

ISSN 2587-814X (print), ISSN 2587-8158 (online)

Russian version: ISSN 1998-0663 (print), ISSN 2587-8166 (online)

# BUSINESS INFORMATICS

**HSE SCIENTIFIC JOURNAL**

## CONTENTS

*E.A. Korobov, Y.V. Semernina,  
A.S. Usmanova, K.A. Odinokova*

Robotizing bond portfolio selection on the Russian  
debt market on the basis of a modified strategy  
of riding the yield curve.....7

*G.N. Zhukova, M.V. Ulyanov*

To the question of restoring symbol sequences  
encoding noisy periodic functions .....22

*G.A. Kalugina, A.V. Ryapukhin*

Methods of digital marketing positioning  
in the global civil passenger aircraft market.....36

*A.V. Sigal*

Using Fishburne's sequences in suitable  
modeling used for sample data .....50

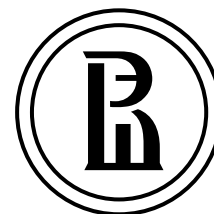
*E.D. Kopnova, L.A. Rodionova*

Modeling the links between institutional  
and actual globalization in the countries  
of the world .....61

*Yu.F. Telnov, V.A. Kazakov, A.V. Danilov*

Technology for designing innovative processes  
for creating products and services of a network  
enterprise using an i4.0 knowledge-based system .....76

**Vol. 15 No 4 – 2021**



Publisher:  
National Research University  
Higher School of Economics

The journal is published quarterly

The journal is included  
into the list of peer reviewed  
scientific editions established  
by the Supreme Certification  
Commission of the Russian Federation

Deputy Editor-in-Chief  
**E. Zaramenskikh**

Computer Making-up:  
**O. Bogdanovich**

Website Administration:  
**I. Khrustaleva**

Address:  
26-28, Shabolovka Street,  
Moscow 119049, Russia

Tel./fax: +7 (495) 772-9590 \*28509  
<http://bijournal.hse.ru>  
E-mail: [bijournal@hse.ru](mailto:bijournal@hse.ru)

Circulation:  
English version – 150 copies,  
Russian version – 150 copies,  
online versions in English and Russian –  
open access

Printed in HSE Printing House  
44, build.2, Izmaylovskoye Shosse,  
Moscow, Russia

© National Research University  
Higher School of Economics

# ABOUT THE JOURNAL

**B**usiness Informatics is a peer reviewed interdisciplinary academic journal published since 2007 by National Research University Higher School of Economics (HSE), Moscow, Russian Federation. The journal is administered by HSE Graduate School of Business. The journal is published quarterly.

The mission of the journal is to develop business informatics as a new field within both information technologies and management. It provides dissemination of latest technical and methodological developments, promotes new competences and provides a framework for discussion in the field of application of modern IT solutions in business, management and economics.

The journal publishes papers in the areas of, but not limited to: modeling of social and economic systems, digital transformation of business, innovation management, information systems and technologies in business, data analysis and business intelligence systems, mathematical methods and algorithms of business informatics, business processes modeling and analysis, decision support in management.

The journal is included into the list of peer reviewed scientific editions established by the Supreme Certification Commission of the Russian Federation.

The journal is included into Scopus, Web of Science Emerging Sources Citation Index (WoS ESCI), Russian Science Citation Index on the Web of Science platform (RSCI), EBSCO.

International Standard Serial Number (ISSN): 2587-814X (in English), 1998-0663 (in Russian).

# EDITORIAL BOARD

## DEPUTY EDITOR-IN-CHIEF

### **Evgeny P. Zaramenskikh**

National Research University Higher School of Economics,  
Moscow, Russia

## EDITORIAL BOARD

### **Habib Abdulrab**

National Institute of Applied Sciences, Rouen, France

### **Sergey M. Avdoshin**

National Research University Higher School of Economics,  
Moscow, Russia

### **Andranik S. Akopov**

National Research University Higher School of Economics,  
Moscow, Russia

### **Fuad T. Aleskerov**

National Research University Higher School of Economics,  
Moscow, Russia

### **Alexander P. Afanasyev**

Institute for Information Transmission Problems (Kharkevich  
Institute), Russian Academy of Sciences, Moscow, Russia

### **Anton A. Afanasyev**

Central Economics and Mathematics Institute, Russian Academy  
of Sciences, Moscow, Russia

### **Eduard A. Babkin**

National Research University Higher School of Economics,  
Nizhny Novgorod, Russia

### **Sergey I. Balandin**

Finnish-Russian University Cooperation in Telecommunications  
(FRUCT), Helsinki, Finland

### **Vladimir B. Barakhnin**

Federal Research Center of Information and Computational  
Technologies, Novosibirsk, Russia

### **Alexander P. Baranov**

Federal Tax Service, Moscow, Russia

### **Jorg Becker**

University of Munster, Munster, Germany

### **Vladimir V. Belov**

Ryazan State Radio Engineering University, Ryazan, Russia

### **Alexander G. Chkhartishvili**

V.A. Trapeznikov Institute of Control Sciences, Russian Academy  
of Sciences, Moscow, Russia

### **Vladimir A. Efimushkin**

Central Research Institute of Communications, Moscow, Russia

### **Tatiana A. Gavrilova**

Saint-Petersburg University, St. Petersburg, Russia

### **Hervé Glotin**

University of Toulon, La Garde, France

### **Alexey O. Golosov**

FORS Development Center, Moscow, Russia

### **Andrey Yu. Gribov**

CyberPlat Company, Moscow, Russia

### **Alexander I. Gromoff**

National Research University Higher School of Economics,  
Moscow, Russia

### **Vladimir A. Gurvich**

Rutgers, The State University of New Jersey, Rutgers, USA

### **Laurence Jacobs**

University of Zurich, Zurich, Switzerland

### **Liliya A. Demidova**

Ryazan State Radio Engineering University, Ryazan, Russia

### **Iosif E. Diskin**

Russian Public Opinion Research Center, Moscow, Russia

### **Nikolay I. Ilyin**

Federal Security Guard of the Russian Federation,  
Moscow, Russia

### **Dmitry V. Isaev**

National Research University Higher School of Economics,  
Moscow, Russia

### **Alexander D. Ivannikov**

Institute for Design Problems in Microelectronics, Russian Academy  
of Sciences, Moscow, Russia

### **Valery A. Kalyagin**

National Research University Higher School of Economics,  
Nizhny Novgorod, Russia

### **Tatiana K. Kravchenko**

National Research University Higher School of Economics,  
Moscow, Russia

### **Sergei O. Kuznetsov**

National Research University Higher School of Economics,  
Moscow, Russia

### **Kwei-Jay Lin**

Nagoya Institute of Technology, Nagoya, Japan

### **Mikhail I. Lugachev**

Lomonosov Moscow State University, Moscow, Russia

### **Svetlana V. Maltseva**

National Research University Higher School of Economics,  
Moscow, Russia

### **Peter Major**

UN Commission on Science and Technology for Development,  
Geneva, Switzerland

### **Boris G. Mirkin**

National Research University Higher School of Economics,  
Moscow, Russia

### **Vadim V. Mottl**

Tula State University, Tula, Russia

### **Dmitry M. Nazarov**

Ural State University of Economics, Ekaterinburg, Russia

### **Dmitry E. Palchunov**

Novosibirsk State University, Novosibirsk, Russia

### **Panagote (Panos) M. Pardalos**

University of Florida, Gainesville, USA

### **Óscar Pastor**

Polytechnic University of Valencia, Valencia, Spain

### **Joachim Posegga**

University of Passau, Passau, Germany

### **Konstantin E. Samouylov**

Peoples' Friendship University, Moscow, Russia

### **Kurt Sandkuhl**

University of Rostock, Rostock, Germany

### **Christine Strauss**

University of Vienna, Vienna, Austria

### **Ali R. Sunyaev**

Karlsruhe Institute of Technology, Karlsruhe, Germany

### **Victor V. Taratukhin**

University of Munster, Munster, Germany

### **José M. Tribolet**

Universidade de Lisboa, Lisbon, Portugal

### **Olga A. Tsukanova**

Saint-Petersburg National Research University of Information  
Technologies, Mechanics and Optics, St. Petersburg, Russia

### **Mikhail V. Ulyanov**

V.A. Trapeznikov Institute of Control Sciences, Russian Academy  
of Sciences, Moscow, Russia

### **Raissa K. Uskenbayeva**

International Information Technology University, Almaty, Kazakhstan

### **Markus Westner**

Regensburg University of Applied Sciences, Regensburg, Germany

# ABOUT THE HIGHER SCHOOL OF ECONOMICS

Consistently ranked as one of Russia's top universities, the Higher School of Economics (HSE) is a leader in Russian education and one of the preeminent economics and social sciences universities in Eastern Europe and Eurasia.

Having rapidly grown into a well-renowned research university over two decades, HSE sets itself apart with its international presence and cooperation.

Our faculty, researchers, and students represent over 50 countries, and are dedicated to maintaining the highest academic standards. Our newly adopted structural reforms support

both HSE's drive to internationalize and the groundbreaking research of our faculty, researchers, and students.

Now a dynamic university with four campuses, HSE is a leader in combining Russian educational traditions with the best international teaching and research practices. HSE offers outstanding educational programs from secondary school to doctoral studies, with top departments and research centers in a number of international fields.

Since 2013, HSE has been a member of the 5-100 Russian Academic Excellence Project, a highly selective government program aimed at boosting the international competitiveness of Russian universities.

# ABOUT THE GRADUATE SCHOOL OF BUSINESS

**H**SE Graduate School of Business was created on September 1, 2020. The School will become a priority partner for leading Russian companies in the development of their personnel and management technologies.

The world-leading model of a ‘university business school’ has been chosen for the Graduate School of Business. This foresees an integrated portfolio of programmes, ranging from Bachelor’s to EMBA programmes, communities of experts and a vast network of research centres and laboratories for advanced management studies. Furthermore, HSE University’s integrative approach will allow the Graduate School of Business to develop as an interdisciplinary institution. The advancement of the Graduate School of Business through synergies with other faculties and institutes will serve as a key source of its competitive advantage. Moreover, the evolution and development of the Business School’s faculty involves the active engagement of three professional tracks at our University: research, practice-oriented and methodological.

What sets the Graduate School of Business apart is its focus on educating and developing globally competitive and socially responsible business leaders for Russia’s emerging digital economy.

The School’s educational model will focus on a project approach and other dynamic methods for skills training, integration of online and other digital technologies, as well as systematic internationalization of educational processes.

At its start, the Graduate School of Business will offer 22 Bachelor programmes (three of which will be fully taught in English) and over 200 retraining and continuing professional development programmes, serving over 9,000 students. In future, the integrated portfolio of academic and professional programmes will continue to expand with a particular emphasis on graduate programmes, which is in line with the principles guiding top business schools around the world. In addition, the School’s top quality and all-encompassing Bachelor degrees will continue to make valuable contributions to the achievement of the Business School’s goals and the development of its business model.

The School’s plans include the establishment of a National Resource Center, which will offer case studies based on the experience of Russian companies. In addition, the Business School will assist in the provision of up-to-date management training at other Russian universities. Furthermore, the Graduate School of Business will become one of the leaders in promoting Russian education.

The Graduate School of Business’s unique ecosystem will be created through partnerships with leading global business schools, as well as in-depth cooperation with firms and companies during the entire life cycle of the school’s programmes. The success criteria for the Business School include professional recognition thanks to the stellar careers of its graduates, its international programmes and institutional accreditations, as well as its presence on global business school rankings.



DOI: [10.17323/2587-814X.2021.4.7.21](https://doi.org/10.17323/2587-814X.2021.4.7.21)

# Robotizing bond portfolio selection on the Russian debt market on the basis of a modified strategy of riding the yield curve\*

**Eugene A. Korobov<sup>a</sup>** 

E-mail: korobovea@yandex.ru

**Yulia V. Semernina<sup>b</sup>** 

E-mail: ysemernina@yandex.ru

**Alina S. Usmanova<sup>b</sup>** 

E-mail: alina.brosalova@yandex.ru

**Kristina A. Odinkova<sup>b</sup>** 

E-mail: odinkova.kristina93@mail.ru

<sup>a</sup> Saratov State University

Address: 83, Astrakhanskaya Street, Saratov 410012, Russia

<sup>b</sup> Yuri Gagarin State Technical University of Saratov

Address: 77, Polytechnicheskaya Street, Saratov 410054, Russia

## Abstract

The modern global debt market features historically low average interest rates, convergence of yields on bonds with different maturities, an increase of yield curve inversion emergence frequency and a large-scale trend to automate financial decision making. The researchers' attention in these fields is mainly focused on designing models that describe the state of the debt market as whole or its individual instruments in particular, as well as on risk management methods. At the same time, the specialized literature offers very few works concerning the topic of computer algorithms for bond portfolio selection based on traditional or advanced investment strategies. The aim of the present research is to create a modification of the existing algorithm of riding the yield curve strategy application, employing, first, average bond yield over the holding period instead of traditional bond yield to maturity; second, a developed algorithm for calculating the market spread on bonds; and, third, alternative risk evaluation indicators (compensation coefficients), which allow us to measure

\* The article is published with the support of the HSE University Partnership Programme

objectively price risk, liquidity risk, transaction costs risk and a general risk. The modification and the development of the algorithm for calculating the market spread were carried out using the direct measurement of the result technique, which entails application of the strategy to the data on bond issues received through the Moscow Exchange API. The selection of financial instruments was conducted in all sectors of the Russian debt market: public bonds, sub-federal and municipal bonds, corporate bonds. The modified algorithm enabled us to get extra yield for each selected bond issue, thereby proving the high effectiveness of the technique compared to the traditional strategy. Software implementation of the algorithm can be integrated into any robotized or semi-robotized stock exchange trading application.

**Key words:** modified riding the yield curve strategy; robotizing bond portfolio selection; algorithm for calculating the market spread on bonds; compensation coefficients; bond price risk; bond liquidity risk; bond transaction costs risk; general bond risk; investment software applications.

**Citation:** Korobov E.A., Semernina Y.V., Usmanova A.S., Odinkova K.A. (2021) Robotizing bond portfolio selection on the Russian debt market on the basis of a modified strategy of riding the yield curve. *Business Informatics*, vol. 15, no 4, pp. 7–21. DOI: 10.17323/2587-814X.2021.4.7.21

## Introduction

The development of financial technologies and financial markets in recent decades has led to the fact that today almost all decisions on purchase or sale of financial instruments are carried out through specialized electronic applications. Their software architecture is largely determined by the financial tasks that a market actor sets for himself. Speculative strategies such as scalping, intraday or swing trading are implemented today mainly by creating trading robots for high-frequency trading. Human participation in making a decision to buy or sell a particular financial instrument, in this case, although it may be present, is usually minimized.

Following the investment strategies today involves engaging three types of software:

- ◆ fully robotized applications that exclude human participation (the first type);
- ◆ semi-robotized complexes, when the program forms investment options, and a per-

son decides on the acceptance of one or another option (the second type);

- ◆ non-robotized systems in which all financial analysis and decision making is a matter of investor's responsibility (the third type).

All three types of investment applications have been actively developing in recent years. The first type of software can be classified as robo-advisors that have been gaining popularity after the financial crisis of 2008. They use artificial neural networks and graph models of decision making. Algorithms of their work [1–4], based on the classical principles of portfolio selection by H. Markowitz, W. Sharpe and J. Tobin, showed high effectiveness in a growing market during the period from February 2009 to February 2020. A “Black Swan” in the form of a collapse of stock indices in March 2020 led to the case when many FinTech providers had to switch the work of their robo-advisors to manual control mode. This once again indicated the limitations of artificial intelligence system applicability for making financial decisions.



Semi-robotized investment applications can include trading systems in which an investor makes transactions based on data received from an integrated screener working with such instruments as stocks, mutual funds, ETFs, bonds, etc. The “intelligence” of such screeners usually comes down to calculating the main economic indicators of a financial instrument or an issuing organization and filtering them according to a given criterion.

Non-robotized applications comprise most of the trading terminals provided by stock brokers and banks to their clients for making transactions in the financial market.

All of the above trends are to be taken into account, first of all, by such institutional investors as mutual funds, as well as more conservative market actors — commercial banks, public and private pension funds, insurance companies and some private investors who face the task of creating low-risk portfolios. The majority of investments in such portfolios are usually government and corporate bonds. One of the possible universal strategies for selecting financial instruments in the debt market is the riding the yield curve strategy (RYCS).

Modern Russian academic literature lacks works analyzing active investment strategies in the bond market; moreover, the modifications of such strategies are practically absent as well. The most common approach to bond investing is the “buy and hold” strategy, which can be traced in the papers by N.F. Korobov and A.A. Beloglazov [5], A.A. Ponomarev [6], N.V. Popova [7], P.A. Lashevsky [8], I.A. Darushin [9], A.N. Zadorozhnaya [10], V.V. Nazarova and I.P. Levichev [11], I.A. Freidina [12], M.V. Fomina, Yu.V. Lakhno and A.P. Pyshnograya [13] and other authors.

Some researchers study the shape of the bond yield curve, but they do not consider its application to RYCS (for example, a similar approach is a characteristic of the papers by

A.Y. Mikhailov [14], A.N. Burenin [15], S.V. Beshenov and V.A. Lapshin [16]).

RYCS is practically not considered in the specialized Russian financial and economic literature, and in foreign literature its most complete description is presented in the fundamental works by F.J. Fabozzi [17–19]. A detailed consideration of this strategy with an analysis of its practical use also can be found in the papers by V. Galvani and S. Landon [20], D. Bieri and L. Chincarini [21], R. Cox and J. Felton [22], Z. Wei, W. Xianhua and W. Guofu [23].

An effectiveness estimation of RYCS under the conditions of low interest rates is presented in a study by the International Monetary Fund [24].

Computer models of bond portfolio selection and evaluation are also presented rather sparsely in the works of Russian and foreign researchers. Russian papers on this topic can be found in [25, 26].

The choice of RYCS as an object of robotization is due to a combination of two factors. On the one hand, under the conditions of low interest rates and the inverted yield curve shape, this strategy (with sufficient credit quality of bond issuers) has a certain “resistance” to the risks associated with the debt market, primarily to the interest rate growth risk; the probability of its realization today is objectively quite high (especially when applying the strategy on a short section of the yield curve). On the other hand, RYCS belongs to the group of well-formalized investing strategies, which allows us to form unambiguous algorithms for robo-trading. In addition, RYCS is available not only to institutional, but also to private investors, whose distinctive feature is the insignificant volume of the portfolio, a priori assuming a limited level of its diversification.

The initial condition determining the possibility of using this strategy is the presence of a

positive difference between the yield of bonds with a longer maturity and the yield of bonds with a shorter maturity (in fact, we are talking about the presence of a “classic” upward slope of the bond yield curve). As an additional condition for its application, it is necessary to note a sufficient level of market liquidity of the bond issue and an appropriate level of transaction costs.

Currently, in the context of a global decline in the average level of interest rates, bond yields with different maturities are “converging” (minimizing the difference in bond yields on different sections of the curve), and the frequency of the bond yield curve inversion is significantly increasing. These trends lead to the need to transform the approach to considering strategy: its use in the classical interpretation makes investors accept increased interest rate risk, liquidity risk and transaction cost risk, while the resulting yield generally does not compensate for the accepted risks.

The purpose of this study is to design a modification of the traditional RYCS, taking into account the interest rate curve shape, as well as the development of an automated algorithm for calculating the market spread in the Russian debt market which allows investors to evaluate the liquidity risk. The algorithms thus obtained can be engaged by exchange trading software of the first and second types.

## 1. Methods

The modification of RYCS and the development of an automated algorithm for calculating the market spread is supposed to be carried out using the method of the result direct measurement. This method involves the application of a modified RYCS to real data (based on information received from the Russian exchange bond market at a specific time – a “market snapshot” for a specific date) and does not involve the design of predictive, descriptive or simulation models for the Russian bond

market as a whole or of its individual sectors in particular.

This method is based on the following assumptions:

- 1) the costs of investors in the purchase and sale of bonds include two types of commissions – the exchange commission and the broker’s commission, charged as a percentage of the amount of actually completed transactions, while other types of costs are not considered;
- 2) bonds are selected for the investment portfolio during one trading session;
- 3) the closing prices of the corresponding trading session are considered as settlement prices for the purchase and / or sale of bonds, regardless of the volume of transactions performed, while in the absence of real transactions during the trading session, the bond issue is deemed not to have a sufficient level of liquidity and is not selected to the bond portfolio;
- 4) when calculating the yield indicators for bond issues, the accumulated coupon income at the time of the relevant transaction is taken into account;
- 5) the credit quality of bond issuers is assumed to be unchanged throughout the analyzed period (this study doesn’t separately assess or analyze it).

Considering modification of RYCS is based on following points:

- 1) the average yield of bonds over the investment period instead of the traditional yield of bonds to maturity;
- 2) the designed author’s algorithm for calculating the market spread on bonds;
- 3) alternative risk assessment indicators (compensation coefficients) that make possible an objective assessment of price risk, liquidity risk and transaction cost risk, as well as overall risk.

The quantitative indicators proposed within the modified strategy are universal for the debt market and can be applied to any coupon and discount bonds. The generalizing nature of the indicators used is due to the chosen technique of their design: when calculating them, fixed parameters of bond issues are used, as a rule, determined at the moment of issue of the latter, data on market transactions with bonds and their market quotations, as well as information on the time bonds are held (investment period).

When using RYCS within the traditional approach, it is assumed that the classical formula is applied to calculate the yield of bonds to maturity:

$$Y_M = \frac{(P_S - P_P)}{P_P} \cdot \frac{365}{t_M} \cdot 100\%, \quad (1)$$

where  $Y_M$  – bond yield to maturity;

$P_S$  – sale price or redemption price of bonds;

$P_P$  – bond purchase price;

$t_M$  – the number of days remaining until the bonds are redeemed.

It should be emphasized that for the classical version of the strategy, strict inequality is mandatory:

$$Y_{LT} > Y_{ST}, \quad (2)$$

where  $Y_{LT}$  – is the yield to maturity of bonds with a long maturity;

$Y_{ST}$  – yield to maturity of bonds with a shorter maturity.

In addition, another strict inequality must also be observed:

$$Y_M > Y_E, \quad (3)$$

where  $Y_E$  – expected return on investment.

Within the consideration modification, it is proposed to focus on an alternative indi-

cator – the yield for the period of holding the bond, which is due to the availability of an objective opportunity for investors to sell bonds without waiting for their redemption. The calculation of this indicator can be done according to the following formula:

$$Y_{t_H} = \frac{(P_S + C_S + \sum_{i=1}^n C_{R_i} - P_P - C_P - \sum_{j=1}^k T_j)}{P_P + C_P} \times \quad (4)$$

$$\times \frac{365}{t_H} \times 100\%,$$

where  $Y_{t_H}$  – yield for the period of holding the bond;

$C_S$  – is accumulated coupon yield at the time of bond sale;

$C_{R_i}$  – is the coupon yield paid by the issuer for the corresponding coupon period (subject to availability of such payments for the holding period);

$n$  – is total number of coupon periods for bonds;

$C_P$  – is accumulated coupon yield at the time of bond purchase;

$T_j$  – transaction costs;

$k$  – total number of transaction cost types;

$t_H$  – number of days the bond is held.

Unlike the traditional technique, the yield indicator for the period of holding the bond should be calculated regularly, for example, at the end of each trading session, and not only at the time of bond purchase. This approach initially implies a higher level of investor trading activity compared to the traditional one.

To assess the effectiveness of the modified RYCS, it is advisable to use the average yield indicator for the investment period, taking into account both the periods of bond holding and the periods of no positions in bonds:

$$\bar{Y} = \frac{\sum_{l=1}^m Y_{t_{H_l}} t_{H_l} + \sum_{q=1}^v Y_{t_{0_q}} t_{0_q}}{\sum_{l=1}^m t_{H_l} + \sum_{q=1}^v t_{0_q}}, \quad (5)$$

where  $\bar{Y}$  – average yield over the investment period;

$Y_{t_{H_l}}$  – yield for the  $l$ -period of bond holding (the period of availability of positions on bonds);

$m$  – number of bond holding periods;

$Y_{t_{0_q}}$  – yield for the period of no positions in bonds (for example, due to the placement of funds on the money market);

$t_{0_q}$  – duration of the period with no positions in bonds;

$v$  – number of periods with no positions in bonds.

Conceptually, it is important to note that for a modified RYCS, the ratio between bond yields of different durations takes the form:

$$Y_{LT} \geq Y_{ST}. \quad (6)$$

Thus, the proposed modified strategy is effective even when the bond yield curve is flat.

In turn, to evaluate the price risk, it is suggested to use the change in the bond market price with an increase in interest rates by 1%:

$$\Delta_+ = |P_{S_{Y+1\%}} - P_P|, \quad (7)$$

where  $\Delta_+$  is the change in the bond market price with an increase in interest rates by 1%;

$P_{S_{Y+1\%}}$  – the bond market price with an increase in interest rates by 1%.

The value  $P_{S_{Y+1\%}}$  is calculated according to formula (4) by selecting such a numerical value of this indicator, at which the value of the bond yield will increase by exactly 1% ( $Y_{M+1\%}$ ). This makes it possible to obtain accurate results (in absolute terms), which is the

main advantage of this indicator compared, for example, with classical or modified duration, which is less accurate. In general, it assumes an approximate calculation of the coupon rate; it does not take into account the daily accrual of accumulated coupon income; does not consider the payment of the latter at the end of the coupon period or pay back by the buyer at the time of bond sale. Duration also assumes complete symmetry of changes in bond prices both in terms of growth and in terms of declining interest rates (the last problem can be solved to a certain extent by using the convexity index of bonds; however, when liquidity is limited, this indicator has limited representativeness).

As a risk indicator that reckons the accrual of current bond yield, we proposed to use the author's price risk compensation coefficient, determined by the formula:

$$K_{CP} = \frac{\Delta_+}{C_D}, \quad (8)$$

where  $K_{CP}$  is price risk compensation coefficient;

$C_D$  – the value of the coupon charged daily.

Thus, the value of the price risk compensation coefficient can be interpreted as the number of days for which the value of the coupon charged on bonds fully compensates for the estimated drop in the bond price (from their current market prices) with an increase in the interest rate average level by 1%.

Another indicator that makes it possible to assess the risk taken by investors in the debt market is the liquidity risk compensation coefficient calculated by the formula:

$$K_{CL} = \frac{\bar{S}}{C_D}, \quad (9)$$

where  $K_{CL}$  is liquidity risk compensation coefficient;

$\bar{S}$  – average value of the market spread for a bond issue.

We propose to interpret the value of the liquidity risk compensation coefficient as the number of days for which the value of the coupon charged on bonds fully compensates for the average market spread on bonds.

The average value of the market spread (in absolute or relative terms) for a certain period of time objectively reflects the level of liquidity of bonds and is more representative compared to the traditionally used spread at a specific time (usually at the time of opening a position in bonds). The fact is that the average value of the market spread considers every change in bond market quotations (both purchase and sale quotations) during market trading; that is, it reflects the average actually observed difference between them. The absolute advantage of this indicator is the possibility of its systematic calculation, especially if the average spread is calculated over a sufficiently long time interval (for example, a month).

The traditionally used market spread calculated at a specific time has two serious drawbacks: firstly, it cannot always be determined (in particular, in the complete absence of quotations for the purchase or sale of bonds), and secondly, the market spread value at a specific time can deviate significantly from its average values in any direction.

In addition, the classical version of calculating the market spread in absolute terms usually does not reckon the volume of the best bid and ask bond orders, i.e. the spread is “mechanically” calculated as the difference between the best ask price and the best bid price according to the following formula:

$$S_M = B_S - B_P, \quad (10)$$

where  $S_M$  – market spread on a bond issue at a specific time;

$B_S$  – the best ask price on a bond issue at a specific time;

$B_P$  – the best bid price on a bond issue at a specific time.

This technique has gained widespread use because, in practice, the correct calculation of the average market spread “manually” is an extremely difficult and time-consuming task. Its solution involves regular depth of market (order book) monitoring for each bond issue, fixing the changes occurring in it and calculating the market spread for each bond issue at a specific time. When forming a diversified bond portfolio, this task is objectively impossible without the use of automated algorithms.

In this study, an algorithm for calculating the market spread on bond issues was designed and programmed in Lua in integration with the QUIK trading terminal (“Quickly Updatable Information Kit”) and assumes the following steps:

- 1) the selection of bond issues in user mode (at this stage, the investor can choose an arbitrary list of bond issues whose issuers satisfy him in terms of credit quality; other criteria for the formation of this list can be set arbitrarily);
- 2) setting up the period for fixing market quotations of selected bond issues (by default, the full duration of the main trading session is set, if necessary, it can be changed arbitrarily);
- 3) setting up conditions for fixing market quotations of selected bonds (default fixing is executed only if there are simultaneous market quotations for the buy and sell; other possible conditions of fixing: if there are market quotations for the buy, if there are market quotations for the sell, permanent fixing);
- 4) recording the market quotations to the information system (carried out in a specialized register in the database);
- 5) setting of the period for calculating the market spread (by default it is 20 trading days; if necessary, it can be set arbitrarily);



- 6) calculation of the average market spread on the bond for the chosen period, carried out according to the formula:

$$\bar{S} = \frac{\sum_{l=1}^w S_{M_l} t_l}{\sum_{l=1}^w t_l}, \quad (11)$$

where  $w$  – the number of price stability periods of bond market quotations (during this period, the value of the market spread remains unchanged);

$S_{M_l}$  – the market spread value during the  $l$ -period of price stability of bond market quotations;

$t_l$  – duration of the price stability period of bond market quotations (calculated with an accuracy of 1 second);

- 7) the output of calculated values of the average market spread to the report (the report form is configurable in user mode).

The use of the single transaction criterion in combination with the developed algorithm allows for a much more accurate assessment of the bond liquidity risk characteristic when using RYCS.

From the automation point of view, the main advantage of the proposed algorithm for fixing and calculation of the market spread on bonds is the regular receipt of objective information about the average actual (actually observed) market spread value on circulating bonds and, if necessary, collecting an array of spread historical data, which allows us much more accurately to characterize the liquidity level of bonds considered as potential investment instruments.

An additional indicator that reflects the level of risk taken by investors is the transaction cost compensation coefficient  $K_{CT}$ :

$$K_{CT} = \frac{\sum_{j=1}^k T_j}{C_D}. \quad (12)$$

The economic interpretation of this indicator is largely similar to the previous ones: it shows the number of days for which the value of the coupon charged on the bonds fully compensates for the amount of transaction costs incurred by the investor when making transactions on the bond market.

Based on the above indicators, you can calculate the overall compensation factor  $K_{CG}$ :

$$K_{CG} = K_{CP} + K_{CL} + K_{CT}. \quad (13)$$

This indicator allows us to get an idea of over what period (number of days) the value of the coupon charged on bonds fully compensates for the price and liquidity risk, as well as the amount of transaction costs. In general, the higher the value of all the above compensation ratios, the higher the level of risk taken by investors, and vice versa.

Obviously, within modified RYCS, maximization of the average yield over the investment period and minimization of the overall compensation coefficient are optimal.

## 2. Results

The method of the result direct measurement in relation to the modified RYCS was carried out in the Russian exchange bond market as of June 30, 2020 (the date of opening bond positions).

The bond portfolio was formed with borrowed funds attracted at a rate of 7.00% per annum for 90 days, while there is a possibility, if necessary, to prolong the loan (the prolongation is carried out on identical terms in terms of cost and terms of attraction).

Bond issues with maturity dates within 13 to 27 December 2020 are considered as investment objects. Bond issues are selected in all segments of the Russian bond market: government bonds, sub-federal and municipal bonds, corporate bonds (the list of bonds traded in the

Russian stock market and having the required maturity is given in *Table 1*).

The formed list of bond issues contains one issue of government bonds and 15 issues of corporate bonds (in the segment of sub-federal and municipal bonds during this period, bond issues were not redeemed).

The next consistently applied criterion for reducing the list of potential investment

objects is the criterion of liquidity, in particular, one that assumes the completion of at least one transaction during the trading session, in which the bond portfolio is formed (data on transactions made on June 30, 2020 in the domestic exchange market are presented in *Table 2*; bold font there marks those bond issues which have no transactions during the trading session).

*Table 1.*

**Bond issues circulating in the Russian stock market  
with maturity dates within 13 to 27 December 2020,  
placed no later than 30 June 2020<sup>1</sup>**

# / item	Bond market segment	Name of bond issue	Maturity date of the bond issue (dd.mm.yyyy)	Nominal value of the issue, billion rubles
1	Government bonds	GSO-36002-PPS	16.12.2020	15.00000
2	Corporate bonds	Ofir-KO-P02	14.12.2020	0.10000
3		DOM.RF-20-ob	15.12.2020	5.00000
4		Sberbank-001-41R-bso	15.12.2020	3.00000
5		FPK Garant-Invest-001R-03	16.12.2020	1.00000
6		Agronova-L-1-ob	17.12.2020	5.00000
7		RSG-Finans-4-bob	17.12.2020	2.00000
8		SoftLain Treid-001R-01	17.12.2020	3.00000
9		Bank VTB-B-1-48	18.12.2020	10.00000
10		UOMZ PO-2-bob	18.12.2020	1.50000
11		Passazhirsky port-1-obn	22.12.2020	0.00009
12		Bank VTB-B-1-48	23.12.2020	5.00000
13		Magnit-003R-03	24.12.2020	10.00000
14		ChTPZ-001R-01-bob	24.12.2020	5.00000
15		Sberbank-001-179R-bso	25.12.2020	3.00000
16		GPB-25-bob	26.12.2020	5.00000

<sup>1</sup> Bond Screener 'RusBonds'. FinMarket Information Agency ([https://www.rusbonds.ru/srch\\_simple.asp](https://www.rusbonds.ru/srch_simple.asp))

Table 2.

