Study on the Fitness of ARIMA Model in Stock Forecasting

Jingwen Li[1*]

[1]School of Mathematical Sciences, Tiangong University, Tianjin, China

Abstract. With the continuous development of the world's financial industry, the international stock market trend prediction has now become one of the most closely related to the actual financial industry, with more and more scholars into the research, the accuracy of asset price prediction is constantly improving. Technical analysis is one of the most important components of securities analysis methods. Among the many prediction models, choosing the most appropriate prediction model for modelling according to the specific objectives is twice as effective and has become the source of motivation for the development of this study. In this study, the ARIMA model, which is one of the most popular models in this field of research, is used to forecast two of the most representative assets in the international stock market. On this basis, the results are discussed and analysed to conclude that compared with the CSI300 index, the accuracy of the model's prediction results for the SP500 is greater, which provides strong support and suggestions for the subsequent research on the two stocks, and also provides a valuable reference for the prediction of the current hot technology methods in the stock markets of different economies.

1 Introduction

The stock market has undergone more than four centuries of development since its birth, and its unique trading mechanism has become an indispensable part of the modern world economic system, while also providing a keyway for major enterprises to raise capital. The volatility of stock prices not only reflects the market's optimism about the future development potential of the enterprise, but also this volatility provides ordinary investors with profit opportunities and loss possibilities, so that countless investors tend to rush, the number of investors, the influence of a wide range of causes of its small fluctuations can be further to the economic stability of the whole society has a much larger impact [1]. Therefore, it is crucial to accurately predict stock price movements as it is the key to efficient stock investment and decision making.

In recent years, mathematical models have been constructed in order to analyse and predict stock prices, thus attempting to predict and reveal the pattern of stock market movements. Therefore, a large number of mathematical models have been put into use by researchers over the past period of time, and further improvements and technological updates have been made to overcome many real-life investment problems [2]. While past research

* Corresponding author: 2210810047@tiangong.edu.cn

has been highly innovative and scientific, it has tended to be less pragmatic in its approach to the highly relevant proposition of asset price prediction. Therefore, choosing the most appropriate forecasting methods and models for specific target assets is often half the battle, and this research is thus carried out.

In this study, two of the most important economies in the world market, i.e., the United States and China, are selected to explore the fitness of the same model for them through the US S&P 500 and CSI 300 indices. The model chosen is the autoregressive differential moving average model, i.e., the ARIMA model. Because the ARIMA model has always been the basic model used by many scholars and widely applied in asset forecasting in recent years' research work. For example, Edson used it to model and forecast the Brazilian stock market index, using the MAPE parameter to compare the results with those of other smoothing models, and the results showed that the model obtained a low MAPE value, meaning that it is highly applicable [3]. This shows that the ARIMA model can be used for time series indices related to stock market index forecasting [4].

While Li Xiaoning in his study based on the stock data of 00001 shares of the above certificate, used two forecasting methods, namely, multiple linear regression model and time series ARIMA model, to forecast the closing price of 00001 shares of the stock in the case of unknown future data, and compared the accuracy of the two models, which has been continuously improved by the scholars in the following [5]. Therefore, this study will expand this idea, explore the proposition of different asset fitness under the same model, compare the prediction error by establishing a model, and then select the smaller error, i.e., higher precision, by comparison, to explore and illustrate that the use of more appropriate models for different assets can achieve more accurate results, and promote the promotion of the subsequent development of research work in the direction of asset prediction by enhancing the fitness of the specific model and the stock. This will facilitate the development of subsequent research work in the direction of improving the accuracy of asset forecasting by enhancing the fit of specific models to stocks.

2 Data

This article selects the historical data of the S&P 500 (SPX) and CSI 300 (CSI300), the representative assets of the two global economies, i.e. the United States and China, for the last 200 days, i.e. 13 November 2023 to 4 September 2024, for the purpose of forecasting and MSE analysis. The data is sourced from the official website of Invesco, which is real and reliable with high reference value. Fig. 1 and Fig. 2 show the closing data of the two assets.

