

Print ISSN: 2288-4637 / Online ISSN 2288-4645
doi:10.13106/jafeb.2022.vol9.no4.0083

A Time Series-Based Statistical Approach for Trade Turnover Forecasting and Assessing: Evidence from China and Russia*

Xiao Wei DING¹

Received: December 15, 2021 Revised: February 27, 2022 Accepted: March 07, 2022

Abstract

Due to the uncertainty in the order of the integrated model, the SARIMA-LSTM model, SARIMA-SVR model, LSTM-SARIMA model, and SVR-SARIMA model are constructed respectively to determine the best-combined model for forecasting the China-Russia trade turnover. Meanwhile, the effect of the order of the combined models on the prediction results is analyzed. Using indicators such as MAPE and RMSE, we compare and evaluate the predictive effects of different models. The results show that the SARIMA-LSTM model combines the SARIMA model's short-term forecasting advantage with the LSTM model's long-term forecasting advantage, which has the highest forecast accuracy of all models and can accurately predict the trend of China-Russia trade turnover in the post-epidemic period. Furthermore, the SARIMA - LSTM model has a higher forecast accuracy than the LSTM-ARIMA model. Nevertheless, the SARIMA-SVR model's forecast accuracy is lower than the SVR-SARIMA model's. As a result, the combined models' order has no bearing on the predicting outcomes for the China-Russia trade turnover time series.

Keywords: Time Series, China-Russia Trade Turnover, Forecast Accuracy, Combined Model

JEL Classification Code: C32, F14, F17, P45

1. Introduction

One of the most important measures of the national economy is the total import and export volume, which represents the country's level of participation in international trade. For a long time, the EU has been Russia's most important trading partner and investment source. Despite this, due to the economic sanctions imposed on Russia by Europe and the United States, economic and trade cooperation between the EU and Russia is dwindling. The implementation of Russia's "Look East" plan gives a chance for China and Russia to enhance their economic and trade relations.

The overall trade turnover between Russia and China was \$72.34 billion from January 2013 to January 2021, with minerals accounting for 33% of total trade turnover and machinery, equipment, and instruments accounting for 33% of total trade turnover (29% of total trade turnover). Among Russia's main trade partners, China ranks first (14% of total trade turnover), and Germany ranks second (9% of total trade turnover) (RU-STAT, 2021). The Russia-China trade index was 195.22 in 2020, up 12.66 percent year on year, indicating that the structure of bilateral trade is still being optimised. Despite the new epidemic's impact, bilateral trade turnover has increased, demonstrating the huge potential for economic and trade cooperation between the two countries (Xinhua Silk Road Database, 2020). In 2021, the bilateral trade turnover hit a new high of over \$14.7 billion, up 41% year on year. China's exports to Russia totaled \$6.75 billion, up 34% year on year, while Russia's imports totaled \$7.93 billion, up 37.5 percent year on year. The ongoing deterioration of the pandemic had no effect on the overall development of bilateral trade, despite the fact that service tourism was heavily hurt. On the contrary, bilateral trade reached a record of \$1,644 billion in December 2021 (GACC, 2021).

In 2021 the Heads of China and Russia announced a joint statement formally deciding to extend the Treaty of

*Acknowledgements:

This research was supported by China Scholarship Council (CSC) Grant #202008090357

¹First Author and Corresponding Author, Department of Economics, Peoples' Friendship University of Russia. Russia [Postal Address: Miklukho-Maklaya 6, Moscow, 117198, Russia]
Email: xiaoweiding06796@gmail.com

© Copyright: The Author(s)

This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<https://creativecommons.org/licenses/by-nc/4.0/>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

Good-Neighborliness and Friendly Cooperation Between the People's Republic of China and the Russian Federation (FCT). China is Russia's first largest trade partner, and its exports to Russia are mainly medicine, electronic equipment, and machinery manufacturing products. Russia is China's tenth-largest trade partner, and its exports to China are primarily raw materials such as agricultural and sideline products and petrochemical products (Nikolay et al., 2021). According to the Cannikin Law in management, the scale of bilateral trade is constrained by the relatively small size of the economy. The economic volume gap between China and Russia is large. China's GDP reached \$14,723 trillion in 2020, while Russia's GDP reached \$1,483 trillion (World Bank Open Data, 2020).

One of the key approaches to boost Russia's GDP is to maintain a long-term stable bilateral trade partnership, expand trade turnover, and promote the renewal and upgrading of Russia's domestic industrial structure. The Russian-Chinese trade turnover time series, as the principal external manifestation of the import and export market, reveals a complex functioning mechanism. Meanwhile, the time series of trade turnover between China and Russia offers important information on the current and future development of bilateral commerce. Understanding the essence of its connotation, as well as the laws that govern its operation and development, is undeniably critical for forecasting, decision-making, and risk management activities in the Russian-Chinese trade market, as well as providing a theoretical and practical foundation for the establishment of the China-Russia Free Trade Area (FTA).

2. Literature Review

The methods used for forecasting International Trade markets are divided into linear and non-linear forecasting models. Linear forecasting models include Gravity Model (GM), Autoregressive (AR), Moving Average (MA), Vector autoregression (VAR), Autoregressive Integrated Moving Average (ARIMA), Generalized AutoRegressive Conditional Heteroskedasticity (GARCH), Holt-Winters (HW), and Exponential Smoothing Models (ESM). The most common method for assessing bilateral trade flows is the Gravity Model (GM). Wang et al. (2019) built the Stochastic frontier gravity model (SFGM) to empirically examine the trade efficiency of China and the Belt and Road countries using import and export trade data. It has been demonstrated that China's GDP growth can help to boost overseas commerce with its trading partners. The effect of trade distance on bilateral trade cooperation is inverse.