**The number of transactions concluded on the selected  
bond issues on June 30, 2020 on the Moscow Exchange  
(in the 'T + Market Transactions' trading mode)<sup>2</sup>**

# / item	Name of bond issue	Transaction volume, RUR	Transaction volume, units	Number of transactions, units
1	<b>GSO-36002-PPS</b>	<b>0</b>	<b>0</b>	<b>0</b>
2	<b>Ofir-KO-P02</b>	<b>0</b>	<b>0</b>	<b>0</b>
3	<b>DOM.RF-20-ob</b>	<b>0</b>	<b>0</b>	<b>0</b>
4	<b>Sberbank-001-41R-bso</b>	<b>0</b>	<b>0</b>	<b>0</b>
5	FPK Garant-Invest-001R-03	63526.60	63	10
6	<b>Agronova-L-1-ob</b>	<b>0</b>	<b>0</b>	<b>0</b>
7	RSG-Finans-4-bob	1438103.50	1439	50
8	SoftLain Treid-001R-01	683453.75	1336	35
9	Bank VTB-B-1-48	9276796.40	9229	27
10	UOMZ PO-2-bob	143036.00	141	5
11	<b>Passazhirsky port-1-obn</b>	<b>0</b>	<b>0</b>	<b>0</b>
12	<b>Bank VTB-B-1-48</b>	<b>0</b>	<b>0</b>	<b>0</b>
13	Magnit-003R-03	177337.70	174	13
14	ChTPZ-001R-01-bob	122784.00	120	2
15	Sberbank-001-179R-bso	4090890.60	4221	23
16	GPB-25-bob	1773540.40	1747	10

The liquidity criterion used made it possible to reduce the number of bond issues from 16 to 9, while the latter were used to calculate the yield for the expected holding period (subject to bond holding until maturity), based on the average broker's commission of 0.01% of the transaction amount and the exchange commis-

sion of 0.01% of the transaction amount (See the calculation results in *Table 3*).

Thus, 7 out of 9 selected bonds (they are marked in bold in *Table 3*) are not advisable to use within the modified RYCS: their estimated yield for the expected holding period, taking

<sup>2</sup> Moscow Exchange MICEX-RTS (<https://www.moex.com/ru/market-data/bulletins/#/date=2020-06-30>)



Table 3.

**Estimated yield over the expected holding period  
of the selected bonds (assuming bond holding until maturity)**

# / item	Name of bond issue	The yield for the expected holding period, % per annum
1	FPK Garant-Invest-001R-03	9.90
2	RSG-Finans-4-bob	10.94
3	<b>SoftLain Treid-001R-01</b>	<b>6.07</b>
4	<b>Bank VTB-B-1-48</b>	<b>4.47</b>
5	<b>UOMZ PO-2-bob</b>	<b>6.11</b>
6	<b>Magnit-003R-03</b>	<b>3.69</b>
7	<b>ChTPZ-001R-01-bob</b>	<b>5.12</b>
8	<b>Sberbank-001-179R-bso</b>	<b>6.65</b>
9	<b>GPB-25-bob</b>	<b>4.43</b>

Table 4.

**Estimated compensation coefficient values for selected bonds**

# / item	Name of bond issue	$K_{CP}$ , days	$K_{CL}$ , days	$K_{CT}$ , days	$K_{CG}$ , days
1	FPK Garant-Invest-001R-03	13.5	27.1	1.2	41.8
2	RSG-Finans-4-bob	14.7	10.6	1.3	26.6

into account the accumulated coupon yield and the transaction costs amount, is less than 7.00% per annum.

For two bond issues with a yield exceeding the cost of raising borrowed funds, compensation ratios (for the case of selling bond issues) were determined as of June 30, 2020.

The calculation of compensation coefficients is performed using classical mathematical and computer modeling methods: determination of the market bond price volatility with an increase in interest rates by 1% (using iterative

methods of value selection), as well as calculation and verification of the significance of the bond market spread average value for bond issues based on small samples, followed by verification by the Bartlett criterion (See the calculation results in Table 4).

For both bond issues, the overall compensation coefficient is less than the expected investment period, so the selected bond issues can be included in the bond portfolio as part of a modified RYCS. The relative share of each bond issue is determined in inverse proportion

to the overall compensation coefficient values, i.e. the share of the bond issue FPK Garant-Invest-001R-03 will be 38.9%, and the share of the issue RSG-Finance-4-bob – 61.1%.

Performance evaluation of the implemented investment strategy was carried out as of September 18, 2020.

The specified date was the date of the bond portfolio review and the adoption of a key investment decision to close the existing bond positions or to continue investment (the calculations are presented in *Table 5*).

The modified strategy made it possible to obtain additional yield on each of the selected bonds, and taking into account the share of bond issues, the average yield growth was 1.45%. Considering the expected yield of 7.00%, the difference in the return on investment is very significant: the increase in profitability amounted to 20.71% of the expected return.

It should be noted that in the future, the complete robotization of the sequence of actions described above when forming a bond investment portfolio will allow institutional and private investors to solve effectively a number of interrelated tasks with minimal time, in particular, the task of selecting potential invest-

ment instruments, the task of structuring an investment portfolio, as well as the task of choosing the optimal moment to open and close positions in bonds.

## Conclusion

Based on the study outcomes, the following conclusions can be drawn:

- 1) the modified riding the yield curve strategy applied to short-term bonds can be used both in the conditions of the classical interest rate curve shape and in the case of a flat interest rate curve;
- 2) the proposed yield indicator for the holding period, which takes into account the accumulation and coupon yield payment as well as the transaction costs that are individual for each investor, is a more accurate and objective indicator compared to the traditional indicator of bond yield to maturity;
- 3) the joint use of the single transaction criterion in combination with the algorithm for calculating the average market spread allows investors to more correctly evaluate the level of market liquidity of bonds and more accurately calculate the liquidity risk compensation coefficient;

*Table 5.*

**Comparative profitability analysis of selected bond issues<sup>3</sup>**

# / item	Name of bond issue	Yield for the expected holding period (subject to bond holding until maturity), % per annum	Yield for the expected holding period (subject to sale on September 18, 2020), % per annum	Error, % per annum
1	FPK Garant-Invest-001R-03	9.90	10.77	+0.86
2	RSG-Finans-4-bob	10.94	12.75	+1.82

<sup>3</sup> Moscow Exchange MICEX-RTS (<https://www.moex.com/ru/marketdata/bulletins/#/date=2020-09-18>)

4) the proposed compensation coefficients (price risk compensation coefficient, liquidity risk compensation coefficient, transaction cost compensation coefficient and general compensation coefficient) make it possible not only to comprehensively characterize the level of risk inherent in bonds according to the selected criteria, but also allow us to structure the bond portfolio (for

example, in inverse proportion to total compensation coefficients).

The algorithms designed in this research for the bond portfolio selection and their software applications can be embedded in robotized and semi-robotized trading terminals and investment applications for use by institutional and private investors. ■

### References

1. Zhengyao J., Xu D., Liang J. (2017) *A deep reinforcement learning framework for the financial portfolio management problem*. Working paper arXiv: 1706.10059v2. Available at: <https://arxiv.org/pdf/1706.10059.pdf> (accessed 22 December 2020).
2. Xiu G., Chan L. (2000) An algorithm for trading and portfolio management using Q-learning and Sharpe ratio maximization. *Proceedings of the International Conference on Neural Information Processing*, pp. 832–837. Available at: <http://www.cse.cuhk.edu.hk/~lwchan/papers/iconip00-qsr.pdf> (accessed 22 December 2020).
3. Almahdi S., Yang S. (2017) An adaptive portfolio trading system: A risk-return portfolio optimization using recurrent reinforcement learning with expected maximum drawdown. *Expert Systems with Applications*, no 87, pp. 267–279. DOI: 10.1016/j.eswa.2017.06.023.
4. Heaton J., Polson N., Witte J. (2016) Deep learning for finance: deep portfolios. *Applied Stochastic Models in Business and Industry* (electronic journal), vol. 33 (1), pp. 3–12. Available at: <http://www.ssrn.com/abstract=2838013> (accessed 25 January 2021). DOI: 10.2139/ssrn.2838013.
5. Korobov N.F., Beloglazov A.A. (2005) Management of corporate bond portfolio. *Bulletin of Taganrog Institute of Management and Economics*, no. 2, pp. 28–30 (in Russian).
6. Ponomarev A.A. (2007) Yield optimization of the sub-federal bond portfolio. *Bulletin of the Russian State Pedagogical University*, no 24, pp. 69–72 (in Russian).
7. Popova N.V. (2012) The influence of the coupon payments' frequency on the bond price. *Finance: Problems and Solutions*, no 3, pp. 40–44 (in Russian).
8. Lashevsky P.A. (2013) Classification of infrastructure bonds. *Theory and practice of social development*, no 10, pp. 381–383 (in Russian).
9. Darushin I.A. (2014) Assessment of the bond reinvestment risk. Complementing duration. *Finance and Credit*, no 24 (600), pp. 9–18 (in Russian).
10. Zadorozhnaya A.N. (2015) The influence of covenants on the corporate bond yield. *Finance and Credit*, no 7 (631), pp. 34–44 (in Russian).
11. Nazarova V.V., Levichev I.P. (2017) Development of the model of improving the effectiveness of investment portfolio. *HSE Economic Journal*, vol. 21, no 3, pp. 451–481 (in Russian).
12. Freydina I.A. (2017) International experience in financing infrastructure projects. *Economic policy*, vol. 12, no 4, pp. 196–203 (in Russian). DOI: 10.18288/1994-5124-2017-4-08.
13. Fomin M.V., Lakhno Yu.V., Pyshnograï A.P. (2019) Municipal bonds as an instrument for infrastructure development. *Issues of public and municipal governance*, no 2, pp. 185–210 (in Russian).
14. Mikhailov A.Yu. (2015) Term structure of public bond yield rates. *Financial analytics: problems and solutions*, no 38, pp. 42–52 (in Russian).

15. Burenin A.N. (2019) Limits of macroeconomic policy from the economic crises' point of view. *Economic policy*, vol. 14, no 1, pp. 76–91 (in Russian). DOI: 10.18288/1994-5124-2019-1-76-91.
16. Beshanov S.V., Lapshin V.A. (2019) Parametric immunization of interest rate risk via term structure models. *HSE Economic Journal*, vol. 23, no 1, pp. 9–31 (in Russian). DOI: 10.17323/1813-8691-2019-23-1-9-31.
17. Fabozzi F.J. (2014) *Bond markets, analysis and strategies*; 9th ed. New York: Pearson.
18. Fabozzi F.J. (2006) *Fixed income mathematics: Analytical and statistical techniques*; 4th ed. New York: McGraw-Hill Education.
19. Fabozzi F.J. (2007) *Fixed income analysis*; 2nd ed. New Jersey: John Wiley & Sons.
20. Galvani V., Landon S. (2013) Riding the yield curve: A spanning analysis. *Review of Quantitative Finance and Accounting*, vol. 40, no 1, pp.135–154. DOI: 10.1007/s11156-011-0267-7.
21. Bieri D.S., Chincarini L.B. (2004) *Riding the yield curve: Diversification of strategies*. Working paper SSRN 547682 (April 12, 2004). Available at: <https://ssrn.com/abstract=547682> (accessed 14 February 2021). DOI: 10.2139/ssrn.547682.
22. Cox R.A.K., Felton J.M. (1994) Performance from riding the yield curve: 1980–1992. *Journal of Business and Economic Perspectives*, vol. 20, pp. 128–132.
23. Wei Z., Xianhua W., Guofu W. (2007) An empirical research on the investment strategy of bond portfolios based on the changes of yield curves. *Management Review*, vol. 6, pp. 128–136.
24. Chami R., Cosimano T.F., Rochon C., Yung J. (2020) Riding the yield curve: Risk taking behavior in a low interest rate environment. *International Monetary Fund Working Paper*, no. 20/53. Available at: <https://www.imf.org/-/media/Files/Publications/WP/2020/English/wpia2020053-print-pdf.ashx> (accessed 10 January 2021).
25. Yakunina A., Varygin S., Nesterenko E., Korobov E., Semernina Y., Yakunin S. (2019) A complex algorithm for selecting instruments to finance mergers and acquisitions. Proceedings of the *Third Workshop on Computer Modelling in Decision Making (CMDM 2018)*, Saratov, Russia, November 14–17, 2018. Advances in Computer Science Research, February 2019, pp. 31–35. DOI: 10.2991/cmdm-18.2019.6
26. Semernina Y., Yakunina A., Nesterenko E., Yakunin S., Korobov E. (2016) Improving the tools used in computer modelling of the bond liquidity assessment on the Russian market. Proceedings of the *Workshop on Computer Modelling in Decision Making (CMDM 2016)*, Saratov, Russia, November 10–11, 2016. CEUR-WS, vol. 1726, pp. 100–108. Available at: <http://ceur-ws.org/Vol-1726/paper-10.pdf> (accessed 23 January 2021).

### About the authors

#### Eugene A. Korobov

Head of Laboratory of IT in Economics, Saratov State University, 83 Astrakhanskaya Street, Saratov 410012, Russia;

E-mail: korobovea@yandex.ru

ORCID: 0000-0001-7198-5941

**Yulia V. Semernina**

Dr. Sci. (Econ.), Associate Professor;

Professor, Department of Finance and Banking, Socio-Economic Institute, Saratov State Technical University, 77 Polytechnicheskaya Street, Saratov 410054, Russia;

E-mail: ysemernina@yandex.ru

ORCID: 0000-0003-2098-795X

**Alina S. Usmanova**

Cand. Sci. (Econ.);

Associate Professor, Department of Finance and Banking, Socio-Economic Institute, Saratov State Technical University, 77 Polytechnicheskaya Street, Saratov 410054, Russia;

E-mail: alina.brosalova@yandex.ru

ORCID: 0000-0002-8488-0425

**Kristina A. Odinkova**

Lecturer, Department of Finance and Banking, Socio-Economic Institute, Saratov State Technical University, 77 Polytechnicheskaya Street, Saratov 410054, Russia;

E-mail: odinkova.kristina93@mail.ru

ORCID: 0000-0002-8191-1337

# To the question of restoring symbol sequences encoding noisy periodic functions

**Galina N. Zhukova**<sup>a</sup> 

E-mail: galinanzhukova@gmail.com

**Mikhail V. Ulyanov**<sup>b,c</sup> 

E-mail: muljanov@mail.ru

<sup>a</sup> National Research University Higher School of Economics

Address: 20, Myasnitskaya Street, Moscow 101000, Russia

<sup>b</sup> Trapeznikov Institute of Control Sciences, Russian Academy of Sciences

Address: 65, Profsoyuznaya Street, Moscow 117997, Russia

<sup>c</sup> Lomonosov Moscow State University

Address: 1, Leninskie Gory, Moscow 119991, Russia

## Abstract

In business informatics, one of the research subjects is the analysis of data on processes in applied subject areas; here problems of qualitative analysis arise. Such problems arise, for example, in the qualitative study of log files of business processes, in the analysis and prediction of time series and other processes of a different nature. Quite often, to represent information about the processes under study, the methods of qualitative analysis use symbolic coding, which makes it possible to remove unnecessary detailing of numerical descriptions. The relevance of this study is due to the fact that when working with the raw data, researchers often face the presence of noise and distortions of the data, which significantly complicates the solution of the problems of qualitative analysis. When working with symbolic representations of the processes under study, which quite often have a periodic nature, we observe noise of deletion, insertion and replacement of symbols, which complicate the solution of the problem of revealing and analyzing the periodicity. This article deals with the problem of recovering periodic symbolic sequences obtained by coding from samples of continuous periodic functions and distorted by noise of insertion, replacement and deletion of symbols. Trigonometric functions are considered as a specific example of synthetic time series data. To encode trigonometric

functions, alphabets of various cardinalities are used. The article presents an experimental study of the dependence of the quality characteristics of the method of period and a periodically repeating fragment recovery, previously proposed by the authors and improved in this study. For alphabets of different cardinalities at fixed sampling intervals, the fraction of sequences with a satisfactorily reconstructed period and the relative error in determining the period are given. The quality of reconstruction of a periodically repeating fragment is estimated by the edit distance from the reconstructed periodic sequence to the original sequence distorted by noise.

**Key words:** symbolic sequence; cardinality of an alphabet; periodic sequence; sequence with noise; noise of insertion; noise of deletion; noise of change.

**Citation:** Zhukova G.N., Ulyanov M.V. (2021) To the question of restoring symbol sequences encoding noisy periodic functions. *Business Informatics*, vol. 15, no 4, pp. 22–35. DOI: 10.17323/2587-814X.2021.4.22.35

## Introduction

One of the subjects of scientific research in business informatics is the analysis of data on processes recorded in applied subject areas [1]. The variety of emerging data analysis problems includes the tasks of qualitative analysis. They arise, for example, during a qualitative study of the log files of business processes [1] and are associated, for example, with determining the correspondence of the log file to the process model [2]. One of the ways to present information about the processes being studied in scientific research is to represent them in the form of time series. At the same time, a significant proportion of the tasks of qualitative analysis relates precisely to research in the field of time series, and is associated both with the analysis of their periodicity and, in general, with the problems of their complex analysis and forecasting [3–8]. Similar qualitative problems arise in the analysis of other processes of a different nature, represented by time series, for example, in environmental monitoring and forecasting environmental changes [9].

The observed values of the process under study, which are elements of the time series,

are quite often exposed to random distortions caused by external factors. “The values compared to the elements of the resulting series also contain measurement errors and, in the general case, are subject to random external influences. Further, such measurement errors and the results of external influences are interpreted as noise” [10].

Let us point out some publications from the field of business informatics and management in which one way or another the influence of noise on the results of forecasting time series is discussed and/or neutralized. The authors in [11–13] note that when creating models and predicting the production and consumption of electricity, noise in the raw data affects the predictive power of models and the quality of forecasts. Stochastic forecasting of risks in business [14], including the risk of operating profit for firms, relies heavily on the assumption of incomplete and noisy data. The authors of [15] point out that when studying the behavior of bank clients by the method of clustering time series, errors and range of the raw data are taken into account. In [16], when assessing the efficiency of forecasting passenger air traffic flows using multiple error indicators, in particular, it was shown that the noise in the raw



data affects the forecasting quality. The authors in [17] try to improve the quality of tourism demand forecasting using deep learning methods in conjunction with image processing of time series visualization, and thereby reduce the influence of errors in the raw data on the forecasting results.

Among other subject areas, we note, for example, microelectronics, where image blurring caused by noise strongly affects the quality of prediction [18] and observations of the biosphere. In [19], the authors note that biometric data (in a broad sense, a kind of data observations of the biosphere) often have observation gaps, outliers and breaks. For one more subject area – remote sensing of the Earth from space, we quote from [20]: “The use of time series of satellite data for monitoring the earth’s surface is associated with the problem of taking into account all sorts of interfering factors leading to partial loss or distortion of information on the dynamics of spectral-reflective characteristics of objects of observation. Such factors include mist and cloudiness, which are opaque in the visible and near-infrared range, shadows from it, as well as measurement errors.”

In order to solve the problems of qualitative analysis, time series data are subjected to symbolic coding, which allows us to remove unnecessary detailing of numerical descriptions [21–23]. In this case, the description of the elements of the time series or steps of the business process is encoded with a word over a finite alphabet, which is the object of further research. Moreover, such coding is relevant for the qualitative analysis of big data. This is due to the fact that the high accuracy of numerical representations of time series elements leads not only to unjustifiably large amounts of information, but also laborious calculations that do not improve the quality of the results obtained [22, 23].

Obviously, when studying time series, working with noisy data causes significant difficul-

ties. This leads to the formulation of the problem of noise elimination. For noise reduction in numerical time series data, various smoothing methods are used, such as moving average, exponential smoothing, etc. [10]. However, these methods are not applicable when dealing with noisy symbolic sequences.

In the aspect of symbolic coding, the arising errors, interpreted as noise, lead to the fact that insertion, deletion and replacement of symbols occur in symbolic sequences. Therefore, for example, errors associated with the adjustment of measuring instruments, inaccuracies in manual data entry and accidental mistakes or deliberate distortion of the values of individual indicators, lead to replacement noise [19]. Registration errors are a source of deletion and insertion noise, and these noises can also occur during data preparation. We also note that the methods for detecting periodicity used for numerical sequences are not applicable when working with symbolic representations [24].

The relevance of this study is due to the fact that noise reduces the efficiency of time series analysis. In this regard, the article discusses the problem of recovering periodic symbolic sequences obtained by coding from samples of periodic functions and distorted by noise of insertion, replacement and deletion of symbols. Trigonometric functions are considered as a specific example of synthetic time series data. To encode trigonometric functions, alphabets of various cardinalities with different granularity of sampling intervals by model time are used.

The article presents an experimental study of the quality characteristics of the method of period and a periodically repeating fragment recovery, previously proposed by the authors in [25] and improved in this study. For alphabets of different cardinalities at fixed sampling intervals according to the model time, the fraction of sequences with a satisfactorily reconstructed value of the period and the relative error in



determining the period are given. The quality of reconstruction of a periodically repeating fragment is measured by the edit distance from the reconstructed periodic sequence to the original sequence distorted by noise.

## 1. Terminology and notation

Further, we will use the notation introduced by us in article [25], which describes a method for recovering a periodic symbolic sequence.

Let  $\Sigma^\sigma$  be the alphabet of cardinality  $\sigma = |\Sigma| \geq 2$ . We will call a word of length  $n$  the symbolic sequence  $q^\sigma = s_1, s_2, \dots, s_n$  over a finite alphabet  $\Sigma^\sigma$ , where  $s$  is an arbitrary symbol of  $\Sigma^\sigma$ , and a subword or a fragment of the word  $q^\sigma$  is any sequence of symbols  $s_k, s_{k+1}, \dots, s_{l-1}, s_l$ ,  $1 \leq k \leq l \leq n$ .

We consider periodic symbolic sequences with period  $p$ . To avoid ambiguity in understanding, we will call the period  $p$  the length of a repeating subword (fragment), and call a part of a periodic sequence of length  $p$  a periodically repeating fragment. Unless otherwise stated, a periodically repeating fragment is a fragment of length  $p$  that begins with the first symbol of the periodic word.

Let  $q^\sigma(m, p)$  be a periodic word containing  $m \geq 8$  repeating fragments of length  $p$ ;  $\tilde{q}^\sigma(m, p)$  – a word over the same alphabet as  $q^\sigma(m, p)$ , but with introduced noise;  $\bar{q}^\sigma(m, p)$  – a periodic word obtained by analysing  $q^\sigma(m, p)$  using the algorithm from [25] with the improvements proposed in this article.

Note that the word  $\bar{q}^\sigma(m, p)$  has the same length as  $\tilde{q}^\sigma(m, p)$  and serves as an approximation of the periodic word  $\tilde{q}^\sigma(m, p)$ .

## 2. Statement of the problem

Let there be some continuous-time periodic process  $g(t)$  from time  $t_0$  to time  $t_0 + n\Delta t$

during which the value  $g(t_i)$ ,  $i = 1, n$ , is measured at regular intervals  $\Delta t$ . Partitioning the range of measured values of  $g(t)$  into  $\sigma$  equal half-segments and encoding the values of  $g(t_i)$  with symbols of alphabet  $\Sigma^\sigma$ , we get the symbolic sequence  $q^\sigma = s_1, s_2, \dots, s_n$  over this alphabet. Note that some symbols of alphabet  $\Sigma^\sigma$  may not be present in the resulting symbolic sequence. We will assume that the period  $p$  of the observed function  $g(t)$  is a multiple of the length of interval  $r\Delta t$  between successive measurements, i. e.  $p = r\Delta t$ , where  $r$  is an integer, due to which, when encoding the periodic function  $g(t)$  on a segment of  $m$  periods, a periodic symbolic sequence is obtained, also containing  $m$  periods, while  $n = mp = mr\Delta t$ .

Now let us introduce random distortions into the measured values of the function  $g(t)$ . We will consider the noises of insertion, replacement, and deletion. Insertion of a new value corresponds to some failure in measurements, when an extraordinary measurement occurred between the scheduled measurements; deletion means the loss of a value when entering or transferring results of measurements; and replacement is considered an incorrect measurement or deliberate distortion of the data.

After introducing all the distortions, we get a noisy sequence, which we encode in the same alphabet  $\Sigma^\sigma$  and consider the problem of recovering a periodic symbolic sequence from a given noisy sequence.

In this article, we consider the symbolic sequences obtained by encoding values of a continuous periodic function (in particular,  $\sin(t)$ ) in points with step  $\Delta t$  on a segment with a length of at least eight full periods ( $m \geq 8$ ) when they are distorted by different types of noise.

When coding a continuous function, an essential role is played by the choice of the cardinality of the alphabet. We encode each function under consideration with symbols of

alphabets of cardinality from 10 to 60 with a step of 10. Despite the fact that when encoding by partitioning the range of values of the encoded function into equal half-segments, some symbols of the alphabet may not occur in an undistorted periodic symbolic sequence. Such symbols can be observed in the sequence obtained by encoding the distorted sequence of measured values of the function.

Statement of the problem: to study the influence of the cardinality of the alphabet, the type of function and the noise level on the quality of reconstruction of the periodic symbolic sequence obtained by coding the values of the periodic function under conditions of its distortion by the noise of insertion, replacement and deletion of values.

### 3. Continuous periodic function encoding

Consider function  $\sin(t)$  on the segment  $[0 - 16\pi]$ , which contains 8 full periods of function  $\sin(t)$ . To construct a numerical sequence with period  $p$ , we partition the interval  $[0 - 16\pi]$  into  $8p$  equal half-segments of length  $\Delta t = \frac{2\pi}{p}$  and evaluate the value of function  $\sin(t)$  in the middle of each half-segment, getting a sequence of  $8p$  real numbers  $y_1, y_2, \dots, y_{8p}$ .

For the purpose of symbolic coding of the values obtained in the alphabet of cardinality  $\sigma$ , we partition the range of sine values – segment  $[-1, 1]$  into  $\sigma$  consecutive half-segments  $I_1, I_2, \dots, I_\sigma$  of equal length; in this case, the number of half-segments is equal to the cardinality of the alphabet. We assign the semi-segments  $I_1, I_2, \dots, I_\sigma$  to the symbols of the alphabet  $\Sigma^\sigma$ . Each number in the sequence  $y_1, y_2, \dots, y_{8p}$  is encoded by the symbol corresponding to the half-segment  $I_j$ , in which this number falls. As a result, we get the symbolic sequence  $q^\sigma(m, p)$ . Note that some symbols of alphabet  $\Sigma^\sigma$  may not appear in this sequence.

### 4. Method for determining the period from a periodic sequence with noise

To understand the proposed improvements to the method for constructing a periodically repeating fragment, we present a brief description of the method for determining the period [25, 26].

The method assumes counting the number of all subwords of length  $k = 10$  in a noisy sequence  $\tilde{q}^\sigma = s_1, s_2, \dots, s_n$ . Subwords of length 10 in  $\tilde{q}^\sigma$  are taken with a shift by one symbol, i. e. the subwords  $s_1, s_2, \dots, s_{10}, s_2, s_3, \dots, s_{11}$ , etc, are considered. Those subwords that have met at least 3 times compose the set  $R$ . Each subword from  $R$  is associated with a list of position numbers of symbols of the sequence  $\tilde{q}^\sigma$ , starting from which this subword is included in  $\tilde{q}^\sigma$ . So, in the sequence “*abcabcdabcedabcabcdabced*” the subword “*abcabcdabc*” is included starting from position numbers 1 and 13.

Further, for each subword of length  $k = 10$ , the differences between the numbers of consecutive occurrences are calculated, after which the set  $\Omega$  is constructed from such differences. Each element of this set is compared the number of times when such a difference in the numbers of consecutive occurrences of the first symbols of subwords of length 10 was observed. For example, in a word “*abcabcdabcedabcabcdabced*” the difference  $r = 13 - 1 = 12$  is observed 3 times (for the words “*abcabcdabc*”, “*bcabcdabcd*”, “*cabcdabced*”). Due to the introduced noise, the differences not always equal the period or its multiples, but most of the differences take values close to the period or its multiple. In this regard, for each value of the difference, it is calculated how many times a difference close to it has occurred in  $\Omega$ . In our experiments, the proximity was defined as falling within the interval  $\pm 20\%$ . The analysis of the obtained differences allows one to obtain an estimate of the period for an unknown strictly periodic sequence [25].

### 5. Improved method for constructing a periodically repeating fragment

The solution that the algorithm from [25] delivers in terms of constructing a periodically repeating fragment is the subword  $\tilde{f}^\sigma = s_1, s_2, \dots, s_{\tilde{p}}$  of the analyzed word  $\tilde{q}^\sigma(m, p)$ , minimizing (on the set of obtained variants of fragments) the edit distance from the symbolic sequence of length  $|\tilde{q}^\sigma(m, p)|$  constructed from this fragment to the sequence  $\tilde{q}^\sigma(m, p)$ .

The construction of an approximating periodically repeating fragment  $\tilde{f}^\sigma$  in [25] is carried out by splitting the distorted sequence into successive subwords of length  $\tilde{p}$  (the last subword of length less than  $\tilde{p}$  was not taken into account) and choosing such one from them, in which the edit distance to one of the remaining subwords is minimal. If there are several such subwords, the first subword with the minimum edit distance to another subword is selected.

The method proposed by us and described below attempts to improve the  $\tilde{f}^\sigma$  fragment in order to obtain a smaller value of the edit distance between the distorted and approximating periodic sequences. The proposed improvement is achieved by using information about the previously determined frequencies of subwords of length  $k = 10$  observed in  $\tilde{q}^\sigma(m, p)$ .

First, based on the fragment  $\tilde{f}^\sigma$ , the word  $\tilde{f}_4^\sigma$  is built, containing the fragment  $\tilde{f}^\sigma$ , written four times in a row. Then each subword of the fragment  $\tilde{f}_4^\sigma$  of length  $k = 10$  ( $(s_1, s_2, \dots, s_{10}; s_2, s_3, \dots, s_{11}$  etc. ) is checked for occurrence in the word  $\tilde{q}^\sigma(m, p)$  at least 3 times, i.e. to belong to the set  $R$ . The first of the checked subwords  $\omega_0 = s_{t+1}, s_{t+2}, \dots, s_{t+10}$ , belonging to the set  $R$ , becomes a start subword, the improved periodic fragment will begin with it. If no such subword was found, the periodic fragment remains unimproved.

After the subword  $\omega_0$  is found, the subwords  $\omega$  from  $\tilde{f}_4^\sigma$  are sequentially scanned with a shift by one character, i. e.  $\omega = s_{t+h}, s_{t+h+1}, \dots, s_{t+h+9}$ ,

$h = 2, 3, \dots$ , membership check is performed of these subwords to the set  $R$ . If the subword belongs to  $R$ , then the corresponding subword of the word  $\tilde{f}_4^\sigma$  remains unchanged, otherwise we start counting consecutive subwords (with a shift by one character) not included in  $R$ .

Since the length of  $\omega$  is 10, until there are 10 consecutive subwords not included in  $R$ , the word  $\tilde{f}_4^\sigma$  remains unchanged. If, after less than 10 consecutive  $\omega$  not included in  $R$ , the next  $\omega$  turned out to be included in  $R$ , then the counter of consecutive subwords not included in  $R$  is reset to zero. If in  $R$  there are no 10 consecutive subwords  $\omega$  but the eleventh  $\omega$  is included in  $R$ , then the search for a subword  $\omega^R$  in  $R$  begins, such that the first  $m$  symbols of  $\omega^R$  coincide with the last  $m$  symbols of  $\omega^-$  — the last of the considered subwords in  $R$ ,  $3 \leq m \leq 9$ . The values of  $m$  are taken from 9 to 3, i. e. first we try to find  $\omega^R$ , in which the first 9 symbols coincide with the last 9 symbols  $\omega^-$ , if such  $\omega^R$  is found, then in  $\tilde{f}_4^\sigma$  replace the symbol following the last by the subword symbol  $\omega^-$ , to the last symbol  $\omega^R$ .

If for  $m = 9$  the subword  $\omega^R$  is not found, we continue with  $m = 8$ , and so on up to  $m = 3$ . Let  $\omega^R$  be found for some  $m$  from 3 to 9, then in  $\tilde{f}_4^\sigma$  we replace the symbol following the last symbol of the subword  $\omega^-$  by the  $(m + 1)$ -st symbol  $\omega^R$ . If no  $\omega^R$  was found for any  $m$  from 3 to 9, then the fragment  $\tilde{f}^\sigma$  remains unimproved.

At some point, either  $2\tilde{p}$  consecutive subwords  $\omega$  will be scanned and the fragment  $\omega_0$  will never be encountered, or the next  $\omega$  subword will match  $\omega_0$ . In the first case, as an improved fragment  $\tilde{f}^\sigma$  we take the first  $\tilde{p}$  symbols of  $\tilde{f}_4^\sigma$  starting from the first symbol  $\omega_0$  (i.e. from  $\tilde{f}_4^\sigma$  cut out  $\omega_0$  and the following symbols in  $\tilde{f}_4^\sigma$ ,  $\tilde{p}$  symbols in total). In the second case,  $\tilde{f}^\sigma$  consists of symbols of  $\tilde{f}_4^\sigma$ , starting from the first occurrence of  $\omega_0$  to the second  $\omega_0$ , while checking the possibility that at the end of the periodic fragment due to insertion or deletion noise, one symbol is lost or added.

The check is done as follows. If after the next  $\omega$  included in  $R$  in  $\tilde{f}_4^\sigma$  there were 9 consecutive subwords not included in  $R$ , after which a word from  $R$  was found and this word coincided with  $\omega_0$ , then perhaps there was a noise of deletion. To check, in the set  $R$ , we look for  $\omega^R$ , in which the first  $m$  symbols coincide with the last  $m$  symbols of  $\omega$ . If such a word  $\omega^R$  is found and its symbols starting from  $m + 2$  coincide with the first symbols  $\omega_0$ , then at the end of  $\tilde{f}^\sigma$  we add the  $(m + 1)$ -st symbol of  $\omega^R$ .

In addition, if after the next  $\omega$  that is included in  $R$ , in  $\tilde{f}_4^\sigma$  there were 9 consecutive subwords not included in  $R$ , after which a word from  $R$  was found and this word coincided with  $\omega_0$ , it is possible that there was a noise of insertion, and we try to eliminate it like that. If the first symbol  $\omega_0$  matches the last symbol  $\omega$ , then we do not include in  $\tilde{f}_4^\sigma$  the last symbol before the second occurrence of  $\omega_0$  in  $\tilde{f}_4^\sigma$ .

