Prediction of Socio-Economic Indicators of the Megapolis Development on the Basis of the Intellectual Forecasting Information System "SHM Horizon"

Kitova, Olga and Dyakonova, Ludmila and Savinova, Victoria

Plekhanov Russian University of Economics

24 July 2020

Prediction of Socio-Economic Indicators of the Megapolis Development on the Basis of the Intellectual Forecasting Information System “SHM Horizon”

Kitova O.V.[1[0000-0002-1820-0954]], Savinova V.M.[1[0000-0002-0036-3675]], Dyakonova L.P. [1[0000-0001-5229-8070]]

[1] Plekhanov Russian University of Economics, Stremanny per, 36, Moscow, Russia, *dyakonova.lp@rea.ru*

Abstract: The article describes a system of hybrid models ‘SGM Horizon’ as intellectual forecasting information system. The system of forecasting models includes a set of regression models and an expandable set of intelligent models, including artificial neural networks, decision trees, etc. Regression models include systems of regression equations that describe the behavior of forecast indicators of the development of the Russian economy in the system of national accounts. The functioning of the system of equations is determined by scenario conditions set by expert. For those indicators whose forecasts do not meet the requirements of quality and accuracy, intelligent models based on machine learning are used.
Using the ‘SHM Horizon’ tools, predictive calculations were performed for a system of 30 indicators of the social sphere of the City of Moscow using hybrid models, and for8 indicators a significant increase in the quality and accuracy of the forecast was achieved with artificial neural network models.
The process of models building requires considerable time, in this regard, the authors see the further development of the system in the application of the multi-criteria ranking method.

Keywords: Regional economics, Forecasting, Socio-economic indicators, Hybrid models, Machine learning, Neural networks, Decision trees

Acknowledgement: This research was performed in the framework of the state task in the field of scientific activity of the Ministry of Science and Higher Education of the Russian Federation, project "Development of the methodology and a software platform for the construction of digital twins, intellectual analysis and forecast of complex economic systems", grant no. FSSW-2020-0008.

1 Introduction

In order to build a competitive and stable economy Russian Federation has been created a system of state strategic management including development of strategic plans on the state and regional levels, a complex of initiatives and target programs in main spheres of economics. A prerequisite for sustainable development and strengthening of Russia and its regions is a strategy that takes into account systemic problems, the best domestic and international experience and is built on the basis of comprehensive modeling and forecasts.

In the context of economic development, the issues of forecasting and decision-making at the federal and regional levels are particularly relevant in connection with the need for strategic planning and budget policy. In connection with the transition of Russia to the System of National Accounts, it became possible to use the experience gained in countries with market economies in modeling the processes of forecasting indicators of socio-economic development.

In forecasting socio-economic indicators, complex mathematical models are implemented in specialized information systems that allow processing large amounts of data. With the development of digital technologies in all sectors of the economy, modern methods of data analysis and mathematical modeling are actively being introduced to improve business performance at both the corporate and state levels.

Currently, the priorities of the program for building a digital economy in the Russian Federation are the following: the development and implementation of modern methods of data mining in the economy and business and the training of specialists who are able to set and solve information and analytical problems in various fields of activity, use modern advanced methods, technologies and systems.

Many modern information and analytical systems include a time series forecasting module. Polymedia company produces information and analytical programs that allow both online (OLAP) and data mining analysis at the regional and state levels. The specified software allows you to build models of the region, including various areas of the state.

The development of information services for solving crisis management tasks is carried out by IBS. The leader in this area was the company "Prognoz". Currently, the Foresight company continues to develop the platform created by this company. Within the framework of this platform, the construction of regional models is used with the further ability to analyze and predict key indicators. However, the described information-analytical systems are very cumbersome, are very expensive and do not provide for the possibility of expanding the

system of models and supplementing them with intelligent models based on artificial neural networks, decision trees, etc.

The team of the Department of Informatics of Plekhanov Russian University of Economics (PRUE) has built and is developing a system of models for short- and medium-term forecasting of indicators of socio-economic development of Russia an Russian Regions of great power. The system of models includes a set of equations, in total consisting of more than 600 indicators. The developed methodology, models, software and technology tools allow for systematic matching macroeconomic development indicators and indicators of the financial system, foreign economic activity, social sphere and other areas. The system of model is implemented in the working prototype of the forecasting information system "System of Hybrid Models Horizon" (SHM Horizon).