The VAR model is a variant of the AR model, which is commonly used to analyze multivariate time series. Solanki et al. (2020), for example, utilize the VAR model to

examine the two-way dynamic causation between economic development and the contributions of the industrial and agricultural sectors. The survey claims that the service sector contributes the most to the Indian economy, followed by industry and agriculture. Zainuri et al. (2021) used the VAR model to capture COVID-19 news, the daily price of the Composite Stock Market Index (IHSG), and interest rates over time. Positive news dominates the IHSG, while negative news reactions reduce the risk of a decline in investor confidence. The ARIMA model is undeniably a traditional time series analysis method, consisting of a linear combination of previous errors and past values of a smooth time series (Fanoodi et al., 2019). Because of the high accuracy of its short-term forecasts, the ARIMA model has been widely used in the economic field (Yue et al., 2019). Suh et al. (2014), for example, used error-adjusted ARIMA to forecast air freight rates. They found that the model is less able to predict small price changes, but it can respond quickly to price changes. However, applying ARIMA models had significantly contributed to forecasting research in financial markets and became a classic control model for various subsequent new forecasting methods (Kim & Won, 2018).

At present, non-linear forecasting models mainly include various models based on Artificial Neural Network (ANN), Support Vector Machine (SVM), Adaptive neuro-fuzzy inference system (ANFIS), etc. The long and short-term memory (LSTM) neural network is an improvement of the recurrent neural network (RNN), which alleviates the gradient disappearance and explosion phenomenon that tends to occur in the training process of traditional RNN (Gers et al., 2000). Using trade data from 10 countries as samples, Shen et al. (2021) proposed a multivariate LSTM-based method for extracting the time variation of trade data. The technique is used to produce accurate trade forecasts using trade data, and it outperforms regression models and standard time series models in terms of forecast accuracy. In recent years, deep neural networks have made an appearance in stock price trend prediction. Fische and Krauss (2018) used an LSTM neural network to forecast the volatility of the S&P 500 index; the results revealed that the LSTM model outperformed the Random Forest(RAF), DNN, and Logistic regression classifier in forecasting the index's volatility (LOG).

The adaptive neuro-fuzzy inference system (ANFIS) combines the benefits of neural networks and fuzzy logic in an organic way. As a result, it possesses both the self-learning ability to quantify data knowledge and the ability to precisely depict the human brain's reasoning capabilities (Jang, 1993). ANFIS is commonly utilized in time series of international trade turnover, and some experts have studied this topic in depth. For example, Tian and

Ju (2019) used the ANFIS to create a forecast model of China-Russia trade turnover. The result showed that when China-Russia relations, the volume of Russian foreign trade, and the volume of Chinese foreign trade are used as input variables, the model has a high forecast accuracy and predicts 8.6% year-on-year growth in the volume of China-Russia trade turnover in 2019. Additionally, a study by Aufar and Sitanggang (2019) pointed out that using the ANFIS for forecasting farmers' terms of trade (NTP) was not suitable for multivariate monthly data but was ideal for monthly time series data.

SVM has a strong generalization performance for pattern classification and regression prediction. Furthermore, it overcomes the drawbacks of traditional forecasting methods that require large samples, extensive forecasting results, over-learning, and under-learning. According to the Firefly algorithm, Kuo and Li (2016) innovatively developed a prediction system for export trade turnover based on the K -means algorithm and the SVR algorithm with wavelet transform. The result shows that the prediction algorithm using wavelet transform and clustering is excellent. In addition, the SVR based on the Firefly algorithm is superior to the other algorithms.

Although ARIMA, LSTM, and SVR models have their advantages in time series forecasting, a single model cannot overcome its shortcomings, so a sequence of combined models for forecasting time series inevitably arises in the import and export market. When using ARIMA and SVM models to analyze information on China's total foreign trade from 1980 to 2016, Zhang and Sun (2019) found that it is not the single model that predicts better; the combined model will be more effective. Therefore, they propose that the optimal combination model could be determined by the relative error of the training samples.

Hence, this study constructed four combination models using a combination of linear and non-linear methods to evaluate the Forecast Accuracy of Single Models and Combined Models to find the best-combined model to predict the time series of China-Russia trade turnover. The study is based on the following hypotheses:

H1: *In single models, the LSTM model has higher forecast accuracy and stability than the SARIMA and SVR models.*

H2: *Although both the SARIMA-LSTM model and the SARIMA-SVR model combine the advantages of linear forecasting with the benefits of non-linear forecasting, the single LSTM model has higher forecast accuracy than the SVR model. Therefore, it is inferred that the SARIMA-LSTM model has higher forecast accuracy than the SARIMA-SVR model.*

H3: *The order of the combined models does not affect the prediction results.*

3. Data and Model Specification

3.1. Data

Geo-relation is an essential factor affecting import and export trade. China and Russia have always maintained good neighborly relations of friendship and cooperation, with steady bilateral economic and trade cooperation development. This paper presents a descriptive analysis using monthly historical data on China-Russia trade volume from 2013–2021 as a sample for the study.

3.2. Forecast Performance Measures

The monthly historical data of China-Russia trade turnover from 2013–2019 were used as the train set. The monthly historical data of China-Russia trade turnover from 2020–2021 were used as the test set. The predicted values of the test set were compared with the actual values for validation. The ratio of the train set to test set is 7:2. The single-step prediction method was used in this paper to reduce the error and improve forecast accuracy. Mean Square Error (MSE), Mean Absolute Error (MAE), Root Mean Square Error (RMSE), and Mean Absolute Percentage Error (MAPE) are used as indicators for the evaluation of the model. The mathematical formula is:

$$\text{MSE} = \frac{1}{n} \sum_{i=1}^n (\hat{y}_i - y_i)^2 \quad (1)$$

$$\text{MAE} = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \quad (2)$$

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n (\hat{y}_i - y_i)^2} \quad (3)$$

$$\text{MAPE} = \frac{1}{n} \sum_{i=1}^n \left| \frac{y_i - \hat{y}_i}{y_i} \right| \quad (4)$$

where y_i is the actual value, \hat{y}_i is the predicted value, and n is the sample size.

