Price Prediction of TSLA, BYD and NIO Based on ARIMA Model

Hanzhuo Liu

Department of Statistics, Shandong University, Beijing, China

201900820195@mail.sdu.edu.cn

Abstract. New energy vehicles are the development direction of the future automobile industry, and various countries are actively deploying new energy vehicle enterprises. Predicting the stock prices of the top three new energy vehicle companies in the market has the high reference value for stock investors. On this basis, this paper uses the ARIMA model to forecast stock prices for the next four trading days of the top three companies, i.e., TSLA, BYD and NIO. The stock prices of the three companies in the first three quarters of 2022 will be used to build suitable models. According to the analysis, the ARIMA model succeeded in predicting stock prices with high accuracy in the short term, but the accuracy decreases over time. In this case, the stock price prediction results of the ARIMA model can provide a reference for stock investors in the short term, but cannot predict the future trend of stocks price. Overall, these results shed light on guiding further exploration of stock price forecasting based on ARIMA model.

Keywords: Time Series Analysis, Stock Price Prediction, New Energy Vehicles.

1. Introduction

As the development trend of future vehicles, new energy vehicles have broad prospects and development opportunities. Germany released its National Electric Vehicle Plan in 2009, the United States published its Electric Vehicle Popularization Plan Blueprint in 2013 [1], and China announced its new energy vehicle industry development plan for the next 15 years in 2020 [2]. In such an environment, the new energy vehicle industry will have more investment injection in the future. Large new energy vehicle manufacturers (e.g., Tesla and BYD) have extended in the international market. It is a great significance to predict the stock prices of new energy vehicle companies [3].

Kumar combined three different hybrids with linear ARIMA and non-linear models and find ARIMA-SVM model is the best forecasting mode [4]. Mondal conducted a study on the effectiveness of ARIMA models [5]. Ali made a comparison on the ARIMA and Artificial neural networks models for stock price prediction [6]. Mustafa forecasted stock prices through univariate ARIMA model [7]. Chen took the stock price of Kweichow Moutai as an example to propose the combination of ARIMA model and support vector machine regression model (SVR) to improve the accuracy of prediction [8]. Xiong's recursively corrected ARIMA-GARCH-M model through different test methods which has good results in short-term stock forecasting and has certain feasibility [9]. Huang took the stock price information of China Merchants Bank as an example using the ARIMA model to make judgments on the future trend of stock prices, which has reference value to obtain income in the stock market [10].

Based on the ARIMA model, this paper makes short-term predictions of the stock prices of the top three new energy vehicle manufacturers, Tesla, BYD and NIO for investors' reference. This article will introduce the sources of stock price data in the data and method section and introduce the ARIMA model. Specifically, this article describes the preprocessing of the data and prediction results, and draws corresponding conclusions. The limitations of this article will also be explained.

2. Data & Method

2.1. Data

The stock price contains four common variables, i.e., opening price, closing price, high price, and low price. The high and low prices can only reflect the supply and demand at a certain time, which are not much helpful for stock forecasting. For the opening and closing prices, investors generally

observe the stock price fluctuation before trading, so the closing price will be chosen as the forecast data. The data is downloaded from https://cn.investing.com/. The daily closing price from January to October 2022 is selected as the initial data, with a total of 209 data. The previous 205 data are used as forecast data to predict the stock price of next four days by ARIMA model. The data of the last four days is used as the test data to test the prediction results. The stock price of the three companies from January to October as shown in Fig. 1 and Fig. 2.