Further we consider the development of models for the regional level starting from country level forecasts and various possible scenarios. Moscow, due to its special role, is positioned as the capital, as an independent subject of the Russian Federation and as the center of the Moscow agglomeration, which determines the spatial development of a vast territory. These three functional roles of Moscow define three levels of consideration and planning: national, regional and local. These roles determined Moscow as an area for testing and verifying our approach.

The purpose of this study was the development of a hybrid system of models and an intellectual system for forecasting indicators of socio-economic development of Moscow, allowing combining various approaches that together provide improved quality and accuracy of the forecast for most indicators.

Methodology

Regression and hybrid approaches to forecasting economic indicators

There are many methods and models used in predicting economic indicators based on the use of regression models (Hastie et al., 2009; Kauffman et al., 2017), artificial neural networks (Dulcet, 2015; Iwaniec & Atink, 2015; Kolesnikov & Konkin, 2015; Laletina, 2015; Lesik, 2015; Medeiros & Pedreira, 2001; Sarajevo & Lezina, 2015; Zhang, 2001), genetic algorithms, and various artificial intelligence methods (Buchatskaya, 2012; Kong et al., 2018; Miao et al., 2007; Wang & Wu, L., 2012). In the framework of this work, a hybrid approach to forecasting is used, which includes the construction of a model based on linear multiple regression equations and the use of intelligent ma-chine learning methods. Hybrid approaches to forecasting are de-scribed in a number of works, including hybrid models based on genetic

algorithms and fuzzy sets (Demidova et al., 2015), dynamic neural networks and evolutionary calculations (Bisoi & Dash, 2015), the support vector method and regression models (Kavousi-Fard et al., 2014), neural networks and linear regression (Nguyen & Cripps, 2001), neural networks and cognitive maps (Papageorgiou & Poczęta, 2017), artificial neural networks and fuzzy logic methods (Singh, 2015; Wang & Ning, 2015).

The basis for the proposed approach is the method of constructing a country model proposed by Robert Lawrence Klein (Klein & Goldberger, 1955). The model is based on the method of building a system of equations of multiple linear regression. This method is widely used to describe economic phenomena, since they are determined by a large number of simultaneously acting factors together.

The constructed model is based on a system of national accounts and includes the main socio-economic indicators of the Russian Federation which form the main blocks (see Kitova et al., 2016; Kitova et al., 2013; Kolmakov & Domozhakov, 2017):

1. Scenario indicators - indicators that are set as scenarios for the development of economic phenomena. In this model, the scenario indicators are the volume of GDP, money supply, interest rate on interbank loans, the price of oil, the dollar exchange rate and the volume of monetary reserves.
2. Macroeconomic indicators are indicators that describe the summary volumes of consumption, production, export, import, etc.
3. State budget indicators - describe the income and expenses of the federal and consolidated budgets.
4. Social indicators - describe the quality of life of the population. The block of social indicators is divided into several blocks:

indicators of the standard of living of the population - characterize the volume of incomes and expenditures of the population, allow assessing the level of poverty;

labor and employment indicators - characterize the state of the labor market;

demographic indicators - include the main indicators of population growth.

5. Indicators of investment and innovation.
6. Foreign trade indicators.

The system of indicators is shown in the Figure 1.