3.3. Building Combination Forecasting Models

3.3.1. Introduction to SARIMA, LSTM, and SVR Model

The SARIMA model is based on the ARIMA model for seasonal or cyclical data. The model expression is:

$$\text{SARIMA} = (p, d, q) \times (P, D, Q)_s \quad (5)$$

where p is the autoregressive orders, P is the seasonal autoregressive orders, d is the differential counts, D is the seasonal differential counts, q is the moving average orders, Q is the seasonal moving-average orders, and s is the number of periods.

The minimum information criterion is a criterion for assessing the complexity of statistical models and measuring the “goodness of fit” of statistical models. In this paper, the model parameters are determined by the Akaike Information Criterion (AIC) (Akaike, 1974). The proposed AIC can effectively compensate for the subjectivity of determining the order based on the Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF) plots. It can find the best-fit model quickly within a limited range of orders.

The input and output gates in the LSTM neural network are used to receive, output, and correct parameters, and the forget gate is used to control the degree of forgetting the cell state at moment $t-1$. The cell records the state of the neuron containing the neural network layers sigmoid and the multiplication operations pointwise. The output value of the function sigmoid is in the range 0 to 1, and the output value of the function tanh is in the range -1 to 1 (Duan et al., 2021). The training process is reflected in the cell state C_t and output information h_t through three different gates to selectively retain and pass information, using a backpropagation algorithm for model training, ultimately controlling the information. C_{t-1} is the cell state at the moment $t-1$, which can retain long-term or short-term memory after different degrees of training.

Support vector machine regression (SVR) is an extended application of SVM, and the SVR algorithm uses ε , and kernel function. The slack variables define the error range of the model and the SVR model based on the insensitive loss function ε is:

$$\min \frac{1}{2} \|\omega\|^2 + C \sum_{i=1}^m (\xi_i + \xi_i') \tag{6}$$

$$\text{s.t.} \begin{cases} -\varepsilon - \xi_i \leq y_i - \omega\phi(x) - b \leq \varepsilon + \xi_i' \\ \xi_i \geq 0, \xi_i' \geq 0 (i = 1, 2, \dots, n) \end{cases} \quad (\text{Constraints})$$

where ω is the normal vector, ξ_i and ξ_i' are the slack variables, and C is the regularization parameters.

According to the method of Lagrange Multipliers and the Karush–Kuhn–Tucker (KKT) conditions. The mathematical formulation of the SVR model is:

$$f(x) = \sum_{i=1}^n (\alpha_i - \alpha_i') K(x_1, x_2) + b \tag{7}$$

where α_i and α_i' are the Lagrange Multipliers, and $K(x_1, x_2)$ is the Kernel function.

3.3.2. Construction of the SARIMA-LSTM and SARIMA-SVR Models

The time series of China-Russia trade turnover Y_t is assumed to have a linear part L_t and a non-linear part N_t , and the mathematical formula is:

$$Y_t = L_t + N_t \tag{8}$$

Using the SARIMA model to fit and predict the time series of China-Russia trade turnover Y_t , the residual series of SARIMA model was acquired:

$$e_t = Y_t - \widehat{L}_t \tag{9}$$

where \widehat{L}_t is the linear predicted values of the SARIMA model, and e_t is the residual series for the SARIMA model.

Using the LSTM model to fit and predict the residual series of the SARIMA model e_t , and the linear predicted values of the SARIMA model are \widehat{L}_t accumulated with the non-linear predicted values of the residual series of the LSTM model \widehat{e}_{t1} . The predicted values of the SARIMA-LSTM model \widehat{Y}_{t11} were acquired (Ding et al., 2020):

$$\widehat{Y}_{t11} = \widehat{L}_t + \widehat{e}_{t1} \tag{10}$$

where \widehat{e}_{t1} is the non-linear predicted values of the residual series of the LSTM model, and \widehat{Y}_{t11} is the predicted values of the SARIMA-LSTM model.

Using the SVR model to fit and predict the residual series of the SARIMA model e_t , and the linear predicted values of the SARIMA model \widehat{L}_t are accumulated with the non-linear predicted values of the residual series of the SVR model \widehat{e}_{t21} . The predicted values of the SARIMA-SVR model \widehat{Y}_{t21} were acquired (Sun et al., 2014):

$$\widehat{Y}_{t21} = \widehat{L}_t + \widehat{e}_{t21} \tag{11}$$

where \widehat{e}_{t21} is the non-linear predicted values of the residual series of the SVR model, and \widehat{Y}_{t11} is the predicted values of the SARIMA-SVR model.

3.3.3. Construction of the LSTM-SARIMA and SVR-SARIMA Models

Using the LSTM model to fit and predict the time series of China-Russia trade turnover Y_t , the residual series of the LSTM model e_{t12} were acquired:

$$e_{t12} = Y_t - \widehat{N}_{t1} \tag{12}$$

where \widehat{N}_{t1} is the non-linear predicted values of the LSTM model, and e_{t12} is the residual series for the LSTM model.

Using the SARIMA model to fit and predict the residual series of the LSTM model e_{t12} , the non-linear predicted

values of the LSTM model \widehat{N}_{t1} are accumulated with the linear predicted values of the residual series of the SARIMA model \widehat{e}_{t12} . The predicted values of the LSTM-SARIMA model \widehat{Y}_{t12} were obtained:

$$\widehat{Y}_{t12} = \widehat{N}_{t1} + \widehat{e}_{t12} \tag{13}$$

where \widehat{Y}_{t12} is the predicted values of the LSTM-SARIMA model, and \widehat{e}_{t12} is the linear predicted values of the residual series of the SARIMA model.