In order to obtain a more accurate value of a period simultaneously with the search for the improved fragment  $\tilde{f}^\sigma$ ,  $2\tilde{p}$ , consecutive subwords  $\omega$  of the word  $\tilde{f}_4^\sigma$ , are searched, until on the next step among the  $\omega$  we meet the word  $\omega_0$ . Then as the improved fragment we regard the subword of the improved  $\tilde{f}_4^\sigma$ , starting from the first occurrence of  $\omega_0$  to the second one, if the period is at least 3. If changes in  $\tilde{f}_4^\sigma$  during the improvement process did not yield another  $\omega_0$ , then after the end of the improvement process, we perform another scan of  $2\tilde{p}$  consecutive subwords  $\omega$  of the word  $\tilde{f}_4^\sigma$  starting from the first occurrence of  $\omega_0$ , and if there is a second occurrence of  $\omega_0$ , then the improved fragment is the subword of the improved  $\tilde{f}_4^\sigma$ , starting from the first occurrence of  $\omega_0$  to the second. If the second occurrence of  $\omega_0$  is not found, then  $\tilde{p}$  consecutive symbols of improved  $\tilde{f}_4^\sigma$ , starting from the first occurrence of  $\omega_0$ , are taken as the improved fragment.

If it is possible to improve a periodic fragment, its cyclic shifts are considered, and the final fragment is taken as the shift, which has

the minimum edit distance to the beginning of the noisy sequence, i.e. its first  $\tilde{p}^*$  symbols, where  $\tilde{p}^*$  is the length of the improved periodic fragment.

If, as a result of improving a fragment, a subword of length more than three is obtained, and the edit distance from the periodic sequence constructed by repeating the improved fragment to the noisy one is less than in the case  $\tilde{f}^\sigma$  before improvement, we use improved  $\tilde{f}^\sigma$  as an approximation of the periodically repeating fragment, otherwise original.

## 6. Evaluation of the quality of the period recovery

The estimation of the accuracy of determining the period and the quality of reconstruction of a periodic fragment will be carried out separately. Let the period of the sequence before introducing distortions be  $p$ , and our algorithm determined that the period is  $\bar{p} = \bar{p}(\bar{q}^\sigma)$ . Then the precision  $\delta$  of determining the period is defined as

$$\delta = \frac{|p - \bar{p}|}{p}. \quad (1)$$

For a periodic sequence obtained by encoding a periodic function on eight periods, we get a series of 100 random noisy sequences. For each of them we determine the period and find the value  $\delta$ , then we calculate the average value  $\delta$  and the median of the sample over 100 noisy sequences.

The quality of reconstruction of a periodic fragment will be estimated by evaluating the ratio of the edit distance  $d(\bar{q}^\sigma, q^\sigma(m, p))$  between the reconstructed and original periodic symbolic sequences, to the length of the original periodic sequence [27]. Let's denote this ratio  $\varepsilon(\bar{q}^\sigma, q^\sigma)$ :

$$\varepsilon(\bar{q}^\sigma, q^\sigma) = \frac{d(\bar{q}^\sigma, q^\sigma(m, p))}{mp}. \quad (2)$$

With this approach, a good estimate will be given by the case when the period obtained by the algorithm is two to three times larger than the initial one, but at the same time the periodic fragment is close to the original, repeated the required number of times.

In addition, the original and reconstructed sequences were compared with the noisy one; in this case, the first  $n$  symbols were taken from each periodic sequence for comparison, where  $n$  is the length of the noisy sequence.

### 7. Computational experiment scheme

In a computational experiment, a study was carried out of the method proposed in [25, 26] and improved in this article on the following functions:

- ◆  $\sin(t)$  on the segment  $[0 - 16\pi]$ ;
- ◆  $\left(t - 8 \left\lfloor \frac{t}{8} \right\rfloor + 1\right) \sin(\pi t)$  on the segment  $[0 - 64]$ ;
- ◆  $\left(t - 16 \left\lfloor \frac{t}{16} \right\rfloor + 1\right) \sin(\pi t)$  on the segment  $[0 - 128]$ ;
- ◆  $\left|7 - t + 16 \left\lfloor \frac{t}{16} \right\rfloor\right| \sin(\pi t)$  on the segment  $[0 - 128]$ , where  $\lfloor \dots \rfloor$  stands for the integer part of the number.

Thus, for all functions, the segment of argument values containing eight full periods was considered. The functions were encoded in the alphabets  $\Sigma^\sigma$  of cardinality  $\delta$  from 10 to 60 with a step of 10. The range of function values was partitioned into  $\delta$  intervals, with each interval  $I_j$  being encoded by the symbol  $s_j$  of the alphabet  $\Sigma^\sigma$ .

The values of the period  $p$  were chosen equal to 20, 30 and 50, while one period of the function was partitioned into  $p$  equal half-segments. Then for each half-segment  $\Delta t_i$  the value of the function in the middle of the half-segment was

evaluated, and this value was used to determine the half-segment  $I_j$  to which this value belongs, after which the symbol  $s_j$ , encoding the half-segment  $I_j$ , was written to the  $i$ -th position of the coding word. Thus, periodic symbolic sequences were obtained in alphabets of cardinality from 10 to 60.

Based on the periodic sequences thus constructed, random noisy sequences were obtained, 100 sequences for each purely periodic one. The noise was introduced in accordance with our earlier proposed probabilistic noise model for periodic symbolic sequences [28].

In the first series of experiments, noise was introduced with a total level of 5%, with the levels of insertion, replacement, and deletion noise taking values from 1 to 5% with a step of 1%, so that the sum of the noise levels is 5%. In the second series, noise was introduced uniformly distributed among the types, i. e. the level of insertion, replacement and deletion noise was the same in a separate experiment, taking values from 1 to 4% in 1% increments.

### 8. Results and discussion

The results of experimental studies for all four functions are shown in *Figures 1–4* and in *Tables 1* and *2*. In *Figures 1–4*, we show the numerical sequences of the values of functions on two periods containing 50 samples per period, with the coding alphabet cardinality of 20. In the figures, the numbers of the samples are plotted along the horizontal axis, and the numbers of the half-segments of the coding corresponding to the symbols of alphabet of cardinality 20 are shown along the vertical axis.

For all functions, as an example, one variant of noise injection is shown from the various variants investigated. Namely, a variant with a total noise level of 5%, consisting of a deletion noise of 2% and an insertion noise of 3%.



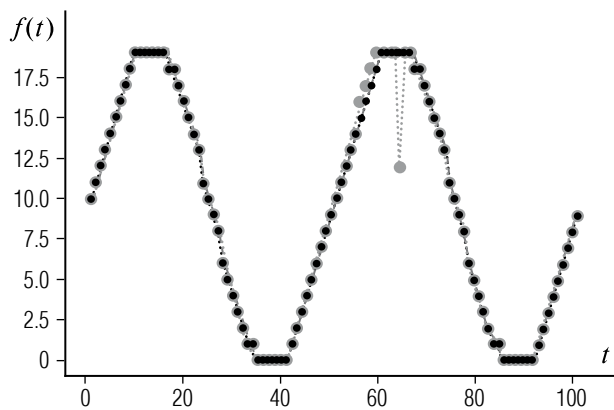


Fig. 1. Two periods of the function  $\sin(t)$ ,  
total noise 5%, deletion noise 2%,  
insertion noise 3%, 50 samples per period,  
coding alphabet cardinality 20 symbols

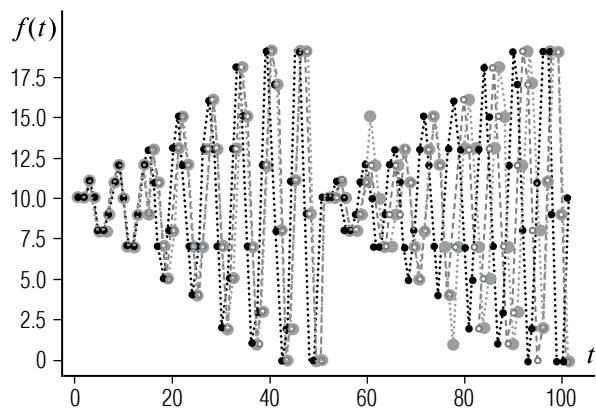


Fig. 3. Two periods of the function  
 $\left(t - 16 \left\lfloor \frac{t}{16} \right\rfloor + 1\right) \sin(\pi t)$ ,  
total noise 5%, deletion noise 2%,  
insertion noise 3%, 50 samples per period,  
coding alphabet cardinality 20 symbols

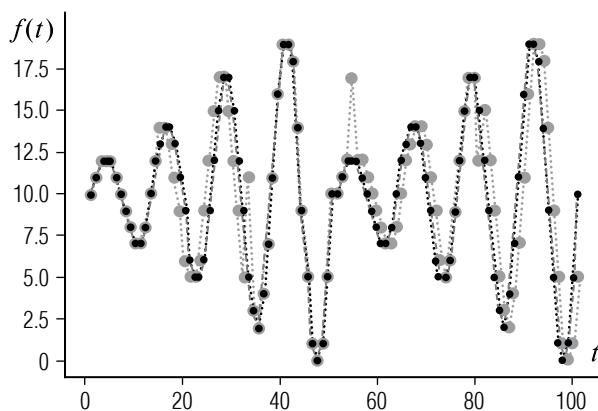


Fig. 2. Two periods of the function  
 $\left(t - 8 \left\lfloor \frac{t}{8} \right\rfloor + 1\right) \sin(\pi t)$ ,  
total noise 5%, deletion noise 2%,  
insertion noise 3%, 50 samples per period,  
coding alphabet cardinality 20 symbols

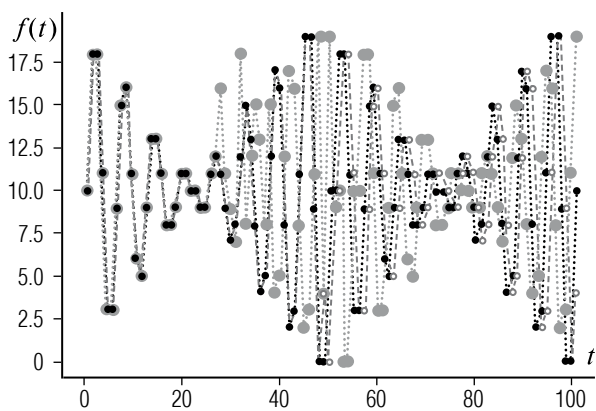


Fig. 4. Two periods of the function  
 $\left|7 - t + 16 \left\lfloor \frac{t}{16} \right\rfloor\right| \sin(\pi t)$ ,  
total noise 5%, deletion noise 2%,  
insertion noise 3%, 50 samples per period,  
coding alphabet cardinality 20 symbols

In all the figures, the black dots correspond to the original periodic sequence obtained from the specified function, the gray dots are the noisy sequence of function values in samples, and the white ones with an outline are the periodic sequence restored by the algorithm under study. The dotted line shows the sequence of points by count.

The results of the study on evaluating the accuracy  $\delta$  of determining the period showed that for all functions the results obtained do not differ much from each other, i. e. the method under investigation is weakly sensitive to the form of a periodic function (at least for these four functions). Therefore, we present the results for only one function.

Table 1.

Fraction of symbolic sequences with a recovered period,  
not more than 2% different from the original,

$$\text{the function is } \left( t - 8 \left\lfloor \frac{t}{8} \right\rfloor + 1 \right) \sin(\pi t)$$

Noise level, %			Alphabet cardinality					
deletion	change	insertion	10	20	30	40	50	60
0	0	5	25	26	33	24	26	31
0	1	4	25	16	22	26	24	26
0	2	3	98	97	94	95	96	93
0	3	2	99	100	98	99	99	100
0	4	1	100	98	100	100	100	100
0	5	0	100	100	100	100	100	100
1	0	4	94	93	94	90	95	96
1	1	3	100	99	96	100	97	97
1	2	2	99	100	99	100	98	97
1	3	1	99	97	97	97	98	99
1	4	0	94	96	97	96	97	95
2	0	3	99	96	99	99	96	97
2	1	2	98	97	99	100	99	98
2	2	1	99	100	99	97	100	97
2	3	0	97	98	97	98	98	98
3	0	2	100	100	98	97	99	100
3	1	1	98	97	98	96	98	98
3	2	0	85	94	91	91	87	90
4	0	1	88	85	88	83	90	88
4	1	0	22	34	28	26	35	31
5	0	0	8	9	7	7	17	12

For better clarity, we give in *Table 1* not the average value  $\delta$  averaged over 100 experiments by introducing random noise with a given level, but the fraction of sequences (out of 100 noisy ones) with a reconstructed period, the value of

which differs by no more than 2% from the initial period.

The results of studies on the quality of reconstruction of a periodically repeating fragment by the improved algorithm are shown in

Table 2.

**Influence of the noise level and cardinality  
of the alphabet on the median  $\varepsilon(\bar{q}^\sigma, q^\sigma) \cdot 100\%$**

Total noise %	Alphabet cardinality	Median of values $\varepsilon(\bar{q}^\sigma, q^\sigma) \cdot 100\%$ of different functions			
		$\sin(t)$	$\left(t - 8 \left\lfloor \frac{t}{8} \right\rfloor + 1\right) \sin(\pi t)$	$\left(t - 16 \left\lfloor \frac{t}{16} \right\rfloor + 1\right) \sin(\pi t)$	$\left 7 - t + 16 \left\lfloor \frac{t}{16} \right\rfloor\right  \sin(\pi t)$
3%	10	0	0	0	0
	20	0	0	0	0
	30	0	0	0	0
	40	0	0	0	0
	50	0	0	0	0
6%	10	2	2	2	2
	20	2	2	2	2
	30	2	2	2	2
	40	2	2	2	2
	50	2	2	2	2
9%	10	4	4	4	4
	20	6	4	4	4
	30	4	4	4	4
	40	4	4	5	4
	50	4	4	5	4
12%	10	7	6	8	6
	20	8	6	8	6
	30	6	6	6	6
	40	8	8	6	8
	50	8	8	7	6

*Table 2.* The values of the median  $\varepsilon(\bar{q}^\sigma, q^\sigma)$  are given for all four functions under study with a noise of uniform structures. In this case, the same noise level of each type (insertion, deletion, replacement) was randomly introduced into each of the 100 ini-

tial sequences. This level varied from 1 to 4% with a step of 1%. The experiments were carried out for all cardinalities of the alphabet – 10, 20, 30, 40 and 50 and all values of the period. *Table 2* shows the results for a period value of  $p = 50$ .



Based on the experimental data obtained, the improved method for reconstructing the period and periodically repeating fragment shows generally satisfactory results. The data in *Tables 1* and *2* show that the method has weak sensitivity in terms of the cardinality of the coding alphabet and in the form of the function, both in determining the period and in recovering a periodically repeating fragment.

When determining the period (see *Table 1*), the method is susceptible to the difference between the levels of insertion and deletion noise, since it is this difference that affects the distances between repeated subwords of length 10 in a noisy sequence. Note that the presence of only 5% replacement noise leads to the best experimentally observed result. When determining a periodically repeating fragment with a noise of uniform structure, neither the form of the function, nor the cardinality of the alphabet has a noticeable effect on the results. The only influencing factor in this case is the overall noise level.

### Conclusion

The use of models of symbolic cycles with noise makes it possible to solve problems of probabilistic forecasting of symbolic noisy sequences and allows you to develop effective methods for forecasting, reconstruction and approximation of data in the form of symbolic codes based on fragmentary, incomplete and distorted information.

This article proposes an improved method for solving the problem of recovering a periodic symbolic sequence based on the original sequence obtained by introducing insertion, deletion and replacement noise into an unknown periodic sequence. The method is based on the study of frequency occurrence and distances between coinciding subwords of fixed length. Symbolic sequences in alphabets of different cardinality encoding noisy periodic functions are consid-

ered as synthetic data. In addition to describing the improved method for finding a periodically repeating fragment, the article contains the results of an experimental study of the dependence of the quality characteristics of the method for restoring the period and periodically repeating fragment on the cardinality of the coding alphabet and noise levels of various types. The study was carried out for noisy symbolic codes of periodic functions, which are models of noisy (quasiperiodic) time series. This kind of input data often occurs in the problems of analysis and forecasting of time series in business informatics and management.

A computational experiment has shown that the quality of the method depends not only on the general noise level, but also on the ratio of the noise level. The proposed method has weak sensitivity in terms of the cardinality of the coding alphabet and in the form of a periodic function, both in determining the period and in recovering a periodically repeating fragment. A study for noise with a uniform structure showed that the only factor affecting quality is the noise level, while neither the type of function, nor the cardinality of the alphabet has a noticeable effect on the results.

The results obtained in the article make it possible to give recommendations on the possible application of the method when solving problems of analyzing symbolic codes of noisy periodic continuous functions in alphabets of small cardinality, with a noise level not exceeding 10–12%. Such problems arise in the analysis of both dynamic processes and time series in business informatics and management, and in the analysis of business processes in conditions of incomplete and fragmentary information. ■

### Acknowledgments

This work was supported by RFBR grants 19-07-00150 and 19-07-00151.

## References

1. Andersen B. (2007) *Business process improvement toolbox*. Milwaukee, Wisconsin: ASQ Quality Press.
2. Mitsyuk A.A., Lomazova I.A., van der Aalst W. (2017) Using event logs for local correction of process models. *Automatic Control and Computer Sciences*, vol. 51, no 7, pp. 709–723. DOI: 10.3103/S0146411617070306.
3. Keogh E.J., Pazzani M.J. (1998) An enhanced representation of time series which allows fast and accurate classification, clustering and relevance feedback. *Proceedings of the Fourth International Conference on Knowledge Discovery and Data Mining (KDD '98), New York, USA, 27–31 August 1998*, pp. 239–241. Available at: <https://www.aaai.org/Papers/KDD/1998/KDD98-041.pdf> (accessed 16 November 2021).
4. Berndt D.J., Clifford J. (1994) Using dynamic time warping to find patterns in time series. *AAAI Technical Report. Workshop on Knowledge Discovery in Databases (KDD '94), Seattle, Washington, USA, 31 July – 1 August 1994*, pp. 359–370. Available at: <https://www.aaai.org/Papers/Workshops/1994/WS-94-03/WS94-03-031.pdf> (accessed 16 November 2021).
5. Wu Y.-L., Agrawal D., el Abbadi A. (2000) A comparison of DFT and DWT based similarity search in time-series databases. *Proceedings of the Ninth International Conference on Information and Knowledge Management (CIKM '00), McLean, Virginia, USA, 6–11 November*, pp. 488–495. DOI: 10.1145/354756.354857.
6. Ding H., Trajcevski G., Scheuermann P., Wang X., Keogh E. (2008) Querying and mining of time series data: Experimental comparison of representations and distance measures. *Proceedings of the VLDB Endowment*, Auckland, New Zealand, 23–28 August 2008, vol. 1, no 2, pp. 1542–1552. DOI: 10.14778/1454159.1454226.
7. Kurbalija V., Radovanović M., Geler Z., Ivanović M. (2011) The influence of global constraints on DTW and LCS similarity measures for time-series databases. *Advances in Intelligent and Soft Computing*, vol. 101, pp. 67–74. DOI: 10.1007/978-3-642-23163-6\_10.
8. Dreyer W., Dittrich A.K., Schmidt D. (1994) Research perspectives for time series management systems. *ACM SIGMOD Record*, vol. 23, no 1, pp. 10–15. Available at: <https://dl.acm.org/doi/abs/10.1145/181550.181553> (accessed 16 November 2021).
9. Rozenberg G.S., Shitikov V.K., Brusilovskij P.M. (1994) *Environmental forecasting (time series functional predictors)*. Tolyatti: IEVB RAS (in Russian).
10. Sklyar A.Ya. (2019) Analysis and elimination of noise component in time series with variable step. *Cybernetics and Programming*, no 1, pp. 51–59 (in Russian). DOI: 10.25136/2306-4196.2019.1.27031.
11. Mauleón I. (2021) Aggregated world energy demand projections: Statistical assessment. *Energies*, vol. 14, no 15, pp. 1–13. Available at: <https://www.mdpi.com/1996-1073/14/15/4657> (accessed 16 November 2021). DOI: 10.3390/en14154657.
12. Suganthi L., Samuel A.A. (2012) Energy models for demand forecasting – A review. *Renewable and Sustainable Energy Reviews*, vol. 16, no 2, pp. 1223–1240. DOI: 10.1016/j.rser.2011.08.014.
13. Boßmann T., Staffell I. (2015) The shape of future electricity demand: Exploring load curves in 2050s Germany and Britain. *Energy*, vol. 90, no 2, pp. 1317–1333. DOI: 10.1016/j.energy.2015.06.082.
14. Akca A., Canakoğlu E. (2021) Adaptive stochastic risk estimation of firm operating profit. *Journal of Industrial and Business Economics*, vol. 48, no 3, pp. 463–504. DOI: 10.1007/s40812-021-00184-z.
15. Abbasimehr H., Shabani M. (2021) A new methodology for customer behavior analysis using time series clustering. A case study on a bank's customers. *Kybernetes*, vol. 50, no 2, pp. 221–242. DOI: 10.1108/K-09-2018-0506.
16. Fildes R., Wei Y., Ismail S. (2011) Evaluating the forecasting performance of econometric models of air passenger traffic flows using multiple error measures. *International Journal of Forecasting*, vol. 27, no 3, pp. 902–922. DOI: 10.1016/j.ijforecast.2009.06.002.
17. Bi J.-W., Li H., Fan Zh.-P. (2021) Tourism demand forecasting with time series imaging: A deep learning model. *Annals of Tourism Research*, vol. 90, article no 103255. DOI: 10.1016/j.annals.2021.103255.
18. Meszmer P., Majd M., Prisacaru A., Gromala P.J., Wunderle B. (2021) Neural networks for enhanced stress prognostics for encapsulated electronic packages – A comparison. *Microelectronics Reliability*, vol. 123, article no 114181. DOI: 10.1016/j.microrel.2021.114181.

19. Deshcherevsky A.V., Zhuravlev V.I., Nikolskij A.N., Sidorin A.Ya. (2016) Time series analysis problems with gaps and methods for solving them in the program WINABD. *Geophysical processes and the biosphere*, vol. 15, no 3, pp. 5–34 (in Russian).
20. Plotnikov D.E., Miklashevich T.S., Bartalyov S.A. (2014) Reconstruction of time series of remote sensing data by the method of polynomial approximation in a sliding window of variable size. *Current Problems in Remote Sensing of the Earth from Space*, vol. 11, no 2, pp. 103–110 (in Russian).
21. Lin J., Keogh E., Wei L., Lonardi S. (2007) Experiencing SAX: A novel symbolic representation of time series. *Data Mining and Knowledge Discovery*, vol. 15, no 2, pp. 107–144. DOI: 10.1007/s10618-007-0064-z.
22. Zhukova G., Smetanin Y., Ulyanov M. (2019) Informative symbolic representations as a way to qualitatively analyses time series. Proceedings of the *2019 International Conference on Engineering Technologies and Computer Science: Innovation & Application*, Moscow, Russia, 26–27 March 2019, pp. 43–47.
23. Lin J., Keogh E., Lonardi S., Chiu B. (2003) A symbolic representation of time series, with implications for streaming algorithms. Proceedings of the *8th ACM SIGMOD Workshop on Research Issues in Data mining and Knowledge Discovery*, San Diego, California, USA, 13 June 2003, pp. 2–11. DOI: 10.1145/882082.882086.
24. Nesterenko A.Yu. (2010) Algorithms for finding the lengths of cycles in sequences and their applications. *Fundamental and Applied Mathematics*, vol. 16, no 6, pp. 109–122 (in Russian).
25. Zhukova G.N., Zhukov A.V., Smetanin Yu.G., Ulyanov M.V. (2020) The method of estimating the period of a symbolic periodic sequence with noise, based on the sub-words positions in the sequence. *Modern information technologies and IT education*, vol. 16, no 1, pp. 23–32 (in Russian). DOI: 10.25559/SITITO.16.202001.23-32.
26. Ulyanov M.V. (2020) An approach to identifying the cycle length in noisy character sequences based on the entropy of words. Proceedings of the *III International Scientific and Technical Forum “Modern Technologies in Science and Education”*, Ryazan, 4–6 March 2020, vol. 4, pp. 120–124 (in Russian).
27. Levenshtejn V.I. (1965) Binary codes with corrected dropouts, insertions and character replacements. *Proceedings of the Academy of Sciences*, vol. 163, pp. 707–710 (in Russian).
28. Zhukova G.N., Zhukov A.V., Smetanin Yu.G., Ulyanov M.V. (2019) Stochastic model of noises for periodic symbol sequences. *Modern information technologies and IT education*, vol. 15, no 2, pp. 431–440. DOI: 10.25559/SITITO.15.201902.431-440 (in Russian).

### About the authors

#### Galina N. Zhukova

Cand. Sci. (Phys.-Math.);

Associate Professor, School of Software Engineering, Faculty of Computer Science, National Research University Higher School of Economics, 20, Myasnitskaya Street, Moscow 101000, Russia;

E-mail: galinanzhukova@gmail.com

ORCID: 0000-0003-1835-7422

#### Mikhail V. Ulyanov

Dr. Sci. (Tech.);

Leading Researcher, Laboratory of Scheduling Theory and Discrete Optimization, V.A. Trapeznikov Institute of Control Sciences, Russian Academy of Sciences, 65, Profsoyuznaya Street, Moscow 117997, Russia;

Professor, Department of Algorithmic Languages, Faculty of Computational Mathematics and Cybernetics, Lomonosov Moscow State University, 1, Leninskie Gory, Moscow 119991, Russia;

E-mail: muljanov@mail.ru

ORCID: 0000-0002-5784-9836

[DOI: 10.17323/2587-814X.2021.4.36.49](https://doi.org/10.17323/2587-814X.2021.4.36.49)

# Methods of digital marketing positioning in the global civil passenger aircraft market

**Galina A. Kalugina** 

E-mail: kalugina@gmail.com

**Anatoly V. Ryapukhin** 

E-mail: anatoliiruapukhin@yandex.ru

Moscow Aviation Institute (National Research University)  
Address: 4, Volokolamskoe Shosse, Moscow 125993, Russia

## Abstract

Under conditions of intense competition, the creation of new products and their promotion to the market requires marketing support. The central element of marketing innovation management is marketing positioning. In contrast to the methodological aspect, there is almost no procedural component of positioning in scientific works and publications. Moreover, due attention is not paid to the issues of market segmentation and the allocation of the competitive advantage of such high-tech products as civil aircraft. This determines the relevance of the task of developing a procedure for marketing positioning within the framework of managing the product offer of the aircraft industry enterprises. From the standpoint of a systematic approach, the necessity of studying the adjacent air transportation market is reasonable, as well as the analysis of the correlation of its segmentation with the allocation of target segments of the aircraft market. When conducting a competitive analysis, we propose to consider the market position not only of certain types of civil aircraft, but also of aircraft families. Modeling marketing management shows the role of positioning in the processes of creating, manufacturing and operating aircraft. A distinctive feature of the procedure is the proposed stage of evaluating the effectiveness of the selected product position implementation, which allows us to analyze the feasibility of marketing decisions throughout the product life cycle. Special attention is paid to the issues of information support. The research results presented here reflect the specifics of the marketing activities of aviation industry enterprises and contribute to the development of the conceptual foundations of industrial marketing.

**Key words:** marketing positioning; marketing; segmentation; market research; aviation; aircraft.

**Citation:** Kalugina G.A., Ryapukhin A.V. (2021) Methods of digital marketing positioning in the global civil passenger aircraft market. *Business Informatics*, vol. 15, no 4, pp. 36–49. DOI: 10.17323/2587-814X.2021.4.36.49

## Introduction

The task of marketing positioning of goods is fundamental in managing the product range of a company. Formation of the marketing complex as a system of measures aimed at creation and promotion of competitive and commercially effective goods is based on determining and fixing in the minds of potential consumers of the target segment how the goods differ from the goods in competitive product offerings, i.e. competent marketing positioning of brand, individual model or product offering of the company as a whole. This problem is also relevant for aviation and industrial enterprises which produce such high-tech products as civil aircraft and promote them on the world market of aviation equipment.

There are many publications on passenger aviation and commerce, as well as many books on marketing touching on certain issues of positioning theory and practice. Some authoritative publications extensively cover the topic of marketing and economics in the aviation industry [1–3], but only partially address the problem of marketing positioning. In other scientific publications [4–6], there are the issues of development of the marketing management system, which allows us to study the market and make marketing decisions that increase competitiveness and market promotion of innovative high-tech products. The issues of the procedural component of the process of determining the position of the goods on the market are not given proper attention. Basically the procedural analysis is limited to the enlarged consideration of stages, such as a choice of ways of positioning, allocation of a system of positioning criteria, fixing the position of the goods among comparable competitive goods by means of marketing tools. It can be difficult to use theoretical developments in practice due to the impersonal nature of the object of positioning. Examples illustrating the procedural aspect of positioning most often refer to B2C goods and do not reveal the peculiarities

of the goods and the product market. There are fewer examples which characterize the process of marketing management in industrial markets. But it is the specifics of these markets that determine the complexity of the choice of competitive advantage, which should not be virtual, which is acceptable for consumer markets, where consumer preferences are often predetermined by the emotional impact of advertising, but are quite concrete. Industrial consumers must be able to verify the claimed position of the product against the possibilities of its effective operation. The process of accepting the product position by market actors requires appropriate procedures for detailed competitive analysis, and they should reflect the specifics of the product life cycle. As Hooley [7] noted, market positioning is, above all, a stage in the development of the marketing strategy of the company as a whole.

The question of information support and formation of databases for positioning of industrial goods is also timely.

The aim of this research is to model the procedure of marketing positioning for such high-tech, science-intensive products as civil aviation equipment. The article presents the results of the study of the above-mentioned issues specifically for the market of civil aircraft.

If in most examples of market positioning the implementation of marketing management is based on the method of deduction, this article illustrates how the classical conceptual approach to determining the position of goods in the market by induction expands the possibilities of industry marketing.

## 1. Methods of digital marketing positioning

In the formation of the positioning procedure, solving the following tasks is supposed, aimed at a deeper and more detailed understanding of the possibilities of marketing management in high-tech industries:



- ◆ defining principles of allocation of positioning stages;
- ◆ substantiation of the logical basis for structuring the process of positioning;
- ◆ formation of a system of requirements for the information necessary to implement the processes of research and decision-making at each stage, as well as critical analysis of information sources on the state of the market to form databases.

The fundamental principle of allocating the stages of the positioning procedure is the implementation of comprehensive and systematic approaches, which in solving the problem set involves a comprehensive analysis of factors reflecting the specifics of the product and its operation by the consumer (airlines) in a dynamically changing environment. In this regard, the procedure should consist of a greater number of stages aimed at solving the problems faced by the manufacturer of civil aircraft than we find in the positioning of consumer goods. Accordingly, more detailed elaboration of individual issues is also required when comparing the positioning of goods from the point of view of their operation.

Each stage must represent a finished action and be aimed at obtaining a result which can be evaluated quantitatively using formal indicators or qualitatively on the basis of individual or collective expertise.

When considering the issue of logical structuring of the decision-making process by the aircraft manufacturer on the declared position of its product, the analysis of a large number of variants of the problem solution with changing factors is assumed. All of this will allow us to evaluate the effectiveness of positioning with different combinations of segmentation criteria, attributes of building perception maps, etc. Thus, the procedure should be iterative. All stages should be interconnected logically so that each subsequent stage required solving the problems of the previous stage.

At the same time, the procedure must be interactive, which implies that there are stages of automated calculation, obtaining and processing of information and stages of human decision-making. This requirement is due to the need to consider a large number of options for combining attributes for different markets when building perception maps, as well as to choose the best method of positioning out of the available alternatives. Because of the long life cycle of such a product as civil passenger aircraft, we have to deal with the possibility of further repositioning, which also involves a large amount of analytical work.

In a view of the above, the procedure for positioning civil passenger aircraft can be shown as follows (*Figure 1*).

When considering the issue of information support of the positioning procedure, the fundamental condition is compliance with the information requirements:

- ◆ it must be from reliable sources;
- ◆ it must be available to market participants.

*Table 1* presents an information map of databases required to implement the process of digital modeling of marketing positioning as a part of the management of civil aircraft creation and development.

## **2. Results of industry research**

We consider in detail the implementation of the procedure of positioning aircraft manufacturers' products on the example of modern civil passenger aircraft.

### **2.1. Study of the external environment**

The primary task is the study of the external environment of the aircraft-industrial enterprise and consumer segmentation, which corresponds to the stage 1, in which the research of economic, political, socio-demographic,

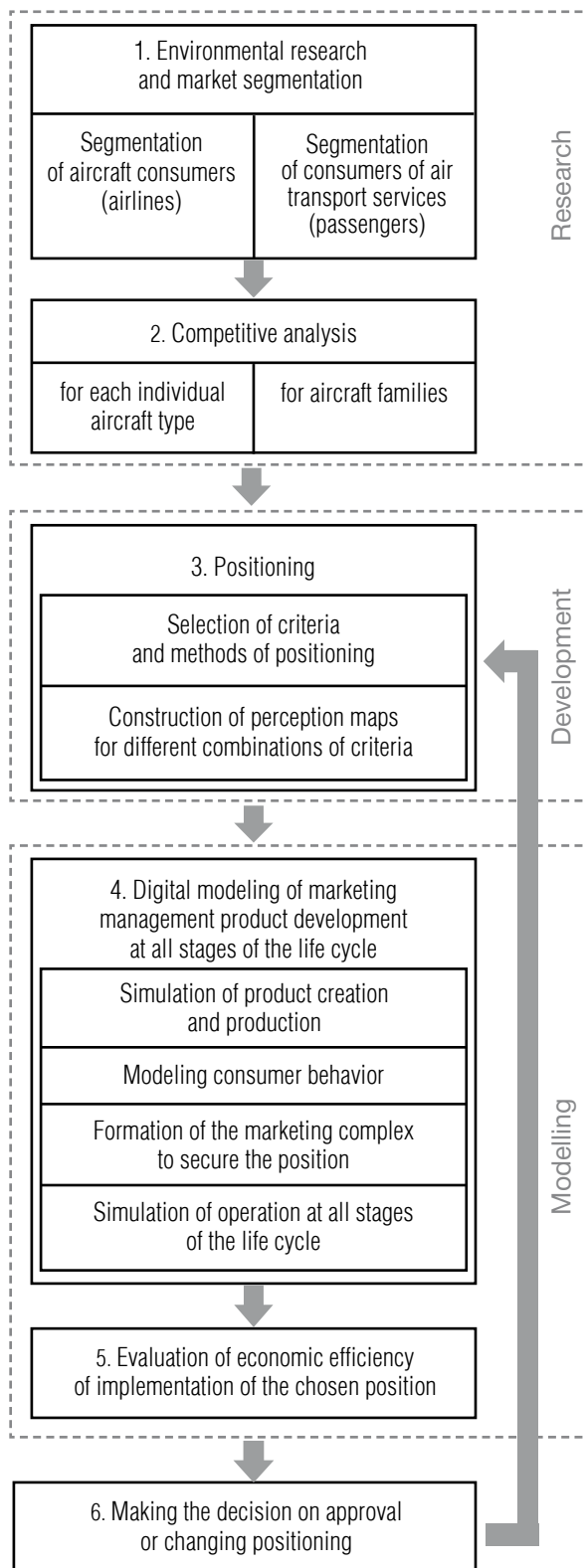


Fig. 1. Enlarged scheme of the marketing positioning procedure within the framework of modeling product creation

environmental and other factors affecting the ability to promote and sell products, as well as market analysis from the perspective of the marketing environment, followed by consumer segmentation. At this stage it is possible to distinguish two directions which are interconnected with each other.

