quantitative measures of recurrence begin to increase during crises, indicating a special state of the market at these points in time. The average white vertical line length shows that crisis events are characterized not only by the determinism of the dynamics of market movement, but also by the dissimilarity of these events to many previous ones, since the length of the white vertical lines is becoming an increasing trend. It can also be said that the market represents a much more deterministic structure than a laminar one. Also, the degree of volatility of these events can knock the market dynamics out of the limits of the epsilon value.

The diagonal line entropy also shows an increasing trend. Since the Shannon entropy is maximal with a uniform distribution, it can be concluded that the collapse events of energy indices are characterized by different horizons of predictability. That is, in the pre-crisis dynamics there is no black diagonal line of the same length, which is the dominant one. During a crisis, horizons of determinism appear, which gain even more weight if compared with the rest.

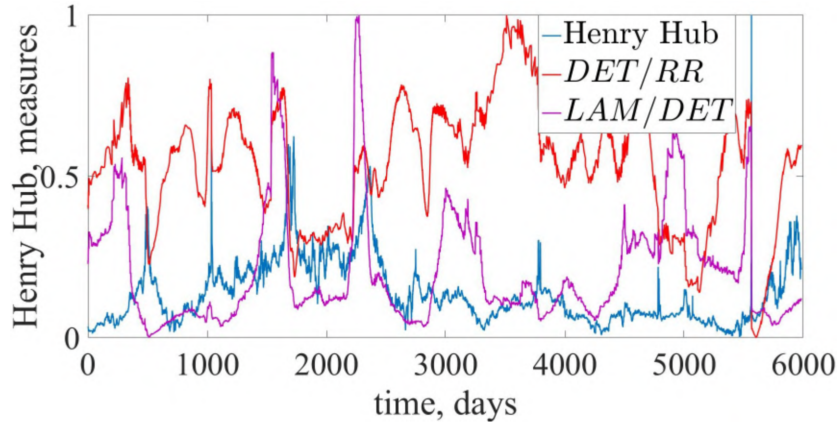
The vertical line entropy increases similarly to $DLEn$. We may assume that similarly to



(a)

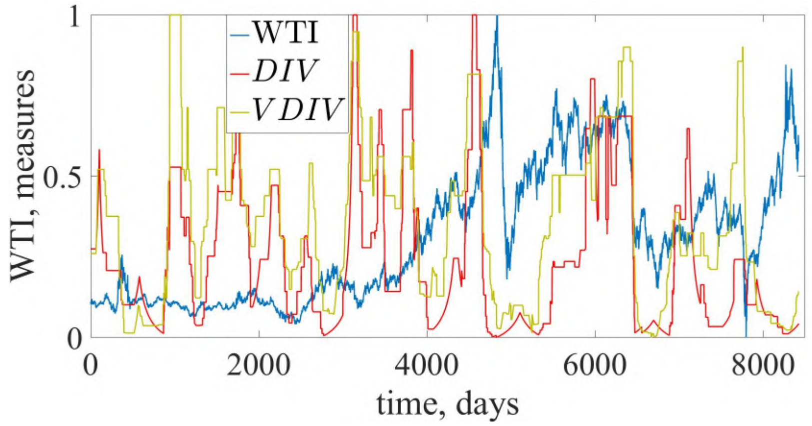


(b)

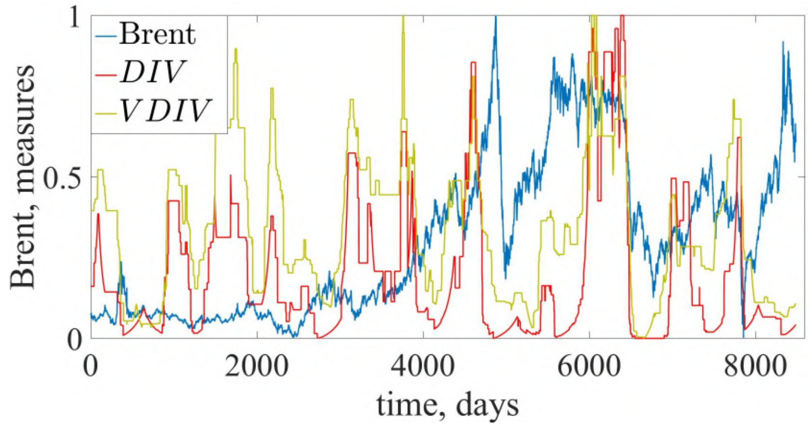


(c)

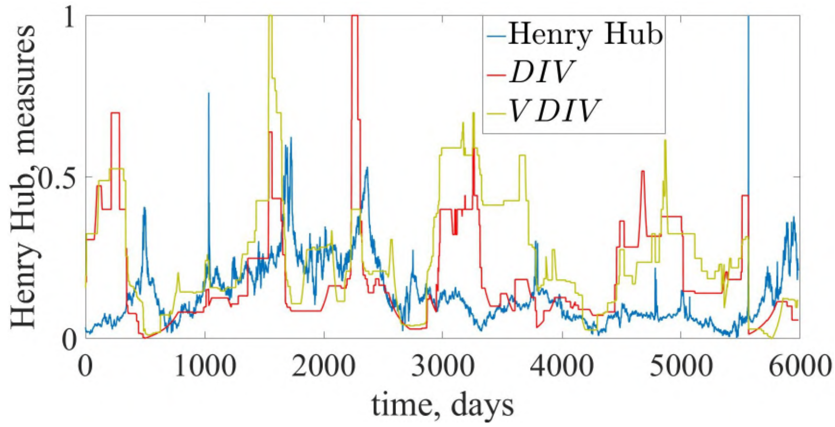
Figure 4: Recurrence measures (DET/RR) and (LAM/DET) calculated for WTI (a), Europe Brent (b), and Henry Hub (c) indices.



(a)

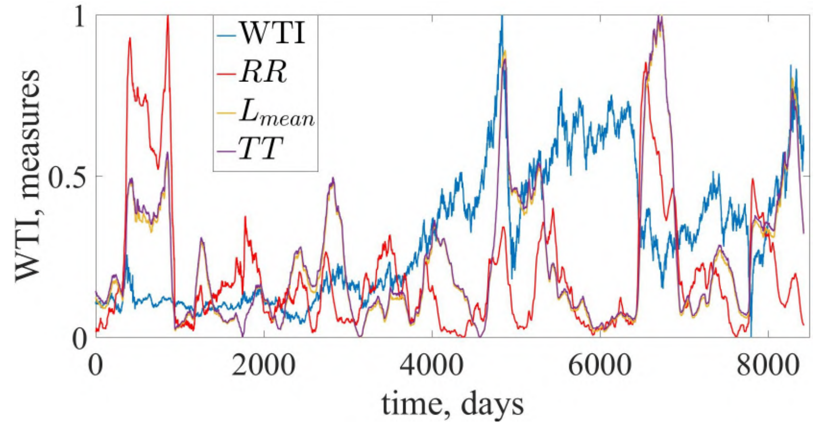


(b)

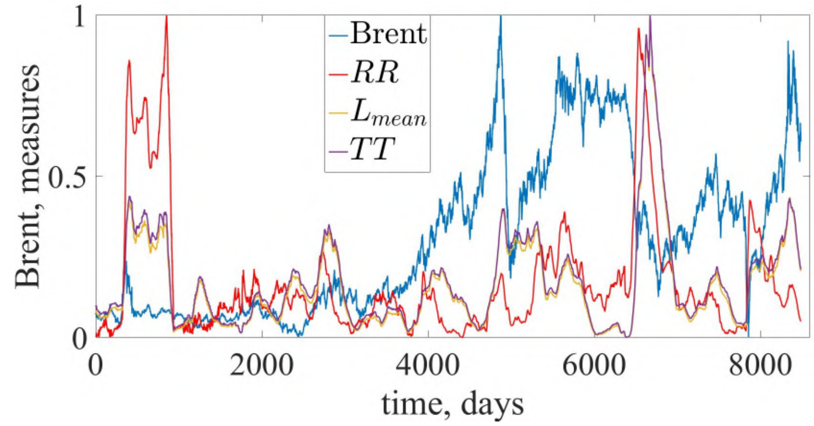


(c)

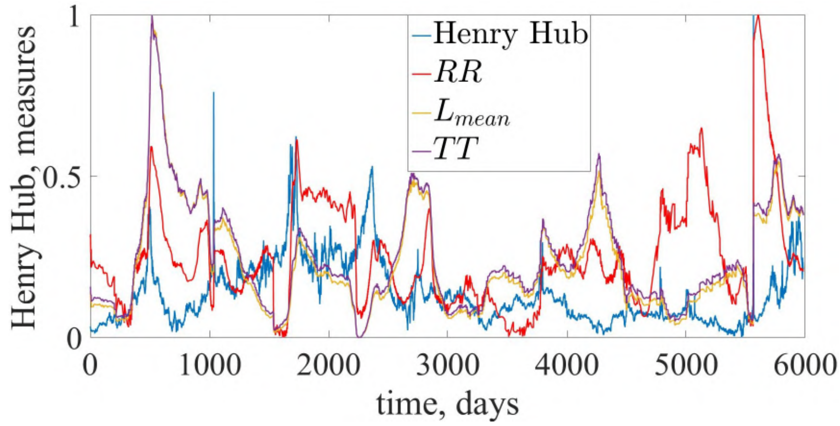
Figure 5: Recurrence measures of diagonal line divergence (DIV) and vertical line divergence ($VDIV$) calculated for WTI (a), Europe Brent (b), and Henry Hub (c) indices.



(a)

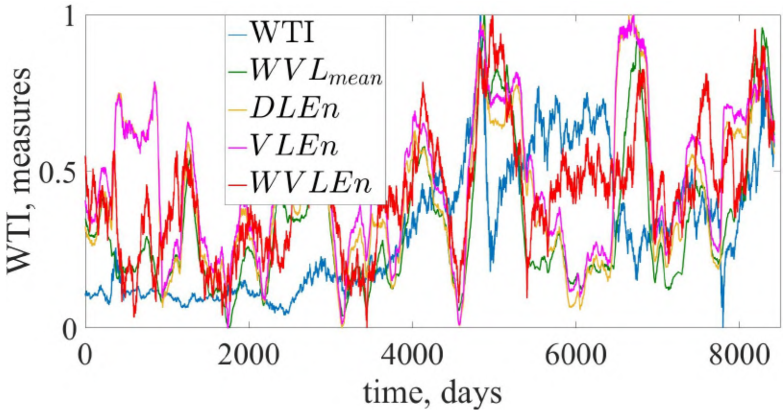


(b)

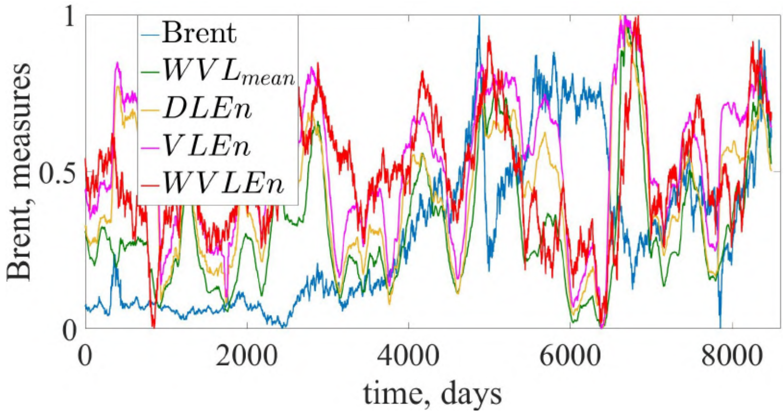


(c)

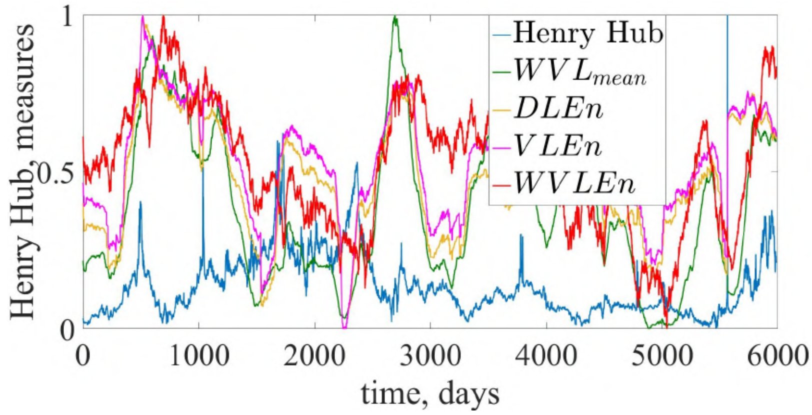
Figure 6: Recurrence measures of recurrence rate (*RR*), average diagonal line length (L_{mean}), and trapping time (*TT*) calculated for WTI (a), Europe Brent (b), and Henry Hub (c) indices.



(a)



(b)



(c)

Figure 7: Recurrence measures of average white vertical line length (WVL_{mean}), diagonal line entropy ($DLEn$), vertical line entropy ($VLEn$), and white vertical line entropy ($WVLEn$) calculated for WTI (a), Europe Brent (b), and Henry Hub (c) indices.

diagonal lines laminar states have different horizons of invariability during crash events, and these horizons of invariability have greater tendency to uniform distribution.

The white vertical line entropy increases similarly to other entropies. This dynamics is consistent with the WVL_{mean} measure.

4. Conclusions

This study has delved into the intricate, nonlinear, and nonstationary dynamics of the oil and gas markets through the lens of recurrence analysis. Leveraging daily data spanning from 7 February 1997 to 18 October 2022 for Henry Hub natural gas spot prices, from 20 May 1987 to 17 October 2022 for WTI spot prices, and corresponding data for Europe Brent spot prices, we draw several significant conclusions from our empirical findings.

Firstly, our analysis of recurrence plots reveals the inherent inhomogeneity within the studied markets. Notably, the nonlinear structures of WTI and Brent exhibit remarkable similarities when contrasted with Henry Hub. Furthermore, recurrence patterns for all indices exhibit temporal variations, demonstrating differences in the frequency, shape, and size of black- and white-dot patterns across time.

From quantitative measures of complexity, the following insights emerge:

1. **Characteristics of Crash Events**: Crashes in energy-related indices exhibit a pronounced degree of both laminarity and determinism, indicating a high level of complexity and determinism during these events.
2. **Determinism vs. Laminarity**: The percentage of recurrence points surpasses that of only diagonal structures during crises. While the overall degree of determinism outweighs laminarity, a higher percentage of diagonal lines during crises highlights their significance.
3. **Divergence During Critical States**: The divergence between deterministic and laminar structures diminishes during critical states, indicating increased repeatability in the dynamics of the studied systems. This suggests that phase-space trajectories converge during financial crises.
4. **Recurrence Measures during Crises**: Measures like recurrence rate, mean diagonal line length, and trapping time increase during crisis periods. This implies a larger number of closely situated trajectories in phase space, raising the probability of recurrence states and predictability. Greater presence of vertical lines signifies the system being 'trapped' in a crisis state.
5. **Entropy-based Measures**: Entropy-based measures, especially white vertical line measures, reveal intricate nonlinear patterns in energy-related indices that encompass determinism, laminarity, and dissimilarities reflected in white lines.

The approach applied to WTI, Brent, and Henry Hub indices underscores the energy market's nature as an open, chaotic, nonlinear system intricately intertwined with diverse technical and fundamental factors. While recurrence plots and recurrence quantification analysis offer promising outcomes for crisis prediction and early-warning indicator construction, their practical application in trading strategies and autonomous trading bots necessitates further refinement.

Furthermore, for accurate crisis forecasting, the integration of proposed indicators (indicators-precursors) with specific forecasting models is imperative [11, 39, 15, 12, 40, 41, 42]. This convergence seems particularly promising at the intersection of artificial intelligence and fuzzy logic methods [43, 44, 45, 46, 47, 48, 49].

Simultaneously, we aim to explore cross-recurrences between energy indices and diverse technical and fundamental indicators utilizing cross- and joint-recurrence quantification analysis [50, 51, 52]. This avenue holds potential for further unraveling the intricate relationships within energy markets.

Acknowledgments

This work was supported by the Ministry of Education and Science of Ukraine (project No. 0122U001694).

References

[1] A. Bielinskyi, V. Soloviev, V. Solovieva, S. Semerikov, M. Radin, Recurrence Measures of Complexity in Energy Market Dynamics, in: S. Semerikov, V. Soloviev, A. Matviychuk, V. Kobets, L. Kibalnyk, H. Danylchuk, A. Kiv (Eds.), Proceedings of 10th International Conference on Monitoring, Modeling & Management of Emergent Economy - M3E2, INSTICC, SciTePress, 2023, pp. 122–133. doi:10.5220/0011931800003432.

[2] Y.-J. Zhang, Y.-B. Wu, The time-varying spillover effect between WTI crude oil futures returns and hedge funds, International Review of Economics & Finance 61 (2019) 156–169. doi:10.1016/j.iref.2019.02.006.

[3] Y. Dong, M. Zhang, R. Zhou, Classification of Network Game Traffic Using Machine Learning, in: H. Yuan, J. Geng, C. Liu, F. Bian, T. Surapunt (Eds.), Geo-Spatial Knowledge and Intelligence, Springer, Singapore, 2018, pp. 134–145. doi:10.1007/978-981-13-0893-2_15.

[4] G. Wu, Y.-J. Zhang, Does China factor matter? An econometric analysis of international crude oil prices, Energy Policy 72 (2014) 78–86. doi:10.1016/j.enpol.2014.04.026.

[5] Q. Ji, E. Bouri, D. Roubaud, L. Kristoufek, Information interdependence among energy, cryptocurrency and major commodity markets, Energy Economics 81 (2019) 1042–1055. doi:10.1016/j.eneco.2019.06.005.

[6] L. Coleman, Explaining crude oil prices using fundamental measures, Energy Policy 40 (2012) 318–324. doi:10.1016/j.enpol.2011.10.012.

[7] R. Sari, U. Soytas, E. Hacihasanoglu, Do global risk perceptions influence world oil prices?, Energy Economics 33 (2011) 515–524. doi:10.1016/j.eneco.2010.12.006.

[8] S. Dées, P. Karadeloglou, R. K. Kaufmann, M. Sánchez, Modelling the world oil market: Assessment of a quarterly econometric model, Energy Policy 35 (2007) 178–191. doi:10.1016/j.enpol.2005.10.017.

[9] Y.-J. Zhang, J. Wang, Exploring the WTI crude oil price bubble process using the Markov regime switching model, Physica A: Statistical Mechanics and its Applications 421 (2015) 377–387. doi:10.1016/j.physa.2014.11.051.

[10] U. Shahzad, S. K. Jena, A. K. Tiwari, B. Doğan, C. Magazzino, Time-frequency analysis between Bloomberg Commodity Index (BCOM) and WTI crude oil prices, Resources Policy 78 (2022) 102823. doi:10.1016/j.resourpol.2022.102823.

[11] T. Yin, Y. Wang, Predicting the price of WTI crude oil futures using artificial intelligence model with chaos, Fuel 316 (2022) 122523. doi:10.1016/j.fuel.2021.122523.

[12] Y. Zhang, M. He, D. Wen, Y. Wang, Forecasting crude oil price returns: Can nonlinearity help?, Energy 262 (2023) 125589. doi:10.1016/j.energy.2022.125589.

[13] L. Kilian, Not All Oil Price Shocks Are Alike: Disentangling Demand and Supply Shocks in the Crude Oil Market, American Economic Review 99 (2009) 1053–69. doi:10.1257/aer.99.3.1053.

[14] R. P. Flood, R. J. Hodrick, On Testing for Speculative Bubbles, Journal of Economic Perspectives 4 (1990) 85–101. doi:10.1257/jep.4.2.85.

[15] J. Li, Q. Wu, Y. Tian, L. Fan, Monthly Henry Hub natural gas spot prices forecasting using variational mode decomposition and deep belief network, Energy 227 (2021) 120478. doi:10.1016/j.energy.2021.120478.

[16] H. Poincaré, The Three-Body Problem and the Equations of Dynamics: Poincaré's Foundational Work on Dynamical Systems Theory, Astrophysics and Space Science Library, 1 ed., Springer, Cham, 2017. doi:10.1007/978-3-319-52899-1.

[17] F. Takens, Detecting strange attractors in turbulence, in: D. Rand, L.-S. Young (Eds.), Dynamical Systems and Turbulence, Warwick 1980, Springer Berlin Heidelberg, Berlin, Heidelberg, 1981, pp. 366–381. doi:10.1007/BFb0091924.

[18] J.-P. Eckmann, D. Ruelle, Ergodic theory of chaos and strange attractors, Rev. Mod. Phys. 57 (1985) 617–656. doi:10.1103/RevModPhys.57.617.

[19] H. Kantz, T. Schreiber, Nonlinear Time Series Analysis, 2 ed., Cambridge University Press, 2003. doi:10.1017/CBO9780511755798.

[20] E. Ott, T. Sauer, J. A. Yorke, Coping with Chaos, Wiley Series in Nonlinear Science, Wiley, 1994.

[21] J.-P. Eckmann, S. O. Kamphorst, D. Ruelle, Recurrence Plots of Dynamical Systems, Europhysics Letters (EPL) 4 (1987) 973–977. doi:10.1209/0295-5075/4/9/004.

[22] C. L. Webber, J. P. Zbilut, Recurrence quantification analysis of nonlinear dynamical systems, in: M. A. Riley, G. C. V. Orden (Eds.), Tutorials in Contemporary Nonlinear Methods for the Behavioral Sciences, National Science Foundation (NSF), 2005, pp. 26–94.

[23] L. W. Charles, C. Ioana, N. Marwan (Eds.), Recurrence Plots and Their Quantifications: Expanding Horizons: Proceedings of the 6th International Symposium on Recurrence Plots, Grenoble, France, 17-19 June 2015, volume 180 of *Springer Proceedings in Physics*, Springer, Cham, 2015. doi:10.1007/978-3-319-29922-8.

[24] N. Marwan, M. Carmen Romano, M. Thiel, J. Kurths, Recurrence plots for the analysis of complex systems, Physics Reports 438 (2007) 237–329. doi:10.1016/j.physrep.2006.11.001.

[25] C. L. Webber, J. P. Zbilut, Dynamical assessment of physiological systems and states using recurrence plot strategies, Journal of Applied Physiology 76 (1994) 965–973. doi:10.1152/jappl.1994.76.2.965.

[26] J. P. Zbilut, C. L. Webber, Embeddings and delays as derived from quantification of recurrence plots, Physics Letters A 171 (1992) 199–203. doi:10.1016/0375-9601(92)

90426-M.
[27] N. Marwan, N. Wessel, U. Meyerfeldt, A. Schirdewan, J. Kurths, Recurrence-plot-based measures of complexity and their application to heart-rate-variability data, Phys. Rev. E 66 (2002) 026702. doi:10.1103/PhysRevE.66.026702.
[28] A. L. Goldberger, L. A. N. Amaral, L. Glass, J. M. Hausdorff, P. C. Ivanov, R. G. Mark, J. E. Mietus, G. B. Moody, C.-K. Peng, H. E. Stanley, PhysioBank, PhysioToolkit, and PhysioNet: components of a new research resource for complex physiologic signals, Circulation 101 (2000) e215–e220. doi:10.1161/01.CIR.101.23.e215.
[29] M. Kirchner, P. Schubert, M. Liebherr, C. T. Haas, Detrended Fluctuation Analysis and Adaptive Fractal Analysis of Stride Time Data in Parkinson's Disease: Stitching Together Short Gait Trials, PLOS ONE 9 (2014) 1–6. doi:10.1371/journal.pone.0085787.
[30] N. Marwan, C. L. Webber, Mathematical and Computational Foundations of Recurrence Quantifications, in: C. L. Webber, N. Marwan (Eds.), Recurrence Quantification Analysis: Theory and Best Practices, Springer International Publishing, Cham, 2015, pp. 3–43. doi:10.1007/978-3-319-07155-8_1.
[31] U.S. Energy Information Administration, Natural gas spot and futures prices (nymex), 1997. URL: https://www.eia.gov/dnav/ng/NG_PRI_FUT_S1_W.htm.
[32] U.S. Energy Information Administration, Spot prices for crude oil and petroleum products, 1986. URL: https://www.eia.gov/dnav/pet/pet_pri_spt_s1_d.htm.
[33] V. N. Soloviev, A. O. Bielinskyi, N. A. Kharadzjan, Coverage of the coronavirus pandemic through entropy measures, CEUR Workshop Proceedings 2832 (2020) 24–42. URL: https://ceur-ws.org/Vol-2832/paper02.pdf.
[34] A. O. Bielinskyi, V. N. Soloviev, Complex network precursors of crashes and critical events in the cryptocurrency market, CEUR Workshop Proceedings 2292 (2018) 37 – 45. URL: https://ceur-ws.org/Vol-2292/paper02.pdf.
[35] A. O. Bielinskyi, A. V. Matviychuk, O. A. Serdyuk, S. O. Semerikov, V. V. Solovieva, V. N. Soloviev, Correlational and Non-extensive Nature of Carbon Dioxide Pricing Market, in: O. Ignatenko, V. Kharchenko, V. Kobets, H. Kravtsov, Y. Tarasich, V. Ermolayev, D. Esteban, V. Yakovyna, A. Spivakovsky (Eds.), ICTERI 2021 Workshops, volume 1635 CCIS of *Communications in Computer and Information Science*, Springer International Publishing, Cham, 2022, pp. 183–199. doi:10.1007/978-3-031-14841-5_12.
[36] A. O. Bielinskyi, O. A. Serdyuk, S. O. Semerikov, V. N. Soloviev, Econophysics of cryptocurrency crashes: a systematic review, in: A. E. Kiv, V. N. Soloviev, S. O. Semerikov (Eds.), Proceedings of the Selected and Revised Papers of 9th International Conference on Monitoring, Modeling & Management of Emergent Economy (M3E2-MLPEED 2021), Odessa, Ukraine, May 26-28, 2021, volume 3048 of *CEUR Workshop Proceedings*, CEUR-WS.org, 2021, pp. 31–133. URL: http://ceur-ws.org/Vol-3048/paper03.pdf.
[37] A. O. Bielinskyi, S. V. Hushko, A. V. Matviychuk, O. A. Serdyuk, S. O. Semerikov, V. N. Soloviev, Irreversibility of financial time series: a case of crisis, in: A. E. Kiv, V. N. Soloviev, S. O. Semerikov (Eds.), Proceedings of the Selected and Revised Papers of 9th International Conference on Monitoring, Modeling & Management of Emergent Economy (M3E2-MLPEED 2021), Odessa, Ukraine, May 26-28, 2021, volume 3048 of *CEUR Workshop Proceedings*, CEUR-WS.org, 2021, pp. 134–150. URL: http://ceur-ws.org/Vol-3048/paper04.pdf.

[38] A. Bielinskyi, S. Semerikov, O. Serdyuk, V. Solovieva, V. N. Soloviev, L. Pichl, Econophysics of sustainability indices, in: A. Kiv (Ed.), 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, volume 2713 of *CEUR Workshop Proceedings*, CEUR-WS.org, 2020, pp. 372–392. URL: http://ceur-ws.org/Vol-2713/paper41.pdf.

[39] T. Fang, C. Zheng, D. Wang, Forecasting the crude oil prices with an EMD-ISBM-FNN model, Energy 263 (2023) 125407. doi:10.1016/j.energy.2022.125407.

[40] Y. Zou, L. Yu, K. He, Forecasting crude oil risk: A multiscale bidirectional generative adversarial network based approach, Expert Systems with Applications 212 (2023) 118743. doi:10.1016/j.eswa.2022.118743.

[41] H. Guliyev, E. Mustafayev, Predicting the changes in the WTI crude oil price dynamics using machine learning models, Resources Policy 77 (2022) 102664. doi:10.1016/j.resourpol.2022.102664.

[42] A. E. Kiv, V. N. Soloviev, S. O. Semerikov, H. B. Danylchuk, L. O. Kibalnyk, A. V. Matviychuk, A. M. Striuk, Machine learning for prediction of emergent economy dynamics III, in: A. E. Kiv, V. N. Soloviev, S. O. Semerikov (Eds.), Proceedings of the Selected and Revised Papers of 9th International Conference on Monitoring, Modeling & Management of Emergent Economy (M3E2-MLPEED 2021), Odessa, Ukraine, May 26-28, 2021, volume 3048 of *CEUR Workshop Proceedings*, CEUR-WS.org, 2021, pp. i–xxxi. URL: http://ceur-ws.org/Vol-3048/paper00.pdf.

[43] A. Bielinskyi, V. Soloviev, S. Semerikov, V. Solovieva, Identifying stock market crashes by fuzzy measures of complexity, Neiro-Nechitki Tekhnolohii Modelyuvannya v Ekonomitsi 2021 (2021) 3–45. doi:10.33111/nfmte.2021.003.

[44] M. Bondarenko, Modeling relation between at-the-money local volatility and realized volatility of stocks, Neiro-Nechitki Tekhnolohii Modelyuvannya v Ekonomitsi 2021 (2021) 46–66. doi:10.33111/nfmte.2021.046.

[45] T. Kmytiuk, G. Majore, Time series forecasting of agricultural product prices using Elman and Jordan recurrent neural networks, Neiro-Nechitki Tekhnolohii Modelyuvannya v Ekonomitsi 2021 (2021) 67–85. doi:10.33111/nfmte.2021.067.

[46] V. Kobets, O. Novak, EU countries clustering for the state of food security using machine learning techniques, Neiro-Nechitki Tekhnolohii Modelyuvannya v Ekonomitsi 2021 (2021) 86–118. doi:10.33111/nfmte.2021.086.

[47] H. Kucherova, Y. Honcharenko, D. Ocheretin, O. Bilska, Fuzzy logic model of usability of websites of higher education institutions in the context of digitalization of educational services, Neiro-Nechitki Tekhnolohii Modelyuvannya v Ekonomitsi 2021 (2021) 119–135. doi:10.33111/nfmte.2021.119.

[48] D. Lukianenko, I. Strelchenko, Neuromodeling of features of crisis contagion on financial markets between countries with different levels of economic development, Neiro-Nechitki Tekhnolohii Modelyuvannya v Ekonomitsi 2021 (2021) 136–163. doi:10.33111/nfmte.2021.136.

[49] I. Miroshnychenko, T. Kravchenko, Y. Drobyna, Forecasting electricity generation from renewable sources in developing countries (on the example of Ukraine), Neiro-Nechitki Tekhnolohii Modelyuvannya v Ekonomitsi 2021 (2021) 164–198. doi:10.33111/nfmte.

2021.164.

[50] S. Ashe, P. Egan, Examining financial and business cycle interaction using cross recurrence plot analysis, Finance Research Letters 51 (2023) 103461. doi:10.1016/j.frl.2022.103461.

[51] Q. He, J. Huang, A method for analyzing correlation between multiscale and multivariate systems—Multiscale multidimensional cross recurrence quantification (MMDCRQA), Chaos, Solitons & Fractals 139 (2020) 110066. doi:10.1016/j.chaos.2020.110066.

[52] M. C. Romano, M. Thiel, J. Kurths, W. von Bloh, Multivariate recurrence plots, Physics Letters A 330 (2004) 214–223. doi:10.1016/j.physleta.2004.07.066.

High-order network analysis for financial crash identification*

Andrii O. Bielinskyi[1,2], Vladimir N. Soloviev[1,3], Serhii V. Hushko[2], Arnold E. Kiv[4,5] and Andriy V. Matviychuk[3,1]

[1]Kryvyi Rih State Pedagogical University, 54 Gagarin Ave., Kryvyi Rih, 50086, Ukraine
[2]State University of Economics and Technology, 16 Medychna Str., Kryvyi Rih, 50005, Ukraine
[3]Kyiv National Economic University named after Vadym Hetman, 54/1 Peremogy Ave., Kyiv, 03680, Ukraine
[4]Ben-Gurion University of the Negev, P.O.B. 653, Beer Sheva, 8410501, Israel
[5]South Ukrainian National Pedagogical University named after K. D. Ushynsky, 26 Staroportofrankivska Str., Odesa, 65020, Ukraine

Abstract
Network analysis is a powerful method to characterize the complexity and dynamics of socio-economic systems. However, traditional network analysis often ignores the higher-order dependencies that arise from the interactions of more than two nodes. In this paper, we propose to use high-order networks, which are generalized network structures that capture the higher-order dependencies, to study the temporal evolution of the Dow Jones Industrial Average (DJIA) index. We construct high-order networks from the DJIA time series using the visibility graph method, and we measure the topological complexity of the high-order networks using various metrics. We find that the complexity of the system changes drastically during crisis events, indicating that high-order network analysis can be used as an indicator (indicator-precursor) of financial crashes. We also show that high-order network analysis and topology can provide more insights into the nonlinear and nonstationary behavior of the DJIA index than traditional tools of financial time series analysis.

Keywords
high-order network analysis, financial crash identification, complex networks, multiplex networks, visibility graph, indicator-precursor

1. Introduction

The proliferation of extensive and finely-grained data, often with temporal resolution, has unlocked unprecedented opportunities to dissect the behaviors of complex systems spanning diverse domains such as biology, technology, finance, and economics [2, 3, 4]. These intricate systems, comprising myriad interacting units, frequently exhibit emergent properties at

M3E2-MLPEED 2022: The 10th International Conference on Monitoring, Modeling & Management of Emergent Economy, November 17-18, 2022, Virtual, Online
*Extended and revised version of paper [1] presented at the 10th International Conference on Monitoring, Modeling & Management of Emergent Economy.
krivogame@gmail.com (A. O. Bielinskyi); vnsoloviev2016@gmail.com (V. N. Soloviev); gushko77@gmail.com (S. V. Hushko); kiv.arnold20@gmail.com (A. E. Kiv); editor@nfmte.com (A. V. Matviychuk)
0000-0002-2821-2895 (A. O. Bielinskyi); 0000-0002-4945-202X (V. N. Soloviev); 0000-0002-4833-3694 (S. V. Hushko); 0000-0002-0991-2343 (A. E. Kiv); 0000-0002-8911-5677 (A. V. Matviychuk)

macroscopic scales due to heterogeneous interactions among their constituents [5]. Complex networks have emerged as a formidable toolset to analyze the structures and dynamics of such systems [5]. However, the tools conventionally employed in network analysis often focus on interactions between pairs of nodes, a limitation at odds with the increasing availability of empirical data illustrating group interactions within heterogeneous systems [6]. Thus, it becomes evident that interactions within systems often extend beyond dyadic connections, manifesting as collective actions involving groups of nodes [7, 8], exerting a notable influence on the interacting systems' dynamics [9, 10].

The notion of higher-order interactions finds historical roots in solid-state physics, where multiparticle potentials and quantum mechanical calculations supplanted paired interactions. Similarly, in thermodynamics and statistical physics, Tsallis introduced nonextensive interactions [11, 12]. However, in contrast to these simpler representations of higher-order interactions, complexity in complex systems demands more intricate mathematical structures like hypergraphs and simplicial complexes.

Diverse models of higher-order networks have surfaced [13], reflecting the growing importance of this domain. Here, we briefly highlight key models that have garnered attention [14, 15, 16].

Multiplex Networks: Multiplex networks, multilayer networks, and networks of networks capture interactions between various entities and have found applicability in systems with diverse interaction types [17]. However, most interactions remain dyadic and can be represented through traditional networks [18]. Their application in financial analysis is well-documented [19, 20, 21, 22, 23, 24, 25, 26, 27, 20] alongside higher-order networks [28, 29, 30, 31, 32, 33, 34, 35].

Hypergraphs and Simplicial Complexes: Algebraic topology's computational techniques, hypergraphs, and simplicial complexes encode units and hyperlinks, allowing explicit consideration of systems beyond pairwise interactions [9, 36, 7, 37].

Higher-Order Markov Models: First-order Markov models have gained traction in describing flows of information, energy, money, etc. within networks [38]. However, many flows exhibit path-dependent behaviors, necessitating higher-order Markov chain models [16].

Higher-Order Graphical Models and Markov Random Fields: Markov random fields, including the Ising model, extended to higher-order models, capture interactions between multiple objects [39, 40].

Recently, Santoro et al. [36] introduced a structure to characterize instantaneously co-fluctuating [41] signal patterns of all interaction orders. They showcased that higher-order measures discern subtleties in space-time regimes in diverse studies: brain activity, stock option prices, and epidemics. In this context, we explore the application of multiplex and higher-order network techniques to model crisis states in the stock market. Section 2 introduces a graph representation based on the visibility graph, while Section 3 presents multiplex networks' theory, including measures. Section 4 elaborates on higher-order networks and encoding methods, describing measures for both classical and high-order networks. Empirical results, including a comparative analysis of measures, are presented in Section 5. Finally, Section 6 outlines our conclusions and future directions.

2. Visibility graph

Visibility graph (VG), which was proposed by Lacasa et al. [42] is typically constructed from a univariate time series. In a visibility graph, each moment in the time series maps to a node in the network, and an edge exists between the nodes if they satisfy a "mutual visibility" condition.

"Mutual visibility" can be understood by imagining two points x_i at time t_i and x_j at time t_j as two hills of a time series, which can be understood as a landscape, and these two points are "mutually visible" if x_i has no any obstacles in the way on x_j. Formally, two points are mutually visible if, all values of x_k between t_i and t_j satisfy:

$$x_k < x_i + \frac{t_k - t_i}{t_j - t_i}\left[x_j - x_i\right], \quad \forall k : i < k < j \tag{1}$$

Horizontal visibility graph (HVG) [43] is a restriction of usual visibility graph, where two points x_i and x_j are connected if there can be drawn a *horizontal* path that does not intersect an intermediate point x_k, $i < k < j$. Equivalently, node x_i at time t_i and node x_j at time t_j are connected if the horizontal ordering criterion is fulfilled:

$$x_k < \inf(x_i, x_j), \quad \forall k : i < k < j. \tag{2}$$

Figure 1 is an approximate illustration of the construction of visibility graphs.

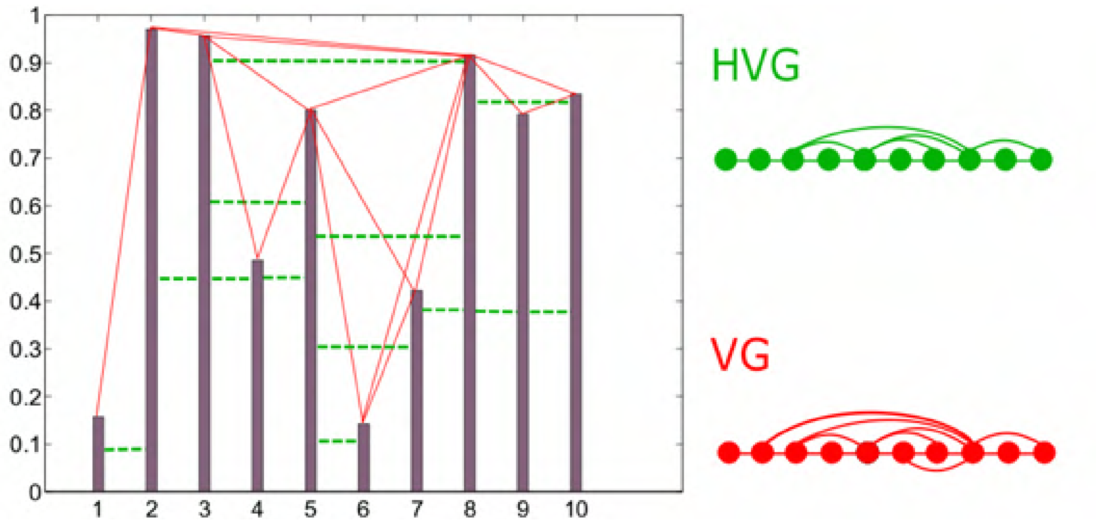


Figure 1: Schematic illustration of the VG (red lines) and the HVG (green lines). Adapted from [44].

3. Multiplex orderness and measures of complexity

Multiplex network [45] is the representation of the system which consists of the variety of different subnetworks with inter-network connections. For working with multiplex financial networks, we set two tasks:

- convert separated time series into network that represent a layer of a multiplex network. The procedure of conversion is presented in section 2;
- create intra-layer connection between each subnetwork.

Figure 2 represents an algorithm for creating a three-layered multiplex visibility graph.

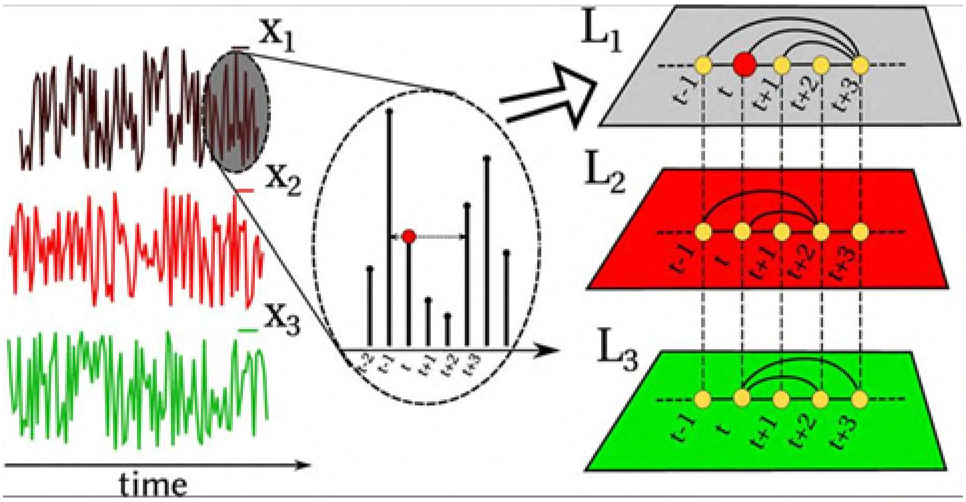


Figure 2: Illustration of the multiplex VG formation on the example of three layers. Adapted from [46].

Multiplex network is the representation of a pair $M = (G, C)$, where $\{G_\alpha \,|\, \alpha \in 1, \ldots, M\}$ is a set of graphs $G_\alpha = (X_\alpha, E_\alpha)$ that called layers and

$$C = \left\{E_{\alpha\beta} \subseteq X_\alpha \times X_\beta \,|\, \alpha, \beta \in 1, \ldots, M, \alpha \neq \beta\right\} \tag{3}$$

is a set of intra-links in layers G_α and G_β ($\alpha \neq \beta$). E_α is intra-layer edge in M, and each $E_{\alpha\beta}$ is denoted as inter-layer edge.

A set of nodes in a layer G_α is denoted as $X_\alpha = \left\{x_1^\alpha, \ldots, x_{N_\alpha}^\alpha\right\}$, and an intra-layer adjacency matrix as $A^{[\alpha]} = (a_{ij}^\alpha) \in \mathrm{Re}^{N_\alpha \times N_\alpha}$, where

$$\alpha_{ij}^\alpha = \begin{cases} 1, & (x_i^\alpha, x_j^\alpha) \in E_\alpha, \\ 0. & \end{cases} \tag{4}$$

for $1 \leq i \leq N_\alpha$, $1 \leq j \leq N_\beta$ and $1 \leq \alpha \leq M$. For an inter-layer adjacency matrix, we have $A^{[\alpha,\beta]}(a_{ij}^{\alpha\beta}) \in \mathrm{Re}^{N_\alpha \times N_\beta}$, where

$$\alpha_{ij}^{\alpha\beta} = \begin{cases} 1, & (x_i^\alpha, x_j^\beta) \in E_{\alpha\beta}, \\ 0. & \end{cases} \tag{5}$$

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

For a multiplex network, the node degree k is already a vector

$$k_i = (k_i^{[1]}, \dots, k_i^{[M]}), \tag{6}$$

with the degree $k_i^{[\alpha]}$ of the node i in the layer α, namely

$$k_i^{[\alpha]} = \sum_j a_{ij}^{[\alpha]}, \tag{7}$$

while $a_{ij}^{[\alpha]}$ is the element of the adjacency matrix of the layer α. Specificity of the node degree in vector form allows describing additional quantities. One of them is the *overlapping degree* of node i:

$$o_i = \sum_{\alpha=1}^{M} k_i^{[\alpha]}. \tag{8}$$

The next measure quantitatively describes the inter-layer information flow. For a given pair (α, β) within M layers and the degree distributions $P(k^{[\alpha]}), P(k^{[\beta]})$ of these layers, we can defined the so-called *interlayer mutual information*:

$$I_{\alpha,\beta} = \sum \sum P(k^{[\alpha]}, k^{[\beta]}) \log \frac{P(k^{[\alpha]}, k^{[\beta]})}{P(k^{[\alpha]}), P(k^{[\beta]})}, \tag{9}$$

where $P(k^{[\alpha]}, k^{[\beta]})$ is the joint probability of finding a node degree $k^{[\alpha]}$ in a layer α and a degree $k^{[\beta]}$ in a layer β. The higher the value of $I_{\alpha,\beta}$, the more correlated (or anti-correlated) is the degree distribution of the two layers and, consequently, the structure of a time series associated with them. We also find the mean value of $I_{\alpha,\beta}$ for all possible pairs of layers – the scalar $\langle I_{\alpha,\beta} \rangle$ that quantifies the information flow in the system.

The *multiplex degree entropy* is another multiplex measure which quantitatively describes the distribution of a node degree i between different layers. It can be defined as

$$S_i = -\sum_{\alpha=1}^{M} \frac{k_i^{[\alpha]}}{o_i} \log \frac{k_i^{[\alpha]}}{o_i}. \tag{10}$$

Entropy is close to zero if ith node degree is within one special layer of a multiplex network, and it has the maximum value when ith node degree is uniformly distributed between different layers.

4. High-order extension of temporal networks

4.1. Time-respecting paths

Financial networks are strongly influenced by the ordering and timing of links. In their context of their temporality, we must consider *time-respecting paths*, an extension of the concept of paths in static network topologies which additionally respects the timing and ordering of time-stamped links [47, 48, 49]. For a source node v and a target node w, a time-respecting path can be presented by any sequence of time-stamped links

$$(v_0, v_1; t_1), (v_1, v_2; t_2), ..., (v_{l-1}, v_l; t_l), \tag{11}$$

where $v_0 = v, v_l = w$ and $t_1 < t_2 < ... < t_l$. Time ordering of temporal financial networks is important since it implies causality, i.e. a node i is able to influence node j relying on two time-stamped links (i, k) and (k, j) only if edge (i, k) has occurred before edge (k, j).

Apart the restriction on networks to have the correct ordering, it is common to impose a maximum time difference between consecutive edges [50], i.e. there is a maximum time difference δ and, example, two time-stamped edges $(i, k; t)$ and $(k, j; t')$ that contribute to a time-respecting path if $0 \leq t' - t \leq \delta$. If $\delta = 1$, we are usually interested in paths with short time scales. For $\delta = \infty$, we impose no restrictions on time-range and consider a path definition where links can be weeks or years apart.

4.2. High-order networks

The key idea behind this abstraction is that the commonly used time-aggregated network is the simplest possible time-aggregated representation, whose weighted links capture the frequencies of time-stamped links. Considering that each time-stamped link is a time-respecting path of length one, it is easy to generalize this abstraction to higher-order time-aggregate networks in which weighted links capture the frequencies of longer time-respecting paths.

There are several variants for encoding high-order interactions [10]. The first concept of high-order links represent *hyperlink*, which can contain any number of nodes. *Hypergraph* is the generalized notion of network which is composed of nodeset V and hyper-edges E that specify which nodes from V participate in which way.

Simplex is another mathematical abstraction to accomplish high-order interaction. Formally, a *k-simplex* σ is a set of $k + 1$ fully interacting nodes $\sigma = [v_0, v_1, ..., v_k]$. Essentially, a node is 0-simplex, a link is 1-simplex, a triangle is 2-simplex, a tetrahedron is 3-simplex, etc. Since a standard graph is a collection of edges, *simplicial complexes* are collections of simplices $K = \{\sigma_0, \sigma_1, ..., \sigma_n\}$.

Figure 3 demonstrates examples of simplices and hyperlinks of orders 1, 2, and 3.

