Proceedings of the 2nd International Conference on Financial Technology and Business Analysis
DOI: 10.54254/2754-1169/65/20231677

An Empirical Study of Shanghai Composite Index Forecasts and Trend Trading Strategies Based on ARIMA Model

Zilan Zhang[1,a,*]

[1]Institute of Accounting, Wuhan Institute of Technology, Wuhan HB 430000, China
a. 22106020087@stu.wit.edu.cn
**corresponding author*

Abstract: ARIMA model is often used to forecast time series, and stock price prediction has always been a concern of investors. Accurate stock price prediction helps to make reasonable decisions in the unpredictable financial market. By modeling historical data, this paper uses ARIMA model to fit the change law of time series data, and then predict future stock changes. This paper selects the closing price data of Shanghai Composite Index on all trading days from August 19, 2019, to August 18, 2023, conducts stationarity test on it, completes the identification and sequencing of ARIMA model, conducts model test, analyzes and forecasts stock prices based on this model. The results show that the ARIMA model has a good forecasting effect on the short-term change rule of stock price time series, and has a certain reference significance for investors to make stock investment. The results show that the selected test data fits the ARIMA model well.

Keywords: Shanghai composite index, ARIMA model, forecast

1. Introduction

Along with several economic and political variables, investing psychology and trading technology also have an impact on the creation and volatility of stock values. The stock price is influenced by numerous things. The stock market operates under the principle that prices fluctuate in accordance with economic performance. But in fact, stock price not only has a close relationship with the internal financial situation of listed companies but also has a close relationship with the status of the whole stock market and even the overall economic operation.

The Shanghai Composite Index is a representative mix of stocks research. It is generally considered to reflect the level of the Chinese stock market. The abnormal reaction of individual company stock prices has a limited impact on the broader index. Therefore, it is more appropriate to use technical means to study stock price volatility to choose the Shanghai Composite Index as the research object. The corresponding also has guiding significance for portfolio operation and institutional or fund investment.

Because there are so many factors that might cause the stock price to fluctuate, the forecast is difficult to come true. To put it another way, it is difficult to predict the stock price with any degree of accuracy, yet we constantly research new methods and models to depict it. Because it is challenging to pinpoint the variables that genuinely influence the projected objects' change, the conventional regression analysis forecasting model is time-consuming and complicated. Additionally, due of the stock market's volatility, its forecast accuracy is no better than that of the Time Series Analysis approach. This method's model is often simple and affordable, making it particularly suitable for data that appear to defy logic. As a result, we simulate stock prices in time series analysis using the ARIMA model.

Shi took the monthly data of Shanghai Stock Composite Index (000001) and RMB/USD exchange rate from January 2000 to February 2019 as the research object [1]. Based on correlation test between sequences, Copula model, ARIMA model, ARIMAX model, and R language, the model is fitted by data sorting, stationarity test, model identification, optimal model selection, model prediction and other steps. By using the Spearman rank correlation test and the Kendall rank correlation test, it is proven that the RMB exchange rate and stock price are correlated and that the RMB exchange rate's volatility can impact the stock price's fluctuation.

Han adopted GARCH-MIDAS-A model, synthesizes the long and short term effects and asymmetric effects, and uses the daily and monthly mixed frequency data of Shanghai Composite Index from January 2016 to March 2021 to study the volatility of the stock market, in order to capture the lag effect and leverage effect in the stock market [2]. The empirical results prove the asymmetric effect of stock return volatility, and explain the long-term and short-term effects of different factors on stock volatility, such as early warning coefficient, consensus index, consumer price index, import volume, export volume, US dollar RMB exchange rate, cash in circulation, stock market turnover and the volatility of Shanghai Composite Index When Model A conducts mixed frequency data analysis, it is found that short-term imbalances occur in consistency index, consumer price index, USD RMB exchange rate and cash in circulation. In addition, the research shows that considering the long-term and short-term equilibrium effects of variables comprehensively, the accuracy of stock volatility prediction is increased.

Zhu and Zhang combined approximate time series representation based on change point detection and realized volatility to identify abnormal time series by clustering [3]. The empirical analysis based on the Shanghai Composite Index data shows that the introduction of realized volatility can further optimize the clustering quality, accurately identify the time series of abnormal volatility, and provide valuable decision support for practical financial analysis.

