Research on the Performance of the ARIMA Model in the Stock Market

Zijun Li[1, a, *]

[1]Maynooth International Engineering College, Fuzhou University, Fuzhou, 350108, China
a. ZIJUN.LI.2022@MUMAIL.IE
**corresponding author*

Abstract: With the rapid development and upgrading transformation of Chinese financial industry and increasing investment quality, the investment enthusiasm of the masses has been improved. The prediction of stock price has an effective reference for investors to determine investment strategies. In this paper, ARIMA model is used to compare the model performance of Shanghai Composite Index, Shenzhen Composite Index and Science and Technology Innovation Board. This model is also fitted to the normal period and the shock period of the stock market respectively. The study found that the Shanghai Composite Index is the most mature market, and the fitting performance of the stock market in the normal period is better than that in the shock period. From the middleof 2015 to the beginning of 2016, the Shanghai Composite Index and Shenzhen Composite Index fluctuated significantly. This study not only helps investors to adopt more reasonable investment strategies, but also has important reference value for market regulators to guide the market effectively, avoid violent fluctuations in the stock market and maintain market stability.

Keywords: time series analysis, financial forecasting, model performs evaluation

1. Introduction

Stock investment is one of the main means of investment. In recent years, an increasing number of people have begun to follow the stock market and take an interest in investing. Despite having a very short history, China's stock market has seen a number of market crashes. One took place in 1996. The Shenzhen Component index increased by 340% and the Shanghai Composite Index by 120% between April 1 and December 9. A second occurred in 2001. The stock market crashed on July 26 of that year, and the Shanghai index dropped 32.55 points as a result of the official launch of fresh share issuance and the decline of state-owned shares. On June 14, the Shanghai index stood at 2,245. By October 19, it had fallen to 1,514, with more than 50 stocks dropping below the daily threshold. In 2001, 80% of investors were locked up, funds lost 40% of their value and brokerage fees fell 30%. For the third time, from November 2007 to October 2008, the Shanghai Composite Index dropped from 6124 points to 1664 points, with a maximum decline of 72.8%; The most recent, from June 2015 to January 2016, saw the Shanghai index fall from 5,178 points to 2,638 points in eight months, with a maximum drop of 49.05%. The moves in the stock market are so disconnected from the fundamentals of the economy that they are doomed to be unsustainable. Some investors speculative mentality is too rich, with less investment experience, eventually may have to bear a certain amount of losses.

Financial time series model is the most commonly used method in stock prediction, and is also considered to be one of the best tools to predict stock market changes. Accurate stock prediction can help investors make correct judgments. For individual investors, the high-speed, simple and accurate lightweight time series model is more suitable for the application of individual investors in daily life, and it is also the research goal of many researchers [1].

Xiong combined the ARIMA model with other models and compared the test results of the ARIMA model, the ARIMA-GARCH model and the ARIMA-GARCH-M model on a single stock. The results show that ARIMA-GARCH-M model with recursive correction has a good effect on short-term stock prediction [1]. In order to anticipate and assess the Shanghai Composite Index of stocks in the future, Zhang et al. improved the ARIMA model. They discovered that the model had a decent prediction impact over the next four trading days [2]. To test and forecast the historical closing price data of a few financial market equities, Liu et al. developed the ARIMA model. They found that the highest error between the anticipated value and the actual value for the next three days was not higher than 4% [3]. Yang et al. and Xu also proposed the improved model of ARIMA to forecast stock prices [4,5]. He combines two different deep learning models of natural language processing NLP and time series processing LSTM to forecast and analyze the CSI 300 index and finds that the prediction accuracy has been significantly improved [6]. Fu et al. used the LSTM model in the neural network to predict the future trend, indicating that in the post-COVID-19 period, the US stock market continued to decline [7]. Tang used the Attention-LSTM method to build a model to forecast the rise and fall trend of the Shanghai 50 Index at its highest price, and the results showed that the constructed quantitative investment method had better forecasting ability and its accuracy could reach 63.42% [8]. Ruan et al. trained ARIMA and LSTM models at the level of individual stock, industry and general market respectively, and incorporated sentiment analysis into the model. Based on the stock prices of 100 companies from 2016 to 2020, the average prediction accuracy of the model was 98% [9]. Zhou also incorporated sentiment analysis into LSTM, improving the effect of the model in predicting stock prices [10]. Chen et al. [11], Song et al. and Li all proposed an improved LSTM model and applied it to stock price prediction, effectively improving the calculation speed and prediction accuracy of the model [12, 13]. Li compared the performance of multi-factor model and BP neural network model on quantitative investment, indicating that the neural network model may not be optimal in terms of obtaining excess returns, but the model is more stable [14]. Yang used four commonly used machine learning algorithms, namely support vector regression, long and short-term memory network, random forest and extreme gradient lifting tree, to construct a price prediction model [15]. This paper studied the prediction of CSI 300 stock index futures prices and used Bayesian algorithm to optimize the hyperparameter of the model. It was found that random forest and extreme gradient lifting tree could achieve accurate prediction of financial time series data, and Bayesian optimization could significantly improve the prediction effect of support vector regression by using Gaussian process and constantly updating the prior.