For a temporal network $G^T = (V^T, E^T)$ we thus formally define a kth order time-aggregated (or simply aggregate) network as a tuple $G^{(k)} = (V^{(k)}, E^{(k)})$ where $V^{(k)} \subseteq V^k$ is a set of node k-tuples and $E^{(k)} \subseteq V^{(k)} \times V^{(k)}$ is a set of links. For simplicity, we call each of the k-tuples $v = v_1 - v_2 - ... - v_k$ ($v \in V^{(k)}, v_i \in V$) a kth order node, while each link $e \in E^{(k)}$ is called a kth order link. Between two kth order nodes v and w exists kth order edge (v, w) if they overlap in exactly $k - 1$ elements. Resembling so-called De Bruijn graphs [51], the basic idea behind this construction is that each kth order link represents a possible time-respecting path of length k in the underlying temporal network, which connects node v_1 to node w_k via k time-stamped links

$$(v_1, v_2 = w_1; t_1), ..., (v_k = w_{k-1}, w_k; t_k). \tag{12}$$

Importantly, and different from a first-order representation, kth order aggregate networks allow to capture *non-Markovian* characteristics of temporal networks. In particular, they allow to represent temporal networks in which the kth time-stamped link $(v_k = w_{k-1}, w_k)$ on a time-respecting path depends on the $k - 1$ previous time-stamped links on this path. With this, we obtain a simple static network topology that contains information both on the presence of time-stamped links in the underlying temporal network, as well as on the ordering in which sequences of k of these time-stamped links occur.

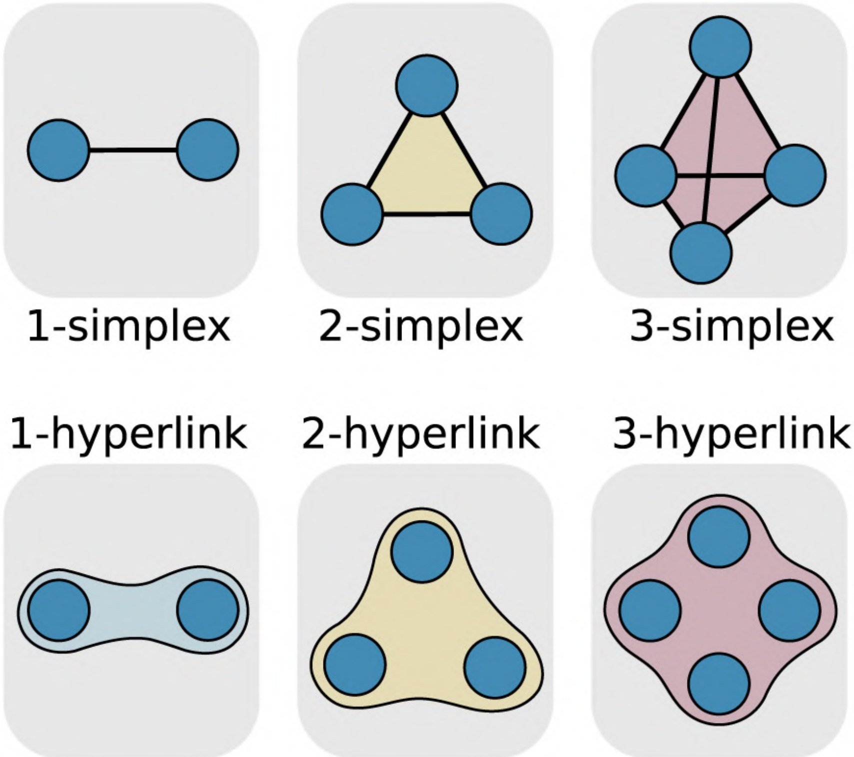


Figure 3: High-order connections in terms of simplices and hyperlinks. Adapted from [9].

4.3. Degree centrality

Network centralities are node-related measures that quantify how “central'' a node is in a network. There are many ways in which a node can be considered so: for example, it can be central if it is connected to many other nodes (degree centrality), or relatively to its connectivity to the rest of the network (path based centralities, eigenvector centrality). One of the simplest centrality measure is the *degree of a node*, which counts the number of edges incident to an ith node.

For any adjacency matrix the degree of a node i can be defined as

$$D_i = \sum_j A_{ij}. \tag{13}$$

High-order degree centrality counts the number of kth-order edges incident to the kth-order node i. To get a scalar value which will serve as an indicator of high-order dynamics, we obtain mean degree D_{mean}:

$$D_{mean} = \frac{1}{N} \sum_{i=1}^{N} D_i. \tag{14}$$

Except this measure, we can calculate nth moment of the degree distribution, which can be defined as

$$\langle k^n \rangle = \sum_{k_{\min}}^{\infty} k^n p_k \approx \int_{k_{\min}}^{\infty} k^n p_k dk. \quad (15)$$

In this study we will present the dynamics of the first moment, which is the mean weighted degree of a network, and its high-order behavior.

4.4. Assortativity coefficient

Assortativity is a property of network nodes that characterizes the degree of connectivity between them. Many networks demonstrate "assortative mixing" on their nodes, when high-degree nodes tend to be connected to other high-degree nodes. Other networks demonstrate disassortative mixing when their high-degree nodes tend to be connected to low-degree nodes. Assortativity of a network can be defined via the Pearson correlation coefficient of the degrees at either ends of an edge. For an observed network, we can write it as

$$r = \frac{M^{-1}\sum_i j_i k_i - \left[M^{-1}\sum_i \frac{1}{2}(j_i + k_i)\right]^2}{M^{-1}\sum_i \frac{1}{2}(j_i^2 + k_i^2) - \left[M^{-1}\sum_i \frac{1}{2}(j_i + k_i)\right]^2}, \quad (16)$$

where $-1 \leq r \leq 1$; j_i, k_i are the degrees of the nodes at the ends of the ith edge, with $i = 1, ..., M$, where M is the number of edges of a network.

This correlation function is zero for no assortative mixing. If $r = 1$, then we have perfect assortative mixing pattern. For $r = -1$, we can observe perfect disassortativity.

Studying financial networks, with time-respecting paths, we can consider four type of assortativity: $r(in, in), r(in, out), r(out, in), r(out, out)$, which will correspond to tendencies to have similar in and out degrees. We can denote one of the studied in/out pairs as (α, β). Suppose, for a given ith edge, we have got the source (i.e. tail) node of the edge and target (i.e. head) node of the edge. We can denote them as α-degree of the source (j_i^α) and β-degree of the target (k_i^β). Assortativity coefficient for degrees of a specific type can be defined as

$$r(\alpha, \beta) = \frac{\sum_i \left(j_i^\alpha - \overline{j^\alpha}\right)\left(k_i^\beta - \overline{k^\beta}\right)}{\sqrt{\sum_i \left(j_i^\alpha - \overline{j^\alpha}\right)^2}\sqrt{\sum_i \left(k_i^\beta - \overline{k^\beta}\right)^2}}, \quad (17)$$

where $\overline{j^\alpha}$ and $\overline{k^\beta}$ are the average α-degree of sources and β-degree of targets.

5. Empirical results

To build indicators (indicators-precursors) based on multiplex and high-order networks, the following is done:

- databases of 6 most influential stock market indices for the period from 02.01.2004 to 18.10.2022 were selected for multiplex analysis (see figure 4). The data were extracted using Yahoo! Finance API based on Python programming language [52];
- the indicators described in the previous sections were calculated using the sliding window procedure [12, 53, 54, 55, 56, 57, 58]. The essence of this procedure is that: (1) a fragment (window) of a series of a certain length w was selected; (2) a network measure was calculated for it; (3) the measure values were stored in a pre-declared array; (4) the window was shifted by a predefined time step h, and the procedure was repeated until the series was completely exhausted; (5) further, the calculated values of the network measure were compared with the dynamics of the stock index. Subsequently, conclusions were drawn regarding the further dynamics of the market. In our case, window length $w = 500$ days and time step $h = 10$ day. The choice of step was limited by the counting time for high-order networks;
- multiplex and high-order indicators are compared with the Dow Jones Industrial Average (DJIA) index.

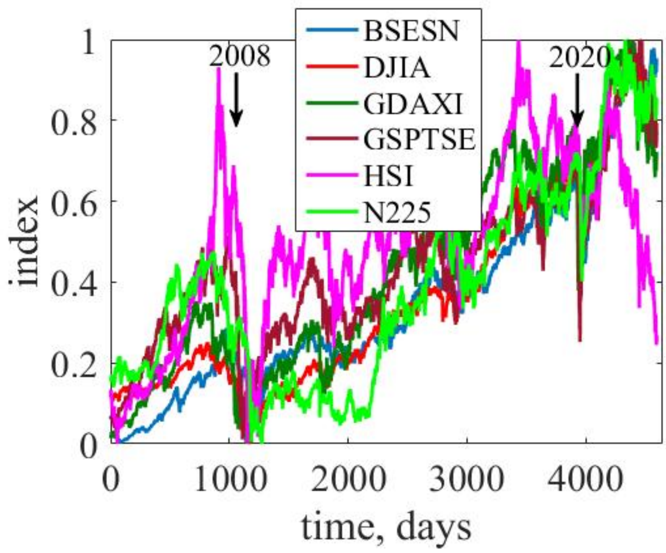


Figure 4: The dynamics of stock market indices for studying multiplex characteristics.

In figure 5 presented the dynamics of inter-layer mutual information (I) and multiplex degree entropy (S) along with the DJIA index.

From figure 5 we can see that multiplex mutual information increases before the crisis of 2008. Also, it noticeably becomes higher before COVID-19 crash. For the last months, it demonstrates decreasing pattern, which indicates that the economies of different countries may be experiencing different evolutions now. Nevertheless, it can be seen that, as a rule, this indicator is characterized by growth, indicating an increase in the interconnection of the economies of different countries. In a crisis, this indicator usually declines, demonstrating different resistance to the collapse events of the stock markets of countries and the difference in the actions that they take. Entropy indicator shows asymmetric behavior.

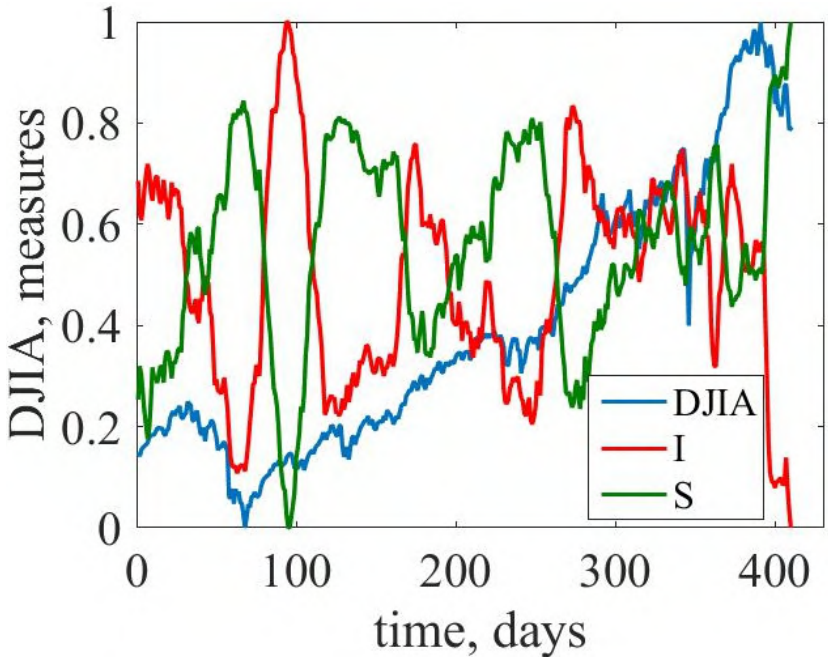


Figure 5: The dynamics of inter-layer mutual information (I) and multiplex degree entropy (S) along with the DJIA index.

Next, we compare one of the multiplex measure, overlapping degree (o), with the mean degree of a network (D_{mean}). Figure 6 represents this result.

In figure 6 we can see that both D_{mean} and o are characterized by similar dynamics. These indicators increase near the crash, which indicates an increase in the concentration of connections for some network nodes, and further, based on the indicators during the crisis, there is a decline in concentration both in the dynamics of the DJIA and the inter-layer connectedness of stock indices. We may see that the multiplex approach does not significantly change the dynamics of the concentration degree indicator in comparison with the indicator based on the classical univariate graph.

Figure 7 demonstrates the dynamics of mean weighted degree (equation (15)) for order 1 and 2 along with the DJIA index.

In figure 7 we can see that the second-order D_{mean} is slightly different from the first-order one. The second-order D_{mean} starts to increase a slightly earlier before the crisis of 2008. We can see that before crisis of 2020 second-order D_{mean} declines more noticeably comparing to the first-order one. However, this difference between the first and second order is still insignificant, what can we say about the fact that the classical visibility graph can reflect all the information that the series under study can represent.

Next, let us present high-order dynamics of the assortativity coefficient for the DJIA index (see figure 8).

Figure 8 presents the assortativity coefficient for first, second, and third orders. Assortativity declines before crashes and increases during them. We see that high-orderness does not change

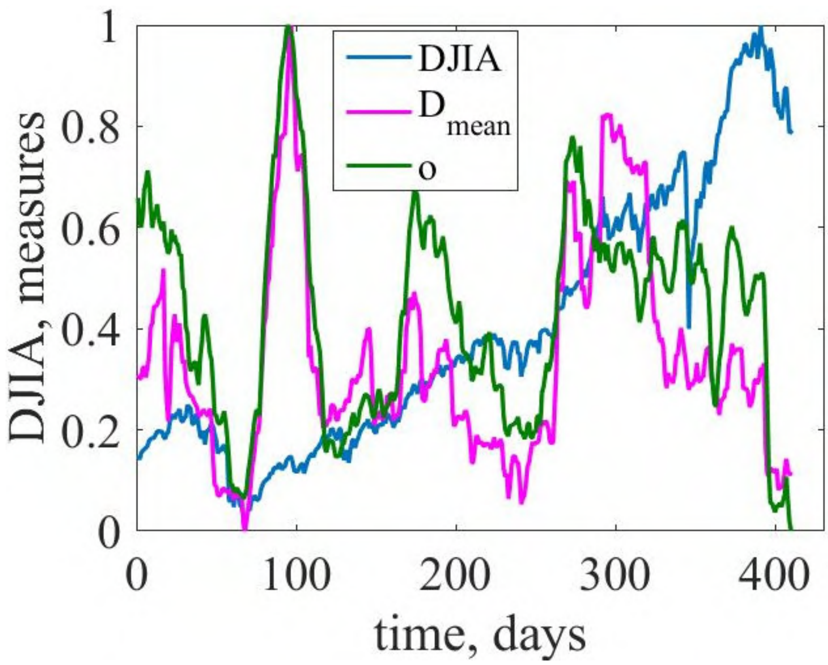


Figure 6: The dynamics of the mean degree (D_{mean}) and overlapping degree (o) along with the DJIA index.

radically change the dynamics of this indicator. Third-order assortativity responds better for the crash of 2008, but worse for the COVID-19 crisis, comparing to first- and second-order assortativity.

6. Conclusions

In this article, we have introduced methods to measure and model systems with causal, multiplex, and high-order interactions. We have shown that these methods can capture the long-range spatio-temporal correlations that characterize non-Markovian, non-stationary, non-linear systems, which are better described by the high-order paradigm. We have used hypergraphs [59, 60, 61] and simplicial complexes [62, 63, 64] as richer types of links that allow us to go beyond typical nodes and encode higher-order clusters and temporal dependencies.

We have presented indicators (indicators-precursors) based on classic visibility graphs, multiplex networks, and high-order networks. We have applied these indicators to the time series of the Dow Jones Industrial Average (DJIA) index and a database of six stock indices from different countries and sectors. We have used the sliding window algorithm to calculate various network measures, such as the mean degree of a node (D_{mean}), the first-moment degree of a network, the assortativity coefficient, the inter-layer mutual information (I), the multiplex degree entropy (S), and the mean overlapping degree of a network (o). We have found that multiplex and high-order networks do not differ significantly from the traditional pairwise visibility model in terms of their dynamics. This may suggest that the classical visibility graph reflects all possible

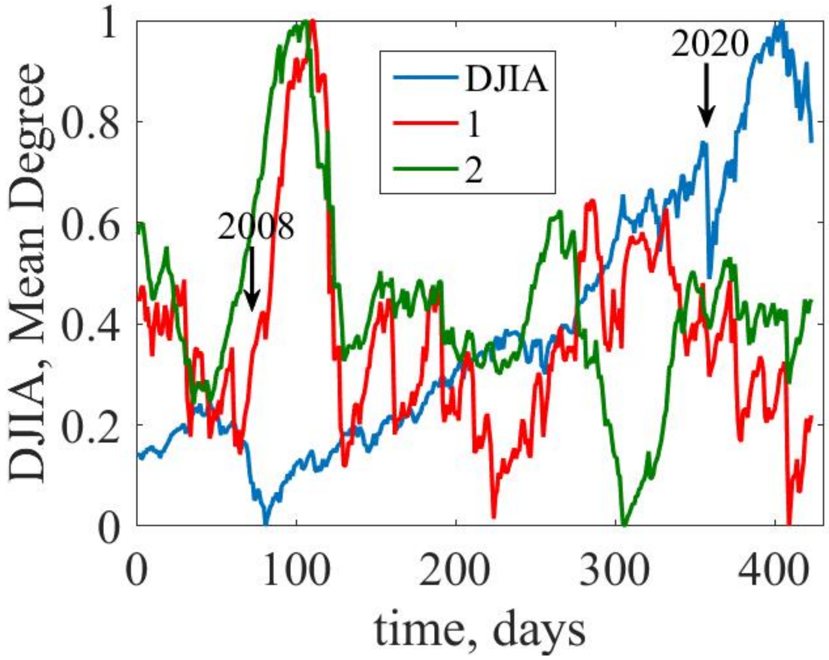


Figure 7: The dynamics of first- and second-order mean (weighted) degree along with the DJIA index.

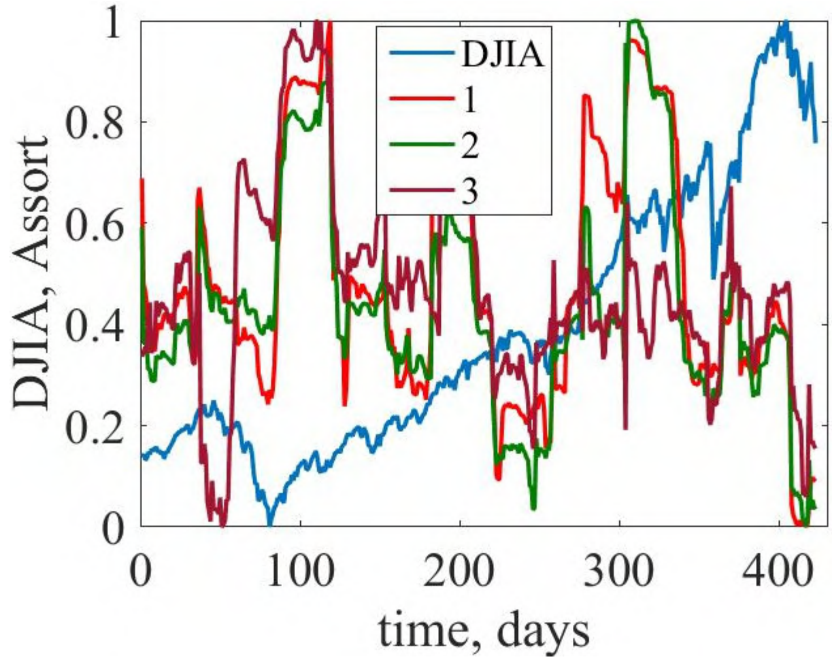


Figure 8: The dynamics of first-, second-, and third-order assortativity along with the DJIA index.

short-term and long-term dependencies in the DJIA index. We have also found that all the presented measures work similarly as indicators (indicators-precursors) of critical financial events, increasing or decreasing before and during them. However, multiplex and high-order network indicators still need further development and improvement for studying complex financial time series. A possible solution may be to combine Markov chains of multiple, higher orders into a multi-layer graphical model that captures temporal correlations in pathways at multiple length scales simultaneously [65]. Another perspective may be to use neuro-fuzzy forecasting and clustering methods of complex financial systems [66, 67, 68, 69, 70, 71, 72].

Acknowledgments

This work was supported by the Ministry of Education and Science of Ukraine (project No. 0122U001694).

References

[1] A. Bielinskyi, V. Soloviev, S. Hushko, A. Kiv, A. V. Matviychuk, High-Order Networks and Stock Market Crashes, in: S. Semerikov, V. Soloviev, A. Matviychuk, V. Kobets, L. Kibalnyk, H. Danylchuk, A. Kiv (Eds.), Proceedings of 10th International Conference on Monitoring, Modeling & Management of Emergent Economy - M3E2, INSTICC, SciTePress, 2023, pp. 134–144. doi:10.5220/0011931900003432.

[2] A.-L. Barabási, M. Pósfai, Network science, 1st ed., Cambridge University Press, Cambridge, 2016. URL: http://barabasi.com/networksciencebook/.

[3] V. Latora, V. Nicosia, G. Russo, Complex Networks: Principles, Methods and Applications, 1st ed., Cambridge University Press, USA, 2017. URL: http://www.complex-networks.net.

[4] O. Burov, Design features of the synthetic learning environment, Educational Technology Quarterly 2021 (2021) 689–700. doi:10.55056/etq.43.

[5] M. E. J. Newman, Networks: an introduction, 2nd ed., Oxford University Press, Oxford; New York, 2010. URL: https://global.oup.com/academic/product/networks-9780198805090.

[6] F. Battiston, G. Petri (Eds.), Higher-Order Systems (Understanding Complex Systems), 1st ed., Springer International Publishing, 2022. doi:10.1007/978-3-030-91374-8.

[7] F. Battiston, E. Amico, A. Barrat, G. Bianconi, G. F. de Arruda, B. Franceschiello, I. Iacopini, S. Kéfi, V. Latora, Y. Moreno, M. M. Murray, T. P. Peixoto, F. Vaccarino, G. Petri, The physics of higher-order interactions in complex systems, Nature Physics 17 (2021) 1093–1098. doi:10.1038/s41567-021-01371-4.

[8] H. Sun, G. Bianconi, Higher-order percolation processes on multiplex hypergraphs, Physical Review E 104 (2021). doi:10.1103/physreve.104.034306.

[9] F. Battiston, G. Cencetti, I. Iacopini, V. Latora, M. Lucas, A. Patania, J.-G. Young, G. Petri, Networks beyond pairwise interactions: Structure and dynamics, Physics Reports 874 (2020) 1–92. doi:10.1016/j.physrep.2020.05.004.

[10] S. Majhi, M. Perc, D. Ghosh, Dynamics on higher-order networks: a review, Journal of The Royal Society Interface 19 (2022) 20220043. doi:10.1098/rsif.2022.0043.

[11] M. L. Lyra, C. Tsallis, Nonextensivity and Multifractality in Low-Dimensional Dissipative Systems, Phys. Rev. Lett. 80 (1998) 53–56. doi:10.1103/PhysRevLett.80.53.
[12] A. O. Bielinskyi, A. V. Matviychuk, O. A. Serdyuk, S. O. Semerikov, V. V. Solovieva, V. N. Soloviev, Correlational and Non-extensive Nature of Carbon Dioxide Pricing Market, in: O. Ignatenko, V. Kharchenko, V. Kobets, H. Kravtsov, Y. Tarasich, V. Ermolayev, D. Esteban, V. Yakovyna, A. Spivakovsky (Eds.), ICTERI 2021 Workshops, volume 1635 CCIS of *Communications in Computer and Information Science*, Springer International Publishing, Cham, 2022, pp. 183–199. doi:10.1007/978-3-031-14841-5_12.
[13] O. Bobrowski, D. Krioukov, Random Simplicial Complexes: Models and Phenomena, in: Understanding Complex Systems, Springer International Publishing, 2022, pp. 59–96. doi:10.1007/978-3-030-91374-8_2.
[14] A. R. Benson, R. Abebe, M. T. Schaub, A. Jadbabaie, J. Kleinberg, Simplicial closure and higher-order link prediction, Proceedings of the National Academy of Sciences 115 (2018) E11221–E11230. doi:10.1073/pnas.1800683115.
[15] C. Bick, E. Gross, H. A. Harrington, M. T. Schaub, What are higher-order networks?, 2021. URL: https://arxiv.org/abs/2104.11329.
[16] R. Lambiotte, M. Rosvall, I. Scholtes, From networks to optimal higher-order models of complex systems, Nat. Phys. 15 (2019) 313–320. doi:10.1038/s41567-019-0459-y.
[17] S. Boccaletti, G. Bianconi, R. Criado, C. del Genio, J. Gómez-Gardeñes, M. Romance, I. Sendiña-Nadal, Z. Wang, M. Zanin, The structure and dynamics of multilayer networks, Physics Reports 544 (2014) 1–122. doi:10.1016/j.physrep.2014.07.001.
[18] P. S. Skardal, L. Arola-Fernández, D. Taylor, A. Arenas, Higher-order interactions can better optimize network synchronization, Phys. Rev. Res. 3 (2021) 043193. doi:10.1103/PhysRevResearch.3.043193.
[19] M. Bardoscia, P. Barucca, S. Battiston, F. Caccioli, G. Cimini, D. Garlaschelli, F. Saracco, T. Squartini, G. Caldarelli, The physics of financial networks, Nature Reviews Physics 3 (2021) 490–507. doi:10.1038/s42254-021-00322-5.
[20] R. M. del Rio-Chanona, Y. Korniyenko, M. Patnam, M. A. Porter, The multiplex nature of global financial contagions, Applied Network Science 5 (2020). doi:10.1007/s41109-020-00301-2.
[21] A. Serguieva, Multichannel contagion vs stabilisation in multiple interconnected financial markets, SSRN Electronic Journal (2016). doi:10.2139/ssrn.2904431.
[22] C. D. Brummitt, T. Kobayashi, Cascades in multiplex financial networks with debts of different seniority, Phys. Rev. E 91 (2015) 062813. doi:10.1103/PhysRevE.91.062813.
[23] J. Cao, F. Wen, H. E. Stanley, X. Wang, Multilayer financial networks and systemic importance: Evidence from China, International Review of Financial Analysis 78 (2021) 101882. doi:10.1016/j.irfa.2021.101882.
[24] Y. Xie, F. Jiao, S. Li, Q. Liu, Y. Tse, Systemic risk in financial institutions: A multiplex network approach, Pacific-Basin Finance Journal 73 (2022) 101752. doi:10.1016/j.pacfin.2022.101752.
[25] Q. Gao, Systemic Risk Analysis of Multi-Layer Financial Network System Based on Multiple Interconnections between Banks, Firms, and Assets, Entropy 24 (2022) 1252. doi:10.3390/e24091252.
[26] I. Aldasoro, I. Alves, Multiplex interbank networks and systemic importance: An applica-

tion to European data, Journal of Financial Stability 35 (2018) 17–37. doi:10.1016/j.jfs.2016.12.008.

[27] T. Squartini, A. Gabrielli, D. Garlaschelli, T. Gili, A. Bifone, F. Caccioli, Complexity in Neural and Financial Systems: From Time-Series to Networks, Complexity 2018 (2018) 13. doi:10.1155/2018/3132940.

[28] S. K. Stavroglou, A. A. Pantelous, H. E. Stanley, K. M. Zuev, Hidden interactions in financial markets, Proceedings of the National Academy of Sciences 116 (2019) 10646–10651. doi:10.1073/pnas.1819449116.

[29] M. O. Jackson, A. Pernoud, Systemic Risk in Financial Networks: A Survey, Annual Review of Economics 13 (2021) 171–202. doi:10.1146/annurev-economics-083120-111540.

[30] S. Saha, J. Gao, R. Gerlach, A survey of the application of graph-based approaches in stock market analysis and prediction, International Journal of Data Science and Analytics 14 (2022) 1–15. doi:10.1007/s41060-021-00306-9.

[31] S. Battiston, G. Caldarelli, R. M. May, T. Roukny, J. E. Stiglitz, The price of complexity in financial networks, Proceedings of the National Academy of Sciences 113 (2016) 10031–10036. doi:10.1073/pnas.1521573113.

[32] K. Huremovic, G. Jiménez, E. Moral-Benito, J.-L. Peydró, F. Vega-Redondo, Production and Financial Networks in Interplay: Crisis Evidence from Supplier-Customer and Credit Registers, Working Papers 1191, Barcelona School of Economics, 2020. URL: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3696358.

[33] F. Franch, L. Nocciola, A. Vouldis, Temporal Networks in the Analysis of Financial Contagion, SSRN Electronic Journal (2022). doi:10.2139/ssrn.4125870.

[34] P. Bartesaghi, G. P. Clemente, R. Grassi, A tensor-based unified approach for clustering coefficients in financial multiplex networks, Information Sciences 601 (2022) 268–286. doi:10.1016/j.ins.2022.04.021.

[35] B. Han, Y. Wei, L. Kang, Q. Wang, Y. Yang, Node Classification in Attributed Multiplex Networks Using Random Walk and Graph Convolutional Networks, Frontiers in Physics 9 (2022). doi:10.3389/fphy.2021.763904.

[36] A. Santoro, F. Battiston, G. Petri, E. Amico, Unveiling the higher-order organization of multivariate time series, 2022. doi:10.48550/arxiv.2203.10702.

[37] C. Berge, Graphs and Hypergraphs, North-Holland mathematical library, North-Holland Publishing Company, 1976. URL: http://compalg.inf.elte.hu/~tony/Oktatas/Algoritmusok-hatekonysaga/Berge-hypergraphs.pdf.

[38] N. Masuda, M. A. Porter, R. Lambiotte, Random walks and diffusion on networks, Physics Reports 716-717 (2017) 1–58. doi:10.1016/j.physrep.2017.07.007.

[39] Y. Shemesh, Y. Sztainberg, O. Forkosh, T. Shlapobersky, A. Chen, E. Schneidman, High-order social interactions in groups of mice, eLife 2 (2013) e00759. doi:10.7554/eLife.00759.

[40] N. Komodakis, N. Paragios, Beyond pairwise energies: Efficient optimization for higher-order MRFs, in: 2009 IEEE Conference on Computer Vision and Pattern Recognition, 2009, pp. 2985–2992. doi:10.1109/CVPR.2009.5206846.

[41] L. Wasserman, Topological Data Analysis, Annual Review of Statistics and Its Application 5 (2018) 501–532. doi:10.1146/annurev-statistics-031017-100045.

[42] L. Lacasa, B. Luque, F. Ballesteros, J. Luque, J. C. Nuño, From time series to complex networks: The visibility graph, Proceedings of the National Academy of Sciences 105

(2008) 4972–4975. doi:10.1073/pnas.0709247105.
[43] B. Luque, L. Lacasa, F. Ballesteros, J. Luque, Horizontal visibility graphs: Exact results for random time series, Phys. Rev. E 80 (2009) 046103. doi:10.1103/PhysRevE.80.046103.
[44] J. Iacovacci, L. Lacasa, Sequential motif profile of natural visibility graphs, Phys. Rev. E 94 (2016) 052309. doi:10.1103/PhysRevE.94.052309.
[45] M. Kivelä, A. Arenas, M. Barthelemy, J. P. Gleeson, Y. Moreno, M. A. Porter, Multilayer networks, Journal of Complex Networks 2 (2014) 203–271. doi:10.1093/comnet/cnu016.
[46] L. Lacasa, V. Nicosia, V. Latora, Network structure of multivariate time series, Scientific Reports 5 (2015). doi:10.1038/srep15508.
[47] P. Holme, J. Saramäki, Temporal networks, Physics Reports 519 (2012) 97–125. doi:10.1016/j.physrep.2012.03.001.
[48] D. Kempe, J. Kleinberg, A. Kumar, Connectivity and Inference Problems for Temporal Networks, Journal of Computer and System Sciences 64 (2002) 820–842. doi:10.1006/jcss.2002.1829.
[49] R. K. Pan, J. Saramäki, Path lengths, correlations, and centrality in temporal networks, Phys. Rev. E 84 (2011) 016105. doi:10.1103/PhysRevE.84.016105.
[50] I. Scholtes, N. Wider, A. Garas, Higher-order aggregate networks in the analysis of temporal networks: path structures and centralities, The European Physical Journal B 89 (2016). doi:10.1140/epjb/e2016-60663-0.
[51] N. G. De Bruijn, A combinatorial problem, Proceedings of the Section of Sciences of the Koninklijke Nederlandse Akademie van Wetenschappen te Amsterdam 49 (1946) 758–764.
[52] R. Aroussi, Yahoo! Finance API, 2022. URL: https://github.com/ranaroussi/yfinance.
[53] V. N. Soloviev, A. O. Bielinskyi, N. A. Kharadzjan, Coverage of the coronavirus pandemic through entropy measures, CEUR Workshop Proceedings 2832 (2020) 24–42. URL: https://ceur-ws.org/Vol-2832/paper02.pdf.
[54] A. O. Bielinskyi, O. A. Serdyuk, S. O. Semerikov, V. N. Soloviev, Econophysics of cryptocurrency crashes: a systematic review, in: A. E. Kiv, V. N. Soloviev, S. O. Semerikov (Eds.), Proceedings of the Selected and Revised Papers of 9th International Conference on Monitoring, Modeling & Management of Emergent Economy (M3E2-MLPEED 2021), Odessa, Ukraine, May 26-28, 2021, volume 3048 of *CEUR Workshop Proceedings*, CEUR-WS.org, 2021, pp. 31–133. URL: http://ceur-ws.org/Vol-3048/paper03.pdf.
[55] A. O. Bielinskyi, S. V. Hushko, A. V. Matviychuk, O. A. Serdyuk, S. O. Semerikov, V. N. Soloviev, Irreversibility of financial time series: a case of crisis, in: A. E. Kiv, V. N. Soloviev, S. O. Semerikov (Eds.), Proceedings of the Selected and Revised Papers of 9th International Conference on Monitoring, Modeling & Management of Emergent Economy (M3E2-MLPEED 2021), Odessa, Ukraine, May 26-28, 2021, volume 3048 of *CEUR Workshop Proceedings*, CEUR-WS.org, 2021, pp. 134–150. URL: http://ceur-ws.org/Vol-3048/paper04.pdf.
[56] A. E. Kiv, V. N. Soloviev, S. O. Semerikov, H. B. Danylchuk, L. O. Kibalnyk, A. V. Matviychuk, A. M. Striuk, Machine learning for prediction of emergent economy dynamics III, in: A. E. Kiv, V. N. Soloviev, S. O. Semerikov (Eds.), Proceedings of the Selected and Revised Papers of 9th International Conference on Monitoring, Modeling & Management of Emergent Economy (M3E2-MLPEED 2021), Odessa, Ukraine, May 26-28, 2021, volume 3048 of *CEUR Workshop Proceedings*, CEUR-WS.org, 2021, pp. i–xxxi. URL: http://ceur-ws.org/Vol-3048/

paper00.pdf.
[57] A. O. Bielinskyi, V. N. Soloviev, Complex network precursors of crashes and critical events in the cryptocurrency market, CEUR Workshop Proceedings 2292 (2018) 37–45. URL: https://ceur-ws.org/Vol-2292/paper02.pdf.
[58] A. Bielinskyi, S. Semerikov, O. Serdyuk, V. Solovieva, V. N. Soloviev, L. Pichl, Econophysics of sustainability indices, in: A. Kiv (Ed.), 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, volume 2713 of *CEUR Workshop Proceedings*, CEUR-WS.org, 2020, pp. 372–392. URL: http://ceur-ws.org/Vol-2713/paper41.pdf.
[59] B. Schölkopf, J. Platt, T. Hofmann, Learning with Hypergraphs: Clustering, Classification, and Embedding, in: Advances in Neural Information Processing Systems 19: Proceedings of the 2006 Conference, MIT Press, 2007, pp. 1601–1608.
[60] G. F. de Arruda, G. Petri, Y. Moreno, Social contagion models on hypergraphs, Phys. Rev. Res. 2 (2020) 023032. doi:10.1103/PhysRevResearch.2.023032.
[61] T. Carletti, F. Battiston, G. Cencetti, D. Fanelli, Random walks on hypergraphs, Phys. Rev. E 101 (2020) 022308. doi:10.1103/PhysRevE.101.022308.
[62] M. T. Schaub, A. R. Benson, P. Horn, G. Lippner, A. Jadbabaie, Random Walks on Simplicial Complexes and the Normalized Hodge 1-Laplacian, SIAM Review 62 (2020) 353–391. doi:10.1137/18M1201019.
[63] J. J. Torres, G. Bianconi, Simplicial complexes: higher-order spectral dimension and dynamics, Journal of Physics: Complexity 1 (2020) 015002. doi:10.1088/2632-072X/ab82f5.
[64] P. S. Skardal, A. Arenas, Higher order interactions in complex networks of phase oscillators promote abrupt synchronization switching, Communications Physics 3 (2020). doi:10.1038/s42005-020-00485-0.
[65] I. Scholtes, When is a Network a Network? Multi-Order Graphical Model Selection in Pathways and Temporal Networks, in: Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD ’17, Association for Computing Machinery, New York, NY, USA, 2017, p. 1037–1046. doi:10.1145/3097983.3098145.
[66] A. Bielinskyi, V. Soloviev, S. Semerikov, V. Solovieva, Identifying stock market crashes by fuzzy measures of complexity, Neiro-Nechitki Tekhnolohii Modelyuvannya v Ekonomitsi 2021 (2021) 3–45. doi:10.33111/nfmte.2021.003.
[67] M. Bondarenko, Modeling relation between at-the-money local volatility and realized volatility of stocks, Neiro-Nechitki Tekhnolohii Modelyuvannya v Ekonomitsi 2021 (2021) 46–66. doi:10.33111/nfmte.2021.046.
[68] T. Kmytiuk, G. Majore, Time series forecasting of agricultural product prices using Elman and Jordan recurrent neural networks, Neiro-Nechitki Tekhnolohii Modelyuvannya v Ekonomitsi 2021 (2021) 67–85. doi:10.33111/nfmte.2021.067.
[69] V. Kobets, O. Novak, EU countries clustering for the state of food security using machine learning techniques, Neiro-Nechitki Tekhnolohii Modelyuvannya v Ekonomitsi 2021 (2021) 86–118. doi:10.33111/nfmte.2021.086.
[70] H. Kucherova, Y. Honcharenko, D. Ocheretin, O. Bilska, Fuzzy logic model of usability of websites of higher education institutions in the context of digitalization of educational

services, Neiro-Nechitki Tekhnolohii Modelyuvannya v Ekonomitsi 2021 (2021) 119–135. doi:10.33111/nfmte.2021.119.
[71] D. Lukianenko, I. Strelchenko, Neuromodeling of features of crisis contagion on financial markets between countries with different levels of economic development, Neiro-Nechitki Tekhnolohii Modelyuvannya v Ekonomitsi 2021 (2021) 136–163. doi:10.33111/nfmte.2021.136.
[72] I. Miroshnychenko, T. Kravchenko, Y. Drobyna, Forecasting electricity generation from renewable sources in developing countries (on the example of Ukraine), Neiro-Nechitki Tekhnolohii Modelyuvannya v Ekonomitsi 2021 (2021) 164–198. doi:10.33111/nfmte.2021.164.

Multidimensional statistical analysis of investment attractiveness and regional changes in the COVID-19 pandemic*

Pavlo M. Hryhoruk[1], Nila A. Khrushch[1], Svitlana S. Grygoruk[1] and Olena R. Ovchynnikova[1,2]

[1]Khmelnytskyi National University, 11 Instytutska Str., Khmelnytskyi, 29016, Ukraine
[2]Center of Migration Research, University of Warsaw, 7 Pasteura Str., Warsaw, 02-093, Poland

Abstract

This paper delves into an in-depth exploration of multidimensional statistical analysis techniques aimed at categorizing regions based on their levels of investment attractiveness, while also scrutinizing the evolving regional structures in light of the persistent and adverse effects of the COVID-19 pandemic. Through a comprehensive review of various approaches to assessing investment attractiveness, the study highlights their respective strengths. Notably, the research underscores the underutilization of multidimensional statistical analysis methodologies in the regional grouping context. The authors undertake the task of clustering Ukrainian regions based on their investment attractiveness levels, employing the well-regarded *k*-means method. This analysis extends to the identification of the regions' investment attractiveness structure in both 2019 and 2020, amid the COVID-19 pandemic. Substantiating the validity of their findings, the authors employ the principal component method in conjunction with the quartimax technique to rotate the space of selected factors. Remarkably, the subsequent regional grouping in this transformed principal component space mirrors the outcomes of the cluster analysis method. The research outcomes hold practical value for potential investors, enabling them to pinpoint key investment areas. Furthermore, local self-governing bodies stand to benefit from these findings, gaining insights into specific regions' relative investment attractiveness levels compared to their counterparts, while also uncovering vulnerabilities in distinct activity domains.

Keywords

socio-economic development, regional analysis, investment attractiveness, clustering, *k*-means method, principal component analysis, quartimax method

1. Introduction

Investment activities are crucial for the sustainable socio-economic development of territories, as they provide the financial basis for enhancing the performance of enterprises, creating new

M3E2-MLPEED 2022: The 10th International Conference on Monitoring, Modeling & Management of Emergent Economy, November 17-18, 2022, Virtual, Online

*Extended and revised version of paper [1] presented at the 10th International Conference on Monitoring, Modeling & Management of Emergent Economy.

violete@ukr.net (P. M. Hryhoruk); nila.ukr@gmail.com (N. A. Khrushch); grygoruk.svitlana@gmail.com (S. S. Grygoruk); o.ovchynnikova@uw.edu.pl (O. R. Ovchynnikova)
0000-0002-2732-5038 (P. M. Hryhoruk); 0000-0002-9930-7023 (N. A. Khrushch); 0000-0003-3047-2271 (S. S. Grygoruk); 0000-0002-7751-2923 (O. R. Ovchynnikova)

employment opportunities, and improving the quality of life. To achieve these goals, it is necessary to use analytical tools that can support the management of regional development. One of the key components of such tools is economic and mathematical modeling, which can use relevant data to fit models that can provide quantitative and qualitative assessments of the state and dynamics of socio-economic development. One of the applications of such modeling is to evaluate the investment attractiveness of regions, which reflects their potential to attract and retain investors.

However, the investment attractiveness of regions is not static, but rather influenced by various factors, such as globalization, market competition, and external shocks. In particular, the COVID-19 pandemic has caused significant structural and technological changes in the economy [2, 3], which have increased the demand for investment, while also limiting the availability of financial resources. Therefore, ensuring the investment attractiveness of regions is a strategic challenge for business development, especially in terms of attracting foreign investment.

In this paper, we address the problem of assessing the investment attractiveness of regions and their grouping based on their similarities and differences. We use multidimensional statistical analysis, a widely used technique in computer science and data science, to reduce the dimensionality of a large set of indicators that measure various aspects of investment attractiveness. We then apply clustering methods to identify groups of regions that share common characteristics and trends in terms of investment attractiveness. We also analyze how the COVID-19 pandemic has affected the regional changes in investment attractiveness and discuss the implications for regional policy and planning.

2. Literature review

Assessing the region's investment attractiveness is essential in developing a strategy for innovation at regional and national levels. Note that the investment volume isn't always directly determined by the high level of investment attractiveness. This is due to many other factors that determine investor decision-making. In particular, such factors are various indices and ratings regularly published by international institutions and characterize the business environment, business conditions, actual investment activities, and the attractiveness of countries for investment. In particular, it can be used such evaluations like the World Bank Ranking "Doing Business" [4]; Index of Economic Freedom, provided by the Heritage Foundation [5]; Corruption Perceptions Index (CPI), compiled by the international anti-corruption organization "Transparency International" [6]; Foreign Direct Investment Confidence Index [7]; Global Innovation Index [8]; European Business Association Investment Attractiveness Index [9]; Credit rating developed by Moody's Investors Service [10]; World Countries' Ranking on the Global Competitiveness Index, provided by the World Economic Forum [11]; World Competitiveness Ranking, provided by the International Institute for Management Development [12]; The KOF Globalization Index, published by the KOF Swiss Economic Institute, reflects the scale of the country's integration into the world [13] and many others.

These ratings provide the necessary information for potential investors on the characteristics of the country's business environment and possible investment risks. Naturally, countries with high ratings are more attractive regarding return on investment. On the other hand, countries with low ratings may also be attractive to investors, particularly for short-term investments, resulting from competition for coverage of developing countries.

These ratings should be noted that characterize the country's business environment. At the same time, investors are usually interested in specific areas, territorial units, markets, sectors of the economy, and business entities. Such assessments of the investment attractiveness of certain regions of Ukraine are provided by the State Statistics Service of Ukraine [14] and the Ministry of Development of Communities and Territories of Ukraine [15].

The issue of investment attractiveness is also the focus of research. The formation of the theoretical basis for the study of the category of investment attractiveness in the context of its relationship with the investment climate, and investment risks, taking into account current trends in economic development, is reflected in the works by Kaminskyi et al. [16], Stadnyk et al. [17], Korenyuk and Kopil [18], Kyshakevych et al. [19], Godlewska-Majkowska [20], Jac and Vondrackova [21]. Researchers presented a modern understanding of the category of investment attractiveness, its content and essential characteristics, and the impact on the socio-economic development of individual regions and the country as a whole.

An important issue in modeling investment attractiveness is forming an information base for calculating various estimates of the studied characteristics. The solution to such problems is considered by Kyshakevych et al. [19], Jac and Vondrackova [21], Bushynskyi [22], Leshchuk [23], Lagler [24], Swidynska and De Jesus [25]. It should be noted that the results of scholars' investigations in this field are differed both in the number of indicators and their focus in the context of reflecting certain aspects of investment activities. At the same time, the authors' positions coincide with the views that the indicators should reflect the economic, financial, and social aspects of regional development.

Modeling the investment attractiveness of regions is mainly based on statistical methods. Their application is based on quantitatively measurable indicators that reflect social, economic, environmental, and investment development components. This approach uses regression models, that presented in papers [26, 27, 28, 29]; correlation analysis techniques [30, 31]; models based on neural networks [32]. At the same time, the issue of identifying the level of investment attractiveness and comparing regions on this indicator is out of the attention of researchers. The approach based on comprehensive index assessment technology is quite common. It is successfully used in solving the problems of ranking regions by socio-economic development [33, 34, 35, 36]. The application of this approach to assessing investment attractiveness is reflected in studies [25, 37, 38, 39].

Among the shortcomings of the comprehensive index assessment technology application presented in these investigations, it should be noted that they use a fairly large set of initial indicators. This makes it difficult to identify the significance of their impact on the final result and eliminates the differentiating ability of the designed composite index. These shortcomings negatively affect the ability to group the set of studied objects due to the high density of values on the composite index scale. Also out of consideration is the definition of the level of investment attractiveness of regions, which complicates the assessment of differences between regions on the calculated index.

The problems of rating regions can also be solved with the application of multidimensional statistical analysis technology, that described, in particular, by Tenreiro Machado and Mata

[40, 41], Meyer and De Jongh [42], De Jongh and Meyer [43], Walesiak [44], Gorbatiuk et al. [45], Hryhoruk et al. [46, 47], Andrusiak et al. [48]. Adaptation of multidimensional analysis methods to assess investment attractiveness is considered by Cheba [49], Danylchuk et al. [50], Shinkarenko et al. [51], Musolino and Volget [52], Roszko-Wójtowicz and Grzelak [53].

At the same time, applying these methods is focused mainly on solving the problem of grouping regions in terms of investment attractiveness. The analysis of structural shifts within the constructed homogeneous groups remains out of the attention of scientists, as well as the comparison of grouping results obtained by different techniques.

According to the results of the analysis of publications, it can be concluded that there is significant diversity in approaches to assessing the investment attractiveness of regions. Among the disadvantages, we can note that the calculations are carried out without considering the dynamic and qualitative changes in the environment.

Also, the use of a large number of baseline partial indicators, to some extent, blurs the study's results and gives only a general description of the socio-economic condition of the region and the characteristics of investment activities.

The significant variety of calculated estimates and the lack of clear conclusions and recommendations for their practical application necessitates the further study of the problem of assessing the investment attractiveness of regions in the context of their grouping by using different techniques to solve this problem with the further comparison of grouping results and structural changes within groups.