Using the SVR model to fit and predict the time series of China-Russia trade turnover Y_t , the residual series of the SVR model e_{t22} were obtained:

$$e_{t22} = Y_t - \widehat{N}_{t2} \tag{14}$$

where \widehat{N}_{t2} is the non-linear predicted values of the SVR model, and e_{t22} is the residual series for the SVR model.

Using the SARIMA model to fit and predict the residual series of the SVR model e_{t22} , the non-linear predicted values of the SVR model \widehat{N}_{t2} are accumulated with the linear predicted values of the residual series of the SARIMA model \widehat{e}_{t22} . The predicted values of the SVR-SARIMA model \widehat{Y}_{t22} were obtained:

$$\widehat{Y}_{t22} = \widehat{N}_{t2} + \widehat{e}_{t22} \tag{15}$$

where \widehat{Y}_{t22} is the predicted values of the SVR-SARIMA model, and \widehat{e}_{t22} is the linear predicted values of the residual series of the SARIMA model.

3.4. Analyze Statistics

This paper was coded in python, and the IDE was pycharm2021. The SARIMA-LSTM and LSTM-SARIMA

models were constructed using the “arima_model” and the “Keras” packages. The SARIMA-SVR and SVR-SARIMA models were constructed using the “arima_model” and the “sklearn.svm” packages.

4. Results and Discussion

4.1. Results of the SARIMA Model

The Seasonal and Trend decomposition using Loess (STL) method was used to decompose the time series of China-Russia trade turnover Y_t . As can be seen from Figure 1 (Trend) and (Resid), there is a clear trend in the China-Russia trade volume time series Y_t . The time series shows an upward trend after 2016, and therefore it may be non-stationary and require different operations. As shown in Figure 2 (Seasonal), there is seasonality in the time series, with a cycle of 12 months. According to the pattern of cycle volatility, we find that the trade turnover in a cycle is the largest in December, the smallest in March, and the most significant fluctuations from June to July and from November to December. The time series of China-Russia trade turnover Y_t has both seasonal and non-seasonal characteristics, so we use the SARIMA multiplicative mixed model to fit and forecast.

After the second-order differences were applied to the time series of China-Russia trade turnover Y_t , the results of the Augmented Dickey-Fuller test (ADF) showed that the p -value is 3.176045087186587e-11, the original hypothesis was rejected, and the series was considered to be smooth after the two differences. When lags are 24, the result of the Ljung-Box test (LB) showed that all p -values are less than 0.05. So, after the second-order difference, the time series of China-Russia trade turnover is smooth and has no white noise.

The optimal parameters were determined using grid search and cross-validation (GridSearchCV) in machine