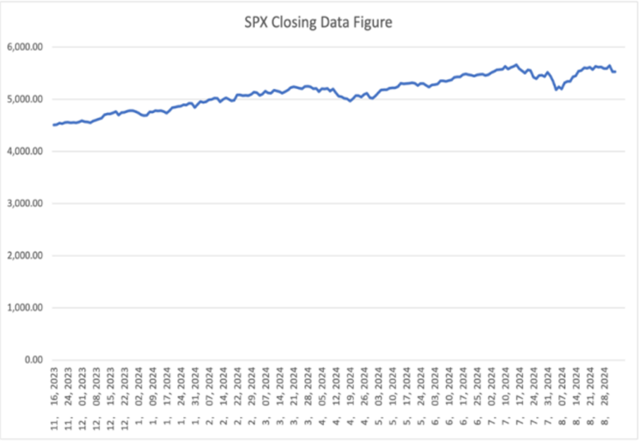


Fig. 1. SP500 Trend

Fig. 1 above shows a chart of the S&P 500, which shows that the stock has shown relatively flat growth over the period. This steady growth reflects the stable earnings growth of the companies behind the stock, as well as the favourable market environment during the 200-day period, which included a stable economy and an improved low interest rate environment. On the other hand, it also reflects the increase in confidence and risk appetite of the majority of investors in the stock [6].

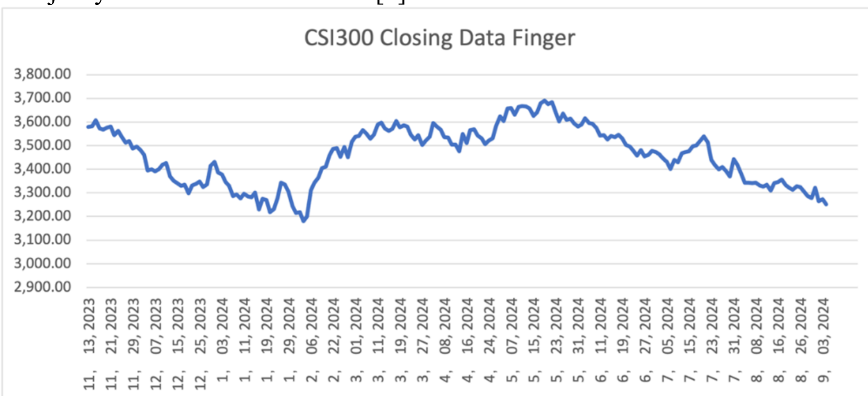


Fig. 2. CSI 300 Trend

Fig. 2 above shows a chart of the CSI 300 and comparing it to the S&P 500 above, it is clear to see that the stock has shown a more volatile and generally declining trend over the period. This volatility on the one hand is often significantly influenced by the factors of the investors behind the stock [7]. In particular, when they are faced with uncertainty about the economic outlook or unexpected social factors, their risk appetite changes tend to follow suit and have a significant impact [8].

At the same time, fluctuations in economic fundamentals during this period, such as a decline in the firm's internal profitability or a slowdown in economic growth in the broader external environment, can also drive a general downward trend in share prices. In addition, changes in capital liquidity can simultaneously reduce the supply of capital to the stock market, exacerbating the downward pressure on share prices.

3 Method

The first 150 days of the 200 days are used to predict the next 50 days, and then the predicted data are compared with the real data for the mean squared error, and then the prediction accuracy of the ARIMA model is investigated for different assets, so as to analyse what kind of prediction is more suitable to use the ARIMA model for the prediction research.

ARIMA is a method for predicting future trends based on historical time series data. It relies on the statistical concept of serial correlation, whereby past data points have an effect on future data points. Any ‘non-seasonal’ time series that exhibits a pattern and is not stochastic can be modelled by an ARIMA model.