The solution to these problems has led to the direction of research in this study.

3. Problem description and methodology

A large number of different indicators characterize modern investment processes. This multidimensionality of the description makes it difficult to solve problems of assessing the various characteristics of these processes, particularly the grouping of regions by the level of investment attractiveness. As noted earlier, one way to solve classification problems is using cluster analysis techniques. Unlike combinational grouping, this approach allows you to create groups of similar objects of observation, considering all the features at once. The degree of similarity, in this case, is usually the Euclidean distance between objects in the multidimensional space of primary indicators. One of the cluster analysis methods is the k-means method, which belongs to the group of iterative clustering ones.

Consider a brief description of the mathematical model of the k-means method [54]. Suppose there are m observations, each characterized by n indicators $X_1, X_2, \ldots, X_n$.

We need to divide these observations into k clusters that do not intersect. At the initial stage, we choose k points-objects that will act as centers of clusters. Denote them by $C_1^{(0)}, C_2^{(0)}, \ldots, C_k{}^{(0)}$. The weight of each cluster will initially be equal to one: $w_1^{(0)} = 1, w_2^{(0)} = 1, \ldots, w_k{}^{(0)} = 1$. The index of the corresponding center will be considered the index of the corresponding cluster. Although the selected centers may move to other clusters during the subsequent iterative procedure, the indexing of the clusters will not change.

In the first step, each of the $(n-k)$ objects that are not the clusters' centers is included in one of the formed clusters. The criterion for such movement is the minimum distance to the cluster's center. The center of the cluster and its weight are recalculated. For example, for a point M_{k+1} with coordinates $(X_{k+1,1}; X_{k+1,2};...; X_{k+1,n})$ the recalculation is performed according to the formulas:

$$C_j^{(0)} = \frac{w_j^{(0)} c_j^{(0)} + M_{k+1}}{w_j^{(0)} + 1}, \tag{1}$$

$$w_j^{(0)} = w_j^{(0)} + 1. \tag{2}$$

In the case of equality of two or more distances to the centers of clusters, the point-object joins the cluster with a smaller sequence number. Note that in practice, such a situation is unlikely.

The resulting centers and corresponding cluster weights are taken as the initial values of the related characteristics for the next iteration.

All stages of the further iterative process use formulas (1) and (2) and the whole set of initial data $M_1, M_2, ..., M_m$. At the same time, the weight of clusters continues to increase.

The objects can also be grouped by expanding them in some new space of latent scales, which reflect the generalizing characteristics. In particular, the principal components method can be constructed in such a space.

In matrix form, the model of the method is described by the formula:

$$Z^T = W \cdot F^T, \tag{3}$$

where Z is an initial standardized indicators matrix;

W – factor loadings matrix; it reflects relations between initial indicators and principal components;

F – principal components matrix.

Factor loadings matrix is calculated using eigenvalues and appropriate eigenvectors of R – initial indicators correlation matrix:

$$W = V \cdot \Lambda^{(-1)}, \tag{4}$$

where W – factor loadings matrix;

V – normalized eigenvectors matrix;

Λ – eigenvalues matrix.

The rule obtains the initial indicators correlation matrix:

$$R = \frac{Z^T \cdot Z}{m - 1}, \tag{5}$$

where R – correlation matrix;

Z – initial standardized indicators matrix.

The standardization procedure for initial indicators uses a formula:

$$Z_{ij} = \frac{X_{ij} - \bar{X}_j}{s_j}, \tag{6}$$

where Z_{ij} – initial standardized indicators values;

X_{ij} – initial indicators values;
$\bar{X}_j$ – average sample value of the indicator X_j;
s_j – sample standard deviation of the indicator X_j;
$i = 1..m; j = 1..n$.
The formula obtains the principal components matrix:

$$F = Z \cdot W \cdot \Lambda^{-1}. \tag{7}$$

This matrix contains the coordinates of objects under study in a principal components space.

Not all the principal components are selected for practical application, but only the most essential part in explaining the variance of the initial indicators (information contained in the initial indicators). Given that the eigenvalues of the correlation matrix are considered to be ordered in descending their values in the calculation procedure, the weight of each subsequent principal component is reduced. Usually, the first two principal components are sufficient to achieve an "acceptable" level of explanation of the information contained in the set of initial indicators, at least 70 %.

Meaningful interpretation of the selected principal components (search for names for them) is carried out by considering the absolute values of the respective factor loads. The initial indicators that will be used to interpret the appropriate principal component include those for which the factor loadings absolute value between them and the corresponding principal component is not less than 0.75. The factor load reflects the correlation between the principal component and the related indicator. To improve the procedure of interpretation of the principal components by the problem's content, the constructed factor space is rotated with a corresponding change in both factor loadings and values of the principal components. The result is a "simple structure" space where the principal components are closely related to some initial indicators and weak to others.

4. Results and discussions

Consider the application of cluster analysis of the grouping of Ukraine's regions by indicators that reflect their investment attractiveness. The choice of the initial indicators set will be made based on the following considerations:

- indicators should reflect both the characteristics of investment activities of the region's business entities and the socio-economic development of the region;
- indicators should be comparable by the values for different regions;
- the indicators must be standardized, i.e., have a sample mean equal to zero and a sample standard deviation equal to one. This procedure is needed because clustering is based on a matrix of differences between the studied points-regions in the multidimensional space of the initial indicators, which is essentially a matrix of Euclidean distances between them. Therefore, for the objectivity of the calculations, it is necessary to remove the measurement units' influence on estimates of distances between objects.

Based on the recommendations by Korenyuk and Kopil [18], Kyshakevych et al. [19], Godlewska-Majkowska [20], Jac and Vondrackova [21], Bushynskyi [22], Leshchuk [23], Lagler

Table 1
The relationships between the quantitative values of the desirability scale and qualitative development levels of group.

Code	Region	Code	Region
C_1	Vinnytsia	C_13	Mykolaiv
C_2	Volyn	C_14	Odesa
C_3	Dnipro	C_15	Poltava
C_4	Donetsk	C_16	Rivne
C_5	Zhytomyr	C_17	Sumy
C_6	Zakarpattia	C_18	Ternopil
C_7	Zaporizhzhia	C_19	Kharkiv
C_8	Ivano-Frankivsk	C_20	Kherson
C_9	Kyiv	C_21	Khmelnytskyi
C_10	Kyrovohrad	C_22	Cherkasy
C_11	Luhansk	C_23	Chernivtsi
C_12	Lviv	C_24	Chernihiv

[24], Vartsaba and Leshuk [28], Dorozynski and Kuna-Marszałek [31], and taking into account the above considerations, we have formed the following set of initial indicators for calculations:

X_1 – Volume of capital investments per capita, UAH;
X_2 – Volume of foreign direct investment per capita, USD;
X_3 – Gross regional product (at actual prices) per capita, UAH;
X_4 – Disposable income per capita, UAH;
X_5 – Volume of exports of goods per capita, USD;
X_6 – Volume of sold industrial products per capita, UAH;
X_7 – Total of construction work per capita, UAH;
X_8 – Employment rate of the population aged 15-70, in percent;
X_9 – Unemployment rate of the population aged 15-70 years (according to the methodology of the ILO), in percent.

We use data for 2019 and 2020 from the materials of the State Statistics Service of Ukraine [55] and the Ministry of Development of Communities and Territories of Ukraine [15] for calculations. Obtained results will also be compared to assess changes in regions' position in the groups in which the regions are located. To present the region's names conveniently and briefly, we point out the correspondence between each region's name and the appropriate code (table 1). Initial data for calculations are written in table 2 and table 3.

Let us cluster Ukraine's regions according to the selected set of indicators using the k-means method. Define the number of clusters for grouping regions as equal to three: a cluster of regions with a high level of investment attractiveness, a cluster of regions with a medium level of investment attractiveness, and a cluster of regions with a low level of investment attractiveness. We make calculations using "Statistica" software. The results of clustering are shown in figure 1 and figure 2. The numbering of clusters, in this case, is determined by the used software arbitrarily. Let us provide a meaningful description of each cluster according to 2019 data.

Table 2
Initial data for calculation for 2019 [15, 55].

Code	Values								
	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9
C_1	9196.8	153.2	71104	64729	937.3	52478.4	6650.5	58.0	9.4
C_2	11800.5	297.5	58297	52879	671.6	30585.3	2259.4	50.9	10.6
C_3	19841.6	1191.1	114784	87130	2477.8	142289.0	6291.1	59.5	7.7
C_4	6789.2	338.4	45959	39141	1116.8	68439.6	1691.2	50.9	13.6
C_5	6095.3	202.7	62911	61961	592.3	37457.1	2227.8	58.5	9.6
C_6	7010.9	288.1	41706	47495	1186.9	19086.0	1770.8	55.4	9.1
C_7	8246.3	538.3	85784	75407	1815.8	114981.1	2270.5	58.1	9.5
C_8	5969.9	529.3	57033	55537	665.1	48750.1	2701.6	56.6	7.2
C_9	27299.0	930.2	112521	75146	1098.1	68058.8	5833.2	59.3	5.9
C_10	7536.1	80.0	67763	58290	752.7	34338.9	2194.2	55.6	11.0
C_11	1303.0	209.0	16301	24477	71.3	10219.1	310.2	58.8	13.7
C_12	10137.4	446.8	70173	65691	874.9	41829.4	4391.2	57.8	6.5
C_13	10394.4	271.6	70336	63685	1912.6	55148.1	3864.1	59.1	9.3
C_14	8372.1	540.3	72738	72805	581.9	25815.1	7557.4	58.3	5.9
C_15	15316.4	841.3	123763	71627	1508.9	120922.5	5472.7	56.6	10.6
C_16	5225.1	116.7	49044	54183	381.1	37058.2	2872.8	58.4	8.3
C_17	6399.4	184.6	62955	65310	821.9	44941.0	1448.5	59.8	7.7
C_18	8016.4	47.7	46833	49843	416.7	19914.5	2325.0	53.8	10.0
C_19	7953.8	287.6	86904	65534	530.6	69605.2	5603.2	62.1	5.0
C_20	11420.8	237.9	52922	57110	259.6	29604.1	1777.3	58.9	9.6
C_21	6812.2	161.6	59583	58008	509.9	34392.0	3061.2	57.0	8.0
C_22	8143.2	298.7	76904	58808	720.1	61514.6	1732.7	59.3	8.3
C_23	3716.9	58.9	37441	48255	236.8	15093.2	2347.4	59.0	6.9
C_24	7965.9	447.4	69725	58904	808.6	34334.3	1907.2	58.9	10.2

Cluster number 1 contains 6 regions: Volyn, Donetsk, Zakarpattia, Kyrovohrad, Luhansk and Ternopil. In our opinion, this cluster can be called a group of regions with a low level of investment attractiveness. Note that the regions of this cluster are not industrially developed, which negatively affects their attractiveness for investment. In addition, the effects of the COVID-19 pandemic have had a significant negative impact on the development of these regions. Cluster number 2 contains regions with an average level of investment attractiveness. It is the most complete and consists of 14 points, which is quite natural in compliance with the essence of the division of the typical characteristics. The cluster with a high level of investment attractiveness includes cluster number 3, which includes Dnipro, Kyiv, Zaporizhzhia, and Poltava regions. For these regions, there are high values of the indicators presented in table 2, particularly the volume of foreign direct investment and relatively high employment rate, which allowed to give the cluster just such an interpretation. In addition, these regions have developed industries, which is also reflected in the indicator's values.

Comparing the cluster's structure obtained from 2020 data (figure 2), we can conclude that the fullness of clusters has not changed compared to the previous year. This indicates that there have been no significant changes in the investment attractiveness of Ukraine's regions in 2020. Although several normative acts have been adopted at the legislative level to facilitate attracting

Table 3
Initial data for calculation for 2020 [15, 55].

Code	Values								
	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9
C_1	6226.8	249.9	83175	70691	896.3	50771.9	7042.7	56.2	10.7
C_2	9319.5	240.8	73215	56603	624.7	31231.1	2465.0	48.9	12.5
C_3	15208.9	1426.0	122379	92083	2403.0	135366.4	5723.9	58.0	8.6
C_4	5355.9	424.0	49422	41662	956.0	62158.9	2470.6	49.2	14.9
C_5	5615.5	266.6	70247	67187	567.0	39163.6	1783.2	55.3	10.9
C_6	3192.3	193.3	48861	51073	1078.1	19249.3	1546.7	53.7	10.6
C_7	5864.6	851.4	91498	81949	1743.8	111716.8	1871.0	56.0	10.7
C_8	3630.9	402.0	63254	60276	555.3	45016.0	2822.3	54.1	8.4
C_9	12929.1	735.8	123267	79263	1102.6	70505.2	7089.9	57.8	6.9
C_10	5562.3	188.4	77816	63472	985.0	37684.9	1486.2	53.1	12.7
C_11	1086.3	74.4	18798	26714	60.9	8904.5	339.1	56.4	15.4
C_12	5880.9	639.3	85198	71150	927.4	44425.4	5709.3	56.0	7.6
C_13	5422.3	318.8	82149	68289	2018.3	55878.6	3017.5	57.3	10.7
C_14	6757.9	470.7	82903	80164	573.4	29687.1	12078.5	56.8	7.1
C_15	11829.0	1411.6	134449	77547	1680.3	115483.4	5940.1	54.8	12.0
C_16	3165.9	229.4	58332	58814	408.0	38908.7	2862.1	56.1	9.3
C_17	4763.8	321.9	70576	71117	918.7	43165.9	1612.1	56.8	9.4
C_18	5510.9	47.5	54833	55570	433.2	20508.6	2511.5	51.6	11.5
C_19	6178.4	344.0	92864	73218	556.1	66393.8	5509.5	59.9	6.2
C_20	3536.8	155.6	59987	63073	275.3	32008.3	1279.4	56.8	11.3
C_21	5784.0	94.8	65916	64824	531.1	37850.9	5301.7	54.8	9.9
C_22	4627.2	176.7	86319	64254	684.1	64414.0	2171.5	57.0	9.5
C_23	2533.5	61.8	46136	53875	187.5	15525.8	2428.7	56.5	8.9
C_24	5599.5	455.0	78118	64933	905.4	35004.1	2501.6	56.4	11.9

investments into Ukraine's economy, their positive impact has not yet manifested itself. On the other hand, it is possible to state a certain stabilization of indicators of socio-economic development of Ukraine's regions during the COVID-19 pandemic.

Let us consider the data of application of the principal components method for grouping Ukraine's regions by the level of investment attractiveness. We create the two-dimensional space of latent indicators obtained by applying this method and project points-regions on this space. To construct latent indicators, we use the quartimax method for rotations of factor space, which will contribute to an adequate representation of points in the new space and the identification of meaningful interpretation of new axes.

Calculations also are performed using Statistica software. The calculations' results of factor loadings are presented in table 4, and the values of points-objects in the new space – are in table 5. Note that in this case, the degree of explanation of the variance of the initial indicators by selected factors (degree of latent indicators informativeness) is 77 % for 2019 and 79 % for 2020 data. These indicators indicate the sufficiency of allocating exactly two principal components as latent indicators for further analysis.

Table 4 analysis allows us to provide such an interpretation of the selected principal components. Component F_1 has high factor loadings values for the initial indicators $X_1 - X_6$, and low

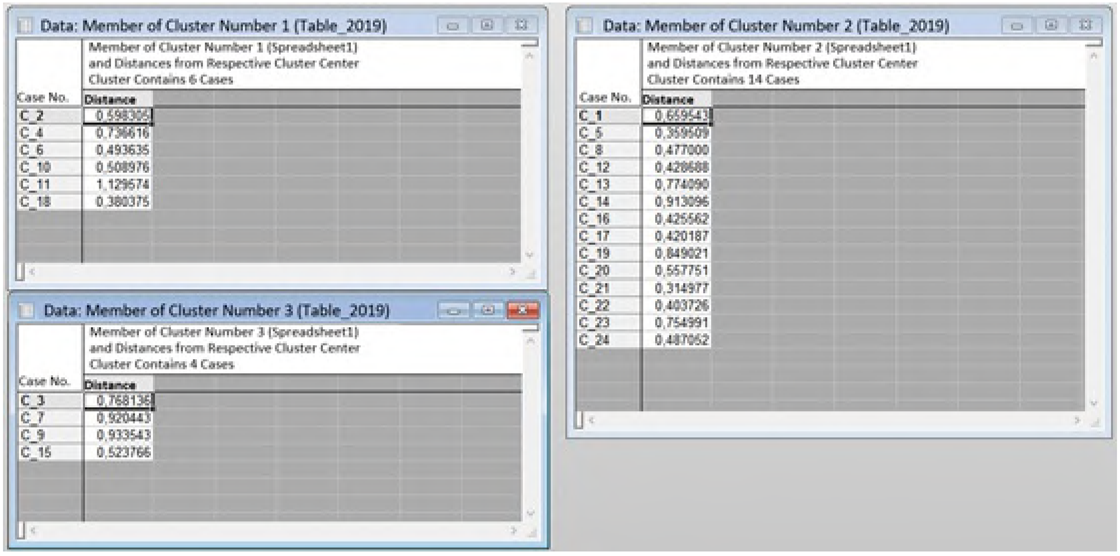


Figure 1: Clustering results of Ukraine‘s regions for data 2019.

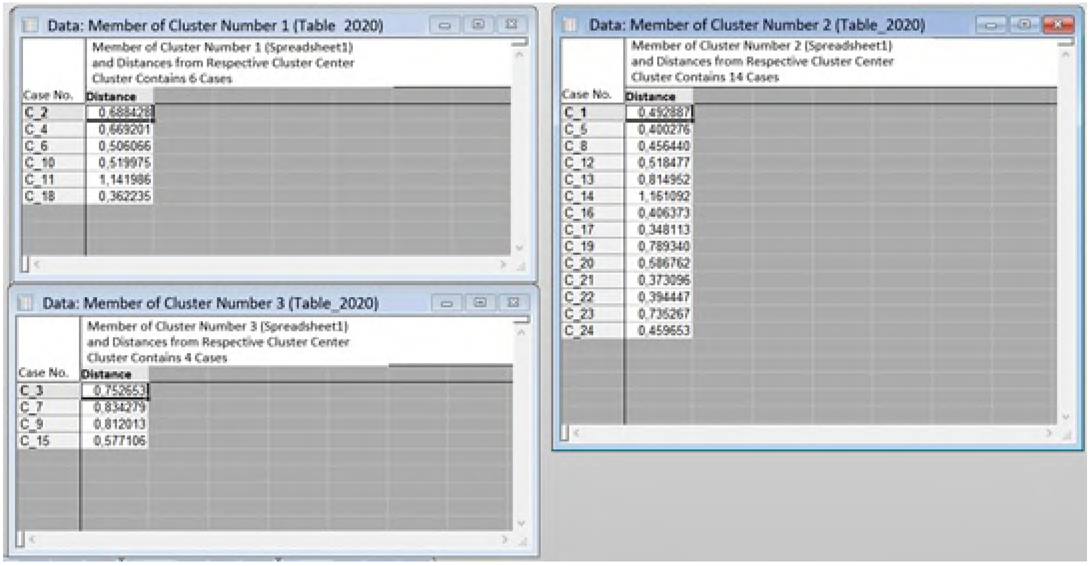


Figure 2: Clustering results of Ukraine‘s regions for data 2020.

for X_8 and X_9. Therefore, we can conclude that F_1 is an economic component of investment attractiveness, as the corresponding indicators $X_1 - X_6$ are characteristics of economic activity. For indicators X_8 and X_9 there are high values of factor loadings for component F_2 and low for F_1. Based on the essence of indicators X_8 and X_9, we can conclude that F_2 can be interpreted as a social component of investment attractiveness. For indicator X_7, the value of factor loadings for the principal component F_1 in 2019 exceeds the corresponding value for component F_2,

Table 4
Principal components' factor loadings values for data 2019 and 2020.

Indicators	Principal components			
	2019		2020	
	F_1	F_2	F_1	F_2
X_1	0.82	0.08	0.89	0.07
X_2	0.89	0.01	0.93	0.01
X_3	0.93	0.22	0.90	0.31
X_4	0.83	0.44	0.77	0.53
X_5	0.84	-0.25	0.87	-0.15
X_6	0.89	-0.13	0.93	-0.02
X_7	0.65	0.49	0.41	0.64
X_8	0.14	0.78	0.16	0.75
X_9	-0.25	-0.88	-0.21	-0.90

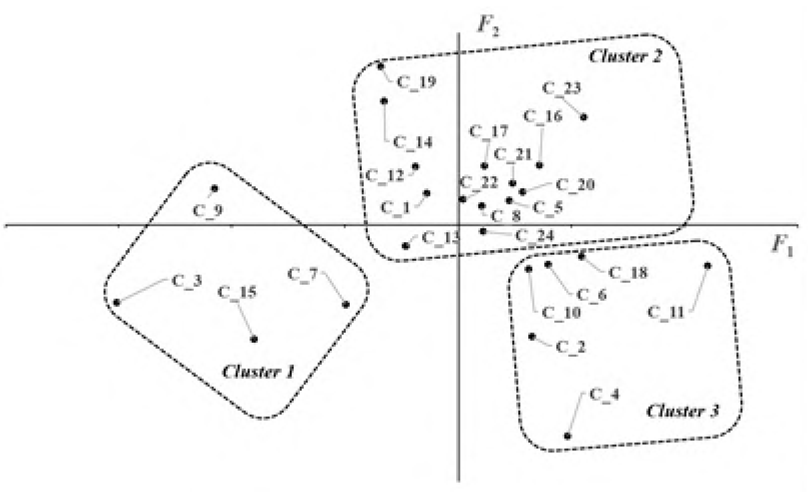


Figure 3: Grouping results of Ukraine's regions in the latent scale space for data 2019.

although this excess is insignificant. In 2020, the situation was reversed. Based on the essence of indicator X_7, we can conclude that it can characterize both the economic and social components of regional development; that is, the interpretation of the results is not essentially affected by this indicator.

Graphic representations of the regions in the space of the identified principal components according to the data of 2019 and 2020 are respectively presented in figures 3 and 4.

Figure 3 analysis allows us to conclude that we can also identify three clusters of regions. The resulting clusters in terms of content correspond to the formed clusters obtained by the *k*-means method. A similar situation occurs in figure 4. The results of a grouping of regions obtained using the principal components method are identical to those obtained from the clustering method.

However, it is worth noting. Figure 3 shows that cluster number 1 can be divided into at least two smaller clusters, including points C_7 (Zaporizhzhia region) and C_15 (Poltava region)

Table 5
Principal components' values for data 2019 and 2020.

Code	2019		2020	
	F_1	F_2	F_1	F_2
C_1	-0.55	0.49	0.04	0.53
C_2	1.31	-1.74	-0.02	-1.46
C_3	-6.03	-1.21	2.75	0.02
C_4	1.94	-3.29	-0.03	-2.32
C_5	0.90	0.38	-0.32	-0.11
C_6	1.59	-0.61	-0.70	-0.56
C_7	-1.98	-1.24	1.23	-0.58
C_8	0.42	0.29	-0.46	0.20
C_9	-4.30	0.56	1.13	1.24
C_10	1.25	-0.69	-0.10	-0.94
C_11	4.42	-0.63	-1.68	-1.16
C_12	-0.75	0.91	0.14	0.87
C_13	-0.92	-0.33	0.42	-0.16
C_14	-1.30	1.93	-0.08	2.09
C_15	-3.60	-1.78	2.25	-0.85
C_16	1.44	0.93	-0.75	0.42
C_17	0.47	0.92	-0.22	0.31
C_18	2.19	-0.49	-0.73	-0.67
C_19	-1.36	2.47	-0.07	1.84
C_20	1.14	0.51	-0.79	0.11
C_21	0.96	0.65	-0.47	0.42
C_22	0.09	0.40	-0.20	0.35
C_23	2.22	1.68	-1.28	0.65
C_24	0.44	-0.10	-0.07	-0.23

to one of them and points C_3 (Dnipro region) and C_9 (Kyiv region) to the other. Similarly, points C_7 and C_15 can be allocated from cluster number 2 to a separate cluster. In the third cluster, we can distinguish points C_4 (Donetsk region) and C_11 (Luhansk region), which form two different clusters. So, the cluster structure may be more complex and require a more complex interpretation of the results. A similar situation occurs in figure 4.

However, according to the task, our study proceeded from a predetermined number of clusters. And the methods used in the study gave identical results.

5. Conclusions

Investment activity is always associated with a specific risk. Potential investors need to conduct a comprehensive study of the investment object to reduce the potential risks. One of the approaches is to assess its investment attractiveness. For individual territorial entities, such as regions, it is advisable to determine the quantitative measures of the level of investment attractiveness and group them according to this indicator. This will identify regions with roughly the same investment climate.

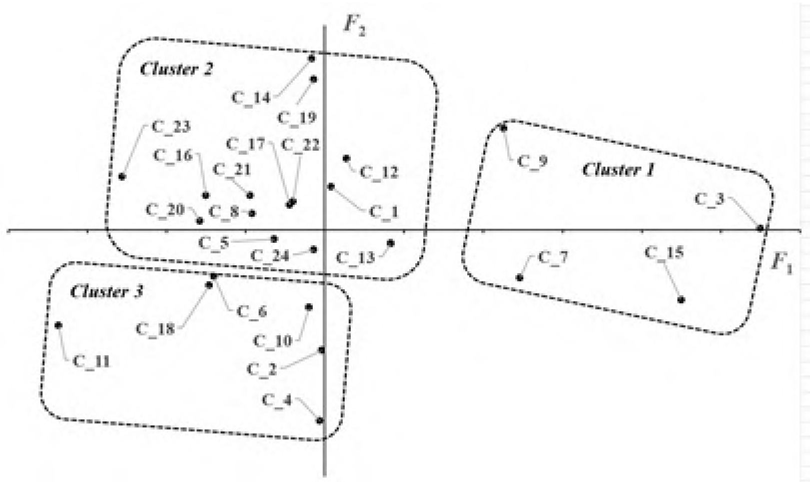


Figure 4: Grouping results of Ukraine's regions in the latent scale space for data 2020.

In this paper, we used the *k*-means method of clustering to divide the regions of Ukraine into relatively homogeneous groups according to the level of investment attractiveness. We used data for 2019 and 2020, which were characterized by the COVID-19 pandemic. We limited ourselves to the selection of three clusters, which were given a meaningful interpretation: the first one is a cluster with regions that have a high level of investment attractiveness, the second cluster contains regions with a medium level of investment attractiveness, and the third cluster includes regions with a low level of investment attractiveness.

Comparing clustering results for selected periods showed that the cluster structure of Ukrainian regions has not changed. To verify the correctness of the obtained grouping of regions, we deployed the regions of Ukraine in the space of latent indicators, which were calculated on the same data set by the method of principal components. A meaningful analysis of factor loads showed that one latent axis of the new space characterizes the economic component of investment attractiveness, and the other is the social component. The results of grouping Ukraine's regions turned out to be identical to those obtained by the *k*-means method.

We conclude that it would be more appropriate to allocate more clusters, which would provide a more accurate picture of the grouping of regions by the level of investment attractiveness. Such an assessment can be helpful for local governments, as it provides information on the relative level of investment attractiveness of a particular region compared to other territorial units and identifies weaknesses in the areas of activity on which the assessment was based. Such results can be used to create and adjust regional socio-economic development programs, particularly in terms of planning to attract investment into the region's economy.

Acknowledgments

This study was supported by the State budget project of Khmelnytskyi National University "Modeling the strategies for safe development of innovation-oriented socio-economic systems",

project's registration number 0122U001212.

References

[1] P. Hryhoruk, N. Khrushch, S. Grygoruk, O. Ovchynnikova, Application of Multidimensional Statistical Analysis Technology for Grouping Regions by the Investment Attractiveness Level, in: S. Semerikov, V. Soloviev, A. Matviychuk, V. Kobets, L. Kibalnyk, H. Danylchuk, A. Kiv (Eds.), Proceedings of 10th International Conference on Monitoring, Modeling & Management of Emergent Economy - M3E2, INSTICC, SciTePress, 2023, pp. 145–155. doi:10.5220/0011932000003432.

[2] S. Semerikov, H. Kucherova, V. Los, D. Ocheretin, Neural Network Analytics and Forecasting the Country's Business Climate in Conditions of the Coronavirus Disease (COVID-19), in: V. Snytyuk, A. Anisimov, I. Krak, M. Nikitchenko, O. Marchenko, F. Mallet, V. V. Tsyganok, C. Aldrich, A. Pester, H. Tanaka, K. Henke, O. Chertov, S. Bozóki, V. Vovk (Eds.), Proceedings of the 7th International Conference "Information Technology and Interactions" (IT&I-2020). Workshops Proceedings, Kyiv, Ukraine, December 02-03, 2020, volume 2845 of *CEUR Workshop Proceedings*, CEUR-WS.org, 2020, pp. 22–32. URL: https://ceur-ws.org/Vol-2845/Paper_3.pdf.

[3] S. S. Iyer, L. Gernal, R. Subramanian, A. Mehrotra, Impact of digital disruption influencing business continuity in UAE higher education, Educational Technology Quarterly 2023 (2023) 18–57. doi:10.55056/etq.29.

[4] World Bank, The World Bank. Doing business 2020, 2021. URL: https://cutt.ly/qU2j9bN.

[5] Heritage Foundation, Index of Economic Freedom 2020, 2020. URL: https://www.heritage.org/index/ranking.

[6] Transparency International, Corruption Perceptions Index 2020, 2020. URL: https://www.transparency.org/en/cpi/2020.

[7] Kearney, Kearney Foreign Direct Investment (FDI) Confidence Index, 2020. URL: https://www.kearney.com/foreign-direct-investment-confidence-index.

[8] WIPO, Global Innovation Index, 2020. URL: https://cutt.ly/dU2kqQw.

[9] EBA, Indeks investytsiinoi pryvablyvosti Ukrainy 2020 [Investment attractiveness index of Ukraine 2020], 2020. URL: https://cutt.ly/MU2j5w5.

[10] Moody's, Credit rating, 2020. URL: https://www.moodys.com/.

[11] WEF, World Countries' Ranking on the Global Competitiveness Index, 2020. URL: https://cutt.ly/rVR9tLo.

[12] IMDWCC, IMD World Competitiveness Center. World Competitiveness Ranking, 2020. URL: https://cutt.ly/RU2kryT.

[13] KOF, KOF Globalization Index, 2020. URL: https://cutt.ly/DU2kuCW.

[14] SSCU, Pro zatverdzhennia Metodyky rozrakhunku intehralnykh rehionalnykh indeksiv ekonomichnoho rozvytku. Nakaz Derzhavnoho komitetu statystyky Ukrainy № 114 vid 15.04.2003 r. [On approval of the Methodology for calculating integrated regional indices of economic development. Order of the State Statistics Committee of Ukraine № 114 of April 15, 2003], 2003. URL: https://zakon.rada.gov.ua/rada/show/v0114202-03#Text.

[15] MCTDU, Reitynhova otsinka rehioniv [Rating assessment of regions], 2021. URL: https://cutt.ly/olpWxkF.
[16] A. Kaminskyi, I. Miroshnychenko, K. Pysanets, Risk and return for cryptocurrencies as alternative investment: Kohonen maps clustering, Neiro-Nechitki Tekhnolohii Modelyuvannya v Ekonomitsi 2019 (2019) 175–193. doi:10.33111/nfmte.2019.175.
[17] V. Stadnyk, P. Izhevskiy, N. Khrushch, S. Lysenko, G. Sokoliuk, T. Tomalja, Strategic priorities of innovation and investment development of the Ukraine's economy industrial sector, in: A. Kiv (Ed.), 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, volume 2713 of *CEUR Workshop Proceedings*, CEUR-WS.org, 2020, pp. 145–166. URL: http://ceur-ws.org/Vol-2713/paper12.pdf.
[18] P. Korenyuk, E. Kopil, Conceptual Approaches and Methods for determining the Investment Attractiveness of the National Economy, Economic Journal of Lesya Ukrainka Eastern European National University 2 (2018) 56–62. doi:10.29038/2411-4014-2018-02-56-62.
[19] B. Kyshakevych, M. Nahayeva, A. Kryński, The economic essence of the investment attractiveness of the regions, Scientific Journal of Polonia University 37 (2020) 46–53. doi:10.23856/3704.
[20] H. Godlewska-Majkowska, Investment Attractiveness of Polish Municipalities in Relation to Local Entrepreneurship, Olsztyn Economic Journal 13 (2018) 103–122. doi:10.31648/oej.2764.
[21] I. Jac, M. Vondrackova, The perception of selected aspects of investment attractiveness by businesses making investments in the Czech Republic, E+M Ekonomie a Management 20 (2018) 103–122. URL: https://dspace.tul.cz/bitstream/handle/15240/20918/EM_3_2017_08.pdf.
[22] Y. Bushynskyi, Kontsept faktoriv, shcho vyznachaiut investytsiinu pryvablyvist rehionu [The concept of factors determining the investment attractiveness of the region], Herald of Khmelnytskyi National University Economic sciences 2 (2020) 185–190. URL: http://journals.khnu.km.ua/vestnik/wp-content/uploads/2021/03/38.pdf.
[23] H. Leshchuk, Systematyzatsiia pidkhodiv do otsiniuvannia investytsiinoi pryvablyvosti rehionu [Systematization of the approaches to estimating the investment attractiveness of a region], Uzhorod National University Herald. International Economic Relations and World Economy (2020) 93–98. doi:10.32782/2413-9971/2020-29-18.
[24] K. Lagler, Use of world experience in increasing investment attractiveness of Ukraine regions, Economics, Finance And Management Review 2 (2020) 63–70.
[25] N. Swidynska, I. De Jesus, Determinants of Potential Investment Attractiveness of a Commune, Olsztyn Economic Journal 15 (2020) 129–140. URL: https://czasopisma.uwm.edu.pl/index.php/oej/article/view/5836/4421.
[26] I. Windhyastiti, S. Hidayatullah, U. Khouroh, Investment attractiveness rating and factors affecting, Accounting (2020) 161–166. URL: http://www.growingscience.com/ac/Vol7/ac_2020_143.pdf.
[27] I. Blahun, D. L., L. H., Simulative model for evaluation of investment processes in the regions of Ukraine, Investment Management and Financial Innovations 14 (2017) 322–329. URL: https://cutt.ly/UVTu9zE.

[28] V. Vartsaba, H. Leshuk, Evaluation of investment attractiveness indicators of regions in Ukraine, Baltic Journal of Economic Studies 3 (2017) 38–44. doi:10.30525/2256-0742/2017-3-5-38-44.

[29] X. N. Kikkas, E. Krasnozhenova, Modeling investment attractiveness of Germany, in: Proceedings of the 5th International Multidisciplinary Scientific Conference on Social Sciences and Arts SGEM 2018, Sofia Bulgaria, 2018, pp. 515–522.

[30] P. Blaschke, Investment Incentives in the Environment of the Czech Republic, E+M Economics and Management 25 (2022) 4–23. doi:10.15240/tul/001/2022-1-001.

[31] T. Dorozynski, A. Kuna-Marszałek, Investment Attractiveness. The Case of the Visegrad Group Countries, Comparative Economic Research. Central and Eastern Europe 19 (2016) 117–138. doi:10.1515/cer-2016-0007.

[32] J. Bruneckiene, R. Jucevicius, I. Zykiene, J. Rapsikevicius, M. Lukauskas, Assessment of Investment Attractiveness in European Countries by Artificial Neural Networks: What Competences are Needed to Make a Decision on Collective Well-Being, Sustainability 11 (2019) 6892. doi:10.3390/su11246892.

[33] P. Hryhoruk, N. Khrushch, S. Grygoruk, The Rating Model of Ukraine's Regions According to the Level of Economic Development, Periodicals of Engineering and Natural Sciences 7 (2019) 712–722. URL: https://cutt.ly/BjAgnAL.

[34] P. Hryhoruk, N. Khrushch, S. Grygoruk, An Approach to Design a Composite Index of Economic Development and Identifying the Bounds of its Levels, in: Proceedings of the 2019 9th International Conference on Advanced computer information technologies ASIT'2019, Ceske Budejovice Czech Republic, 2019, pp. 48–51.

[35] M. Mazziotta, A. A. Pareto, Composite index for measuring Italian regions' development over time, Rivista Italiana di Economia Demografia e Statistica LXVIII (2014) 127–134. URL: https://cutt.ly/UVTo3op.

[36] D. F. Meyer, J. De Jongh, N. Meyer, The Formulation of a Composite Regional Development Index, International Journal of Business and Management Studies 8 (2016) 100–116. URL: https://cutt.ly/cf5q8J2.

[37] B. Kyshakevcyh, M. Nakhaeva, Intehralna otsinka investytsiinoi pryvablyvosti rehioniv Ukrainy [Integrated assessment of investment attractiveness of regions in Ukraine], Herald of Khmelnytskyi National University Economic sciences (2021) 51–58. URL: https://cutt.ly/MVTpdJ3.

[38] Y. Yelnikova, Reitynhuvannia rehioniv Ukrainy za rivnem pryvablyvosti dlia vidpovidalnykh investytsii [Ranking of the Ukraine regions by the level of responsible investments attractiveness], Investytsiyi: praktyka ta dosvid (2020) 63–68. URL: http://www.investplan.com.ua/?n=17-18&y=2020.

[39] N. D. Sizova, O. O. Petrova, G. V. Solodovnik, M. U. Perun, Otsinka investytsiinoi pryvablyvosti z vykorystanniam informatsiinoi systemy [Evaluation of investment attractiveness using information system], Young Scientist 4.4 (2017) 90–93. URL: http://molodyvcheny.in.ua/files/journal/2017/4.4/21.pdf.

[40] J. A. Tenreiro Machado, M. E. Mata, Analysis of World Economic Variables Using Multidimensional Scaling, PLoS ONE 10 (2015) e0121277. URL: https://cutt.ly/nH4a5HP.

[41] J. A. Tenreiro Machado, M. E. Mata, Multidimensional scaling analysis of the world economy during the period 1972-2012, Acta Polytechnica Hungarica 12 (2015) 67–82. URL:

https://cutt.ly/4VTsSjT.
[42] D. F. Meyer, J. De Jongh, An Alternative Multi-Dimensional Regional Economic Development Index: a Provincial Application in South Africa, International Journal of eBusiness and eGovernment Studies 10 (2018) 97–113. URL: https://cutt.ly/VVTaO7l.
[43] J. J. De Jongh, D. F. Meyer, The Multidimensional Regional Economic Development Index (MREDI) Applied in the North-West Province: A Rural Regional Application, Administratio Publica 27 (2019) 162–185. URL: https://cutt.ly/EVTaGVA.
[44] M. Walesiak, 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 (2017) 9–25. URL: https://cutt.ly/OVTaMeB.
[45] K. Gorbatiuk, O. Mantalyuk, O. Proskurovych, O. Valkov, Analysis of Regional Development Disparities in Ukraine Using Fuzzy Clustering, in: A. Kiv, S. Semerikov, V. N. Soloviev, L. Kibalnyk, H. Danylchuk, A. Matviychuk (Eds.), Proceedings of the Selected Papers of the 8th International Conference on Monitoring, Modeling & Management of Emergent Economy, M3E2-EEMLPEED 2019, Odessa, Ukraine, May 22-24, 2019, volume 2422 of *CEUR Workshop Proceedings*, CEUR-WS.org, 2019, pp. 194–210. URL: https://ceur-ws.org/Vol-2422/paper16.pdf.
[46] P. Hryhoruk, N. Khrushch, S. Grygoruk, Using Multidimensional Scaling for Assessment Economic Development of Regions, Journal of Industrial Engineering & Production Research 31 (2020) 597–607. URL: https://cutt.ly/hVTsxRJ.
[47] P. Hryhoruk, S. Grygoruk, N. Khrushch, T. Hovorushchenko, Using non-metric multidimensional scaling for assessment of regions' economy in the context of their sustainable development, in: A. Kiv (Ed.), 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, volume 2713 of *CEUR Workshop Proceedings*, CEUR-WS.org, 2020, pp. 315–333. URL: https://ceur-ws.org/Vol-2713/paper35.pdf.
[48] N. O. Andrusiak, V. M. Andrusiak, H. B. Danylchuk, Assessment of ecological and economic competitiveness of regions using factor analysis, IOP Conference Series. Earth and Environmental Science 1049 (2022) 012084. doi:10.1088/1755-1315/1049/1/012084.
[49] K. Cheba, Multidimensional analysis of investments attractiveness of selected European countries from Japanese investors' perspective, in: Proceedings of Carpathian Logistics Congress - CLC 2017, TANGER Ltd, Ostrava Slovakia, 2017, pp. 204–209.
[50] H. Danylchuk, N. Chebanova, N. Reznik, Y. Vitkovskyi, Modeling of investment attractiveness of countries using entropy analysis of regional stock markets, Global J. Environ. Sci. Manage 5 (2019) 227–235.
[51] V. Shinkarenko, M. Matskul, D. Linok, Investment attractiveness modeling using multidimensional statistical analysis methods, SHS Web of Conferences 65 (2019) 04007. doi:10.1051/shsconf/20196504007.
[52] D. Musolino, S. Volget, Towards a multidimensional approach to the study of territorial attractiveness, 2020. URL: https://hal.archives-ouvertes.fr/hal-02501582.
[53] E. Roszko-Wójtowicz, M. M. Grzelak, Multi-dimensional analysis of regional investment attractiveness in Poland, Equilibrium. Quarterly Journal of Economics and Economic Policy 16 (2021) 103–138. URL: https://cutt.ly/yVTdK70.
[54] P. Hryhoruk, N. Khrushch, S. Grygoruk, Environmental safety assessment: a regional

dimension, IOP Conference Series: Earth and Environmental Science 628 (2021) 012026. doi:10.1088/1755-1315/628/1/012026.

[55] SSSU, Derzhavna sluzhba statystyky Ukrainy [State Statistics Service of Ukraine], 2022. URL: http://www.ukrstat.gv.ua.

A comparative study of deep learning models for sentiment analysis of social media texts*

Vasily D. Derbentsev[1], Vitalii S. Bezkorovainyi[1], Andriy V. Matviychuk[1,2], Oksana M. Pomazun[1], Andrii V. Hrabariev[1] and Alexey M. Hostryk[3]

[1] *Kyiv National Economic University named after Vadym Hetman, 54/1 Peremogy Ave., Kyiv, 03680, Ukraine*
[2] *Kryvyi Rih State Pedagogical University, 54 Gagarin Ave., Kryvyi Rih, 50086, Ukraine*
[3] *Odessa National Economic University, 8 Preobrazhenskaya Str., Odessa, 65082, Ukraine*

Abstract
Sentiment analysis is a challenging task in natural language processing, especially for social media texts, which are often informal, short, and noisy. In this paper, we present a comparative study of deep learning models for sentiment analysis of social media texts. We develop three models based on deep neural networks (DNNs): a convolutional neural network (CNN), a CNN with long short-term memory (LSTM) layers (CNN-LSTM), and a bidirectional LSTM with CNN layers (BiLSTM-CNN). We use GloVe and Word2vec word embeddings as vector representations of words. We evaluate the performance of the models on two datasets: IMDb Movie Reviews and Twitter Sentiment 140. We also compare the results with a logistic regression classifier as a baseline. The experimental results show that the CNN model achieves the best accuracy of 90.1% on the IMDb dataset, while the BiLSTM-CNN model achieves the best accuracy of 82.1% on the Sentiment 140 dataset. The proposed models are comparable to state-of-the-art models and suitable for practical use in sentiment analysis of social media texts.

Keywords
sentiment analysis, social media, deep learning, convolutional neural networks, long short-term memory, word embeddings

1. Introduction

The swift evolution of electronic mass media and social networks has spurred the advancement of automated Natural Language Processing (NLP) systems.

NLP resides at the crossroads of Computer Science, Artificial Intelligence, and Linguistics, dedicated to unraveling the intricacies of computer-based analysis of human language models.

The spectrum of challenges that NLP addresses is extensive. It encompasses tasks like machine

M3E2-MLPEED 2022: The 10th International Conference on Monitoring, Modeling & Management of Emergent Economy, November 17-18, 2022, Virtual, Online

*Extended and revised version of paper [1] presented at the 10th International Conference on Monitoring, Modeling & Management of Emergent Economy.

derbvd@gmail.com (V. D. Derbentsev); retal.vs@gmail.com (V. S. Bezkorovainyi); editor@nfmte.com (A. V. Matviychuk); oksp@ukr.net (O. M. Pomazun); andr.grab@gmail.com (A. V. Hrabariev); alexeyGostrik@gmail.com (A. M. Hostryk)
0000-0002-8988-2526 (V. D. Derbentsev); 0000-0002-4998-8385 (V. S. Bezkorovainyi); 0000-0002-8911-5677 (A. V. Matviychuk); 0000-0001-9697-1415 (O. M. Pomazun); 0000-0001-6165-0996 (A. V. Hrabariev); 0000-0001-6143-6797 (A. M. Hostryk)

translation, speech recognition, named entity recognition, text classification and summarization, sentiment analysis, question answering, autocomplete, predictive text input, and more [2, 3, 4].

Central to NLP is Sentiment Analysis (SA), also known as opinion mining. SA endeavors to distill subjective attributes from text, such as emotions, sarcasm, confusion, and suspicion.

The crux of SA revolves around classifying the polarity of a given document, determining whether the sentiment expressed is positive, negative, or neutral.

Being a potent text classification technique, sentiment analysis can unveil a wealth of insights about viewpoints on discussed subjects. It facilitates comprehensive analysis of feedback, message polarity, and reactions. Notably, SA finds extensive utility among business professionals, marketers, and politicians.

In dissecting public sentiment regarding sensitive social and political matters, discerning prevailing themes and tonalities within discussions significantly eases the tasks of sociologists, political scientists, and journalists [5, 6].

In the face of ever-mounting information volumes, conventional methodologies have begun to falter. Swiftly monitoring and controlling public sentiment remains pivotal for success.

Historically, this challenge has been met with dictionary or rule-based approaches [7, 8, 9, 10]. These methods are statistical, relying on precompiled sentiment lexicons that pair words with respective polarities to categorize them as "positive" or "negative".

However, construction complete dictionaries for a large amount of unstructured data generated by modern electronic media and social networks are quite a tedious task.

Machine Learning (ML) methods [11, 12, 13] help solves this problem. Such approaches are based on algorithms for classifying words according to the corresponding sentiment marks. That's why ML models are preferred for SA due to their ability to processing with the large amount of texts compared to dictionary-based approaches.

Over the past decade, Deep Neural Networks (DNNs) have emerged as formidable tools in solving numerous NLP challenges, including SA [14, 15, 16]. This surge is underpinned by:

- Progress in crafting diverse DNN architectures (recurrent, convolutional, encoder-decoder, transformer, hybrid).
- Escalating computational prowess, bolstered by graphics processing units and a profusion of cloud computing services.
- Availability of labeled datasets tailored to various NLP tasks.
- Emergence of pre-trained word vector representations (word embeddings) like Word2Vec, FastText [17, 18, 19], extending across multiple languages.

Recent years have seen the ascendancy of colossal pre-trained models rooted in the Transformer architecture and the Attention mechanism—think GPT-3, BERT, ELMo [20, 21, 22, 23]. These models embody language models, encapsulating probability distributions across word sequences.

These models are all-encompassing, extracting features from text pivotal for solving diverse text analysis conundrums. However, they come at a computational cost—bearing hundreds of millions of parameters, necessitating formidable computational resources.

Hence, for the majority of practical NLP applications, conventional ML and Deep Learning (DL) methodologies persist as stalwarts.

Our **research aims** to architect a suite of sentiment classification models grounded in varied DNN architectures, scrutinizing their efficacy across the IMDb and Sentiment 140 Twitter datasets.

2. Related works

Drus and Khalid [24] provided a report of review on sentiment analysis in social media that explored the common methods and approaches which used in this domain. This review contains an analysis of about 30 publications published during 2014-2019 years. According to their results most of the articles applied opinion-lexicon method to analyses text sentiment in social media in such domain as world events, healthcare, politics and business.

Recently Jain et al. [25] published report on ML applications for consumer sentiment analysis in the domain of hospitality and tourism. This report based on 68 research papers, which were focused on sentiment classification, predictive recommendation decisions, and fake reviews detection.