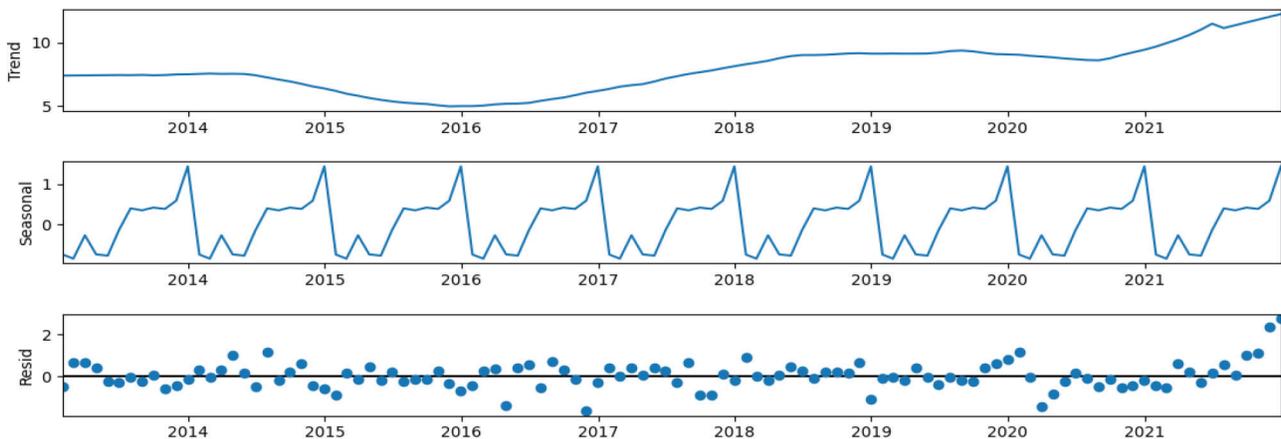


Figure 1: The Illustration of the STL Decomposition

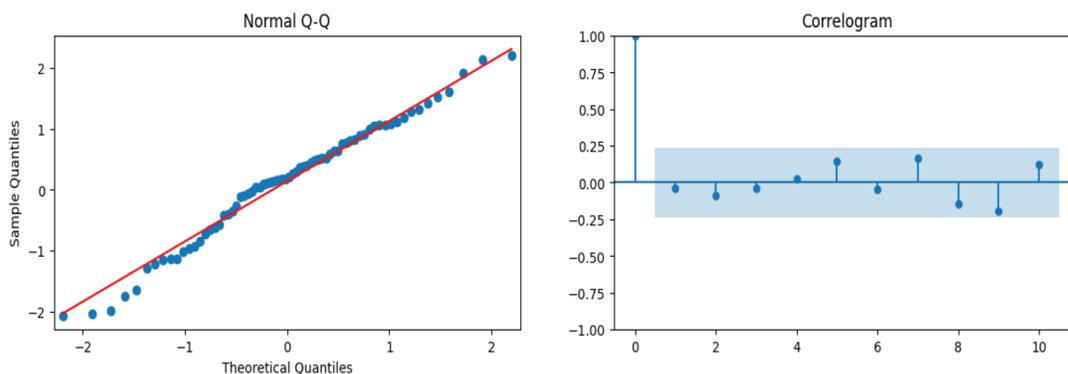


Figure 2: Residual Information of SARIMA Model

Table 1: Significance Tests for SARIMA Model Coefficients

Parameters	coef	std err	z	$P > z $	[0.025	0.975]
ar.L1	-0.6635	0.114	-5828	0.000***	-0.887	-0.440
ar.L2	-0.3756	0.125	-3004	0.003***	-0.621	-0.131
ma.L1	-0.9862	0.386	-2552	0.011***	-1744	-0.229
ar.S.L12	-0.6757	0.132	-5128	0.000***	-0.934	-0.417
ar.S.L24	-0.6387	0.144	-4438	0.000***	-0.921	-0.357
ar.S.L36	-0.5860	0.137	-4273	0.000***	-0.855	-0.317
sigma2	0.3717	0.165	2.259	0.024***	0.049	0.694

Note: *, $p < 0.1$; **, $p < 0.05$; and *** $p < 0.01$. Significant at the 0.05 level.

learning. The process of parameter search is divided into two stages: firstly, the optimal parameters of the ARIMA model are determined, and secondly, the optimal seasonal parameters are searched by using the optimal parameters of the ARIMA model. Based on the ACF and PACF plots of the smooth series, set the range of search for the parameters: $p \in [0, 5]$, $d = 2$, $q \in [0, 5]$, $P \in [0, 5]$, $D \in [0, 2]$, $Q \in [0, 5]$. During the search process, the model must be smooth and reversible, and the combination of parameters is automatically skipped if any of the conditions are not satisfied. Figure 1 (Seasonal), which shows that the number of periods is 12, the optimal combination of parameters for the model is SARIMA = (2, 2, 1) × (3, 1, 0)₁₂. In addition, the model was obviously optimized with the addition of seasonal parameters to the ARIMA model, and the AIC values were significantly reduced.

The SARIMA model is trained to forecast the time series of China-Russia trade turnover Y_t and the linear predicted values of the SARIMA model \hat{L}_t and the residual series of the SARIMA model e_t , are acquired. The significance test of the model is to verify that the fitted residual series of China-Russia trade volume is white noise and obeys a Gaussian distribution (Babu & Reddy, 2014).

Figure 2 (Normal Q-Q) shows that the values of the residuals are approximately on a straight line, which

indicates that the residuals are normally distributed; Figure 2 (Correlogram) shows that 95% of the autocorrelation falls within the confidence interval. Therefore, the residual series of the SARIMA model is white noise. As shown in Table 1, all coefficients of the model are statistically significant.

4.2. Results of SARIMA-LSTM and LSTM Models

The residual series of the SARIMA model and the time series of China-Russia trade turnover were respectively used as input variables for the LSTM model, normalized in the range 0 to 1 using the MinMaxScaler from the scikit-learn library. The dataset is transformed into a supervised learning dataset, and then the value at the moment t is predicted by the value at moment $t-1$.

This paper uses the keras deep learning framework to quickly build an LSTM neural network model, build a Sequential model and add LSTM layers. In the training process of deep learning networks, neurons are temporarily discarded from the network in a particular proportion, weakening the joint adaptation between neuron nodes, so a Dropout layer is added after the LSTM layer (Krizhevsky et al., 2017). Dropout is 0.5, randomly generating the most

network structure, which effectively enhances the model's generalization ability and prevents over-fitting (Song et al., 2021). The output dimension of the Dense layer is 1. When the activation function is set to Sigmoid, it is easy to produce gradient explosion and gradient disappearance, and the practical effect is not as good as Relu. Therefore, Relu is used for the activation function.

The Adam algorithm combines the advantages of both the AdaGrad and RMSProp optimization algorithms which adaptively adjust the learning rate (Dokkyun et al., 2020). So, in this paper, the Adam optimizer is used for optimization training. The loss function is set to MAE. When the parameters of the SARIMA-LSTM model were set as: the number of units in the LSTM layers is 80, the number of units in the dense layers was 1, and batch size is 100, the loss rate of the model was reduced to 0.1104 after 50 iterations. In the same way, the parameters of the LSTM model are set as: the number units in LSTM layers are 65, the number of units in Dense layers are 1, and batch size is 40, the loss rate of the model is reduced to 0.0615 after 300 iterations. The trained LSTM model is applied to the test set, and the predicted values are renormalized. Finally, the non-linear predicted values of the LSTM model \widehat{N}_{t1} and the predicted values of the SARIMA-LSTM model \widehat{Y}_{t11} are acquired.

4.3. Results of the SARIMA-SVR and SVR Models

The kernel function is the most important parameter in the SVR model, so the choice of kernel function is the key to solving non-linear problems. The Gaussian radial basis kernel function (RBF) has a strong learning ability, high recognition rate, and good performance. In this paper, the RBF is chosen for regression prediction. Since different parameter values have an important impact on the fit and prediction results, the SVR model requires identifying three parameters: the insensitive loss function ε , the penalty parameter C , and the parameter σ in the kernel function.

The input variables of the SVR model are normalized and transformed to the supervised learning dataset in the same way as the input variables of the LSTM model are pre-processed in the previous section. The hyperparameter optimization algorithms include Grid Search, Genetic Algorithm, Particle Swarm Optimization (PSO), etc. This paper used the GridSearchCV function in Scikit-learn for optimal hyperparameter determination. PSO is also the best method for determining the optimal Hyperparameters (Viet & Nhat, 2020). The search results of the optimal hyperparameters for the SVR model: C is 5, σ is 0.1, ε is 0.2; the search results of the optimal hyperparameters for the SARIMA-SVR model: C is 4, σ is 0.25, ε is 0.0625. So, the non-linear predicted values of the SVR model \widehat{N}_{t2} and the predicted values of the SARIMA-SVR model \widehat{Y}_{t21} are acquired.