Liu and Zheng first compared three current mainstream co-jump test methods: LM based co-jump test, BLT co-jump test and FHLL co-jump test [4]. The results show that there is a significant difference in the number of co-jumps identified by the three methods, but the overlapping part of the results of the three methods is basically the market boom and slump, indicating that co-jump recognition is more sensitive to the aggregation of violent market fluctuations. Based on the aggregation problem of jumps and co-hops, this paper introduces Hawkes process into the study of jumps and co-hops, and constructs a factor model based on Hawkes process. The results show that the MJ statistics, CJ statistics and empirical data based on Hawkes factor model have a good fit, indicating that the factor model can better describe the aggregation of jumps and co-hops Sex.

Chen used deep learning for the first time to forecast volatility out of sample to improve the accuracy of volatility prediction, and the prediction results are compared with 19 classical models to evaluate the prediction effect in the field of high-frequency volatility prediction [5]. It is found that the prediction accuracy of deep learning ranks first under 5 kinds of loss functions. Compared with the second comparison model, the prediction accuracy increases by 13.16% and 9.72% respectively under different loss functions. Finally, deep learning is less affected by the change of historical days of key parameters. Under most historical days, LSTM model still has the best prediction effect in the test model, and the prediction effect of the model tends to be stable with the increase of historical days.

The central parity rate of the RMB exchange rate is the Granger cause of the return rate of the Shanghai Composite Index, and the return rate of the Shanghai Composite Index is also the Granger cause of the return rate of the RMB exchange rate, according to Wei and Fang's empirical research conducted in 2018 on the relationship between the Shanghai Composite Index and the central parity rate of the RMB exchange rate [6]. Further details on their quantitative relationship are provided, along with recommendations.

Huang analyzed and predicted Ping An's stock price based on the ARIMA model, selected 244 sample data of P/E ratio of Ping An from January 1, 2019 to December 31, 2019 as the research object, established the ARIMA model with R language, and predicted the return rate in the next 5 working days based on the model [7]. The forecast results are available for investors and management Provides decision-making reference.

Zheng and Du chose China's overnight lending rate as the research object and establishes ARIMA model to forecast it in the short term, and obtains the ideal short-term forecasting effect, thus determining the interest rate forecasting model suitable for the national conditions of the interbank lending market [8]. The research results can help financial institutions to price financial products reasonably and prevent risks.

Kuang and Liang used the ARIMA model to predict the S&P500 index [9]. The original data was collected from Yahoo finance database, and the research data was the closing price of S&P500 index, ranging from 1990-1-3 to 2012-3-26. The data was divided into modeling data and test data. The results show that the prediction model is in ARIMA (5,1,4) form, and the average prediction accuracy of the model is 1.8%. The research results can provide theoretical and empirical reference for financial investment.

Liu used the time series ARIMA model to conduct quantitative analysis of Shenzhen Stock Exchange index [10]. Taking the closing price of Shenzhen Stock Exchange Index from July 1, 2009 to June 30, 2010 as the original data, ARIMA (6,1,6) is finally established after the data is processed by stationary and zero-averaging, model identification and model ordering, and parameters are estimated by the least square method, and after the model test proved effective, short-term prediction of future data was carried out.

2. Methods

2.1. Data Source

Since July 15, 1991, the Shanghai Stock Composite Index, which includes all the equities listed on the Shanghai Stock Exchange, including A shares and B shares, has been publicly available. This index reflects price fluctuations of the stocks listed on the Shanghai Stock Exchange. The first index to be made public was the Shanghai Composite Index, which is a weighted composite stock price index based on all the equities listed on the Shanghai Stock Exchange and uses the weight of circulation. With a base date of December 19, 1990 and a base day index of 100 points, the index has been available in real time since July 15, 1991. It is calculated using the following equation.

$$\text{Shanghai Composite Index} = \frac{\text{Total market value of stocks in the reporting period}}{\text{Total market value of stocks in the base period}} \tag{1}$$

In this paper, the daily trading data of Shanghai Composite index from August 19, 2019 to August 18, 2023 were obtained from CSMAR Data Online as a sample. Finally, a total of 972 sets of daily data of Shanghai Composite Index were obtained. The following Figure 1 shows original data.

Figure 1: Shanghai Composite Index Points.

2.2. Method Introduction

The autoregressive, integrated, and moving average (ARIMA) model is a time series analysis technique. Time series data forecasting and modeling frequently use the ARIMA model. Through analysis and fitting previous time series data, it can forecast current patterns and changes. Data ordered chronologically, changing over time, and connected to one another are all examples of time series data. We can assess and forecast future data by analyzing trends in historical data. To check the stationarity of the sequence, we often start with the ADF test. Partial autocorrelation analysis may be carried out on a stationary sequence. In the event that the sequence is non-stationary, the data is differentially treated before the stationarity test is run.