There are many methods for stock price prediction, including time series, neural network, deep learning, etc., for a single stock prediction, or for the same sector using different prediction methods. However, there is little research on using the same method to analyze and compare different plates or comparing the same plates at different time periods. This paper will make up for this gap. This paper will use the ARIMA model and discuss two aspects. The first is to explore the fitting degree of the ARIMA model to different plates, and the second is to explore the performance of the ARIMA model in normal times and shock times. It aims to evaluate the maturity of the stock market and provide reference information for investors to make investment strategies. For regulators, provide possible early warning, to make plans when the stock market turbulence. The

DOI: 10.54254/2754-1169/46/20230322

research of this paper is also a useful extension of the existing theoretical research on the field of stock price prediction.

The rest of this paper is structured as followings: Section II introduces the data and methods, Section III presents the results, and Section IV is the conclusions.

2. Data and Method

2.1. Data

2.1.1. Data Indicators and Sources

The closing prices of Shanghai Stock Index, Shenzhen Stock Index and Science and Innovation Board index in CSMAR are used as the data sample. This paper selects the daily market data from early 2013 to April 28, 2023 (starting at the end of 2019 of the Science and Technology Innovation Board).

2.1.2. Descriptive Statistical Results

As can be seen from Figure 1, the Shanghai Stock Exchange showed an obvious upward trend in 2014-2015, a marked decline from 2015 to 2016, and a steady fluctuation thereafter. The peak was within 2015 years.

From 2014 to 2015, the Shenzhen Stock Exchange showed an obvious upward trend, and from 2015 to 2016, it decreased significantly, which was the period of the domestic stock market crash. After 2018 years, it also fell. The overall fluctuation range was sharp. It can also be seen from Table 1 shows that the Shenzhen Stock Exchange's standard deviation was higher than the Shanghai Stock Exchange's. The peak was within 2015 years.

The index of the Science and Innovation Board showed a downward trend at the end of 2021, and the overall fluctuation range was not obvious. It can be seen from Table 1 that the standard deviation of the Science and Innovation board was not large.