The first step in constructing an ARIMA (d,p,q) model is to make the time series smooth and to determine the difference order d. The stability of the series is observed by plotting the series or by performing an ADF test [9]. The most common method is to perform a differencing, i.e., subtracting the previous value from the current value. The difference order d is the minimum number of differences required to make the series smooth, while both d=0, as seen in Fig. 3 below, reduces to an ARIMA (p, q) model [10].

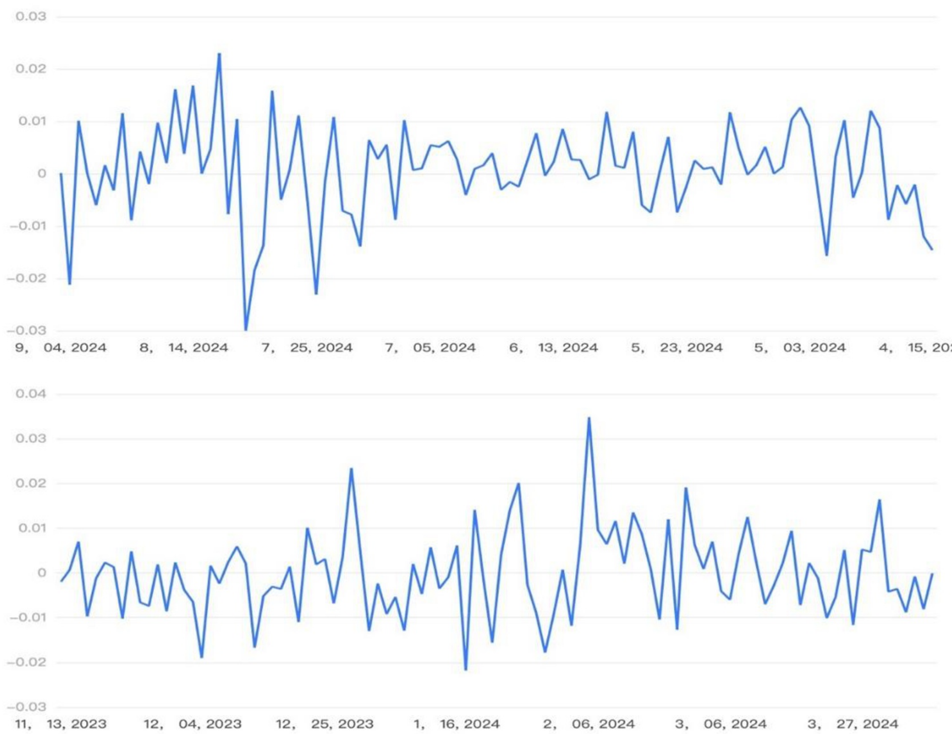


Fig. 3. Figure of the best differential series of two stocks (top SPX500; bottom CSI300)

Afterwards, the order p of the regression and the order q of the moving average are determined by comparing autocorrelation and partial autocorrelation.

$$ACF(k) = \rho_k = \frac{Cov(y_t, y_{t-k})}{Var(y_t)}, q_{(SP500)} = 1,, q_{(CSI300)} = 15 \tag{1}$$

$$PACF(k) = \frac{Cov[(Z_t - \overline{Z_t}), (Z_{t-k} - \overline{Z_{t-k}})]}{\sqrt{Var(Z_t - \overline{Z_t})}\sqrt{Var(Z_{t-k} - \overline{Z_{t-k}})}}, p_{(SP500)} = 6, p_{(CSI300)} = 12 \tag{2}$$

Then, an SP500 ARIMA (6,1) model and a CSI300 ARIMA (12,15) model can be constructed based on the respective values of p, q, and d.