3. Results and Discussion

3.1. Parameter Determination

In order to collect the Shanghai Composite Index's daily trading data from August 19, 2019, to August 18, 2023, this article uses CSMAR as a sample. Finally, 972 sets of the Shanghai Composite Index's daily data are presented. Figure 1 below illustrates how well the base period data fit.

The matching ARIMA (p, d, q) model must first be built after the parameters p, d, and q have been specified. As can be seen from Table 1, for closing price, the t statistic of the ADF test of this time series data is -2.465, the p value is 0.124, and the critical values of 1%, 5%, and 10% are, respectively, -3.437, -2.865, and -2.568. $p=0.124>0.1$, null hypothesis cannot be rejected, and the sequence is unstable. The ADF test is run after the first difference in the sequence. The sequence is stable, the null hypothesis is rejected with greater than 99% certainty, and the ADF test findings for the data after first-order difference show that p=0.0000.01.

The values of p and q can be calculated to be 0 and 0, respectively, from the autocorrelation coefficient (ACF) diagram in Figure 2 and the partial correlation coefficient (PACF) diagram in Figure 3, resulting in the ARIMA (1,0,0) model.

Table 1: ACF and PACF result tables.

Order of lag	AC value	PAC value	Q statistics	p value
1	0.908	0.908	84.893	0.000
2	0.812	-0.066	153.580	0.000
3	0.709	-0.097	206.474	0.000
4	0.608	-0.046	245.812	0.000
5	0.521	0.017	274.998	0.000
6	0.428	-0.094	294.920	0.000
7	0.366	0.105	309.596	0.000
8	0.291	-0.122	318.998	0.000
9	0.223	-0.029	324.551	0.000
10	0.139	-0.147	326.741	0.000
11	0.081	0.109	327.488	0.000
12	0.029	-0.052	327.585	0.000
13	-0.017	0.007	327.620	0.000
14	-0.042	0.024	327.833	0.000
15	-0.046	0.124	328.089	0.000

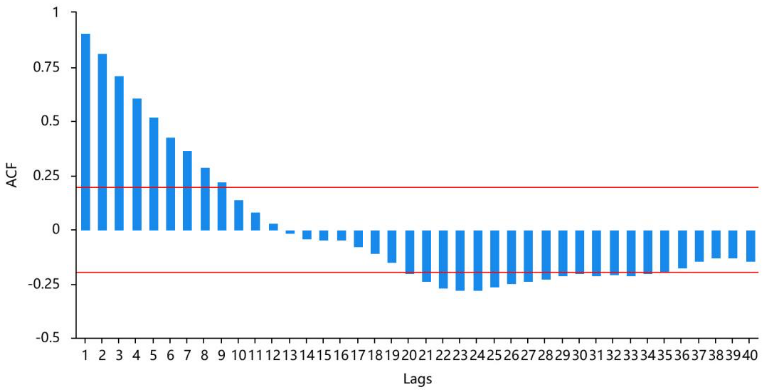

Figure 2: ACF Chart.

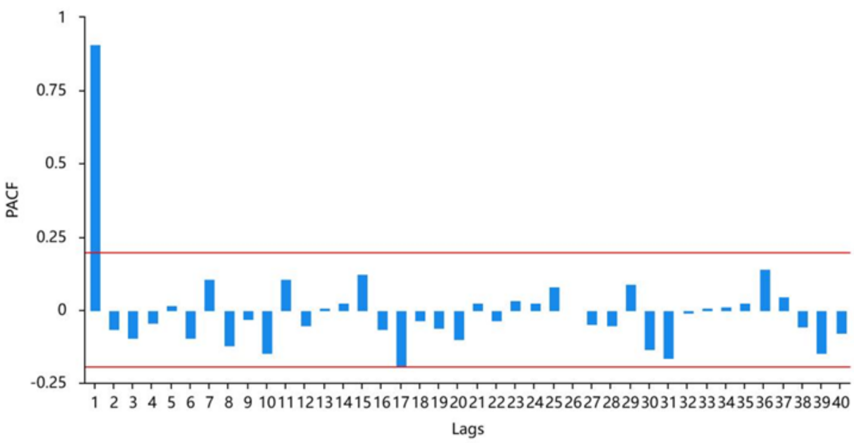

Figure 3: PACF Chart.

3.2. Estimated Values

We choose the data for the period 2023-8-21 to 2023-9-6 were used for testing. The results show that the price matching results are good. So we can come to a conclusion that the ARIMA model can accurately predict the stock price of Shanghai Composite Index, and the forecast results can provide reference for investors and policy makers. Table 2 provides the results of the analysis and Figure 4 visualizes the results as follows. Note that, Root mean square error (RMSE) is 34.8121, Mean square error (MSE) is 1211.8823, Mean absolute error (MAE) is 24.2574 and Mean absolute percentage error (MAPE) is 0.0075.

