Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 221 (2023) 442–449

Procedia Computer Science

www.elsevier.com/locate/procedia

10th International Conference on Information Technology and Quantitative Management

Comparison of Forecasting Power of Statistical Models for GDP Growth Under Conditions of Permanent Crises for Application in Strategic Risk Controlling

Maria Lashina[a], Sergey Grishunin[a]*

[a]Higher School of Economics, 11 Pokrovsky Boulevard, Moscow, Russia

Abstract

The study evaluates the effectiveness of combining different forecasting models to predict Russia's GDP growth rates for the upcoming quarter. The ensemble model utilized in this study consists of a dynamic factor model (DFM) and a neural network with long- and short-term memory (LSTM). The research compared the root-mean-squared errors (RMSE) of the ensemble model with other popular models such as ARIMA, VAR, SVR, and CatBoost, and found that the proposed ensemble model performed better than the LSTM and competitor models but did not improve upon the DFM forecasts. Additionally, the study identified key indicators with high predictive power for the Russian economy by analyzing the DFM eigenvectors and LSTM integrated gradient coefficients.

Keywords: macroeconomic forecasting; dynamic factor model; machine learning; neural network; integrated gradients; uncertainty indices.

1. Introduction

Anticipating the future state of the economy is crucial for authorities to effectively manage the country, and for economic agents to make strategic decisions while being aware of the potential consequences in upcoming environments. To tackle this challenge, various techniques, including statistical and machine learning methods, have been utilized in macroeconomic forecasting, particularly for GDP growth rate. However, traditional models have proven to be less accurate during periods of economic crises due to the complexity of the forecasting task, which involves rare and diverse structural gaps that depend on numerous factors and can be caused by exogenous events. This paper proposes an ensemble model that combines the Dynamic Factor Model (DFM), as in (Bok, Caratelli, Giannone, Sbordone, & Tambalotti, 2017), and Long Short-Term Memory (LSTM) neural network to capture both

* Corresponding author.
E-mail address: mdlashina@edu.hse.ru (M. Lashina), sgrishunin@hse.ru (S. Grishunin)

10.1016/j.procs.2023.07.059

linear and nonlinear dependencies in the data. The model's nowcasting performance for Russian GDP growth is evaluated through a one quarter forecast horse-race between autoregressive moving average with exogenous factors (ARMAX), vector autoregressive (VAR), support vector regression (SVR), boosting, DFM, LSTM, and the ensemble model. Furthermore, the study examines the predictive potential of 75 monthly and quarterly predictors, including economic policy uncertainty and geopolitical risk indices (Baker, Bloom, Davis, & Terry, 2020). The neural network individual forecasts are interpreted using integrated gradients (IG) to identify relatively important indicators.