4 Results

In this paper, the effective historical data of the rise and fall of the two assets from 13 November 2023 to 24 June 2024 is used as the training set to predict the rise and fall data of the two stocks from 27 June 2024 to 4 September 2024 by ARIMA model. Table 1 and Table 2 are the two stock indices of each predictive indicators and diagrams

Table 1. S&P 500 Index Various Forecast Indicators

Indicator	Value
MAE	0.0050
MAPE	Inf
RMSE	0.0065
R2	-567.6277
MSE	3.00642E-05

Table 2. CSI 300 Index Forecast Indicators

Indicator	Value
MAE	0.0068
MAPE	1.0432
RMSE	0.0087
R2	-138.4928
MSE	5.89844E-05

From each of the above prediction indicators, it can be clearly seen that for the values predicted by both using the ARIMA model, all the individual error values of the S&P 500 index are smaller than those of the CSI 300 index, indicating that the model can be better suited to the work of predicting the SP300 index to provide a more informative prediction value.

In the following section, the forecast data of the two stocks will be analysed more visually with images to observe the fitness of the ARIMA model for them.

In Fig. 4, the top panel shows the ARIMA forecast root mean square error (RMSE) plot for the S&P 500 index. It is obvious from its histogram that compared to the RMSE image of the CSI 300 index in the lower figure, it is more concentrated in the central region and the distribution of errors shows some positive and negative symmetry, i.e., most of the prediction errors are smaller and the ARIMA model predicts them more accurately. In other words, according to the RMSE index: the ARIMA model predicts the SP500 index more accurately than the CSI300 index.

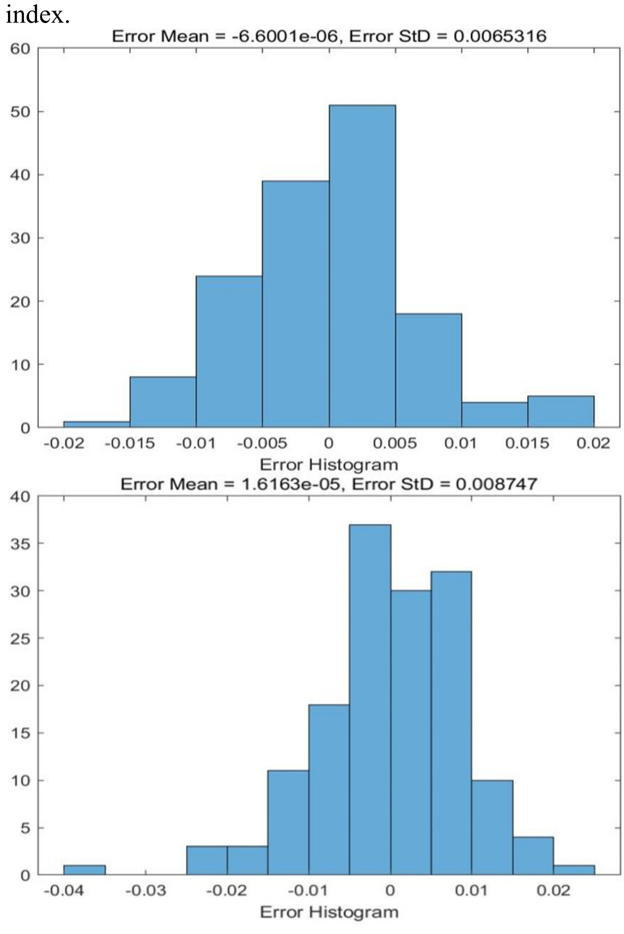
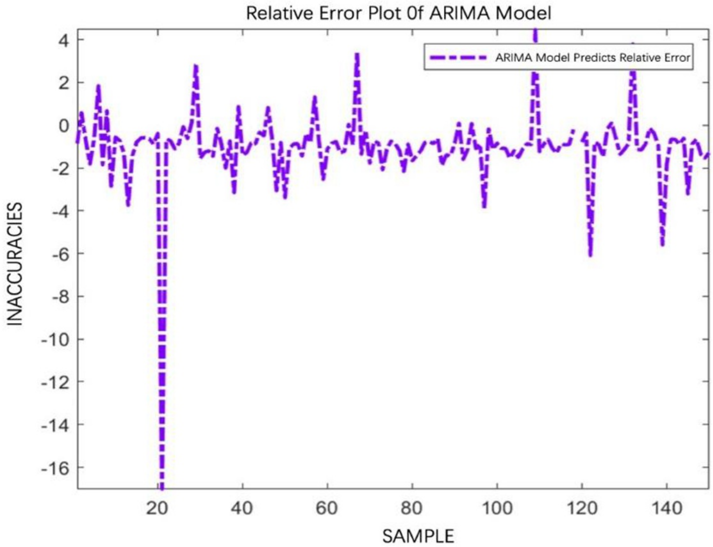