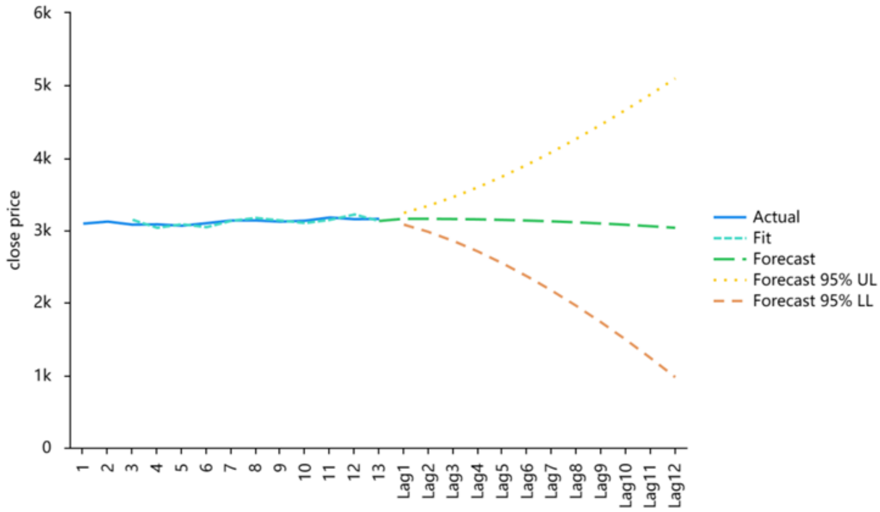

Figure 4: Fitted results.

Table 2: Predicted value (12 phases).

period	1	2	3	4	5	6	7	8	9	10	11	12
Value	3132	3133.531	3134.309	3135.080	3135.842	3136.598	3137.346	3138.087	3138.821	3139.547	3140.267	3140.9

4. Conclusion

In this paper, a time series forecasting method of financial market is proposed. The data are divided into two parts, where the data from 2019-8-19 to 2023-8-18 are used for modelling, and the second stage is to use the data from 2023-8-21 to 2023-9-6 for verification. The results show that the selected test data fits the ARIMA model very well. Therefore, when making stock price forecasts, it may be a good choice to apply the ARIMA model.

References

[1] Shi, Y. (2019) An Analysis Based on Copula-ARIMA Model of the Correlation Between Stock Price and RMB Exchange Rate. Journal of Shanxi University of Finance and Economics, 41(S2), 14-19.

[2] Han, Y.L. (2021) An Volatility Analysis of Shanghai Composite Index Based on Asymmetric GARCH-MIDAS Model. Statistics and Decision, 5(24), 157-160.

[3] Zhu, Y.Q. and Zhang, B. (2021) An Anomaly Timing Detection of Shanghai Composite Index Based on Realized Volatility. Systems Engineering Theory and Practice, 41(03), 625-635.

[4] Liu, Z.D. and Zheng, X.F. (2018) A Common Research based on Hawkes jump factor model shares [J]. Journal of Management Science in China, 26(7), 18-31.

DOI: 10.54254/2754-1169/65/20231677

[5] *Chen, W.H. (2018) A Comparative Research on Volatility Prediction Effect of Shanghai Composite Index Based on Deep Learning. Statistics and Information Forum, 33(05), 99-106.*

[6] *Wei, C. and Fang, H. (2020) An Empirical Study on the Return Rate of Shanghai Composite Index and RMB Central Parity Based on VAR Model. Economic Research Guide, 3, 58-61.*

[7] *Huang, L.X. (2020) Stock Price Analysis and Prediction based on ARIMA Model -- Taking Ping An of China as an example [J]. Science and Technology Economic Market, (10),62-63.*

[8] *Zheng, Y.T. and Du, Z.P. (2007) Application of ARIMA model in interbank lending rate prediction. Journal of Taiyuan Normal University (Natural Science Edition), (01), 45-47+74.*

[9] *Kuang, Y. and Liang, Z.J. (2012) A Prediction Analysis of S&P500 Index Based on ARIMA Model. Modern Trade Industry, 24(14), 100-102.*

[10] *Liu, M.X. (2011) An Analysis and Prediction of Shenzhen Stock Exchange Index Based on ARIMA Model. Chinese Urban Economy, (18), 70-71.*