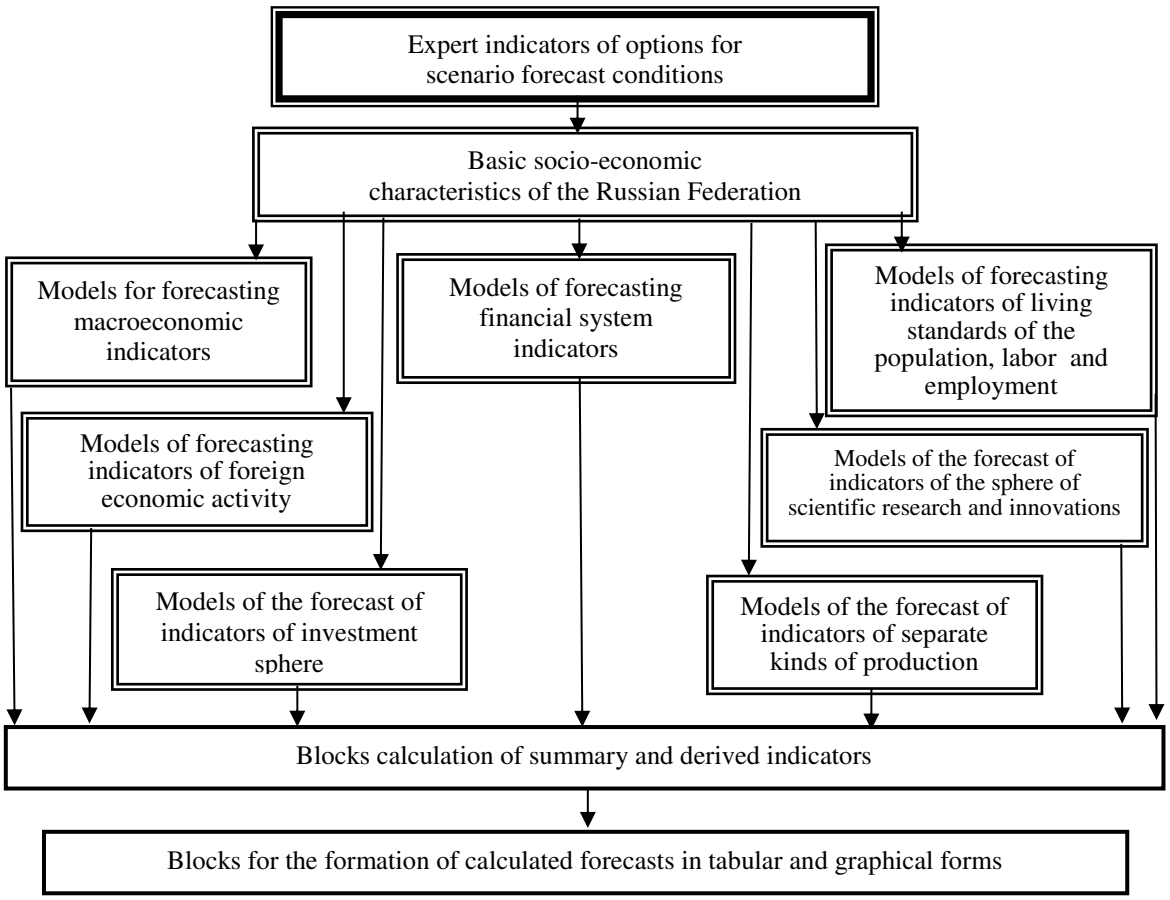

Figure 1: A conceptual scheme of the system of models. Source: Authors

Based on the model for forecasting indicators of the Russian Federation, regional-level models are being built. Based on the scenario indicators of the Russian Federation, forecasting of regional indicators is carried out. In this paper, we consider a model for forecasting indicators of the social sphere using the case of the city of Moscow.

Moscow is one of the key regions of the Russian Federation. About 80 percent of the country's financial turnover, the same share of investments in commercial real estate, 70% of the share of Moscow banks in the total assets of the Russian banking system and a significant part of export-import deductions are concentrated in the region.

The socio-economic sphere of the region can be characterized as unstable. In Moscow, there is a strong dependence of economic development on fluctuations in export prices and the location of the tax base. Fluctuations in indicators arising under the crisis conditions of the economy are difficult to compensate due to the lack of a high export potential and new active drivers of economic and innovative development of the region.

There is an urgent need to build reliable short-term forecasts of the socio-economic activity of the region, which would make it possible to prematurely determine the likelihood of crisis

situations. This would allow adjusting the development strategy of the region depending on the changing key economic indicators.

As a part of the work, a study was conducted on the indicators of the social sphere of the city of Moscow, depending on the scenario conditions of the Russian Federation. Social indicators are grouped into the following categories: demography, living standards, labor.

Statistics published by the Federal State Statistics Service of the Russian Federation from 2000 to 2018 was used as data sources.

The accuracy and quality of forecasts is determined in the forecast verification module, which calculates quality factors (determination coefficient, Darbin-Watson coefficient and Fisher coefficient) and accuracy (average relative error). The values of indicators are set expertly. cators are set expertly.