4.4. Results of the LSTM-SARIMA and SVR-SARIMA Models

The residual series of the LSTM model e_{t12} and the residual series of the SVR model e_{t22} are used as input variables of the SARIMA model, respectively, and the STL method is used to decompose the residual series to show that the residual series e_{t12} and the residual series e_{t22} are seasonal and non-smooth. By ADF test, the p -value is 1.78E-11, and when lags are 24, the LB test shows that p -value less than 0.05. So the residual series e_{t12} is a smooth and non-white noise after the first-order difference. Also, by ADF test and LB test, the residual series e_{t22} is a smooth and non-white noise after the second order difference.

Based on the ACF and PACF plots after series difference, the parameter search range of the LSTM-SARIMA model is set: $p \in [0, 5]$, $d = 1$, $q \in [0, 5]$, $P \in [0, 5]$, $D \in [0, 2]$, $Q \in [0, 5]$. The parameter search range of the SVR-SARIMA model is set: $p \in [0, 2]$, $d = 2$, $q \in [0, 5]$, $P \in [0, 5]$, $D \in [0, 2]$, $Q \in [0, 5]$. After two-stage grid search, the optimal hyperparameters of the LSTM-SARIMA are shown as: SARIMA = (4, 1, 3) \times (3, 1, 0)₁₂; the optimal hyperparameters of the SVR-SARIMA are shown as SARIMA = (0, 2, 5) \times (0, 1, 1)₁₂. In addition, the model was optimized with the addition of seasonal parameters to the ARIMA model, and the AIC values were significantly reduced. There is no autocorrelation in the residual series by Durbin-Watson (DW) test. Finally, the predicted values of the LSTM-SARIMA model \widehat{Y}_{t12} and the predicted values of the SVR-SARIMA model \widehat{Y}_{t22} were acquired.

The experimental results show that the time series of China-Russia trade turnover Y_t is complex, with obvious seasonal volatility. Therefore, when the SARIMA model is used to fit and forecast in this paper, the time series must be transformed into smooth and non-white noise by difference. To show the comparative effectiveness, in this paper, we innovatively use a two-stage grid search to determine the optimal hyperparameters: first, the ARIMA optimal parameters are determined, and then the seasonal parameters. Using the ARIMA model as a control model, we can observe that it is necessary to add seasonal parameters based on the change of AIC value. The generation of combined models effectively compensates for the limitations of single models, but it does not mean that the forecast accuracy of a single model is certainly worse than that of a combined model.

Therefore, from Table 2, we can make the following results: (1) Comparative analysis of the forecast accuracy of all models. The SARIMA-LSTM model has the highest forecast accuracy: RMSE = 0.861833, MAPE = 0.067733. When predicting the time series of trade turnover, the SARIMA-LSTM model can be used in preference.

Comparative analysis of forecast accuracy between single models. The SARIMA model has the lowest forecast

accuracy: RMSE = 2.629184, MAPE = 0.2274, and the long-term forecast effect are not good. Although the LSTM model and the SVR model have their own advantages in the long-term forecast, the LSTM model has the highest forecast accuracy. SVR can be used for time series analysis but is not the best choice. The best choice is to use LSTM neural network to handle the time series data. This conclusion is consistent with H1.

(3) Comparative analysis of the forecast accuracy between the single models and the combined models. The SARIMA model was compared with the SARIMA-LSTM, LSTM-SARIMA, SARIMA-SVR, and SVR-SARIMA models for forecast accuracy and the SARIMA model came out with the worst results. Therefore, when using the SARIMA model to forecast the time series of trade turnover, it is better to use combined models which contains an LSTM or an SVR model; The LSTM model is compared with the SARIMA-LSTM and the LSTM-SARIMA models for forecast accuracy, and the LSTM model was the worst: RMSE = 1.04661, MAPE = 0.083345. Therefore, when using LSTM models to forecast the time series of trade turnover, the combined model which contains LSTM models is more effective; The SVR model was compared with the SARIMA-SVR

and the SVR-SARIMA models for forecast accuracy, and it came out best: RMSE = 1.759563, MAPE = 0.134896. The result proves that the combined model is not certainly better than the single model, and the combined model which contains the SVR model demonstrates the "1 + 1 < 1" negative prediction effect. Therefore, when using the SVR model to forecast the time series of trade turnover, it is more accurate to use a single model.

(4) Comparative analysis of forecast accuracy between combined models. The SARIMA-LSTM model has the highest predictive accuracy of all the combined models: RMSE = 0.861833, MAPE = 0.067733. The SARIMA-LSTM model has higher forecast accuracy than the SARIMA-SVR model, and this conclusion is consistent with H2. Therefore, when using a combined model to forecast the time series of trade turnover, the SARIMA-LSTM model is the most effective; The SARIMA-LSTM model has a higher forecast accuracy than the LSTM-SARIMA model, but the SARIMA-SVR model has a lower forecast accuracy than the SVR-SARIMA model. In conclusion, the order of the combined models has no impact on the forecast effect. This conclusion is consistent with H3.

(5) The coefficients of the LSTM-SARIMA and SVR-SARIMA models were not all statistically significant when the SARIMA model was tested for significance of coefficients. Although the order of combined models has no effect on forecast accuracy, the model coefficients are non-significant. As a result, LSTM-SARIMA and SVR-SARIMA models are rarely employed to forecast financial time series.