The process of market research of aviation equipment is one of the key processes, the main purpose of which is the analysis of market conditions and forecast of long-term trends of supply and demand for aviation equipment for a period of up to 30 years, which is conditioned by the long life cycle of the product. At the same time, it is necessary to simulate the main prospective changes in the composition of the aircraft fleet both in the whole world and for individual regions, countries and airlines. The forecast should reflect the flight-technical, economic and operational characteristics of all types and classes of aircraft.

To solve this problem, one should create a database of the world aircraft fleet, which contains comprehensive information on passenger aircraft: model, type and manufacturer of aircraft, date of manufacture, status and history of operation, including the resource indicators of the airframe and power plant, maintenance history, owner and operator of the aircraft, configuration with basic flight and technical characteristics, estimated cost, etc. Additionally, a database of commercial aircraft orders, including information on realized, concluded and probable (option) contracts, is required. The study uses sorting, sampling, classification, decomposition and extrapolation methods.

Consideration of the related market of air transportation is the second important direction of the study of the external environment. The main task is also the analysis of market conditions and forecasts of development trends, analysis of development of airlines-operators of aircraft, as well as the study of the typology and behavior of end users – passengers. In the course of the study, it is necessary

Table 1.

**Database information map**

Stage/substage	Type and method of information collection	Source of information
External environment research	Statistical data Expert evaluation	World Bank, United Nations, Central Intelligence Agency, Eurostat, Federal State Statistics Service (Russia), etc.
Segmentation of consumers (airlines)	Statistical data Interview/survey Expert evaluation	Federal Air Transport Agency, International Civil Aviation Organization (ICAO), International Air Transport Association (IATA), Statistical and analytical organizations (Cirium, CAPA, Airfinance, etc.)
Segmentation of end users (passengers)	Statistical data Interview/survey Expert evaluation	ICAO, IATA, airlines, airports
Competitive analysis	Statistical data	Documentation of aircraft manufacturers, aviation authorities, certification organizations, analytical agencies, etc.
Positioning	Calculated analytical data Expert evaluation	Database on market research results
Marketing modeling	Calculated analytical data Expert evaluation	Database on market research results
Assessment of economic efficiency	Calculated analytical data Expert evaluation	Database on market research results
Making a decision to approve or change the position	Calculated analytical data Expert evaluation	Database on market research results

to use statistical databases containing data on the volume of air transportation in different perspectives, namely by routes, flights, capacity and passenger flows. Information will also be required on flights between airports with classification by aircraft type, operator, class of flights themselves, as well as financial indicators allowing for profitability and cost analysis in their performance. Such an extensive analysis serves as a basis for forecasting traffic volume trends with a detailed classification by class and type of aircraft, by operator and by geographical distribution.

Studies of secondary markets, such as railroads, automobile and other types of passenger transportation, are also conducted in addition. High-speed rail transport is becoming a significant competitor to air transportation, and on certain routes it has become equal to air

transportation in terms of availability, speed, travel time and volume. As a rule, high-speed railroads are created on Russian routes with very high regular passenger traffic and cannot occupy the entire share of the local market. Typical examples are rail transportation in Europe, Japan, China, as well as in Russia on the Moscow – St.Petersburg route and other new destinations.

End user research is aimed at identifying the factors that determine trends in air travel and involves analysis of demographic and socio-economic components at the level of a detailed study of the population of countries and regions of the world in the framework of the consumer behavior concept. At the same time, the issues of studying the lifestyle and economic situation of the population as preconditions for the change in effective demand for air trans-



portation are relevant. It is also necessary to take into account the geographical features of regions and individual cities, the level of development of industry and tourism.

As a result of the study, it becomes possible to calculate and forecast the index of population mobility up to individual settlements, as well as to identify the correlation of the studied indicators and their impact on the projected volume of transportation with detail down to individual routes. The solution of the above tasks requires a large volume of statistical data on many quantitative and qualitative indicators. Correlation and regression analysis is appropriate to identify the relation between the factors and assess the degree of interdependence of the criteria selected for analysis, while the expert assessment is appropriate to determine the weighting parameters of the impact of input data on the final indicators.

In the study of the external environment, consideration of the global political situation is very important. Despite the overall globalization of developed countries, which are the main and in many segments the only producers of aviation equipment, one can observe not only examples of political and economic sanctions against developing countries and third world countries (such as the ban on the supply of Western aircraft to Iran and Cuba), but also an open struggle between developed countries, using various political, customs and economic barriers in the aviation equipment and air transportation market. The most attractive and capacious United States market, where conflicts between regional aircraft producers (Embraer, Bombardier) [8, 9] and long-haul aircraft producers (Airbus, Boeing) [10, 11] periodically escalate, is a vivid example of competition with political and economic methods in the segment of regional and long-haul passenger aircraft. The study requires not only factual data on existing constraints or barriers, but also expert analysis and forecasting of developments in the field of politics and economics.

## 2.2. Consumer segmentation

One part of the study of the external environment is customer segmentation. When considering civilian passenger aircraft as a B2B commodity, it is necessary to segment not only customers (i.e. airlines), but also the segmentation of end consumers - passengers, as the market for aircraft equipment and the market for passenger air transportation are closely linked. Airline market trends are directly reflected in aircraft market trends.

Segmentation of the markets of aviation equipment and air transportation involves the following tasks:

- ◆ formation of a database structure on the characteristics of airlines and passengers;
- ◆ collection of necessary information and filling the database;
- ◆ analysis of the characteristics and needs of potential buyers;
- ◆ selection of criteria for segmentation of consumers of the market of aviation equipment and air transportation market;
- ◆ implementation of market segmentation, taking into account the correlation between the markets of aircraft equipment and air transportation;
- ◆ clarification of boundaries of market segments of aircraft equipment, taking into account the related market of air transportation;
- ◆ selection of target segments of both markets.

Due to the fact that the product in question is high-tech and technically complex and belongs to the B2B market, it has a huge number of consumer characteristics analyzed by the customer in product selection, as well as considered by the manufacturer in segmentation and positioning. As Guzhva [12] noted, research objects related to B2B products require a deeper analytical approach in market research,

competitive analysis and customer segmentation. As a part of the segmentation, one should create a multidimensional matrix, where each criterion under consideration forms a dimension consisting of a sequence of values. *Figure 2* shows the conceptual model of the formation of two-dimensional segmentation matrices reflecting the relation between segments of the airline and aviation market. Each matrix has a multidimensional structure ( $N = 1, \dots, i$ ;  $M = 1, \dots, ii$ ) with a large array of data ( $N_i, M_{ii}$ ). In practice, the formation of this matrix when selecting target segments is associated with the processing of a huge amount of information.

As a result of data processing, the target segments and groups of segments in multidimensional matrices should be identified, the correlation between segments of different matrices

should be determined (*Figure 2*) and a characterization of each multidimensional segment should be given.

### 2.3. Competitive analysis

Competitive analysis, an enlarged conceptual model of which is presented in *Figure 3*, includes the study of functional, generic and subject competitors, which corresponds to stage 2 of the marketing positioning procedure. The analysis includes the study of consumer characteristics of services of functional and generic competitors, as they are necessary for further comparison (transportation parameters, price, etc.). When analyzing the subject competitors (similar passenger airplanes), the study of not only consumer but also techni-

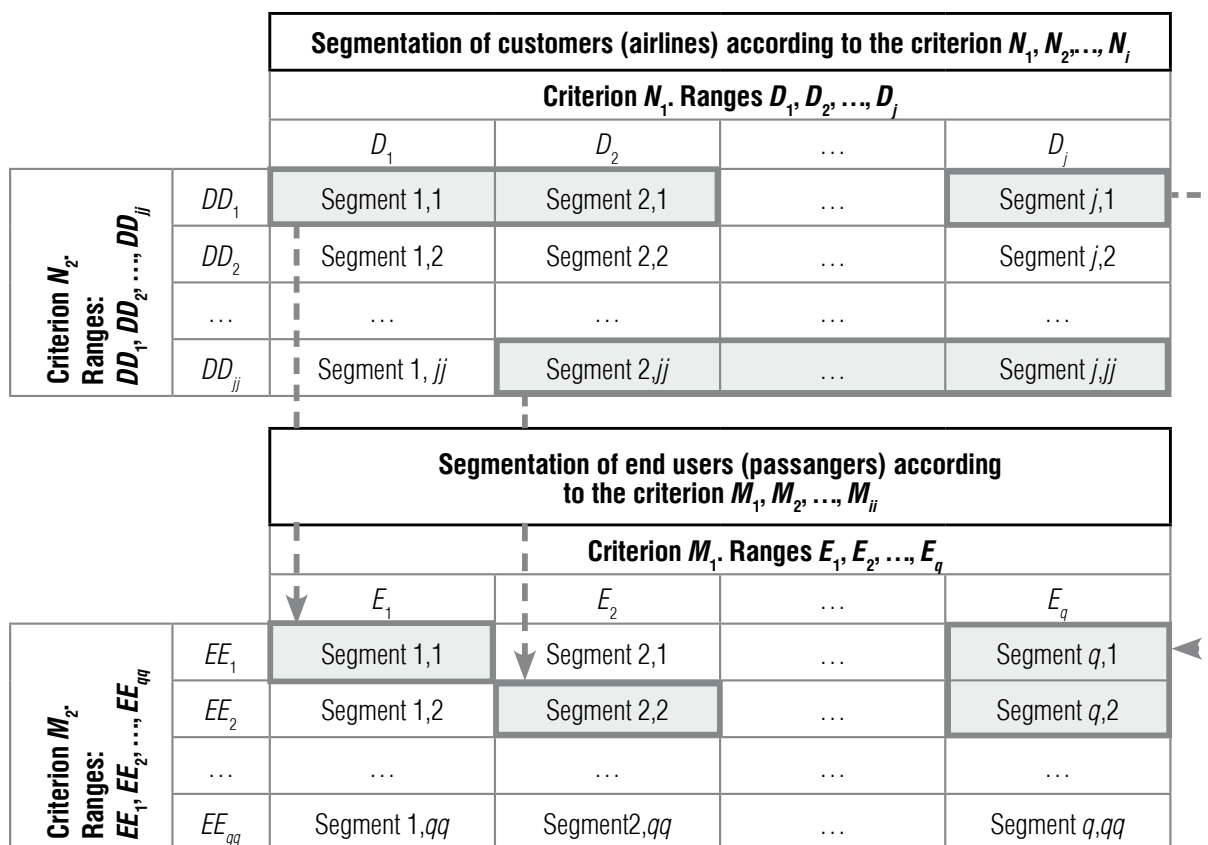


Fig. 2. Correlation of customers' segmentation of aviation equipment (airlines) and consumers of air transportation (passengers)

cal parameters is carried out for a more complete characterization of competitive products and forecast of their development. At the same time, not only airplanes, but their families are analyzed, with different types and modifications of airplanes of one type grouped into a family. The unification has higher consumer characteristics and allows coverage of a larger market segment. The topic of aircraft families is discussed in detail by Kalugina [13].

In the course of competitive analysis, a database of product characteristics is formed, their interrelation is determined, derived indicators are calculated as a ratio of initial indicators and parameter indicators which require more complex calculations (economic and operational characteristics) are singled out. According to the results of the analysis, the competitive options are compared with the designed aircraft by the entire set of parameters.

#### 2.4. Positioning an aircraft family

The marketing positioning of a new aircraft or family of aircraft corresponds to stage 3 of the enlarged positioning procedure. Based on

the results of environmental research and segmentation, using the characteristics of competitive products and the results of comparative analysis of aircraft and their families and taking into account the forecast of development for the coming 20–40 years, the choice of positioning criteria is made. The information model of the process of marketing positioning of airplanes is reflected in *Figure 4*.

Positioning criteria can be any consumer characteristics of the products that best reflect the competitive advantages of the aircraft from the point of view of consumers. The selected criteria are then combined into pairs to build perception maps. The perception map or positioning map is a graph that visually shows the positions of certain brands in terms of the most important product characteristics for the consumer. The map is built using two defining parameters as *X* and *Y*-axes. The most common parameters for comparing products and building the brand perception map are the following: quality, price, safety, reliability and functional characteristics. The main advantage of the brand position visualization method with perception maps is that it is very easy to construct and interpret.

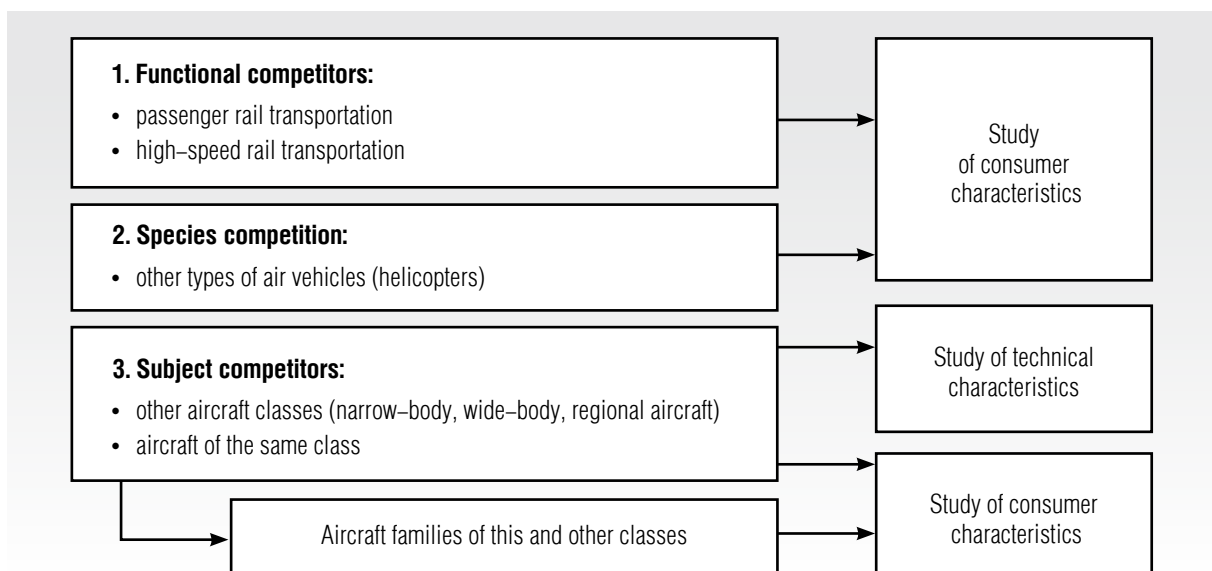


Fig. 3. Conceptual model of competitive analysis

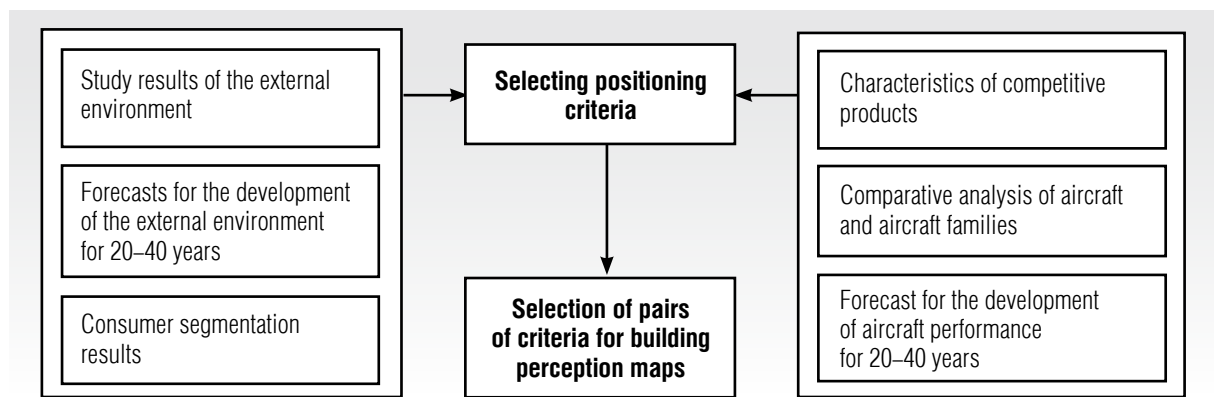


Fig. 4. Information model of aircraft marketing positioning

It is important to note that for perception maps even indicators in which products are inferior to competitors can be selected, in case:

- ◆ choice is due to the high correlation of the indicators of the selected pair;
- ◆ position of the product on the perception map is above the “line” of the market trend;
- ◆ position of the product is insignificantly inferior to “recognized” competitors in the market which are already a success;
- ◆ positioning on the perception map allows one to occupy a larger area than competitors when comparing aircraft families, which is one of the positioning objectives.

### 2.5. Examples of perception maps

Figure 5 shows a conditional example of a perception map to illustrate the marketing positioning of commercial aircraft according to the criteria of range (criterion 1) and passenger capacity of the aircraft (criterion 2).

The perception map identifies the positions of 15 aircraft, grouped by manufacturer and the families to which they belong. There are four manufacturers (A, B, C, D) with their families of aircraft (1, 2, 3, 4, 5). The map shows a trend line, characterizing the average trend of

dependence of criterion 2 on criterion 1 calculated on the basis of analysis of the characteristics of the aircraft on the market, and a corridor of deviation from the trend line of  $-10\%$  and  $+10\%$ . The advantages and disadvantages of aircraft positioning are not considered for each aircraft separately by criterion 1 and 2, but by its location within the family and the manufacturer’s product offering as a whole, as well as depending on the positions of competitors.

Positioning families, as opposed to positioning individual aircraft, enhances the ability to meet demand in more market segments, thereby improving the company’s commercial performance.

Manufacturer A offers Family 1 and Family 2, thereby covering multiple market segments. Although Family 1 is inferior to its competitors, it is nevertheless in the trend of the market. Family 2 is noticeably above the trend line, which strengthens the position of Producer A.

Manufacturer B offers one Family 3, consisting of four aircraft. So, the family occupies more market segments than its competitors.

Manufacturer C offers one Family 4, which does not occupy as many market segments as Family 3, but outperforms it in criteria 1 and 2.

Manufacturer D offers Family 5 consisting of only two aircraft. This family is below the trend

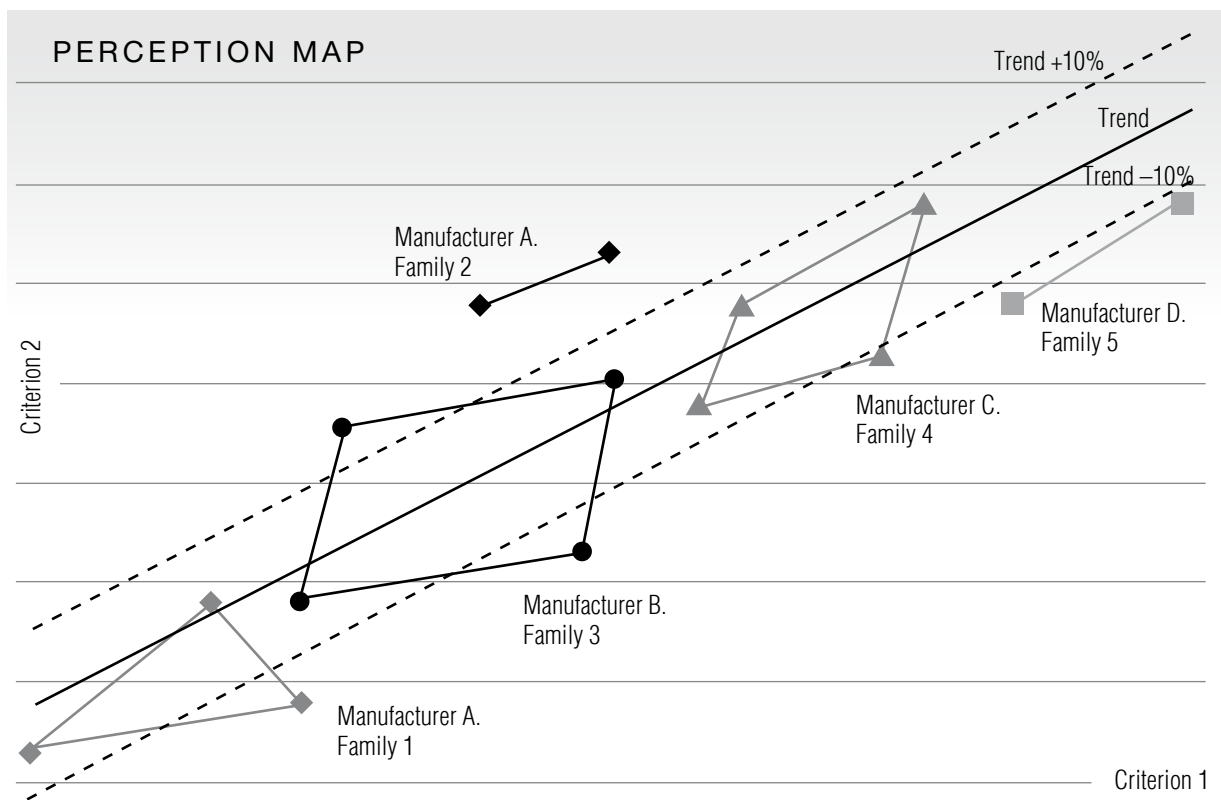


Fig. 5. Conditional example of a perception map

line, but it targets its market segments, outperforming its competitors on criterion 1.

In terms of market coverage of segments covered, Manufacturer A leads the overall product offering. Manufacturer B leads in terms of the number of covered segments within a single product family.

Figure 6 shows a real-world example of a perception map for the modern wide-body twin-engine segment excluding discontinued aircraft (Boeing 767, Airbus A330ceo (200/300) and those in the design phase (for example, CRJ929). Boeing 777, which are being replaced by the Boeing 777x, are also not considered.

As we can see on the perception map, A330-900, A350-900/1000 and 787-9 are in trend. They are the most in-demand at the moment in the selected segment, with the largest number of orders placed on them. 787-8/10 air-

craft are less in demand, but increase the market coverage within the family. The situation is similar with A330-800 aircraft. Demand for them is low, but they expand the family's ability to capture the market. The 777x family (777-8/9) aircraft are more spacious and noticeably different in other characteristics, so they are not generally considered direct competitors to the above aircraft. However, on the perception map, the increased capacity can be presented as an advantage.

## 2.6. Digital simulation of product development

Figure 7 schematically shows stage 4, which involves multivariate marketing modeling of the creation and development of the product, taking into account the selected position based on the data obtained in the first and second

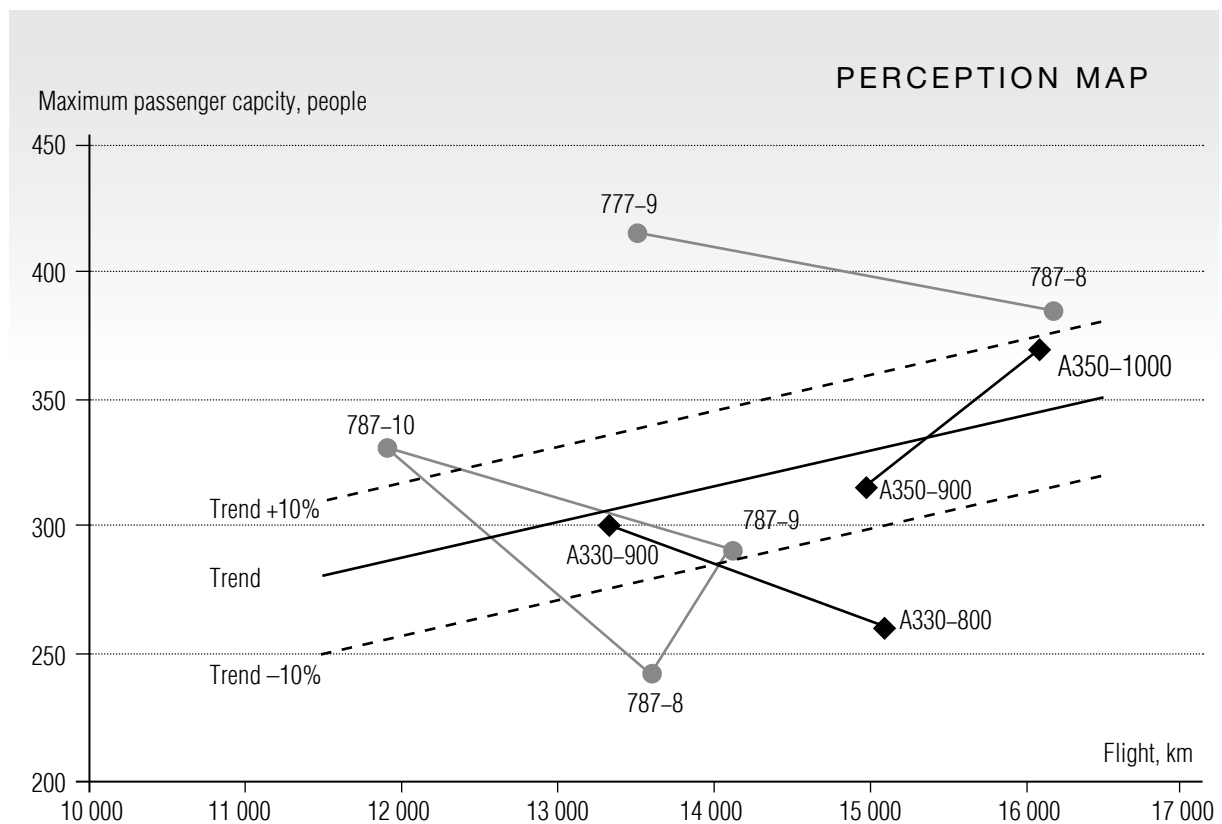


Fig. 6. Example of a perception map for Airbus and Boeing aircraft

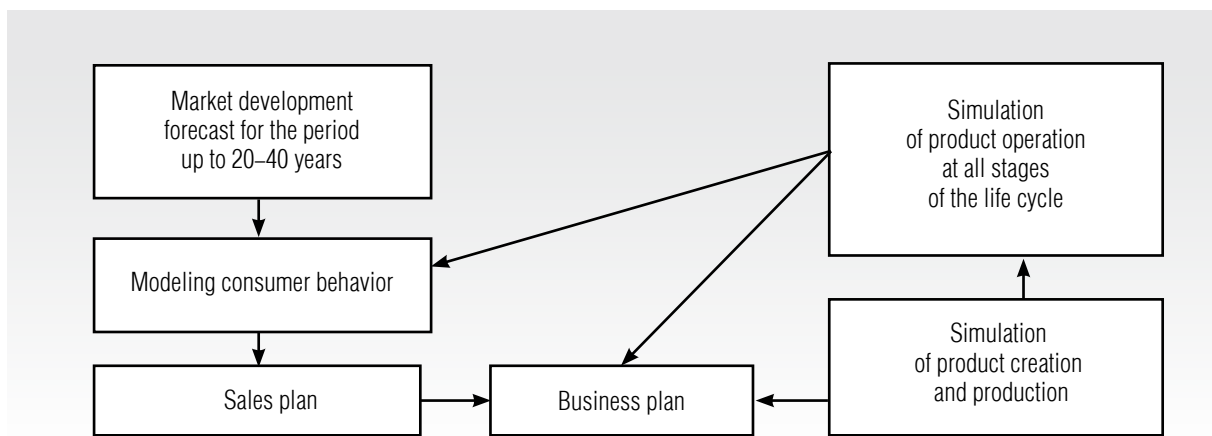
stages. The final result of the stage is a business plan of the project to create a new product, which takes into account the following:

- ◆ ales plan made taking into account the modeling of consumer behavior in the projected long-term conditions within the formed marketing positioning of the product. It is necessary to take into account the different types of business models of airlines, as described in the work by Cook [14];
- ◆ results of modeling the creation and production of the product based on market requirements that provide a competitive advantage;
- ◆ results of modeling the operation of the product at all stages of the life cycle based on the evaluation of the economics of airlines – target potential consumers of the product. As Riordan [15] noted, the duration of the life cycle of the aircraft as a product is an important feature that determines the specifics of organizational processes. The results obtained are taken into account when modeling consumer behavior.

Stage 5 in continuation of stage 4 involves assessing the cost-effectiveness of implementing the chosen position of the new aircraft as a part of business planning. To this end, a base of business plan options necessary for further decision-making is formed.

Stage 6 involves a decision to approve or modify the position. In the second case, a return to stage 3 is required, and, if necessary, to stage 1 or 2 for additional research on the external environment, customer segmentation and competitive analysis.





*Fig. 7.* Digital modeling of marketing management of product development at all stages of the life cycle

### 3. Discussion of marketing positioning

Marketing positioning of passenger commercial aircraft by the criteria pair “flight range – passenger capacity” that was described in previous paragraph illustrates the possibilities of demand management in a fiercely competitive environment. In this case, even not good enough positions of the product, presented on the map of perception, can be used by the manufacturer as an advantage, for example, in the development of pricing policy, providing a better price-quality ratio. The issue of pricing is considered in detail by Kalugina [16].

It is important to note that during the construction of perception maps one should not always strive to achieve the maximum indicators according to the chosen positioning criteria. It is necessary to meet the needs of the chosen target segment. In aviation, there are many examples of targeting a narrow market segment. For example, when segmenting customers by geography, as Czinkota [17] discovered, E-Jets, Canadair Regional Jet, Mitsubishi Regional Jet on the United States market and ARJ21 and SSJ-100 on national markets were successfully positioned. In line with the type of business model, the 737 and

A320 families are positioned for low-cost airlines, while 747 and A330 wide-body jets are positioned for charter airlines. As Clark [18] wrote, the A380 is intended for use by airlines in the Middle East and Asia to serve flights between these regions on lines with very high traffic. The 747-400D is intended for short-haul flights on high-traffic lines in the Japanese market. It should be noted that other methods of positioning Airbus and Boeing aircraft are used in parallel with this method. The method of positioning based on building of perception maps with criteria which reflect specific requests of target segments to the aircraft is also used by airlines on the related market of air transportation as mentioned by Malaval [19].

### Conclusion

This article deals with the issues of modeling marketing positioning in the process of creating and producing such high-tech, science-intensive products as civil aircraft. The emphasis is placed on the procedural component, which is less represented in theoretical developments and scientific works. At the same time, the procedure of marketing positioning is reflected in the framework of modeling the innovative pro-

cess of creating aviation equipment and its operation at all stages of the life cycle, which characterizes the novelty of the very statement of the management problem.

In developing the procedure according to the formulated principles, six main stages aimed at market research and positioning of the product offer of the aviation industry enterprise were identified.

In connection with the specifics of civil aircraft as a commodity, it was proposed to study not only the market of aircraft, but also the related market of air transportation, which depends on the condition and development of the demand for aircraft. In the allocation of target segments, it is proposed to analyze the correlation segmentation of these markets, which is not typical of the markets B2C, but necessary for most industrial markets.

The peculiarity of the procedure is also the proposal of the stage of assessment of the effectiveness of implementing positioning of a new product within the framework of business planning.

The question of information support of positioning, including the definition of categories of data necessary for decision-making, information sources, as well as methods of collection and processing of information, is aimed at the possibility of implementing the proposed procedure in a particular environment.

The significance of the results of the research is determined by the development of the concept of marketing positioning. Within the limits of the classical methodical approach to definition of a position of the goods, the procedure offered here reflects the specificity of the market of civil planes and expands possibilities of branch marketing. ■

### References

1. Scheinberg R. (2017) *The commercial aircraft finance handbook*. London: Routledge. DOI: <https://doi.org/10.4324/9780203713303>.
2. Spreen W.E. (2016) *Marketing in the international aerospace industry*. London: Routledge. DOI: <https://doi.org/10.4324/9781315249681>.
3. Vasigh B., Gorjidoz J. (2016) *Engineering economics for aviation and aerospace*. London: Taylor & Francis. DOI: <https://doi.org/10.4324/9781315644257>.
4. Zaripov R.N., Murakaev I.M., Novikov S.V., Ryapukhin A.V. (2020) Corporate structure for innovative enterprises. *Russian Engineering Research*, vol. 40, no 2, pp. 137–139. DOI:10.3103/S1068798X20020239.
5. Tikhonov G.V., Lavrova L.A., Kolosova V.V., Zemlyanskaya N.B., Kazakova N.V. (2020) Marketing as an effective control of progressive innovations. *TEM Journal*, vol. 9, no 3, p. 1094. DOI: <https://doi.org/10.34069/AI/2021.41.05.18>.
6. Ryapukhin A.V., Kabakov V.V., Zaripov R.N. (2019) Risk management of multimodular multi-agent system for creating science-intensive high-tech products. *Espacios*, vol. 40, no 34, p. 19. DOI: <http://dx.doi.org/10.21533/pen.v9i4.2388>.
7. Hooley G.J., Piercy N., Nicoulaud B. (2008) *Marketing strategy and competitive positioning*. New York: Pearson Education.
8. Embraer S.A., the Official Web site (2021) *Media and downloads*. Available at: <https://www.embraercommercialaviation.com/media-downloads/> (accessed 01 April 2021)
9. Bombardier Inc., the Official Web site (2021) *Airport emergency publications*. Available at: <https://u.to/7JCEGg> (accessed 01.04.2021).