2. Literature Review

The fundamental concept behind static factor analysis, which preceded DFMs, is to identify several unobservable factors that account for the volatility of a larger set of time series (Lawley & Maxwell, 1971). The intertemporal dynamics of these unobservable factors were then considered in DFMs. Chamberlain & Rothschild (1983) introduced the notion of an "approximate" factor structure, which relaxed the assumption that there is no cross-correlation of idiosyncratic errors. Additionally, principal component analysis (PCA) was proposed for estimating unobservable factors by Forni, Hallin, Lippi, & Reichlin (2000), allowing DFMs to handle larger numbers of variables. Despite the recent availability of vast amounts of data for researchers, time series often have different frequencies. For example, national accounts are published quarterly, while many economic indicators are released monthly. To address this issue and simultaneously model both periodicities within a single framework, Mariano & Murasawa (2003) suggested a specification that treats quarterly series as monthly series with missing values. Following recent advances in machine learning, intelligent methods have been proposed to predict a financial downturn in the economy. For example, Agu, Onu, Ezemagu and Oden (2022) proved principal component regression to forecast GDP more accurately compared to traditional statistical methods. Nevertheless, among intelligent approaches, neural networks (NNs) attract the attention of researchers most of all due to their natural application to time series. For example, Jena, et al. (2021) proposed the design of RNN for predicting financial time series. Several studies have shown that LSTM is the most preferred method due to the ability to capture the dependencies between the values of a series at different points in time. Shijun, Xiaoli, Yunbin, & Chong (2021) applied LSTM to a data set of the futures index of agricultural products, which is believed to be able serve as an early warning signal about a possible crisis. The study demonstrates the superiority of LSTM over the common statistical model – ARIMA. However, complex NNs comprise many neurons and layers, which leads to problems with interpretability and their rejection in favor of "classical" models (Kurihara & Fukushima, 2019). Nonetheless, significant progress has been made in the field of NNs' results interpretation in recent years. Global interpreting methods aim to describe the general behavior of models by using a separate understandable model that mimics the original algorithm. However, this approach is not applicable for deep learning models that rely on non-linear relationships in the data. Therefore, local interpreted algorithms such as the Local Interpretable Model-agnostic Explanations method (Ribeiro, Singh, & Guestrin, 2016) or Shapley values are used. In this study, we use the IG method (Sundararajan, Taly, & Yan, 2017) to interpret LSTM forecasts of GDP growth rates and identify the most relevant indicators. Approach to improving forecast accuracy is through ensemble modeling, which combines the strengths of multiple models. Longo, Riccaboni, & Rungi (2022) demonstrated that an ensemble consisting of DFM and LSTM outperforms other modeling methods, including random walk, VAR, random forest and boosting. The method of combining models typically involves obtaining a weighted average of model forecasts (Clark & McCracken, 2010). Some researchers use an additional model to determine the weights of the ensemble components (Qingwen, Chengming, & Guangxi, 2022). However, studies demonstrating the effectiveness of using model errors to assign weights to ensemble components seem more promising (Lu, 2021; Siqi, 2022). This study contributes to the discussion on improving forecast accuracy through ensemble models. The model is constructed in accordance with the following structure: the final forecast of the ensemble is the sum of the forecasts of DFM trained to predict Russia's GDP growth, and LSTM that predicts the DFM's forecast error.

3. Data and Methodology

The study utilizes monthly indicators and quarterly values for certain indicators, including real GDP, confidence indices in the Construction and Services sectors, consumer confidence index, and real investments in fixed assets index. The data is collected from 1995 to the present time. In total, 75 explanatory indicators were collected, which could be divided into four categories: (1) leading indicators – 35 variables, for example, the indicators of the quarterly bulletin "Russian Economic Barometer" and the Rosstat business confidence indices; (2) real sector indicators – 15 variables: industrial production indices by industry, the volume of exports and imports, the unemployment rate; (3) financial indicators – 23 variables: interest rates, monetary aggregates, the exchange rate, the MOEX stock index, business activity indicators in trading partner countries (China, Germany, USA), export goods prices; and (4) uncertainty – 2 variables, namely the economic policy uncertainty and geopolitical risk indices. In this study, the DFM described in the study by Bok, Caratelli, Giannone, Sbordone, & Tambalotti (2017) is utilized. The N observed variables ($y1,t$, ⋯ , yN,t) are determined by r unobservable dynamic factors ($f1,t$, ⋯ , fr,t), while the features of individual series, for example, measurement errors, are captured by idiosyncratic errors ($e1,t$, ⋯ , eN,t). The empirical model can be generalized by the following equation:

$$y_{i,t} = \sum_{j=1}^{r} \lambda_{i,j} f_{j,t} + e_{i,t},\ i = 1, \ldots, n \tag{1}$$

Common factors and idiosyncratic components are modeled as Gaussian autoregressive processes to account for their sequential correlation:

$$f_{j,t} = \sum_{l=1}^{p} a_{j,t} f_{j,t-1} + u_{j,t},\ u_{j,t} \sim i.i.d.\ N\left(0, \sigma^2_{u_j}\right) for\, j = 1, \ldots, r \tag{2}$$

$$e_{i,t} = \sum_{k=1}^{q} \rho_{i,k} e_{i,t-k} + \varepsilon_{i,t},\ \varepsilon_{i,t} \sim i.i.d.\ N\left(0, \sigma^2_{\varepsilon_i}\right) for\, i = 1, \ldots, n \tag{3}$$

Eq. 1 is known as the measurement equation and relates predicted data to unobservable dynamic factors. Eq. 2 and 3, known as the transition equations, describe the dynamics of the system.