They have shown a systematic literature review to compare, analyze, explore, and understand ML possibilities to find research gaps and the future research directions.

Sudhir and Suresh [26] published comparative study of various approaches, applications and classifiers for sentiment analysis. They have discussed the advantages and disadvantages of the different approaches such as Rule-based, ML and DL approaches used for SA as well as compared the performances of the classification models on the IMDb dataset.

The authors note that, in general, ML-based approaches provide greater accuracy than Rule-based ones. At the same time, Conventional ML models (Support Vector Machine, Decision Trees, and Logistic Regression) provide classification accuracy at the level of 85-87% for the IMDb dataset. DL-based models (CNN, LSTM, GRU) shows higher accuracy: about 89% on the IMDb dataset.

Trisna and Jie [15], presented a comparative review of DL approaches for Aspect-Based SA. The results of their analysis show that the use of pre-trained embeddings is very influential on the level of accuracy. They also found that every dataset has a different method to get better performance. It is still challenging to find the method that can be flexible and effective for using in several datasets.

There are several papers devoted to developing new methods of word embeddings.

Thus, Biesialska et al. [27] proposed a novel method which uses contextual embeddings and a self-attention mechanism to detect and classify sentiment. They performed experiments on reviews from different domains, as well as on languages (Polish and German).

Authors have shown that proposed approach is on a par with state-of-the-art models or even outperforms them in several cases.

Rasool et al. [28] proposed a novel word embedding method novel word-to-word graph (W2WG) embedding method for the real-time sentiment for word representation. He noted that performance evaluation of proposed word embedding approach with integrated LSTM-CNN outperformed the other techniques and recently available studies for the real-time sentiment classification.

Recently have been published several research papers devoted using DNNs different architecture based on CNN-LSTM models for SA task [29, 30, 31, 32, 33].

Elzayady et al. [29] presented two powerful hybrid DL models (CNN-LSTM) and (CNN-BILSTM) for reviews classification. Experimental results have shown that the two proposed models had superior performance compared to baselines DL models (CNN, LSTM).

Khan et al. [31] evaluated the performance of various word embeddings for Roman Urdu and English dialects using the CNN-LSTM architecture and compare results with traditional ML classifiers. Authors mentioned that BERT word embedding, two-layer LSTM, and SVM as a classifier function are more suitable options for English language sentiment analysis.

Priyadarshini and Cotton [32] proposed a novel LSTM-CNN grid search-based DNN model for sentiment analysis. As to the experimental results they observed proposed model performed relatively better than other algorithms (LSTM, Fully-connected NN, K-nearest neighbors, and CNN-LSTM) on Amazon reviews for sentiment analysis and IMDb datasets.

Haque et al. [33] analyzed different DNNs for SA on IMDb Movie Reviews. They have compared between CNN, LSTM and LSTM-CNN architectures for sentiment classification in order to find the best-suited architecture for this dataset. Experimental results have shown that CNN has achieved an $F1-score$ of 91% which has outperformed LSTM, LSTM-CNN and other state-of-the-art approaches for SA on IMDb dataset.

Quraishi [34] evaluated of four ML algorithms (Multinomial Naïve Bayes, Support Vector Machine, LSTM, and GRU) for sentiment analysis on IMDb review dataset. He found that among these four algorithms, GRU performed the best with an accuracy of 89.0%.

Derbentsev et al. [35] also explored the performance of four ML algorithms (Logistic Regression, Support Vector Machine, Fully-connected NN, and CNN) for SA on IMDb dataset. They used two pre-trained word embeddings GloVe and Word2vec with different dimensions (100 and 300) as well as TF-IDF representation. They reported that the best classification accuracy (90.1%) was performed by CNN model with Word2vec-300 embedding.

3. Base concept of NLP applying to sentiment analysis

3.1. ML approach of NLP

To solve NLP problems using ML methods, it is necessary to represent the text in the form of set feature vectors. The text can consist of words, numbers, punctuation, special characters of additional markup (for example, HTML tags). Each such “unit” can be represented as a vector in various ways, for example, using unitary codes (one-hot encoding), or context-independent (depended) vector representations.

The base idea of applying ML to NLP was introduced by Bengio et al. [36]. They proposed to jointly learn an “embedding” of words into an n-dimensional numeric vector space and to use these vectors to predict how likely a word is given its context.

In the case of text, features represent attributes and properties of documents including their content and meta-attributes, such as document length, author name, source, and publication date. Together, all document features describe a multidimensional feature space to which ML methods can be applied.

Thus, in the most general terms, the application of ML to SA problems consists of the following: text data preprocessing, feature extraction, classification, and interpretation of results.

3.2. Data pre-processing

The quality of the result depends on the input data. Therefore, it is important that they are prepared in the best possible way. In general, pre-processing stage consists of the following steps [37, 38, 39]:

- *Text cleaning*. First of all, we need to clean up the text. Depending on the task, cleaning includes removing non-alphabets, various tags, URLs, punctuation, spaces, and other markup elements;
- *Segmentation and tokenization*. They are relevant in the vast majority of cases, and provide division of the text into separate sentences and words (tokens). As a rule, after tokenization all words are converted to lower case;
- *Lemmatization and stemming*. Typically, texts contain different grammatical forms of the same word, and there may also be words with the same root. Lemmatization is the process of reducing a word form to a lemma – its normal (dictionary) form. Stemming is a crude heuristic process that cuts off "excess" from the root of words, often resulting in the loss of derivational suffixes. Lemmatization is a subtler process that uses vocabulary and morphological analysis to eventually reduce a word to its canonical form, the lemma;
- *Definition of context-independent features* that characterize each of the token, which not dependent on adjacent elements;
- *Refining significance and applying a filter to stop words*. Stop words are frequently used words that do not add additional information to the text. When we apply ML to texts, such words can add a lot of noise, so it is necessary to get rid them;
- *Dependency parsing*. The result is the formation of a tree structure, where the tokens are assigned to one parent, and the type of relationship is established;
- *Converting text content to a vector representation* that highlights words used in similar or identical contexts.

3.3. Features extraction

ML algorithms cannot work directly with raw text, so it is necessary to convert the text into sets of numbers (vectors) – construct a vector representation. In ML this process is called feature extraction.

Vector representation is a general name for various approaches to language modeling and representation training in NLP aimed at matching words (and possibly phrases) from some dictionary of vectors.

The most common approaches for construction vector representations are Bag of Words, TF-IDF, and Word Embeddings [38].

3.3.1. Bag of words

Bag of words (Bow) is a popular and simple feature extraction technique used in NLP. It describes the occurrences of each word in the text.

Essentially, it creates a matrix of occurrences for a sentence or document, ignoring grammar and word order. These frequencies ("occurrences") of words are then used as features for learning.

The basic idea of applying Bow is that similar documents have similar content. Therefore, basis on content, we can learn something about the meaning of the document.

For all its simplicity and intuitive clarity, this approach has a significant drawback. The Bow encoding uses a corpus (or set, collection) of words and represents any given text with a vector of the length of the corpus. If a word in the corpus is present in the text, the corresponding element of the vector would be the frequency of the word in the text.

If individual words are encoded by one-hot vectors, then the feature space will have a dimension equal to the cardinality of the collection's dictionary, i.e. tens or even hundreds of thousands. This dimension rises along with the increasing of the amount of dictionary.

3.3.2. N-grams

Another, more complex way to create a dictionary is to use grouped words. This will resize the dictionary and give Bow more details about the document.

This approach is called "N-gram". An N-gram is a sequence of any entities (words, syllable, letters, numbers, etc.). In the context of language corpora, an N-gram is usually understood as a sequence of words.

A unigram is one word, a be-gram is a sequence of two words, a trigram is three words, and so on. The number N indicates how many grouped words are included in the N-gram. Not all possible N-grams get into the model, but only those that appear in the corpus.

3.3.3. TF-IDF

Term Frequency (TF) is the ratio of the number of appearing a certain word to the total number of words in the document. Thus, the importance of a word t within a single document d_i is evaluated:

$$TF(t, d_i) = \frac{n_t}{\sum_k n_k}, \tag{1}$$

where n_t is the number of occurrences of the word t in the document d_i, and in the denominator of the fraction is the total number of words in the document.

But frequency scoring has a problem: words with the highest frequency have, accordingly, the highest score. There may not be as much information gain for the model in these words as there is in less frequent words.

One way to remedy the situation is to downgrade a word that appears frequently in all similar documents. This metric is called $TF - IDF$ (short for Term Frequency – Inverse Document Frequency).

In this metric IDF is the inverse of the frequency with which a certain word occurs in the documents of the collection:

$$IDF(t, d_i, D) = \log \frac{|D|}{|\{d_i \in D | t \in d_i\}|}. \tag{2}$$

Here $|D|$ is the number of documents in the collection (corpus), $\{d_i \in D | t \in d_i\}$ is the number of documents in the collection D that contain word t.

There is only one IDF value for each unique word within a given collection of documents. IDF metric reduces the weight of commonly corpusused words.

$TF - IDF$ is a statistical measure for estimating the importance of a word in a document that is part of a collection or corpus:

$$TF\text{-}IDF(t, d_i, D) = TF(t, d_i) \times IDF(t, d_i, D). \tag{3}$$

$TF - IDF$ scoring increases in proportion to the frequency of occurrence of the word in the document, but this is compensated by the number of documents containing this word.

The disadvantage of the frequency approach based on this metric is that it does not take into account the context of a single word. Moreover, it does not distinguish the semantic similarity of words. All vectors are equally far from each other in the feature space.

3.3.4. Word embedding

Word embedding is one of the most popular representations of document's vocabulary. This is a technique that maps words into number vectors, where words which have similar meanings will be close to each other with their vector representation in terms of some distance metric in the vector space.

Word embedding gives the impressive performance of DL methods on challenging NLP problem. Recently, several powerful word embedding models have been developed:

- Word2vec (short from Words to Vectors, provided by Google in 2013) [17];
- GloVe (short from Global Vectors, provided by Stanford University in 2014) [18];
- FastText (provided by Facebook in 2017) [19];
- BERT (short from Bidirectional Encoder Representations from Transformers, provide by Google in 2018) [40].

These models are pre-trained on large corpuses of texts, including Wikipedia and specific domain.

Word2vec is a set of ANN models designed to obtain word embedding of natural language words. It takes a large text corpus as input and maps each word to a vector, producing word coordinates as output. It first generates a dictionary of the corpus and then calculates a vector representation of the words by learning from the input texts.

The vector representation is based on contextual proximity: words that occur in the text next to the same words (and therefore have a similar meaning) will have close (by cosine distance) vectors.

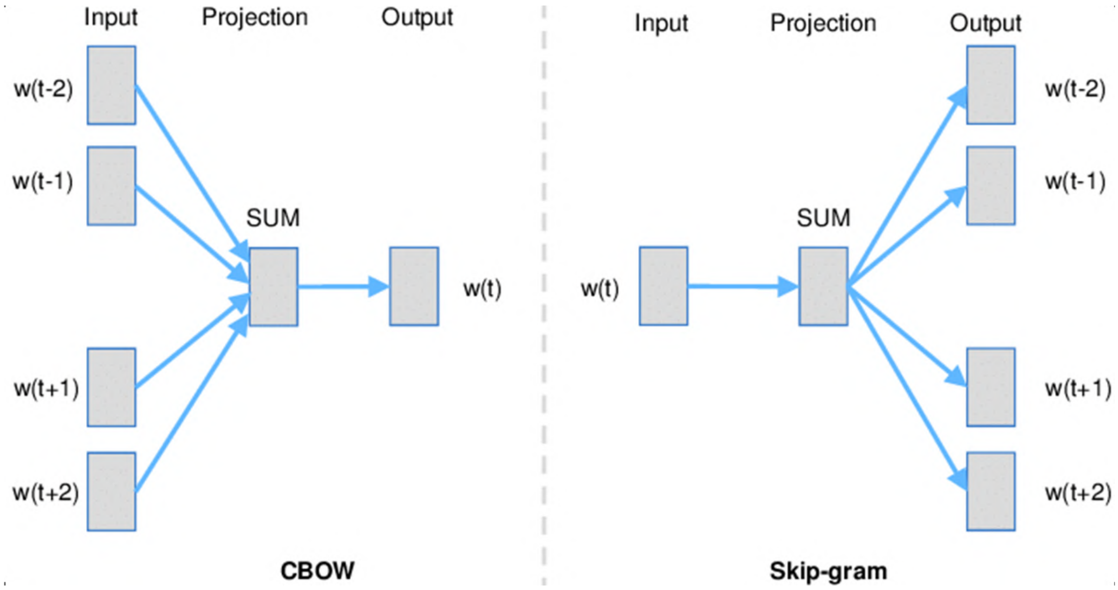


Figure 1: Simplified representation of the CBoW and Skip-gram models [17].

Word2vec implements two main learning algorithms: CBoW (Continuous Bag of Words) and Skip-gram (figure 1).

CBoW is an architecture that predicts the current word based on its surrounding context. Architecture like Skip-gram does the opposite: it uses the current word to predict surrounding words.

Building a Word2vec model is possible using these two algorithms. The word order of the context does not affect the result in any of these algorithms.

GloVe focuses on words co-occurrences over the whole corpus. Its embeddings relate to the probabilities that two words appear together. So, GloVe combines features of Word2vec and singular co-occurrence matrix decomposition.

In the present study, we applied both Word2vec and GloVe models to obtain vector representations of words.

The main application effect of using pre-trained language models is to obtain high-quality vector representations of words that take into account contextual dependencies and allow you to achieve better results on targets.

4. DNNs classification models design

After previous stage, we can start building a classification model. The model type and architecture depends on the research task of SA which can be performed at different hierarchical levels of text documents (document-level, sentence-level, word or aspect-level), domains (reviews about travel agencies, hotels, movies, election opinion prediction, analysis of public opinion on acute social and political issues), binary or multiclass classification.

If we have a dataset of texts with class labels (for example, with binary labels "positive" and "negative"), we could apply Supervised ML techniques, in particular, binary classification algorithms.

Mathematically, this problem can be formulated as follows: given training sample of texts $X = \{x_1, x_2, ...x_m\}$, for each text there is a class label $Y = \{y_i\}$, $y_i \in \{0, 1\}, i = 1, 2, ...m$.

It is necessary to build a classifier model $b(X, w): X \to Y$, where w is a vector of unknown parameters or weights.

At the same time, it is necessary to minimize the *Loss* function that determines the total deviation of real class labels from those predicted by the classifier. For binary classification problems, the most common is binary cross-entropy:

$$Loss = -\frac{1}{N}\left[\sum_{i=1}^{N}(y_i \log(p_i) + (1 - y_i)\log(1 - p_i))\right] \tag{4}$$

where N is the size of the training sample, $y_i = \{0, 1\}$ is the true class label for the i-th data sample, p_i is the probability of belonging to the positive class for the i-th data sample provided by the classifier.

4.1. Logistic regression

Since the task of SA in the general case is reduced to the binary classification problem (negative, positive), we chose the Logistic Regression (LR) model as the baseline classifier $b(\cdot)$:

$$b(X, w) = \sigma(\langle w, x \rangle), \tag{5}$$

where $\langle w, x \rangle$ – denotes the scalar product, $\sigma(\cdot)$ is a *Sigmoid* (logistic) function

$$\sigma(z) = \frac{1}{1 + \exp(-z)}. \tag{6}$$

LR has such advantage as it can be used to predict the probability to belong a training sample (in our case, tokenized and vectorized text) to one of the two target classes.

4.2. CNN model

CNNs are a class of DNNs that were originally designed for image processing [41]. But these models have shown their efficiency for many other tasks, such as time series forecasting [42].

Kim [43] has shown that CNNs are efficient for classifying texts on different datasets. Recently, they have also been used for various NLP tasks (speech generation and recognition, text summarization, named entity extraction).

The architecture of CNNs consists of convolutional and subsampling layers (figure 2).

The convolutional layer performs feature extraction from the input data and generates feature maps. The feature map is computed through an element-wise multiplication of the small matrix of weights (kernel) and the matrix representation of the input data, and the result is summed.

This weighted sum then passed through the non-linear activation function. One of the most common is the function *ReLu*, which is given as $ReLu(x) = \max(0, x)$.

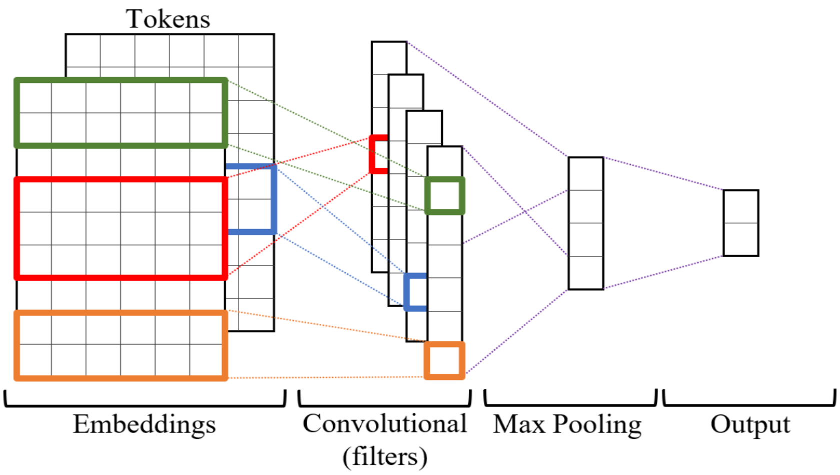


Figure 2: Convolutional Neural Network (CNN) Architecture for Text Processing [43].

The pooling (subsampling) layer is a non-linear compaction of the feature maps. For example, max-pooling takes the largest element from the feature map and extracts the sum of all its elements.

After max-pooling, feature maps are concatenated into a flatten vector, which will then be passed to a fully connected layer.

The input data for the most NLP problems is text which consists of sentences and words. So we need represent the text as an array of vectors of a certain length: each word mapped to a specific vector in a vector space composed of the entire vocabulary.

As these vectors, we can use word frequencies (for example, obtained using the $TF - IDF$ metric), or pre-trained embeddings (Word2vec, GloVe, FastText).

Unlike images processing, text convolution is performed using one-dimensional filters (1D Convolution) on one-dimensional input data, for example, sentences, using convolution kernels of different size (widths).

Applying of multiple kernels widths and feature maps is analogous to the use of N-grams.

For image processing, convolutions are usually performed on separate channels that correspond to the colors of the image: red, green, blue. Set of different filters is applied for each channel, and the result of this operation is then merged into a single vector.

For text processing as channels we can consider, for example, the sequence of words, or words embeddings. Then different kernels applied to the words can be merged into a single vector.

The final result of sentiment analysis is obtained by applying *Sigmoid* activation function (binary classification task) or *Softmax* (in the case of multi-class task).

4.3. LSTM and BiLSTM model

Sequential information and long-term dependencies in NLP traditionally performed with Recurrent Neural Networks (RNNs) which could compute context information, for example, in dependency parsing.

The most common and efficient for many ML tasks, including NLP, were architectures based on LSTM (Long Short Term Memory) or GRU (Gated Recurrent Unit) cells [37, 16].

4.3.1. LSTM

LSTM model proposed by Hochreiter and Schmidhuber [44] introduces the concept of a state for each of the layers of a RNN which plays the role of memory.

The input signal affects the state of the memory, and this, in turn, affects the output layer, just like in a RNN. But this state of memory persists throughout the time steps of a sequence (for example, time series, sentence, or text document). Therefore, each input signal affects the state of the memory as well as the output signal of the hidden layer.

LSTM cell includes several units or gates: the inputs, output, and forget gates (figure 3). These gates are used to control a memory cell that is carrying the hidden state h_t to the next time step.

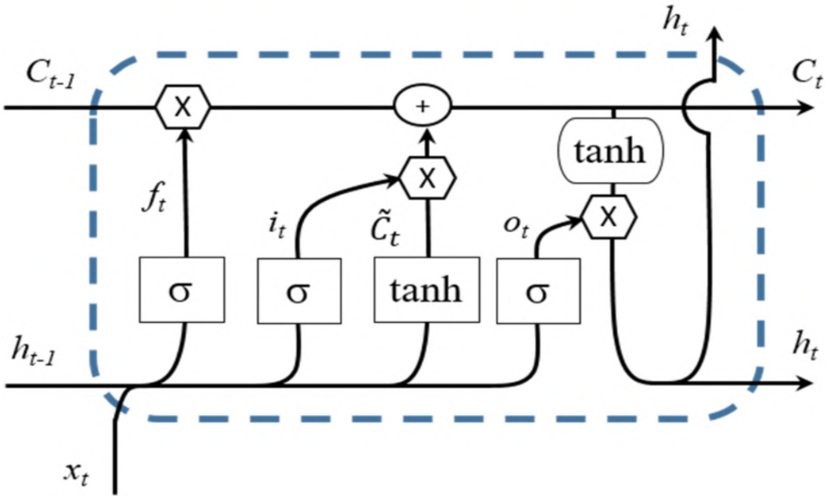


Figure 3: Diagram of a LSTM cell.

The LSTM cell is formally defined as:

$$f_t = \sigma(\mathbf{W}_f \cdot (h_{t-1}, \mathbf{x}_t) + b_f, \tag{7}$$

$$i_t = \sigma(\mathbf{W}_i \cdot (h_{t-1}, \mathbf{x}_t) + b_i, \tag{8}$$

$$\tilde{C}_t = \tanh(\mathbf{W}_c \cdot (h_{t-1}, \mathbf{x}_t) + b_c), \tag{9}$$

$$o_t = \sigma(\mathbf{W}_o \cdot (h_{t-1}, \mathbf{x}_t) + b_o), \tag{10}$$

$$a_t = i_t \otimes \tilde{C}_t, \tag{11}$$

$$C_t = f_t \otimes C_{t-1} + a_t, \tag{12}$$

where $\mathbf{x}_t$ – is the vector of input sequence at time t; C_{t-1}, h_{t-1} – state (long-term content) and hidden state in previous time step $(t-1)$ respectively; $\sigma(\cdot)$, $\tanh(\cdot)$ are the *Sigmoid* and

Hyperbolic tangent activation functions; $\otimes$ – the Kronecker product; $\mathbf{W}_f$, $\mathbf{W}_i$, $\mathbf{W}_o$ – the weight matrices for input, forget, output of the gates respectively; b_f, b_i, b_o – biases for the gates.

The *input gate* i_t determines which values need to update. Then the hyperbolic tangent layer builds a vector $\tilde{C}_t$ of new values that can be added to the state of the cell C_t.

The *forget gate* f_t controls how much is remembered (what part of the information is kept and what is erased) from step to step. Decision what information can be thrown out of the cell state is made by a sigmoid layer.

The *output gate* o_t receives an input signal (which is the concatenation of the input signal at time step t and the cell output signal at time step $(t-1)$ and passes it to the output. Thus, this gate determines which part of the long-term content C_t should be transferred to the next time step.

Each of these gates is a feed-forward neural network layer consisting of a sequence of weights fitted by the network with an activation function. This allows the network to learn the conditions for forgetting, ignoring, or keeping information in the memory cell.

Due its structure LSTM can learn and remember representations for variable length sequences, such as sentences, documents, and speech samples.

4.3.2. BiLSTM

Unidirectional (standard) LSTM only preserves information of the past because the only inputs it has seen are from the past. Unlike standard LSTM, in BiLSTM (Bidirectional LSTM) model the input flows in both directions and it's capable of utilizing information from both sides.

So BiLSTM is a sequence processing model that consists of two LSTMs layers: one taking the input in a forward direction (from "past to future"), and the other in a backwards direction (from "future to past") (figure 4).

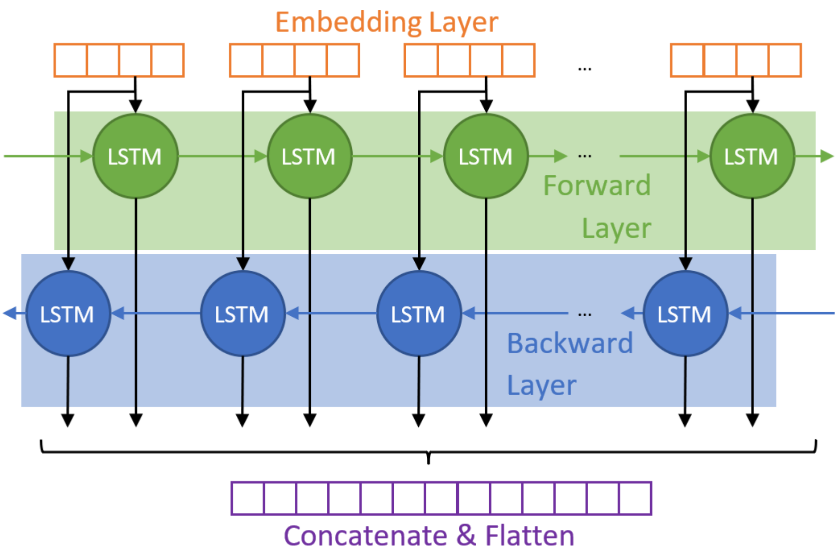


Figure 4: Diagram of a BiLSTM model.

For example, if we want to predict a word by context (the central word), the network takes a given number of words to the left of it as the context – the Forward layer performs it, as well as the words to the right of it – Backward layer performs it.

Then we can combine the outputs from both LSTM layers in different ways: as sum, average, concatenation or multiplication. This output contains the information or relation of past and future word.

BiLSTM increase the amount of information available to the network, improving the context.

It's also more powerful tool for modeling the sequential dependencies between words and phrases in both directions of the sequence than standard LSTM.

BiLSTM is usually used when we have the sequence to sequence tasks but it should be noted that BiLSTM (compared to LSTM) is a much "slower" model and requires more time for training.

4.4. CNN+LSTM model

Both basic DNNs architectures CNN and LSTM have own advantages and disadvantages. Thus, LSTM networks can capture long-term dependencies and find hidden relationships in the data. CNNs are able to extract features using different convolutions and filters.

Therefore, the combination of convolutional and recurrent layers in the model turns out to be effective in many applied problem such as simulation of various natural processes, image processing, time series forecasting, and different NLP tasks [45, 46, 47, 31, 28, 48].

So we developed two models based on modifications of CNN+LSTM architecture which final design and hyperparameters settings are given in the Section 6.

Our proposed models exploit the main features of both LSTM and CNN. In fact, LSTM could accommodate long-term dependencies and overcome the key issues with vanishing gradients. For this reason, LSTM is used when longer sequences are used as inputs. On the other hand, CNN appears able to understand local patterns and position-invariant features of a text.

5. Datasets and software implementation

All developed DNNs (CNN, CNN-LSTM, BiLSTM-CNN), and LR as the baseline, were implemented in the Python 3.8 programming language using Scikit-learn library for LR, estimation classification accuracy, and for designing DNNs models we used Keras library and TensorFlow as backend.

We evaluate the performance of our models on two datasets: Stanford's IMDb dataset (Stanford's Large Movie Review Dataset), which contains 50,000 movie reviews as well as Sentiment 140 dataset [49] with 1.6 million tweets.

Both datasets are intended for binary classification: they contain for each text (review or tweet) a sentiment class binary label. They are also balanced, i.e. contain the same number of texts for the positive and negative classes.

6. Empirical results

6.1. Pre-processing and words embeddings

For text pre-processing the Python library package NLTK [50] was used, as well as customers regular expressions.

The pre-processing stage included removing punctuations, markup tags, html and tweet addresses, removing stop words and converting all words to lower case.

Tokenization was performed by using Keras preprocessing text library. After tokenization we got the length of the vocabulary in 92393 unique tokens for IMDb dataset and 507702 for Sentiment140 respectively to which one token was added for representation out of vocabulary words.

It should be noted that the selected datasets are characterized by different average length of texts (number of words). Thus the length of most reviews does not exceed 500 words, and tweets – 50.

Since DNNs work with fixed-length input sequences we padded zero tokens all reviews and tweets which length are less than average to fixed length 500 and 50 words (tokens) respectively, and cut longer texts to these fixed sizes.

For words vector representation was used GloVe word embeddings with word vectors of dimension 100 provided by Gensim library [51].

6.2. DNNs models design and hyperparameters setting

To initialize the weights of the first layer (Embedding Layer) for all models, pre-trained GloVe embeddings of size 100 were used. These weights were frozen and did not change during training.

The first model, CNN, consists of three sequential Convolutional layers with filter sets of different kernel widths. These layers are interspersed with Maxpooling layers. Behind them are a Flatten and a Fully connected (Dense) layer.

The second, CNN-LSTM model differs from the CNN by the presence of an LSTM layer instead of a Flatten after Convolutional and Maxpooling. The base idea of such architecture is that CNN can be used to retrieve higher-level word feature sequences and LSTM to catch long-term correlations across window feature sequences, respectively.

The third, BiLSTM-CNN model contains two BiLSTM layers (forward and backward), followed by a Convolutional and Maxpooling layers. After that, two Fully connected layers were used to reduce the output dimension and make prediction.

For all models Dropout layers were also used to prevent overfitting. As the *Loss*-function Binary Cross-Entropy (4) was chosen, which can be calculated as the average cross-entropy over all data samples [52].

The final parameters of DNNs architecture are shown in table 1.

6.3. Evaluating Performance Measures

The datasets were divided in the proportion of: 64% for training, 20% for validation, and 16% for test subsets respectively.

Table 1
Final DNNs models hyperparameters setting.

Model	Layers	Parameters
CNN	Embedding	emb_dim 100, sent_len 500(50)
	Dropout	0.3
	Convolutional 1D	100 filters of size 2, act_func ReLu
	Max pooling	pool_size 2
	Convolutional 1D	100 filters of size 3, act_func ReLu
	Max pooling	pool_size 2
	Convolutional 1D	100 filters of size 4, act_func ReLu
	Max pooling	pool_size 2
	Flatten	
	Dropout	0.3
	Fully connected	1 neuron, act_func Sigmoid
CNN-LSTM	Embedding	emb_dim 100, sent_len 500(50)
	Dropout	0.3
	Convolutional 1D	50 filters of size 2, act_func ReLu
	Max pooling	pool_size 2
	Convolutional 1D	100 filters of size 2, act_func ReLu
	Max pooling	pool_size 2
	Convolutional 1D	200 filters of size 2, act_func ReLu
	Max pooling	pool_size 2
	LSTM	64 neurons, reccur_dropout 0.3
	Dropout	0.3
	Fully connected	32 neurons
	Fully connected	1 neuron, act_func Sigmoid
BiLSTM-CNN	Embedding	Emb_dim 100, sent_len 500(50)
	Dropout	0.3
	Bidirectional	LSTM with 100 neurons
	Dropout	0.3
	Bidirectional	LSTM with 100 neurons
	Dropout	0.3
	Convolutional 1D	100 filters of size 3, act_func ReLu
	Global Max pooling 1D	
	Fully connected	10 neurons, act_func ReLu
	Fully connected	1 neuron, act_func Sigmoid

All DNNs models were trained over 5 epochs with a minibatch size of 256 and 1024 samples for IMDb and Sentiment 140 respectively. To compare classification performance of the developed models we used the *Accuracy* metrics given by:

$$Accuracy = \frac{TP + TN}{P + N} \times 100\%, \tag{13}$$

where TP and TN are the number of correctly predicted values of the positive and negative classes, respectively; P and N are the actual number of values for each of the classes.

We also calculated $F1$-*score* which is harmonic average between *Precision* (the percentage of objects in the positive class, which were classified as positive, are correctly classified), and

Recall (percentage of objects of the true positive class which we correctly classified):

$$F1\text{-}score = \frac{2TP}{2TP + FP + FN}, \tag{14}$$

$$Precision = \frac{TP}{TP + FP}, \tag{15}$$

$$Recall = \frac{TP}{TP + FN}. \tag{16}$$

Here FP (False Positive) and FN (False Negative) are numbers of times (data samples) where the model incorrectly classified these samples as belonging to the positive and negative classes respectively.

The final results of classification performance are presented in tables 2-3.

Table 2
Classification performance on IMDb dataset,%.

Models	*Precision*	*Recall*	*F1-score*	*Accuracy*
LR (baseline)	86.62	85.54	86.08	85.90
CNN	90.04	90.31	**90.18**	**90.09**
CNN-LSTM	**90.90**	84.84	87.76	88.08
BiLSTM-CNN	83.08	**93.25**	87.87	87.03

Table 3
Classification performance on Sentiment 140 dataset, %.

Models	*Precision*	*Recall*	*F1-score*	*Accuracy*
LR (baseline)	71.61	74.63	73.09	74.23
CNN	76.17	79.47	77.78	77.24
CNN-LSTM	78.98	77.47	78.23	78.37
BiLSTM-CNN	**79.54**	**84.41**	**81.91**	**82.10**

Classification performance on IMDb dataset for all developed DNN models is better than baseline. The best *Accuracy* metric was obtained using the CNN model (90.09%). At the same time, models based on the combination of Convolutional and LSTM layers showed an *Accuracy* of 2-3% less (table 2).

It should be noted that obtained results are comparable or even superior in accuracy to the results given by other researchers [33, 34, 53] for IMDb dataset.

All models showed significantly lower accuracy (on average 10% less) on the dataset Sentiment 140 (table 3). The best result was achieved for the BiLSTM-CNN model – *Accuracy* 82.1%.

At the same time, the complication of models by adding new layers did not lead to a significant increase in accuracy, but prolonged the training time.

In our opinion, lower accuracy may be due to the fact that Sentiment 140 dataset contains many slang words that are out of vocabulary. So, if for IMDb dataset the part of the missing words was about 30 percent, then for the Sentiment 140 this part was more than 70.

7. Discussion

Our research sheds light on the effectiveness of relatively uncomplicated Deep Neural Networks (DNNs) architectures with a modest layer count for sentiment analysis of social media texts, particularly within binary classification scenarios. These models exhibit a level of accuracy that is sufficiently practical for real-world applications.

In the case of the English-language datasets, IMDb and Sentiment 140, our models showcased the following classification accuracy rates: Logistic Regression (Baseline) achieved 85.9% (74.23%), CNN achieved 90.09% (77.24%), CNN-LSTM reached 88.01% (78.36%), and BiLSTM-CNN attained 87.03% (82.10%).

Notably, preprocessing steps like lemmatization or stemming can likely boost classification accuracy. This becomes especially relevant for tweets, which frequently feature an array of user-generated vocabulary.

Another avenue for potential improvement involves utilizing word embeddings weighted by their Term Frequency-Inverse Document Frequency (TF-IDF) metric. Addressing out-of-vocabulary words could involve strategies like employing the weighted average value of neighboring word embeddings within a designated window length or substituting missing words with normalized TF-IDF embeddings transformed via principal component analysis (SVD decomposition of the sparse TF-IDF matrix to reduce dimensionality).

In our perspective, an exciting trajectory for advancing sentiment analysis in social media involves the utilization of models rooted in deep convolutional networks or the amalgamation of convolutional and recurrent networks. Coupling these models with pre-trained embeddings, such as those founded on GloVe, Word2Vec, and FastText, holds promise. Leveraging pre-trained embeddings allows the initialization of DNNs with parameters that are already somewhat attuned to the text classification task, accelerating the learning process and enhancing the generalization capabilities of classifiers founded on deep networks.

8. Conclusion

In conclusion, our research illuminates the efficacy of employing relatively straightforward Deep Neural Network (DNN) architectures for sentiment analysis in the context of social media text. Our findings underscore that even DNNs with limited complexity can yield accuracy levels suitable for practical applications in binary sentiment classification.

Through experimentation on the IMDb and Sentiment 140 datasets, we observed compelling classification accuracy results: Logistic Regression (Baseline) achieved 85.9% (74.23%), CNN achieved 90.09% (77.24%), CNN-LSTM reached 88.01% (78.36%), and BiLSTM-CNN attained 87.03% (82.10%).

Enhancing the preprocessing steps with techniques like lemmatization and incorporating weighted word embeddings via TF-IDF are potential strategies to further refine classification accuracy. Additionally, the combination of deep convolutional and recurrent networks, complemented by pre-trained embeddings, emerges as a promising avenue for advancing sentiment analysis in social media. Pre-trained embeddings not only expedite learning but also enhance the classifier's ability to generalize.

References

[1] V. Derbentsev, V. Bezkorovainyi, A. Matviychuk, O. Pomazun, A. Hrabariev, A. Hostryk, Sentiment analysis of electronic social media based on deep learning, in: S. Semerikov, V. Soloviev, A. Matviychuk, V. Kobets, L. Kibalnyk, H. Danylchuk, A. Kiv (Eds.), Proceedings of 10th International Conference on Monitoring, Modeling & Management of Emergent Economy - M3E2, INSTICC, SciTePress, 2023, pp. 163–175. doi:10.5220/0011932300003432.

[2] M. Azlinah, B. W. Yap, J. M. Zain, M. W. Berry (Eds.), Soft Computing in Data Science 6th International Conference, SCDS 2021, volume 1489 of *SCDS: International Conference on Soft Computing in Data Science*, Springer, Singapore, 2021. doi:10.1007/978-981-16-7334-4.

[3] W. Mayur, C. S. R. Annavarapu, K. Chaitanya, A survey on sentiment analysis methods, applications, and challenges, Artificial Intelligence Review 55 (2022) 5731–5780. doi:10.1007/s10462-022-10144-1.

[4] M. Silberztein, F. Atigui, E. Kornyshova, E. Métais, F. Meziane (Eds.), Natural Language Processing and Information Systems: Proceedings of 23rd International Conference on Applications of Natural Language to Information Systems, volume 10859 of *Lecture Notes in Computer Science*, Springer, Cham, 2018. doi:10.1007/978-3-319-91947-8.

[5] C. A. Iglesias, A. Moreno (Eds.), Sentiment Analysis for Social Media, MDPI, 2020. doi:10.3390/books978-3-03928-573-0.

[6] F. Pozzi, E. Fersini, E. Messina, B. Liu, Sentiment Analysis in Social Networks, Elsevier Science, 2016.

[7] H. Karamollaoğlu, İ. A. Doğru, M. Dörterler, A. Utku, O. Yıldız, Sentiment Analysis on Turkish Social Media Shares through Lexicon Based Approach, in: 2018 3rd International Conference on Computer Science and Engineering (UBMK), 2018, pp. 45–49. URL: https://ieeexplore.ieee.org/document/8566481.

[8] C. Dhaoui, C. M. Webster, L. P. Tan, Social media sentiment analysis: lexicon versus machine learning, Journal of Consumer Marketing 34 (2017) 480–488. doi:10.1108/JCM-03-2017-2141.

[9] C. S. Khoo, S. B. Johnkhan, Lexicon-based sentiment analysis: Comparative evaluation of six sentiment lexicons, Journal of Information Science 44 (2018) 491–511. doi:10.1177/0165551517703514.

[10] D. Alessia, F. Ferri, P. Grifoni, T. Guzzo, Approaches, tools and applications for sentiment analysis implementation, International Journal of Computer Applications 125 (2015) 26–33.

[11] A. Kiv, S. Semerikov, V. N. Soloviev, L. Kibalnyk, H. Danylchuk, A. Matviychuk, Experimental Economics and Machine Learning for Prediction of Emergent Economy Dynamics, in: A. Kiv, S. Semerikov, V. N. Soloviev, L. Kibalnyk, H. Danylchuk, A. Matviychuk (Eds.), Proceedings of the Selected Papers of the 8th International Conference on Monitoring, Modeling & Management of Emergent Economy, M3E2-EEMLPEED 2019, Odessa, Ukraine, May 22-24, 2019, volume 2422 of *CEUR Workshop Proceedings*, CEUR-WS.org, 2019, pp. 1–4. URL: https://ceur-ws.org/Vol-2422/paper00.pdf.

[12] V. Derbentsev, A. Matviychuk, V. N. Soloviev, 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, Springer, Singapore, 2020, pp.

211–231. doi:10.1007/978-981-15-4498-9_12.
[13] P. V. Zahorodko, S. O. Semerikov, V. N. Soloviev, A. M. Striuk, M. I. Striuk, H. M. Shalatska, Comparisons of performance between quantum-enhanced and classical machine learning algorithms on the IBM Quantum Experience, Journal of Physics: Conference Series 1840 (2021) 012021. doi:10.1088/1742-6596/1840/1/012021.
[14] H. Li, Deep learning for natural language processing: advantages and challenges, National Science Review 5 (2017) 24–26. doi:10.1093/nsr/nwx110.
[15] K. W. Trisna, H. J. Jie, Deep Learning Approach for Aspect-Based Sentiment Classification: A Comparative Review, Applied Artificial Intelligence 36 (2022) 2014186. doi:10.1080/08839514.2021.2014186.
[16] U. Kamath, J. Liu, J. Whitaker, Deep Learning for NLP and Speech Recognition, Springer, Cham, 2019. doi:10.1007/978-3-030-14596-5.
[17] T. Mikolov, K. Chen, G. Corrado, J. Dean, Efficient Estimation of Word Representations in Vector Space, in: Y. Bengio, Y. LeCun (Eds.), 1st International Conference on Learning Representations, ICLR 2013, Scottsdale, Arizona, USA, May 2-4, 2013, Workshop Track Proceedings, 2013. URL: https://arxiv.org/abs/1301.3781.
[18] J. Pennington, R. Socher, C. Manning, GloVe: Global Vectors for Word Representation, in: Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP), Association for Computational Linguistics, Doha, Qatar, 2014, pp. 1532–1543. doi:10.3115/v1/D14-1162.
[19] P. Bojanowski, E. Grave, A. Joulin, T. Mikolov, Enriching Word Vectors with Subword Information, Transactions of the Association for Computational Linguistics 5 (2017) 135–146. doi:10.1162/tacl_a_00051.
[20] A. K. Durairaj, A. Chinnalagu, Transformer based Contextual Model for Sentiment Analysis of Customer Reviews: A Fine-tuned BERT, International Journal of Advanced Computer Science and Applications 12 (2021). doi:10.14569/IJACSA.2021.0121153.
[21] M. P. Geetha, D. Karthika Renuka, Improving the performance of aspect based sentiment analysis using fine-tuned Bert Base Uncased model, International Journal of Intelligent Networks 2 (2021) 64–69. doi:10.1016/j.ijin.2021.06.005.
[22] H. Deng, D. Ergu, F. Liu, Y. Cai, B. Ma, Text sentiment analysis of fusion model based on attention mechanism, Procedia Computer Science 199 (2022) 741–748. doi:10.1016/j.procs.2022.01.092.
[23] S. Tabinda Kokab, S. Asghar, S. Naz, Transformer-based deep learning models for the sentiment analysis of social media data, Array 14 (2022) 100157. doi:https://doi.org/10.1016/j.array.2022.100157.
[24] Z. Drus, H. Khalid, Sentiment Analysis in Social Media and Its Application: Systematic Literature Review, Procedia Computer Science 161 (2019) 707–714. doi:10.1016/j.procs.2019.11.174.
[25] P. K. Jain, R. Pamula, G. Srivastava, A systematic literature review on machine learning applications for consumer sentiment analysis using online reviews, Computer Science Review 41 (2021) 100413. doi:10.1016/j.cosrev.2021.100413.
[26] P. Sudhir, V. D. Suresh, Comparative study of various approaches, applications and classifiers for sentiment analysis, Global Transitions Proceedings 2 (2021) 205–211. doi:https://doi.org/10.1016/j.gltp.2021.08.004.

[27] M. Biesialska, K. Biesialska, H. Rybinski, Leveraging contextual embeddings and self-attention neural networks with bi-attention for sentiment analysis, Journal of Intelligent Information Systems 57 (2021) 601–626. doi:10.1007/s10844-021-00664-7.
[28] A. Rasool, Q. Jiang, Q. Qu, C. Ji, WRS: A Novel Word-embedding Method for Real-time Sentiment with Integrated LSTM-CNN Model, in: 2021 IEEE International Conference on Real-time Computing and Robotics (RCAR), 2021, pp. 590–595. doi:10.1109/RCAR52367.2021.9517671.
[29] H. Elzayady, M. S. Mohamed, S. Badran, Integrated bidirectional LSTM-CNN model for customers reviews classification, Journal of Engineering Science and Military Technologies 5 (2021). doi:10.21608/EJMTC.2021.66626.1172.
[30] N. Hernández, I. Batyrshin, G. Sidorov, Evaluation of deep learning models for sentiment analysis, Journal of Intelligent & Fuzzy Systems (2022) 1–11. doi:10.3233/JIFS-211909.
[31] L. Khan, A. Amjad, K. M. Afaq, H.-T. Chang, Deep Sentiment Analysis Using CNN-LSTM Architecture of English and Roman Urdu Text Shared in Social Media, Applied Sciences 12 (2022) 2694. doi:10.3390/app12052694.
[32] I. Priyadarshini, C. Cotton, A novel LSTM–CNN–grid search-based deep neural network for sentiment analysis, The Journal of Supercomputing 77 (2021) 13911–13932. doi:10.1007/s11227-021-03838-w.
[33] M. R. Haque, H. Salma, S. A. Lima, S. M. Zaman, Performance Analysis of Different Neural Networks for Sentiment Analysis on IMDb Movie Reviews, 2020. URL: https://www.researchgate.net/publication/343046458.
[34] A. H. Quraishi, Performance Analysis of Machine Learning Algorithms for Movie Review, International Journal of Computer Applications 177 (2020) 7–10. doi:10.5120/ijca2020919839.
[35] V. Derbentsev, V. Bezkorovainyi, R. Akhmedov, Machine Learning Approach of Analysis Emotion Polarity Electronic Social Media, Neiro-Nechitki Tekhnolohii Modelyuvannya v Ekonomitsi 9 (2020).
[36] Y. Bengio, R. Ducharme, P. Vincent, C. Jauvin, A neural probabilistic language model, Journal of Machine Learning Research 3 (2003) 1137–1155. URL: https://proceedings.neurips.cc/paper/2000/file/728f206c2a01bf572b5940d7d9a8fa4c-Paper.pdf.
[37] J. Brownlee, Develop Deep Learning Models for Natural Language in Python. Deep Learning for Natural Language Processing, 2017. URL: http://ling.snu.ac.kr/class/AI_Agent/deep_learning_for_nlp.pdf.
[38] L. Hobson, H. Cole, H.Hannes, Natural Language Processing in Action: Understanding, analyzing, and generating text with Python, Manning Publications, 2019.
[39] J. Camacho-Collados, M. T. Pilehvar, On the Role of Text Preprocessing in Neural Network Architectures: An Evaluation Study on Text Categorization and Sentiment Analysis, in: Proceedings of the 2018 EMNLP Workshop BlackboxNLP: Analyzing and Interpreting Neural Networks for NLP, Association for Computational Linguistics, Brussels, Belgium, 2018, pp. 40–46. doi:10.18653/v1/W18-5406.
[40] J. Devlin, M.-W. Chang, K. Lee, K. Toutanova, BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding, 2018. URL: https://arxiv.org/abs/1810.04805.
[41] Y. LeCun, Y. Bengio, Convolutional Networks for Images, Speech, and Time Series, in: The Handbook of Brain Theory and Neural Networks, MIT Press, Cambridge, MA, USA,

1998, p. 255–258.
[42] Y. LeCun, Y. Bengio, G. Hinton, Deep learning, Nature 521 (2015) 436–444.
[43] Y. Kim, Convolutional Neural Networks for Sentence Classification, in: A. Moschitti, B. Pang, W. Daelemans (Eds.), Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing, EMNLP 2014, October 25-29, 2014, Doha, Qatar, A meeting of SIGDAT, a Special Interest Group of the ACL, ACL, 2014, pp. 1746–1751. doi:10.3115/v1/d14-1181.
[44] S. Hochreiter, J. Schmidhuber, Long Short-Term Memory, Neural Computation 9 (1997) 1735–1780. doi:10.1162/neco.1997.9.8.1735.
[45] N. Chen, P. Wang, Advanced Combined LSTM-CNN Model for Twitter Sentiment Analysis, in: 2018 5th IEEE International Conference on Cloud Computing and Intelligence Systems (CCIS), 2018, pp. 684–687. doi:10.1109/CCIS.2018.8691381.
[46] V. Derbentsev, V. Bezkorovainyi, M. Silchenko, A. Hrabariev, O. Pomazun, Deep Learning Approach for Short-Term Forecasting Trend Movement of Stock Indeces, in: 2021 IEEE 8th International Conference on Problems of Infocommunications, Science and Technology (PIC S&T), 2021, pp. 607–612. doi:10.1109/PICST54195.2021.9772235.
[47] M. Z. Islam, M. M. Islam, A. Asraf, A combined deep CNN-LSTM network for the detection of novel coronavirus (COVID-19) using X-ray images, Informatics in Medicine Unlocked 20 (2020) 100412. doi:10.1016/j.imu.2020.100412.
[48] L. Shang, L. Sui, S. Wang, D. Zhang, Sentiment analysis of film reviews based on CNN-BLSTM-attention, Journal of Physics: Conference Series 1550 (2020) 032056. doi:10.1088/1742-6596/1550/3/032056.
[49] Kaggle, Sentiment140 dataset with 1.6 million tweets, 2022. URL: https://www.kaggle.com/datasets/kazanova/sentiment140.
[50] NLTK Project, Natural language toolkit, 2022. URL: https://www.nltk.org/.
[51] R. Řehůřek, Gensim: Topic modelling for humans, 2022. URL: https://radimrehurek.com/gensim/.
[52] A. Geron, Hands-On Machine Learning with Scikit-Learn and TensorFlow, O'Reilly Media, Inc., 2017.
[53] N. M. Ali, M. M. A. El Hamid, A. Youssif, Sentiment Analysis for Movies Reviews Dataset Using Deep Learning Models, International Journal of Data Mining & Knowledge Management Process (IJDKP) 9 (2019). URL: https://aircconline.com/abstract/ijdkp/v9n3/9319ijdkp02.html.