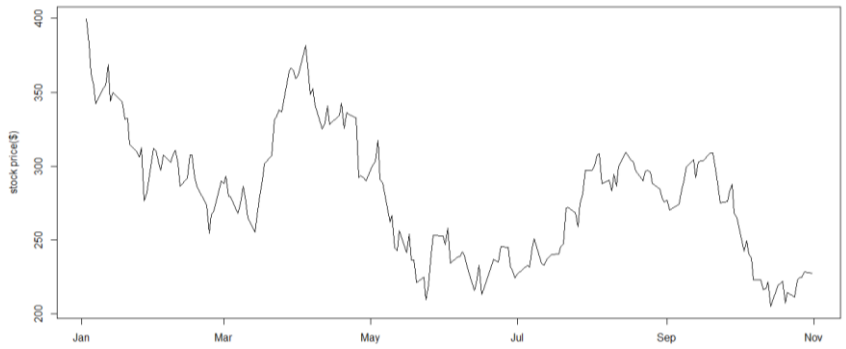

Figure 1. Stock price for TSLA

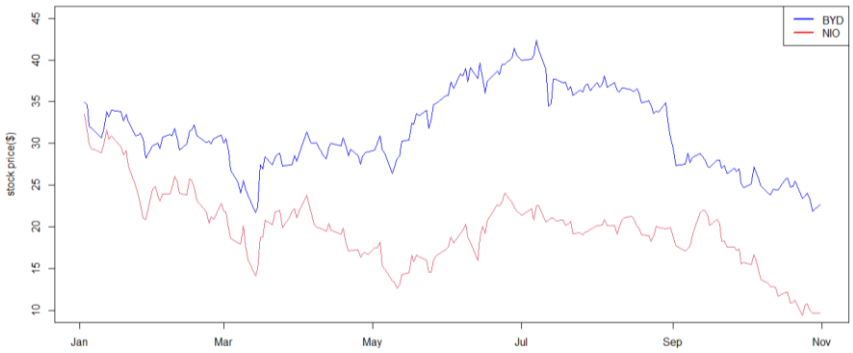

Figure 2. Stock price for BYD and NIO

2.2. ARIMA

ARIMA model is the abbreviation of Autoregressive Integrated Moving Average Model, which was proposed by American statistician Jenkins and British statistician Box in the 1970 s. This model is mainly used in the short-term prediction of time series variables. A single random variable is not measurable, but the overall time series has a certain regularity, which can be predicted by the model. ARIMA model is composed of AR model and MA model. The AR(p) model is an autoregressive model, where p is the autoregressive order. The MA(q) model is a moving average model, where q is the moving average order. In ARIMA(p, q, d) model, d is the difference number from non-stationary sequence to stationary sequence.

After differential processing of non-stationary sequences, the difference degree d is known. The ARMA(p, q) model is defined as $\{\varepsilon_t\}$~WN(0, σ^2) (i.e., white noise sequence with the mean of 0), Polynomials A(z) and B (z) have no common roots. Formulas as follows are satisfied.

$$A(z) = 1 - \sum_{j=1}^{p} a_j z^j \neq 0 \,, |z| \leq 1 \tag{1}$$

$$B(z) = 1 - \sum_{j=1}^{q} b_j z^j \neq 0 \,, |z| \leq 1 \tag{2}$$

And differential equation

$$X_t = \sum_{j=1}^{p} a_j X_{t-j} + \varepsilon_t + \sum_{j=1}^{q} b_j \varepsilon_{t-j} \tag{3}$$

The model satisfying the above equation is called ARMA model. It requires time series to be stationary series, non-stationary series can be obtained after d times difference. Thus, one can establish an ARIMA model (p, q, d).

2.3. Procedure of ARIMA

ARMIA modeling can be divided into the following steps:

- Preprocessing the time series data. Completing the smooth processing by differential operation, and using the unit root ADF test to test the stability of the time series.
- Determine the order of the model. Using the R language code to determine the order of time series, get the value of p, q to determine the ARIMA model.
- Calculate forecast results according to the model and analyze the results.

3. Results & Discussion

3.1. The Stationary Test Based on ADF

According to the stationary sequence requirements of the ARIMA model, the original data needs to be checked for stationarity. If the data is not stable, the original data needs to be differentiated to make it a stationary sequence, and the model can be established only after passing the stationarity test. For the time series, the preliminary discrimination can be made by the graph. According to the characteristics of the sequence diagram and the autocorrelation diagram, it can be preliminarily judged whether the series has stationarity. Then, use hypothesis testing to check whether the data is stationarity. The acf function of the tseries package in R is used to plot the autocorrelation graph, and the adf.test function is used to test the data. The following is the autocorrelation chart and test results of the stock price data of the three companies. The test result as shown in Fig. 3.