Fig. 4. Comparison figure of forecast RMSE for two stocks (top SP500; bottom CSI300)

In Fig. 5, in the relative error analysis of the ARIMA model predictions of the CSI 300 and the S&P 500, it can be clearly seen that the CSI 300 has a higher volatility in the error curve, which indicates that when applying the ARIMA model to predict the CSI 300, it is not as stable and accurate in its predictions as the S&P 500. Specifically, it shows more violent fluctuations in the prediction error graph, which means that the ARIMA model has some limitations in capturing the short-term fluctuations of the CSI 300 index.



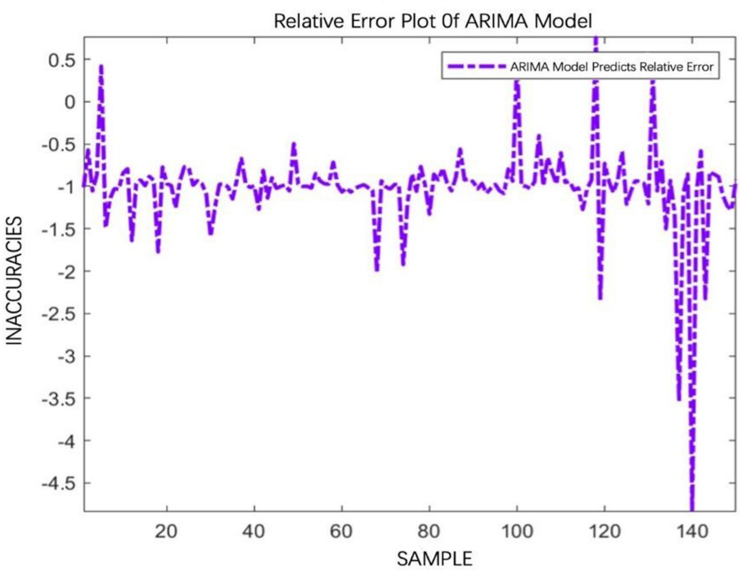


Fig. 5. Relative error comparison chart (top SP500; bottom CSI300)

On the contrary, the S&P 500 forecast curve reflects a smoother fluctuation state, with the forecast error mainly concentrated in a small range above and below -1, which indicates that the model has a higher stability and accuracy in forecasting the S&P 500. This is because this smaller error fluctuation indicates that the model is able to adapt better to the volatility characteristics of the S&P 500, thus providing more reliable forecasting results.

Fig. 6 shows that the prediction trend of the two curves is basically the same as their original curves in Fig. 1, and the curve of SP500 is gentler, while the fluctuation of CSI300 curve is bigger, which indicates that the model establishment and prediction are reasonable. Comprehensive analysis of the above values and images shows that the ARIMA model is more suitable for the SP500 index than the CSI300 index in the prediction work.