The impact of the war in Ukraine on globalization processes and world financial markets: a wavelet entropy analysis*

Hanna B. Danylchuk[1], Liubov O. Kibalnyk[1], Oksana A. Kovtun[2], Oleg I. Pursky[3], Yevhenii M. Kyryliuk[1] and Olena O. Kravchenko[1]

[1]The Bohdan Khmelnytsky National University of Cherkasy, 81 Shevchenko Blvd., Cherkasy, 18031, Ukraine
[2]University of Educational Management, 52A Sichovykh Striltsiv Str., Kyiv, 04053, Ukraine
[3]State University of Trade and Economics, 9 Kyoto Str., 02156, Kyiv, Ukraine

Abstract
This paper is an applied research that aims to model and analyze how the war in Ukraine influenced the globalization processes and the world financial markets. This topic is relevant but underexplored in the literature. We used the wavelet entropy method to build models for the markets of natural gas, oil, gasoline, and currency pairs EUR/USD, GBP/USD. Wavelet entropy is a measure of complexity and uncertainty of unsteady signals or systems in both time and frequency domains. Our results show that the war in Ukraine was a source of crises in the studied markets and a factor that reshaped the world economic space.

Keywords
globalization processes, global financial markets, oil market, natural gas market, currency markets, crisis impact, wavelet entropy, war in Ukraine

1. Introduction

The turn of the 20th and 21st centuries has witnessed intensified scholarly interest in the challenges posed by globalization processes amid various crisis phenomena, fostering theoretical and methodological explorations in forecasting, analysis, and modeling. The trajectory of globalization theories, transitioning from Keynesian paradigms to neoliberal constructs throughout the 20th century, has underpinned the establishment of the post-industrial economy. The prevalence of contemporary crises—ranging from warfare and the COVID-19 pandemic to the quest for national self-determination, hunger, income disparities, and ecological, energy,

M3E2-MLPEED 2022: The 10th International Conference on Monitoring, Modeling & Management of Emergent Economy, November 17-18, 2022, Virtual, Online

*Extended and revised version of paper [1] presented at the 10th International Conference on Monitoring, Modeling & Management of Emergent Economy.

abdanilchuk@gmail.com (H. B. Danylchuk); liubovkibalnyk@gmail.com (L. O. Kibalnyk); kovtun.oa71@gmail.com (O. A. Kovtun); Pursky_O@ukr.net (O. I. Pursky); en_kirilyk@ukr.net (Y. M. Kyryliuk); olena_kravchenko17@ukr.net (O. O. Kravchenko)
0000-0002-9909-2165 (H. B. Danylchuk); 0000-0001-7659-5627 (L. O. Kibalnyk); 0000-0002-0159-730X (O. A. Kovtun); 0000-0002-1230-0305 (O. I. Pursky); 0000-0001-7097-444X (Y. M. Kyryliuk); 0000-0002-8776-4462 (O. O. Kravchenko)

raw material, food, and demographic quandaries—attests to the state of crisis confronting the modern world.

These multifaceted predicaments have prompted conjectures that 21st-century economic growth will persist but chart novel directions, characterized by qualitative shifts in services, digitization, and transformations in scientific and technological progress. The currents of globalization are manifest in trends such as the division of world markets into core and peripheral domains, spawning divergent interests between hegemonic and peripheral nations. The consolidation of national economies and societies within robust regional frameworks, the ebb and flow of income polarization tied to increased productivity, swift capital mobility, financial elite-driven speculation, and the paradoxical dynamics between the virtual and real economic sectors typify globalization's canvas. Furthermore, the imperative to coalesce against terrorism and global crises compounds these intricate trends.

Maurice Allais, the renowned French economist and Nobel laureate, contemplated the comprehensive globalization of trade between nations with disparate wage levels. He envisioned an outcome marked by unemployment, declining economic growth, inequality, and poverty—a perspective that stirs questions concerning globalization's necessity and desirability [2].

This dichotomous nature of globalization, yielding both positive and adverse repercussions for the world economy, is exemplified in the erosion of sovereign, economic, political, and energy independence among nations. The ripples of financial crises travel swiftly across regions, exerting significant influence on dependent economies, amplified by political, food, and energy upheavals. The ongoing war in Ukraine, coupled with the concomitant blockade of its seaports, has underscored the vulnerability of grain-importing nations to potential food crises [3]. Concurrently, the surge in forced migration and escalating unemployment further underscores globalization's complex fallout.

Today, financial and investment spheres exemplify the zenith of globalization. Financial flows cascade through the global economy, primarily via financial markets—largely detached from tangible goods and services markets. This intricate interplay intermittently begets financial crises, eroding financial systems, and culminating in socio-economic, demographic, and financial instabilities. The ripples of regional financial crises [4], the resounding echo of the global COVID-19 crisis [5], and the recurring turmoil in markets like the USA and China illuminate this reality.

The military-industrial complex's upswing following the 9/11 attacks, propelled by NATO-led wars in Afghanistan and Iraq, demonstrates the interdependence between war, crises, and economic dynamics. The ongoing war in Ukraine, coupled with military assistance, has spurred heightened production within the military-industrial complexes of the US and specific European countries. The associated demand for financial investment reverberates in global and regional financial markets. Thus, the salience of this research topic is evident, demanding a comprehensive toolkit to decipher evolving trends within the globalized financial system, particularly financial markets.

In this context, the analysis and modeling of globalization processes during crisis phases, with ramifications for financial market states and trajectories, assumes paramount importance. A substantial body of work, from both domestic and international scholars, attends to this scientific challenge. For instance, the bankruptcy rates of Turkish banking institutions vis-à-vis the deep-rooted financial crisis were examined using diverse performance indicators via stochastic methods, such as frontier analysis and data coverage analysis [6].

The European sovereign debt crisis prompted statistical examinations of financial relationships to model bond market yield movements [7]. This elucidated long-term and short-term contagion effects, particularly pronounced in peripheral countries after the crisis's acute phase. Many investigations have delved into modeling the yield, volatility, and risk profiles of diverse financial instruments within financial markets, employing an array of methodologies.

Cross-quantile analyses explored the intricate relationships between developed and emerging market stock returns, unveiling nuanced time-varying characteristics [8]. The dynamics of illiquidity within developed stock markets during and post the global financial crisis were modeled using a multiplicative error model, revealing pronounced interdependencies in volatility and illiquidity, especially during crises [9]. GARCH models unraveled the far-reaching impact of COVID-19 on the precious metals market, exposing its long memory effect [10].

The high-dimensional conditional Value-at-Risk (CoVaR), which is based on the LASSO-VAR model, is used to study the systemic risks of financial contagion in crisis situations using the example of oil markets and G20 stock markets [11]. The authors proved that in the event of a crisis in the oil markets, the stock markets of those countries that are connected with oil production will experience the greatest shocks.

Changes in the environment and depletion of natural resources have led to investment in renewable energy sources, and therefore to the need to analyze herd (collective) behavior in this market [12]. In the article, the authors presented the results of testing the collective behavior of the renewable energy market using an empirical model during the periods of the global financial crisis and the coronavirus crisis. The authors proved the herd behavior of market participants during periods of crises in the oil markets. As a result, there is an invigoration of collective behavior in the stock markets as well. Attention is also paid to the study of contagion and the emergence of risks from fossil fuel energy markets to renewable energy stock markets.

The burgeoning interest in monitoring, modeling, and forecasting financial markets during crisis episodes has propelled the adoption of nonlinear dynamics tools. Fractal and entropy analyses uncovered trends in the cryptocurrency market [13] during the COVID-19 pandemic, serving as effective crisis identifiers [14]. Quantum models, exemplified by the heterogeneous economic model, offered insights into the flow and aftermath of various crises, thereby enriching comparisons.

The articles [15, 16] are devoted to the identification of special conditions in the cryptocurrency market. The authors classified and adapted quantitative indicators to this market, analyzed their behavior in the conditions of critical events and well-known cryptocurrency market crashes.

Danylchuk et al. [17] use entropy methods to determine the investment attractiveness of countries. For this purpose, regional stock markets are studied, as they are a reflection of the economies of countries.

Quantum modeling, namely the heterogeneous economic model, has been applied to stock markets [18]. With the help of "measurement of the temperature of the series" crisis periods in the markets were detected. This model made it possible to adequately compare the features of the flow and consequences of various crises.

Modeling the impact of geopolitical risks on the state and dynamics of financial markets under conditions of crises of various natures is a little-researched field. This issue becomes especially relevant in the context of the creation of political and economic alliances and recent political crises. Choi [19] presents the results of using the method of multiple and partial wavelet-coherent analysis regarding the influence of geopolitical problems on stock markets in the countries of Northeast Asia.

Abdel-Latif and El-Gamal [20] investigate the global dynamic interrelationship between the prices of petroleum products, oil, financial liquidity, geopolitical risk and economic indicators of the economies of countries dependent on oil exports. For this purpose, the authors use the global vector autoregression (GVAR) model.

The full-fledged war in Ukraine has spurred inquiries into its financial market implications. Empirical evidence substantiates the war's deleterious impact on global stock market profitability, significantly affecting markets in geographically proximate countries, as well as those denouncing the war. Boungou and Yatié [21] provide empirical evidence of the negative impact of the war in Ukraine on the profitability of the global stock market. The largest decrease in the indicator was demonstrated by the markets of those countries geographically bordering Ukraine and Russia, as well as countries that condemned the war.

The war's influence on financial markets, particularly concerning countries reliant on Russian goods, is also studied [22]. Results indicate increased instability across markets, proportionate to a country's dependence on Russian goods.

The extent of globalization's influence on financial markets during crises remains a subject necessitating thorough investigation. While globalized markets appear more vulnerable, the reactions of US and Asian markets vary [23].

The study of modern crises—political, social, military, and pandemic—has engendered shifts in globalization patterns within financial markets, warranting a rigorous exploration. Classical analytical methods, however, often fall short in fully assessing and predicting these intricate dynamics. Consequently, the exigency of a comprehensive, interdisciplinary approach is evident in addressing this complex scientific endeavor.

2. Research methods

In this study, the wavelet entropy method is used to model and analyze the impact of the war in Ukraine on globalization processes using the example of the gas, oil, petroleum products, and currency markets. The method of wavelet transformations is proposed for the analysis of periods in time series with the aim of detecting the evolution of parameters [24]. Wavelet analysis based on wavelet entropy allows obtaining information about dynamic complexity [25].

We can describe wavelet entropy based on the work of Zunino et al. [26]. When studying the time series, which consists of sample values x_i, $i = 1, ..., M$, when using a set of scales $1, ..., N$, we will get a wavelet transformation (expansion)

$$X(t) = \sum_{j=1}^{N} \sum_{k} C_j \psi_{j,k}(t) = \sum_{j=1}^{N} r_j(t), \tag{1}$$

$r_j(t)$ contains information about the series X in scale 2^{j-1} and 2^j.

Application of the theory of Fourier expansions allows us to determine the energy on each scale using

$$E_j = \|r_j\|^2 = \sum_k |C_j(k)|^2. \tag{2}$$

The total energy of the series can be calculated by

$$E_{tot} = \|X\|^2 = \sum_{j=1}^{N} \sum_k |C_j(k)|^2 = \sum_{j=1}^{N} E_j. \tag{3}$$

The next step is to determine the relative wavelet energy

$$p_j = \frac{E_j}{E_{tot}}, \tag{4}$$

which provides hidden characteristics of the series in time and frequency spaces.

Using the concept of Shannon entropy, we can determine the normalized total wavelet entropy

$$E_{WT} = \frac{-\sum_{j=1}^{N} p_j \ln p_j}{X_{max}}. \tag{5}$$

The improvement of the wavelet entropy calculation algorithm was the use of a window procedure [27]. The following formula is used to calculate the wavelet energy for a time window

$$E_j^{(i)} = \sum_{k=(i-1)L+1}^{i\cdot} |C_j(k)|^2, i = 1, ..., N_T. \tag{6}$$

The total energy in the window is calculated by

$$E_{tot}^{(i)} = \sum_{j=-N}^{-1} E_j^{(i)}. \tag{7}$$

The change in time of relative wavelet energy and normalized total wavelet entropy is obtained by

$$p_j^{(i)} = \frac{E_j^{(i)}}{E_{tot}^{(i)}}, E_{WT}^{(i)} \sum_{j=-N}^{-1} p_j^{(i)} \cdot \frac{\ln p_j^{(i)}}{X_{max}}. \tag{8}$$

3. Results and discussions

Oil is considered to be the benchmark of world economic activity. The price of crude oil reflects such market properties as stability/volatility and liquidity.

The article examines the oil, gas and gasoline market. The most popular grades of oil are Brent and West Texas Intermediate (WTI). For this purpose, daily values of Brent and WTI brand oil indices, natural gas and gasoline for the period from January 2015 to September 2022 were used. All calculations were performed in Matlab. Calculation parameters: window width

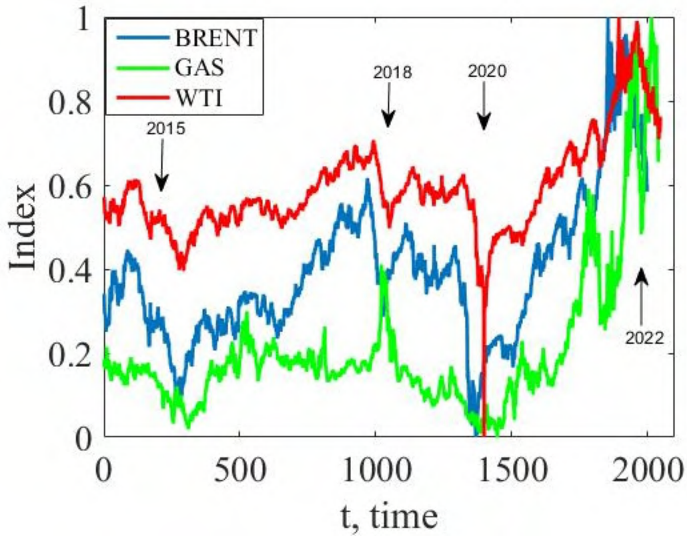


Figure 1: Comparative dynamics of oil (Brent and WTI) and gas indices.

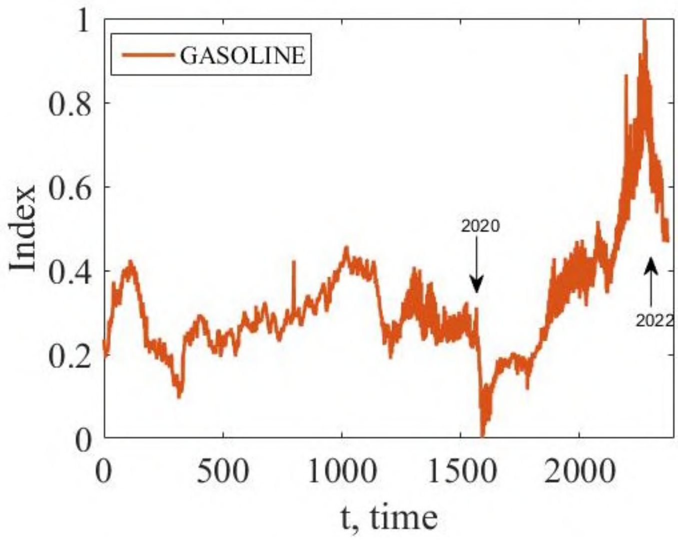


Figure 2: Dynamics of the gasoline index.

100 points, step – 10 points. Calculations were made according to the official website Yahoo Finance [28].

In figures 1, 2 shows the dynamics of indices. Arrows indicate the periods of 2020 (the beginning of the coronavirus pandemic) and 2022 (the beginning of the war in Ukraine).

From figures 1, 2 we can note 2020 a drop in oil and gasoline indices. And in 2022, all indices experienced a rapid decline. The situation regarding 2020 is quite obvious and understandable. The announcement of the pandemic halted and slowed down economic activity. Demand for oil and gasoline fell.

The fall in 2022 is due to various factors, but in our opinion, the war in Ukraine should be considered the main one. Although the events unfold on the territory of Ukraine, the consequences are felt by almost all countries. European Union countries, Great Britain, the USA, Turkey, etc. support Ukraine not only with military aid, but also with the introduction of political and economic sanctions. Russia was a strong player in the oil and gas markets. The introduction of sanctions, the refusal of Russian gas forces the market and all market participants to quickly reorient themselves and reformat connections (e.g. increasing oil production in Norway, expected deliveries from Nigeria and Venezuela).

The use of wavelet entropy is due to the illustrative nature of this indicator and its predictive properties. The formation of three increasing entropy wavelet waves is a proven indicator-precursor of crisis phenomena of various natures [29]. As soon as the third wave exceeded the maximum of the second wave, it can be argued that the market is waiting for a crisis ahead. The maximum of the third wave is a crisis itself. Therefore, the use of such an indicator allows for predicting a crisis and having time to take measures that can mitigate the consequences of the crisis. In addition, the wavelet transform provides a time-frequency representation of the signal, which allows you to obtain additional information that is not reflected in the time representation of the signal.

In figures 3–10 shows the results of wavelet entropy calculation for the gas, oil, and gasoline markets.

Analysis of the energy surface of the wavelet coefficients (figure 3) allows us to draw conclusions about the crisis situation in the gas market. On a small scale, there is a manifestation of disturbance. In wavelet analysis, small scales correspond to high frequencies.

Figure 4 shows the dynamics of wavelet entropy. We observe the formation of three waves in a neighborhood of 1750-2000 points, which is an indicator of the crisis. This crisis is the

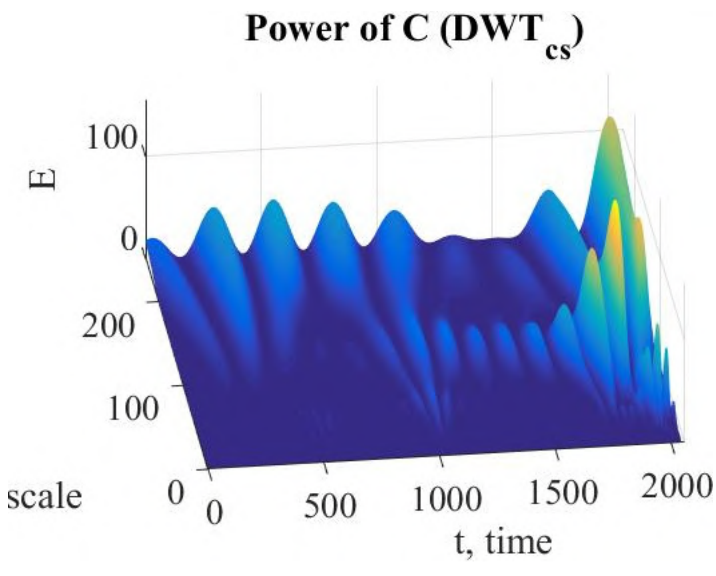


Figure 3: Wavelet coefficient energy for gas index.

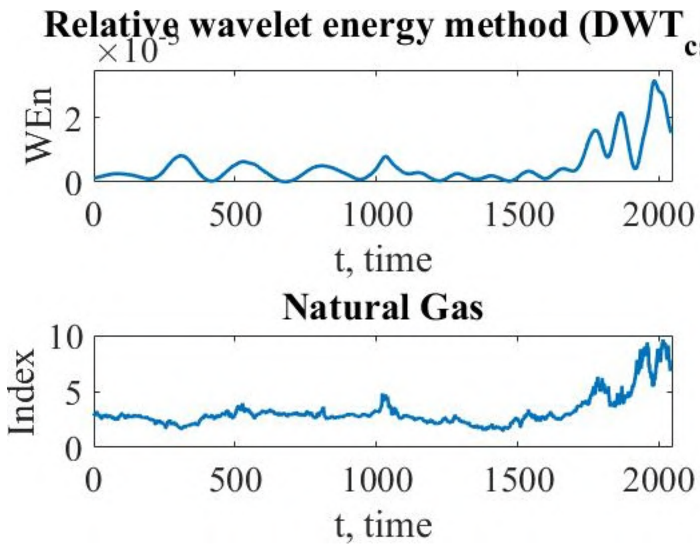


Figure 4: Wavelet entropy and dynamics of the gas index.

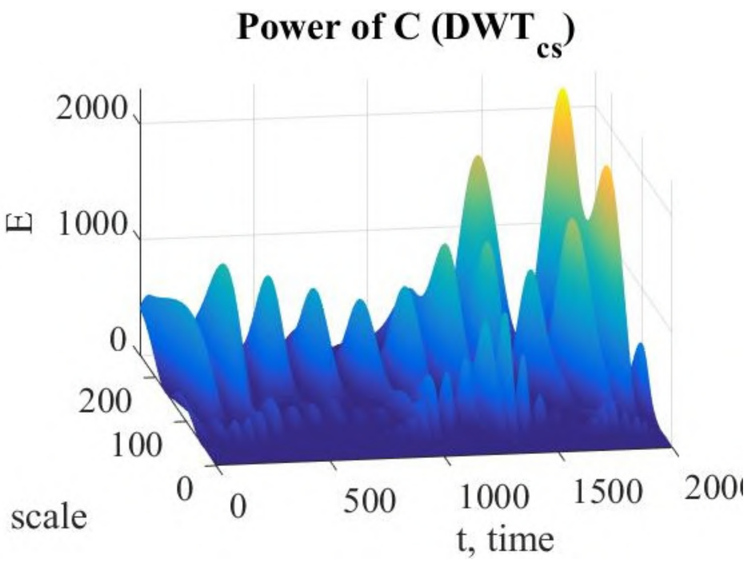


Figure 5: Wavelet coefficient energy for oil Brent index.

market's reaction to Russia's refusal to supply natural gas to Europe and the introduction of sanctions.

In figures 5, 6 shows the results of calculations for Brent oil, and figures 7, 8 – for WTI oil.

The energy of the wavelet coefficients shows a different situation for these two oil brands. This can be explained by the fact that Brent oil is traded on the markets of Europe and Asia, while WTI oil is traded on the US markets. But for the current time, the situation for these two brands of oil is similar. We see the formation of stable three waves, which indicates a

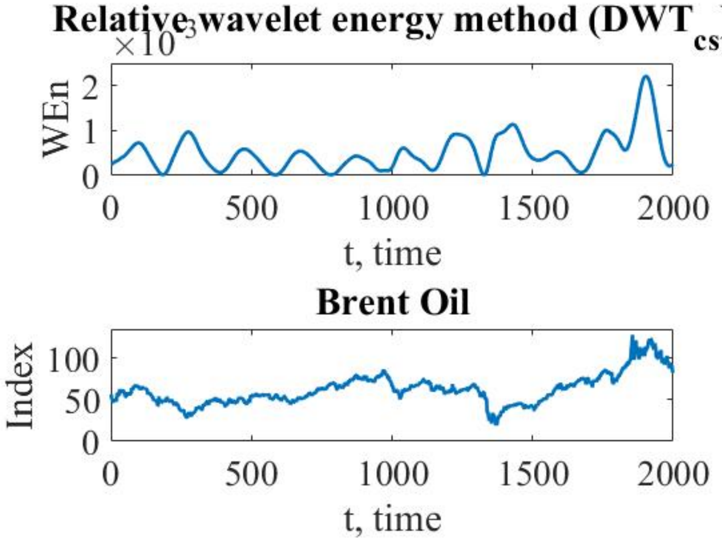


Figure 6: Wavelet entropy and dynamics of the oil Brent index.

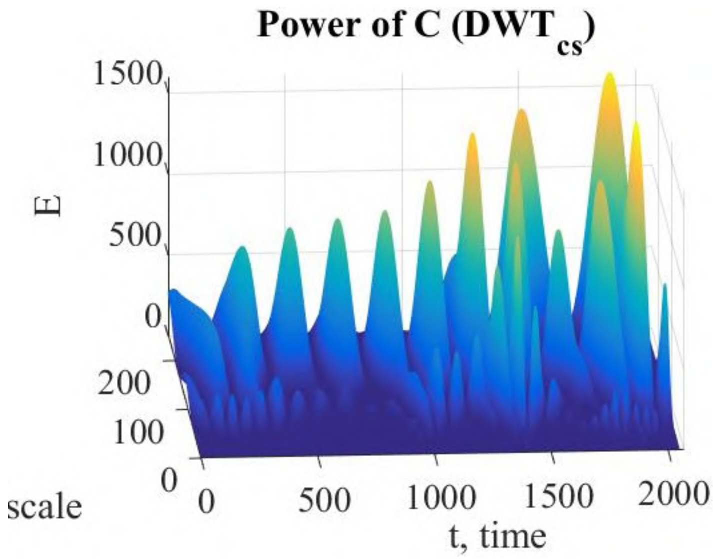


Figure 7: Wavelet coefficient energy for oil WTI index.

crisis. What is happening in the oil market? It can be seen that the price of Brent and WTI oil benchmarks continue to fall. In our opinion, this is related to the war in Ukraine and the risk of recession. The European Union in the eighth package of anti-Russian sanctions "included a ceiling" on oil prices. In addition, the EU plans to ban sea imports of crude and refined oil from Russia. In response to the EU sanctions, Russia decided to reduce oil production by 3 million barrels per day, arguing that this is a lever to increase oil prices on the market. For

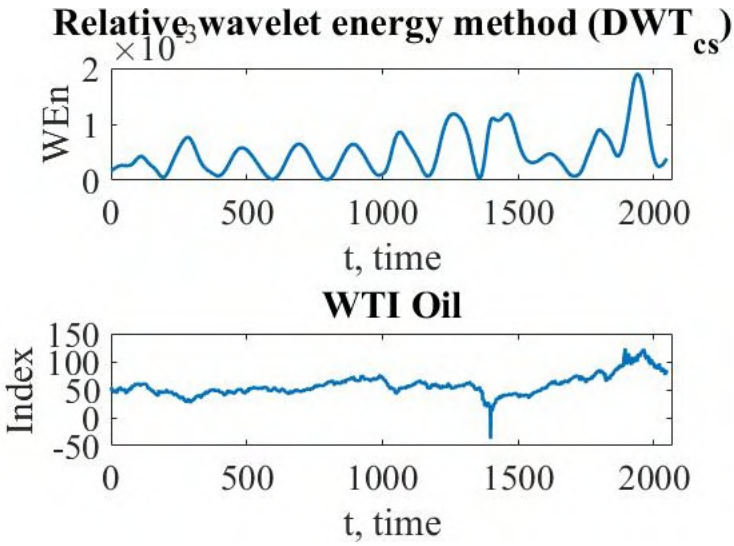


Figure 8: Wavelet entropy and dynamics of the oil WTI index.

Russia, the imposition of sanctions is a blow, as this is a budget-forming article (about 40% of budget revenues are in the form of taxes on hydrocarbon exports, and direct and indirect revenues related to this export make up to 60%). That is, the consequence of the introduction of sanctions will be a reduction in revenues from oil and gas. That is, it is precisely in this sector that Russia's "Achilles' heel" is, but the refusal of Saudi Arabia and other large Middle Eastern players to replace the Russian share of the oil market leads to fluctuations in its price, which in some way neutralizes the measures of the EU and the US countries regarding the oil embargo against Russia. They are trying to regulate the oil market. Thus, OPEC+'s decision is to reduce oil production by 2 million barrels per day, which should lead to an increase in oil prices. However, such a decision by OPEC+ has a reverse side. In particular, the United States began selling oil from reserves.

So, according to the results of the calculations, it can be stated that the oil and gas market is in a state of crisis, which was formed as a result of the war in Ukraine and the efforts of the main players to carry out its transformation, blocking Russia and reducing its influence on the world market. One such move by the global anti-Putin coalition (producing countries account for 60% of global GDP) is the declared creation of a buyers' cartel that has set a "price ceiling" for Russian oil and oil products. Even if India and China do not join the "price ceiling", the path of Russian oil to the world market will be difficult in December 2022, as the EU, Switzerland and Great Britain will not only ban their factories and traders from buying it, but will also introduce sanctions on insurance, financing and ship freight, which will lead to the need for Russia not only to look for new sales markets, but also to build alternative supply chains to the world market from scratch.

In figures 9, 10 shows calculations for the gasoline market. Gasoline is a derivative of oil. Therefore, the behavior of the gasoline market should be similar to the behavior of the oil market. If oil becomes cheaper, then the price of gasoline should also fall.

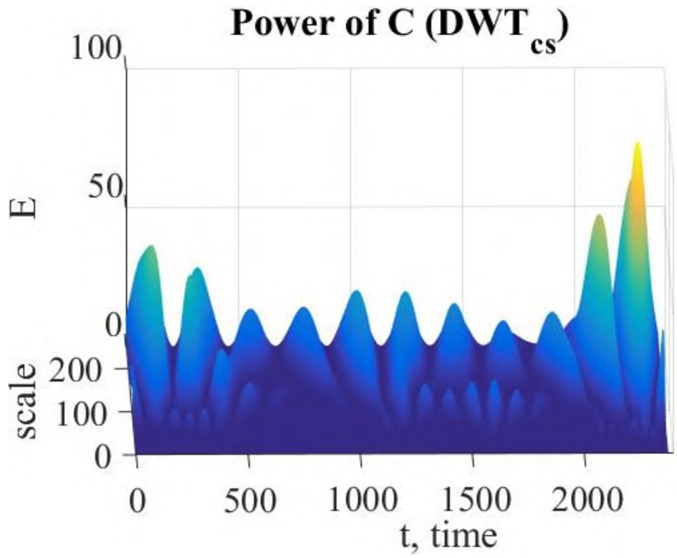


Figure 9: Wavelet coefficient energy for gasoline index.

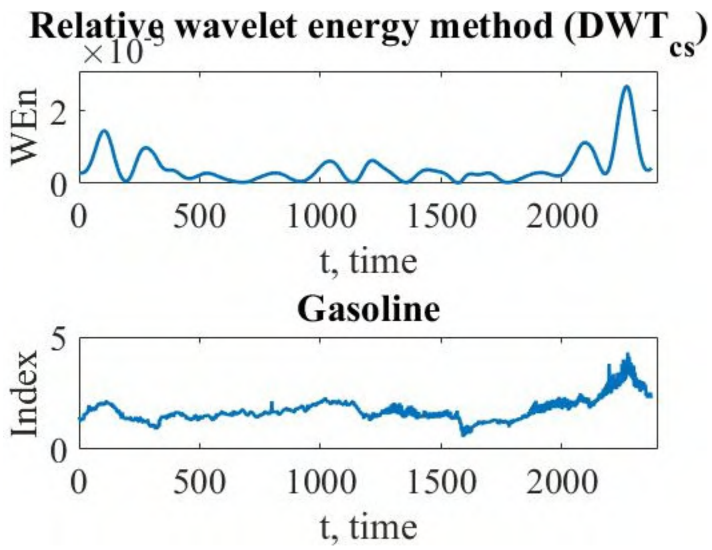


Figure 10: Wavelet entropy and dynamics of the gasoline index.

Comparing figure 9 from figure 5 and figure 7, we see that the energy surface for the gasoline market differs from the energy surfaces for oil. As you can see, the gasoline market is not stable. But starting from around the point of 1800, which corresponds to the year 2022 (figure 10), we observe the appearance of a triad of growing waves. And from this period, the behavior of the gasoline market becomes similar to the oil and gas market. And we state the crisis state of the market. What is the impact of the war in Ukraine? The world market of oil, oil products, and gas is being reformatted, and connections are changing. Ukrainian markets are also undergoing transformation, reorienting themselves towards the EU. It is obvious that the change of players in the market (both strong and not so) leads to instability, problematic issues of redistribution of resources.

The foreign exchange market is an important component of the financial market. Modeling and analysis of the currency market will allow an understanding of the economic and organizational relations between the participants.

In figure 11 shows the comparative dynamics of currency pairs EUR/USD and GBP/USD. These currency pairs are the most traded, which influenced the selection for the study.

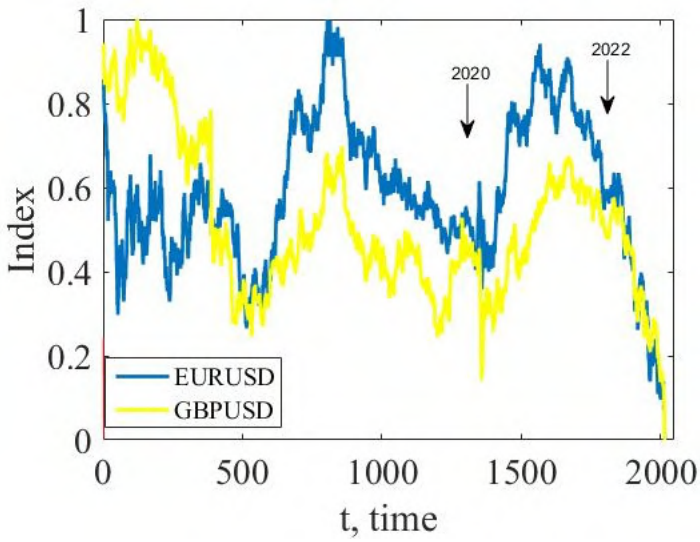


Figure 11: Comparative dynamics of indices of currency pairs EUR/USD and GBP/USD.

Figure 11 shows the sharp decline of currency pair indices in 2020. As for 2022, there is a drop in indices, but it is not of a rapid nature.

Applying the wavelet entropy method to the currency market allows you to get an answer to the question of the existence of a crisis in it. For both currency pairs, the formation of three waves, which is an indicator-precursor of the crisis phenomenon, was observed during 2015-2017 (within points 50-520, see figures 13, 15). The same situation is observed for the currency pair GBP/USD during the pandemic period (figure 15). The current situation for both currency pairs is marked by a gradual drop in the index values. The reasons for the subsidence may be the war in Ukraine, sanctions against Russia, the dependence of European states on Russian gas supplies, the political crisis in the EU regarding the support of sanctions and aid to Ukraine. The euro is the base currency, but it is also a tool for speculation.

Therefore, the simulation results indicate the absence of a crisis state at the time of the study. This market needs further monitoring, as the next wave is still in the process of formation.

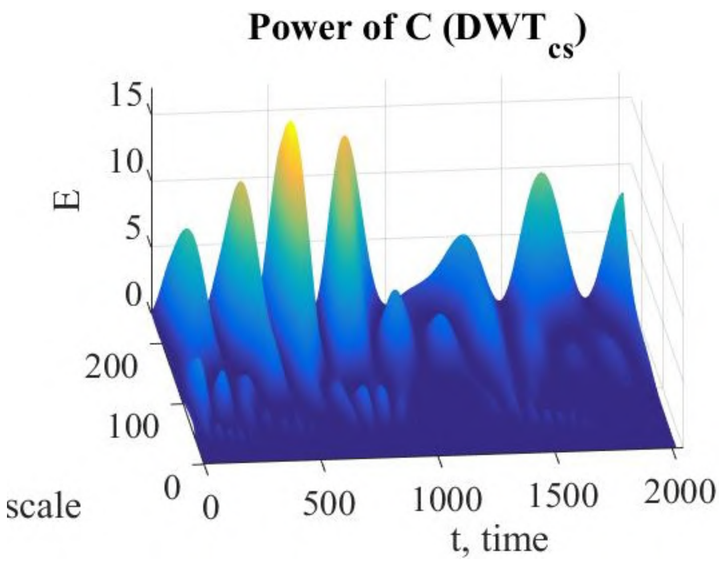


Figure 12: Wavelet coefficient energy for the currency pair EUR/USD.

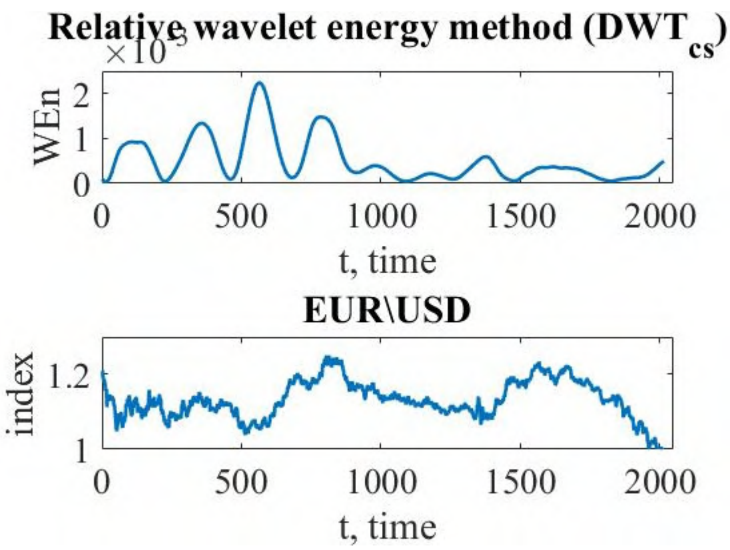


Figure 13: Wavelet entropy and dynamics of the currency pair EUR/USD.

4. Conclusion

In summation, the utilization of the wavelet entropy method for modeling and analyzing the oil, gas, oil products, and foreign exchange markets unveils the war in Ukraine as a potent force driving extant or emerging crisis phenomena within these domains. The wavelet entropy models distinctly underscore a crisis presence in the oil, gas, and gasoline markets. Concurrently, the primary currency pairs within the foreign exchange market exhibit a gradual yet protracted

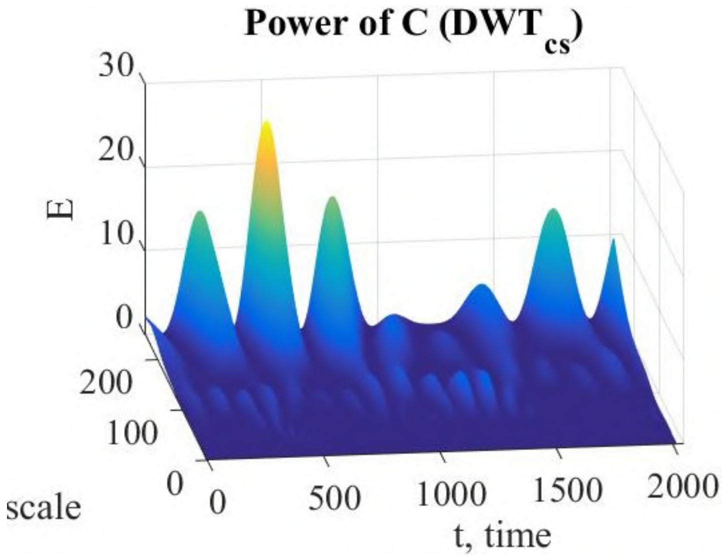


Figure 14: Wavelet coefficient energy for the currency pair GBP/USD.

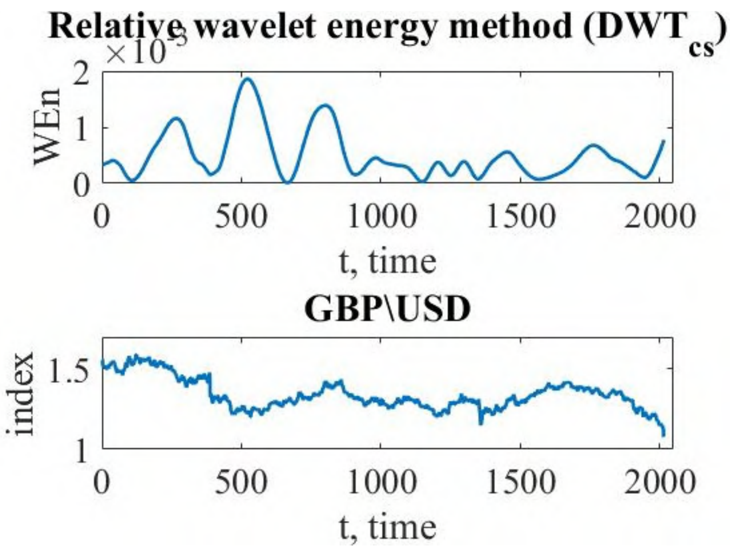


Figure 15: Wavelet entropy and dynamics of the currency pair GBP/USD.

decline. Notably, the intricacies of the currency market necessitate continuous vigilance, and the employment of the wavelet entropy method for its modeling holds promise in preemptively flagging crisis states.

These findings harmonize with existing conclusions, such as the heterogeneous impact of the oil market on diverse financial assets, peaking during the Ukraine conflict, and the greater susceptibility of globalized markets to its ramifications. The realm of globalization within the world economic landscape, while conferring advantages, is fraught with inherent perils. Today, these threats reverberate within the energy sectors, spanning oil, gas, and related commodities. The war ignited by Russia in Ukraine—a manifestation of its aspirations for supremacy, territorial aggrandizement, and fear of relinquishing its standing—compels the global community to reevaluate the architecture, interconnections, and dynamics of globalization within the world economic tapestry.

In essence, the research underscores that crises—whether triggered by political strife, military engagements, or global pandemics—resonate across financial markets and globalization's intricate fabric, demanding a holistic comprehension fostered by interdisciplinary methodologies. The wavelet entropy method, by detecting the foreboding ripples of crisis at early junctures, emerges as an indispensable tool for anticipating and navigating the complex intersections of financial markets, globalization processes, and geopolitical convulsions. The war in Ukraine, emblematic of contemporary geopolitical tensions, serves as a poignant illustration of the profound reverberations that crises can engender, catalyzing a transformative reevaluation of economic interconnectedness and collective security in a rapidly evolving world.

References

[1] H. Danylchuk, L. Kibalnyk, O. Kovtun, O. Pursky, Y. Kyryliuk, O. Kravchenko, Analysis and Modeling of Globalization Processes in the Period of Crisis: The Impact of Military Actions in Ukraine on World Financial Markets, in: S. Semerikov, V. Soloviev, A. Matviychuk, V. Kobets, L. Kibalnyk, H. Danylchuk, A. Kiv (Eds.), Proceedings of 10th International Conference on Monitoring, Modeling & Management of Emergent Economy - M3E2, INSTICC, SciTePress, 2023, pp. 176–184. doi:10.5220/0011932400003432.

[2] S. V. Vedernikova, Modern trends of globalization, Problemy systemnoho pidkhodu v ekonomitsi–Problems of a systemic approach in economics 5 (2017) 16–22.

[3] I. Kholoshyn, L. Burman, T. Nazarenko, S. Mantulenko, N. Panteleeva, Geographic particulars of the world's population food ration, E3S Web of Conferences 166 (2020) 13007. doi:10.1051/e3sconf/202016613007.

[4] V. N. Soloviev, A. Bielinskyi, V. Solovieva, Entropy Analysis of Crisis Phenomena for DJIA Index, in: V. Ermolayev, F. Mallet, V. Yakovyna, V. S. Kharchenko, V. Kobets, A. Kornilowicz, H. Kravtsov, M. S. Nikitchenko, S. Semerikov, A. Spivakovsky (Eds.), Proceedings of the 15th International Conference on ICT in Education, Research and Industrial Applications. Integration, Harmonization and Knowledge Transfer. Volume II: Workshops, Kherson, Ukraine, June 12-15, 2019, volume 2393 of *CEUR Workshop Proceedings*, CEUR-WS.org, 2019, pp. 434–449. URL: https://ceur-ws.org/Vol-2393/paper_375.pdf.

[5] S. S. Iyer, L. Gernal, R. Subramanian, A. Mehrotra, Impact of digital disruption influencing business continuity in UAE higher education, Educational Technology Quarterly 2023 (2023) 18–57. doi:10.55056/etq.29.

[6] I. Isik, O. Uygur, Financial crises, bank efficiency and survival: Theory, literature and emerging market evidence, International Review of Economics & Finance 76 (2021) 952–987.

[7] S. Campos-Martins, C. Amado, Financial market linkages and the sovereign debt crisis, Journal of International Money and Finance 123 (2022) 102596.

[8] C. Labidi, M. L. Rahman, A. Hedström, G. S. Uddin, S. Bekiros, Quantile dependence between developed and emerging stock markets aftermath of the global financial crisis, International review of financial analysis 59 (2018) 179–211.
[9] Y. Xu, N. Taylor, W. Lu, Illiquidity and volatility spillover effects in equity markets during and after the global financial crisis: An mem approach, International Review of Financial Analysis 56 (2018) 208–220.
[10] S. R. Bentes, On the stylized facts of precious metals' volatility: A comparative analysis of pre-and during covid-19 crisis, Physica A: Statistical Mechanics and its Applications 600 (2022) 127528.
[11] B.-Y. Liu, Y. Fan, Q. Ji, N. Hussain, High-dimensional covar network connectedness for measuring conditional financial contagion and risk spillovers from oil markets to the g20 stock system, Energy Economics 105 (2022) 105749.
[12] C.-L. Chang, M. McAleer, Y.-A. Wang, Herding behaviour in energy stock markets during the global financial crisis, sars, and ongoing covid-19, Renewable and Sustainable Energy Reviews 134 (2020) 110349.
[13] V. N. Soloviev, A. Belinskiy, Complex Systems Theory and Crashes of Cryptocurrency Market, in: V. Ermolayev, M. C. Suárez-Figueroa, V. Yakovyna, H. C. Mayr, M. Nikitchenko, A. Spivakovsky (Eds.), Information and Communication Technologies in Education, Research, and Industrial Applications, Springer International Publishing, Cham, 2019, pp. 276–297. doi:10.1007/978-3-030-13929-2_14.
[14] H. Danylchuk, L. Kibalnyk, O. Kovtun, A. Kiv, O. Pursky, G. Berezhna, Modelling of cryptocurrency market using fractal and entropy analysis in COVID-19, in: A. Kiv (Ed.), 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, volume 2713 of *CEUR Workshop Proceedings*, CEUR-WS.org, 2020, pp. 352–371. URL: http://ceur-ws.org/Vol-2713/paper40.pdf.
[15] A. O. Bielinskyi, O. A. Serdyuk, S. O. Semerikov, V. N. Soloviev, Econophysics of cryptocurrency crashes: a systematic review, in: A. E. Kiv, V. N. Soloviev, S. O. Semerikov (Eds.), Proceedings of the Selected and Revised Papers of 9th International Conference on Monitoring, Modeling & Management of Emergent Economy (M3E2-MLPEED 2021), Odessa, Ukraine, May 26-28, 2021, volume 3048 of *CEUR Workshop Proceedings*, CEUR-WS.org, 2021, pp. 31–133. URL: http://ceur-ws.org/Vol-3048/paper03.pdf.
[16] V. Derbentsev, A. Matviychuk, V. N. Soloviev, 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, Springer, Singapore, 2020, pp. 211–231. doi:10.1007/978-981-15-4498-9_12.
[17] H. Danylchuk, N. Chebanova, N. Reznik, Y. Vitkovskyi, Modeling of investment attractiveness of countries using entropy analysis of regional stock markets, Global J. Environ. Sci. Manage 5 (2019) 227–235.
[18] E. Kuzu, A. Süsay, C. Tanrıöven, A model study for calculation of the temperatures of major stock markets in the world with the quantum simulation and determination of the crisis periods, Physica A: Statistical Mechanics and its Applications 585 (2022) 126417.
[19] S.-Y. Choi, Evidence from a multiple and partial wavelet analysis on the impact of geopolitical concerns on stock markets in north-east asian countries, Finance Research Letters

46 (2022) 102465.
[20] H. Abdel-Latif, M. El-Gamal, Financial liquidity, geopolitics, and oil prices, Energy Economics 87 (2020) 104482.
[21] W. Boungou, A. Yatié, The impact of the ukraine–russia war on world stock market returns, Economics Letters 215 (2022) 110516.
[22] G.-D. Lo, I. Marcelin, T. Bassène, B. Sène, The russo-ukrainian war and financial markets: The role of dependence on russian commodities, Finance Research Letters 50 (2022) 103194.
[23] S. Boubaker, J. W. Goodell, D. K. Pandey, V. Kumari, Heterogeneous impacts of wars on global equity markets: Evidence from the invasion of ukraine, Finance Research Letters 48 (2022) 102934.
[24] G. Foster, Wavelets for period analysis of unevenly sampled time series, The Astronomical Journal 112 (1996) 1709–1729.
[25] S. Sello, Wavelet entropy and the multi-peaked structure of solar cycle maximum, New Astronomy 8 (2003) 105–117. doi:10.1016/S1384-1076(02)00192-6.
[26] L. Zunino, D. G. Pérez, M. Garavaglia, O. A. Rosso, Wavelet entropy of stochastic processes, Physica A: Statistical Mechanics and its Applications 379 (2007) 503–512.
[27] R. Q. Quiroga, O. A. Rosso, E. Başar, Wavelet entropy: a measure of order in evoked potentials, Electroencephalography and clinical neurophysiology. Supplement 49 (1999) 299–303.
[28] Yahoo Finance, 2022. URL: https://finance.yahoo.com/.
[29] V. Soloviev, A. Serdyuk, D. Chabanenko, Wavelet entropy as a critical phenomena precursor, in: Information Technologies, Management and Society, Information System Institute, Riga, Latvia, 2010, pp. 48–49.