Table 1: Criteria of quality and accuracy. Source: Authors

Quality assessment settings

coefficient of determination (R^2),	> 0,4
values of Fisher statistics (F-stat)	> 5,0
The Durbin-Watson criterion (DW)	0,8 < DW< 3,2

Accuracy estimation settings (Δ)

High	Middle	Low
<0,06	0,06< Δ <0,16	>0,16

However, not for all indicators the application of the linear multiple regression model allows us to obtain the necessary level of quality and accuracy of forecasts. For these indicators intelligent models, such as neural networks, decision trees, and others are built.

The described approach is currently implemented in the information system "SHM Horizon" (Kitova et al., 2019).

The Fundaments and Functionality of Horizon Hybrid Model

The SHM Horizon system is an analytical forecasting system in which various machine learning models can be used. The system implements the powerful country model described above.

Currently, the main modules are functioning within the system:

Data loading module

Module for constructing regression equations

Module for the construction of artificial neural networks of direct distribution

Module for building regression decision trees (CART).

The system has a microservice architecture, which determines its scalability from the point of view of the developed services (modules). Currently, modules are being developed based on the following fore-casting methods: decision trees, cognitive maps, fuzzy neural networks. However, setting up and calculating the simulation result requires considerable time, since it is not known in advance which model will be preferable for predicting the data set under study. Thus, it is necessary to develop an algorithm that would allow ranking of models and choosing the most appropriate ones before calculating forecasts.

In the framework of this study, the further development of the system using the hybrid decision-making method is proposed, based on the integration of the knowledge representation model of a poorly structured subject area in the form of a fuzzy hierarchy and a multicriteria ranking method for fuzzy objects. A detailed description of the method is described in the works of A.N. Averkin (see Averkin et al., 2007).

The proposed method will allow us to present the process of selecting the relevant model in the form of a fuzzy hierarchy, which it will allow us to compare the models according to given criteria in terms of the possibility of predicting the studied data set. The fuzzy hierarchy method is hybrid and is based on the integration of the knowledge representation model with decision-making methods based on subjective expert assessments.

Implementation of the fuzzy hierarchy method in the SHM Horizon system will allow you to select alternative forecasting options according to the specified criteria, which will reduce the time costs of building models. The functioning algorithm of the developed system is presented in Figure 2.

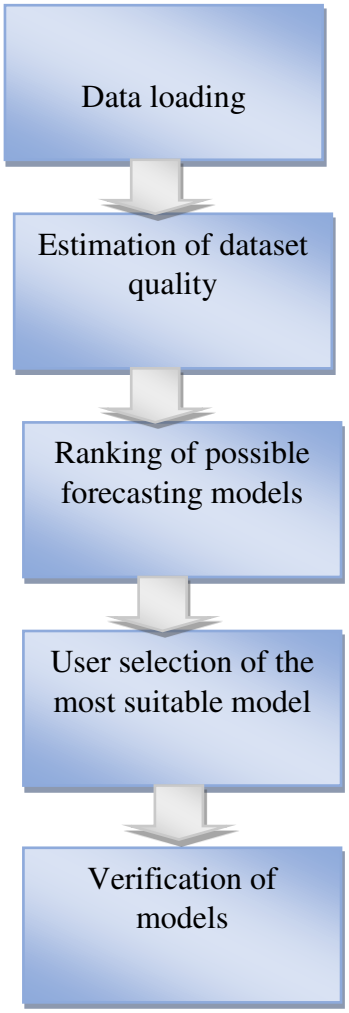

Figure 2: Forecast models building process in the system "SHM Horizon". Source: Authors

Results

Further an example of forecasting indicators of the social sphere for Moscow in the SHM Horizon system is described.

As a result of the ranking of forecasting models for the studied indicators, the model with the highest priority were identified: linear regression and artificial neural network.

At the first stage, models for predicting multiple linear regression for 30 social indicators were built. An example of a calculation result based on the multiple linear regression equations is shown in Figure 3.