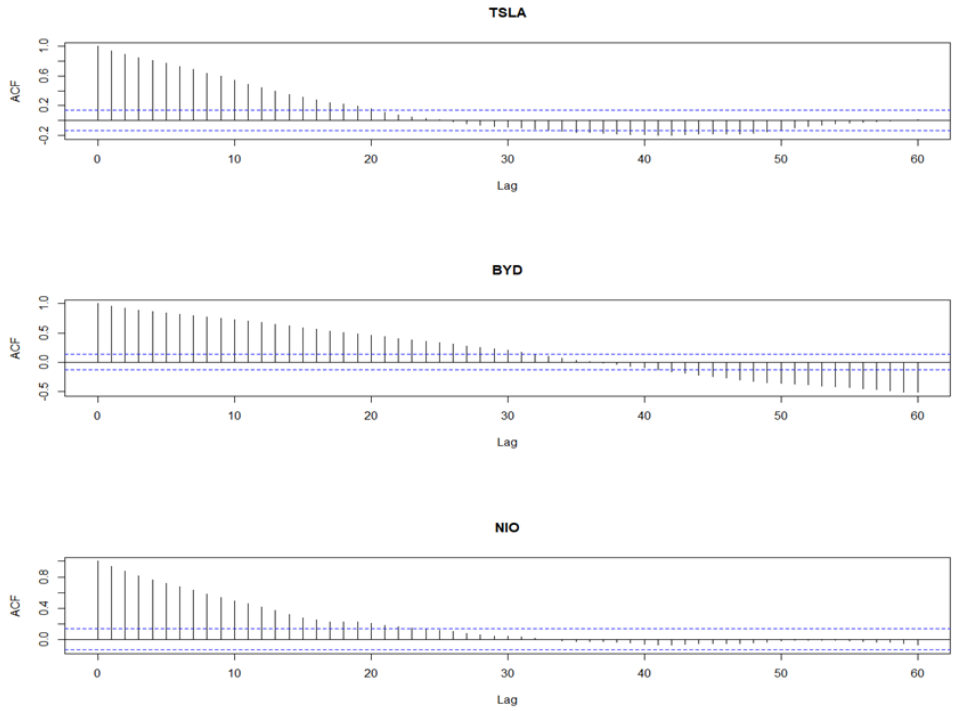

Figure 3. ADF test results

According to the autocorrelation chart of the three enterprises, the autocorrelation coefficient of all the series decreases to zero quite slowly. In a long delay period, the autocorrelation coefficient has been positive, and the overall triangular symmetry is also obvious. Therefore, the three sets of data are nonstationary columns, and there is a certain monotonic trend. Based on the results of hypothesis testing, the p-values of the three sets of data are 0.4823, 0.9656, and 0.5275, all of which are greater than 0.05. Accepting the null hypothesis, three sets of data are non-stationary sequences.

3.2. Differential Operational

After performing first-order differences on the three sets of data, the timing diagram is presented and the unit root test is performed again. As illustrated in the Fig. 4, the three sets of differential data basically swim around 0, which can be regarded as a stationary series. As shown in the Table 1, the p-values of the three sets of differential data are all less than 0.05, hence one rejects the null hypothesis and regard three them as stationary sequence.