Nonlinear dynamics of electric vehicle sales in China: a fractal analysis*

Serhii Kurkula, Nataliia Maksyshko, Dmytro Ocheretin and Serhii Cheverda

Zaporizhzhia National University, 66 Zhukovskogo Str., Zaporizhzhia, 69600, Ukraine

Abstract

Electric vehicles (EVs) are rapidly growing in the global automobile market, especially in China, which accounted for 45% of EV sales in 2020. However, forecasting the sales of EVs is challenging due to the complex and nonlinear nature of the market dynamics. In this paper, we apply three methods of nonlinear analysis to investigate the properties of the monthly sales volumes of the leading EV manufacturers in China from January 2016 to June 2022. The methods are: the Hurst normalized range method, phase analysis, and recurrence plots. We use the R software environment to perform the calculations and visualize the results. We find that the sales dynamics exhibit fractal features, trend stability, long-term memory, cyclicity, quasi-cycles, and determinism. These findings can inform the selection of relevant forecasting methods and their parameters for the EV market in China.

Keywords

electric vehicles, China, nonlinear dynamics, fractal analysis, phase analysis, recurrence plots, Hurst exponent

1. Introduction

Transportation is one of the major consumers of energy and a significant source of greenhouse gas emissions. To reduce the dependence on fossil fuels and mitigate the environmental impact, many developed countries have been promoting the adoption of electric vehicles (EVs) as a cleaner and more efficient alternative. EVs are vehicles that use electric motors powered by batteries or fuel cells, instead of internal combustion engines. EVs have been gaining popularity in the global automobile market, especially in China, which is the largest and fastest-growing EV market in the world.

The main drivers for the increasing demand for EVs can be classified into three categories. The first category is legislative factors, such as subsidies, discounts, free parking, free charging, and other incentives offered by governments to encourage EV purchases. The second category is environmental factors, such as the awareness of the negative effects of carbon dioxide emissions and the social responsibility of consumers to choose eco-friendly vehicles. The third category

M3E2-MLPEED 2022: The 10th International Conference on Monitoring, Modeling & Management of Emergent Economy, November 17-18, 2022, Virtual, Online

*Extended and revised version of paper [1] presented at the 10th International Conference on Monitoring, Modeling & Management of Emergent Economy.

sergeysergey0093@gmail.com (S. Kurkula); maxishko@ukr.net (N. Maksyshko); odvisua@gmail.com (D. Ocheretin); cheverdaserega@gmail.com (S. Cheverda)
0000-0003-0717-0291 (S. Kurkula); 0000-0002-0473-7195 (N. Maksyshko); 0000-0001-6705-6381 (D. Ocheretin); 0000-0003-2161-037X (S. Cheverda)

is energy security factors, such as the volatility of oil and gasoline prices and the vulnerability of supply chains. In contrast, electricity generation is more diversified and less dependent on external factors.

The competition in the EV market has stimulated the development of new technologies, enterprises, business models, and markets. The global EV market is still in its formative stage, with a large amount of investments in EV production and infrastructure. The decisions made during this period will shape the future architecture of the global market, from educational and production standards, urban infrastructure design, to new business models and market regulation conditions.

The EV market is an important and dynamic object of study, as it has significant implications for the global economy and the individual countries. According to Bloomberg rating agency estimates, EV sales will account for two-thirds of the global automobile market by 2040 [2]. Therefore, it is essential to understand the nature and dynamics of the EV market.

The global EV market is evolving, so it is necessary to determine the models for its evolution. Based on the statistical analysis of the EV market, it can be observed that China is the dominant player in EV sales and market penetration. In particular, in 2013, China achieved phenomenal growth in vehicle sales in the segment of battery electric vehicles (BEV) and plug-in hybrid vehicles (PHEV). For six consecutive years from 2012 to 2017, the annual growth rate of the market volume was at least 45 per cent [3]. And in 2020, according to the International Energy Agency [4], the Chinese market accounted for almost 45 per cent of global sales. Thus, the study of the development dynamics of the EV market in China is necessary as a basis for further research in the markets of other countries.

2. Related works

Zhang et al. [5] presents Singular Spectral Analysis (SSA) as a one-dimensional time series model and Vector Autoregressive Model (VAR) as a multivariate model that displays the sales volume of automobiles with electric and hybrid engines in China. Empirical calculation results show that SSA satisfactorily indicates the market trend. The VAR model, which contains exogenous parameters related to the market, according to the authors, can significantly improve the accuracy of the results when used to build forecasts.

The price of charging the automobile is important for owners during its operation. Zhang et al. [6] proposes a pricing model for public-private partnership projects of automobile charging infrastructure in China, which is based on the use of the system dynamics (SD) method. In paper [7], based on predictive data on the number of automobiles, a simulation of the spread of electric vehicles is presented using the example of France and Germany.

Articles [8, 9] are devoted to predicting the dynamics of the distribution of electric vehicles within the European Union. For this, logistic models are used, in particular, the logistic and Bass diffusion model [8], which is used in [9] to predict the number of cars used in Beijing.

An overview of the methods that are used to predict the penetration of electric vehicles into the passenger vehicle market is presented in [10]. Two groups of models are distinguished: econometric models with disaggregated data (such as discrete choice) and simulation models based on agents. Some methods have been found to have a stronger methodological basis, while others require complex datasets or can be more flexibly combined with other methods. Despite the absence of a dominant method, Jochem et al. [10] justify the advantage of hybrid approaches and managed data that take into account micro and macro aspects, which allows obtaining more accurate results.

In [11], using a logistic growth model, a long-term forecast of stocks of electric vehicles in 26 countries on five continents is provided. The findings show that in 2032, 30 per cent of the global vehicle fleet will be electric vehicles. However, the results obtained by the authors also demonstrate significant differences between countries, which may be due to differences in government support.

Electric vehicle sales are influenced by many factors (especially in China) and there are not many sales forecasting models available. In particular, Wan et al. [12] used decomposition and integration procedures based on the TEI@I methodology. So, in the forecasting model, principal component regression analysis (PCR) was used to work with a linear relationship. Then a BP neural network and a support vector machine (SVM) were used to work with non-linear dependence. In the last step, all models were integrated together. The Granger causality test and the degree of gray correlation are used to quantify the factors that affect EV sales through consumer network data analysis. On the example of two automobile models, it was found that the PCR-BP models and the PCR-SVM models have better predictive performance than one model. According to the authors, this approach is more suitable for making decisions about forecasting markets for similar products.

Dingab and Li [13] proposes to use the modified gray model as a promising tool for predicting sales of electric vehicles.

The use of different approaches to forecasting the sales of electric vehicles indicates that the quality of the results is not satisfactory. A common feature of almost all almost all forecasting methods that are presented in the review is that they provide for the subordination of volume dynamics to a linear paradigm. However, today it is a recognized fact that the dynamics of most markets does not obey the law of normal distribution, and therefore their modeling by traditional methods leads to significantly unsatisfactory results. The linear paradigm has been replaced by a nonlinear paradigm [14, 15], which is based on the recognition of the fractal nature of the market and is actively developed for analysis and modeling [16, 17, 18, 19]. This statement is based on such features of time series (TS) of indicators characterizing financial markets: the lack of independence of levels, the presence of long-term memory, and others [20, 21, 22, 23]. The use of statistical methods for their research and further forecasting (as the ultimate goal of the analysis) turns out to be inadequate. Therefore, there is a need to use new, different from statistical, methods of analysis.

The purpose of this research is to diagnose the nature and properties of the dynamics of sales of electric vehicles in the Chinese market using non-linear analysis tools for further use in choosing a relevant forecasting method.

3. Materials

The object of analysis of this research is the sales volumes of cars, which are contained in the reports of the China Association of Automobile Manufacturers [24] and published by the online publication "Chinese Cars" [25].

An analysis of the structure of the electric vehicle market in China revealed that in the period from January 2016 to June 2022, 37.5 per cent of the electric vehicle market belongs to five automakers, namely: BYD, Mercedes-Benz, Roewe, Geely, Chery. Most of these companies are representatives of the Chinese automotive industry, which is due, in particular, to state support for manufacturers of this type of transport [26]. Let's characterize these companies in more detail.

BYD is the only automobile manufacturer that has mastered batteries, electric motors, and vehicle control technologies. BYD was founded in 1995 as a pioneer in the battery technology industry. Its stated goal is to change the world by creating a complete zero-emission ecosystem that runs on clean energy and reduces dependence on oil. BYD's innovative products are leaders in many sectors, including battery electric vehicles, buses, medium and heavy duty trucks and forklifts. In 2003, the company entered the automotive business, and in 2005, the first BYD brand automobile went on sale [27]. The company holds 16 per cent of the electric vehicle market in China.

Mercedes-Benz is a world-famous automaker that in recent years has been investing more resources in its advanced research and design capabilities in China as the new center of gravity for the auto industry [28]. The company holds 9 per cent of the electric vehicle market in China.

Roewe is owned by the Shanghai Automotive Industry Corporation (SAIC) and is one of the few Chinese luxury brands that actually manufacture modernized copies of older Rover models [29]. The company holds 6 per cent of the electric vehicle market in China.

Geely Auto Group is a leading automobile manufacturer that was founded in 1997 as a subsidiary of Zhejiang Geely Holding Group. For the past five years, the company has maintained its position as the best-selling Chinese brand [30]. The company holds 4 per cent of the electric vehicle market in China.

Chery was founded in 1997 under the patronage of state-owned companies and holdings, as well as smaller investors. In 2006, Ukraine was one of the first countries to introduce the assembly of automobiles of this brand outside China. In 2012, in pursuit of a globalization strategy, Chery and Jaguar Land Rover Motors jointly invested in the establishment of Chery Jaguar Land Rover Motors Co., Ltd., which is China's first Sino-British automobile joint venture [31]. The company holds 3 per cent of the electric vehicle market in China.

Thus, we will analyze the nature of the dynamics of the behavior of agents of the electric car market in China on the basis of time series (TS) of monthly sales volumes of automobile companies (manufacturers) BYD, Chery, Geely, Mercedes-Benz and Roewe. These automakers were selected based on the fact that they are among the top 9 most popular electric mobile brands in terms of sales for the period from January 2016 to June 2022 [25] and have sufficient data for analysis for this period. When analyzing the dynamics, we will identify the sales volumes of electric vehicles with the volume of demand for them.

4. Methodology

To identify the nonlinear (chaotic) behavior of economic data, various methods of time series analysis are used [32]. In particular, tests for deterministic chaos have been developed for this purpose, which allow one to study the main features of chaotic phenomena: nonlinearity, a fractal attractor, and sensitivity to initial conditions.

In this research, to diagnose the nature and properties of the dynamics of sales of electric vehicles in the Chinese market, we will use three tools for analyzing nonlinear dynamics, namely: traditional R/S-analysis – the Hurst normalized range method, phase analysis and recurrence analysis.

For the purpose of a general assessment of the fractal properties of time series, we use the Hurst normalized range algorithm for analysis [14]. It is known that if the system gives the Hurst statistics for a sufficiently long period, then this indicates the result of interrelated events. As is known, a measure of the mutual connection of events is the correlation coefficient. The influence of the present on the future can be represented by the following correlation:

$$C = 2^{2H-1} - 1, \tag{1}$$

where C – measure of correlation,

H – Hurst exponent.

The range of the Hurst exponent (H) is the interval [0; 1]. The indicator value allows classifying all time series into three groups:

1) $H = 0{,}5$;
2) $0 \le H < 0{,}5$;
3) $0{,}5 < H \le 1$.

The value $H = 0{,}5$ indicates a random time series: the events are random and not correlated ($C = 0$ according to (1)). The present does not affect the future.

If $H \in (0{,}5; 1]$, then the considered time series is persistent or trend-resistant and is characterized by the effect of long-term memory. Events are the more correlated, the closer the value is to 1 (correspondingly, C also approaches 1 or 100 per cent correlation according to (1)).

The value $H \in [0; 0{,}5)$ corresponds to antipersistent or ergodic time series. In a loose definition, antipersistence means reverting to the mean or, in other terminology, reversing (alternating positive and negative increments) more often than in a random process. Thus, the Hurst exponent (H) is decisive in diagnosing the nature of the development of a system or process.

To check the validity of the results on the presence of long-term memory based on the value of the Hurst exponent (H), we will use a test for random mixing of the levels of the time series.

Phase analysis is one of the effective methods for obtaining information about the nature of the dynamics of the system under consideration [16]. To the time series ($X = (x(t), t = \overline{1, n})$) that characterizes the dynamics of demand in the market of electric vehicles, we will apply such a presentation method, which can be used to return from the observed state of the system to its previous state. This "return" is implemented by the method of time delays and is produced by constructing a phase trajectory (phase portrait) of dimension ρ:

$$\Phi_\rho(X) = \left\{(x(t), x(t+1), ..., x(t+\rho-1)), t = \overline{1, n}\right\}, \tag{2}$$

which is a set of points called "ρ-history". For any time series, the list of all its M-histories determines the corresponding set of points in the pseudo-phase (or lag) space. In this case, when using the terms "phase portrait" or "phase trajectory" it means that the neighboring points of the set (2) are connected by segments of a straight or curved line for clarity.

Thus, the graphic representation of the system on the phase plane (or in the phase space), along the coordinate axes of which the values of the variables of the system (TS levels) are plotted, is called the phase portrait of the system. The behavior of phase points in time, which is described by the phase trajectory and the set of such phase trajectories for any initial conditions form a phase portrait. A phase portrait is a mathematical method for representing the behavior of a system and a geometric representation of individual movements, and also displays the state of equilibrium, periodic and chaotic movement of a phase point, the logic of the system's behavior and its dependence on external and internal influences.

Objective information about the nature of the behavior of a dynamic process can be obtained by observing the time series X, based on the Takens theorem [33]: if the system generating the time series is m-dimensional and inequality $\rho \geq 2m + 1$ is satisfied, then in the general case, phase trajectories reflect the dynamics of the system under study. There is a dipheomorphism between the phase trajectories and the true data generated by the system. This result allows one to draw conclusions about the behavior of the system based on observational data, and, moreover, to obtain information to predict this behavior.

Analysis of the phase portrait makes it possible to determine the type and characteristic features of the dynamics of a particular system. To deepen such an analysis, Eckmann et al. [34] proposed in 1987 a new diagnostic tool, the recurrence plot.

The recurrence plot is a projection of the ρ-dimensional pseudo-phase space onto the surface. Let point x_i-correspond to the point of the phase trajectory (2), which describes a dynamical system in m-dimensional space at times $t = i$, for $i = 1, ..., n$. Then the recurrence plot is an array of points, where non-zero elements with coordinates (i, j) correspond to the case when the distance between x_i and x_j is less then γ:

$$\begin{aligned} &RP_{i,j} = \theta\left(\gamma - \|x_i - x_j\|\right), \\ &x_i,\ x_j\ \in R^m,\ i, j = 1, ..., n, \end{aligned} \tag{3}$$

where γ – size of the point x_i,

$\|x_i - x_j\|$ – distance between points,

$\theta(\cdot)$ – Heaviside function.

For the practical reconstruction of the attractor for a given time series, it is necessary to determine the values of the parameters: ρ – the embedding dimension of the time series, d – the time lag of the time series [35].

To determine the time lag of the time series, the function (S) – the adjusted mutual information function (AMI) was used for the time series under research, which takes into account non-linear correlations [36]:

$$S = -\sum_{ij} p_{ij}(\Phi_\rho(X)) \cdot \ln \frac{p_{ij}(\Phi_\rho(X))}{p_i p_j}, \tag{4}$$

where $p_{ij}(\Phi_\rho(X))$ – joint probability that an observation falls into the i-th interval and the observation time d later falls into the j-th;

p_i – the probability to find a time series value in the i-th interval;

p_j – the probability to find a time series value in the j-th interval.

To calculate the optimal time lag of the time series (d), we will use the tseriesChaos library of the R environment.

To determine the embedding dimension of the time series, the false nearest neighbor method given in [37] was used. This method is based on the assumption that at the next iterations the neighboring points of the phase trajectory remain sufficiently close. But if the nearest points move away from one another, then they are called false nearest neighbors. The task of the method is to choose such a dimension of the time series (ρ), in which the proportion of points that have false neighbors is minimized.

Based on the calculated parameters of the embedding dimension and time lag, recurrence diagrams of time series are built. The analysis of the statistical characteristics of the recurrence diagram makes it possible to determine the measures of complexity of the structure of the recurrence diagrams [38]:

- percent recurrence (%REC),
- percent determinism (%DET),
- average (ADL) and maximum (MDL) diagonal lines lengths of the recurrence diagram.

The construction and determination of the statistical characteristics of recurrence diagrams will be implemented in the R environment using the tseriesChaos and nonlinearTseries libraries.

Based on the analysis of the statistical characteristics of the recurrence diagram, it is possible to determine the presence of homogeneous processes with independent random values; processes with slowly changing parameters; periodic and oscillating processes that correspond to nonlinear systems. Thus, the analysis of the recurrence surface makes it possible to evaluate the characteristics of a non-linear object on relatively short time series, which makes it possible to make prompt decisions regarding the control of the object.

5. Results

The analysis of the behavior of Chinese electric automobiles market agents was carried out on the basis of monthly sales data from January 2016 to June 2022 of five automobile companies (BYD, Chery, Geely, Mercedes-Benz, Roewe) (figure 1).

Time series of sales of electric vehicles in the Chinese market denoted by $X_k = (x(t),\ t = \overline{1,n}),\ k = \overline{1,5}$ where n is the length of the time series, k is the index assigned to the corresponding manufacturer (in order of priority): BYD, Chery, Geely, Mercedes-Benz, Roewe.

Table 1 shows the results of the Hurst exponent calculations (H) for these time series and the value of the Hurst exponent (H_{mixing}) obtained after applying the mixing test.

According to table 1, we can conclude that all time series of sales volumes (demand for electric automobiles) of all manufacturers have signs of persistence, that is, they have a long-term memory. This is evidenced by the following:

a) the value of the Hurst exponents for all time series are in the interval $H \in [0,817;\ 0,873]$, which corresponds to the area of black noise;
b) the results of the mixing test ($H_{mixing} \in [0,546; 0,597]$) confirm the significance of the time series structure: its violations lead to the complete destruction of the trace of long-term memory.

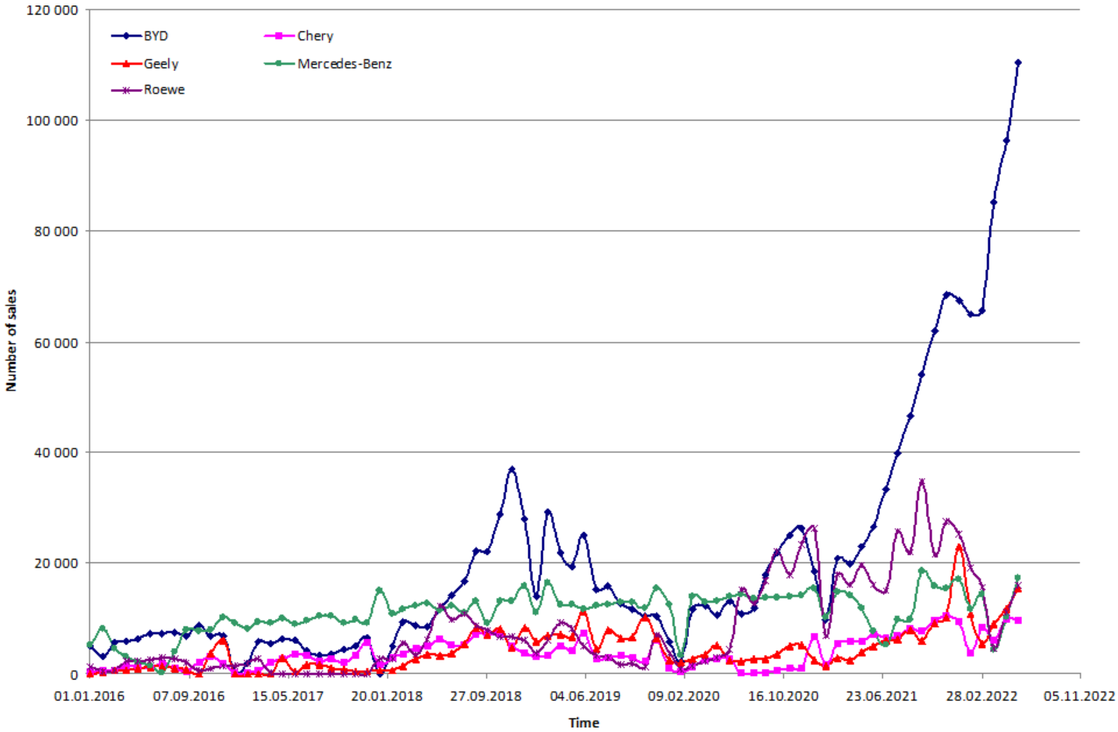


Figure 1: Number of sales of electric vehicles in the Chinese market from January 2016 to June 2022.

Table 1
The value of the Hurst exponent for the series of dynamics of sales volumes of electric automobiles of manufacturing companies for the period from January 2016 to June 2022.

Manufacturer (TS)	H	H_{mixing}
BYD (X_1)	0,84655	0,56659
Chery (X_2)	0,82696	0,58156
Geely (X_3)	0,81668	0,57214
Mercedes-Benz (X_4)	0,86762	0,54563
Roewe (X_5)	0,87330	0,59666

The presence of significant Hurst statistics for the time series of sales of electric vehicles is explained by the following reasoning.

The change in the volume of demand for electric vehicles is based on an increase in the overall demand for vehicles, the perception of buyers of a certain expediency to follow the trend in energy security (increased charging stations), legislative incentives and social responsibility (concern for the environment). The demand for electric vehicles is partly determined by fundamental information such as the state of the energy market, public discussion of environmental issues, current economic circumstances, expectations, and so on. This information is often useful in making decisions when purchasing a type of vehicle. Of great importance in this belongs to the marketing activities of manufacturing companies, the volume and quality of their offers on the market. Another important component of demand volumes is the extent to which buyers are able to pay for a new and usually more expensive product (an electric car). This "sensory component" is also analyzed, and as a result, a certain range of demand volume is formed around the existing one. This combination of information and thoughts results in displacement of volumes. If buyers see that the trend is in line with their positive expectations for a particular electric vehicle, they start buying like others. Yesterday's activity has an impact on today – the market remains mindful of yesterday's trend. The bias will change when demand reaches the upper limit of some actual value. At this point, the offset will change. The interesting thing is that the "range" of demand does not remain constant, but changes. New information regarding a particular electric vehicle (innovations and shortcomings) or the market as a whole can change this range and cause a sharp increase in sales volumes of the manufacturer (in particular, the introduction of breakthrough innovations) or a negative turn in the market situation, or for an individual seller (in particular, in case of deficiencies, and so on).

Let's proceed to the consideration of the results of the phase analysis of time series X_k, $k = \overline{1,5}$ of sales of electric vehicles in the Chinese market. Figure 2 shows phase portraits in a two-dimensional pseudo-phase (lag) space $\Phi_2(X_k) = \{(x(t),\ x(t+1))\}$, $k = \overline{1,5}$.

A more detailed analysis of phase portraits makes it possible to identify the following individual features.

In the dynamics of sales of the automobile company BYD (figure 2a)), at the beginning of the observation period for the first 5 years (from January 2016 to February 2021), almost stable quasi-cycles of length 7 were observed, which indicates the presence of long-term memory in them (confirmed by the value $H \approx 0{,}85$). However, since February 2021, the dynamics has changed dramatically in the direction of increasing sales volumes and almost no cyclicity when moving along the bisector of the coordinate angle. This indicates an increase in the memory depth of the time series.

The dynamics of sales of automobile companies Chery and Gelly (figure 2b), c)) are characterized by shorter quasi-cycles (length 4 or 5), and there is an increase in the amplitude of these quasi-cycles in the final interval of the time series (from February 2021 to June 2022), but no significant movement along the bisector of the coordinate angle is observed. The dynamics is characterized by less trend resistance, which is confirmed by the values $H \approx 0{,}83$) and $H \approx 0{,}82$) for the respective manufacturers.

The dynamics of sales of automobile companies Mercedes-Benz and Roewe (figure 2d), e)) is characterized by the presence of the longest quasi-cycles (length 9), their slow movement along the bisector of the coordinate angle (increase in volumes) and an increase in amplitude. This is evidence that the dynamics of sales volumes of these manufacturers is characterized by the greatest trend resistance (confirmed by the value of the Hurst exponent $H \approx 0{,}87$) for both companies).

Thus, the analysis of phase portraits $\Phi_2(X_k)$ in a two-dimensional pseudo-phase (lag) space makes it possible to identify the characteristic features of the dynamics of sales volumes of each agent in the Chinese electric car market.

At the first stage, using the tseriesChaos library of the R environment, the values of the embedding dimension (ρ) and the time lag (d) of the considered time series were calculated (table 2).

At the second stage, using the tseriesChaos and nonlinearTseries libraries in the R environ-

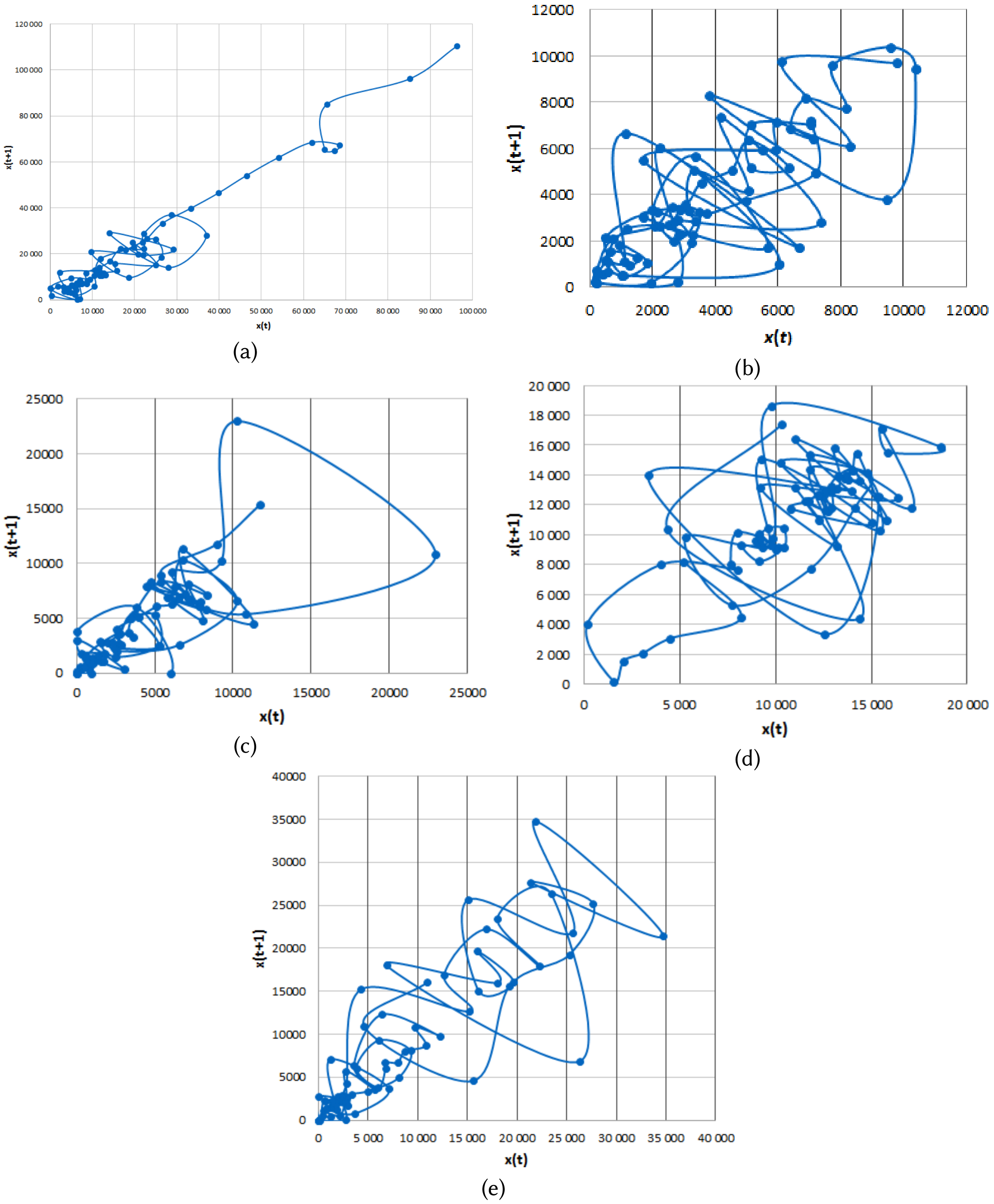


Figure 2: Phase portraits in a two-dimensional pseudo-phase space $\Phi_2(X_k) = \{(x(t),\ x(t+1))\}$, $k = \overline{1,5}$ for time series X_k, $k = \overline{1,5}$ from January 2016 to June 2022: a) BYD, b) Chery, c) Gelly, d) Mercedes-Benz, e) Roewe.

Table 2
The value of the embedding dimension (ρ) and time lag (d) for the series of dynamics of sales volumes of electric automobiles of manufacturing companies for the period from January 2016 to June 2022.

Manufacturer (TS)	The embedding dimension (ρ)	The time lag (d)
BYD (X_1)	5	9
Chery (X_2)	4	3
Geely (X_3)	4	3
Mercedes-Benz (X_4)	4	1
Roewe (X_5)	6	2

Table 3
Statistical characteristics of recurrence plots of electric automobiles sales in China from January 2016 to June 2022.

Manufacturer (TS)	%REC	%DET	ADL	MDL
BYD (X_1)	2,381	100	0	42
Chery (X_2)	1,429	100	0	70
Geely (X_3)	1,429	100	0	70
Mercedes-Benz (X_4)	1,333	100	0	75
Roewe (X_5)	1,471	100	0	68

ment, recurrence plots were constructed (figure 3a)-f)) and their statistical characteristics were determined (table 3).

The topology of the recurrence plots for electric automobiles sales in China shows abrupt changes in the dynamics of the system that generates the time series and causes white areas or bands to appear. On the recurrence plots, there is a gradual change in the parameters of the behavior of the agents of the automobile market, and there is also a drift of the attractor (white lower and upper corners of the diagram, crosses). The absence of short diagonal stripes on the recurrence plots indicates the absence of a stochastic process and the non-return of the trajectory to the same region of the phase space in different time periods.

The determinism of the behavior of companies in the automobile market is confirmed by the calculated statistical characteristics of recurrence plots, which are shown in table 3.

The value of the %REC indicator for all time series falls within the interval from 1% to 5%, which indicates the regular behavior of the time series.

The measure of determinism (%DET) of the recurrence plot characterizes the level of system predictability. Diagonal structures show the time during which a segment of the trajectory comes very close to another segment of the trajectory. For all five recurrence plots, the level of predictability is 100%. Note that this measure does not determine the real determinism of the process.

The average diagonal lines lengths (ADL) characterizes the average time during which two sections of the trajectory pass close to each other, and can be considered as the average predictability time of the system. An interesting fact is that, according to the calculation results, the smallest average predictability time of time series is 0.

The maximum diagonal lines lengths (MDL) characterizes the length of the trend. The shortest trend is in the BYD time series (42 points), and the longest is in Mercedes-Benz (75

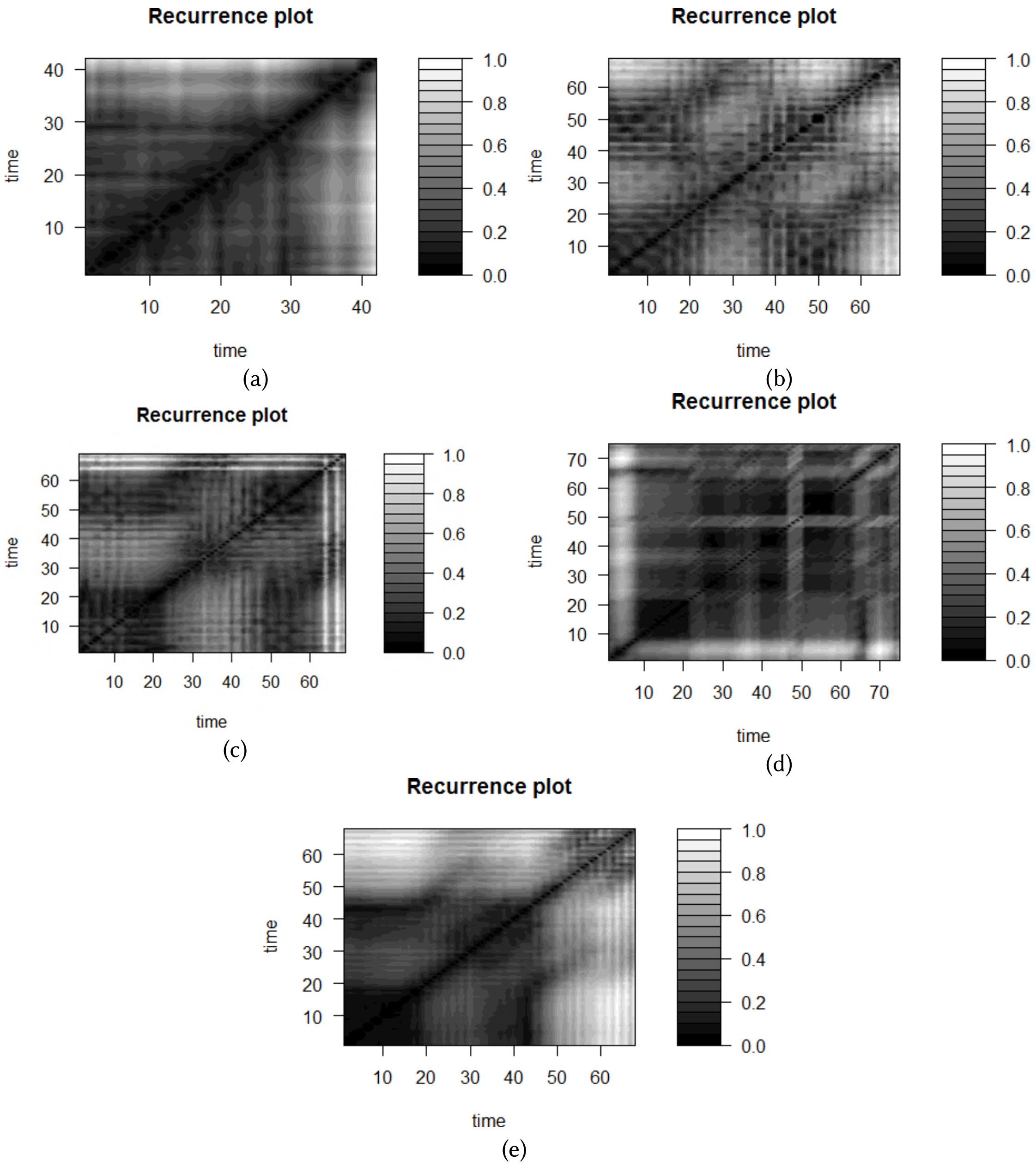


Figure 3: Recurrence plots of electric automobiles sales in China from January 2016 to June 2022: a) BYD, b) Chery, c) Gelly, d) Mercedes-Benz, e) Roewe.

points).

6. Conclusion

This paper presents a nonlinear analysis of the sales dynamics of electric automobiles in the Chinese market, which is the largest and fastest-growing EV market in the world.

The data used for the analysis are the monthly sales volumes of five EV manufacturers in China: BYD, Chery, Geely, Mercedes-Benz and Roewe, from January 2016 to June 2022.

The paper employs three methods of nonlinear dynamics: the traditional R/S-analysis, the phase analysis, and the recurrence plots.

The R/S-analysis reveals the trend stability and the long-term memory of the sales time series, indicating their nonlinear (fractal) nature. This implies that the classical forecasting methods are not suitable and may lead to poor results. The forecasting methods and their parameters should consider the long-term memory and its features.

The fractal analysis based on the R/S-analysis, however, only provides qualitative insights into the properties of the EV market and the trend stability of each time series. The quantitative characteristics obtained by this method are averaged over the entire series. Therefore, to obtain more differentiated characteristics of the memory, it is promising to apply fractal analysis methods based on the sequential R/S analysis algorithm [16].

The phase analysis in a two-dimensional phase space allows to identify the cyclicity and the attractors (quasi-cycles) of the sales dynamics for each EV manufacturer. The results provide a basis for further research on the features of the dynamics by decomposing the phase portrait into quasicycles, determining their characteristics, and analyzing the dynamics of their sizes and centers.

The recurrence plots in ρ-dimensional phase space and their topological analysis confirm the attractor drift for all EV manufacturers. A gradual change in the behavior parameters of each manufacturer is also detected.

The quantitative analysis of recurrence plots based on the complexity measures of their structure (such as %REC and %DET) confirms the fractal (deterministic) nature of the sales dynamics of EVs in China. It should be noted that the data used for this study are short time series, which may affect the possibilities, features, and results of applying these methods. However, their application – separately or in combination – enables to gain new knowledge about the characteristics of the dynamics in a new and rapidly developing market with global implications – the EV market.

The results of this study can be used to select relevant forecasting methods and their parameters for the EV market in China.

References

[1] S. Kurkula, N. Maksyshko, D. Ocheretin, S. Cheverda, Nonlinear Analysis of the Dynamics of Sales of Electric Automobiles in the Chinese Market, in: S. Semerikov, V. Soloviev, A. Matviychuk, V. Kobets, L. Kibalnyk, H. Danylchuk, A. Kiv (Eds.), Proceedings of 10th International Conference on Monitoring, Modeling & Management of Emergent Economy - M3E2, INSTICC, SciTePress, 2023, pp. 185–195. doi:10.5220/0011932500003432.

[2] C. McKerracher, S. Wagner, At Least Two-Thirds of Global Car Sales Will Be Electric by 2040, 2021. URL: https://cutt.ly/vVTyg5R.
[3] L. Jin, H. He, Comparison of the electric car market in China and the United States, Working paper 2019-10, International Council on Clean Transportation, 2019. URL: https://theicct.org/wp-content/uploads/2021/06/ICCT_US-China_EV-mkt-comp_20190523.pdf.
[4] L. Paoli, A. Dasgupta, S. McBain, Electric Vehicles, Tracking report, International Energy Agency, 2022. URL: https://www.iea.org/reports/electric-vehicles.
[5] Y. Zhang, M. Zhong, N. Geng, Y. Jiang, Forecasting electric vehicles sales with univariate and multivariate time series models: The case of China, PLOS ONE 12 (2017) e0176729. doi:10.1371/journal.pone.0176729.
[6] L. Zhang, Z. Zhao, H. Xin, J. Chai, G. Wang, Charge pricing model for electric vehicle charging infrastructure public-private partnership projects in China: A system dynamics analysis, Journal of Cleaner Production 199 (2018) 321–333. doi:10.1016/j.jclepro.2018.07.169.
[7] T. Dhakal, K.-S. Min, Macro Study of Global Electric Vehicle Expansion, Foresight and STI Governance 15 (2021) 67–73. doi:10.17323/2500-2597.2021.1.67.73.
[8] Z. Zhu, H. Du, Forecasting the Number of Electric Vehicles: A Case of Beijing, IOP Conference Series: Earth and Environmental Science 170 (2018) 042037. doi:10.1088/1755-1315/170/4/042037.
[9] A. Ensslen, C. Will, P. Jochem, Simulating Electric Vehicle Diffusion and Charging Activities in France and Germany, World Electr. Veh. J. 10 (2019) 73. doi:10.3390/wevj10040073.
[10] P. Jochem, J. J., G. Vilchez, A. Ensslen, J. Schäuble, W. Fichtner, Methods for forecasting the market penetration of electric drivetrains in the passenger car market, Transport Reviews 38 (2018) 322–348. doi:10.1080/01441647.2017.1326538.
[11] N. Rietmann, B. Hügler, T. Lieven, Forecasting the trajectory of electric vehicle sales and the consequences for worldwide CO2 emissions, Journal of Cleaner Production 261 (2020) 121038. doi:10.1016/j.jclepro.2020.121038.
[12] J. P. Wan, L. Q. Xie, X. F. Hu, Study on the electric vehicle sales forecast with TEI@I methodology, International Journal of Knowledge Engineering and Data Mining 7 (2021) 1–38. doi:10.1504/IJKEDM.2021.119836.
[13] S. Dingab, R. Li, Forecasting the sales and stock of electric vehicles using a novel self-adaptive optimized grey model, Engineering Applications of Artificial Intelligence 100 (2021) 104148. doi:10.1016/j.engappai.2020.104148.
[14] E. E. Peters, Fractal Market Analysis: Applying Chaos Theory to Investment and Analysis, John Wiley & Sons, Inc., 1994.
[15] V. N. Soloviev, N. Moiseienko, O. Tarasova, Modeling of Cognitive Process Using Complexity Theory Methods, in: V. Ermolayev, F. Mallet, V. Yakovyna, V. S. Kharchenko, V. Kobets, A. Kornilowicz, H. Kravtsov, M. S. Nikitchenko, S. Semerikov, A. Spivakovsky (Eds.), Proceedings of the 15th International Conference on ICT in Education, Research and Industrial Applications. Integration, Harmonization and Knowledge Transfer. Volume II: Workshops, Kherson, Ukraine, June 12-15, 2019, volume 2393 of *CEUR Workshop Proceedings*, CEUR-WS.org, 2019, pp. 905–918. URL: https://ceur-ws.org/Vol-2393/paper_356.pdf.
[16] V. Perepelitsa, N. Maksyshko, Analiz i prognozirovanie evolyuczii ekonomicheskikh sistem: problemy strukturirovaniya dannykh v usloviyakh neopredelennosti i predprognoznogo

analiza [Analysis and forecasting of the economic systems evolution: problems of data structuring in conditions of uncertainty and pre-forecast analysis], Lambert Academic Publishing (2012).
[17] N. Maksyshko, O. Vasylieva, I. Kozin, V. Perepelitsa, Comparative analysis of the attractiveness of investment instruments based on the analysis of market dynamics, in: A. Kiv (Ed.), 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, volume 2713 of *CEUR Workshop Proceedings*, CEUR-WS.org, 2020, pp. 219–238. URL: http://ceur-ws.org/Vol-2713/paper18.pdf.
[18] A. O. Bielinskyi, I. Khvostina, A. Mamanazarov, A. Matviychuk, S. Semerikov, O. Serdyuk, V. Solovieva, V. N. Soloviev, Predictors of oil shocks. Econophysical approach in environmental science, IOP Conference Series: Earth and Environmental Science 628 (2021) 012019. doi:10.1088/1755-1315/628/1/012019.
[19] A. Bielinskyi, S. Semerikov, O. Serdyuk, V. Solovieva, V. N. Soloviev, L. Pichl, Econophysics of sustainability indices, in: A. Kiv (Ed.), 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, volume 2713 of *CEUR Workshop Proceedings*, CEUR-WS.org, 2020, pp. 372–392. URL: https://ceur-ws.org/Vol-2713/paper41.pdf.
[20] V. Derbentsev, H. Velykoivanenko, N. Datsenko, Machine learning approach for forecasting cryptocurrencies time series, Neiro-Nechitki Tekhnolohii Modelyuvannya v Ekonomitsi 2019 (2019) 65–93. doi:10.33111/nfmte.2019.065.
[21] T. Kmytiuk, G. Majore, Time series forecasting of agricultural product prices using Elman and Jordan recurrent neural networks, Neiro-Nechitki Tekhnolohii Modelyuvannya v Ekonomitsi 2021 (2021) 67–85. doi:10.33111/nfmte.2021.067.
[22] V. N. Soloviev, A. Belinskiy, Complex Systems Theory and Crashes of Cryptocurrency Market, in: V. Ermolayev, M. C. Suárez-Figueroa, V. Yakovyna, H. C. Mayr, M. Nikitchenko, A. Spivakovsky (Eds.), Information and Communication Technologies in Education, Research, and Industrial Applications, Springer International Publishing, Cham, 2019, pp. 276–297. doi:10.1007/978-3-030-13929-2_14.
[23] V. Derbentsev, A. Matviychuk, V. N. Soloviev, 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, Springer, Singapore, 2020, pp. 211–231. doi:10.1007/978-981-15-4498-9_12.
[24] CAAM, China Association of Automobile Manufacturers (CAAM), 2022. URL: http://en.caam.org.cn/.
[25] Car Sales in China, 2022. URL: https://www.chinamobil.ru/eng/sales/sales.
[26] L. Qian, J. M. Grisolía, D. Soopramanien, The impact of service and government-policy attributes on consumer preferences for electric vehicles in China, Transportation Research Part A: Policy and Practice 122 (2019) 70–84. doi:10.1016/j.tra.2019.02.008.
[27] BYD North America, Bus: BYD Motors - BYD USA, 2022. URL: https://en.byd.com/bus/bus-byd-motors.
[28] N. Shirouzu, Home from home: Mercedes-Benz doubles down on China, 2021. URL: https://tinyurl.com/mvkksnn6.

[29] Roewe, Roewe Brand - SAIC Roewe official website, 2022. URL: https://www.roewe.com.cn/about-us.
[30] Geely, Overview : Geely Global, 2022. URL: http://global.geely.com/overview.
[31] Chery, Introduction, 2022. URL: https://www.cheryinternational.com/pc/aboutchery/introduction/.
[32] M. Faggini, Chaotic time series analysis in economics: Balance and perspectives, Chaos: An Interdisciplinary Journal of Nonlinear Science 24 (2014) 042101. doi:10.1063/1.4903797.
[33] F. Takens, Detecting strange attractors in turbulence, Dynamical systems and turbulence 898 (1981) 366–382. doi:10.1007/BFb0091924.
[34] J.-P. Eckmann, S. O. Kamphorst, D. Ruelle, Recurrence Plots of Dynamical Systems, Europhysics Letters 4 (1987) 973–977. doi:10.1209/0295-5075/4/9/004.
[35] A. P. Rotshtein, D. I. Katelnikov, Fuzzy-chaotic time series prediction, Optoelectronic information-power technologies 27 (2014) 42–55. URL: https://oeipt.vntu.edu.ua/index.php/oeipt/article/view/332.
[36] R. Hegger, H. Kantz, Practical implementation of nonlinear time series methods: The TISEAN package, Chaos: An Interdisciplinary Journal of Nonlinear Science 9 (1999) 413–435. doi:10.1063/1.166424.
[37] M. B. Kennel, R. Brown, H. D. I. Abarbanel, Determining embedding dimension for phase-space reconstruction using a geometrical construction, Phys. Rev. A 45 (1992) 3403. doi:10.1103/PhysRevA.45.3403.
[38] S. Wallot, Recurrence Quantification Analysis of Processes and Products of Discourse: A Tutorial in R, Discourse Processes 54 (2017) 382–405. doi:10.1080/0163853X.2017.1297921.