Quarterly indicators are assumed to allow the same representation as the monthly indicators (Eq. 1). To link the measurement equation x_t with the observed quarterly data x_t^Q, partially observed monthly series are constructed using the Mariano & Murasawa (2003) approximation:

$$x_t^Q = \begin{cases} \gamma x_t + 2\gamma x_{t-1} + 3\gamma x_{t-2} + 2\gamma x_{t-3} + \gamma x_{t-4},\ t = 3, 6, 9\ldots \\ unobservable, otherwise \end{cases} \tag{4}$$

In practice, DFM parameter estimates are calculated iteratively with two stages. At the first stage, $\lambda i,j$ and fj,t are estimated using PCA. The coefficients of the transition equations are estimated with the ordinary least squares (OLS) procedure. This is a good approach, especially when working with big data, given that the main components are reliable estimates of common factors (Stock & Watson, 2002). The second stage uses Kalman smoothing (Durbin & Koopman, 2012) for coefficient revaluation, given the parameters obtained at the previous stage. The maximum likelihood estimate is obtained by repeating two steps until convergence. The forward passing process of artificial neural networks involves multiplying a set of numerical inputs by weights, resulting in linear combinations. These linear combinations are then transmitted through the network, activating neurons with nonlinear activation functions. This activation process occurs in one or more layers until the output signal is calculated. In a multilayer perceptron, which is one of the most common NNs, at each time step t, the neuron receives a set of input data measured at time t. In RNN, neurons receive input data measured at time t, as well as output data generated in $t - 1$. In this sense, RNN stores in memory the previous output signal, which is a nonlinear combination of the input data measured in the previous step. Eq. 5 represents the output value generated at time t:

$$\boldsymbol{h}_t = \varphi\left(\boldsymbol{X}_t \cdot \boldsymbol{W}_X + \boldsymbol{h}_{t-1} \cdot \boldsymbol{W}_y + \boldsymbol{b}\right) \tag{5}$$

In this model, ht is a function of Xt , $ht{-}1$ is a function of $Xt{-}1$ and $ht{-}2$, and so on. This makes the output at time t a function of all previous inputs with a time step. The recursive RNN structure is optimal for time series analysis because it stores previous steps information. The LSTM model, introduced by Hochreiter & Schmidhuber (1997), is one of the RNN types. The main advantage of LSTM is the ability to store information about the previous values of the input data for a long time, which allows it to better process complex data sequences. The integrated gradients (IG) methodology was used to understand the role of indicators in the creation of NN forecasts. IG coefficients are calculated by comparing the prediction of a trained model with the prediction of a model without certain indicator.

4. Results

The time series of all indicators have been transformed to ensure stationarity, which is the main assumption for DFMs, ARMAX and VAR models. The data was split into three sets: training, validation, and test. The training set covers 84 observations from 1996Q1 to 2016Q4, while the validation set includes the last 16 observations of the training set (2013Q4-2016Q4), and the test set comprises the last 24 observations (2017Q1-2022Q4). Normalization of the data was performed to simplify the analysis of patterns of underlying input data with different scaling. Median values were used to fill gaps in the data. Additionally, a rolling window method was employed to generate data samples for training NN and forecasting. In this study, four unobservable dynamic factors were used for the DFM: (1) global, which includes all indicators, should reflect interactions between indicators of different nature; (2) leading includes leading indicators, as well as uncertainty indices; (3) real – real sector indicators, uncertainty indices, as well as GDP growth rate; (4) financial comprise financial indicators and uncertainty indices. To prevent overfitting and speed up calculations, a single lag was selected for the transition equations of dynamic factors and idiosyncratic errors. The model's hyperparameters were optimized using Random and Grid Search procedures. Table 1 displays the RMSE of the model's quarter-ahead predictions on both the training and test samples. The results indicate that the DFM outperforms other statistical methods, which is consistent with previous literature (Chernis & Sekkel, 2017; Ponomarev & Pleskachev, 2018). However, previous research (Loermann & Maas, 2019; Hopp, 2022) did not report the DFM's superiority over machine and deep learning methods. Another potential explanation for the DFM's superior performance is its ability to handle data of different frequencies (i.e., quarterly and monthly) without compromising their structure, while other models require monthly indicators to be converted into three separate indicators representing each month of the quarter.