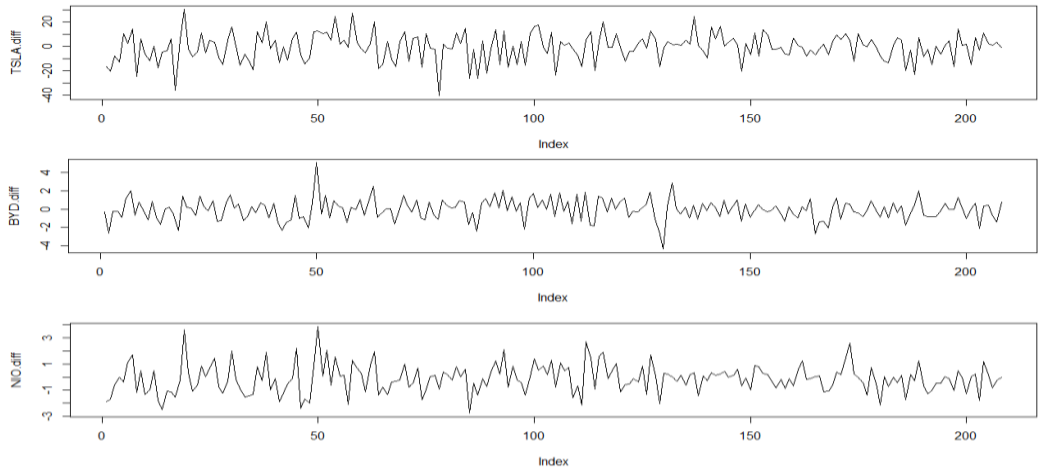

Figure 4. First-order differences results

Table 1. Results of the unit root test

	Dickey-Fuller	Lag order	p-value
TSLA	-6.2732	5	0.01
BYD	-6.3254	5	0.01
NIO	-6.6524	5	0.01

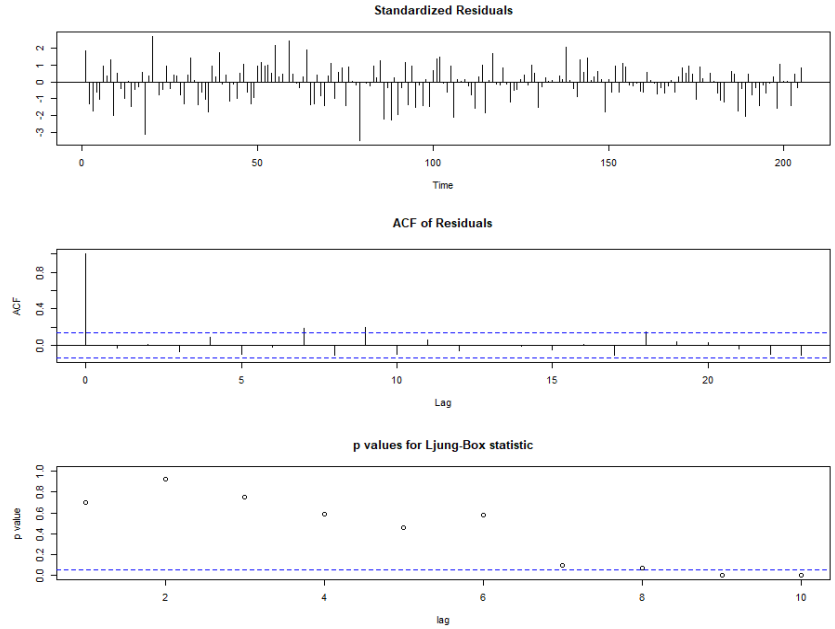

Figure 5. T test result for ARIMA (2, 0, 1)

3.3. Identification and Establishment of ARIMA Models

Now, one can identify the model order of the resulting stationary sequences to determine theirs, p, q, d. The auto.arima function of the forecast package in R language can automatically identify the order of the time series, and the method is more accurate and objective according to the BIC criterion. As a result of the code operation, the orders of TSLA are 2, 0, and 1. Model ARIMA (2, 0, 1) is based