A cognitive approach to modeling sustainable development of complex technogenic systems in the innovation economy*

Sultan K. Ramazanov[1], Bohdan O. Tishkov[1], Oleksandr H. Honcharenko[1] and Alexey M. Hostryk[2]

[1]*Kyiv National Economic University named after Vadym Hetman, 54/1 Peremohy Ave., Kyiv, 03057, Ukraine*
[2]*Odessa National Economic University, 8 Preobrazhenska Str., Odesa, 65000, Ukraine*

Abstract

Sustainable development of ecological, economic and socio-humanitarian systems is a crucial challenge in the modern world of instability and crises. To address this challenge, integrated models based on mathematical methods, models and innovative technologies are needed to manage and predict the nonlinear dynamics of these systems. Moreover, these models should incorporate humanitarian and cognitive variables that affect the behavior and decision-making of the system agents. In this paper, we present and develop a cognitive approach to modeling sustainable development of complex technogenic production systems in the innovation economy. We propose an integration model of sustainable development as a family of models for creating integrated information systems of ecological, economic and socio-humanitarian management of various social and organizational systems, especially economic objects of anthropogenic nature. We also present a cognitive model of nonlinear system dynamics that takes into account the dynamics of the humanitarian component with management in general. Furthermore, we introduce a model of innovation capital dynamics for the eco-economic and socio-humanitarian system (EESHS), as innovation capital is broader than intellectual capital by its nature and content. We derive an extended integral model of nonlinear stochastic dynamics of EESHS in the innovation space. The theoretical foundations and paradigms of our research are based on: systems of type "X", integral models and the problem of sustainable development, models such as "NMSSD" and systems such as "SEEHS", convergent technologies "NBIC" and "NBIC⊕SG".

Keywords

sustainable development, complex technogenic system, cognitive factor, innovation economy, knowledge-intensive enterprise, Industry 4.0, convergence, stochastic, human capital

1. Introduction

Sustainable development is defined as "development that meets the needs of the present without compromising the ability of future generations to meet their own needs" [2]. It is a multidimen-

M3E2-MLPEED 2022: The 10th International Conference on Monitoring, Modeling & Management of Emergent Economy, November 17-18, 2022, Virtual, Online

* Extended and revised version of paper [1] presented at the 10th International Conference on Monitoring, Modeling & Management of Emergent Economy.

sramazanov@i.ua (S. K. Ramazanov); tishcov_b@ukr.net (B. O. Tishkov); alexgontcharenko@gmail.com (O. H. Honcharenko); AlexeyGostrik@gmail.com (A. M. Hostryk)
0000-0002-8847-6200 (S. K. Ramazanov); 0000-0003-3381-9103 (B. O. Tishkov); 0000-0003-4861-2439 (O. H. Honcharenko); 0000-0001-6143-6797 (A. M. Hostryk)

sional concept that encompasses ecological [3], economic [4] and socio-humanitarian aspects [5, 6]. Sustainable development aims to achieve a balance between environmental protection, social equity and economic growth [7]. It is also a global challenge that requires collective action and cooperation among all stakeholders.

The United Nations has adopted 17 Sustainable Development Goals (SDGs) as part of the 2030 Agenda for Sustainable Development, which sets out a 15-year plan to achieve them [7]. The SDGs cover various areas such as poverty eradication, health and well-being, education, gender equality, clean energy, climate action, peace and justice. The SDGs are interrelated and interdependent, meaning that progress in one area affects and depends on progress in other areas. The SDGs also reflect the complexity and diversity of the world's problems and solutions. Currently, there is some progress in many areas, but in general, actions to implement the goals have not yet reached the necessary pace and scale. These goals have also been adapted and accepted for implementation in Ukraine [8].

One of the key challenges for achieving sustainable development is to understand and manage the complex dynamics of ecological, economic and socio-humanitarian systems in the modern conditions of instability and crises. These systems are characterized by nonlinearity, uncertainty, feedback loops, emergent properties, self-organization and adaptation [9]. They are also influenced by various factors such as technological innovations, human behavior, social norms, cultural values, political decisions and environmental changes [10]. Therefore, to effectively address the problems and opportunities related to sustainable development, integrated models based on mathematical methods, models and innovative technologies are needed.

Moreover, these models should take into account not only the physical and material aspects of the systems, but also the humanitarian and cognitive aspects that affect the behavior and decision-making of the system agents. Humanitarian factors include ethical, moral, legal, social and psychological dimensions that shape the attitudes, values and preferences of individuals and groups [11]. Cognitive factors include mental processes such as perception, memory, learning, reasoning, problem-solving and creativity that enable individuals and groups to acquire, process and apply information [12]. These factors play an important role in determining the outcomes and impacts of sustainable development initiatives.

In this paper, we present and develop a cognitive approach to modeling sustainable development of complex technogenic production systems in the innovation economy. A technogenic production system is a system that consists of human-made elements such as machines, tools, materials, products and processes that interact with natural elements such as resources, energy and environment to produce goods or services. An innovation economy is an economy that is driven by technological innovations that create new products or services or improve existing ones [13]. A cognitive approach is an approach that focuses on understanding how human cognition influences or is influenced by system dynamics [14].

The main contributions of this paper are extension of the results presented earlier in [15, 16]:

- We propose an integration model of sustainable development as a family of models for creating integrated information systems of ecological, economic and socio-humanitarian management of various social and organizational systems.
- We present a cognitive model of nonlinear system dynamics that takes into account the dynamics of the humanitarian component with management in general.

- We introduce a model of innovation capital dynamics for the eco-economic and socio-humanitarian system (EESHS), as innovation capital is broader than intellectual capital by its nature and content.
- We derive an extended integral model of nonlinear stochastic dynamics of EESHS in the innovation space.

2. Results

Currently, the determining factors of a knowledge-intensive enterprise (KE) are not so much production capacity, but rather knowledge, know-how, research and development.

The theory of production factors (PF) by the beginning of the 21st century became one of the actual research directions, covering the methodology of economic analysis and management of economic subjects. The main postulate of the theory of production factors is that the ratio of external factors of production and the internal state of the economic entity determines its strategic position in a complex and multidimensional market space, i.e. its organizational, economic and structural sustainability [17, 18, 19, 20].

The main provisions of the modern theory of PF can be formulated as follows: organizational and economic sustainability of the economic entity is determined by the ratio of available factors of production and their effective management; competitive advantages of the economic entity depend on the availability (including ownership) of strategic resources; effective management of available factors of production is provided by organizational capabilities of KE; taking into account cognitive, stochastic, humanitarian and "NOT-" factors.

A logical question arises: what properties should the factors of production have, so that the innovative development of the KE could be effective, intensive and adaptive?

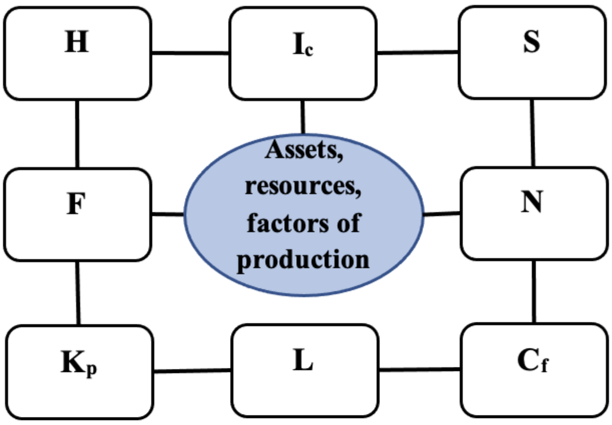
To answer this question, it is necessary to clarify the list of PF, which play a key role for the sustainable functioning and development of KEs, to introduce the concept and give a definition of cognitive factors of production; to develop a classification of cognitive factors of production, etc.

To implement this task, we will use the system paradigm, analyze the known concepts of PF and identify the main characteristics of cognitive production factors, which determine the organizational and economic sustainability of KE[1] (figure 1) [15].

Cognitive production factors (CPF). The analysis of the development of the theory of production factors and the emergence of their new types shows that the composition and role of production factors are most closely connected both with changes in production itself, and with the development of economic science, identifying and explaining the emergence and purpose of certain production factors by increasing opportunities for economic growth of knowledge-intensive enterprise.

Thus, according to the theory of human capital (the term was introduced by G. Becker [21, 22, 23]), the stock of knowledge, abilities and motivation embodied in a person contributes to the growth of human productive power. Human resources are to a certain extent similar to natural resources and physical capital, but in this interpretation they are divided into two parts.

[1]KE – knowledge-intensive enterprises of the high-tech sector of the economy. Knowledge-intensive enterprises (in other words, high-tech enterprises – HE) are technological leaders in the national innovation economy.



S	*Social capital*
F	*Financial capital*
N	*Natural capital (land, water, etc.)*
$\mathbf{K_p}$	*Physical capital (main production assets)*
L	*Labor resources (manpower)*
H	*Human / intellectual capital*
$\mathbf{C_f}$	*Cognitive factors of production*
$\mathbf{I_c}$	*Innovative capital*

Figure 1: "Octagon" of basic assets/resources that support the sustainability and safety of the system.

The unit of "human capital" is not the worker himself, but his knowledge. However, this capital does not exist outside of its bearer. And this is the fundamental difference between human capital and physical capital – machines and equipment.

By its economic essence, human capital is closer to the intangible fixed assets of an enterprise. According to the theory of human capital, investments in human beings are regarded as a source of economic development, no less important than "ordinary" capital investments. This means that an economic dimension is applied to a person.

The modern stage of KE development is characterized by qualitative changes in the types of socially significant human activity: labor characteristic of an industrial society is replaced by creativity in a post-industrial society. Machine technology gives way to "intellectual technology". As a result, knowledge and information become the leading factors of production, which leads to a decrease in the role of material factors of production. Radical changes in production relations have led to special requirements for the quality of human resources, highlighting their intellectual component and making them an independent factor of production.

Let us introduce the concept of cognitive production factor (CPF) – it is an embodied in an economic entity totality of knowledge, abilities, skills, which contribute to the growth of human productive power in the creation of an intellectual product demanded by the market.

The convergence of intellectual resources and information technology as a productive force causes the emergence of new types of factors of production – cognitive production factors (CPF, C_f) – which means specific, difficult to imitate resources of an industrial enterprise to create a product and added value, demanded by the market.

CPF are considered as a productive force arising from the convergence of human cognitive abilities and information technology.

Cognition as a scientific-cognitive action, is moving to a new quality, providing relevant knowledge for complex research. Artificial intelligence, neurocomputers, technologies of various interfaces based on the use of the properties of the human brain [24, 25] – a fundamentally new environment of human productive activity. The use of cognitive principles in economics allows to bring the main production processes to an intellectually new level.

CPF provide internal (endogenous) opportunities for the development of industrial enterprises and, in fact, become one of the sources of endogenous economic growth [18, 19]. The management of CPF means the emergence in the practice of industrial enterprises of a specific type of organizational and economic activity associated with their identification, ranking, analysis, evaluation and monitoring at all stages of the reproduction cycle to achieve the goals of long-term economic growth.

The allocation of CPF as a new type of productive force necessitates the development of appropriate methods and models of their management, the practical implementation of which is possible due to the mechanism of integration into the overall management circuit of the industrial enterprise.

The effectiveness of methods used in the management of traditional factors of production is becoming less effective, since it does not take into account the dynamics of modern changes, the need to process a large amount of data, the structural complexity of management tasks, the need to use coordination mechanisms.

The study of theoretical and practical results of production factor management allowed us to conclude that CPF management should be integrated into the overall management circuit of a high-tech enterprise and be supported primarily by end-to-end activities implemented through appropriate business processes.

The increasing intellectualization of industrial production contributes to the fact that the distinctive features of enterprises become:

- significant individualization of products in conditions of high flexibility of high-volume production;
- the modern vector of civilizational development of society is represented by the intensive spread of global technologies: nano-, bio-, information and communication technologies. Cognitive technologies refer to the technologies of the global level, the transformative effect of which gives a new quality of interacting elements and leads to the formation of a fundamentally new technological platform for economic development;
- integration of consumers and manufacturers in end-to-end processes of the entire product lifecycle and value chain;
- integration of information and data within production networks, reflecting all aspects of requirements, design, development, production, logistics, operation, service, etc., i.e.

creation of "production intelligence";

- globalization of product/goods development teams, as the complexity of products requires a variety of competencies;
- formation of a networked production "ecosystem" through cooperation and reduction of barriers between enterprises and customers;
- development of cloud technologies as a way to implement customized production on demand; use the production capabilities of virtual production networks based on united production sites, and support them with special software;
- isolation and accumulation of intangible functions, such as research and forecasting of the market and demand, formation of the product concept, formation of technical requirements, etc.; since intangible components take an increasing share in the cost and price of the finished product;
- formation of the market value of enterprises due to the knowledge of employees, know-how, knowledge-intensive technologies, inventions, industrial designs and other intangible assets. The qualitative change of production factors puts forward a set of interrelated tasks for industrial enterprises [19, 20];
- the integration into Industry 4.0, increasing the continuity and flexibility of production, the transition to flexible production systems that ensure the adaptation of the production infrastructure to innovative activities, changes in market requirements demand different approaches to the composition and configuration of key factors of production [26];
- increased consistency in the duration and productivity of all interrelated subdivisions of industrial enterprises causes the accounting of results not only at the place of application of production factors, but also in related units from the perspective of their impact on the economic performance of enterprises;
- rational increase in the growth of R&D costs, which ensures the implementation of scientific and technological policy directly in the process of scientific and production activities, determines the assessment of their relationship with the share of revenues from new types of products;
- the uncertainty of the economic environment, high risks in the development of innovative products create the preconditions for the development of economic-mathematical models that are adequate to the object of research and improve the quality of the effectiveness of industrial enterprises.

Thus, sustainable economic growth and development of modern industrial enterprises determines not so much the number of personnel, but the presence of workers who are able to conduct scientific and technological development at the modern level, to create competitive products and services on their basis, to propose new ways of organizing production, to determine the process of forming new trends in technological development in the market environment. In this regard, we need a different system of productive forces, surpassing the capabilities of industrial type of production and other ways of combining human and material labor.

The convergence of intellectual resources and information technologies as a productive force causes the emergence of new types of production factors – cognitive factors of production – which are understood as specific, difficult to imitate resources of an industrial enterprise that allow creating a product that is in demand by the market.

Cognitiveness, as a scientific and cognitive action, is moving into a new quality, providing appropriate knowledge for comprehensive research. Artificial intelligence, neurocomputers, technologies of various interfaces based on the use of the properties of the human brain are a fundamentally new environment for human production activities. The use of cognitive principles in the economy allows you to bring the main production processes to an intellectually new level.

Cognitive production factors provide internal opportunities for the development of industrial enterprises and, in fact, become one of the sources of endogenous economic growth [18, 19, 20]. Cognitive production factors management means the emergence in the practice of industrial enterprises of a specific type of organizational and economic activity related to their identification, ranking, analysis, evaluation, monitoring at all stages of the reproduction cycle in order to achieve the goals of long-term economic growth.

The identification of cognitive factors of production as a new type of productive force necessitates the development of appropriate methods and models of their management, the practical implementation of which is possible due to the mechanism of integration into the overall control loop of an industrial enterprise [15, 16, 27, 28].

So, cognitive production factors (CPF, C_f) – are the result of the convergence of intellectual resources / intellectual capital and information technology:

"IR/IC" & "IT",

where & – here is a conditional symbol of convergence.

Cognitive basis of high-tech activity, which includes the unity of knowledge, experience, creativity and information technology. Structural elements of CPF are: knowledge, experience, creativity and skills in the use of information technology, i.e. *CPF – is a tuple <knowledge, experience, creativity, level of use of IT, ...>.*

One of the variants of correlations of cognitive production factors (CPF), human capital (HC) and intellectual capital (IC) by three comparison parameters.

1. Structural elements:
   - CPF: Knowledge, experience, creativity, skills, in the use of information systems and technology.
   - HC: Level of education, health status.
   - IC: Market assets, human assets, intellectual property, infrastructure assets.
2. Methods of evaluation and measurement:
   - CPF: Indicator based on up-to-date financial and accounting statements.
   - HC: Aggregated indices, the calculation of which requires an extensive information base.
   - IC: Ratio of market value to book value; Intellectual coefficient of value added.
3. Correlation with performance results:
   - CPF: Production function.
   - HC: The balanced scorecard system.

- IC: Aggregate of IC and capital involved.

Note that the presented list of CPF is not exhaustive, it can and should be supplemented and improved.

So, CPF is a set of both active and intensional, as well as tangible and intangible factors of production:

- tangible-active can include those CPF, which are embodied and directly used in the economic turnover. These include local computer networks for information exchange, flexible manufacturing systems (FMS), simple/complex robots, automated information storage and retrieval systems, planning systems (ERPI, ERPII), design systems (CFD, CAE, PLM), electronic document management systems, vision systems;
- intangible assets include objects of intellectual property: know-how, technical solutions, licenses, patents, databases, information about customers and suppliers, etc;
- material-intentional cognitive factors include the potential use of advanced technologies, such as augmented reality technologies, artificial intelligence technologies: Internet of Things technologies, big data, cloud computing, deep learning, 5G, etc;
- intangible-intentional include personal characteristics of employees, experience, culture of thinking, ability to learn, creativity, insight, intuition, level of education, level of digital literacy, ability to cognitive activity, analysis, reflection, self-regulation, communication abilities, compliance with ethical and social norms.

Let us also note now that innovation capital is one of the most important and specific forms of capital, reflecting the ability of industrial enterprises as participants in the innovation cluster to generate income due to the development of innovative activity and acquisition of a special status due to the dynamics of innovation potential as an institution capable of transformation into capital as a result of the synergistic effect of interaction between economic entities in the process of innovation development. Innovation capital from the point of view of classical economic theory is characterized by three essential features, namely, it is a product of past labor, the role of which is played by innovation potential; it is a production or product stock in the form of innovations produced and ready for implementation, as well as innovations requiring further improvement and innovations that can be accumulated in the form of innovation potential; it is a source of income based on the effective commercialization of innovation [29, 30].

By its nature and content, innovation capital is wider than intellectual capital, which according to the concept presented in the works of Milner [29, 23], consists of three elements: 1) human capital; 2) organizational (structural) capital; 3) consumer capital. Machlup [22] in 1966, analyzing the processes of knowledge production and dissemination in the United States, without downplaying the role and importance of material production, reasonably proved that the economic development of the "new age" is determined not so much by the availability and productivity of material resources as by the availability and speed of information distribution in society and the amount of intellectual capital [17, 18, 20].

Let us present a cognitive model of the nonlinear dynamics of the system, taking into account the dynamics of the humanitarian component with control (as an extension of the integral model [15, 16]), in general terms it can be represented as stochastic differential equations:

$$\frac{dH_{\mathcal{U}}(t)}{dt} = \chi_+ H_{\mathcal{U}}^+(t) - \chi_- H_{\mathcal{U}}^-(t) + \sigma_{H_{\mathcal{U}}}(H_{\mathcal{U}}, t) dW_{H_{\mathcal{U}}}(t) + b_{H_{\mathcal{U}}} U_{H_{\mathcal{U}}}(t). \tag{1}$$

$$\frac{dC_f(t)}{dt} = \vartheta_+ C_f^+(t) - \vartheta_- C_f^-(t) + \sigma_{C_f}(C_f, t) dW_{C_f}(t) + \vartheta_{C_f} U_{C_f}(t). \tag{2}$$

The model of the dynamics of innovativeness of the eco-economic and socio-humanitarian system (EESHS) can also be represented in the form of an equation of dynamics:

$$\frac{dI_c(t)}{dt} = \varsigma_+ I_c^+(t) - \varsigma_- I_c^-(t) + \sigma_{I_c}(I_c, t) dW_{I_c}(t) + \vartheta_{I_c} U_{I_c}(t). \tag{3}$$

In equations (1)-(3) the variable $H_{\mathcal{U}}(t)$ is a humanitarian variable, $C_f(t)$ – cognitive variable, $I_c(t)$ – variable (level) of innovativeness in the integral model EESHS [16]; $\chi_+, \chi_-, \vartheta_+, \vartheta_-, \varsigma_+, \varsigma_-$ – parameters, and other designations are given in the same work.

So, supplementing the system of equations of the integral model [15, 16, 31] with equations (1) – (3) we obtain an extended (generalized) integral model of nonlinear stochastic dynamics of EESHS in the innovation space.

The generalized production and technological function (PTF) can now be represented as:

$$Y(t) = F[K(t, L(t), H(t), N(t), \Phi(t), S(t), I_c(t), C_f(t); \vec{c}]. \tag{4}$$

It can be used to study sustainable development.

In the general case, the integral level of sustainable development can be represented as a nonlinear function:

$$Y_{sdl}(t) = F_{sdl}[K(t), L(t), H(t), N(t), \Phi(t), S(t), I_c(t), C_f(t), \vec{c}]. \tag{5}$$

Private versions of the PTF model:

a) Mankiw-Romer-Weil model. Option of accounting for human capital H in the production function (PF), along with physical capital (K), labor (L) and natural (N) resources:

$$Y(t) = K^{\alpha}(t) \cdot H^{\beta}(t) \cdot [A(t) \cdot L(t)]^{1-\alpha-\beta}, \tag{6}$$

where $\alpha, \beta > 0$, $\alpha + \beta < 1$; H; $A(t)$ – function of scientific and technological progress. Note that α – is a part of capital provided by investment growth (capital costs); β is similar.

b) Model of accounting for all fixed assets:

$$Y(t) = A(t) K^{\alpha}(t) \cdot L^{\beta}(t) \cdot H^{\gamma}(t) \cdot S^{\rho}(t) \cdot \Phi^{q}(t) \cdot N^{\tau}(t) \cdot I^{\nu}(t), \tag{7}$$

where $\alpha, \beta, \gamma, \rho, q, \tau, \nu > 0$ and $\alpha + \beta + \gamma + \rho + q + \tau + \nu = 1$.

The following notations are also used here: K – physical capital, L – labor (labor), H – human capital, S – social capital, Φ – financial capital, N – natural resources (land, water, etc.), $A(t)$ is a function of the level of scientific, technical and technological development, for example, $A(t) = aT^S(t)$, where $T(t)$ – volume of innovative technologies (resources).

In [16], the equation of the dynamics of the potential of the R&D sector in the integral model is presented as:

$$\frac{d}{dt}[\dot{\varphi}(t)] - \delta_\varphi \varphi(t) = G[\varphi(t)]^{\gamma_1} \cdot [\alpha^1_{L_1}(t) L_1(t)]^{\gamma_2} \cdot [\alpha^1_K(t) K(t)]^{\gamma_3} \cdot [s(t)]^{\gamma_4} + \sigma_\varphi(\varphi, t) e_\varphi(t), \quad (8)$$

where $\varphi(t)$ – stock of knowledge and technologies in the economy – the number of inventions that have not lost their relevance by the year t; $\dot{\varphi}(t)$ – increase in the stock of knowledge per unit of time – the number of new inventions per year t minus obsolete; $L_1(t)$ – the volume of skilled (more precisely – highly skilled) labor (skilled labor force with qualifications, i.e. the product of the number of skilled workers $L_1(t)$ and the level of qualification of the average employee $h(t)$, i.e. $h(t)L_1(t)$); s(t) – social index; δ_φ – the rate of knowledge attrition due to its obsolescence $\delta_\varphi > 0$; $\alpha^1_{L_1}(t)$ – share of skilled labour employed in the R&D sector $0 \leq \alpha^1_{L_1}(t) \leq 1$; $\gamma_1, \gamma_2, \gamma_3$ – static parameters $0 \leq \gamma_1 \leq 1$, $0 \leq \gamma_2 \leq 1$, $0 \leq \gamma_3 \leq 1$; G – scale parameter: $G > 0$. Here $\{e_\varphi(t), t \in T\}$ – white noise with continuous time; $\sigma_\varphi(\varphi, t)$ – volatility coefficient.

From [15, 16] we have a more general equation of dynamics, i.e. the equation of the STP index (STP weight), which shows the growth and efficiency of the use of labor, capital and technology in production, i.e. $\tau(t)$:

$$\frac{d}{dt}[\dot{\tau}(t)] + \delta_\tau \tau(t) = B[\dot{\varphi}(t) + \delta_\varphi \varphi(t)]^{\beta_1} \cdot [\dot{\sigma}(t) + \delta_\sigma \sigma(t)]^{\beta_2} [\dot{s}(t) + \delta_s s(t)]^{\beta_3} [\dot{z}(t) + \delta_z z(t)]^{\beta_4} \quad (9)$$

where $\dot{\tau}(t)$ is the increase of the STP index caused by the change in the number of advanced production technologies used in production per unit of time, δ_τ – the rate of decrease of the STP index due to the obsolescence of advanced production technologies, $\delta_\tau > 0$; $\beta_1, \beta_2, \beta_3, \beta_4$ – static parameters, $0 \leq \beta_1 \leq 1, 0 \leq \beta_2 \leq 1, 0 \leq \beta_3 \leq 1, 0 \leq \beta_4 \leq 1$; B – scale parameter; $B > 0$.

Note that $\tau(t)$ – STP index, dependent on the number of advanced production technologies $w(t)$ and used in production, for example, $\tau(t) = [w(t)]^d$, where $d - const$.

Now in this generalized and integral variant we can use the conditions of development stability from [15, 16, 32].

This construction of the indicator will reflect the importance of each of the considered components: eco-economic and socio-humanitarian subsystems (spheres) in the performance of the objective function. A change in any of the private indicators leads to a change in the value of the aggregate indicator and captures a change in the steady state of the region. In the general case, all indicators change over time, i.e. have a certain dynamic.

Simple conditions for sustainable development (SD) are defined as follows.

1) Condition of weak stability:

$$\frac{dF[\cdot]}{dt} \geq 0 \quad or \quad F_{t+1}[\cdot] \geq F_t[\cdot], \quad (10)$$

where

$$F_t[\cdot] = F[K(t), L(t), H(t), N(t), \Phi(t), S(t), I_c(t), C_f(t), \vec{c}] \quad (11)$$

2) Condition of strong stability:

$$\frac{dF[\cdot]}{dt} \geq 0 \quad , N = N^C + N^S \quad and \quad \frac{dN^C}{dt} \geq 0, or \quad N^C_{t+1} \geq N^C_t, \quad N = 1...4 \quad (12)$$

where N^C – critical part of natural capital, and N^S – natural capital, which can be replaced by artificial.

For example, given critical natural capital N^C, sustainable development can be supplemented by a time limit on depletion of this value. For a time-decreasing production function, the arguments of which are aggregated variables: labor – L, capital – K and natural – resource N, we will have the ratio:

$$F_t(K, L, N) \leq F_{t+1}(K, L, N) \tag{13}$$

or, in the general case:

$$F(K(t), L(t), H(t), N(t), \Phi(t), S(t), I_c(t), C_f(t), \vec{c}) \leq$$

$$\leq F(K(t+1), L(t+1), H(t+1), N(t+1), \Phi(t+1), S(t+1), I_c(t+1), C_f(t+1), \vec{c}) \tag{14}$$

And it also requires compliance with the condition of not decreasing in time the value of N^C, i.e. $N_t = N_t^C + N_t^S$, as well as the condition of partial replacement of natural capital N by artificial N^S (or non-renewable resource for renewable resource): $N_t = N_t^C + N_t^S$.

The integrated level of sustainable development for all capital (resources) can be defined, for example, in the case of linear dependence as:

$$Y_{sdl}(t) = c_1 K(t) + c_2 L(t) + c_3 H(t) + c_4 N(t) + c_5 \Phi(t) + c_6 S(t) + c_7 I_c(t) + c_8 C_f(t), \tag{15}$$

where $c_1, c_2, c_3, c_4, c_5, c_6, c_7, c_8$ are weight (normalizing and scaling) coefficients.

3. Conclusion

This paper presents and develops a cognitive approach to modeling sustainable development of complex technogenic production systems in the innovation economy. We propose an integration model of sustainable development as a family of models for creating integrated information systems of ecological, economic and socio-humanitarian management of various social and organizational systems, especially economic objects of anthropogenic nature. We also present a cognitive model of nonlinear system dynamics that takes into account the dynamics of the humanitarian component with management in general. Furthermore, we introduce a model of innovation capital dynamics for the eco-economic and socio-humanitarian system (EESHS), as innovation capital is broader than intellectual capital by its nature and content. We derive an extended integral model of nonlinear stochastic dynamics of EESHS in the innovation space.

Our approach is based on the theoretical foundations and paradigms of systems of type “X”, integral models and the problem of sustainable development, models such as “NMSSD” and systems such as “SEEHS”, convergent technologies “NBIC” and “NBIC$\oplus$SG”. We show how these concepts can be applied to understand and manage the complex dynamics of technogenic production systems in the context of innovation economy.

We also demonstrate how our approach can address the challenges posed by the transition to an information society, which leads to a change in the structure of total capital in favor of human capital, an increase in intangible flows, knowledge flows, intellectual and innovative capital. We investigate the problem of sustainable development based on 8 important assets that support the sustainability and viability of EESHS.

We claim that our approach can increase the efficiency of solutions in the management of technogenic production systems, enhance the utilization of innovations and identify areas of innovation strategies for the regions.

The presented result requires further research, generalizations and computer experiments on real data. We plan to extend our approach to other types of complex systems and domains, as well as to incorporate more cognitive factors and methods into our models. We also aim to develop practical applications and tools based on our approach that can support decision-makers and stakeholders in achieving sustainable development goals.

References

[1] S. K. Ramazanov, B. Tishkov, O. Honcharenko, A. Hostryk, The Problem of Estimating the Sustainable Development of Technogenic Production System in According to Cognitive Factors in the Innovation Economy, in: S. Semerikov, V. Soloviev, A. Matviychuk, V. Kobets, L. Kibalnyk, H. Danylchuk, A. Kiv (Eds.), Proceedings of 10th International Conference on Monitoring, Modeling & Management of Emergent Economy - M3E2, INSTICC, SciTePress, 2023, pp. 196–203. doi:10.5220/0011932800003432.

[2] World Commission on Environment and Development, Our Common Future, Oxford University Press, Oxford, 1987. URL: OurCommonFuture.

[3] V. Morkun, S. Semerikov, S. Hryshchenko, K. Slovak, Environmental Geo-information Technologies as a Tool of Pre-service Mining Engineer's Training for Sustainable Development of Mining Industry, in: V. Ermolayev, N. Bassiliades, H. Fill, V. Yakovyna, H. C. Mayr, V. S. Kharchenko, V. S. Peschanenko, M. Shyshkina, M. S. Nikitchenko, A. Spivakovsky (Eds.), Proceedings of the 13th International Conference on ICT in Education, Research and Industrial Applications. Integration, Harmonization and Knowledge Transfer, ICTERI 2017, Kyiv, Ukraine, May 15-18, 2017, volume 1844 of *CEUR Workshop Proceedings*, CEUR-WS.org, 2017, pp. 303–310. URL: https://ceur-ws.org/Vol-1844/10000303.pdf.

[4] S. Semerikov, S. Chukharev, S. Sakhno, A. Striuk, A. Iatsyshyn, S. Klimov, V. Osadchyi, T. Vakaliuk, P. Nechypurenko, O. Bondarenko, H. Danylchuk, Our sustainable pandemic future, E3S Web of Conferences 280 (2021) 00001. doi:10.1051/e3sconf/202128000001.

[5] O. Lavrentieva, V. Pererva, O. Krupskyi, I. Britchenko, S. Shabanov, Issues of shaping the students' professional and terminological competence in science area of expertise in the sustainable development era, E3S Web of Conferences 166 (2020) 10031. doi:10.1051/e3sconf/202016610031.

[6] E. Komarova, T. Starova, Majority values of school biological education in the context of education for sustainable development, E3S Web of Conferences 166 (2020) 10029. doi:10.1051/e3sconf/202016610029.

[7] United Nations, Sustainable development goals, 2015. URL: https://sdgs.un.org/goals.

[8] Sustainable Development Goals: Ukraine, Technical Report, Ministry of Economic Development and Trade of Ukraine, 2017. URL: https://me.gov.ua/Documents/Download?id=05822f66-290b-4b51-a392-347e76ebeb5f.

[9] D. H. Meadows, Thinking in Systems: A Primer, Chelsea Green Publishing, White River Junction, VT, 2008. URL: https://wtf.tw/ref/meadows.pdf.

[10] J. D. Sterman, System Dynamics: Systems Thinking and Modeling for a Complex World, Working Paper Series ESD-WP-2003-01.13, Massachusetts Institute of Technology, 2002. URL: https://www.researchgate.net/publication/44827001.

[11] I. Kant, Groundwork of the Metaphysics of Morals, Cambridge Texts in the History of Philosophy, Cambridge University Press, Cambridge, 1998. URL: https://cpb-us-w2.wpmucdn.com/blog.nus.edu.sg/dist/c/1868/files/2012/12/Kant-Groundwork-ng0pby.pdf.

[12] J. R. Anderson, Cognitive Psychology and its Implications, 9 ed., Worth Publishers, New York, NY, 2020.

[13] J. A. Schumpeter, Capitalism, Socialism and Democracy, Routledge, London and New York, 2003. URL: https://periferiaactiva.files.wordpress.com/2015/08/joseph-schumpeter-capitalism-socialism-and-democracy-2006.pdf.

[14] J. D. W. Morecroft, Strategic Modelling and Business Dynamics: A Feedback Systems Approach, 2 ed., Wiley, Chichester, UK, 2015.

[15] S. Ramazanov, O. Honcharenko, The problem of estimating the sustainable development of the integrated technogenic industrial system, SHS Web of Conferences 107 (2021) 11002. doi:10.1051/shsconf/202110711002.

[16] S. Ramazanov, O. Stepanenko, S. Ustenko, Information and innovative management technologies in ecological and economic systems, Kyiv national economic univeristy named after Vadym Hetman, 2020.

[17] N. Moiseev, Mathematical problems of systems analysis, Nauka, Moscow, 1981.

[18] V. Kolemaev, Economic and mathematical modeling. Modeling of macroeconomic processes and systems, Unity-Dana, Moscow, 2005.

[19] M. Krass, B. Chuprinov, Mathematical Methods and Models for Master's Students in Economics, Saint-Petersburg, 2006.

[20] A. V. Vorontsovsky, L. F. Vyunenko, Forecasting of economic development based on a stochastic model of economic growth, taking into account the pivot point, Vestnik of the St. Petersburg University. Series 5. Economics (2016) 4–32. URL: https://econpapers.repec.org/article/scn003571/16949689.htm.

[21] V. L. Makarov, A. R. Bahtizin, N. V. Bahtizina, A computable model of the knowledge economy, Ekonomika i matematicheskie metody 45 (2009) 70–82. URL: https://naukarus.com/vychislimaya-model-ekonomiki-znaniy.

[22] F. Machlup, The Production and Distribution of Knowledge in the United Statess, Princeton University Press, Princeton, New Jersey, 1962.

[23] B. Z. Milner, Knowledge Management: The Evolution and Revolution in the Organization, Infra-M, Moscow, 2003.

[24] O. M. Haranin, N. V. Moiseienko, Adaptive artificial intelligence in RPG-game on the Unity game engine, CEUR Workshop Proceedings 2292 (2018) 143–150.

[25] D. Lukianenko, I. Strelchenko, Neuromodeling of features of crisis contagion on financial markets between countries with different levels of economic development, Neiro-Nechitki Tekhnolohii Modelyuvannya v Ekonomitsi 2021 (2021) 136–163. doi:10.33111/nfmte.2021.136.

[26] V. Kobets, V. Yatsenko, Influence of the Fourth industrial revolution on divergence and convergence of economic inequality for various countries, Neiro-Nechitki Tekhnolohii Modelyuvannya v Ekonomitsi 2019 (2019) 124–146. doi:10.33111/nfmte.2019.124.

[27] V. Pankratov (Ed.), Strategy of socio-economic systems development based on foresight and cognitive modeling methodologies, Kyiv, 2017.
[28] E. Gorlacheva, Metodologiia upravleniia kognitivnymi faktorami proizvodstva vysokotekhnologichnykh promyshlennykh predpriiatii, D.Sc. thesis, specialty 08.00.05, N.E. Bauman Moscow State Technical University, Moscow, 2020.
[29] B. Z. Milner, Innovative Development: Economics, Intellectual Resources, and Knowledge Management, Infra-M, Moscow, 2009.
[30] V. M. Geets, V. P. Seminozhenko, Innovative prospects of Ukraine, Constanta, Kharkiv, 2006.
[31] V. Halitsin, S. K. Ramazanov, Integral stochastic nonlinear model of dynamics of innovation economy, Modeling and information systems in the economy (2016) 50–64. URL: https://ir.kneu.edu.ua/handle/2010/22805.
[32] M. Z. Zgurovsky, Sustainable development in the global and regional dimensions: analysis according to 2005, Kyiv, 2006.

University competitiveness in the knowledge economy: a Kohonen map approach*

Dmytro H. Lukianenko[1], Andriy V. Matviychuk[1,2], Liubov I. Lukianenko[1] and Iryna V. Dvornyk[1]

[1]Kyiv National Economic University named after Vadym Hetman, 54/1 Peremogy Ave., Kyiv, 03680, Ukraine
[2]Kryvyi Rih State Pedagogical University, 54 Gagarin Ave., Kryvyi Rih, 50086, Ukraine

Abstract

In the post-industrial knowledge economy, universities play a key role in the generation and dissemination of innovations. They are also becoming the drivers of digital transformation in science, business, countries, and society as a whole. This paper studies the factors of university competitiveness in the knowledge economy. A clustering approach is used to group countries based on their university competitiveness. The level of significance of normalized parameters is also assessed. The results of the study are used to propose an organizational design for a competitive model of the university. The key factors of the university's success in the system of open science, education, and innovation are also discussed. The findings of this study contribute to the understanding of the factors that drive university competitiveness in the knowledge economy. The proposed organizational design and key factors of success can be used by universities to improve their competitiveness and become drivers of innovation and transformation.

Keywords

university, competitiveness, knowledge economy, Kohonen map, clustering, open science

1. Introduction

Universities are essential institutions for generating and disseminating innovations in the knowledge economy. However, they face increasing competition and challenges in the global market of educational services, especially in the era of digital transformation [2]. Therefore, it is important to assess and enhance the competitiveness of universities using reliable and objective methods [3].

Many existing methods for measuring university competitiveness are based on expert opinions, subjective criteria, or simple statistical techniques. These methods often produce inconsistent, biased, or incomplete results. Moreover, they do not capture the complex and dynamic nature of university performance and its relation to various factors.

M3E2-MLPEED 2022: The 10th International Conference on Monitoring, Modeling & Management of Emergent Economy, November 17-18, 2022, Virtual, Online

*Extended and revised version of paper [1] presented at the 10th International Conference on Monitoring, Modeling & Management of Emergent Economy.

lukianenko@kneu.edu.ua (D. H. Lukianenko); editor@nfmte.com (A. V. Matviychuk); Lukolga555@gmail.com (L. I. Lukianenko); i_dvornik@ukr.net (I. V. Dvornyk)
0000-0002-3475-2126 (D. H. Lukianenko); 0000-0002-8911-5677 (A. V. Matviychuk); 0000-0001-6997-1575 (L. I. Lukianenko); 0000-0002-2558-9654 (I. V. Dvornyk)

Avralev and Efimova [4] have conducted a survey of students over the years, which showed that place in the university rankings is an increasingly important criterion for students when choosing a university. At the same time, most researchers criticize the widely used rating systems. Thus, Sayed [5] demonstrates that according to some of the world's leading ranking systems, a university may be at the top of the ranking, while in others it may not be ranked at all. Many researchers note [6, 7] that most of the global university rankings focus primarily on research, while at the same time not paying enough attention to the quality of teaching, student competences and learning outcomes, social responsibility, etc.

At the same time, most scientists agree that the main criteria that determine the competitiveness of universities are research and teaching [8, 5, 9, 10]. In addition, some authors emphasize the importance of other criteria, such as international cooperation with university research networks, involving foreign teachers and students, increasing international citation [11, 12, 13], quality of pedagogical staff [12, 14], social and environmental responsibility [15], digitization of all university functioning processes [16, 17, 18], expenditure on higher education per student [19], employability of graduates [20, 21]. The importance of cooperation with business to improve the competencies and employability of students and, as a result, the competitiveness of the university, is emphasized in the papers [20, 17, 22, 23].

As can be seen from the above review, all these works are aimed either at the analysis and criticism of known rating systems, or at the study of factors that affect the competitiveness of universities, or, at most, at the creation of own methods for calculating university ratings, which are based on the simplest statistical methods.

There are works in which advanced artificial intelligence technologies are used to analyze and rank universities according to certain areas of activity. For example, in [16] developed a fuzzy logic model for assessment and ranking of universities' websites by criterion of usability.

However, the analysis of developments in this direction did not allow to identify studies on the modeling of university competitiveness based on cutting-edge artificial intelligence technologies, moreover, which would not be based in the rating on the expertly set weights of the evaluation criteria.

2. Modeling method

Solving the task of evaluating the international competitiveness of universities is associated with a number of specific problems, because competitiveness does not have generally accepted evaluation indicator, units or measurement scales. This is a subjective category that depends on many factors affecting it. Moreover, the set of these factors and the degree of influence of each of them are also not determined by any objective circumstances and can be chosen by analysts and researchers depending on their own understanding of the essence of the category "competitiveness of universities", the development of the educational process, their own priorities, etc. All this imposes a significant imprint of subjectivism on the formation of methods of their evaluation.

It is possible to reduce the dependence on the subjective opinions of individual experts with the use of special modeling methods capable of revealing regularities in the structure of an array of heterogeneous data, when there are no predetermined values of the resulting indicator, such as for the international competitiveness of universities.

Under such conditions, the clustering approach is the most appropriate means of searching for hidden regularities in sets of explanatory variables. The main feature of this approach is that with its application, objects that belong to one cluster are more similar to each other than to objects that are included in other clusters. As a result, it becomes possible to form fairly homogeneous groups of researched objects that are characterized by similar properties.

There is a wide range of cluster analysis methods: K-means [25], K-medoids [26], Principal Component Analysis [27], Spectral Clustering [28], Dendrogram Method [29], Dendrite Method [30], Self-Organizing Maps – SOM [31, 32], Density-Based Spatial Clustering of Applications with Noise – DBSCAN [33], Hierarchical DBSCAN – HDBSCAN [34], Ordering Points to Identify the Clustering Structure – OPTICS [35], Uniform Manifold Approximation and Projection – UMAP [36], Balanced Iterative Reducing and Clustering Using Hierarchies – BIRCH [37], etc.

Each of these methods has its advantages and areas of application and tasks, where it reveals itself in the best way. Experimental studies on comparative analysis of the effectiveness of various clustering methods are described, in particular, in scientific works [38, 39, 40, 41].

Taking into account the capabilities of each of the mentioned methods and the specifics of this study, the Kohonen self-organizing maps toolkit was used to cluster countries by the level of competitiveness of universities, which, in addition to forming homogeneous groups of researched objects, provide a convenient tool for visual analysis of clustering results. In particular, in contrast to other clustering methods, the location of an object on the Kohonen map immediately indicates to the analyst how developed the investigated property is compared to others, because the best and worst objects according to the analyzed indicator are located in opposite corners of the self-organizing map.

The result of constructing the Kohonen map is a visual representation of a two-dimensional lattice of neurons that reflect the organizational structure of the countries of the world, forming clusters in which countries are similar to each other according to the group of indicators of evaluating the competitiveness of universities (figure 1).

The Kohonen self-organizing algorithm is a clustering method that reduces the dimension of multidimensional data vectors. It can be used to visualize clusters and to detect nonlinear patterns in input data structures. The main feature of such neural networks is unsupervised learning, when information about the desired network response is not needed to correctly set the parameters. In this study, self-organizing maps are used to summarize a complex set of data and clustering of countries by indicators that have the greatest impact on the international competitiveness of universities.

Thus, each neuron of the Kohonen layer receives information about the research object in the form of a vector $\mathbf{x}$, which consists of n explanatory variables (in our case, these are the characteristics that determine the competitiveness of universities). When a new data vector arrives at the input layer of the network, all neurons of the self-organization map participate in the competition to be the winner. As a result of such a competition, the winner is the neuron

$$o = \operatorname{argmin} \left\{ \left\| \mathbf{x} - \mathbf{w}^j \right\| \right\} \tag{1}$$

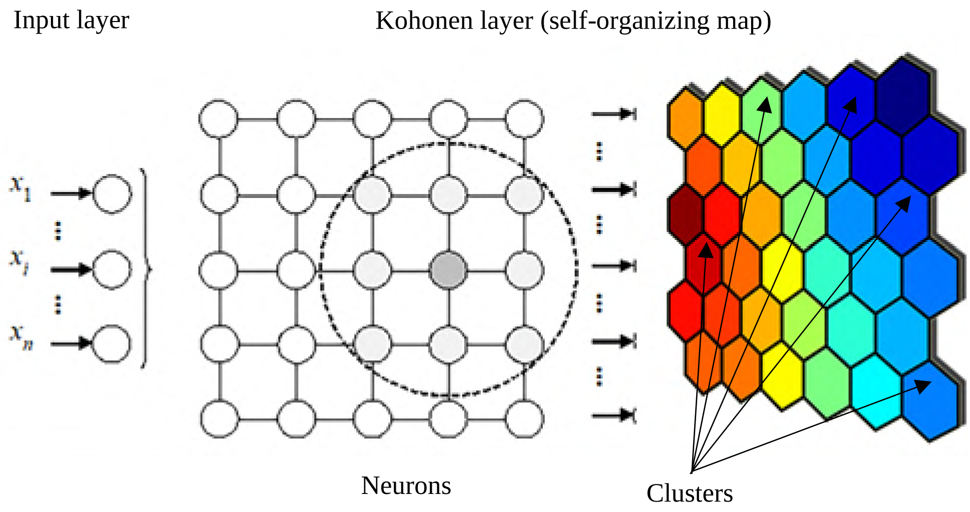


Figure 1: Visual representation of clusters on the self-organizing map [42].

that is more similar to the input data vector than others, usually by Euclidean distance:

$$\|\mathbf{x} - \mathbf{w}^j\| = \sqrt{\sum_{i=1}^{n} \left(x_i - w_i^j\right)^2}, j = \overline{1, K} \tag{2}$$

where $\mathbf{x}$ is a vector of input data consisting of indicators $\{x_1, \dots, x_i, \dots, x_n\}$ that describe the objects under study; $\mathbf{x}^j$ is the vector of parameters of j^{th} neuron of the Kohonen map, which consists of elements $\{w_1^j, \dots, w_i^j, \dots, w_n^j\}$; K is the number of neurons of the Kohonen map.

After determining the neuron-winner, we adjust the vector of its parameters and its neighbors according to the input vector:

$$\mathbf{w}^j(t+1) = \mathbf{w}^j(t) + \alpha(t) \cdot h_{oj}(t) \cdot \left[\mathbf{x}(t) - \mathbf{w}^j(t)\right], j = \overline{1, K} \tag{3}$$

where $\alpha(t)$ is the rate of learning ($0 < \alpha(t) \leq 1$), which decreases with each learning epoch t; $h_{oj}(t)$ is the strength of mutual influence for any pair of neurons o and j, determined as a function (usually Gaussian) of the distance between them on the map topology:

$$h_{oj}(t) = \exp\left[-\frac{\|\mathbf{r}_o - \mathbf{r}_j\|^2}{2 \cdot \sigma^2(t)}\right] \tag{4}$$

where $\mathbf{r}_o$, $\mathbf{r}_j$ are the two-dimensional vectors of coordinates of geometric location of the neuron-winner o and the j^{th} neuron on the map; $\sigma(t)$ is the effective width of the topological neighborhood (a specially chosen function of time that monotonically decreases in the learning process).