Table 1. Comparison of model’s forecasts RMSE.

	RMSE	
	Train set	Test set
DFM	2,76	1,59
LSTM	2,21	3,19
Ensemble DFM, LSTM (error correction)	1,32	2,42
Ensemble DFM, LSTM (equal weights)	2,35	1,95
ARMAX	0	8,56
VAR	5,91	3,95
SVR	0,49	3,49
CatBoost	0,27	3,23

Regarding the test sample, the DFM's forecasts (Fig. 1) almost match the actual values during the period of economic stability from 2017Q1 to 2019Q4, and even during the COVID-19 outbreak in 2020Q2. However, the forecasts deviate from the actual values from 2021Q1 to 2022Q1, which may be due to the unusual nature of the crisis and the subsequent recovery period of the Russian economy. Nevertheless, the model provided accurate forecasts for the decline in economic activity in 2022.

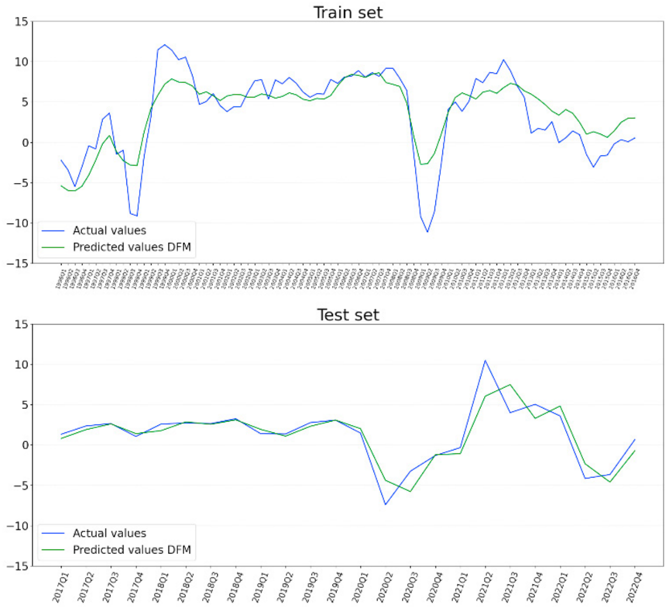

Fig. 1. DFM forecasts for the Russia's real GDP growth rate.

The RMSE of machine learning models forecasts (SVR, CatBoost, LSTM) is less than that of forecasts of econometric models (ARIMAX, VAR). The relative superiority in prediction accuracy of machine learning models is consistent with previous findings in the literature (Ahmed et al. (2010); Richardson et al. (2020); Teräsvirta et al. (2005)). The poor quality of the ARIMAX model forecasts is due to its non-adaptive nature, which requires periodic re-evaluation and even re-identification of the "naive" model. VAR might ambiguously determine the influence of variables on GDP growth due to the large number of estimated parameters. The instability of SVR to outliers in the data may have caused the model to assign significant weights to noise sources instead of significant indicators for crisis periods during training. Boosting algorithms are sensitive to the distribution of data, and more complex methods of filling missing values could be used to improve quality. The neural network's inability to accurately predict recent crises may be due to insufficient data to train the network. Fig. 2 shows that ensemble models, incorporating both averaging and error-correction structures, did not enhance the forecasting accuracy of the DFM. The error-correction ensemble model exhibited a decrease in RMSE on the training sample, yet an increase in error on the test sample. Consequently, this ensemble structure is susceptible to overfitting and requires simpler error-correction models than neural networks. Conversely, the ensemble model, combining weighted average forecasts of the DFM and LSTM, demonstrated better performance on the test sample than on the training sample. However, the neural network was underfitted due to the lack of enough data, and the accuracy of the ensemble forecasts is lower than that of the DFM alone.