10. Airbus SE, the Official Web site (2021) *Aircraft characteristics, airport operations and tech data*. Available at: <https://www.airbus.com/aircraft/support-services/airport-operations-and-technical-data/aircraft-characteristics.html> (accessed 01.04.2021).
11. The Boeing Company, the Official Web site (2021) *Airplane characteristics for airport planning*. Available at: [https://www.boeing.com/commercial/airports/plan\\_manuals.page](https://www.boeing.com/commercial/airports/plan_manuals.page) (accessed 01.04.2021).
12. Guzhva V. S., Raghavan S., D'Agostino D. J. (2018) *Aircraft leasing and financing: Tools for success in international aircraft acquisition and management*. Amsterdam: Elsevier.
13. Kalugina G.A. (2018) Civil passenger aircraft prices determination features. *Economics and management in mechanical engineering*, no 3, pp. 30–34 (in Russian).
14. Cook G.N., Billig B. (2017) *Airline operations and management. Management textbook*. London: Routledge. DOI: <https://doi.org/10.4324/9781315299594>.
15. Riordan D., Kundu A. K., Price M. A. (2016) *Theory and practice of aircraft performance*. New York: John Wiley & Sons Limited. DOI: <https://doi.org/10.1017/S0001924000091636>.
16. Kalugina G.A. (2013) Marketing research of aviation-industrial enterprises in the process of an aircraft family development. *MAI Works*, no 69 (in Russian).
17. Czinkota M.R., Ronkainen I.A. (2013) *International marketing*. 10th edition. Boston, MA: Cengage Learning. DOI: <https://doi.org/10.1080/08975931003644630>.
18. Clark P. (2017) *Buying the big jets - fleet planning for airlines*. London: Routledge. DOI: <https://doi.org/10.4324/9781315195780>.
19. Malaval P., Benaroya C., Aflalo J. (2013) *Aerospace marketing management: A handbook for the entire value chain*. Berlin: Springer Science & Business Media. DOI: <http://dx.doi.org/10.1007/978-3-319-01354-1>.

### About the authors

#### **Galina A. Kalugina**

Ph.D. (Econ.);

Associate Professor, Department of Management and Marketing of High-tech Industries, Moscow Aviation Institute (National Research University), 4, Volokolamskoe Shosse, Moscow 125993, Russia;

E-mail: [kalugina@gmail.com](mailto:kalugina@gmail.com)

ORCID: 0000-0002-2085-8851

#### **Anatoly V. Ryapukhin**

Senior Lecturer, Department of Industrial Economics, Moscow Aviation Institute (National Research University), 4, Volokolamskoe Shosse, Moscow 125993, Russia;

E-mail: [anatoliiruapukhin@yandex.ru](mailto:anatoliiruapukhin@yandex.ru)

ORCID: 0000-0002-2208-6875

# Using Fishburne's sequences in suitable modeling used for sample data\*

**Anatoliy V. Sigal** 

E-mail: ksavo3@cfuv.ru

V.I. Vernadsky Crimean Federal University

Address: 4, Prospekt Vernadskogo, Simferopol 295007, Russia

## Abstract

This article deals with probabilistic and statistical modeling of managerial decision-making in the economy based on sample data for the previous periods of time. For better definition, the study is limited to Markowitz's models in the problem of finding an effective portfolio of the field in the third information situation. The third information situation is a widespread decision-making situation and is characterized by the fact that the decision-maker sets, according to his opinion, are a linear order relation on the components of an unknown probabilistic distribution of the states of the economic environment. Often, from the point of view of the decision-maker, the components of an unknown probability distribution of the states of the economic environment must satisfy a partially reinforced linear order relation. As a result, the use of traditional statistical estimates turns out to be impossible, while the following question arises, which is practically not studied in the scientific literature. In this case, what formulas should be used to find statistical estimates and, above all, estimates of unknown probabilities of the state of the economic environment? As an estimate of an unknown probability distribution, we proposed to use the Fishburne sequence that satisfies all available constraints, while corresponding to the opinion of the decision maker and the linear order relation given by him. Fishburne sequences are a generalization of the well-known Fishburne formulas. It is fundamentally important that any Fishburne sequence satisfies a simple linear order relation, and under certain conditions, a partially strengthened linear order relation. Particular attention is paid to the entropic properties of generalized Fishburne progressions, which represent the most important class of Fishburne sequences, as well as the use of generalized Fishburne progressions to take into account the opinion of the decision maker. Such a scheme for estimating an unknown probability distribution has been developed, which makes it possible to achieve the correctness of probabilistic and statistical modeling, as well as appropriate consideration of the opinion of the decision-maker, uncertainty and risk.

---

\* The article is published with the support of the HSE University Partnership Programme

**Key words:** Fishburne sequence; managerial decision-making; Markowitz model; linear order relation; entropy; Fishburne progression.

**Citation:** Sigal A.V. (2021) Using Fishburne's sequences in suitable modeling used for sample data. *Business Informatics*, vol. 15, no 4, pp. 50–60. DOI: 10.17323/2587-814X.2021.4.50.60

## Introduction

In game theory modeling, a special role is played by the statistical game [1], which is a game of two participants: a decision-maker (DM) who consciously chooses his behavior, and “nature” (the economic environment in the case of economic modeling), which randomly turns out to be in its possible states.

In the case of economic modeling, it is often necessary to establish what type of order relation is performed on a set of states of the economic environment which is characteristic of the third information situation (IS) [1, p. 13].

Modeling of the economy requires taking into account a number of specific features inherent in the economy, primarily uncertainty and economic risk (for example, [2]). Failure to take into account these features entails management inefficiency. The appropriateness and correctness of modeling is largely determined by the degree of consideration of uncertainty, economic risk and the opinion of the DM about the specifics of the decision-making situation.

Methods and models of probability theory and mathematical statistics are widely used in economic modeling. A classic example of probabilistic-statistical modeling of the economy is the portfolio theory, which began with the works of Harry Markowitz [3, 4].

According to Markowitz's approach, the rate of return (profitability) of an arbitrary asset / portfolio is characterized by a corresponding random variable (RV), the values of which are determined by the conditions in which the

economic environment may be (in fact, the stock market). For the sake of certainty and convenience, we will assume that the set of possible states of the economic environment is finite. In this case, the probability distribution of the states of the economic environment  $\mathbf{q} = (q_1; \dots; q_n)$  is a vector whose components are non-negative numbers satisfying the normalization property, while the RV characterizing the rate of return of the selected asset is a discrete RV (DRV). In practice, the available sample data are used as possible values of these DRV, namely the previously observed values of the corresponding profit margins, while it is customary to use traditional point estimates as estimates of the numerical characteristics of these DRV, the values of which are calculated based on the use of a uniform law, i.e. a vector  $\mathbf{q}^* = (q_1^*; \dots; q_n^*)$ , where  $q_1^* = \dots = q_n^* = \frac{1}{n}$ , as it estimates of the probability distribution.

Suppose that the distribution  $\mathbf{q} = (q_1; \dots; q_n)$  satisfies a linear relation of order (LRO) of one type or another. The main types of LRO were studied by Peter Fishburn [5–7], (and also reviewed, for example, in [1, pp. 77–80]. Fishburne formulas and their generalizations, called Fishburne sequences in the article, are used to estimate the corresponding distributions. For example, in [8] the author proposed the use of Fishburne sequences to bring generalized models of the optimal portfolio search problem to the classical (traditional) Markowitz model.

The purpose of this study is to develop a scheme for constructing such an estimate of the probability distribution which makes it possible to achieve the correctness of probabil-

istic and statistical modeling, as well as to best take into account the opinion of the DM about the type of LRO, to which the elements of the set of states of the economic environment obey, including for cases where the use of traditional point estimates is impossible because the most characteristic estimate of the probability distribution should differ from the uniform law.

The main objectives of the study are to develop:

- ◆ concepts of suitable modeling of managerial decision-making in the economy based on sample data, in particular, portfolio decision-making in the field of the third IP;
- ◆ a method for estimating the probability distribution based on the use of Fishburne sequences, primarily generalized Fishburne progressions;
- ◆ schemes for constructing such an estimate of the probability distribution that best takes into account the opinion of the DM.

### 1. Formulation of the research task

We introduce the following designations:  $r_{ij}$  – the value of the  $i$ -th rate of return of the asset in conditions when the economic environment was in  $j$ -th state,  $i = \overline{1, k}$ ,  $j = \overline{1, n}$ ;  $x_i$  – the share of the asset in the portfolio,  $i = \overline{1, k}$ ;  $\mathbf{x} = (x_1; \dots; x_k)$  – portfolio (more precisely its structure);  $R_i$  – DRV, characterizing the rate of  $i$  return of the asset,  $i = \overline{1, k}$ ;  $R_x = \sum_{i=1}^k R_i \cdot x_i$  – DRV, characterizing the rate of return of the portfolio  $\mathbf{x}$ ;  $\mathbf{q} = (q_1; \dots; q_n)$  – probability distribution of the states of the economic environment;  $m_i = \mathbf{M}(R_i) = \sum_{j=1}^n r_{ij} \cdot q_j$ ,  $m_x = \mathbf{M}(R_x)$  – mathematical expectations of the corresponding DRV;  $\sigma_i^2 = \mathbf{D}(R_i) = \sum_{j=1}^n r_{ij}^2 \cdot q_j - m_i^2$ ,  $\sigma_x^2 = \mathbf{D}(R_x)$  – variance of the corresponding DRV;  $c_{il} = \text{cov}(R_i; R_l) = \sum_{j=1}^n r_{ij} \cdot r_{lj} \cdot q_j - m_i \cdot m_l$  – covariance between the specified DRV,  $i = \overline{1, k}$ ,  $l = \overline{1, k}$ .

If the exact true values  $r_{1l}, \dots, r_{in}$ , of the corresponding DRV  $R_i$  are known and the first IC takes place when the probabilities ( $q_1; \dots; q_n$ ) of states are known, then the values of the numerical characteristics  $m_i$ ,  $\sigma_i^2$ ,  $c_{ij}$ , can be found by the above formulas, and the numerical characteristics  $m_x$  and  $\sigma_x^2$  DRV  $R_x = \sum_{i=1}^k R_i \cdot x_i$  of the  $\mathbf{x}$  portfolio are functions of fractions  $x_1, \dots, x_k$ . The classical Markowitz model (the model of the problem of finding an effective portfolio in the field of the first IS) can be presented in the following form:

$$m_x = \sum_{i=1}^k m_i \cdot x_i \rightarrow \max_{\mathbf{x}}, \quad (1)$$

$$\sigma_x^2 = \sum_{i=1}^k \sum_{l=1}^k c_{il} \cdot x_i \cdot x_l \rightarrow \min_{\mathbf{x}}, \quad (2)$$

$$\sum_{i=1}^k x_i = 1, \quad (3)$$

$$x_i \geq 0, \quad i = \overline{1, k}. \quad (4)$$

An effective portfolio (in the Markowitz model) is usually called a portfolio whose structure is a Pareto optimal solution to the problem (1)–(4).

The main task of the DM (investor) is to find the structure of the optimal portfolio, i.e. the structure of such an effective portfolio, which, according to the DM, has all the desired properties, first of all, the best combination of the values of its numerical characteristics.

Let us now assume the possible values  $r_{1l}, \dots, r_{in}$ , of an arbitrary DRV  $R_i$  are known, and the probabilities  $q_1, \dots, q_n$  are not known, then the numerical characteristics  $m_i$ ,  $\sigma_i^2$ ,  $c_{ij}$  are probability functions  $q_1, \dots, q_n$ , and the numerical characteristics  $m_x$  and  $\sigma_x^2$  DRV  $R_x = \sum_{i=1}^k R_i \cdot x_i$  of the  $\mathbf{x}$  portfolio are probability functions  $q_1, \dots, q_n$  and fractions  $x_1, \dots, x_k$ . This assumption corresponds to the approach used in practice when using available sample data, i.e. when the values  $r_{1l}, \dots, r_{in}$  are used as the observed val-



ues of the  $i$ -th rate of return of the asset, while unknown values of numerical characteristics  $m_i, \sigma_i^2, c_{ij}$ , as a rule, evaluate them with traditional point estimates, i.e. the values of the numerical characteristics of the sample

$$\bar{r}_i = \frac{1}{n} \cdot \sum_{j=1}^n r_{ij}, i = \overline{1, k},$$

$$\sigma_i^{*2} = \bar{r}_i^2 - \bar{r}_i^2 = \frac{1}{n} \cdot \sum_{j=1}^n r_{ij}^2 - \bar{r}_i^2, i = \overline{1, k},$$

$$c_{il}^* = \frac{1}{n} \cdot \sum_{j=1}^n r_{ij} \cdot r_{lj} - \bar{r}_i \cdot \bar{r}_l, i = \overline{1, k}, l = \overline{1, k},$$

accordingly. We emphasize, now  $r_{ij}$  – this is the value of the  $i$  rate of return of the asset observed in the period (moment) of time, and to calculate the values  $r_i, \sigma_i^{*2}, c_{il}^*$ , a uniform law was used as an estimate of the probability distribution.

The use of traditional point estimates may contradict the opinion of the DM about the significance (in formativeness) of various points in time. The opinion of the DM about the significance of various moments of time must necessarily be reflected in the generalized Markowitz model: in this case, the system of constraints of the task of finding an effective portfolio may contain such constraints for possible values  $q_1, \dots, q_n$  components of the probability distribution that the uniform law does not satisfy.

The generalized Markowitz model of the problem of finding an effective portfolio in the field of the third OF the simple gallop can be written in the following form [8]:

$$H(\mathbf{q}) = -\sum_{j=1}^n q_j \cdot \ln q_j \rightarrow \max_{\mathbf{q}}, \quad (5)$$

$$m_x = \sum_{i=1}^k m_i \cdot x_i \rightarrow \max_x, \quad (6)$$

$$\sigma_x^2 = \sum_{i=1}^k \sum_{l=1}^k c_{il} \cdot x_i \cdot x_l \rightarrow \min_x, \quad (7)$$

$$\sum_{i=1}^k x_i = 1, \quad (8)$$

$$x_i \geq 0, i = \overline{1, k}, \quad (9)$$

$$q_1 \leq q_2 \leq \dots \leq q_n, \quad (10)$$

$$\sum_{j=1}^n q_j = 1, \quad (11)$$

$$q_j \geq 0, j = \overline{1, n}. \quad (12)$$

The ratios (10) reflect the essence of a simple LRO and the opinion of the DM that the situation that developed in a later period has greater significance i.e. has a more significant impact on the present and future than the situation that developed in an earlier period of time. Many researchers express this opinion, which certainly corresponds to the realities of the economy. So, V.K. Semenychev and E.V. Semenychev note that “when forecasting in conditions of rapidly changing socio-economic phenomena, the information of later time periods is more important, more significant than the information of earlier periods” [9, p. 60]. We emphasize that in the case of portfolio decisions in the field of the third IP, with the fairness of a simple LRO, it is possible to use traditional point estimates, since a uniform law satisfies a simple LRO.

The generalized Markowitz model of the problem of finding an effective portfolio in the field of the third IC, with the validity of a partially enhanced LRO, can be written in the following form [8]:

$$H(\mathbf{q}) = -\sum_{j=1}^n q_j \cdot \ln q_j \rightarrow \max_{\mathbf{q}}, \quad (13)$$

$$m_x = \sum_{i=1}^k m_i \cdot x_i \rightarrow \max_x, \quad (14)$$

$$\sigma_x^2 = \sum_{i=1}^k \sum_{l=1}^k c_{il} \cdot x_i \cdot x_l \rightarrow \min_x, \quad (15)$$

$$\sum_{i=1}^k x_i = 1, \quad (16)$$

$$x_i \geq 0, i = \overline{1, k}, \quad (17)$$

$$\begin{cases} q_2 \geq q_1, \\ q_3 \geq q_1 + q_2, \\ \dots \\ q_n \geq q_1 + q_2 + \dots + q_{n-1}, \end{cases} \quad (18)$$

$$\sum_{j=1}^n q_j = 1, \quad (19)$$

$$q_j \geq 0, j = \overline{1, n}. \quad (20)$$

The ratios (18) reflect the essence of the partially enhanced LRO and the opinion of the DM that socio-economic conditions are changing extremely rapidly, while the sample data under consideration are time series characterized by a high rate of change. The DM is obliged to adhere to such an opinion in cases when there is either a pre-crisis (crisis) situation, or a sharp growth of the economy (of the relevant sector of the economy), etc. [10]. If the DM believes that the probability distribution satisfies the partially enhanced LRO, then this means that, in the opinion of the DM, the significance of the current time period is not less than the total significance of all previous time periods. In this case, it is impossible to use traditional point estimates, because the uniform law does not satisfy the partially enhanced LRO.

This indicates the relevance and necessity of developing such a method of constructing an estimate of the probability distribution that best takes into account the opinion of the DM about the type of order ratio on a set of states of the economic environment, which is in fact his opinion about the type of LRO that the probability distribution should satisfy.

It can be argued that the above situations of portfolio decision-making are characterized by a statistical game: with tasks (1)–(4), (5)–(12), (13)–(20) a statistical game is connected, given by a matrix  $\mathbf{R} = \mathbf{R}_{k \times n} = (r_{ij})$ , where  $r_{ij}$  is the corresponding value of the  $i$  rate of return of

the asset. Note that portfolio decision-making is possible by solving the corresponding game [11]: when certain requirements are met, the solution of the antagonistic game given by the matrix  $\mathbf{R} = \mathbf{R}_{k \times n} = (r_{ij})$  allows you to find an effective case  $\mathbf{x} = (x_1^*, \dots, x_k^*)$ , while its structure does not depend on the probability distribution.

## 2. Basic notion and definitions

Here are the definitions of LRO of the main types [1, p. 78]:

1) a simple LRO is an order relation given by inequalities of the form  $q_1 \geq q_2 \geq \dots \geq q_n$  or  $q_1 \leq q_2 \leq \dots \leq q_n$ ;

2) a partially enhanced LRO is an order relation given by inequalities of the form  $q_j \geq q_{j+1} + \dots + q_n, j = \overline{1, n-1}$  or  $q_j \geq q_1 + \dots + q_{j-1}, j = \overline{2, n}$ ;

3) an enhanced LRO is an order relation given by inequalities of the form

$$q_{j+1} + \dots + q_{j+\alpha(j)} \leq q_j \leq q_{j+1} + \dots + q_{j+\alpha(j)} + q_{j+\alpha(j)+1},$$

$$j = \overline{1, n-2}, \alpha(j) \in \{1; 2; \dots; n-1-j\}, \text{ or}$$

$$q_{j-\alpha(j)} + \dots + q_{j-1} \leq q_j \leq q_{j-\alpha(j)-1} + q_{j-\alpha(j)} + \dots + q_{j-1},$$

$$j = \overline{3, n}, \alpha(j) \in \{1; 2; \dots; j-2\}, \text{ where } \alpha(j) - \text{given natural numbers taking values from the specified sets.}$$

The *Fishburne sequence* (FS) will be called the sequence  $(q_1; q_2; \dots; q_j; \dots)$  the elements of which are equal

$$q_j = \frac{a_j}{\sum_i a_i}, \forall j, \quad (21)$$

where the sequence  $(a_1; a_2; \dots; a_j; \dots)$  generating this FS is a given monotonic sequence of non-negative numbers, the sum of which is a positive number [12, p. 132]. Further, for convenience, we will limit ourselves to the consideration of finite FS  $(q_1; q_2; \dots; q_n)$ .

Externally, formula (21) coincides with the so-called third Fishburne formula. The third Fishburne formula is used to estimate the probability distribution when it must satisfy an enhanced LRO. In this case, the formulas for calculating probability estimates contain a parameter, while using formula (21) allows us to find a value of this parameter for which the values of probability estimates satisfy both the enhanced LRO and condition (3). Obviously, the set of all FS is significantly wider than the set of all sequences satisfying the enhanced LRO and condition (3).

In the author's publications (for example, [8, 10, 12–16]), the following properties of FS and their special cases are investigated:

- 1) *Fishburne arithmetic progression, given by the first Fishburne formula:*

$$q_j = \frac{2 \cdot (n - j + 1)}{n \cdot (n + 1)}, j = \overline{1, n}; \quad (22)$$

- 2) *Fishburne geometric progression, given by the second Fishburne formula:*

$$q_j = \frac{2^{n-j}}{2^n - 1}, j = \overline{1, n}; \quad (23)$$

- 3) *Fishburne's increasing arithmetic progression:*

$$q_j = \frac{2 \cdot j}{n \cdot (n + 1)}, j = \overline{1, n}; \quad (24)$$

- 4) *the increasing geometric progression of Fishburne:*

$$q_j = \frac{2^{j-1}}{2^n - 1}, j = \overline{1, n}; \quad (25)$$

- 5) *generalized Fishburne arithmetic progression:*

$$\begin{aligned} q_j &= \frac{1}{n} - \frac{(n-1) \cdot x}{2} + (j-1) \cdot x = \\ &= \frac{2 - n \cdot (n-2 \cdot j + 1) \cdot x}{2 \cdot n}, j = \overline{1, n}; \end{aligned} \quad (26)$$

for which its difference  $x$  satisfies the inequality

$$|x| \leq \frac{2}{n \cdot (n-1)};$$

- 6) *generalized Fishburne geometric progression:*

$$q_j = \frac{x-1}{x^n-1} \cdot x^{j-1} = \frac{1-x}{1-x^n} \cdot x^{j-1}, j = \overline{1, n}; \quad (27)$$

for which its denominator is  $x$  satisfies the inequality  $x > 0$ .

Often, the use of formula (24) or (25), for example, when the index values represent time points, is preferable to the use of formula (22) or (23), respectively. Formulas (22)–(25) are used in a wide variety of studies, as evidenced by numerous publications (for example, [17–35]).

Uniform law, e.g. vector  $\mathbf{q}^* = (q_1^*; \dots; q_n^*)$ , where  $q_1^* = \dots = q_n^* = \frac{1}{n}$  is a special case of the generalized arithmetic Fishburne progression (for the difference  $x = 0$ ), and the generalized Fishburne geometric progression (for the denominator  $x = 1$ ).

### 3. Estimation of probability distribution based on the use of Fishburne sequences

Thus, generalized Markowitz models of the problem of finding an effective portfolio can be brought to the classical Markowitz model by using FS. The problem is how to make a correct estimate of the probability distribution. The proposed scheme for constructing a correct estimate of the probability distribution is based on the properties of FS, primarily on the properties of generalized Fishburne progressions.

The correctness of using formulas (26), (27) depends on the answers to such questions. 1. When does the generalized Fishburne progression satisfy a simple LRO? 2. When does the generalized Fishburne progression satisfy a partially enhanced LRO? 3. When does the generalized Fishburne progression satisfy the Gibbs–Jaynes principle?

Recall that the Gibbs–Jaynes principle considers the vector to be the most charac-

teristic estimate of an unknown distribution  $\mathbf{q}^* = (q_1^*; \dots; q_n^*)$  maximizing the entropy value  $H(\mathbf{q}) = -\sum_{j=1}^n q_j \cdot \ln q_j$  when all constraints are met, e.g. constraints (3), (4) and, possibly, one or more constraints that, in the opinion of the DM, the probability distribution should satisfy. The entropy approach significantly enriches the tools of analysis and modeling of economic risk. Obviously, if only the constraints (3), (4) are met, then the maximum value of entropy  $H(\mathbf{q}^*) = \ln n = \max_{\mathbf{q}} H(\mathbf{q})$  is achieved for a uniform law, e.g. for a vector  $\mathbf{q}^* = (q_1^*; \dots; q_n^*)$ , where  $q_1^* = \dots = q_n^* = \frac{1}{n}$ . In particular, on the set of all generalized Fishburne progressions, the maximum value of entropy is also achieved for a uniform law.

Note that an arbitrary FS, and hence an arbitrary generalized Fishburne progression, satisfies the corresponding simple LRO.

If generalized arithmetic Fishburne progressions almost always do not satisfy partially enhanced LRO, then generalized geometric Fishburne progressions, depending on the value of their denominator, in some cases satisfy, and in other cases do not satisfy partially enhanced LRO.

Consequently, the above questions have trivial answers on the set of all generalized Fishburne arithmetic progressions. However, on the set of all generalized Fishburne geometric progressions, the answers to these two questions are not so obvious. The theorems containing these answers are given and proved in the monograph of A.V. Sigal, E.S. Remesnik [12, pp. 119–127]. The essence of these theorems is as follows: an arbitrary generalized geometric Fishburne progression  $(q_1; q_2; \dots; q_n)$  satisfies the corresponding partially amplified LRO if and only if the value of its denominator belongs to the set  $x \in (0; \alpha_n] \cup [\beta_n; +\infty)$ , where  $\alpha_n$  – the root of the equation  $x^n - 2 \cdot x + 1 = 0$

owned by  $\alpha_n \in (0,5; 1]$ ,  $\beta_n = \frac{1}{\alpha_n}$ , at the same time  $x^* = \alpha_n$  ( $x^* = \beta_n$ ) – the only value of the denominator  $x$  progressions (27), maximizing the value of entropy  $H(\mathbf{q})$  for all non-increasing (non-decreasing, respectively) generalized geometric Fishburne progressions.

One of the natural methods of solving problems (5)–(12) and (13)–(20) is their reduction to problem (1)–(4) by using the most characteristic estimate of the probability distribution, i.e. by using the vector  $\mathbf{q} = (q_1; \dots; q_n)$ , maximizing the entropy value on the set of all FS satisfying the LRO of the corresponding type. Practically without loss of generality, instead of the set of all FS, we can limit ourselves to considering only generalized Fishburne progressions, the most important and very broad special case of FS.

The vector maximizing the entropy value on the set of all FS (and, consequently, on the set of all generalized Fishburne progressions), by definition always satisfying a simple LRO, is a uniform distribution, i.e. the vector  $\mathbf{q} = (q_1; \dots; q_n)$ , где  $q_1^* = \dots = q_n^* = \frac{1}{n}$ . Therefore, in the case of problem (5)–(12), the use of traditional point estimates is advisable, both from the standpoint of mathematical statistics and from the standpoint of the suitability of modeling.

And a vector maximizing the entropy value on the set of all non-decreasing generalized Fishburne geometric progressions  $(q_1; q_2; \dots; q_n)$ , satisfying a partially enhanced LRO is the progression (27), whose denominator is equal to  $x = \beta_n$ , where  $\beta_n = \frac{1}{\alpha_n}$ ,  $\alpha_n$  the root of the equation  $x^n - 2 \cdot x + 1 = 0$  belongs to  $\alpha_n \in (0,5; 1]$ .

The estimation of an unknown probability distribution and the use of this estimate can be carried out according to the following five-step scheme.

**Step 1.** Choosing the type of LRO, which, according to the DM, the distribution should satisfy  $(q_1; q_2; \dots; q_n)$  probabilities.

**Step 2.** Choosing the sequence  $(a_1; a_2; \dots; a_n)$  non-negative numbers, which, according to the DM, it is advisable to use as a sequence generating FS with the desired properties. This sequence  $(a_1; a_2; \dots; a_n)$  must satisfy the LRO of the selected type, while it can be a sequence whose elements form, for example, a certain progression of natural numbers (including a constant  $a_1 = a_2 = \dots = a_n = 1$ , strictly monotone arithmetic or strictly monotone geometric progression), Fibonacci numbers or Mersenne numbers.

**Step 3.** Construction of the FS, which, according to the DM, it is advisable to use as an estimate of the probability distribution, i.e. calculation by formula (21) of the values of the FS components  $(q_1; q_2; \dots; q_n)$ , generated by the selected sequence  $(a_1; a_2; \dots; a_n)$ .

**Step 4.** Applying the constructed FS as an estimate of the probability distribution, and, in particular, calculating the values of the estimates of the corresponding numerical characteristics according to formulas for calculating these numerical characteristics of the DRV, in which the values of the corresponding elements of the constructed FS are used instead of the probability values.

**Step 5.** Choosing the optimal solution, i.e. choosing to implement such a solution that has the best combination of the calculated values of the estimates of the numerical characteristics under consideration.

### Conclusion

Fishburne formulas for calculating point estimates of probability distributions are well known and widely used in theoretical and practical research. For a correct assessment of the probability distribution of the states of the economic environment, it is advisable to use Fishburne sequences (FS) generalizing Fishburne formulas: as an estimate of the probability distribution, it is advisable to use an FS that has

all the desired properties and, in particular, satisfies the type of linear order relation (LRO), which, according to the decision maker (DM), the probability distribution should satisfy. The most important types of LRO are simple LRO and partially reinforced LRO.

The DM can evaluate the probability distribution and use this estimate according to the following scheme, first given in the article.

1. Choosing the type of LRO that the probability distribution should satisfy.
2. The choice of a sequence that is appropriate to use as a sequence generating FS.
3. Construction of the FS, which is advisable to use as an estimate of the probability distribution.
4. Application of the constructed FS as an estimate of the probability distribution.
5. Choosing the optimal solution, i.e. the choice for the implementation of such a solution that has the best combination of the calculated values of the estimates of the numerical characteristics under consideration.

The proposed scheme makes it possible to achieve correctness in probabilistic and statistical modeling, as well as to best take into account the opinion of the DM about the type of LRO to which the elements of a variety of states of the economic environment are subject, including for cases where the use of traditional point estimates is impossible because the most characteristic estimate of the probability distribution should differ from the uniform law.

The properties of the FS completely coincide with the properties of the sequence generating it (with the exception of the normalization condition, which the sequence generating the FS does not have to satisfy). When choosing an FS that has all the desired properties, we can limit ourselves to considering the set of



generalized Fishburne progressions, which are FS that are arithmetic or geometric progressions. In the case of a simple LRO, we can limit ourselves to considering generalized arithmetic Fishburne progressions, and in the case of a partially enhanced LRO, generalized geometric Fishburne progressions. Finally, if the DM adheres to the Gibbs–Jaynes principle, then either the uniform law (with the validity of a simple LRO) or the generalized geometric Fishburne progression maximizing the entropy value (with the validity of the corresponding partially enhanced LRO) should be used as an estimate of the unknown probability distribution.

The cases when FS should be used that satisfy partially enhanced LRO include cases when the sample data under consideration are time series characterized by a high rate of change. The DM is obliged to adhere to such an opinion in cases where, for example, there is either a pre-crisis (crisis) situation, or a sharp growth of the economy (of the relevant sector of the economy).

In cases where, according to the DM, the probability distribution must satisfy a partially enhanced LRO, it is impossible to use traditional point estimates, because the uniform law used to calculate the values of traditional point estimates does not satisfy a partially enhanced LRO. It can be said that in these cases there is a contradiction between the traditional probabilistic and statistical tools and the peculiarities of the decision-making situation, and the correctness and appropriateness of modeling require using a strictly monotonic sequence as an estimate of the probability distribution, i.e. a FS satisfying the corresponding partially enhanced LRO, for example, the generalized geometric Fishburne progression maximizing the entropy value (if the corresponding partially enhanced LRO is fair). Although the desired statistical properties of point estimates are lost (unbiased, etc.), the use of a strictly monotonic FS (instead of a uniform law) allows us to achieve the desired levels of correctness and appropriateness of modeling, as well as the best consideration of the opinion of the DM. ■

### References

1. Trukhaev R.I. (1981) *Models of decision-making in conditions of uncertainty*. Moscow: Nauka (in Russian).
2. Vitlinsky V.V., Verchenko P.I., Sigal A.V., Nakonechny Ya.S. (2002) *Economic risk: Game models*. Kiev: KNEU (in Ukrainian).
3. Markowitz H.M. (1952) Portfolio selection. *Journal of Finance*, March 1952, vol. 7, no 1, pp. 77–91.
4. Markowitz H.M. (1959) *Portfolio selection: Efficient diversification of investments*. N.Y.: John Wiley & Sons.
5. Fishburn P.C. (1964) *Decision and value theory*. N.Y.: John Wiley & Sons.
6. Fishburn P.C. (1965) Analysis of decisions with incomplete knowledge of probabilities. *Operations Research*, vol. 13, no 2, pp. 217–237.
7. Fishburn P.C. (1965) Independence in utility theory with whole product sets. *Operations Research*, vol. 13, no 1, pp. 28–45.
8. Sigal A.V. (2017) On the approximation of the generalized model of Markowitz in the field of information third of the situation to the classical model of Markowitz. *Proceedings of the Seventh International conference system analysis and information technology*, pp. 159–167 (in Russian).
9. Semenychev V.K., Semenychev E.V. (2011) *Parameter identification of time series: structures, models, evolution*. Monograph. Samara: SamSC RAS (in Russian).
10. Sigal A.V. (2019) Statistical estimates taking into account the features of digital transformation. Collection of scientific works of the *II International scientific and practical forum “Russia, Europe, Asia: Digitalization of the global space”*, Stavropol, 09-12 October 2019 (ed. V.A. Korolev). Stavropol: SEKVOYA, pp. 134–137 (in Russian).