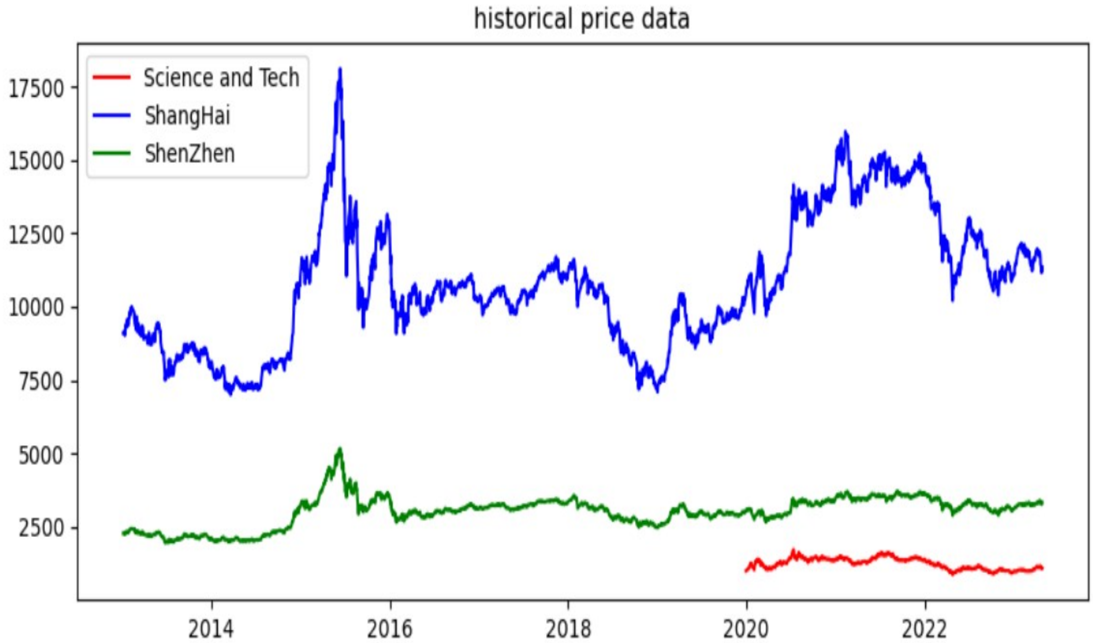

Figure 1: Volatility of Shanghai Stock Exchange, Shenzhen Stock Exchange and Science and Technology Innovation Board.

DOI: 10.54254/2754-1169/46/20230322

Table 1: Mean value, maximum value, minimum value and standard deviation of Shanghai Stock Index, Shenzhen Stock Index and Science and Technology Innovation Board.

	Mean	Maximum	Minimum	Standard Deviation
Shanghai	3025.18	5166.35	1950.01	521.11
Shenzhen	10816.79	18098.27	6998.19	2203.66
Science and Innovation	1246.33	1721.98	868.07	190.21

2.2. Method

2.2.1. ARIMA Model

One of the time series prediction analytic techniques is the ARIMA Model, also known as the Autoregressive Integrated Moving Average Model, differential integrated moving average autoregressive model, and integrated moving average autoregressive model (movement is sometimes referred to as sliding). Is a famous time series prediction method proposed by Box and Jenkins in the early 1970s, also known as Box-Jenkins model. In essence, ARIMA model simulates random changes of data through autocorrelation and partial autocorrelation functions, so as to achieve the purpose of predicting future trends of data [1]. ARIMA(p, d, q) model can be expressed by the formula(1)

$$dX_t = c + \sum_{i=1}^{p} \varphi_i dX_{t-i} + \varepsilon_t + \sum_{j=1}^{q} \theta_j \varepsilon_{t-j} \tag{1}$$

where p, d and q are the autoregressive order, difference order and moving average order respectively, X_t is the time series of time t, φ_i and θ_j are the stationary polynomial operators of order p and order q respectively, and ε_t is a random error term.

2.2.2. Experimental Procedure

This paper will forecast the Shanghai Stock Index, Shenzhen Stock Index and Science and Innovation Board index. The main steps are as follows:

(1) Import the dataset;

(2) Use the ADF test to completes the test of stock price stability;

(3) If the stock sequence is non-stationary, differential processing is performed on the sequence to convert it into stationary sequence, and then ADF is used the second time to verify the stability of the processed data;

(4) Determine the values of ARIMA parameters p,d and q to build the optimal ARIMA model;

(5) Divide the data set into training set and test set according to 7:3;

(6) Use test sets to predict correlation values;

(7) Compare the predicted value with the actual value by using mean error as the standard to measure the quality of the predicted result;

(8) Repeat the above steps to test the Shanghai Stock Index, Shenzhen Stock Index and Science and Innovation Board index respectively;

(9) Repeat the above steps by dividing the Shanghai Stock Index and Shenzhen Stock Index into 17-19, 20-22 respectively, and adding MSE as the standard to measure the predicted results;

DOI: 10.54254/2754-1169/46/20230322

2.2.3. Measurement Standards

(1) For the purpose of research 1, due to the different bases of different plates, the same measurement standard is selected and formula (2) is used for measurement. The smaller the mean error is, the smaller the prediction error is and the higher the prediction accuracy will be.