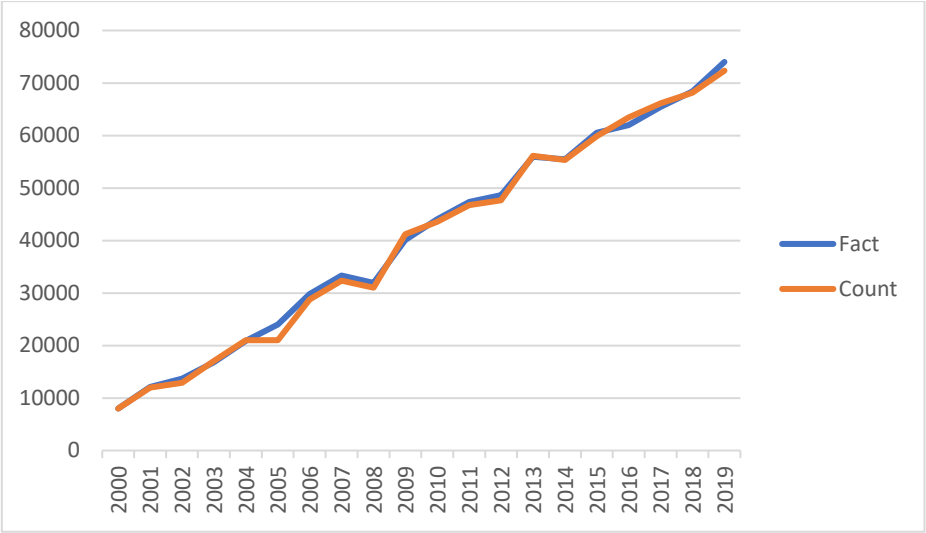

Figure 3: Per capita cash income (per month), rubles. Source: Authors

For the calculations performed, verification of the obtained models was carried out using the previously described criteria. The verification results are shown in the table 2.

Table 2: Forecasts verification results in SHM Horizon. Source: Authors

Accuracy / Quality	High	Medium	Low
High	20	1	1
Low	5	3	

As can be seen from the table, not for all indicators it is possible to build regression equations that would allow to obtain a forecast of high accuracy and quality. For these indicators, models of artificial neural networks were constructed using the architecture of a multilayer perceptron and regression decision trees. As a result, for 8 indicators, improving the quality and accuracy of the forecast was achieved by using models of artificial neural networks. Examples of calculations of retro-forecasts using regression and neural network models for the indicator “Real Estate Acquisition” are presented in Figures 4 and 5.

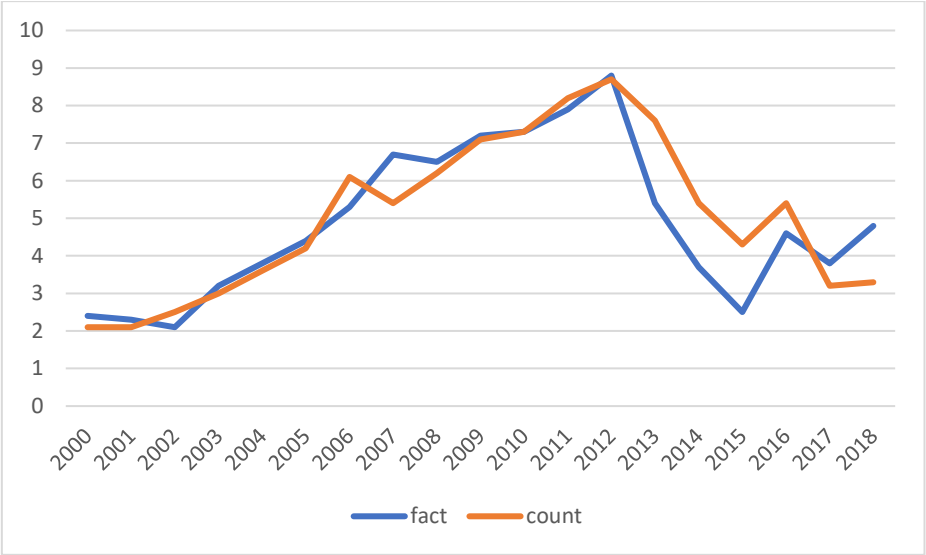

Figure 4: Real Estate Acquisition, as a percentage of total cash income, linear regression model. Source: Authors