11. Sigal A.V. (2018) On the effectiveness of portfolios found by the game-theoretic method. *Economics and Mathematical Methods*, vol. 54, no 1, pp. 125–144 (in Russian).
12. Sigal A.V., Remesnik E.S. (2018) *Fishburne sequences and their application in modern portfolio theory. Monograph*. Simferopol: IP Kornienko A.A. (in Russian).
13. Sigal A.V. (2019) On managerial decision-making in the field of the second information situation. *Proceedings of the XVIII All-Russian scientific and practical conference with international participation Actual problems and prospects of economic development, Simferopol–Gurzuf, October 24–26, 2019* (eds. N.V. Apatova, T.V. Zueva), pp. 56–58 (in Russian).
14. Sigal A.V., Makeeva G.N. (2015) Generalized Fishburne progressions. *Proceedings of the IX International school-symposium analysis, modeling, management, development of socio-economic systems (AMUR-2015), Sevastopol, September 12–21, 2015*, pp. 343–350 (in Russian).
15. Sigal A.V., Remesnik E.S. (2018) Sequences satisfying linear order relations: application in economics and properties. *Drucker's Bulletin*, no 1, pp. 44–58 (in Russian). DOI: 10.17213/2312-6469-2018-1-44-58.
16. Sigal A.V., Remesnik E.S. (2019) Fishburne point estimates and their generalizations. *Proceedings of the International scientific school modeling and analysis of safety and risk in complex systems MASR-2019, Saint Petersburg, June 19–21, 2019* (eds. E.D. Solojntsev, V.V. Karasev). St. Petersburg: GUAP, pp. 85–92 (in Russian).
17. Aliev A.A., Gordienko M.S., Petelina A.V. (2020) Multipurpose assessment of the financial competitiveness of a publishing company. *Vestnik Universiteta*, no 10, pp. 113–121 (in Russian). DOI: 10.26425/1816-4277-2020-10-113-121.
18. Aliev A.A., Litvishko O.V., Yusifova A.I., Fateeva A.A. (2021) Analysis of the effectiveness of financial and economic activities of Russian football clubs. *Vestnik Universiteta*, no 5, pp. 85–92 (in Russian). DOI: 10.26425/1816-4277-2021-5-85-92.
19. Babenkov V.I., Gasyuk D.P., Dubovsky V.A. (2020) Method of risk assessment at the weapons and military equipment samples life cycle stages. *Vooruzheniye i ekonomika*, no 3 (53), pp. 59–65 (in Russian).
20. Belozertsev O.V., Belozertsev V.N. (2020) Assessment of the impact of environmental factors on the economic security of the enterprise. *Economic Bulletin of Donbass State Technical University*, no 5, pp. 5–13 (in Russian).
21. Vinogradova T.A., Kuvshinov M.S. (2021) Implementation of the assessment and analysis of the level of employees' innovation behavior. *Bulletin of the South Ural State University. Ser. Economics and Management*, vol. 15, no 2, pp. 132–139 (in Russian). DOI: 10.14529/em210215.
22. Dolzhenko A.I., Shpolyanskaya I.Yu., Glushenko S.A. (2020) Fuzzy production network for quality analysis of microservice architecture. *Business Informatics*, vol. 14, no 4, pp. 36–46 (in Russian). DOI: 10.17323/2587-814X.2020.4.36.46.
23. Zheishev R.S., Nikitin Yu.A. (2020) Evaluation of the military and economic effectiveness of the system of food supplies for the armed forces in the arctic zone of Russian federation. *Ekonomicheskij vektor*, no 2 (21), pp. 96–102 (in Russian). DOI: 10.36807/2411-7269-2020-2-21-96-102.
24. Kostyrev A.P. (2020) Evaluation of the results of the implementation of industrial policy based on a multilevel approach. *Financial Economy*, no 8, pp. 299–304 (in Russian).
25. Kuvshinov M.S., Kalacheva A.G. (2015) Development of the state of analysis of investment attractiveness of industrial enterprises. *Bulletin of the South Ural State University. Ser. Economics and Management*, vol. 9, no 2, pp. 74–81 (in Russian).
26. Lyamin B.M., Mottaeva A.B. (2020) Assessment of the potential of commercialization of the results of innovative activity in higher education. *Economic Sciences*, no 191, pp. 110–115 (in Russian). DOI: 10.14451/1.191.110.
27. Nedosekin A.O. (2003) *Methodological foundations of modeling financial activity using fuzzy multiple descriptions*. St. Petersburg: UNECON (in Russian).
28. Potapov D.K., Evstafyeva V.V. (2008) On the methods of determining weight coefficients in the problem of assessing the reliability of commercial banks. *Socio-economic situation of Russia in the new geopolitical*,

- financial, and economic conditions: realities and prospects of development: collection of scientific articles.* St. Petersburg: Institute of Business and Law, no 5, pp. 191–196 (in Russian).
29. Rodionov D.G., Konnikov E.A., Mugutdinov R.M. (2020) System analysis of the competitiveness of a digital enterprise within the information environment. *Economic Sciences*, no 193, pp. 394–401 (in Russian). DOI: 10.14451/1.193.394.
  30. Sazonov A.E., Osipov G.S. (2017) Linguistic assessment of the perfection level of safety management system of shipping companies. *Vestnik Gosudarstvennogo universiteta morskogo i rechnogo flota imeni admirala S.O. Makarova*, vol. 9, no 1, pp. 7–16 (in Russian). DOI: 10.21821/2309-5180-2017-9-1-7-16.
  31. Sakharova L.V., Akperov G.I. (2020) Fuzzy-multiple method of complex assessment of the state of socio-economic systems of the region. *Intellektual'nyye resursy – regional'nomu razvitiyu*, no 2, pp. 137–143 (in Russian).
  32. Somov V.L., Tolmachov M.N. (2017) Methods for determining the weighting coefficients of dynamic integral indicators. *Questions of Statistics*. Moscow: Statistics of Russia, no 6, pp. 74–79 (in Russian).
  33. Tolstykh T.O., Gamidullayeva L.A., Shmeleva N.V. (2020) Methodological aspects of project portfolio formation in the innovation ecosystem. *Models, Systems, Networks in Economics, Technology, Nature and Society*, no 1 (33), pp. 5–23 (in Russian). DOI: 10.21685/2227-8486-2020-1-1.
  34. Tyutyukina E.B., Kapranova L.D., Sedash T.N. (2014) Identification of priority areas and investment support for the development of the Russian economy. *Economic Analysis: Theory and Practice*, Moscow, no 38 (389), pp. 2–11 (in Russian).
  35. Yashin S.N., Borisov S.A. (2020) Methodological approaches to determining the rating of economic and innovative development of industrial enterprises in the region. *Voprosy innovatsionnoy ekonomiki*, vol. 10, no 2, pp. 819–836 (in Russian). DOI: 10.18334/vinec.10.2.100921.

### About the author

#### Anatoliy V. Sigal

Doct. Sci. (Econ.), Professor;

Professor, Department of Business Informatics and Mathematical Modeling, V.I. Vernadsky Crimean Federal University, 4, Prospekt Vernadskogo, Simferopol 295007, Russia

E-mail: ksavo3@cfuv.ru

ORCID: 0000-0003-2090-4464

DOI: [10.17323/2587-814X.2021.4.61.75](https://doi.org/10.17323/2587-814X.2021.4.61.75)

# Modeling the links between institutional and actual globalization in the countries of the world

**Elena D. Kopnova** 

E-mail: [ekopnova@hse.ru](mailto:ekopnova@hse.ru)

**Lilia A. Rodionova** 

E-mail: [lrodionova@hse.ru](mailto:lrodionova@hse.ru)

National Research University Higher School of Economics  
Address: 20, Myasnitskaya Street, Moscow 101000, Russia

## Abstract

The paper is devoted to modeling the links between the institutional and actual level of globalization in the countries of the world. Vector models of error correction, quantile regression, and a stochastic frontier model are considered. As a measure of globalization and its components, the KOF-index of globalization system is used, which allows us to analyze individual globalization processes in the economy, social sphere and politics. According to 2020 data, we determine the dynamic relations between the actual and institutional components of globalization, and the priority of the institutional component for informational and financial globalization is revealed. The example of financial globalization shows the uneven degree of influence of the institutional component on the actual globalization, in particular, its prevailing importance for less globalized countries, indicating the alignment of the degree of internationalization in the global financial system. The degree of effectiveness of the impact of institutional measures, together with the overall level of well-being on the actual financial globalization is analyzed. It is shown that the spread across the countries of the world in the efficiency indicator is almost 70%. Almost 10% of countries have a low efficiency of up to 50%. One third of the countries has average efficiency (50–75%). The share of countries with high efficiency over 75% is about 60%.

**Key words:** KOF index of globalization; cointegration; vector error correction model; forecast errors variance decomposition; quantile regression; stochastic frontier model.

**Citation:** Kopnova E.D., Rodionova L.A. (2021) Modeling the links between institutional and actual globalization in the countries of the world. *Business Informatics*, vol. 15, no 4, pp. 61–75. DOI: [10.17323/2587-814X.2021.4.61.75](https://doi.org/10.17323/2587-814X.2021.4.61.75)

## Introduction

**G**lobalization is the most important factor of social progress. It is determined by the strengthening of economic, social and political interactions of countries and peoples, regardless of national borders [1]. According to some data [2], globalization is capable of raising world GDP per capita by almost a third. The integration of trade and investment flows, the convergence of markets and the development of multinational corporations characterize economic globalization. Social globalization is expressed in the development of communication technologies, international cultural centers and personal contacts. Political globalization manifests itself in the activities of organizations that, in accordance with the principles of international law, make it possible to unite the forces of the countries of the world in the fight against global problems. The positive effects of globalization include the optimization of production due to the cross-country division of labor and access to innovation [3], the diversification of financial risks with the attraction of foreign investment [4], the increase in the level of human capital development due to the development of information technology and the international education system [5]. Researchers usually identify the main problem of globalization as an increase in income inequality [6]. Among the negative effects, there are also risks to economic security [7], violation of human rights [8] and loss of ethnic identity [9].

To measure globalization, index systems reflecting its structure are used [10]. The most popular of them is the system of the KOF-index of globalization from the Swiss Economic Institute<sup>1</sup>. The appearance in 2018 of the latest edition of the methodology of its calculation with a radical expansion of the structure and informa-

tion base significantly expanded the possibilities of studying the problems of globalization [11]. In addition to the integral indicator, the system of this index contains sub-indices of economic, social and political globalization, which in turn are divided into separate components. The economic sub-index includes sub-indices of trade and financial globalization, and the social sub-index includes personal, informational, and cultural globalization. Each of these indicators is further divided into *de facto* and *de jure* categories. The indicators *de facto* measure the actual flows between countries (for example, the amount of imports), while *de jure* – their institutional capabilities (for example, import taxes). Each indicator is formed according to world official statistics since 1970, published with a delay of two years, measured on a 100-point scale. A total of 42 variables are used. The principal component method is used for calculation, as well as the panel normalization method<sup>2</sup>.

Most of the works on the study of globalization are devoted to analysis of its impact on the well-being of the population. The most interesting of them are the works [12–18], which use econometric tools for data analysis. However, it should be noted that these studies are limited by lack of results from studying the structure of the globalization process itself, the relationship between its individual components. This reorganization of the KOF-index of globalization in 2018 marks noticeable progress in the development of the statistical methodology for measuring globalization and expands the possibilities of its systematic analysis. Taking into account this reorganization, the authors set a goal to analyze the links between institutional and actual levels of globalization. Two main tasks were put forward. The first is to investigate the trends of dynamic relationships between de

---

<sup>1</sup> KOF (Konjunkturforschungsstelle) Globalisation Index: <https://www.kof.ethz.ch/en/forecasts-and-indicators/indicators/kof-globalisation-index.html>

<sup>2</sup> The structure of the KOF index indicating the weights of individual indicators is given in the Appendix (*Table A1*)

jure and de facto sub-indices. And the second is to study the degree of influence of the institutional component and the effectiveness of its application to form the actual component of globalization for the countries of the world.

## 1. Methods

### 1.1. Data

Globalization was measured by the KOF-sub-indices – the de facto and de jure globalization index. All indicators were considered in the period from 1970 to 2018<sup>3</sup>. *Figures 1–4* show graphs of the analyzed time series averaged by countries of the world for the Globalization KOF-index (gl) and its components: economic (ec), social (soc) and political (pol) globalization; trade (tr) and financial (fin) globalization; personal (per), information (inf) and cultural (cul) globalization. The corresponding sub-indices are de facto and de jure denoted with the addition of the symbols \_df, \_dj.

*Figure 1* shows that the acceleration of globalization in the world began in 1994. At the same time, if previously the sub-index was de facto superior to de jure, now the de jure sub-index prevails. There is a noticeable divergence of trends. De jure, in general, is growing faster. *Figure 2* shows that this is due to the social and political index de jure.

At the same time, *Figure 2* shows that the de facto political sub-index remains noticeably smaller. Among economic sub-indices, on the contrary, the de facto sub-index is superior. The connection of the components of economic globalization is noticeable. It can be seen how the de facto sub-index follows the de jure sub-index. The connection of the sub-indices of social globalization is also visible, but not so noticeable. The connection between the components of political globalization is even less noticeable.

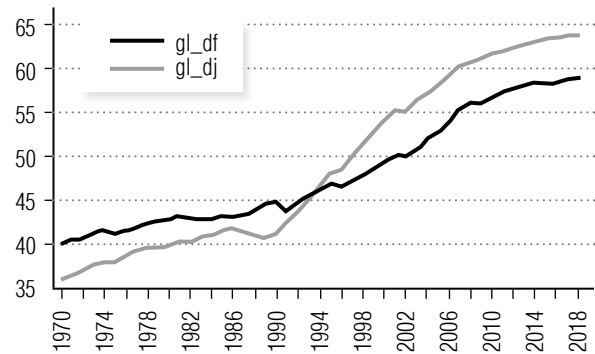


Fig. 1. Sub-indices of the de facto and de jure KOF-index of globalization

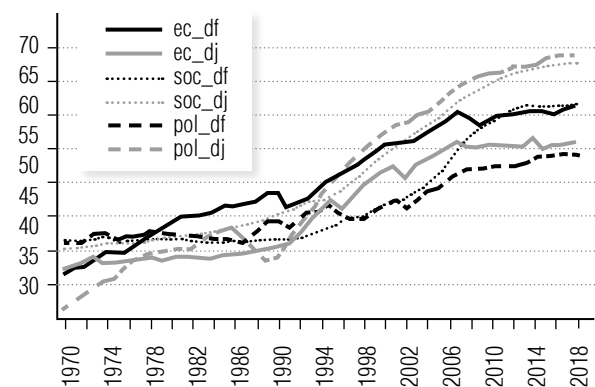


Fig. 2. De facto and de jure sub-indices of the components of the globalization KOF-index

The de facto predominance in the economic index is due to the relatively strong growth of the corresponding component of financial globalization. As part of the trade globalization index, in recent years, the values of the de facto and de jure sub-indices have been converging.

*Figure 4* shows that the prevalence of the de jure sub-index in the indicator of social globalization is achieved at the expense of all components. However, a strong acceleration is noticeable for the dynamics of the de facto information globalization index in recent years. At the same time, the relationship of this index with the de jure index is visible. The cultural component of globalization is characterized by a relatively low value of the de facto sub-index.

<sup>3</sup> <https://kof.ethz.ch/en/forecasts-and-indicators/indicators/kof-globalisation-index.html>

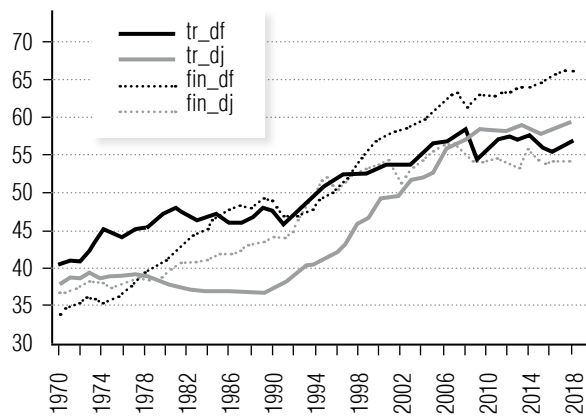


Fig. 3. De facto and de jure sub-indices of the components of the economic sub-index of the KOF-index of globalization

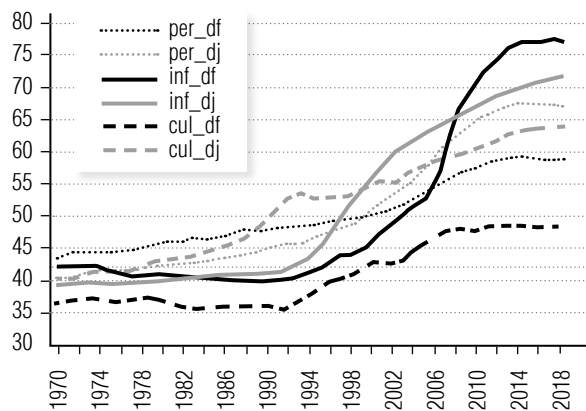


Fig. 4. Sub-indices of de facto and de jure components of the social sub-index of the KOF-index of globalization

Figure 5 shows the values of the de jure (abscissa axis) and de facto (ordinate axis) sub-indices of the KOF-index of globalization at the time of 2018 for 196 countries of the world. It can be seen from the figure that the sub-indices are strongly correlated. The sample value of the paired coefficient of correlation was 0.87.

Additionally, when analyzing the spatial sample at the time of 2018, the income index was used, obtained on the basis of the logarithm of

gross national income (GNI) per capita at purchasing power parity (PPP) in 2017 prices in US dollars<sup>4</sup> and measured on a 100-point scale. The indicator was calculated as the ratio of the logarithm of the GNI growth index to the logarithm of its maximum value<sup>5</sup>. The minimum value of GNI was assumed to be equal to \$100, as the minimum fixed in official statistics. The maximum value was set at \$75,000 in accordance with the phenomenon of the immutability of the level of well-being for countries with a higher level of GNI [19]. The Appendix contains a list of countries used in the analysis for which the index values are available. Figure 5 shows its values for these countries at the time of 2018. The figure shows that the spread of values is quite large and amounts to almost 70%, which indicates a high differentiation of well-being in the countries of the world.

## 1.2. Procedures

The methodology for the study of dynamic relationships was based on the idea of cointegration analysis of random processes using the Vector Error Correction Model (VECM) [20]. Nine models of the relationship between the de facto and de jure sub-indices were constructed for the KOF-index of globalization and its components.

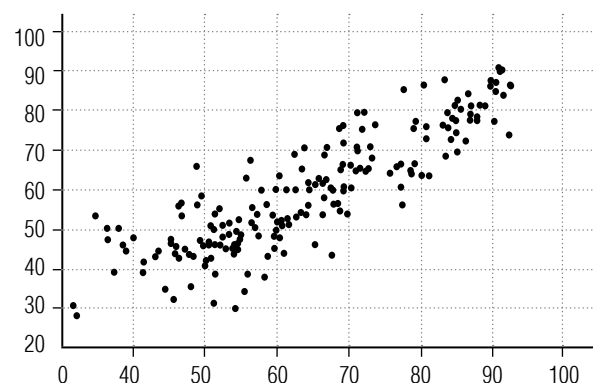


Fig. 5. De jure and de facto sub-indices of globalization 2018 for the countries of the world

<sup>4</sup> <https://data.worldbank.org/indicator/NY.GNP.PCAP.PP.KD?view=chart>

<sup>5</sup>  $\text{Income index} = \ln(\text{GNI}/\text{GNI}_{\min}) / \ln(\text{GNI}_{\max}/\text{GNI}_{\min}) \times 100$



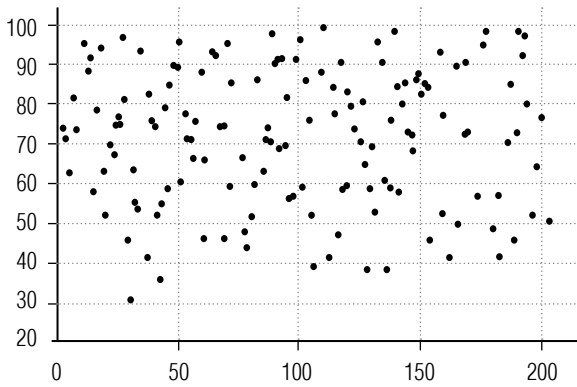


Fig. 6. Income index in the countries of the world, 2018

The general form of the model:

$$\Delta X_t = \mu_t + \alpha \beta' X_{t-1} + \sum_{j=1}^p \Gamma_j X_{t-j} + U_t, \\ t = 1, 2, \dots, T.$$

$$\Delta X_t = X_t - X_{t-1}, \alpha = \|\alpha_{ij}\|_{k \times r}, \beta = \|\beta_{ij}\|_{k \times r}, \\ \Gamma_j = \|\gamma_{ij}\|_{k \times k}, k = 2,$$

$$U \sim N(0, I_T \otimes \Sigma_u), U = (U_1, \dots, U_T),$$

$$\Sigma_u = \|\sigma_{ij}\|_{k \times k}, \text{cov}[U_{it}, U_{js}] = \begin{cases} \sigma_{ij}, & t = s \\ 0, & t \neq s \end{cases},$$

where the components of the vector  $X_t$  are the processes analyzed in the work; the vector  $\mu_t$  contains deterministic components for each of these processes: a trend and a constant;  $r$  – cointegration rank;  $U_t$  – errors vector.

For each time series, tests were applied for the presence of a single root of the characteristic equation of the corresponding process in accordance with the algorithm of the Dolado procedure [21]. ADF (Augmented Dickey–Fuller) and KPSS (Kwiatkowski–Schmidt–Shin) tests [20] were used. The Johansen approach [22] was used to evaluate the parameters of the cointegration ratio and the error correction model. The optimal specification of the models was selected based on the Bayesian Information Criterion (BIC) and the model's compliance with its assumptions. The remnants of the mod-

els were tested for the absence of autocorrelation and compliance with the normal distribution law (multidimensional analogues of the Breush–Godfrey and Jarque–Bera LM test). Durnik–Hansen orthogonalization [23] was used for the remnants of the VEC model.

The characteristics of long-term dynamic relationships for cointegrated processes were determined using testing of variables for weak exogeneity relative to the parameters of the error correction model. For this purpose, the statistical significance of the estimates of the components of the correction matrix  $\alpha$  was analyzed, since the insignificance of the estimate  $\alpha_{ij}$  means that when the processes deviate from long-term equilibrium, the corresponding  $i$ -th variable is not corrected. To draw conclusions about strict exogeneity, the Toda–Yamamoto approach [24] was used with the choice of the number of lags in the test model according to the BIC criterion. The degree of impact of the analyzed processes on each individual process was measured using the decomposition of the variance of its prediction error, in which the selected process occupied the last position in the recursive order of causality by Wold (Wold-causality) [25]. As an impact measure for each individual process, the corresponding proportion of the estimate of the variance of the forecast error, the maximum for 10 years, was considered.

To analyze the degree of influence of the institutional factors of globalization on the actual globalization, regression models were used, estimated from the prolog data of the de facto and de jure sub-indices for 2018, averaged across the countries of the world. Taking into account the heteroscedasticity of the errors of the usual regression

$$E[Y_i | X_i] = X_i \beta (i = 1, 2, \dots, n)$$

quantile regression [26] was used for quantiles corresponding to probabilities 0.25, 0.5 and 0.75:

$$q_\tau[Y_i | X_i] = X_i \beta_\tau, P\{Y_i \leq q_\tau\} = \tau.$$

To analyze the heterogeneity of the influence of the factors of the formation of actual financial globalization in the world, the Stochastic Frontier Model (SFM) [27] was evaluated:

$$Y_i = X_i\beta + V_i - U_i, U_i \sim iidN^+(0, \sigma_u^2), \\ U_i \geq 0, \text{cov}[U_i, X_i] = 0, \text{cov}[U_i, V_i] = 0.$$

Additionally, the logarithm of the income index was taken into account as a regressor. To justify the use of the model, the remnants of the usual regression were tested for the statistical significance of the asymmetry coefficient; its sign was checked. The hypothesis about the ineffective influence of factors  $H_0: \sigma_u^2 = 0$  was also tested. The evaluation of the distribution of the efficiency indicator (Ef) of the factors under consideration for the countries of the world was interpreted:

$$Ef_i = \left( \frac{1 - \Phi(\sigma_* - \mu_{*i} / \sigma_*)}{1 - \Phi(-\mu_{*i} / \sigma_*)} \right) \exp\left(-\mu_{*i} + \frac{1}{2}\sigma_*^2\right), \\ \mu_{*i} = \frac{-s\varepsilon_i\sigma_u^2}{\sigma^2}, \sigma_* = \frac{\sigma_u\sigma_v}{\sigma}, \sigma = \sqrt{\sigma_u^2 + \sigma_v^2}, \\ \Phi(x) = \frac{1}{2\pi} \int_{-\infty}^x \exp\left(-\frac{x^2}{2}\right) dx.$$

The efficiency distribution was compared with the ranking of the de facto index.

## 2. Results

### 2.1. Analysis of dynamic relationships between de jure and de facto sub-indices

Based on the results of testing processes for stationarity (*Table 1*), with a probability of 0.95, it can be argued that all the series under consideration are realizations of random processes that are stationary in the first differences.

*Table 2* shows the results of testing the de facto and de jure sub-indices for cointegration. The table shows that the hypothesis of the absence of cointegration is rejected at the significance level

of 0.05 for most sub-indices with the definition of one cointegration ratio for them. The presence of a long-term relationship is not evident for the global index and sub-indices of cultural and political globalization.

The results of quality control of the evaluated VEC models showed their sufficiently high validity and allowed them to be used for analysis and interpretation. In particular, the hypothesis of the absence of autocorrelation of residues up to and including the 3-rd order was not rejected at the significance level of 0.05 for all models. *Table 3* shows the results of testing cointegrated sub-indices for weak exogeneity. Statistically insignificant at the level of 0.05 estimates for the short-term ratio for the de jure sub-indices of financial and information globalization indicate a weak exogeneity of these values relative to the parameters of VEC models.

*Tables 4* and *5* present the results of an analysis of the short-term relationship between the de facto and de jure sub-indices. *Table 4* shows that such a relationship, as well as a long-term one, is practically not manifested for the political and cultural sub-indices. The economic sub-index is characterized by the influence of the de jure sub-index on the de facto. This is especially noticeable for the financial sub-index. For trade globalization, the de facto influence on de jure prevails a little. For the sub-index of social globalization, there is a relatively weak mutual influence of the de facto and de jure sub-indices. However, for the indices of personal and information globalization, the priority influence of de jure on de facto is noticeable.

The data in *Table 5* confirm the fact that there is no short-term connection between the de facto and de jure sub-indices for political and cultural globalization. For both economic and social globalization, the defining role of the de jure sub-index is manifested. For financial and information globalization, this indicator is a highly exogenous variable relative to the parameters of the VEC model.

Table 1.

**Results of the analysis  
of processes for stationarity**

		Y		dY	
		ADF	KPSS	ADF	KPSS
1	gl_df	0.99	0.21	0.00	0.34
2	gl_dj	0.42	0.16	0.03	0.17
3	ec_df	0.58	0.08	0.00	0.17
4	ec_dj	0.99	0.13	0.00	0.20
5	soc_df	0.52	0.22	0.06	0.40
6	soc_dj	0.36	0.21	0.25	0.40
7	pol_df	0.62	0.21	0.00	0.25
8	pol_dj	0.46	0.12	0.01	0.09
9	tr_df	0.32	0.09	0.00	0.16
10	tr_dj	0.19	0.19	0.15	0.29
11	fi_df	0.99	0.06	0.00	0.09
12	fi_dj	0.999	0.149	0.000	0.214
13	in_df	0.53	0.22	0.06	0.44
14	in_dj	0.10	0.20	0.22	0.34
15	per_df	0.42	0.19	0.01	0.17
16	per_dj	0.03	0.21	0.32	0.30
17	cul_df	0.62	0.19	0.00	0.27
18	cul_dj	0.09	0.08	0.00	0.11

Notes:

1. For all series it is indicated: Y is the initial series, dY is the first difference.
2. For the ADF test, one-sided MacKinnon P-values are given.
3. For the KPSS test, test statistics are compared with critical values at the significance level of 0.05: 0.146 – for the initial series (taking into account the trend) and 0.463 – for the first difference (taking into account the constant).

**2.2. Analysis of the degree  
and heterogeneity of the influence  
of institutional factors  
on the actual globalization**

Further, using the example of the financial globalization sub-index, the results of the analysis of the degree of influence of the de jure index on the de facto for the 2018 globalization KOF-

Table 2.

**Results of the analysis  
of processes for cointegration\***

Test	Trace		Max-eigenvalue	
Grade	0	1	0	1
gl	0.54	0.76	0.76	0.43
ec	0.04	0.07	0.11	0.07
soc	0.02	0.22	0.02	0.22
pol	0.44	0.98	0.36	0.98
tr	0.06	0.71	0.04	0.71
fin	0.01	0.92	0.01	0.92
in	0.03	0.62	0.02	0.62
per	0.06	0.69	0.03	0.69
cul	0.13	0.44	0.14	0.44

\* MacKinnon–Haug–Michelis P-values with minimum BIC value

Table 3.

**Results of the analysis  
of sub-indices for weak exogeneity\***

	ec	soc	tr	fin	inf	per
df	–2.92	–2.04	–2.98	–4.22	–3.94	–2.37
dj	–2.05	–3.90	2.80	–1.57	1.01	–4.83

\* The value of T-statistics for estimating the coefficient at the remainder of the cointegration ratio in the ECM (error correction model) for the de facto (df) and de jure (dj) sub-indices is given.

index with the involvement of the income index are demonstrated. Taking into account the omissions in the data, 145 observations were used in the models. The data was used in logarithms. There was no multicollinearity problem, since the correlation coefficient between the regressors turned out to be statistically insignificant and equal to 0.03. *Table 6* shows some results of estimating the usual regression model, quantile regression with quantiles corresponding to probabilities 0.25, 0.5, 0.75, and the stochastic frontier model.

Table 4.

**Estimates of the decomposition  
of the variance of the forecast error in VAR/VEC models, %\***

Endogenous	d(gl)	ec	soc	d(pol)	tr	fin	inf	per	d(cu)
df	14.98	55.81	27.76	2.15	24.18	54.50	44.00	62.01	10.23
dj	4.24	15.12	29.76	2.58	29.87	6.83	15.75	27.85	3.85

\* The maximum value for a 10-year period is given. For each of the de facto (df) and de jure (dj) sub-indices, in accordance with the Cholesky Ordering, the proportion of variation due to a change in the alternative variable va and he is shown.

Table 5.

**Results of the analysis of processes for causality by Granger\***

Endogenous	gl	ec	soc	pol	tr	fin	inf	per	cul
df	0.00	0.01	0.00	0.45	0.00	0.03	0.00	0.01	0.99
dj	0.00	0.45	0.44	0.50	0.00	0.87	0.95	0.06	0.61

\* The P-value of  $\chi^2$ -statistics is given to test the hypothesis that each of the de facto and de jure sub-indices is not a Granger reason for an alternative de jure and de facto sub-index.

The use of quantile regression was due to the high heteroscedasticity of the remnants of the usual regression – the P-value of the  $\chi^2$ -statistics of the Breusch–Pagan test was equal to 0.000. This is also noticeable in *Figure 5*. It can be seen that there are countries with a low de jure and high de facto sub-index, but with a high de jure, high de facto is determined almost unambiguously.

*Table 6* shows that all models describe the direct statistically significant impact of institutional factors, together with the indicator of the state of health, on the actual globalization in the financial sector. According to estimates of conventional regression, the de facto index value changes by almost a third of a percent on average when the de jure index changes by 1 percent. A comparison of the estimates for the dj indicator and their graphic illustration in *Figure 7* shows that for countries with a higher level of globalization, this influence is weaker than for countries with a low level.

The application of the stochastic frontier model was facilitated by a statistically significant (at the level of 0.001), rather high modulo negative value of the residual asymmetry coefficient (–0.788). The relatively low BIC value compared to its value for conventional regression also supports the use of this model. The hypothesis of the absence of inefficiency of factors was rejected at the significance level of 0.05. *Figures 8* and *9* show the results of calculating the effectiveness of the factors of the formation of actual globalization in the financial sphere for 145 countries.

### 3. Discussion

The results of the calculations showed that there is both a long-term and a short-term dynamic relationship between the processes of globalization in the institutional sphere and its actual manifestation. And although it does not manifest itself at the global level, for the general index of globalization – perhaps due to its com-

Table 6.

## Results of evaluation of regression models\*

	Normal regression	Quantile regression			Stochastic frontier model
		0.25	0.5	0.75	
dj	0.29***	0.51***	0.27***	0.16**	0.15**
gni	0.44***	0.61***	0.39***	0.41***	0.37***
const	1.10***	−0.62	1.03**	1.94***	2.27***
Prob F/ $\gamma^2$	0.00	0.00	0.00	0.00	0.00
R <sup>2</sup> /Pseudo R <sup>2</sup>	0.38	0.28	0.24	0.23	—
BIC	35.58	—	—	—	27.48

\* The statistical significance of coefficient estimates at the level of 0.01 is marked \*\*\*, 0.05 – \*\*, 0.1 – \*.

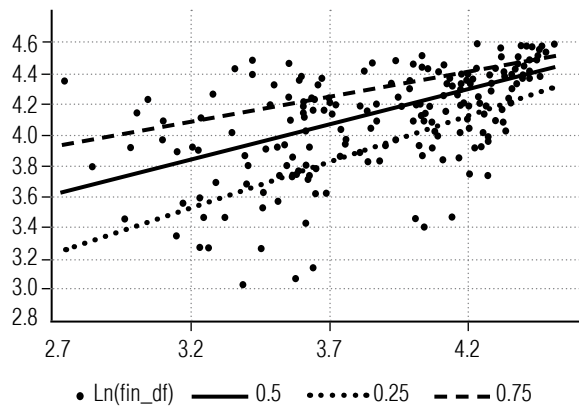


Fig. 7. Initial and model values of the logarithm of the de facto financial globalization index for quantile regression, excluding the income index

posite structure – for most of its components it is clearly traceable. The special role of institutional factors for the process of globalization has manifested itself for the financial sphere and the sphere of information technology. It turned out that their formation in these areas occurs relatively independently, without significant reliance on the results of de facto globalization in them, and they are the determining basis for actual internationalization.