Table 2: The Prediction Performance of Each Model

Model	MAE	MSE	RMSE	MAPE
SARIMA	7.85855	2.629184	2.80331	0.2774
LSTM	1.095391	0.823569	1.04661	0.083345
SVR	3.096062	1.23827	1.759563	0.134896
SARIMA-LSTM	0.742756	0.66663	0.861833	0.067733
LSTM-SARIMA	0.884691	0.740579	0.94058	0.080129
SARIMA-SVR	4.354601	1.908101	2.086768	0.202942
SVR-SARIMA	2.546738	1.24051	1.59585	0.134972

5. Conclusion and Implications

The time series of China-Russia trade turnover is seasonal, although it was in a downward trend from January 2020 to March 2020 due to the impact of the COVID-19 epidemic. However, through the cooperation of the Chinese and Russian governments and enterprises, the development

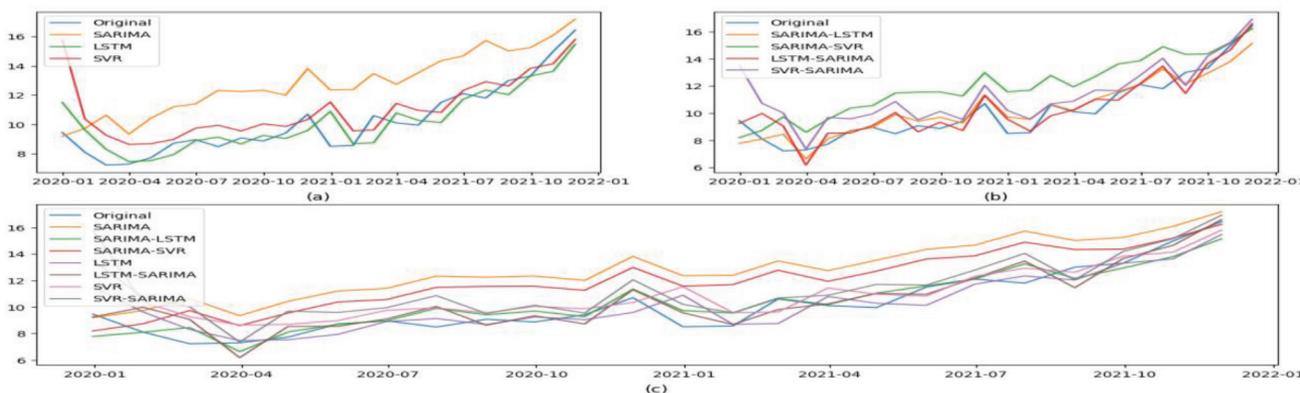


Figure 3: China-Russia Trade Turnover Forecast

of China-Russia trade quickly returned to normal. From Figure 3, we can analyze the prediction of the trend for each model.

- (1) The ARIMA-LSTM and SARIMA-SVR models are influenced by the SARIMA model, and accurately capture the pattern of fluctuation in the time series of China-Russia trade turnover in the following periods: June 2020 to July 2020, March 2021 to April 2021, and June 2021 to July 2021. In addition, the predicted values of the SARIMA-LSTM model during these periods were exactly consistent with the true values.
- (2) The LSTM-SARIMA model is affected by the hysteresis of the LSTM model, and the changing trends are almost the same as the LSTM model; the SVR-SARIMA model is affected by the hysteresis of the SVR model, and the changing trends are almost the same as the SVR model. Using the Empirical Mode Decomposition (EMD) decomposition algorithm can effectively resolve hysteresis phenomena while also improving the forecast accuracy of financial time series (Chen et al., 2019). The main purpose of this paper is to compare the forecast accuracy between the models. Therefore, the time series of China-Russia trade turnover was not transformed into a stationary series when used as input variables for the LSTM and SVR models. Although stacking LSTM hidden layers can increase the neural network's depth, it also efficiently reduces the number of neurons, improves training efficiency, and improves forecast accuracy. The LSTM layer, Dropout layer, and Dense layer are the sole layers in the LSTM model built in this paper. However, for the time series of China-Russian trade turnover, the model is highly effective.
- (3) All models accurately capture the fluctuation pattern of the time series of China-Russia trade turnover from November 2020 to January 2021 and from November to December 2021. In addition, trade turnover for December is predicted as the highest during the cycle, which is consistent with the true result.

References

- Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, 19(6), 716–723. <https://doi.org/10.1109/TAC.1974.1100705>
- Aufar, Y., & Sitanggang, I. S. (2019). Adaptive neuro-fuzzy inference system implementation for farmer's term of trade forecasting in West Sumatra. *IOP Conference Series: Earth and Environmental Science*, 335(1), 012010. <https://doi.org/10.1088/1755-1315/335/1/012010>
- Babu, C. N., & Reddy, B. E. (2014). A moving-average filter-based hybrid Arima-ANN model for forecasting time series data. *Applied Soft Computing*, 23(10), 27–38. <https://doi.org/10.1016/j.asoc.2014.05.028>
- Chen, L., Chi, Y. G., Guan, Y. Y., & Fan, J. I. (2019). A hybrid attention-based EMD-LSTM model for financial time series prediction. In 2nd International Conference on Artificial Intelligence and Big Data, Chengdu, China, May 25–28, 2019 (pp. 113–118). Manhattan, NY: IEEE Publications. <https://doi.org/10.1109/ICAIBD.2019.8837038>
- Ding, R., Li, W., & Wang, R. Z. (2020). Combined forecasting model based on SARIMA and LSTM. *Computer & Digital Engineering*, 48(2), 304–308. <https://doi.org/10.3969/j.issn.1672-9722.2020.02.007>
- Dokkyun, Y., Jaehyun, A., & Sangmin, J. (2020). An effective optimization method for machine learning based on ADAM. *Applied Sciences*, 10(3), 1073. <https://doi.org/10.3390/app10031073>
- Duan, L. Y., Liu, Z. Y., Yu, W., Chen, W., Jin, D., Li, D., Sun, S., & Dai, R. (2021). Modeling analysis and comparison of neural network simulation based on ECM and LSTM. *Journal of Physics: Conference Series*, 2068(1), 012041. <https://doi.org/10.1088/1742-6596/2068/1/012041>
- Fanoodi, B., Malmir, B., & Jahantigh, F. F. (2019). Reducing demand uncertainty in the platelet supply chain through artificial neural networks and Arima models. *Computers in Biology and Medicine*, 113, 103415. <https://doi.org/10.1016/j.combiomed.2019.103415>
- Fischer, T., & Krauss, C. (2018). Deep learning with long short-term memory networks for financial market predictions. *European Journal of Operational Research*, 270(2), 654–669. <https://doi.org/10.1016/j.ejor.2017.11.054>
- GACC. (2021). *About GACC*. <http://tjs.customs.gov.cn/tjs/sjgb/tjyb/4127455/index.html>
- Gers, F. A., Schmidhuber, J., & Cummins, F. (2000). Learning to forget: Continual prediction with LSTM. *Neural Computation*, 12(10), 2451–2471. <https://doi.org/10.1162/089976600300015015>
- Jang, J. S. R. (1993). ANFIS: Adaptive-network-based fuzzy inference system. *IEEE Transactions on Systems, Man, and Cybernetics*, 23(3), 665–685. <https://doi.org/10.1109/21.256541>
- Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2017). Image net classification with deep convolutional neural networks. *Communications of the ACM*, 60(6), 84–90. <https://doi.org/10.1145/3065386>
- Kuo, R. J., & Li, P. S. (2016). Taiwanese export trade forecasting using firefly algorithm based K-means algorithm and SVR with wavelet transform. *Computers and Industrial Engineering*, 99(C), 153–161. <https://doi.org/10.1016/j.cie.2016.07.012>
- Kim, H. Y., & Won, C. H. (2018). Forecasting the volatility of stock price index: A hybrid model integrating LSTM with multiple Garch-type models. *Expert Systems with*