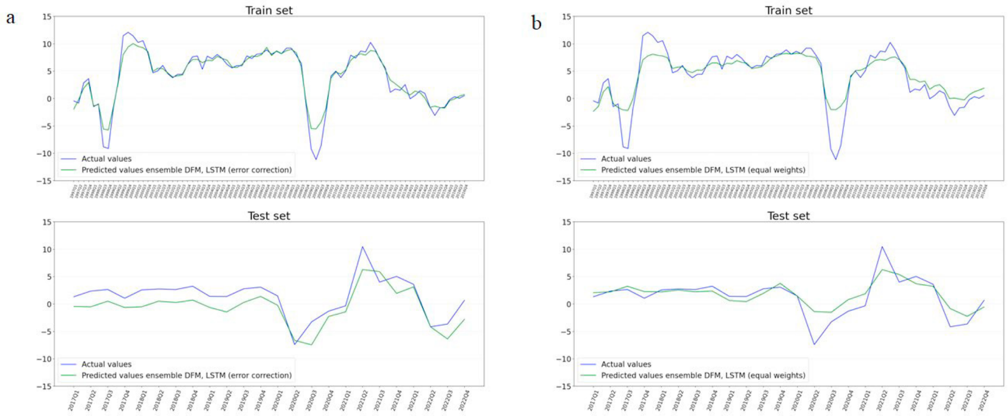

Fig. 2. (a) forecasts of the ensemble model with error correction structure for the Russia's real GDP growth rate; (b) forecasts of the ensemble model with weighted average structure for the Russia's real GDP growth rate.

Therefore, the use of ensemble models did not improve the one-quarter ahead GDP growth rate forecasts for Russia compared to the DFM due to insufficient data for neural network training. To evaluate the impact of uncertainty indices on forecast accuracy, similar models were trained and tested on a data sample without these indicators. Table 2 presents the results of comparing the forecast quality of models in both samples. In general, excluding uncertainty indices from the data led to a decline in most models' accuracy. However, the DFM maintained its forecasting accuracy level without these indicators, indicating limited predictive power.

Table 2. Comparison of model's forecasts RMSE using data with and without uncertainty indices.

	RMSE on test set	
	With uncertainty data	Without uncertainty data
DFM	1,59	1,59
LSTM	3,19	3,32
Ensemble DFM, LSTM (error correction)	2,42	2,37
Ensemble DFM, LSTM (equal weights)	1,95	2,11
ARMAX	8,56	7,45
VAR	3,95	3,96
SVR	3,49	3,51
CatBoost	1,59	1,59

The significance of factors in a DFM is determined through the application of the principal components' algorithm, which assigns the extent of variability in the data explained by the variables to their eigenvectors. Table 3 displays the 10 most valuable monthly indicators for the unobserved Global factor, with six of them being leading.

Table 3. The 10 most important monthly indicators for the DFM unobservable factor Global.

Variable	Eigenvector in Global unobservable factor
Diffuse employment index, expected changes (share of enterprises with an indicator growing in 3 months)	0,27
Diffuse index of equipment purchases, expected changes (share of enterprises with an indicator growing in 3 months)	0,26
Diffuse output index, expected changes (share of enterprises with an indicator growing in 3 months)	0,25
Overdue accounts receivable	-0,24
Overdue accounts payable	-0,24
Unemployment rate	-0,23

Diffuse index of the order portfolio, actual changes (the share of enterprises with a growing indicator for 1 month)	0,22
Commercial cargo turnover of transport	0,22
Diffuse wage index, actual changes (share of enterprises with an indicator growing in 1 month)	0,21
Diffuse output index, actual changes (share of enterprises with an indicator growing in 1 month)	0,20

To evaluate the contribution of indicators to the GDP forecast of a neural network during the business cycle, the integrated gradients approach is utilized. This algorithm provides local coefficients for all variables in the test dataset, which are then aggregated by averaging over groups of variables. A high coefficient indicates that a group of indicators has a significant (positive or negative) impact on Russia's GDP growth rate. Fig. 11 illustrates the coefficients of integrated coefficients for groups of indicators that were formed by indicators reflecting similar areas of the economy.