The analysis of dynamic relations between de jure and de facto factors revealed the ambiguity of mutual influence for individual components of the globalization process. If economic

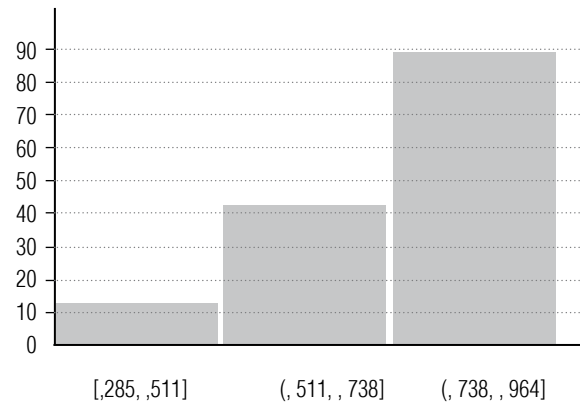


Fig. 8. Distribution of efficiency of factors of formation of actual globalization in the financial sphere

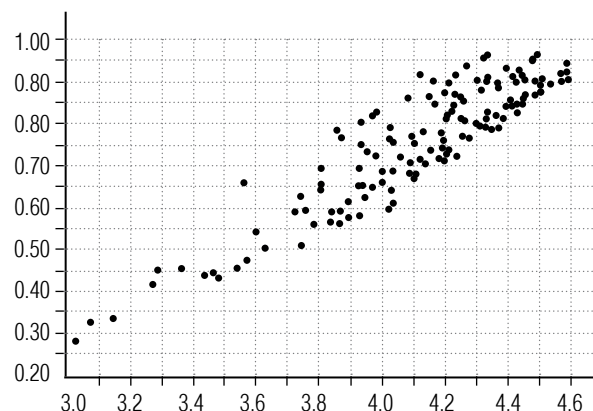


Fig. 9. The relationship of the indicator of the effectiveness of the factors of the formation of actual globalization with the sub-index of de facto financial globalization

globalization is characterized by a noticeable influence of institutional factors on de facto globalization, then for social globalization this apparent orientation is replaced by their mutual influence. For the process of political globalization, the connection between de jure and de facto factors has not been found, both in the long term and in the short term.

The analysis of the spatial sample according to the data of the financial globalization KOF-index at the time of 2018 indicated that in countries with a lower manifestation of actual globalization, the role of institutional factors in shaping the level of globalization is higher than in more globalized countries. Perhaps this indicates a tendency to equalize the level of financial globalization in the countries of the world.

According to the results of calculating the effectiveness of the factors of the formation of actual globalization in the financial sphere, *Figure 8* shows a high heterogeneity of countries in this indicator: the spread of values occurs from 0.285 to 0.964. Almost 10% (8.28%) of countries have low efficiency up to 50%. These include countries such as Iran, Bangladesh, Ethiopia. *Figure 7* shows that these are countries with a low level of actual financial globalization – noticeably lower than the first quantile of the logarithm of this indicator (3.937). A third of the countries (30.34%) have average efficiency (50–75%), for example, Turkey, Russia, Brazil. These are mainly countries with an average level of actual financial globalization that does not exceed its median (4.199). More than 60% (61.38%) of countries have high efficiency. Note that among the countries with high efficiency there are representatives from all groups according to the de facto index level. A complete list of countries with the valu-

es of efficiency and the logarithm of the corresponding globalization index is given in the Appendix (*Table A2*).

## Conclusion

The application of the methodology of cointegration analysis to the 2020 data of the globalization KOF-index system made it possible to determine the dynamic relationships between the actual and institutional components of globalization. It is shown that the institutional factors of globalization are the determining basis for the actual internationalization of the financial sphere and the sphere of information technology.

Modeling of the relationship between the de facto and de jure sub-indices of globalization revealed a significant heterogeneity of influence for the countries of the world. Using the example of financial globalization, the difference in the degree of influence of the institutional component on actual globalization is shown, in particular, its predominant importance for less globalized countries with a level of financial globalization less than the first quartile.

The application of the stochastic frontier model to the data of financial globalization made it possible to analyze the degree of effectiveness of the impact of institutional measures together with the overall level of well-being on the actual globalization in the financial sphere. It is shown that the level of efficiency varies in the range from 28 to 96 percent. Almost 10% of countries have low efficiency of up to 50%. One third of the countries has average efficiency (50–75%). Thus, the share of countries with high efficiency over 75% is about 60%. ■

## References

1. Akhter S.H. (2004) Is globalization what it's cracked up to be? Economic freedom, corruption, and human development. *Journal of World Business*, vol. 39, no 3, pp. 283–295. DOI: 10.1016/j.jwb.2004.04.007.
2. Weiß J., Sachs A., Weinelt H. (2018) *Globalization report 2018: Who benefits most from globalization*. Bertelsmann Stiftung. Available at: <https://ged-project.de/globalization/globalization-report-2018-who-benefits-most-from-globalization/> (accessed 01 July 2021).



3. Tsai C. (2007) Does globalization affect human well-being? *Social Indicators Research*, no 81, pp. 103–126. DOI: 10.1007/s11205-006-0017-8.
4. Ghosh A. (2017) How does banking sector globalization affect economic growth? *International Review of Economics & Finance*, no 48, pp. 83–97. DOI: 10.1016/j.iref.2016.11.011.
5. Manyika J., Lund S., Bughin J., Woetzel J., Stamenov K., Dhingra D. (2016) *Digital globalization: The new era of global flows. Report*. McKinsey Global Institute. Available at: <https://www.mckinsey.com/business-functions/digital-mckinsey/our-insights/digital-globalization-the-new-era-of-global-flows> (accessed 01 July 2021).
6. Bergh A., Nilsson T. (2010) Do liberalization and globalization increase income inequality? *European Journal of Political Economy*, vol. 26, no 4, pp. 488–505. DOI: 10.1016/j.ejpoleco.2010.03.002.
7. Krylova I.A. (2016) Russia in the context of globalization: New threats. *Russian Journal of Philosophical Sciences*, no 4, pp. 30–44 (in Russian).
8. Dreher A., Gassebner M., Siemers L. (2012) Globalization, economic freedom, and human rights. *Journal of Conflict Resolution*, no 56, pp. 516–546. DOI: 10.2139/ssrn.1695446.
9. Yanitzky O.N. (2019) Challenges and risks of globalization. Seven Theses. *Sociological Studies*, no 1, pp. 29–39 (in Russian). DOI 10.31857/S013216250003745-2.
10. Cherkashina T.Yu. (2011) Indices of globalization: Indicators and the calculation scheme. *Sociology: methodology, methods, mathematical modeling*, no 33, pp. 136–165 (in Russian).
11. Gygli S., Haelg F., Potrafke N. (2019) The KOF globalisation index – revisited. *Review of International Organizations*, no 14, pp. 543–574. DOI: 10.1007/s11558-019-09344-2.
12. Rajan R. G., Zingales L. (2003) The great reversals: The politics of financial development in the twentieth century. *Journal of Financial Economics*, vol. 69, no 1, pp. 5–50. DOI: 10.1016/S0304-405X(03)00125-9.
13. Baltagi B.H., Demetriades P.O., Law S.H. (2009) Financial development and openness: Evidence from panel data. *Journal of Development Economics*, vol. 89, no 2, pp. 285–296. DOI: 10.2139/ssrn.1808903.
14. Tovar-García E. (2012) Financial globalization and financial development in transition countries. *Economía: Teoría y Práctica*, no 36, pp. 155–178. DOI: 10.24275/ETYPAM/NE/362012/Tovar.
15. Atif S.M., Srivastav M., Sauyrbekova M., Arachchige U.K. (2012) Globalization and income inequality: A panel data analysis of 68 countries. *EconStor Preprints*, no 65664. ZBW – Leibniz Information Centre for Economics. Available at: <http://hdl.handle.net/10419/65664> (accessed 01 July 2021).
16. Law S.H., Azman-Saini W.N.W., Tan H.B. (2014) Economic globalization and financial development in East Asia: A panel cointegration and causality analysis. *Emerging Markets Finance and Trade*, vol. 50, no 1, pp. 210–225. DOI: 10.2753/REE1540-496X500112.
17. Le T.H., Kim J., Lee M. (2016) Institutional quality, trade openness, and financial sector development in Asia: An empirical investigation. *Emerging Markets Finance and Trade*, vol. 52, no 5, pp. 1047–1059. DOI: 10.1080/1540496X.2015.1103138.
18. Muye I.M., Muye I.Y. (2017) Testing for causality among globalization, institution and financial development: Further evidence from three economic blocs. *Borsa Istanbul Review*, vol. 17, no 2, pp. 117–132. DOI: 10.1016/j.bir.2016.10.001.
19. Kahneman D., Deaton A. (2014) High income improves evaluation of life but not emotional well-being. *Proceedings of National Academy of Sciences*, vol. 107, no 38, pp. 16489–16493. DOI: 10.1073/pnas.1011492107.
20. Mills T.C. (2008) *The econometric modeling of financial time series*. Cambridge. New York. DOI: 10.2307/2329254.
21. Dolado H., Jenkinson T., Sosvilla-Rivero S. (1990) Cointegration and unit roots. *Journal of Economic Surveys*, no 4, pp. 243–273.

22. Johansen S. (1988) Statistical analysis of cointegration vectors. *Journal of Economic Dynamics and Control*, vol. 12, no 2–3, pp. 231–254.
23. Doornik J.A., Hansen H. (2008) An omnibus test for univariate and multivariate normality. *Oxford Bulletin of Economics and Statistics*, no 70, pp. 927–939. DOI: 10.1111/j.1468-0084.2008.00537.x.
24. Toda H.Y., Yamamoto T. (1995) Statistical inferences in vector autoregression with possible integrated processes. *Journal of Econometrics*, no 66, pp. 225–250.
25. Lütkepohl H. (2007) *New introduction to multiple time series analysis*. New York: Springer-Verlag. DOI: 10.1017/S0266466606000442.
26. Koenker R., Hallock K.F. (2021) Quantile regression. *Journal of Economic Perspectives*, vol. 15, no 4, pp. 143–156. DOI: 10.1257/jep.15.4.143.
27. Kumbhakar S., Parmeter C., Zelenyuk V. (2017) *Stochastic frontier analysis: Foundations and advances. Working Papers 2017–10*, University of Miami, Department of Economics. DOI: 10.1017/CBO9781139174411.

## Appendix

Table A1.

### Structure of the KOF index of globalization

Index	Weight	Index	Weight
<b>1. Economic globalization</b>			33.3
<b>1.1. Trade globalization</b>			50.0
<i>De facto</i>	50.0	<i>De jure</i>	50.0
Trade in goods	38.8	Import barriers	26.8
Trade in services	44.7	Average tariff level	25.6
Variety of trading partners	16.5	Taxes on trade	24.4
		Trade agreements	23.2
<b>1.2. Financial globalization</b>			50.0
<i>De facto</i>	50.0	<i>De jure</i>	50.0
Foreign direct investment	26.7	Barriers to investment	33.3
Portfolio investments	16.5	Capital account openness	38.5
International debt	27.6	Investment agreements	28.2
International reserves	2.1		
International income payments	27.1		
<b>2. Social globalization</b>			33.3
<b>2.1. Personal globalization</b>			33.3
<i>De facto</i>	50.0	<i>De jure</i>	50.0
Phone traffic	20.8	Access to telephone communication	39.9
Money transfers	21.9	Freedom of visits	32.7
International tourism	21.0	Number of airports	27.4
Migration	17.2		
International education	19.1		

<b>2.2. Information globalization</b>				33.3
<i>De facto</i>	50.0	<i>De jure</i>		50.0
Internet Usage	37.2	Access to television		36.8
Export of high technologies	34.5	Internet access		42.6
International patents	28.3	Freedom of the press		20.6
<b>2.3. Cultural globalization</b>				33.3
<i>De facto</i>	50.0	<i>De jure</i>		50.0
Trade in cultural goods	28.1	Gender equality		24.7
Exchange of services	24.6	Human capital		41.4
International trademarks	9.7	Civil liberties		33.9
Number of McDonalds	21.6			
Number of IKEA stores	16.0			
<b>3. Political globalization</b>				33.3
<i>De facto</i>	50.0	<i>De jure</i>		50.0
Embassies in the country	36.5	Membership in organizations		36.2
Participation in UN missions	25.7	International agreements		33.4
Public organizations	37.8	Variety of partners		30.4

Table A2.

**The effectiveness of the formation of financial globalization  
in the countries of the world and the logarithm  
of the corresponding de facto globalization sub-index\***

No	country	ef	ln_d	No	country	ef	ln_d
1	Iran. Islamic Republic	0.28	3.030	17	China	0.56	3.835
2	Comoros	0.33	3.075	18	Turkey	0.58	3.896
3	Bangladesh	0.33	3.147	19	Dominican Republic	0.58	3.930
4	Ethiopia	0.42	3.271	20	Bolivia	0.59	3.839
5	Kenya	0.43	3.479	21	Cameroon	0.59	3.722
6	Pakistan	0.44	3.437	22	Gabon	0.59	3.866
7	Haiti	0.44	3.461	23	Rwanda	0.59	3.754
8	Chad	0.45	3.284	24	Korea. Rep	0.59	4.020
9	Nepal	0.45	3.363	25	Romania	0.61	4.033
10	Algeria	0.46	3.532	26	Ecuador	0.61	3.896
11	Iraq	0.47	3.573	27	Indonesia	0.62	3.943
12	India	0.5	3.624	28	Tanzania	0.63	3.738
13	Guatemala	0.51	3.744	29	Botswana	0.64	4.027
14	Myanmar	0.54	3.600	30	Egypt. Arab Republic	0.65	3.971
15	Sri Lanka	0.56	3.783	31	Bhutan	0.65	3.806
16	Paraguay	0.56	3.866	32	Morocco	0.65	3.933

No	country	ef	ln_d	No	country	ef	ln_d
33	Nigeria	0.65	3.924	73	United States	0.79	<b>4.371</b>
34	Brazil	0.65	3.938	74	New Zealand	0.79	4.330
35	Benin	0.65	3.806	75	Chile	0.79	4.310
36	Burundi	0.66	3.559	76	Greece	0.8	4.303
37	Philippines	0.66	4.003	77	Madagascar	0.8	<b>3.937</b>
38	Mexico	0.67	4.104	78	Serbia	0.8	4.261
39	Argentina	0.68	4.108	79	Armenia	0.81	4.256
40	Peru	0.68	4.089	80	Jordan	0.81	4.250
41	Sudan	0.69	3.999	81	Czech Republic	0.81	4.381
42	El Salvador	0.69	4.031	82	Slovenia	0.81	4.333
43	Mali	0.69	3.807	83	Nicaragua	0.81	4.205
44	Uganda	0.69	3.926	84	Italy	0.82	4.360
45	Russian Federation	0.7	4.139	85	Niger	0.82	3.971
46	Belarus	0.71	4.090	86	Vietnam	0.82	4.205
47	Maldives	0.71	4.124	87	Malaysia	0.82	4.336
48	Poland	0.72	4.199	88	United Arab Emirates	0.83	4.432
49	Costa Rica	0.72	4.179	89	Solomon Islands	0.83	3.984
50	Djibouti	0.72	4.061	90	South Africa	0.83	4.220
51	Gambia	0.72	3.976	91	Moldova	0.84	4.224
52	Israel	0.72	4.237	92	Estonia	0.84	4.416
53	Saudi Arabia	0.73	4.206	93	Panama	0.84	4.400
54	Zimbabwe	0.73	3.948	94	Latvia	0.84	4.397
55	Albania	0.74	4.152	95	Honduras	0.84	4.164
56	Uruguay	0.74	4.216	96	Germany	0.85	4.446
57	Oman	0.74	4.211	97	Spain	0.85	4.429
58	Bulgaria	0.74	4.200	98	Hungary	0.85	4.411
59	Guinea	0.75	3.934	99	Canada	0.86	4.450
60	Cote d'Ivoire	0.75	4.035	100	Equatorial Guinea	0.86	4.253
61	Bosnia and Herzegovina	0.76	4.102	101	Namibia	0.86	4.248
62	North Macedonia	0.76	4.198	102	Liberia	0.86	4.089
63	Ghana	0.76	4.027	103	Congo	0.86	4.15
64	Lithuania	0.77	4.282	104	Eswatini	0.87	4.232
65	Sierra Leone	0.77	3.871	105	Austria	0.87	4.479
66	Kyrgyz Republic	0.77	4.095	106	Portugal	0.87	4.456
67	Croatia	0.77	4.255	107	Vanuatu	0.87	<b>4.199</b>
68	Colombia	0.78	4.189	108	France	0.87	4.483
69	Tunisia	0.78	4.130	109	Finland	0.87	4.491
70	Congo. Dem. Republic	0.78	3.859	110	Denmark	0.87	4.499
71	Mauritania	0.79	4.027	111	Cabo Verde	0.88	4.315
72	Japan	0.79	4.343	112	Sweden	0.88	4.508

No	country	ef	ln_d	No	country	ef	ln_d
113	Norway	0.88	4.506	130	Montenegro	0.91	3.274
114	Bahrain	0.89	4.507	131	Mongolia	0.91	4.493
115	Belgium	0.89	4.533	132	Cambodia	0.91	3.274
116	Jamaica	0.89	4.369	133	Lebanon	0.91	4.493
117	United Kingdom	0.89	4.533	134	Guinea-Bissau	0.91	3.274
118	Kazakhstan	0.89	4.369	135	Angola	0.92	4.493
119	Senegal	0.89	4.493	136	Cyprus	0.92	3.274
120	Georgia	0.89	3.274	137	Malta	0.92	4.493
121	Switzerland	0.9	4.493	138	Bahamas	0.92	3.274
122	Burkina Faso	0.9	3.274	139	Ukraine	0.93	4.493
123	Brunei Darussalam	0.9	4.493	140	Lesotho	0.94	3.274
124	Belize	0.9	3.274	141	Mauritius	0.94	4.493
125	Luxembourg	0.9	4.493	142	Marshall Islands	0.95	3.274
126	Netherlands	0.9	3.274	143	Togo	0.96	4.493
127	Kiribati	0.9	4.493	144	Mozambique	0.96	3.274
128	Kuwait	0.91	3.274	145	Timor-Leste	0.96	4.493
129	Ireland	0.91	4.493				

\* The quartiles of the de facto index are marked in bold

### About the authors

#### Elena D. Kopnova

Cand. Sci. (Tech.);

Associate Professor, Department of Statistics and Data Analysis, National Research University Higher School of Economics, 11 Pokrovsky Bulvar, Moscow 109028, Russia;

E-mail: [ekopnova@hse.ru](mailto:ekopnova@hse.ru)

ORCID: 0000-0002-8429-141X

#### Lilia A. Rodionova

Cand. Sci. (Econ.);

Associate Professor, Department of Statistics and Data Analysis, National Research University Higher School of Economics, 11 Pokrovsky Bulvar, Moscow 109028, Russia;

E-mail: [lrodionova@hse.ru](mailto:lrodionova@hse.ru)

ORCID: 0000-0002-0310-6359

[DOI: 10.17323/2587-814X.2021.4.76.92](https://doi.org/10.17323/2587-814X.2021.4.76.92)

# Technology for designing innovative processes for creating products and services of a network enterprise using an i4.0 knowledge-based system

**Yury F. Telnov** 

E-mail: Telnov.YUF@rea.ru

**Vasily A. Kazakov** 

E-mail: Kazakov.VA@rea.ru

**Andrey V. Danilov** 

E-mail: Danilov.AV@rea.ru

Plekhanov Russian University of Economics  
Address: 36, Stremyanny Lane, Moscow 117997, Russia

## Abstract

The creation of network enterprises based on the digital technologies of the Industrie 4.0 (the 4th Industrial Revolution, i4.0) opens broad opportunities for increasing production flexibility, customer focus and continuous innovation in products and services provided. At the same time, new opportunities necessitate the development of new methods and technologies for designing innovative processes in the context of digital i4.0 platforms, all of which highlights the relevance of the presented research topic. This work aims to define technologies for designing innovative processes to create products and services using i4.0 systems which are based on multi-agent interaction of asset administration shells (AAS), displaying digital twins of product components, and the use of ontological and cognitive methods for forming and justifying design decisions. The work presented here uses the Domain-Driven Design approach, an architectural framework for building i4.0 systems, methods of ontological engineering, quality function deployment (QFD), analysis of the types and consequences of potential inconsistencies (FMEA) and processing of fuzzy sets. The paper proposes principles for identifying bounded contexts of the domain under the design activities for the stages of the life cycle and products' subsystems (components). For bounded contexts of the domain, it is envisaged to create AAS of i4.0 systems, with the help of which the innovative process is supported and the multi-agent interaction of its



participants is carried out. As cognitive tools for making design decisions, we proposed to use services for assessing the importance of the determined quality characteristics of products and minimizing deviations of the proposed solutions from the formed functional and non-functional requirements. The methods of ontological engineering and data modelling allow us to dynamically develop an innovative project and support various versions of the project in the design process. Application of the proposed technology for designing innovative processes to create products and services at network enterprises using i4.0 systems will improve the quality of design decisions, increase the dynamism and continuous design of innovative projects.

**Key words:** innovative process; digital twin; domain-driven design; bounded context; value-added chain; ontology; multi-agent interaction; i4.0 system; i4.0 platform; asset administration shell (AAS).

**Citation:** Telnov Yu.F., Kazakov V.A., Danilov A.V. (2021) Technology for designing innovative processes for creating products and services of a network enterprise using an i4.0 knowledge-based system. *Business Informatics*, vol. 15, no 4, pp. 76–92. DOI: 10.17323/2587-814X.2021.4.76.92

## Introduction

The innovative process for creating products and services in a network enterprise is an iterative process of designing a value-added chain in which the conceptual study of the product design is accompanied by the selection of project participants and agreeing with them the conditions for performing work at the implementation stage. The architecture of a knowledge-based system (KBS) that supports the innovative process for creating products and services was proposed by the authors in [1]. Its distinctive feature is the use of digital threads [2, 3] and digital twins [4–6] technologies for reflecting the design process in the system of product models and related processes, multi-agent technology for organizing the interaction of participants in the value-added chain [7–9] and the ontological approach [10–12] for automating the search for relevant information sources and organizing information exchange between software agents. For further development of technology to design innovative processes that create products and services, it is necessary to justify the choice of a software platform for multi-agent

implementation of a knowledge-based system and determine the design technology in more detail, considering the selected software implementation environment.

The methodological foundations of the conceptual design of such complex innovative processes of product and service design are laid down in the Domain-Driven Design approach [13, 14], which allows one to structure the subject area, considering its objective structure and subjective component that determines the organization of design teams working efficiently. In this regard, this article develops the application of the Domain-Driven Design approach to identifying bounded contexts of the innovative process under the stages of the life cycle and with the value-added chain for decomposing the process into separate components and establishing various types of interfaces between them following the needs of multi-agent interaction of the innovative process participants.

The multi-agent implementation of a knowledge-based system to support the innovative process for creating products and services is aimed at ensuring the stability of the entire structure and the autonomy and interoperabil-

ity of its components, distributed in a computing system, and operating on a single software platform. Since the proposed KBS implements the technology of digital threads and digital twins [1], the article discusses the use of the 4th generation industry software platform (the i4.0 platform) [15] and asset administration shells (AAS) [16, 17] to represent the bounded contexts of software agents corresponding to individual components of the value-added chain. At the same time, the composition of functional and platform (infrastructure) services necessary for the implementation of the innovative process to create products and services in a networked distributed environment is determined.

To develop mechanisms for supporting planning processes and organizing a value-added chain which is implemented through the interaction of software agents (AAS), the article proposes to use the ontological approach [11, 12] and cognitive approach [18, 19]. They both provide algorithms formalization for choosing business partners and coordination of conditions of interaction based on semantic search methods and fuzzy assessment of requirements correspondence and their fulfilment possibilities. The co-use of ontological and cognitive approaches will allow us to build effective value-added chains aimed at promising and executable product design implementation. The mechanisms of multi-agent interaction of business partners based on digital twins' technology on the i4.0 platform will reduce the time and increase the flexibility of designing innovative processes.

### **1. The domain-driven design approach for the structuring of innovative processes domain design**

During the design of innovative processes to create products and services, there are problems of rational definition of the boundaries of project tasks, decision-making during the for-

mation of project groups, organization of interaction between project groups. As a methodology for designing innovative processes to create products and services, it is proposed to use the domain-driven design approach that is best focused on solving cognitive problems in project teams that arise during the coordination and justification of project solutions [13, 14].

The domain-driven design approach to innovative design aims at conducting an effective decomposition of the design process according to the subject and contextual (cognitive) principles and allows one to develop a common language of interaction in project groups of diverse specialists. From our point of view, this language and corresponding design solution concepts can be reflected in computer-supported models in the form of ontologies.

At its core, the domain-driven design approach is designed for the development of a special type of innovative product (software) and the use of flexible iterative software development technologies in which each iteration involves the creation of a working version of the software. Considering that modern production of conventional products is becoming digital (that is, products, on the one hand, are saturated with embedded software, and on the other hand, are supported by remote computer models – digital twins) then the principles of domain-driven design can be fully extended to traditional products and services with a cascade (sequential) development model.

The decomposition and modelling of the subject area for the innovative processes design are considered from two points of view: from the side of individual subsystems, the so-called subdomains, and from the point of view of solving individual tasks, the so-called bounded contexts, for the joint solution of which project groups are created.

From the point of view of the allocation of subsystems (subdomains) and following modelling, the functional principle of highlighting

the functions that the product should perform is usually applied; cross-functional areas of integration can be distinguished in addition. Often, the functional subsystems development results in the corresponding physical parts of the product development. For example, for a car design, the chassis, fuel system, electrical equipment, security system and other car subsystems are allocated as subdomains, which can be divided into components corresponding to individual nodes or parts. In *Figure 1*, subsystems and components (subdomains) are shown with rounded rectangles, and the arrows reflect the decomposition order of the subject area.

Bounded contexts are associated with specific tasks to be solved and have clearly defined boundaries. Specialists, primarily of two types, are involved in solving innovative tasks in a bounded context: developers and domain experts. Depending on the task complexity, developers and experts of different profiles can be involved in the innovative process.

In the innovative processes of product creation, it is advisable to allocate bounded contexts according to the life cycle stages: concept formation, product requirements formation, structural design, value-added chain design, prototype production and testing, industrial production and maintenance, disposal. At the same time, if the same group is engaged in several works throughout the life cycle, these works can be combined into one bounded context. For example, the work on the concept (product quality characteristics) formation and requirements formation, structural design and value-added chain design can be combined in the same bounded context, provided that the same specialists work together on the listed tasks. In *Figure 1*, individual works are shown as ovals, and their associations into bounded contexts are shown as rectangles. At the same time, the implementation of bounded contexts may differ significantly from each other depending on the functional subsystems or even the components of the product.

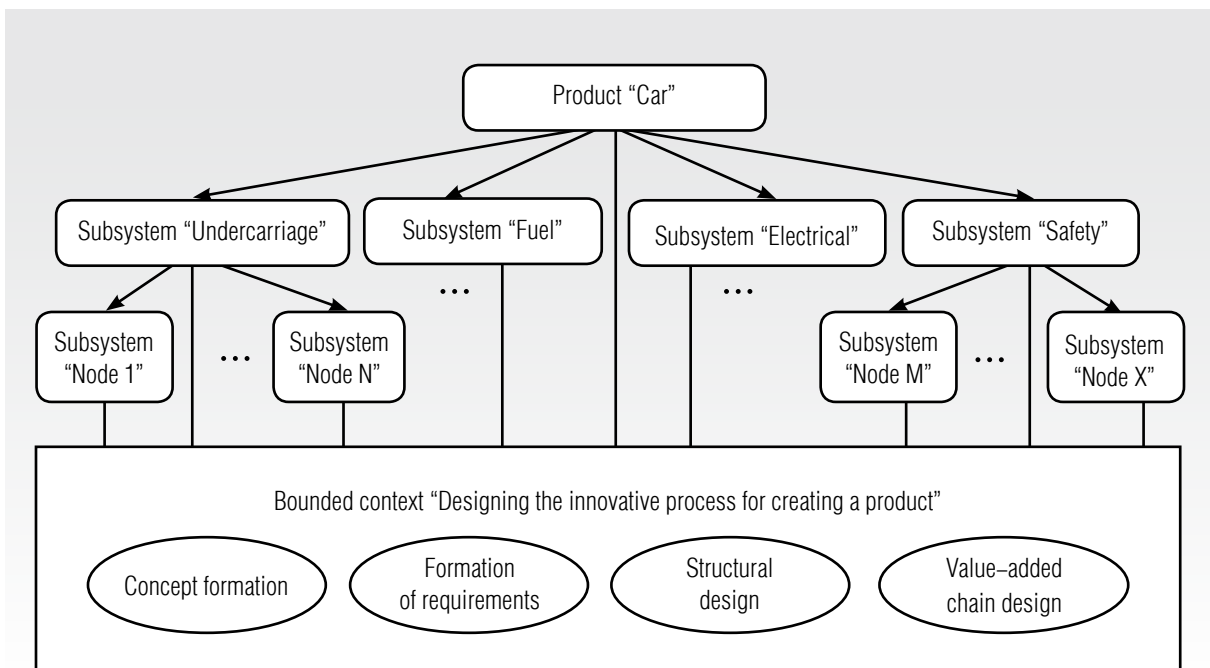


Fig. 1. Example of interaction between subsystems and bounded contexts

Within one bounded context, the participants in the product creation process work on an equal basis, forming sets of properties and rules (procedures) of behavior. At the same time, standard ontologies, reference books, and any other external information resources can be used within one bounded context. In the process of joint work, a common language of interaction is formed, formalized in the form of an iteratively developed ontology. In the course of work, mathematical and simulation models can be developed to justify the adoption of optimal (or rational) design decisions.

Bounded contexts of the innovative process for creating a product can be set both for the entire product and its components (for a detailed study). In the bounded context of product design, the value-added chain is being worked out, causing the need to select the product components contractors. In this case, interfaces between bounded contexts for the entire product and its components should be organized, as a rule, in the “requirement – execution” paradigm.

Interfaces between bounded contexts can be implemented in various ways [14]:

- ◆ **Partnership.** In this case, groups of process performers working on related subject sub-domains actively exchange the results of joint work. Group members work on common parts of the ontology and should exchange both the parts of the ontology being developed and the project results.
- ◆ **Shared kernel.** This method of interaction can be considered as a special case of partnership organization. In this case, a common bounded context is organized for several groups that have access to it with different modes of resource sharing and results adjustment.
- ◆ **Customer – supplier development.** This method can be considered as a way of interaction along the value-added chain when a higher-level product group representing the customer forms specifications of requirements for the contractor of the order (supplier). The supplier must cognitively and economically evaluate the received requirements and, following the requirements, must give consent to the execution and fulfil them accordingly. In this case, it is possible to interact with external suppliers (manufacturers) in the mode of sending messages and deferred decision-making. To coordinate requirements between customers and suppliers, it is also necessary to coordinate interacting fragments of ontologies.
- ◆ **Conformist.** This method of interaction is a specific case of the customer-supplier interface. In this case, the contractor is in a subordinate position and fulfils the customer’s requirements without agreeing with the latter on the possibilities of fulfilling the order. This way of establishing the interface is characteristic of the internal relations of the company’s divisions when a superior structure is well aware of the capabilities of subordinate structures, and subordinate structures fully own the ontology or are directly involved in its development.
- ◆ **Anticorruption layer.** This method of interaction is also a specific case of the customer-supplier interface. But in this case, direct access to the supplier’s data for criterion selection is prohibited. Getting answers to the queries of interest is possible only through sending messages, usually in offline mode. Thus, the subordinate level of the organization protects itself from possible customer access to confidential information. In this case, requirements adjustment through ontology is the most difficult and requires the translation of the shared knowledge representation often.
- ◆ **Open host service.** In this case, a protocol is used that provides access to a bounded context through a set of services that are expanded and refined as the project develops.

- ◆ **Published language.** This method is a specific case of the “service with an open protocol” method. The use of a public language involves the use of a common project ontology. In some cases, it is possible to translate concepts in both directions of interaction.
- ◆ **Separate ways.** The unrelated use of bounded contexts may be due to competing product and service development. In this case, decisions are made as the project is completed.
- ◆ **Big ball of mud.** This is the interaction of a new product with an already existing part of the product which can be performed without complying with the requirements of subject-oriented design.

From the point of view of creating some material product, the structural model of bounded context is transformed into the program code of a digital twin. At the same time, the digital twin will correspond to one bounded context, reflecting either the entire product or its specific component and represent an autonomous software agent in a multi-agent system based on knowledge to solve the problems of designing the innovative process for creating products and providing services. At the same time, the presence of several subsystems (subdomains) in the product makes it possible to specialize the development of software modules for the corresponding subdomains and the need to obtain integrated solutions at the level of the entire product.

## 2. Software implementation of a knowledge-based multi-agent system for designing innovative processes using the asset administration shell

The knowledge-based system proposed in [1] for designing innovative processes to create products and services can be implemented in the architecture of the 4th generation industry system (the i4.0 system) [20–22], which

includes a set of interacting i4.0 components – Asset Administration Shells (AAS), which are accessed by i4.0 applications of innovative processes participants. AAS is a software implementation of the concept of digital twins of created products or parts thereof, as well as related resources: machine tools, production lines, supply chains, actors executing production processes, information assets and organizational and management assets. From the point of view of the domain-driven design approach, AAS of products and their components within the design of innovative processes for creating products and services will constitute the semantic core of KMS, and AAS of the remaining resources will be auxiliary entities.

AAS services are categorized into application component services and infrastructure services of the AAS platform. *Figure 2* shows the structure of an asset administration shell for products (product components) interacting with the i4.0 platform.

An asset administration shell of a digital twin reflects data about a specific physical entity – any resource and includes a set of services that implement functional modules for solving design problems.

Asset data determine the state of this entity at various stages of the life cycle. It can be planned, actual, predicted, simulated, etc. Data can be categorized into sub-models, such as identification, technical, operational and documentation sub-models:

- ◆ identification data, which must be immutable throughout the entire life cycle of the project;
- ◆ technical data – quantitative and qualitative characteristics that meet the business requirements for the product to perform the functions of the product;
- ◆ functional data – requirements for the functional subsystems of the product in case of a complex structure;

- ◆ project documented data that reflects the characteristics of the implementation of the functional requirements in terms of the technologies used and process executors;
- ◆ operational data that reflects the performance of products at the operational stage or the results of simulation and mathematical modelling at the design and development stages.

AAS application component services correspond to functional modules of applications. In the case of a bounded context associated with the design of innovative processes to create products and services, services of application components call application modules for the formation of quality characteristics, product requirements, design of structures and

value-added chains, as well as a planning module that coordinates the execution of all functional modules [1].

Infrastructure services of the i4.0-platform are auxiliary. They ensure the functioning of the AAS itself and the interaction between AAS, and from this point of view, they are a platform and standardized. Infrastructure services refer to the i4.0 platform. The i4.0 system is developing on the i4.0 platform basis.

First of all, infrastructure services include AAS management services, which perform the functions of creating AAS components, registering and creating service registries, searching and presenting services, searching and presenting AAS data and submodels.