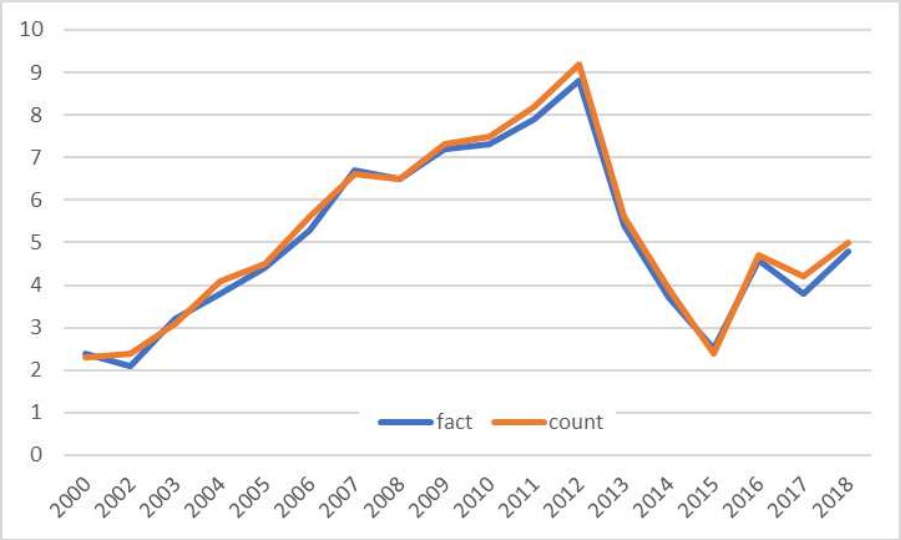

Figure 5: Real Estate Acquisition, as a percentage of total cash income, artificial neural network model. Source: Authors

However, in order to achieve these high quality results, it is necessary to build various network models in several configurations, which requires considerable time and a large load on the system. The ranking method will highlight the most appropriate models based on expert judgment.

Conclusion

The presented study allowed to obtain the following results:

- it was established that in the models describing the system of economic indicators of the country's development, it is necessary to use, along with the multiple regression

model, intelligent methods, such as neural networks and decision trees, to achieve the necessary quality and accuracy values across the entire set of indicators;

- a hybrid approach to the construction of models and forecasts was developed, in which at the first stage a regression model is built for all the studied indicators, then a multiple regression model is verified based on expert estimates of quality and accuracy values, and at the third stage, intelligent models are used for indicators with unsatisfactory values based on machine learning;
- the author's forecasting information system "SHM Horizon" has been improved, which has a microservice architecture, which determines its scalability from the point of view of the modules being developed: modules based on models of decision trees and neural networks have been improved;
- by means of the SGM Horizon system, forecasting was performed for a system of 150 indicators of the social sphere using hybrid models;
- for 112 of 150 indicators, the regression model showed high and satisfactory values of quality and accuracy; for 20, the improvement in the quality and accuracy of the forecast was achieved by using artificial neural network models and models of regression decision trees.

In order to predict the values of scenario indicators in a crisis, it is also possible to take into account the political situation in the world. For this purpose, in the future it is planned to include into the system modules for analysis of the content of news portals and for big data processing.

References

Averkin, A.,N., Titova, N.,V., and Agrafonova T.,V. (2007). Synthesis of distributed fuzzy hierarchical model in decision support systems in fuzzy environment. New Dimensions in Fuzzy Logic and Related Technologies. *Proceedings of the 5th EUSFLAT 2005 Conference*, 377-379.

Bisoi, R., and Dash, P. (2015). Prediction of financial time series and its volatility using a hybrid dynamic neural network trained by sliding mode algorithm and differential evolution. *International Journal of Information and Decision Sciences,* 7 (2), 166–191.

Buchatskaya, V., V. (2012). Algorithm of forecasting methods choice in the study of complex systems. *Bulletin of Adyghe state University. Series 4: Natural-mathematical and technical Sciences*, 3, 136-140.

Demidova, L.,A., Pylkin, A.,N., Skvortsov, S.,V., and Skvortsova, T.,S. (2015). Hybrid models for predicting short time series. Moscow: Hotline - Telecom.

Dulcet, E., I. (2015). Forecasting using artificial neural networks. *Applied mathematics and fundamental Informatics,* 2, 118-126.