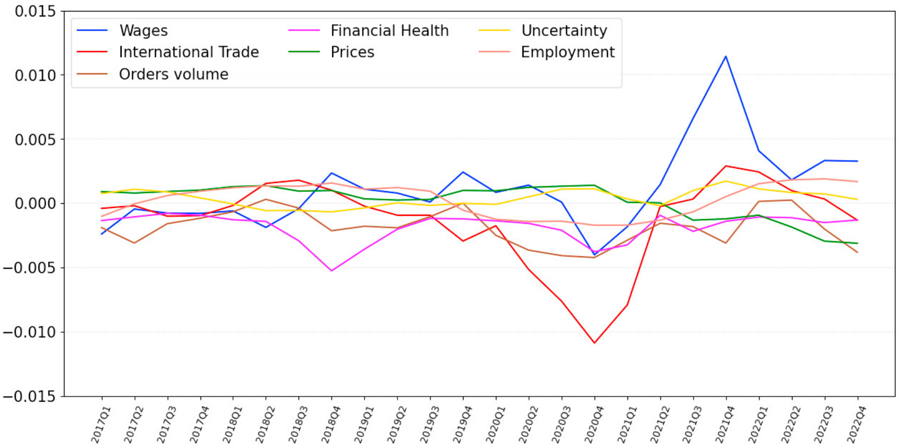

Fig. 3. IG coefficients for groups of indicators for LSTM forecasts.

5. Conclusion

In this study, our aim was to evaluate the forecasting effectiveness of ensemble methods compared to a set of models, including traditional statistical and machine learning algorithms. We found that combining DFM and LSTM in an ensemble provided higher accuracy of forecasts than the LSTM and competitor models. However, it did not improve the quality of forecasts compared to a separate DFM. The DFM demonstrated the most accurate forecasts for the Russia's GDP growth rates, given the limited time horizon of data. Additionally, we analyzed the predictive power of various indicators. The dynamic factor model identified the indices of expected employment, equipment purchases, and output as the most significant indicators. To interpret the role of various indicators in neural network forecasts, we used the IG method, which showed that the set of the most influential indicators varies over time. Indicators of international trade, indices of actual and expected employment and wages were more important for predicting the crisis of the COVID-19. On the other hand, the dynamics of the crisis in 2022 is well predicted with the indicators of the orders volume in the industry and price indices for manufactured and purchased products as they could capture the decline in economic activity after the introduction of sanctions.

6. Acknowledgement

This work has been supported by the grants the Russian Science Foundation, RSF 23-18-00756

References

[1] Agu, S. C., Onu, F. U., Ezemagu, U. K., & Oden, D. (2022). Predicting gross domestic product to macroeconomic indicators. Intelligent Systems with Applications, 14.

[2] Ahmed, N. K., Atiya, A. F., Gayar, N. E., & El-Shishiny, H. (2010). An empirical comparison of machine learning models for time series forecasting. Econom Rev, 594-621.

[3] Aruoba, S., Diebold, F. X., & Scotti, C. (2009). Real-time measurement of business conditions. Journal of Business and Economic Statistics, 417-427.

[4] Baker, S. R., Bloom, N., Davis, S. J., & Terry, S. J. (2020). Covid-induced economic uncertainty.

[5] Bok, B., Caratelli, D., Giannone, D., Sbordone, A., & Tambalotti, A. (2017). Macroeconomic Nowcasting and Forecasting with Big Data.

[6] Burns, A. F., & Mitchell, W. C. (1946). Measuring Business Cycles. NBER Book Series Studies in Business Cycles.

[7] Chamberlain, G., & Rothschild. (1983). Arbitrage, factor structure and mean-variance analysis in large asset markets. Econometrica, 1305-1324.

[8] Chernis, T., & Sekkel, R. (2017). A dynamic factor model for nowcasting Canadian GDP growth. Empirical Economics, 53(4).

[9] Chu, B., & Qureshi, S. (2022). Comparing Out-of-Sample Performance of Machine Learning Methods to Forecast U.S. GDP Growth. Computational Economics.

[10] Clark, T. E., & McCracken, M. W. (2010). Averaging Forecasts from VARs with Uncertain Instabilities. Journal of Applied Econometrics, 25(1), 291-311.