In the process of self-organization of the Kohonen map, the topological neighborhood narrows. This is caused by a gradual decrease in the width of the function $\sigma(t)$. The neuron-winner is located in the center of the topological neighborhood. It affects neighboring neurons, but this effect decreases with increasing distance to them according to (4). As a result, closely located map nodes acquire similar characteristics.

The result of the learning process will be the tuning of parameters of the Kohonen layer neurons, which will correspond to different examples from the training set. Thus, the self-organization of the structure of the Kohonen map is carried out, which acquires the ability to combine multidimensional data vectors in a cluster by identifying similar statistical characteristics in them. As a result, the initial high-dimensional space is projected onto a two-dimensional map. Since self-organization maps are characterized by the generalization property, they can recognize input examples on which they have not previously been tuned – the new input data vector corresponds to the map element to which it is mapped.

3. Collection of data for modeling

In order to correctly identify regularities in the development of the scientific and educational sphere, it is necessary to select the key properties that characterize the processes under study, taking into account the task. That is, it is necessary not only to choose the maximum possible set of characteristics of the objects of study, but to form a set of those features that describe the most significant aspects of activity in the context of the analysis. In this case, the selected features will make it possible to group the studied objects or processes according to their similarity. That is, if the task of analyzing the competitiveness of universities is being solved, then it is necessary to determine a set of characteristics of countries that will influence this indicator. And as a result of clustering the countries of the world according to these characteristics, we will get a number of clusters, each of which will group countries with a similar level of international competitiveness of universities (since they will have fairly close values of the characteristics that determine this competitiveness).

Therefore, we will conduct an analysis of publicly available databases that contain information on indicators that can influence the level of competitiveness of universities.

Thus, the World Bank's "World Development Indicators" database contains the ranking of the world's countries by the level of "Government expenditure on education, total (% of GDP)" indicator [43]. The indicator is calculated annually (for 266 countries) based on data from national statistics and international organizations, including data from the UN. Information on individual countries has been available in this database since 1970, in the last decade the data is presented quite fully, but only until 2018 (later data by countries is much less). Other indicators presented in this database are much poorer and less related to higher education.

In the Human Development Reports of UNDP [44] there are data for 195 countries for 2021 according to the indicators: "Human Development Index (HDI)" (both in general and by male and female sexes, in addition, by this indicator also shows the dynamics and increases in dynamics since 1990), "Government expenditure on education, % of GDP", "High-skill to low-skill ratio", "Research and development expenditure, % of GDP" (during 2014-2018), "Ratio of education and health expenditure to military expenditure" (during 2010-2017), "Foreign direct investment, net inflows, % of GDP", "International student mobility, % of total tertiary enrollment", indicators of employment and unemployment both in general and among young people, migrants, population by age group, etc.

The Global Competitiveness Index from the World Economic Forum for 2019 [45] can also be informative in assessing the international competitiveness of the country's universities. On this resource, this index is given for 141 countries. Later, in 2020, the Global Competitiveness Index has been paused.

Another resource with information on competitiveness is the annual reports of the European Commission [46], in particular in the areas of: "Competitiveness & Innovation", which contains separate reports and the following sections: "Global Innovation Index", "Global Attractiveness Index", "Global Talent Competitiveness Index", "Elcano Global Presence Index", "Innovation Output Indicator"; "Learning & Research", which presents reports: "European Skills Index", "European Lifelong Learning Indicators (ELLI-Index)", "Higher Education Rankings", "Composite Learning Index".

The work "Global Talent Competitiveness Index: 2019" [47] contains integrated assessments and ranking places of countries for a number of top-level indices, as well as for basic indicators.

To assess the competitiveness of world universities, the resource [48] can be useful, which provides fairly detailed country-level aggregated information on the research and educational activities of universities in 50 countries for 2020. Here are the indicators grouped into four generalized categories – "Resources", "Environment", "Connectivity", "Output". Each of these categories consists of a set of basic indices, all of which are listed in the header of the table 1.

In addition, we add to the database the overall competitiveness score and rank number in the general list (these indicators will not be taken into account when clustering countries, but will serve as a reference when analyzing clusters).

To carry out clustering based on Kohonen maps, it is necessary to avoid gaps in the data. Since there are only 50 countries in this database, moreover, the scores for each individual indicator for different countries are quite close to each other, so we will not divide countries into groups and replace the blanks with the corresponding average values for all countries. This will not lead to distortions of the clustering results, since the percentage of gaps in this database is very small.

4. Modeling the university competitiveness

The construction of Kohonen self-organizing maps in our study was carried out using the analytical platform Deductor Studio Academic. In the process of constructing a map, the task of finding its optimal dimension (number of neurons) arises, which is implemented experimentally on the basis of statistical data. The dimension of the self-organizing map was chosen from various options according to the mean weighted quantization error criterion, which reflects the average distance between the data vector given to the map inputs and neurons' parameters.

A hexagonal lattice of neurons with dimensions of 8 by 8 was determined as the most adequate structure of a self-organizing map for this task according to a given set of indicators (table 1). Self-organization occurs over 1500 learning epochs.The map parameters are initialized with small random variables. Gaussian (4) was chosen as a function of the neighborhood of neurons. Since all indicators for assessing the competitiveness of universities are already presented on an identical scale from 0 to 100, none of them will have a decisive influence on the clustering process.

Table 1
Indicators of evaluation of international competitiveness of countries' universities.

	OVERALL RANKING				RESOURCES 2020 SCORES					ENVIRONMENT 2020 SCORES					CONNECTIVITY 2020 SCORES					OUTPUT 2020 SCORES								
Country	Rank 2020	Rank 2019	Score 2020	Score 2019	Government expenditure on tertiary education as a percentage of GDP	Total expenditure on tertiary education as a percentage of GDP	Total expenditure per student USD PPP	Expenditure in tertiary institutions for R&D as a percent of GDP	Expenditure in tertiary institutions for R&D per head of population	Proportion of female students	Proportion of female academic staff	Data quality	Qualitative index of environment	WEF Survey	Proportion of international students	Proportion of articles with international collaborators	Webometrics VISIBILITY index divided by population	Rating of knowledge transfer between university and companies	Percentage of university research publications co-authored with industry	Total number of documents produced by higher education institutions	Total documents produced per head of population	Average impact of articles	Weighted Shanghai ranking scores for universities per head of population	Shanghai scores for best three universities	Tertiary enrollment rates	Percentage of population aged 24-64 with a tertiary qualification	Number of researchers in the nation per head of population	Unemployment rate of the tertiary educated compared with school leavers
Argentina	40	38	46	45,1	56,7	48,4	13,8	13,4	4,7	100	97,1	100	67,8	51,3	9,1	52,4	7	54,1	19,5	2,1	6,2	45,9	2,6	13,1	90	61,6	14,9	30,9
Australia	9	8	82,2	80,9	37,7	70,6	42,9	64	51,2	100	91,3	100	98,1	81,9	78,9	72,8	56	68,1	41,4	15,9	85,3	84,3	76,8	39,3	100	79	55	32,6
Austria	12	12	79,3	77,2	81,9	64,8	48,7	68,6	63,7	100	84,7	100	72	68,3	63,1	86,9	54	84,8	100	3,3	49,4	86	57	22	85,1	56,5	62,5	31,3
Belgium	13	13	75,6	73,6	63,8	55,4	48,3	53	45	100	97,1	100	75,8	82,2	31,8	89,3	28,4	82,5	78,8	4,8	56,6	94,2	51,4	31,4	79,7	70,2	59,9	39,1
Brazil	41	40	45,6	44,1	49,9	66,5	37,9	n.a.	n.a.	100	91,4	88,6	63,8	41,8	0,9	43,8	6,9	40,3	26,2	11,6	7,5	45,3	3,8	21,2	51,3	31,8	10,7	39,7
Bulgaria	45	44	42,7	41,8	32,3	40,3	17,8	4,3	1,6	100	97,9	93,2	53,1	54,7	16,8	57,5	10,9	46	44,3	0,8	14,6	55,2	3,3	2,7	71,2	38,1	25,8	45,2
Canada	7	6	83,2	81,9	62,5	86,8	62,9	63,6	53	100	88,6	90,9	73,3	87,1	47,4	68,8	69,2	86,3	59	17,2	62,6	82	44	42,9	88,2	100	51,8	33,7
Chile	31	32	54,3	51,3	48,4	100	22,3	14,8	6	100	85,1	100	81,4	54,8	1,4	81,6	14,3	62,3	28,3	2,2	16,1	63,4	8,2	11,8	88,5	43,5	6,1	30,2
China	26	27	56,8	54,7	42,5	50,7	20	15	4,4	100	n.a.	88,6	76,6	73	1,3	34,1	8,4	65,9	32	70,7	6,8	59,3	7,5	39,6	49,1	16,7	15	n.a.
Croatia	43	43	43,6	42,1	49,9	36,8	18	24,9	11,2	100	97,8	93,2	47,3	47	1,6	58,3	11	35,5	50,8	0,9	30,4	52,9	17,2	8	66,5	39,2	22,6	31,2
Czech Rep.	29	26	54,8	55,2	36,3	35,1	26,6	34,4	22,2	100	76,9	100	69,3	61,1	46,1	62	29,3	53,6	55,7	2,9	37	62,8	22	14,7	64,1	41,9	44,7	40,4
Denmark	3	5	85,7	82,5	80,4	62,7	44,9	100	91,4	100	88,6	95,5	67,4	80,6	39,5	85,3	47,5	89,2	85,2	4,3	100	97,1	83,3	38,8	80,6	65,7	95,7	21
Finland	8	9	82,8	80,4	80,8	61,9	46,6	68,5	54,6	100	100	100	81,6	93,8	30	82,3	64,7	90	77	2,9	70,8	86	72,2	23,9	88,2	78,1	81,3	41,3
France	17	17	68,6	67,6	57	53,6	43	44,3	35	100	87,9	100	73,1	69,6	37,4	77,7	23,8	70,3	68,8	13,5	28,2	75,6	28,9	40,6	65,6	63,7	53,8	39,8
Germany	16	16	70,5	69,6	51,3	44,9	46,3	51,2	46	97	78,6	100	61,6	86,8	30,8	67,8	38,6	87,9	76	21	34,2	79,1	32,9	39,7	70,2	50,2	61	37
Greece	37	37	47,4	47	35,1	26	10,9	31,8	15,5	97,1	68,6	93,2	26,9	49,2	12,5	68,7	35,2	43,7	61,5	2,5	31,2	73,3	21	14,1	100	54,8	38,2	36
Hong Kong	14	15	72,7	70,2	50,1	55,6	64,7	39,8	43,3	100	n.a.	90,9	97,2	76,7	42	54,3	48,2	82,5	35,8	3,5	63,7	95,9	54,9	26,3	74,3	50,9	41,4	41,4
Hungary	33	35	51,3	48,5	34,7	39,7	30	17,6	8,8	100	80,5	100	51,6	47	36,6	70,9	22,1	58,6	82,8	1,6	22,3	69,2	14,4	10,7	48,5	43,4	35,4	54,6
India	49	49	39,6	38,8	54,8	59,1	13	2,4	0,3	96,2	81,2	90,9	58,1	74,6	0,5	27,2	0,9	57,8	19	14,9	1,5	47,1	0,6	12,5	27,4	18,3	2,6	12,6
Indonesia	50	50	35	33,5	25,7	25	7,9	4,6	1	100	86,2	100	64,7	71,6	0,3	23,6	4,4	72,5	31,4	3,1	1,6	45,3	0	0	36,4	20,5	2,6	26,4
Iran	47	48	42,2	39,2	50,2	51,9	15	n.a.	n.a.	92,1	62,2	81,8	67	52,8	1,6	33,7	5,1	52,1	10,6	7,3	12	51,7	5	15,2	69,6	36,9	8,1	n.a.
Ireland	19	19	66	64,7	28,7	29,6	35,2	25,2	33,7	100	90	100	68,6	87,6	32,6	75,1	60,1	88,2	63	2,4	64,8	80,8	47,6	18,7	77,8	81,1	49,8	36,8
Israel	18	18	67,4	67,3	39,4	52,1	29,6	50,9	34,6	100	n.a.	95,5	73,3	74,9	10,6	66,3	34,6	91,1	49,5	3,2	48,5	77,9	51,5	30,6	63,4	88	100	34,6
Italy	30	30	54,5	53,4	28,6	33	30,8	32,1	22,5	100	74,2	100	63,8	60	19,5	62,9	18	60,5	54,2	15,7	34,9	77,3	29,4	24,6	61,9	33,4	27,8	35,6
Japan	20	20	61,9	61,7	21,2	51,2	51	37,6	28,9	95,4	56,8	100	83,2	70,8	15,7	39	18,9	57	78	17,2	18,4	50,6	14,5	42,9	63,6	89,7	64,3	34,5
Korea	24	23	58	57,4	32,7	64,4	27,8	37,7	25,7	83,4	70,2	100	58	56,3	8,3	37,9	14,8	62,6	61,4	12,4	32,3	56,4	24,1	24,8	94,3	84,7	91,1	25,2
Malaysia	27	28	56,1	54,5	56,5	75,1	39,4	48	22,9	100	100	95,5	78,6	83,7	29,6	59,5	7,5	79,4	16,3	3,6	15,1	55,8	5,8	14	43,7	37,7	28,6	21,6
Mexico	48	47	41,7	41,1	47,2	50,5	19,5	12,7	4,1	100	n.a.	95,5	82,4	48,5	2,1	53	3,8	52,9	19,5	3	3,2	42,4	0,8	11,1	40,2	31,1	3	20,7
Netherlands	10	10	81,6	80,2	59,8	62,7	51,8	58,3	54,3	100	91,7	100	79,3	88	40,4	82,5	47,5	96,7	85,4	9	70,7	97,7	59,4	37,5	85	66,2	60,7	34,7
New Zealand	14	14	72,7	71,5	44,1	64,5	39,7	33,6	21,4	100	99,7	100	89,7	86,5	72	77,3	55,8	76,1	46,4	2,2	59,5	79,1	64,6	18,4	82	67,9	49,1	33,9
Norway	11	11	80,5	77,8	89,9	70,7	58,4	68,8	74,8	100	92,6	100	66,9	85,9	11,6	80,8	58,9	81	61,8	3,1	78,2	87,2	63	28,1	82	75,3	78,5	32,2
Poland	32	31	52,6	52,2	48,4	43,9	23,8	33,3	17,2	100	90	100	81,9	58,3	15,1	42,3	17,3	60,4	32,3	6,4	22,8	58,1	7,3	14,1	67,8	53,4	30,6	49,1
Portugal	25	25	57,6	56,8	39,4	42,6	29,3	55,3	31	100	88,6	100	60,9	71,7	23,5	71,9	33,9	64,5	41,5	3,6	47,2	66,9	26,7	18,7	63,9	43,1	52	33,7
Romania	44	45	43	41,7	32,9	42,8	28,9	5,2	2,4	100	100	95,5	76	45,2	17,7	36,6	10,2	54	32,1	2,3	15,9	50,6	2,7	5,9	48,2	29,6	10,8	45,8
Russia	35	35	49,1	48,5	37,3	42,5	22,5	9,8	4,6	100	100	100	70,2	60,1	15	36,2	8,4	43,9	20,1	8,8	8,3	47,7	2,9	21,7	81,9	97,9	34,6	47,7
Saudi Arabia	22	22	59,3	59,3	100	77,7	53,1	n.a.	n.a.	96,3	81,7	79,5	50,5	69,3	17,1	100	3,9	68,9	29,6	3,1	12,7	76,7	7,8	24,8	69,7	41,2	n.a.	9,4
Serbia	42	41	44,2	43,4	55,8	48,7	17,4	32,9	8,8	100	93,1	90,9	42,3	52,9	16,3	62,1	8,5	52,1	23,7	1,1	20,5	53,5	9,3	7,4	66,5	37,2	25,2	28,7
Singapore	4	7	84,5	81,3	50,1	53,7	100	63,4	100	100	74,1	95,5	82	94	100	87,7	36,7	91,7	38,6	2,8	66,5	94,8	41,4	26,5	84,8	86,5	81,6	30,6
Slovakia	38	33	47,2	49,6	35,8	37	30,3	21,2	11,9	100	91,5	100	64,2	44,8	25,3	57,7	16,8	35,6	64,4	1	24,2	63,4	4,7	2,9	46,6	42,5	33,9	46,4
Slovenia	28	29	55,4	53,6	44,5	38,3	29,9	20,3	12,6	100	85,1	100	63,7	65,3	14,3	71,1	25,1	63,1	53,6	0,7	48,8	67,4	31,4	7,4	78,6	56,1	54,2	35,9
SAR	34	34	49,7	48,7	37,4	49,9	28,9	26,2	6,1	100	n.a.	88,6	86,7	45,3	11,9	68,6	3,7	54,8	36,9	3,7	8,6	69,8	5,8	18,6	22,4	12,4	6	100
Spain	23	24	58,6	57,3	41,6	45,9	33,5	32	21,6	100	86,9	100	69,9	59,5	11,9	61,6	30,7	57	46,2	12,7	37	65,7	29,9	22	88,9	64,4	34,8	39,7
Sweden	5	4	84,3	82,9	71,2	59,7	64,6	83,2	73,7	100	89,7	100	75,2	76,9	24,8	86,8	59,6	83,1	86,2	6,3	83,5	89,5	82,5	38,8	67	74,7	92	24,4
Switzerland	2	2	90,1	88,6	64,9	51,7	75,4	88	97,6	99,4	71	100	69,5	100	65,2	91,3	79,7	100	76,9	5,8	91,7	100	100	44,2	59,6	75,6	63,7	30,4
Taiwan	21	21	60,5	60,5	33,5	51,8	32,8	29,3	26,2	100	72	93,2	86,9	72,3	16,2	45,4	44,3	80	38,3	5,1	29,1	55,2	20,3	19,7	84,5	84,5	76,1	25
Thailand	46	46	42,3	41,2	32,1	34,8	13,7	13,7	4,2	100	100	95,5	71,9	60,1	4,8	57,8	10,2	65,5	34,7	2,1	4,3	53,5	1,8	11,3	49,3	28,1	14,7	18,2
Turkey	39	42	46,3	43,3	71,1	70,5	27,9	31,5	14,8	92	88	100	44,9	51,3	5,5	30,6	7,6	57,4	16,6	7,2	11,9	44,2	4,1	11,2	94,7	35,9	16,8	23,2
Ukraine	36	38	47,8	45,1	76,4	63,9	10,8	3,2	0,5	100	n.a.	90,9	60,6	62,4	11,6	41,2	8	45,8	60,4	1,3	4,2	33,1	0	0	83,4	84,4	12	58,3
UK	6	3	83,6	84,5	28,2	64,7	63,1	38,8	29,7	100	89,5	100	89,5	75,5	65,8	72,1	63,7	82,1	68,9	31,1	63,1	86	58,1	73,7	60	79,1	53,1	34,4
USA	1	1	100	100	42,6	91,2	80,1	35,7	37,7	100	98,2	100	100	90,8	19	45	100	92,3	58,4	100	41,2	78,5	43,4	100	88,2	81,9	51,6	48,7

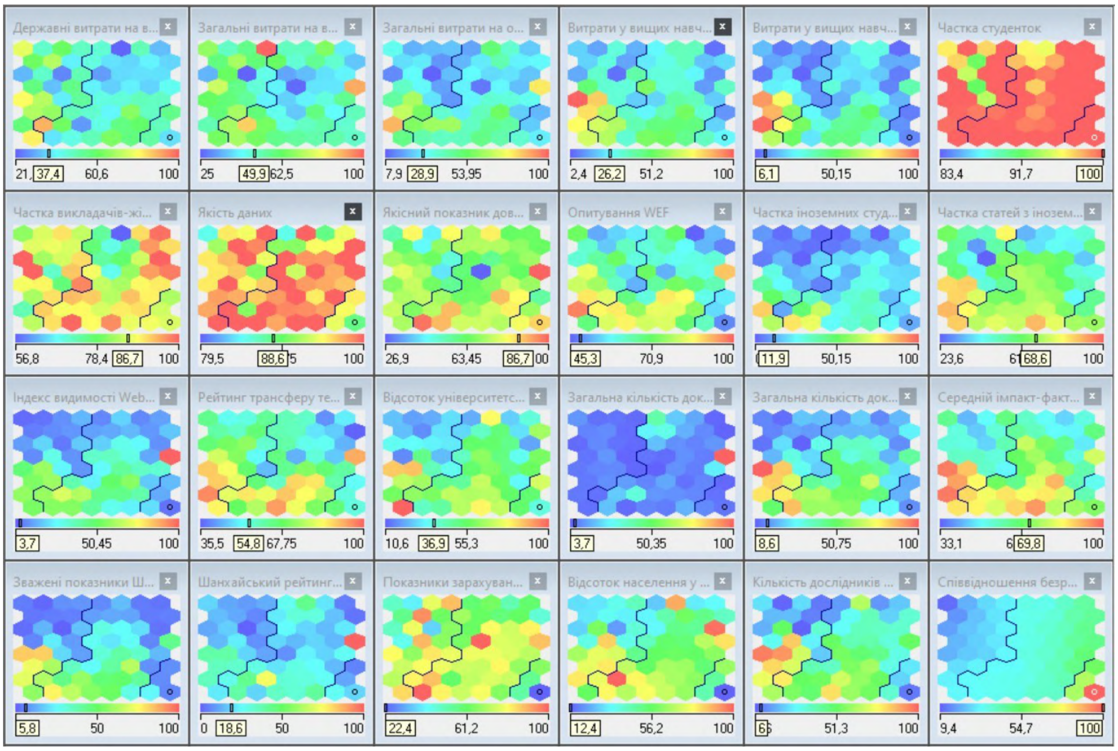


Figure 2: Kohonen topological maps for all indicators of university competitiveness assessment.

Therefore, it was decided to build Kohonen maps on the original data without processing them. As a result of the process of self-organization, the countries from the table 1 were distributed among three clusters, which can be seen in figure 2.

As can be seen from the topological maps for all indicators in figure 2, for the vast majority of them there is no clear demarcation of their levels between clusters. That is, their low, medium and high values are evenly distributed throughout the map, which, together with the low levels of significance of many indicators (figure 3), does not contribute to the quality of the countries segmentation process.

Given the low significance of a large number of indicators selected for the study, a series of experiments was conducted on the construction of Kohonen maps on different sets of input variables, when various combinations of the least influential factors were alternately removed. However, each time the same low quality of the distribution of countries by the levels of university competitiveness evaluation indicators remained. For example, for all clustering options, Bulgaria, South Africa, Poland, the Russian Federation, Romania, Slovakia, Hungary, and Croatia were located next to Ukraine on Kohonen map, but the United States was also a neighbor in this cluster. Of course, such segmentation of countries cannot be considered acceptable.

Therefore, it was decided to apply z-score standardization to process the initial values of the variables. As a result of forming a map on the full set of standardized explanatory variables, 5

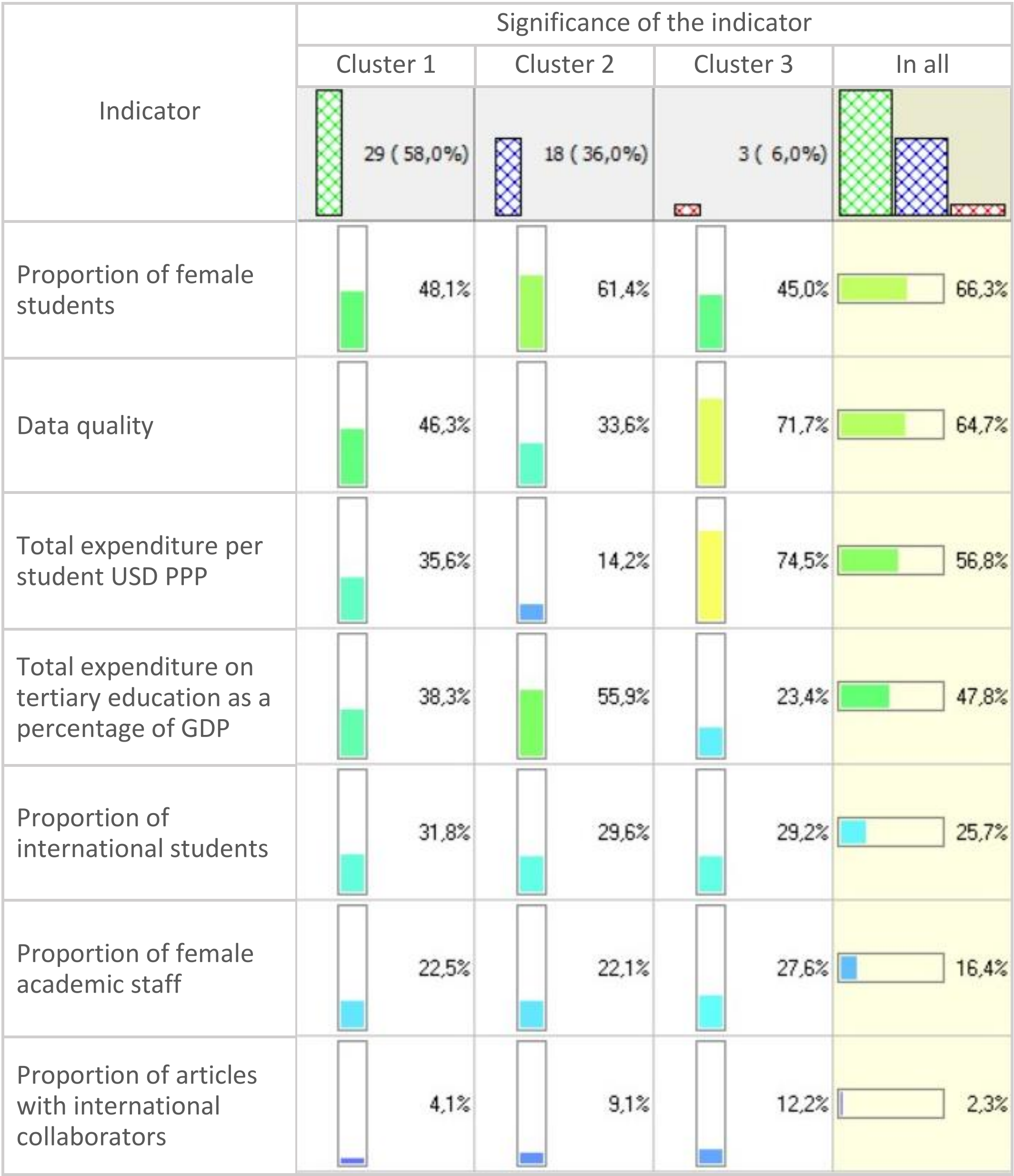


Figure 3: Levels of significance of a number of indicators for evaluating the competitiveness of universities.

clusters were obtained (figure 4).

Figure 4 shows that the levels of indicators change when crossing from cluster to cluster, which indicates a successful delimitation of countries based on a given set of explanatory variables. Ukraine got to the upper right corner of the Kohonen map next to Argentina, Bulgaria, Poland, the Russian Federation, Serbia, Turkey, Croatia, and Chile. Somewhat lower in the same cluster

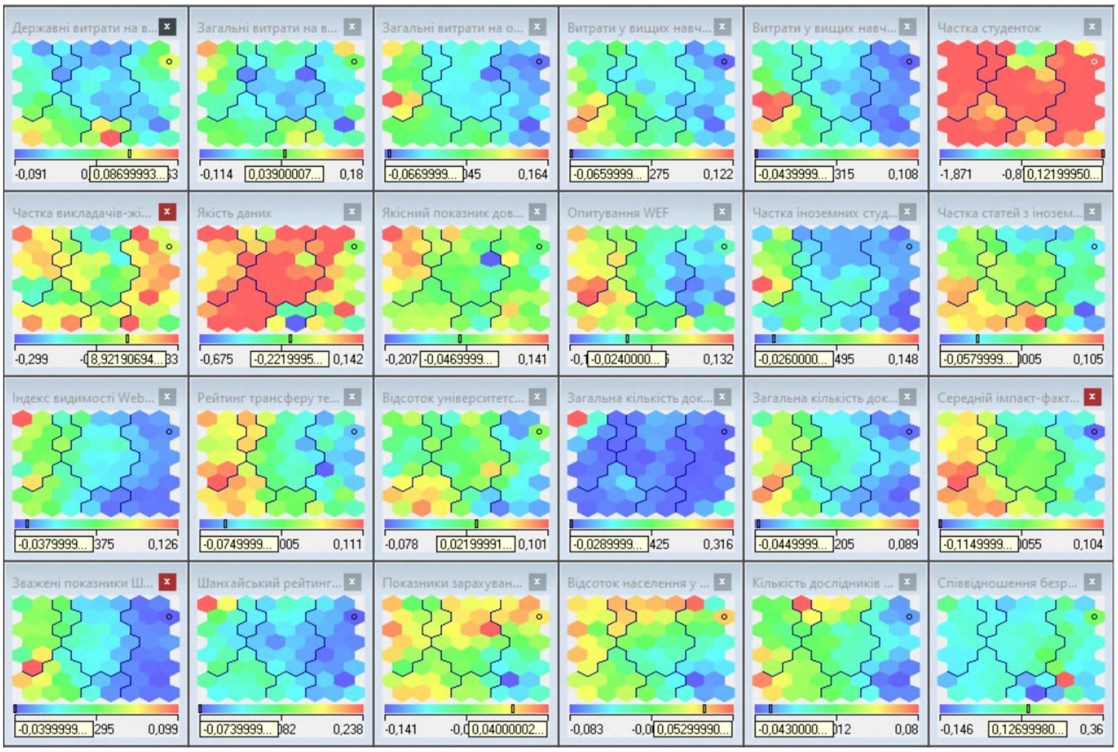


Figure 4: Kohonen topological maps according to the normalized indicators of university competitiveness assessment.

were Brazil, India, Indonesia, Iran, China, Malaysia, Mexico, South Africa, Romania, Slovakia, and Thailand.

Austria, Denmark, the Netherlands, Norway, Singapore, Finland, Switzerland, Sweden are located in the opposite corner of the map from Ukraine (bottom left). The United States and Great Britain were located in the upper left corner of the map. They are surrounded by Australia, Hong Kong, Israel, Canada, and Taiwan.

It should be noted that since, in accordance with the given task, polar objects are located on the Kohonen map in opposite corners, this self-organization of countries indicates that the competitiveness of Ukrainian universities is currently quite far from the competitiveness of universities in developed countries.

The analysis of the characteristics of the universities of the countries of the most developed cluster makes it possible to determine the priority areas of development and tasks that must be solved in order to increase the international competitiveness of Ukrainian universities.

Research and generalization of traditional, entrepreneurial, innovative and creative models of universities, their selection depending on objective endogenous and exogenous conditions and imperatives of the development of Ukrainian higher education made it possible to substantiate the most adaptive competitive model of the university, which is shown in figure 5.

Critically important in the proposed model is the development of strategic partnership in the triangle “science – business – education”, public-private partnership and consolidated social

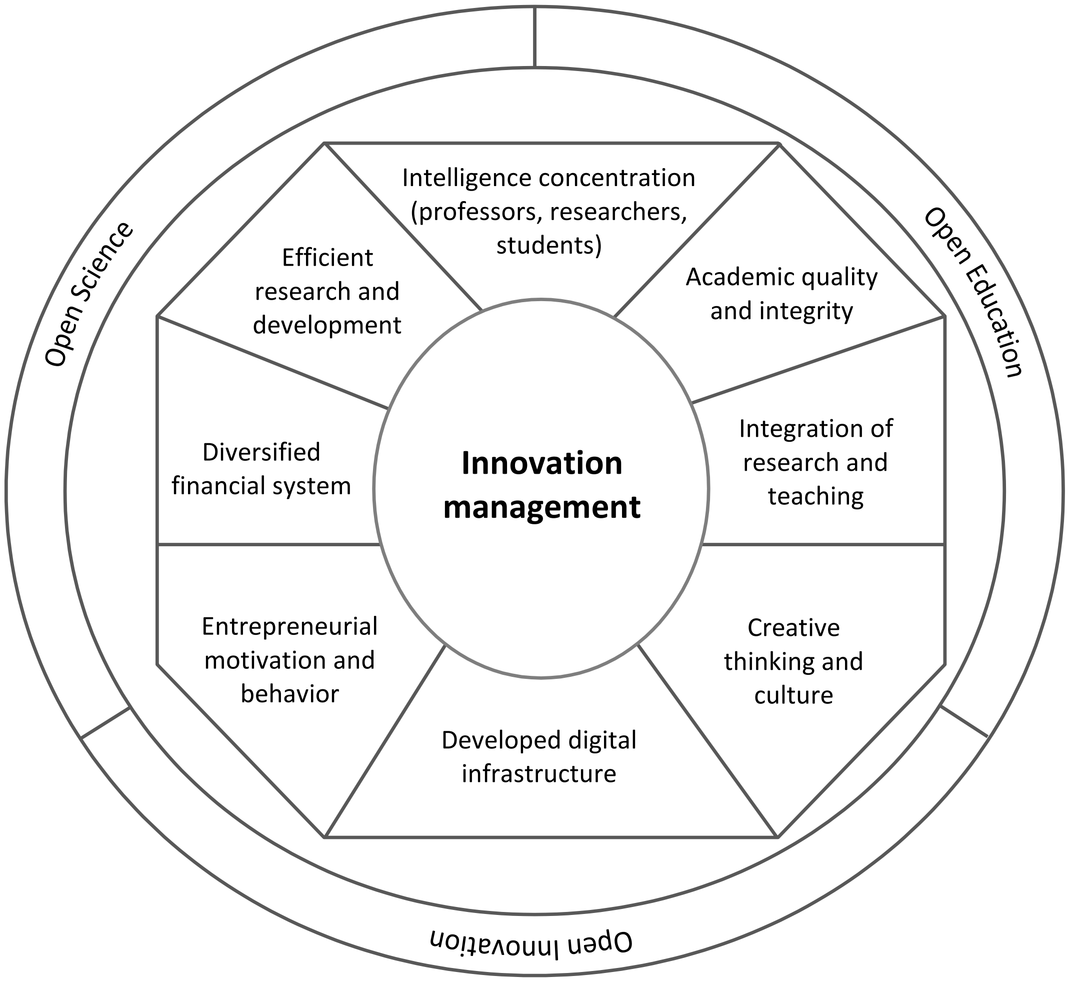


Figure 5: Competitive model of the university.

responsibility.

5. Conclusions

The evolving global landscape of university education presents fresh challenges for educational authorities and university administrations, urging them to bolster competitiveness in the international educational services market. Amidst the modern era of globalization, there arises a need to efficiently manage universities, necessitating the assessment of their international competitiveness.

In today's ever-changing world, organizations like corporations and universities must navigate political, market, and social turbulence. This calls for continuous generation of unconventional ideas, strategic concepts, and behaviors to drive innovation.

This research is driven by a mission to establish a fresh methodological approach for evaluating the often elusive indicator of university competitiveness. Given the lack of standardized evaluation indicators and measurement scales, a clustering approach was chosen to uncover hidden patterns within a set of explanatory variables.

The study involved a comprehensive exploration of existing approaches to assessing university competitiveness and identified unresolved issues in the field. Diverse clustering methods were analyzed, comparing their strengths and characteristics to identify the most fitting approach.

The innovation proposed in this paper lies in the application of artificial neural networks, particularly Kohonen maps, for modeling university competitiveness. Kohonen maps facilitate data clustering based on similarity and visual representation in a lower-dimensional space. Using these maps, we clustered countries based on parameters such as research output, teaching quality, internationalization, social responsibility, digitization, expenditure, and employability. We also employed Kohonen maps to rank the significance of these parameters for distinct country clusters.

The utilization of Kohonen self-organizing maps demonstrated its worth, not just in forming homogenous groups of research subjects, but also as a powerful tool for visually dissecting clustering outcomes. Moreover, this methodology aids in identifying lagging indicators, enabling strategic interventions to enhance the competitiveness of Ukrainian universities in the global educational services market.

This method boasts multiple advantages over existing ones. Firstly, it operates without depending on expert judgments or predefined parameter weights. Secondly, it accommodates vast and diverse datasets with differing variable types. Thirdly, it uncovers latent patterns not immediately apparent with conventional techniques. Lastly, it offers intuitive, interactive visualizations, facilitating comprehension and communication.

The culmination of this research yielded a competitive university model, unraveling the competitive strengths of universities within the most competitive cluster countries. This underscores the practical applicability and potency of the proposed approach in assessing and elevating university competitiveness in the contemporary educational landscape.

References

[1] D. Lukianenko, A. Matviychuk, L. Lukianenko, I. Dvornyk, Modelling the Design of University Competitiveness, in: S. Semerikov, V. Soloviev, A. Matviychuk, V. Kobets, L. Kibalnyk, H. Danylchuk, A. Kiv (Eds.), Proceedings of 10th International Conference on Monitoring, Modeling & Management of Emergent Economy - M3E2, INSTICC, SciTePress, 2023, pp. 204–214. doi:10.5220/0011933200003432.

[2] T. Sych, Y. Khrykov, O. Ptakhina, Digital transformation as the main condition for the development of modern higher education, Educational Technology Quarterly 2021 (2021) 293–309. doi:10.55056/etq.27.

[3] O. Aleksandrova, I. Hroznyi, N. Vinnikova, N. Chuvasova, Control of the quality assurance system at the modern Ukrainian university, Naukovyi Visnyk Natsionalnoho Hirnychoho Universytetu 2019 (2019) 153–162. doi:10.29202/nvngu/20192/18.

[4] N. Avralev, I. Efimova, University Rankings as a Tool for Assessing the Quality of Education

in the Context of Globalization, Asian Social Science 11 (2015) 292–298. doi:10.5539/ass.v11n10p292.
[5] O. H. Sayed, Critical Treatise on University Ranking Systems, Open Journal of Social Sciences 07 (2019) 39–51. doi:10.4236/jss.2019.712004.
[6] F. Anowar, M. A. Helal, S. Afroj, S. Sultana, F. Sarker, K. A. Mamun, A Critical Review on World University Ranking in Terms of Top Four Ranking Systems, in: K. Elleithy, T. Sobh (Eds.), New Trends in Networking, Computing, E-learning, Systems Sciences, and Engineering, Springer International Publishing, Cham, 2015, pp. 559–566. doi:10.1007/978-3-319-06764-3_72.
[7] S. Marginson, M. van der Wende, To Rank or To Be Ranked: The Impact of Global Rankings in Higher Education, Journal of Studies in International Education 11 (2016) 306–329. doi:10.1177/1028315307303544.
[8] G. Dimitrova, T. Dimitrova, Competitiveness of the universities: measurement capabilities, Trakia Journal of Sciences 15 (2017) 311–316. doi:10.15547/tjs.2017.s.01.055.
[9] P. Taylor, R. Braddock, International University Ranking Systems and the Idea of University Excellence, Journal of Higher Education Policy and Management 29 (2008) 245–260. doi:10.1080/13600800701457855.
[10] K. F. Tee, Suitability of performance indicators and benchmarking practices in UK universities, Benchmarking: An International Journal 23 (2016) 584–600. doi:10.1108/BIJ-07-2014-0069.
[11] N. V. Avralev, I. N. Efimova, University Rankings as a Tool to Enhance Competitiveness, Clustering and Transnational Governance of Higher Education in the Context of Globalization, Middle-East Journal of Scientific Research 16 (2013) 357–361. URL: https://www.idosi.org/mejsr/mejsr16(3)13/7.pdf.
[12] H. Chládková, R. Skýpalová, V. Blašková, Strengthening the university competitiveness in the Czech Republic, Tuning Journal for Higher Education 9 (2021) 127–155. doi:10.18543/tjhe-9(1)-2021pp127-155.
[13] R. Deem, K. H. Mok, L. Lucas, Transforming Higher Education in Whose Image? Exploring the Concept of the 'World-Class' University in Europe and Asia, Higher Education Policy 21 (2008) 83–97. doi:10.1057/palgrave.hep.8300179.
[14] V. Kukharenko, B. Shunevych, H. Kravtsov, Distance course examination, Educational Technology Quarterly 2022 (2022) 1–19. doi:10.55056/etq.4.
[15] R. Lukman, D. Krajnc, P. Glavič, University ranking using research, educational and environmental indicators, Journal of Cleaner Production 18 (2010) 619–628. doi:10.1016/j.jclepro.2009.09.015.
[16] H. Kucherova, Y. Honcharenko, D. Ocheretin, O. Bilska, Fuzzy logic model of usability of websites of higher education institutions in the context of digitalization of educational services, Neiro-Nechitki Tekhnolohii Modelyuvannya v Ekonomitsi 2021 (2021) 119–135. doi:10.33111/nfmte.2021.119.
[17] D. Lukianenko, O. Lukianenko, O. Mozghalli, I. Dvornyk, M. Oriekhov, Digital imperative of university activities transformation, Financial and credit activity: problems of theory and practice 4 (2020) 449–458. doi:10.18371/fcaptp.v4i35.222478.
[18] I. Sannikova, E. Prikhodko, A. Muhitdinov, Assessment of the universities impact on global competitiveness based on rankings, E3S Web Conferences 296 (2021) 08009. doi:10.1051/

e3sconf/202129608009.
[19] V. Satsyk, Global Competitiveness of Universities: Key Determinants and Strategies (International and Ukrainian cases), in: Int. Conf. “The Education and Science and their Role in Social and Industrial Progress of Society”, 12-15 June 2014, Kyiv, Ukraine, 2014. URL: https://www.academia.edu/10233950.
[20] O. Jurášková, M. Juříková, J. Kocourek, Innovation of Educational Process as a Factor of Enhancing Competitiveness, The Turkish Online Journal of Educational Technology (2015) 301–305. URL: http://www.tojet.net/special/2015_7_2.pdf.
[21] O. Jurášková, M. Juříková, J. Kocourek, Brand Building of a University as an Integral Part of the Educational Process, The Turkish Online Journal of Educational Technology (2015) 100–105. URL: http://tojet.net/special/2015_9_1.pdf.
[22] B. Stoimenova, Regional innovation systems and university competitiveness, International Journal of Innovation 7 (2019) 227–239. doi:10.5585/iji.v7i2.353.
[23] A. Teixeira, A. Oliveira, A. Daniel, M. Preto, G. Brás, C. Rodrigues, The Impact of Universities on Regional Competitiveness: A Review of the Main Theoretical and Methodological Approaches, in: A. Daniel, A. Teixeira, M. Preto (Eds.), Examining the Role of Entrepreneurial Universities in Regional Development, IGI Global, 2020, pp. 67–92. doi:10.4018/978-1-7998-0174-0.ch004.
[24] D. Lukianenko, I. Strelchenko, Neuromodeling of features of crisis contagion on financial markets between countries with different levels of economic development, Neiro-Nechitki Tekhnolohii Modelyuvannya v Ekonomitsi 2021 (2021) 136–163. doi:10.33111/nfmte.2021.136.
[25] J. A. Hartigan, M. A. Wong, Algorithm AS 136: A K-Means Clustering Algorithm, Journal of the Royal Statistical Society. Series C (Applied Statistics) 28 (1979) 100–108. doi:10.2307/2346830.
[26] L. Kaufman, P. J. Rousseeuw, Partitioning Around Medoids (Program PAM), Wiley Series in Probability and Statistics, John Wiley & Sons, Ltd, 1990, pp. 68–125. doi:10.1002/9780470316801.ch2.
[27] I. T. Jolliffe, Principal Component Analysis, Springer Series in Statistics, 2nd ed., Springer, 2002. URL: https://link.springer.com/book/10.1007/b98835.
[28] U. Von Luxburg, A tutorial on spectral clustering, Statistics and computing 17 (2007) 395–416. doi:10.1007/s11222-007-9033-z.
[29] R. Sokal, F. Rohlf, The comparison of dendrograms by objective methods, Taxon 11 (1962) 33–40. doi:10.2307/1217208.
[30] T. Caliński, J. Harabasz, A dendrite method for cluster analysis, Communications in Statistics 3 (1974) 1–27. doi:10.1080/03610927408827101.
[31] T. Kohonen, Self-organized formation of topologically correct feature maps, Biological Cybernetics 43 (1982) 59–69. doi:10.1007/BF00337288.
[32] T. Kohonen, Self-organizing maps, volume 30 of *Springer Series in Information Sciences*, 3 ed., Springer Berlin, Heidelberg, 2001. doi:10.1007/978-3-642-56927-2.
[33] E. Schubert, J. Sander, M. Ester, H. Kriegel, X. Xu, DBSCAN revisited, revisited: Why and how you should (still) use DBSCAN, ACM Transactions on Database Systems 42 (2017) 1–21. doi:10.1145/3068335.
[34] R. J. G. B. Campello, D. Moulavi, J. Sander, Density-Based Clustering Based on Hierarchical

Density Estimates, in: J. Pei, V. S. Tseng, L. Cao, H. Motoda, G. Xu (Eds.), Advances in Knowledge Discovery and Data Mining, Springer Berlin Heidelberg, Berlin, Heidelberg, 2013, pp. 160–172. doi:10.1007/978-3-642-37456-2_14.

[35] M. Ankerst, M. M. Breunig, H.-P. Kriegel, J. Sander, OPTICS: Ordering Points to Identify the Clustering Structure, in: Proceedings of the 1999 ACM SIGMOD International Conference on Management of Data, SIGMOD '99, Association for Computing Machinery, New York, NY, USA, 1999, p. 49–60. doi:10.1145/304182.304187.

[36] L. McInnes, J. Healy, UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction, 2018. doi:10.48550/arXiv.1802.03426.

[37] T. Zhang, R. Ramakrishnan, M. Livny, BIRCH: an efficient data clustering method for very large databases, in: Proceedings of the 1996 ACM SIGMOD international conference on Management of data, ACM, 1996, p. 103–114. doi:10.1145/233269.233324.

[38] V. Kobets, O. Novak, EU countries clustering for the state of food security using machine learning techniques, Neiro-Nechitki Tekhnolohii Modelyuvannya v Ekonomitsi 2021 (2021) 86–118. doi:10.33111/nfmte.2021.086.

[39] V. Kobets, V. Yatsenko, Influence of the Fourth industrial revolution on divergence and convergence of economic inequality for various countries, Neiro-Nechitki Tekhnolohii Modelyuvannya v Ekonomitsi 2019 (2019) 124–146. doi:10.33111/nfmte.2019.124.

[40] A. Subasi, Clustering examples, in: Practical Machine Learning for Data Analysis Using Python, Elsevier, 2020, pp. 465–511. doi:10.1016/b978-0-12-821379-7.00007-2.

[41] H. Velykoivanenko, V. Korchynskyi, Application of Clustering in the Dimensionality Reduction Algorithms for Separation of Financial Status of Commercial Banks in Ukraine, Universal Journal of Accounting and Finance 10 (2022) 148–160. doi:10.13189/ujaf.2022.100116.

[42] A. Matviychuk, O. Lukianenko, I. Miroshnychenko, Neuro-fuzzy model of country's investment potential assessment, Fuzzy economic review 24 (2019) 65–88. doi:10.25102/fer.2019.02.04.

[43] The World Bank, World Development Indicators, 2022. URL: https://data.worldbank.org/indicator/SE.XPD.TOTL.GD.ZS.

[44] United Nations Development Programme, Human Development Reports, 2022. URL: http://hdr.undp.org/en/content/download-data.

[45] World Economic Forum, Global Competitiveness Index 4.0, 2019. URL: http://reports.weforum.org/global-competitiveness-report-2019/competitiveness-rankings/.

[46] European Commission, Statistical Audits, 2022. URL: https://knowledge4policy.ec.europa.eu/composite-indicators/statistical-audits_en.

[47] B. Lanvin, F. Monteiro, Global Talent Competitiveness Index 2019, INSEAD, 2019. URL: https://nonews.co/wp-content/uploads/2019/06/GTCI2019.pdf.

[48] UNIVERSITAS 21, Rankings Raw Data 2020, 2021. URL: https://universitas21.com/media/966.