$$\text{mean error} = \frac{1}{n}\sum_{i=1}^{n}\frac{|\text{real}_i - \text{predict}_i|}{\text{real}} \times 100\% \tag{2}$$

(2) For the second purpose of the study, formula (3) is used for measurement while formula (2) is used. The smaller the value of MSE, the smaller the prediction error and the higher the prediction accuracy will be.

$$\text{MSE} = \frac{1}{n}\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2 \tag{3}$$

3. Results

3.1. Stationarity Test Results

Because time series analysis only works with stationary data, it is necessary to determine if a sequence is stationary. One of the most common statistical tests is the ADF (Augmented Dickey-Fuller) Test. It can be used to assess whether a series of integers has unit roots, which will help readers establish whether the sequence is stationary. Null hypothesis: The sequence has a unit root (a =1). Alternative hypothesis: The sequence has no unit root. It suggests that the sequence is not steady even if the null hypothesis is not rejected. The sequence can thus be differentially stationary or linear. The series becomes smooth if the mean and standard deviation are both flat lines (constant mean and constant variance). According to Table 2,3, and 4, it can be seen that the p value is too large before stabilization to reject the null hypothesis, and the data is not stable. After the stabilization, the null hypothesis is received and the data is stabilized.

Table 2: Shanghai Composite Index.

	Test Statistics	p-value	Critical value(1%)	Critical value(5%)	Critical value(10%)
Before stabilization	-2.328005	0.163095	-3.432988	-2.862706	-2.567391
After stabilization	-10.085487	0.000000			

Table 3: Shenzhen Composite Index.

	Test Statistics	p-value	Critical value(1%)	Critical value(5%)	Critical value(10%)
Before stabilization	-2.299523	0.172122	-3.432988	-2.862706	-2.567391
After stabilization	-13.325838	0.000000			

Table 4: Science and Innovation Board index.

	Test Statistics	p-value	Critical value(1%)	Critical value(5%)	Critical value(10%)
Before stabilization	-2.098577	0.245087	-3.432988	-2.862706	-2.567391
After stabilization	-27.306414	0.000000			

3.2. Prediction Results of Different Plates

It can be seen from Table 5-8 that the mean error of Shanghai Composite Index is the smallest (0.6%), and that of Science and Innovation board is the largest (1.1%), indicating that ARIMA model is the best in fitting and predicting the price of Shanghai Composite Index, followed by Shenzhen Composite index, and science and Innovation Board index has the worst performance. This shows that the Shanghai Index market is more mature and competitive than the Shenzhen index and the Science and Innovation Board index. The poor performance of the Science and Innovation Board index indicates that the market is not mature enough, the stock market is volatile, and investors need to take more risks.

Table 5: Comparison of 5 real values and predicted values of Shanghai Composite Index test set.

	Real Price	Forecast	Error_%
2023/2/17	3224.02	3248.92	0.8
2023/2/20	3290.34	3229.78	1.8
2023/2/21	3306.52	3291.85	0.4
2023/2/22	3291.15	3298.62	0.2
2023/2/23	3287.48	3290.67	0.1

Table 6: Comparison of 5 real values and predicted values of Shenzhen Index test set.

	Real Price	Forecast	Error_%
2023/2/17	11715.77	11914.75	1.7
2023/2/20	11954.13	11732.45	1.9
2023/2/21	11968.60	11968.23	0.0
2023/2/22	11900.12	11948.75	0.4
2023/2/23	11884.30	11888.64	0.0

Table 7: Comparison between real values and predicted values of 5 items in the test set of Science Innovation Board.

	Real Price	Forecast	Error_%
2023/2/17	986.33	1005.50	1.9
2023/2/20	998.83	984.89	1.4
2023/2/21	995.73	998.83	0.3
2023/2/22	989.65	995.99	0.6
2023/2/23	992.46	991.78	0.1

DOI: 10.54254/2754-1169/46/20230322

Table 8: mean error results of Shanghai Stock Index, Shenzhen Stock Index and Science and Innovation Board index.