on TSLA's stock price on the previous 205 days. Model should be test for significance to ensure that it is reasonable. Fig. 5 is the standard deviation of the model residuals. The self-co-assisted variance of the residuals is within twice the standard deviation, and the p-value is more than 0.05, indicating that the residuals are not correlated. So, the residual terms belong to the white noise sequence, and the valid information in the residuals has been extracted completely. According to the results of the t-test, the p-values of the model are all less than 0.05. The model is reasonable. The ARIMA model is suitable for short-term forecasting, and highly accurate in short-term forecasting, so the stock price will be forecasted for the next four trading days. Forecast results are as shown in Fig. 6 and Table 2. As a result of the code operation, the orders of BYD are 1, 0, and 1. The test is the same as above and the results are depicted in Fig. 7. Forecast results are presented in Fig. 8 and Table 3. According to the code operation results, the orders of NIO are 2, 0, and 2, respectively. The test is the same as above and the results are exhibited in Fig. 9. Forecast results are given in Fig. 10 and Table. 4.

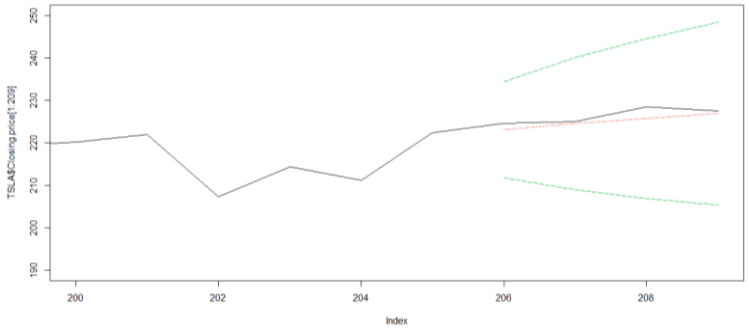

Figure 6. Forecast results of TSLA

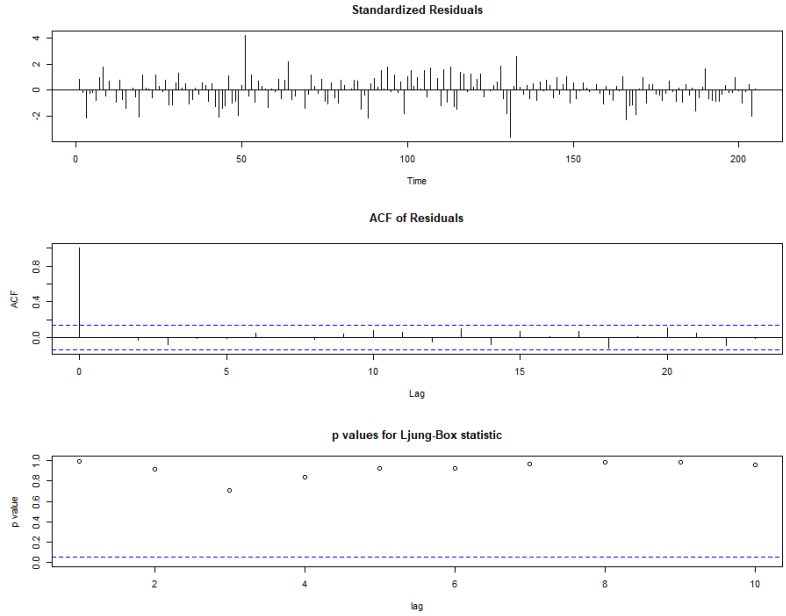

Figure 7. T test result for ARIMA (1, 0, 1)