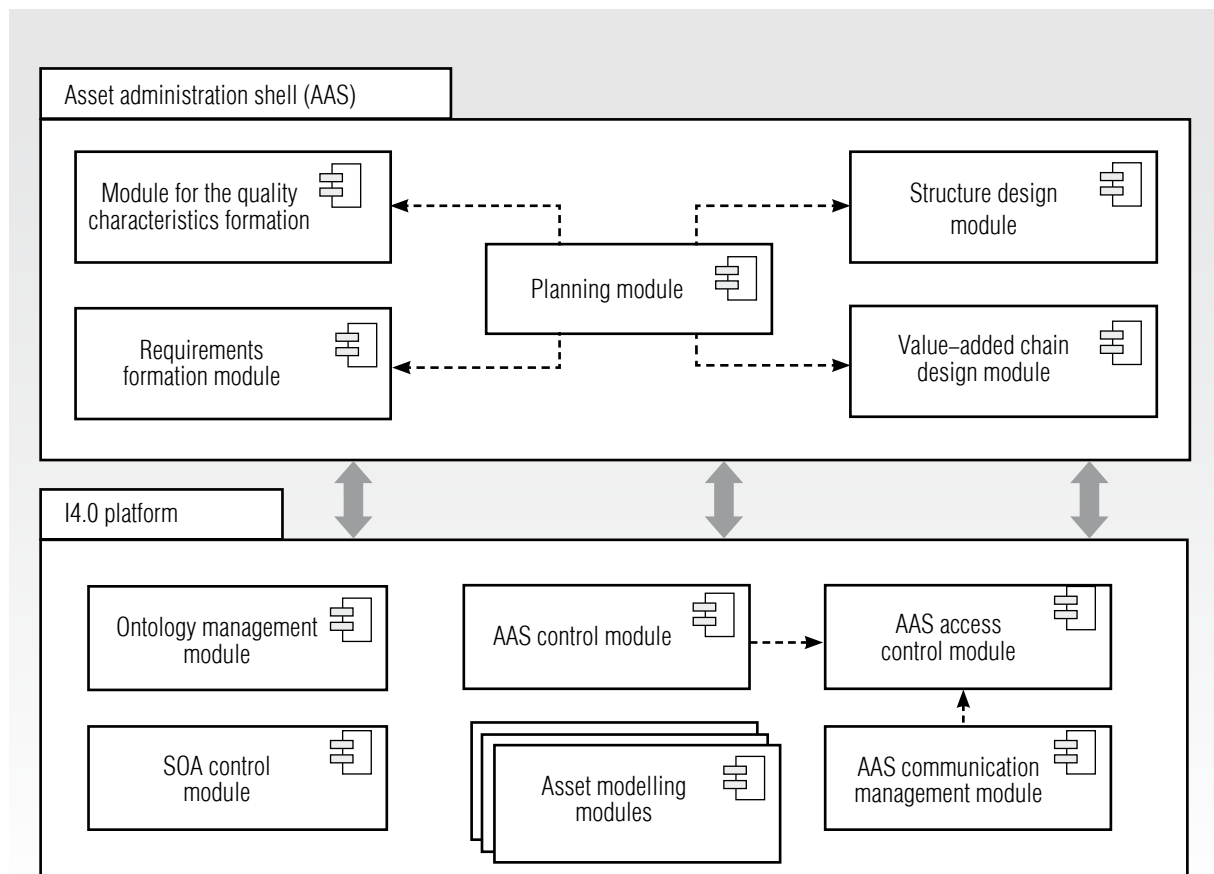


Fig. 2. Structure of the asset administration shell for products (product components) interacting with the i4.0-platform



A set of standardized asset modelling services enables the generation and use of asset data using modelling tools. Standard modelling tools are used that implement:

- ◆ Quality Function Deployment (QFD) methods to prioritize business and functional requirements implementation;
- ◆ Failure Mode and Effects Analysis (FMEA) methods for analyzing the types and consequences of potential non-conformities;
- ◆ methods of processing fuzzy sets;
- ◆ simulation methods;
- ◆ statistical modelling methods;
- ◆ machine learning and neural network modelling, etc.

Thus, on the i4.0 platform, you can implement the general functionality of various types of modelling (asset modelling modules) required to build a value-added chain - simulation mechanisms, QFD and FMEA algorithms, and other mathematical and statistical modelling algorithms. At the same time, any AAS participating in various stages of the value-added chain can call these services to implement the corresponding functionality (*Figure 2*), substituting the necessary input data for modelling from its submodels and obtaining required results.

Infrastructure services of the i4.0 platform are used to control access of application components services to data and models in specific AAS. Infrastructure services set access restrictions for various application components, access ontologies, determine the context for sharing application components, check (classify) the compliance of calls to application components services. In *Figure 2*, the infrastructure platform services implement the functions of ontology management, service-oriented architecture, interoperability with other AAS and access control to AAS.

AAS can provide service interfaces for software applications to access their data and invoke commands or run models. Service interfaces (for example, implemented in RESTful API) provide communication and interoperability between AAS and applications.

Typically, modern industrial products have a complex structure. In this case, for the entire product and its specific parts, separate i4.0 components (AAS) are created, which are connected by service interfaces (*Figure 3*). Both external subcontractors and internal executors can act as launch actors for functional applications and corresponding i4.0 components.

These i4.0 components exchange information with each other using three modes of interaction:

- ◆ passive mode – file exchange;
- ◆ reactive mode – by organizing data exchange and calling services through API;
- ◆ proactive mode (peer-to-peer), when digital twins automatically jointly carry out some production processes.

Since the article deals only with the initial stages of creating products and services associated with concept formation and value-added chain design, then either passive or reactive interaction modes are considered as modes of interaction, involving the active participation of people in the design process. From the point of view of the multi-agent implementation of the interaction process of i4.0 components belonging to external participants in the value-added chain, the reactive interaction mode corresponds to various modes of sending messages and subscribing to information, as well as service interaction (see the section “The domain-driven design approach for the structuring of designing innovative processes domain”).

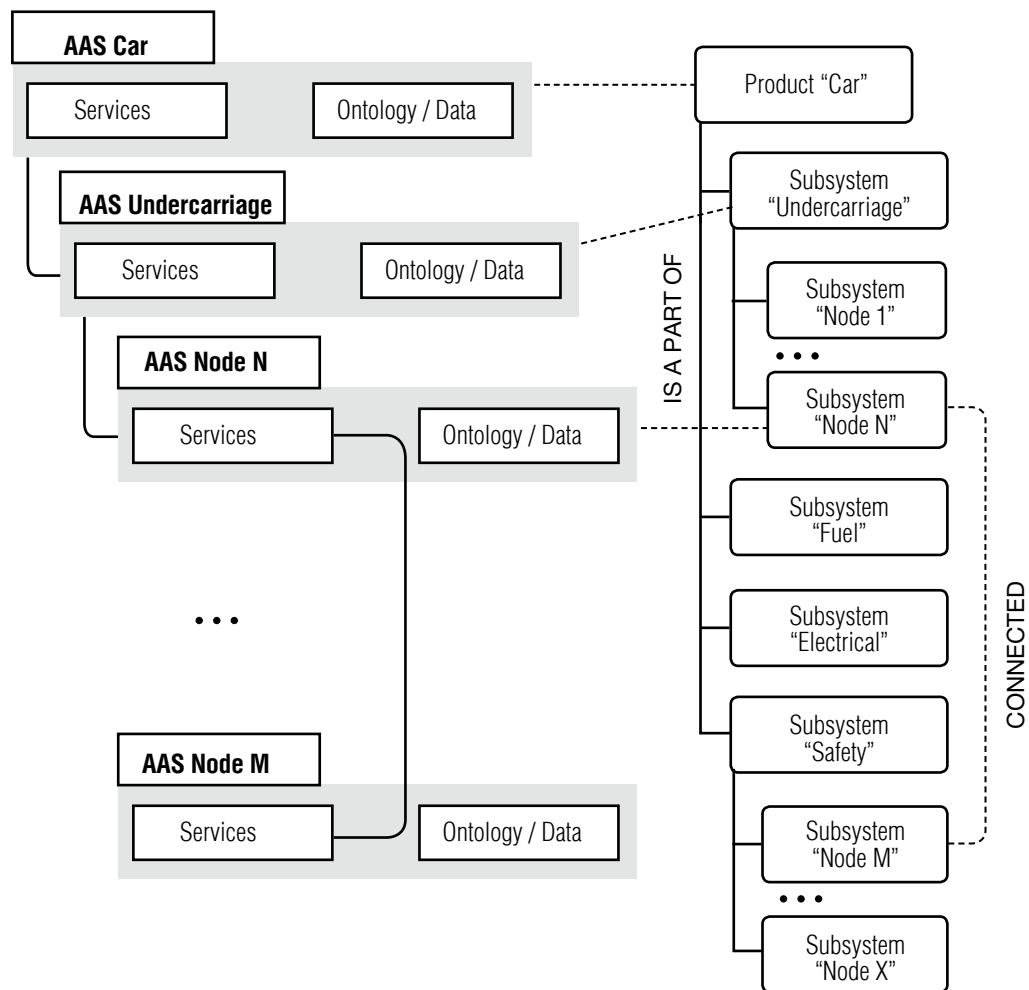


Fig. 3. An example of a layered AAS system (based on [21])

### 3. Technology for designing innovative processes to create products using ontological and cognitive approaches in the i4.0 system

The scheme of interaction between participants of a network enterprise in the i4.0 system which implements the functions of a knowledge-based system to support innovative processes of products creation is shown in *Figure 4*. Based on the reference model of the domain ontology, project ontologies are formed both for the entire project and its components. As noted in the first section, ontologies are formed iteratively and are used to

solve search problems in knowledge bases and exchange knowledge between separate AAS. Standardized services of QFD [23], FMEA [24], fuzzy set processing [25] are used in functional modules of AAS, realizing cognitive decision-making mechanisms. Let's consider the technology for designing innovative processes to create products in the i4.0 system in detail.

The core of AAS is a **planning module** that implements interface services for interacting with participants of an innovative process of product creation and provides coordination of functional services. In particular, the module supports the decision-making process

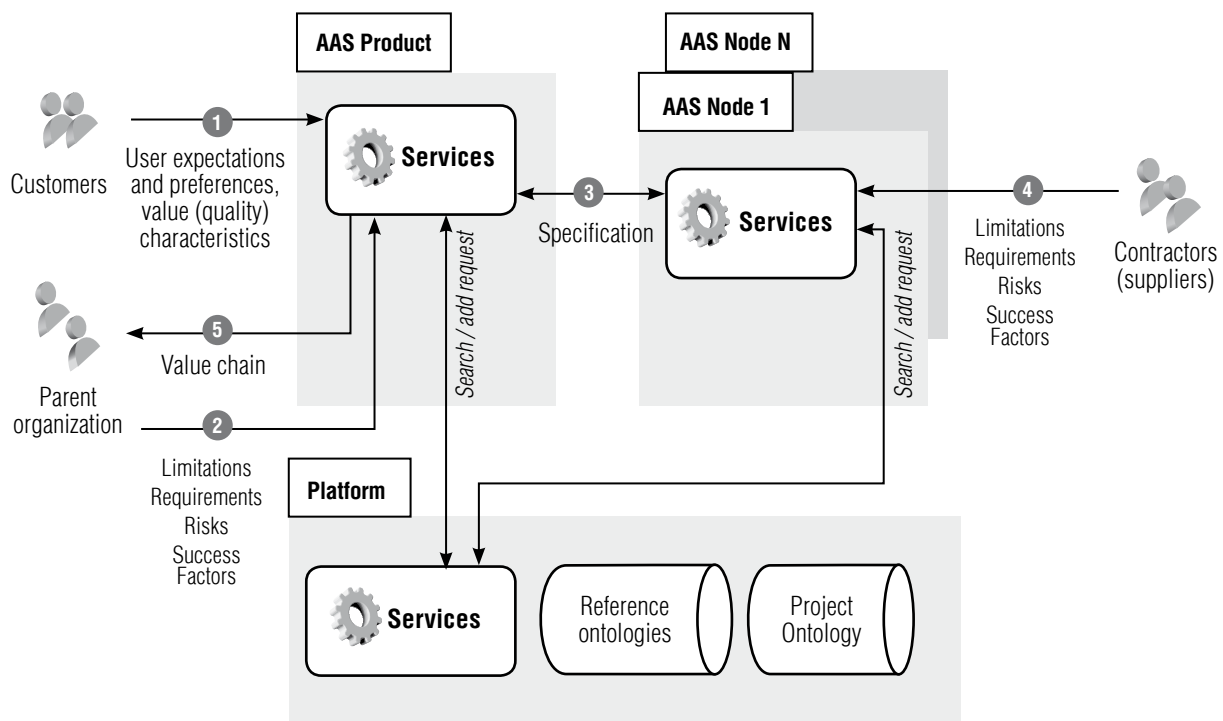


Fig. 4. Scheme of the interaction between participants of a network enterprise in the i4.0 system

for the choice of product design and value-added chain alternatives based on a combination of QFD and fuzzy logic methods. Within the QFD method, when describing each value (quality) characteristic (business requirement), a priority (weight) is established in an expert way, where the most important characteristic with the greatest consumer interest has the maximum weight. The further procedure involves the collection of requirements and an assessment of their impact on value (quality) characteristics, and the selection of the most important requirements. The subsequent assessment of structural elements in terms of their impact on the implementation of requirements is associated with the choice of a product design alternative that better meets the selected requirements. The assessment of risks associated with suppliers and(or) contractors takes into account the generated design alternatives. As a result, the choice of the optimal

value-added chain alternative is carried out, in which requirements and associated risks are determined for all participants, and access to the data of AAS models (distributed repository) is organized. A detailed description of the methodology for using the QFD method is presented in the article [26].

The formation of product design alternatives is initiated in the AAS and goes through a series of iterations [1, 26], within which calls to the services implemented by the AAS modules are made.

The product value for the consumer is determined through the value (quality) characteristics (or business requirements), which are described in the project ontology using the services of the **quality characteristics formation module**, as well as a set of functional and non-functional requirements, the criteria base for assessing the implementation

of requirements, which are described using the services of **the requirements formation module**.

An integrated expression of the cognitive understanding of the market situation for the specific product type finds its representation in the SWOT analysis methodology, which reveals the advantages and disadvantages of the specific product type manufacturing in generalized form, as well as the opportunities and threats of the market implementation of the project [27].

The formation of product quality characteristics is a poorly formalized task of analyzing the external and internal factors of the specific market for products and services. Specialists of various profiles take part in this process: marketers, researchers of new technologies and materials, designers and technologists in the area of similar product types. Solving the problem of selecting the most significant properties of a designed product can use the methods of Big Data analysis of markets of products, technologies and materials, predictive forecasting of development trends, analysis of resource constraints and risks of mastering new types of products and technologies.

Product classification, standard (mandatory and optional) value characteristics, product parameters and some logical rules presented in the reference ontology are used to describe an innovative product (a new class of products) in the project ontology. The product inherits some value characteristics of the category to which it belongs and/or has some unique characteristics not typical for this product category.

With the help of existing services, expected and attractive characteristics are evaluated additionally and classified in the project ontology (according to Kano [28]). Furthermore, the presence of indifferent, contradictory and repellent characteristics is minimized. In this case, the highest priority (weight) according to

QFD will have the mandatory (expected) and attractive characteristics, and the contradictory or repellent characteristics – the least. In the course of the project, some of the optional characteristics, the implementation of which leads to significant project costs, as well as repellent and contradictory characteristics that reduce the value of the product for the consumer, can be excluded from the set.

Information about the value (quality) product characteristics can be refined during the project and is used in the **requirements formation module**, where for each of the identified value (quality) characteristics, at least one requirement is determined, the implementation of which provides the corresponding characteristic. Functional requirements should be distributed across subsystems (subdomains of the subject area).

Typical requirements inherent in specific product categories and specified by standards (for example, technical specifications and GOSTs) are transferred to the project ontology from the reference ontology, and the original requirements and project constraints derived from these requirements (or independent of them) are introduced.

To identify mandatory requirements, the degree of connection of each requirement with the implementation of a quality characteristic is assessed. As a result of using the fuzzification function, the confidence coefficient for assessing such a relationship is reduced to the scale [0, 1]. The total importance for the implementation of the entire set of quality characteristics is calculated for each requirement using the formula [29]:

$$\hat{T}_j = \sum_f \varphi_f(C_{jf}P_f), \quad (1)$$

where  $\hat{T}_j$  – fuzzy assessment of the importance of the  $j$ -th requirement for the implementation of quality characteristics (business requirements) of an innovative product;

$C_{jf}$  – level of connectivity of the  $j$ -th requirement and the  $f$ -th characteristic on the scale [0, 1, 3, 9];

$P_f$  – preset priority of the  $f$ -th characteristic on the scale [1, N];

$\varphi_j$  – fuzzification function on the scale [0, 1]: the more the importance of the  $j$ -th requirement for the implementation of the  $f$ -th characteristic, the closer to 1;

$\sum$  – operation of additive summation of fuzzy numbers.

To determine the final set of groups of mandatory and optional requirements which must ensure the presence of all mandatory and expected characteristics, the specified threshold values of the assessment  $\hat{T}_j$  are used. Expected and optional requirements that are below the threshold are not considered further.

The requirements are detailed down to the level of specific functions and structural elements. They are described in the project ontology within **the structure design module**.

A condition for including a structural element in a product design alternative (bill of materials, BOM) is to ensure the ability to perform some product functions or to implement mandatory interface requirements. The product breakdown structure (PBS) can also be inherited from the reference product ontology. Thus, each design element and function of a product is directly or indirectly related to one or more requirements which define the product design alternative.

Ranking (comparing) different product design alternatives is possible by using QFD services to assess the impact of structural elements on meeting the requirements. In this case, the estimates of the importance of requirements obtained according to the above formula are considered when calculating the specific structural element estimates.

BOM specification is a poorly formalized, iterative and innovative task. In the course of this task, the BOM is defined in such a way as to ensure, on the one hand, full compliance with existing requirements (a design alternative must always include at least one element for each mandatory requirement) and, on the other hand, an appropriate level of costs. A threshold value can be set and used to select optional elements and/or elements with high costs.

The determining factor is the possibility of purchasing the selected types of structural elements and/or their production. To assess a possibility, the generated design variant is passed to the planning module, which calls the services implemented by the value-added chain design module. The market analysis, the choosing of the innovative process alternatives and the selection of its participants are carried out with the help of these services.

The absence from suppliers of a product design structural elements related to mandatory requirements can lead to one of the following organizational scenarios triggered by the planning module:

- ◆ searching for contractors who could provide the manufacturing of the missing structural element using the specified technologies, tools and design constraints;
- ◆ selection according to the rank of another product design alternative, which does not contain the missing elements;
- ◆ revision of the product concept to clarify the characteristics and requirements to eliminate the missing structural element;
- ◆ termination of work due to the lack of the possibility of revising the concept of the product.

To identify businesses that are capable of custom-made missing structural elements, it is necessary to form a value-added chain alternative that includes manufacturing, design, assembly, processing, setting up, maintenance

and delivery of products. The set and connections of activities are contained in the project ontology, where possible tools and technologies can also be determined, as well as compliance with the specified design constraints and established requirements.

Assessment of the total impact of all activities in the value-added chain on ensuring quality characteristics and meeting requirements allows one to select participants in the process, taking into account their competencies (ability to perform certain types of work, possession of tools and technologies). The formula for calculating the deviations of requirements from the capabilities of the enterprise, based on fuzzy logic, is considered below while determining the possibility of performing some types of work on their own or outsourcing them [29]:

$$\hat{V}_{iu} = \sum_j \theta_j (T_{ijj} - R_{ijj}), \quad (2)$$

where  $\hat{V}_{iu}$  – fuzzy assessment of the competitiveness of the  $u$ -th product component on the market supplied (produced) by the  $i$ -th contractor;

$T_{ijj}$  – a given level of the  $j$ -th requirement for the implementation of the  $u$ -th component by the  $i$ -th contractor;

$R_{ijj}$  – level of implementation of the  $j$ -th requirement for the  $u$ -th component by the  $i$ -th contractor;

$\theta_j$  – fuzzification function (0, 1) on the scale, the smaller the deviation, the higher the competitiveness;

$\sum$  – operation of additive summation of fuzzy numbers.

The level of implementation of the  $j$ -th requirement for the  $u$ -th component by the  $i$ -th contractor is determined in the process of the requirement adjustment with the contractor by interpreting the data received from the contractor (on a natural or fuzzy scale [0, 1]).

Accordingly, the choice of the  $i$ -th contractor (supplier or manufacturer) is carried out according to the maximum assessment of the competitiveness of the  $u$ -th supplied (produced) component in the market:

$$\hat{V}_u^* = \max_i \hat{V}_{iu} \hat{K}_{iu}, \quad (3)$$

where  $\hat{V}_u^*$  – maximum assessment of the competitiveness of the  $u$ -th supplied (produced) component;

$\hat{V}_{iu}$  – fuzzy assessment of the competitiveness of the  $u$ -th product component on the market supplied (produced) by the  $i$ -th contractor;

$\hat{K}_{iu}$  – risk ratio related with delivery of the  $u$ -th component by the  $i$ -th contractor.

The risk ratio is calculated as a multiplicative product of fuzzy numbers corresponding to specific risk types, which are determined on a fuzzy scale [0, 1]:

$$\hat{K}_{iu} = \prod_k K_{iuk}, \quad (4)$$

where  $\hat{K}_{iu}$  – risk ratio related with delivery of the  $u$ -th component by the  $i$ -th contractor;

$K_{iuk}$  – risk ratio for the  $k$ -th type of risk related with delivery of the  $u$ -th component by the  $i$ -th contractor;

$\prod$  – fuzzy multiplication operation.

During the search for contractors, **platform services** for ensuring the interaction of AAS are used. By means of them, requests are sent to the contractors. In particular, the specification of the structural elements set in which the initiating organization is interested, the description of the corresponding tools and technologies, the structural element requirements and the design constraints are sent. Through an existing or created AAS, the contractor uses calculation models and the information provided in the enterprise databases to assess the production possibilities. For this, similar services are called, which are imple-



mented at the structural element level by the module for the formation of quality characteristics, the requirements formation module, the structure design module.

If the enterprise does not fully meet the specified conditions, then the appropriate response is sent through the platform services. Otherwise, the results of the calculation models and the structural element characteristics that can be developed are sent. The absence of positive responses to the mailing from the contractors is the basis for switching to other process alternatives according to the previously determined rank.

The selected process alternative with a list of activities and participants, a detailed description of the product and structural elements are saved by platform services in the submodel of the project documentary data of AAS. They are used at the production stage to coordinate work (direct communication between the process participants services), as well as during the maintenance of the product and during the development of new types of products.

### Conclusion

As a result of the research into the technology for designing innovative processes to create products of a network enterprise using the i4.0 knowledge-based system, the following conclusions can be drawn:

It is advisable to decompose the subject area of designing innovative processes for creating products and services based on the Domain-driven design approach, which provides, on the one hand, an objective reflection of the structure of products and services and value-added chains, and on the other hand, the formation of project teams capable of effectively interacting within the assigned tasks.

The knowledge-based system for designing innovative processes to create products and

services can be based on Plattform Industrie 4.0 (i4.0 platform), which provides a multi-agent implementation of the design process in a distributed network architecture using digital twins and digital threads technologies.

Domain-driven design allows us to highlight the bounded contexts of creating products and services by stages of the life cycle, and for the design phase - the sequence of activities of concept formation, functional and non-functional requirements formation, product design and value-added chain design. Accordingly, the structure of an i4.0 system based on domain-driven design decomposition will include sets of interacting Asset Administration Shells (AAS) that implement bounded contexts for design.

The feature of the proposed technology for designing innovative processes to create products and services is its iterative nature. During the process, AAS are dynamically created for digital twins of products and their components, forming and evaluating alternatives of value-added chains for the best fulfilment of functional and non-functional requirements.

The use of standardized ontological services of the i4.0 platform provides dynamic creation of ontologies from reference models and their actualization as an innovative project of creating products and services develops.

Standardized methods of the i4.0 platform in terms of information exchange between AAS allow for active interaction between participants in the innovative process, aimed at well-grounded and coordinated design alternatives.

The functional modules of AAS are based on standardized services that implement cognitive methods: QFD to determine priorities of business requirements and functional requirement implementation; FMEA to risk assessment; a method for processing fuzzy sets to obtain multi-criteria assessments of the advis-

ability of making specific design decisions. The application of these methods allows for multi-criteria optimization of the innovative design process.

Characterizing the proposed technology for designing innovative processes to create products and services, we note its novelty in terms of the proposed methods for implementing multi-agent interaction between participants in the innovative process within the dynamic construction of an i4.0 system and the use of ontological and cognitive methods for devel-

oping design solutions. In the future, there are plans to continue research into the application of the developed methods and mechanisms of i4.0 systems for the implementation of subsequent stages of the life cycle of creating products and services. ■

### Acknowledgments

This study was funded by the Russian Foundation for Basic Research according to the research projects No. 19-07-01137a.

### References

1. Telnov Yu.F., Kazakov V.A., Trembach V.M. (2020) Developing a knowledge-based system for the design of innovative product creation processes for network enterprises. *Business Informatics*, vol. 14, no 3, pp. 35–53. DOI: 10.17323/2587-814X.2020.3.35.53.
2. Bajaj M., Hedberg T. (2018) System lifecycle handler – Spinning a digital thread for manufacturing. *Proceedings of the 28th Annual INCOSE International Symposium. Washington, DC, 7–12 July 2018*, vol. 28, pp. 1636–1650. DOI: 10.1002/j.2334-5837.2018.00573.x.
3. NIST (2020) *Digital thread for smart manufacturing*. Available at: <https://www.nist.gov/programs-projects/digital-thread-smart-manufacturing> (accessed 26 July 2021).
4. Grieves M. (2016) *Origins of the digital twin concept*. Available at: [https://www.researchgate.net/publication/307509727\\_Origins\\_of\\_the\\_Digital\\_Twin\\_Concept](https://www.researchgate.net/publication/307509727_Origins_of_the_Digital_Twin_Concept) (accessed 26 July 2021). DOI: 10.13140/RG.2.2.26367.61609.
5. Minaev V.A., Mazin A.V., Zdiruk K.B., Kulikov L.S. (2019) Digital twins of objects in the solution of control problems. *Radioindustry*, vol. 29, no 3, pp. 68–78 (in Russian). DOI: 10.21778/2413-9599-2019-29-3-68-78.
6. NTI (2019) *Digital twins in the high-tech industry. Expert and analytical report*. Available at: [http://assets.fea.ru/uploads/fea/news/2019/12\\_december/28/cifrovoy\\_dvoynik.pdf](http://assets.fea.ru/uploads/fea/news/2019/12_december/28/cifrovoy_dvoynik.pdf) (accessed 26 July 2021).
7. Wooldridge M. (2009) *An introduction to multi-agent systems*. New York: Wiley, 2009.
8. Gorodetskiy V.I., Grushinskiy M.S., Khabalov A.V. (1998) Multi-agent systems (review). *Artificial Intelligence News*, no 2, pp. 64–116 (in Russian).
9. Telnov Yu.F., Danilov A.V., Diveev R.I., Kazakov V.A., Yaroshenko E.V. (2018) Development of a prototype of multi-agent system of network interaction of educational institutions. *Open Education*, vol. 22, no 6, pp. 14–26 (in Russian). DOI: 10.21686/1818-4243-2018-6-14-26.
10. Uschold M., King M., Moralee S., Zorgios Y. (1998) The enterprise ontology. *Knowledge Engineer Review*, vol. 13, no 1, pp. 31–89. DOI: 10.1017/S0269888998001088.
11. Gavrilova T.A., Kudryavtsev D.V., Muromtsev D.I. (2016) *Knowledge engineering. Models and methods*. Saint Petersburg: Lan' (in Russian).
12. Efimenko I.V., Khoroshevsky V.F. (2011) *Ontological modeling of enterprises and markets of modern Russia. Part 1. Ontological modeling: Approaches, models, methods, tools, solutions. Working paper WP7/2011/08 (part 1)*. Moscow: HSE Publishing House (in Russian).

13. Evans E. (2003) *Domain-driven design: Tackling complexity in the heart of software*. Boston: Addison-Wesley.
14. Vernon V. (2016) *Domain driven design distilled*. Boston: Addison-Wesley.
15. Industrial Internet Consortium, Plattform Industrie 4.0 (2017) *Architecture alignment and interoperability. Joint Whitepaper*. Available at: [https://www.iiconsortium.org/pdf/JTG2\\_Whitepaper\\_final\\_20171205.pdf](https://www.iiconsortium.org/pdf/JTG2_Whitepaper_final_20171205.pdf) (accessed 26 July 2021).
16. Plattform Industrie 4.0. (2019) *Details of the asset administration shell from idea to implementation*. Available at: [https://www.plattform-i40.de/IP/Redaktion/EN/Downloads/Publikation/vws-in-detail-presentation.pdf?\\_\\_blob=publicationFile&v=12](https://www.plattform-i40.de/IP/Redaktion/EN/Downloads/Publikation/vws-in-detail-presentation.pdf?__blob=publicationFile&v=12) (accessed 26 July 2021).
17. Industrial Internet Consortium, Plattform Industrie 4.0 (2020) *Digital twin and asset administration shell concepts and application in the industrial internet and Industrie 4.0. Joint Whitepaper*. Available at: <https://www.iiconsortium.org/pdf/Digital-Twin-and-Asset-Administration-Shell-Concepts-and-Application-Joint-Whitepaper.pdf> (accessed 26 July 2021).
18. Kuznetsov O.P. (2012) Cognitive semantics and artificial intelligence. *Artificial Intelligence and Decision Making*, no 4, pp. 95–105 (in Russian).
19. Trembach V.M. (2017) Cognitive approach to developing the intelligent modules of organizational and technical systems. *Open Education*, vol. 21, no 2, pp. 78–87 (in Russian). DOI: 10.21686/1818-4243-2017-2-78-87.
20. The Plattform Industrie 4.0 (2020) *Details of the asset administration shell. Part 1 – The exchange of information between partners in the value chain of Industrie 4.0 (Version 3.0RC01)*. Available at: [https://www.plattform-i40.de/IP/Redaktion/EN/Downloads/Publikation/Details\\_of\\_the\\_Asset\\_Administration\\_Shell\\_Part1\\_V3.pdf?\\_\\_blob=publicationFile&v=5](https://www.plattform-i40.de/IP/Redaktion/EN/Downloads/Publikation/Details_of_the_Asset_Administration_Shell_Part1_V3.pdf?__blob=publicationFile&v=5) (accessed 26 July 2021).
21. Plattform Industrie 4.0 (2019) *Usage view of asset administration shell. Discussion paper*. Available at: [https://www.plattform-i40.de/IP/Redaktion/DE/Downloads/Publikation/2019-usage-view-asset-administration-shell.pdf?\\_\\_blob=publicationFile&v=4](https://www.plattform-i40.de/IP/Redaktion/DE/Downloads/Publikation/2019-usage-view-asset-administration-shell.pdf?__blob=publicationFile&v=4) (accessed 26 July 2021).
22. Plattform Industrie 4.0 (2019) *Functional view of the asset administration shell in an Industrie 4.0 system environment. Discussion paper*. Available at: [https://www.plattform-i40.de/IP/Redaktion/DE/Downloads/Publikation/Functional-View.pdf?\\_\\_blob=publicationFile&v=5](https://www.plattform-i40.de/IP/Redaktion/DE/Downloads/Publikation/Functional-View.pdf?__blob=publicationFile&v=5) (accessed 26 July 2021).
23. Vashukov Yu.A., Dmitriyev A.Ya., Mitroshkina T.A. (2012) *QFD: Product and process development based on customer requirements and expectations: A methodological guide*. Samara: SSAU (in Russian).
24. Vashukov Yu.A., Dmitriyev A.Ya., Mitroshkina T.A. (2008) *Type, consequence and cause analysis of potential nonconformities (FMEA): A methodological guide*. Samara: SSAU (in Russian).
25. Averkin A.N. (2013) *Fuzzy sets in management and artificial intelligence models*. Moscow: Kniga po Trebovaniyu (in Russian).
26. Telnov Yu.F., Trembach V.M., Danilov A.V., Yaroshenko E.V., Kazakov V.A., Kozlova O.A. (2019) Constructing network enterprise structure to create innovative products. *Open Education*, vol. 23, no 6, pp. 59–73 (in Russian). DOI: 10.21686/1818-4243-2019-6-59-73.
27. Thompson A.A. Jr., Strickland III A.J. (2003) *Strategic management: Concepts and cases*. Boston, Mass.: McGraw-Hill/Irvin.
28. Kano N., Seraku N., Takahashi F., Tsuji S. (1984) Attractive quality and must-be quality. *Journal of the Japanese Society for Quality Control*, vol. 14, no 2, pp. 147–156. DOI: 10.20684/quality.14.2\_147.
29. Kalachikhin P.A., Telnov Yu.F. (2018) Value-added chains formation in network structures of interaction based on intelligent technologies. *Proceedings of the XVI National Scientific Conference on Artificial Intelligence with International Participation (KII–2018), Moscow, 24–27 September 2018*, vol. 1, pp. 106–115 (in Russian).

### About the authors

#### **Yury F. Telnov**

Dr. Sci. (Econ.), Professor;

Head of the Department of Applied Information Technologies and Information Security, Plekhanov Russian University of Economics, 36, Stremyanny Lane, Moscow 117997, Russia;

E-mail: Telnov.YUF@rea.ru

ORCID: 0000-0002-2983-8232

#### **Vasily A. Kazakov**

Cand. Sci. (Econ.);

Leading Researcher, Research Institute “Strategic Information Technologies”, Plekhanov Russian University of Economics, 36, Stremyanny Lane, Moscow 117997, Russia;

E-mail: Kazakov.VA@rea.ru

ORCID: 0000-0001-8939-2087

#### **Andrey V. Danilov**

Senior Lecturer, Department of Applied Information Technologies and Information Security, Plekhanov Russian University of Economics, 36, Stremyanny Lane, Moscow 117997, Russia;

E-mail: Danilov.AV@rea.ru

ORCID: 0000-0002-0433-9701