	Shanghai	Shenzhen	Science and tech
Mean error	0.6%	0.7%	1.1%

3.3. Prediction Results Before and After the Epidemic

Through the above experiments, mean error and mse in normal period are both smaller than those in shock period (table 9 and table 10). It can be concluded that the ARIMA model performs best in fitting and forecasting for normal times. It shows that the stock market fluctuates greatly during the shock period, which is not conducive to investors investment.

(1) Mean error.

Table 9: The mean error performance of Shanghai Composite Index and Shenzhen Composite Index in different periods.

	Normal period17-19	Shock period20-22
Shanghai	0.5%	0.7%
Shenzhen	0.8%	0.9%

(2)Mse.

Table 10: The mse performance of Shanghai Composite Index and Shenzhen Composite Index in different periods.

	Normal period17-19	Shock period20-22
Shanghai	378.7783	904.0055
Shenzhen	8238.9072	18109.3140

4. Conclusion

This paper uses ARIMA model to study Shanghai Stock Index, Shenzhen Stock index and Science and Innovation board index. The main research conclusions include that, firstly, the Shanghai Composite Index market is the most mature and competitive, secondly, the model performance in normal period is better than that in shock period. In addition, it is also found that the Shanghai and Shenzhen indices fluctuated significantly from the middle of 2015 to the beginning of 2016.

The conclusions of this paper are helpful for investors to make more reasonable investment strategies, and have important implications for market regulators to effectively guide the market, avoid violent fluctuations in the stock market, and maintain market stability.

However, there are several limitations in this paper. This study only used one model of ARIMA for fitting, and there was no validation of other models, which needed to be further studied in the future.

References

[1] Xiong Z., Che W. (2022)Application of ARIMA-GARCH-M Model in Short-term stock prediction. Journal of Shaanxi University of Technology (Natural Science Edition), 38(4), 69-74.

[2] Zhang Y., Sun Y. (2019) Empirical Research on Shanghai Composite Index Analysis and Prediction Based on ARIMA Model. Economic Research Guide, 11, 131-135.

[3] Liu S., Zhang S. (2021) An Empirical Study on Stock Price Prediction Based on ARIMA Model. Economic Research Guide, 25, 76-78.

[4] Yang Q., Cao X. (2016) Stock price analysis and prediction based on ARMA-GARCH model. Practice and understanding of mathematics, 46(6), 80-86

[5] Xu S., Liang X. (2019) Research on Stock Forecasting Based on ArMIa-Garch Model. Journal of Henan Institute of Education, 28(4), 20-24

[6] He Y. (2021)Application of deep learning NLP technology and LSTM algorithm model in quantitative investment strategy. [Master Dissertation].

[7] Fu W., Li Z., Zhang Y., Zhou X. (2022) Time Series Analysis in American Stock Market Recovering in Post COVID-19 Pandemic Period. Cornell University Library, arXiv.org

[8] Tang C. (2021) Quantitative Investment Based on Attention-LSTM Deep Learning method. Roundtable Forum,1671-6728, 36-0155-03.

[9] Ruan J.,Wu W., Luo J. (2021) Stock Price Prediction Under Anomalous Circumstances. Cornell University Library, arXiv.org

[10] Zhou L. (2018) Research on stock quotation prediction based on LSTM and investor sentiment. Wuhan: Central China Normal University.

[11] Chen J., Liu D.,Wu D. (2019) Based on feature selection and LSTM model Research on the Method of Finger Prediction [J]. Computer Engineering and Applications, 55(6), 108 -112.

[12] Song G.,Zhang Y., Bao F., et al. (2019) Unit based on particle swarm optimization LSTM. Journal of Beijing University of Aeronautics and Astronautics,45 (12), 2533-2542.

[13] Li X. (2023) Research on Stock price Prediction based on SSA-LSTM neural Network, Information System Engineering, No.351(03), 48-50.

[14] Li X. (2018) Comparative Analysis of Neural network and multi-factor model in the field of quantitative investment. Market Forum, 1672-8777, 08-065-06.

[15] Yang X. (2022) Research on price prediction model of stock index futures based on machine learning algorithm, SOFTWARE ENGINEERING, 2096-1472, 12-01-07.