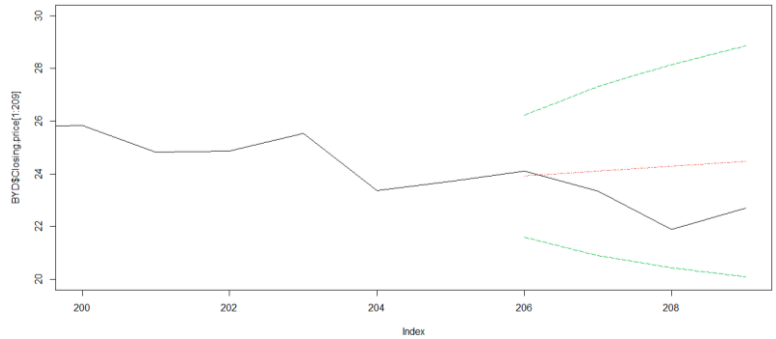

Figure 8. Forecast results of BYD

Table 2. Predicted value and error estimates for TSLA

Date	Predicted value	Std	Actual value	Error estimates
2022/10/26	223.15	11.35	224.64	0.66%
2022/10/27	224.56	15.57	225.09	0.24%
2022/10/28	225.71	18.86	228.52	1.23%
2022/10/31	226.93	21.53	227.54	0.27%

Table 3. Predicted value and error estimates for BYD

Date	Predicted value	Std	Actual value	Error estimates
2022/10/26	23.92	1.16	24.10	0.75%
2022/10/27	24.11	1.60	23.34	3.31%
2022/10/28	24.30	1.93	21.91	10.93%
2022/10/31	24.49	2.19	22.70	7.87%

Table 4. Predicted value and error estimates for NIO

Date	Predicted value	Std	Actual value	Error estimates
2022/10/26	10.82	1.09	10.81	0.14%
2022/10/27	11.11	1.60	9.98	11.33%
2022/10/28	11.22	2.01	9.69	15.76%
2022/10/31	11.42	2.32	9.67	18.07%

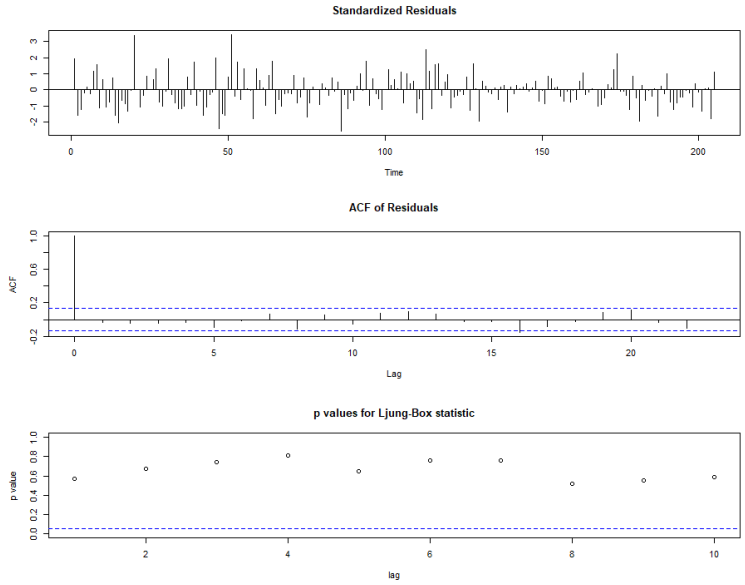

Figure 9. T test result for ARIMA (2, 0, 2)

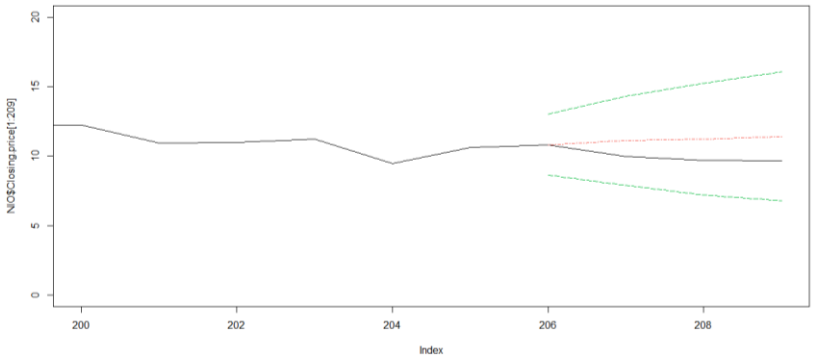

Figure 10. Forecast results of NIO

3.4. Comparison

Based on the results of the above forecasts, it is ideal to make stock price predictions through the ARIMA model. The error in the stock price forecast for the next day is within one percent. As the date increases, the error gradually becomes larger, but the actual value is still within twice the standard deviation of the predicted value. According to the ARIMA model, it can reflect the change of the closing price of new energy vehicle companies in the future, which can predict the future trend of stock prices. Comparing the forecast data of the three companies, the accuracy of the forecast results is more affected by orders of magnitude of the stock price. The higher the stock price, the more accurate the prediction will be. The large data prediction model for the user's electricity consumption is implemented in the Clementine software.