Hastie, T., Tibshirani, R., and Friedman, J., and (2009). The Elements of Statistical Learning. Data Mining, Inference, and Prediction. Second edition. NewYork: Springer

Iwaniec, D.,V., and Atink, D., K. (2015). Forecasting using neural networks. *Applied mathematics and fundamental Informatics*, 2, 126-131.

Kauffman, R.,J., Kim, K., Lee, S-Y.,T., Hoang, A-P., and Ren, J. (2017). Combining machine-based and econometrics methods for policy analytics insights. *Electronic Commerce Research and Aplications*, 25 (C), 115-114.

Kavousi-Fard, A., Samet, H., and Marzbani, F. (2014). A new hybrid modified firefly algorithm and support vector regression model for accurate short term load forecasting. *Expert Syst. Appl.*, 41, 6047-6056.

Kitova, O.,V., Savinova, V.,M., Dyakonova, L.P., and Kitov, V.V. (2019). Development of hybrid models and a system for forecasting the indicators of the russian economy. *Espacios*, 40 (10), 18–29.

Kitova, O.,V., Kolmakov, I.,B., and Sharafutdinova, A.,R. (2013). Analysis of accuracy and quality of a short-term forecast of indicators of socio-economic development of Russia. *Bulletin of the Russian Economic University G.V. Plekhanov*, 9 (63), 111-119.

Kitova, O.,V., Kolmakov I.,B., Dyakonova, L.,P., Grishina, O.,A., Sekerin, V.,D., and Danko, T.,P. (2016). Hybrid intelligent system of forecasting of the socio-economic development of the country. *IJABER*, 14 (9), 5755-5766.

Klein, L.,R., and Goldberger A.,S. (1955). An econometric model of the United States, 1929-1952. Amsterdam: North-Hollard.

Kolesnikov, A., N., and Konkin, U., V. (2015). Neural Network Modeling for Time Series Prediction. *Dynamics of complex systems - XXI Century*, 9 (3), 10-13.

Kolmakov, I.,B., and Domozhakov, M.,V. (2017). Methodology for predicting indicators of the sphere of scientific research and innovation using neural networks. *Management Sciences*, 7(1), 53-62.

Kong, W. et al. (2018). Short-Term Residential Load Forecasting based on Resident Behaviour Learning. *IEEE Transactions on Power Systems*, 33 (1), 1087-1088.

Laletina, A.,I. (2015). Forecasting financial markets using artificial neural networks. *Electronic tools and control systems*, 1-2, 214-217.

Lesik, A., I. (2015). The solution of the problem prediction using neural networks of direct distribution, for example, forecasting the growth rate of the STOCK. *Software products and systems*, 2, 70-74.

Medeiros, M.,C., and Pedreira, C.,E. (2001). What are the effects of forecasting linear time series with neural networks? *Engineering Intelligent Systems,* 9, 237–424.

Miao, K., Chen, F., and Zhao, Z.,G. (2007). Stock price forecast based on bacterial colony RBF neural network. *Journ. of QingDao University,* 20, 50–54

Nguyen N., and Cripps, A. (2001). Predicting housing value: a comparison of multiple regression analysis and artificial neural networks, *JRER*, 3, 314–336.

Papageorgiou, E.,I., and Poczęta, K. (2017). A two-stage model for time series prediction based on fuzzy cognitive maps and neural networks. *Neurocomputing*, 232, 113.

Sarajevo, K., V., and Lezina, V., I. (2015). Forecasting oil prices with neural network of Volterra. In the collection: *Royal read the XIII international youth scientific conference, proceedings*, 104.

Singh, P. (2015). Big data time series forecasting model: a fuzzy-neuro hybridize approach. Computational Intelligence for Big Data Analysis, 55–72.

Wang, J., S., and Ning, C., X. (2015). ANFIS Based Time Series Prediction Method of Bank Cash Flow Optimized by Adaptive Population Activity PSO Algorithm. *Information*, 6(3), 300-313.

Wang, S., and Wu, L. (2012). An improved PSO for bankruptcy prediction. *Adv. Comput. Math. Appl.*, 1, 1-6.

Zhang, G.,P. (2001). An investigation of neural networks for linear time-series forecasting. *Computers & Operations Research*, 28, 1183–1202.