[11] Coulombe, P. G., Marcellino, M., & Stevanovic, D. (2021). Can Machine Learning Catch the COVID-19 Recession? National Institute Economic Review, 256, 71-109.

[12] Durbin, J., & Koopman, S. J. (2012). Time series analysis by state space methods. Oxford university press.

[13] Forni, M., Hallin, M., Lippi, M., & Reichlin, L. (2000). The generalized dynamic-factor model: Identification and estimation. Review of Economics and Statistics, 540-554.

[14] Foroni, C., Marcellino, M., & Stevanovic, D. (2022). Forecasting the Covid-19 recession and recovery: Lessons from the financial crisis. International Journal of Forecasting, 38.

[15] Giannone, D., Reichlin, L., & Small, D. (2008). Nowcasting: the real-time informational content of macroeconomic data. J Monet Econ, 55(4), 665-676.

[16] Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. Neural Comput, 9(8), 1735-1780.

[17] Hopp, D. (2022). Economic Nowcasting with Long Short-Term Memory Artificial Neural Networks (LSTM). Journal of Official Statistics, 847-873.

[18] Jena, P. R., Majhi, R., Kalli , R., Managi, S., & Majhi, B. (2021). Impact of COVID-19 on GDP of major economies: Application of the artificial neural network forecaster. Economic Analysis and Policy, 69.

[19] Kurihara, Y., & Fukushima, A. (2019). AR Model or Machine Learning for Forecasting GDP and Consumer Price for G7 Countries. Applied Economics and Finance.

[20] Lawley, D. N., & Maxwell, A. E. (1971). Factor Analysis in a Statictical Method.

[21] Loermann, J., & Maas, B. (2019). Nowcasting US GDP with artificial neural networks.

[22] Longo, L., Riccaboni, M., & Rungi, A. (2022). A neural network ensemble approach for GDP forecasting. Journal of Economic Dynamics & Control, 134.

[23] Lu, S. (2021). Research on GDP Forecast Analysis Combining BP Neural Network and ARIMA Model. Computational Intelligence and Neuroscience.

[24] Mariano, R., & Murasawa, Y. (2003). A new coincident index of business cycles based on monthly and quarterly series. Journal of Applied Econometrics, 427-443.

[25] Ponomarev, Y., & Pleskachev, Y. (2018). Short-term GDP forecasting using a dynamic factor model. Economic development of Russia.

[26] Porshakov, A., Ponomarenko, A., & Sinyakov, A. (2016). Estimation and forecasting of Russia's GDP using a dynamic factor model. Journal of the New Economic Association, 60-76.

[27] Qingwen, L., Chengming, Y., & Guangxi, Y. (2022). A New Multipredictor Ensemble Decision Framework Based on Deep Reinforcement Learning for Regional GDP Prediction. IEEE Access.

[28] Ribeiro, M. T., Singh, S., & Guestrin, C. (2016). "Why should I trust you?" explaining the predictions of any classifier. 1135-1144.

[29] Richardson, A., van Florenstein Mulder, T., & Vehbi, T. (2020). Nowcasting GDP using machine-learning algorithms: a real-time assessment. Int J Forecast.

[30] Shijun, C., Xiaoli, H., Yunbin, S., & Chong, Y. (2021). Application of Improved LSTM Algorithm in Macroeconomic Forecasting. Computational Intelligence and Neuroscience.

[31] Siqi, H. (2022). Back-Propagation Neural Network and ARIMA Algorithm for GDP Trend Analysis. Wireless Communications and Mobile Computing.

[32] Stock, J., & Watson, M. (2002). Macroeconomic forecasting using diffusion indexes. Journal of Business and Economic Statistics, 147-162.

[33] Sundararajan, M., Taly, A., & Yan, Q. (2017). Axiomatic attribution for deep networks. International Conference on Machine Learning, 3319-3328.

[34] Teräsvirta, T., Van Dijk, D., & Medeiros, M. C. (2005). Linear models, smooth transition autoregression, and neural networks for forecasting macroeconomic time series: a re-examination. Int J Forecast, 755-774.