4. Limitations & Prospects

Predicting stock prices is a classic problem, and forecasting using the ARIMA model is a good solution. However, the ARIMA model is still a relatively basic mathematical model. The ARIMA model uses a stationary sequence, and the stock price will be affected by many other factors. Hence, not all of them may meet the requirements of a stationary sequence. These lead to errors in the results, which are not obvious in short-term forecasting, but affect the accuracy of long-term forecasting. In the era of big data, ARIMA models combined with machine learning methods such as neural networks are better options for predicting stock prices, and more data added to the calculation will make the prediction results more accurate.

5. Conclusion

In summary, this paper investigates the prediction of the stock price of new energy vehicle enterprises based on ARIMA model. Specifically, ARIMA model can be used for Short-term forecasting. The stock prices in the last 205 days of three company are utilized to build a reasonable ARIMA model. First, the timeseries is checked for stationarity, and then the auto.arima() code in R language is used to automatically identify the model order. Meanwhile a suitable ARIMA model is constructed, and the model is correlated tested, and finally the predict( ) code in R language is used to predict the stock price in the future trading day. According to the analysis, TSLA, BYD and NIO’s forecast results for the next four days are relatively accurate. Nevertheless, this paper has limitations. ARIMA is the basic mathematical model, and which combined with the deep learning neural network will get more accurate prediction results. Overall, these results offer a guideline for stock investors in the field of new energy vehicles.

References

[1] Wang Huifang, Shi Shuling. Research on the enlightenment of foreign new energy vehicle policies on the development of China's automobile industry [J]. Inner Mongolia Science and Technology and Economy, 202, 24: 9-10.

[2] Yuan Bo. Development of China's new energy vehicle industry under the goal of carbon neutrality [J]. Management Engineer, 2022, 27(05): 5-10.

[3] Liu Xueping. Research on stock price prediction of new energy automobile industry based on LSTM model. Chongqing University, 2021.

[4] Kumar M, Thenmozhi M. Forecasting stock index returns using ARIMA-SVM, ARIMA-ANN, and ARIMA-random forest hybrid models [J]. International Journal of Banking, Accounting and Finance, 2014, 5(3): 284-308.

[5] Mondal P, Shit L, Goswami S. Study of effectiveness of time series modeling (ARIMA) in forecasting stock prices [J]. International Journal of Computer Science, Engineering and Applications, 2014, 4(2): 13.

[6] Afeef M, Ihsan A, Zada H. Forecasting stock prices through univariate ARIMA modeling [J]. NUML International Journal of Business & Management, 2018, 13(2): 130-143.

[7] Adebiyi A A, Adewumi A O, Ayo C K. Comparison of ARIMA and artificial neural networks models for stock price prediction [J]. Journal of Applied Mathematics, 2014, 2014.

[8] Chen Dengjian, Du Feixia, Xia Xian. Stock prediction based on ARIMA and SVR rolling residual model combination [J]. Computer Age, 2022, 5: 76-81.

[9] Xiong Zheng, Chi Wengang. Application of ARIMA-GARCH-M model in short-term stock forecasting [J]. Journal of Shaanxi University of Technology (Natural Science Edition), 2022, 38(04): 69-74.

[10] Huang Shimin. Stock price analysis and prediction based on ARIMA model: Taking China Merchants Bank as an example [J]. SME Management and Technology, 2022, 11: 184-187.