- Applications*, 103(1), 25–37. <https://doi.org/10.1016/j.eswa.2018.03.002>
- Nikolay, Y., Ekaterina, K., Tatiana, M., & Lazar, B. (2020). Features of foreign trade between Russia and China and prospects for its development. In *E3S Web of Conferences*, 220(2), (SES-2020). <https://doi.org/10.1051/e3sconf/202022001020>
- RU-STAT. (2021). *Export and import of Russia by goods and countries*. <https://ru-stat.com/date-M201301-202101/RU/trade/CN>
- Solanki, S., Inumula, K. M., & Chitnis, A. (2020). Sectoral contribution to economic development in India: A time-series co-integration analysis. *Journal of Asian Finance, Economics, and Business*, 7(9), 191–200. <https://doi.org/10.13106/jafeb.2020.vol7.no9.191>
- Suh, S. S., Park, J. W., Song, G., & Cho, S. G. (2014). A study of air freight forecasting using the Arima model. *Journal of Distribution Science*, 12(2), 59–71. <https://doi.org/10.13106/jds.2014.vol12.no2.59>
- Sun, Y. X., Shao, C. F., & Ji, X. (2014). Urban traffic accident time series prediction model based on the combination of Arima and information granulation SVR. *Journal of Tsinghua University (Science and Technology)*, 54(3), 348–353. <https://doi.org/10.16511/j.cnki.qhdxxb.2014.03.004>
- Song, C. H., Li, Z. X., & Yu, H. X. (2021). An oil well fault identification method base on improved GoogleNet. *Journal of Jiangsu University of Science and Technology (Natural Science Edition)*, 35(2), 52–58. <https://doi.org/10.11917/j.issn.1673-4807.2021.02.008>
- Shen, M., Lee, C., Liu, H., Chang, P., & Yang, C. (2021). Effective multinational trade forecasting using LSTM recurrent neural network. *Expert Systems with Applications*, 182(6). <https://doi.org/10.1016/J.ESWA.2021.115199>
- Tian, C. Y., & Ju, G. H. (2019). China–Russia trade forecasting based on adaptive neuro-fuzzy inference system. *Modern Business*, 36, 70–71. <https://doi.org/10.14097/j.cnki.5392/2019.36.027>
- Viet, H. N., Nhat, D. H., Van, B. D., Hong, D. V., & Dieu, T. B. (2020). A hybrid computational intelligence approach for predicting soil shear strength for urban housing construction: A case study at Vinhomes Imperia project, Hai Phong city (Vietnam). *Engineering with Computers*, 36(2), 603–616. <https://doi.org/10.1007/s00366-019-00718-z>
- Xinhua Silk Road Database. (2020). *Welcome to the Belt and Road information service platform*. <https://en.imsilkroad.com/db/#/home>
- Wang, P., Li, G., & Pang, S. (2019). An analysis of the potential of import and export trade between China and the countries along with the belt and road advances in economics. In *2019 4th International Conference on Financial Innovation and Economic Development*, Sanya, China, January 18–20, 2019 (pp. 342–345). Paris: Atlantis Press. <https://doi.org/10.2991/icfied-19.2019.64>
- World Bank Open Data. (2020). *World Bank Data*. <https://data.worldbank.org.cn/indicator/NY.GDP.MKTP.CD?locations=CN-RU>
- Yue, L. X., Zhou, X. Y., & Chen, Y. L. (2019). Thematic trend prediction of information architecture based on the Arima model. *Library and Information Knowledge*, 5, 54–63. <https://doi.org/10.13366/j.dik.2019.05.054>
- Zainuri, Z., Viphindartin, S., & Wilantari, R. N. (2021). The impacts of the COVID-19 pandemic on the movement of composite stock price index in Indonesia. *Journal of Asian Finance, Economics, and Business*, 8(3), 1113–1119. <https://doi.org/10.13106/jafeb.2021.vol8.no3.1113>
- Zhang, L., & Sun, D. S. (2019). Forecast of total import and export trade based on combination model. *Jiangsu Commercial Forum*, 2, 57–59. <https://doi.org/10.13395/j.cnki.issn.1009-0061.2019.02.011>