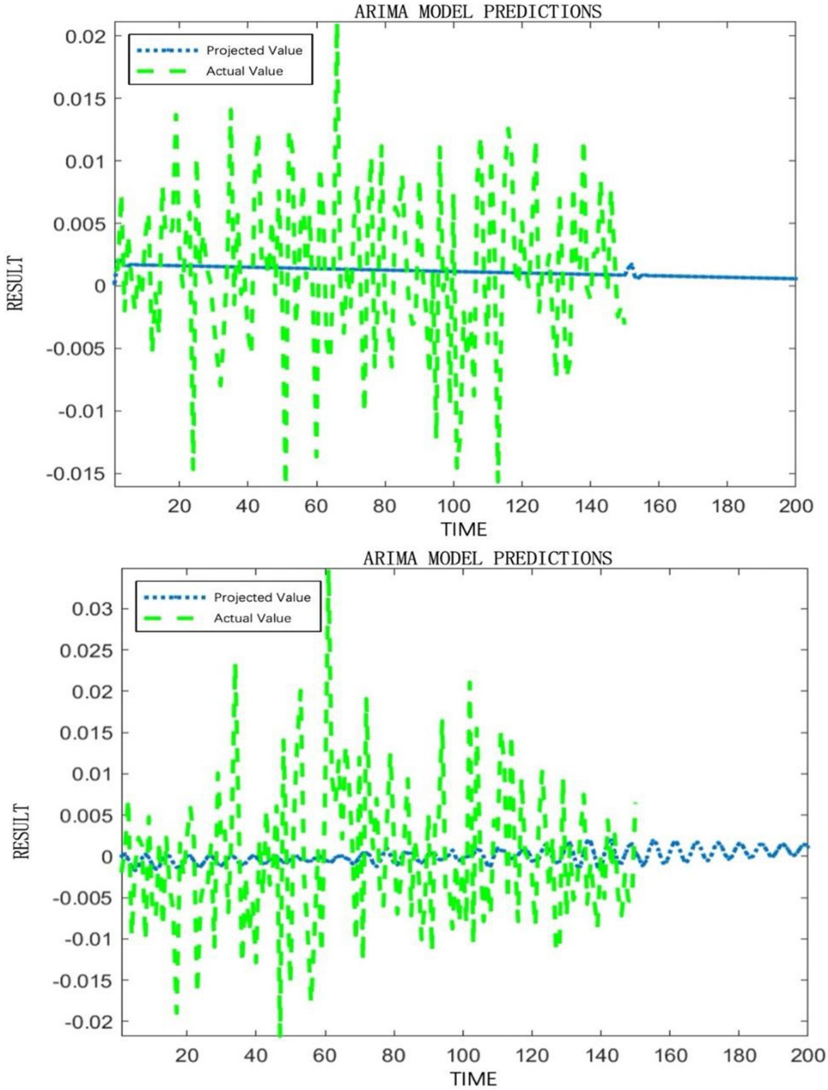


Fig. 6. Final forecast chart (top SP500; bottom CSI300)

5 Conclusion

From the above analyses, it can be seen that the forecasting of international capital markets has been one of the most critical research themes in the continuing progress of the financial industry.

This study firstly focuses on the widely used ARIMA model, and then selects two representative assets in the international stock market for forecasting analysis. After a comparative analysis, it is found that the accuracy of the model in forecasting the Standard & Poor's 500 Index (SP500) is better than that of the China Shanghai-Shenzhen 300 Index (CSI300), i.e., the ARIMA model is more suitable for forecasting the SP500 index.

This finding not only provides strong data support and suggestions for further research on these two stocks, but also provides a valuable reference for exploring the application of current popular forecasting tools in stock market forecasting in different economies. In conclusion, through this study, the effectiveness of ARIMA model in international stock market forecasting is further verified and new perspectives and methodological guidance are provided for future research.

References

1. M. F. Cheng, S. P. Gao. Multiscale stock prediction based on deep migration learning. Computer Engineering and Applications. 58, 249-259 (2022).
2. W. X. Yan. Application of machine learning based in stock prediction. Information Systems Engineering. 4, 40-43 (2024).
3. P. R. Junior, F. L. R. Salomon, E. de Oliveira Pamplona. ARIMA: An applied time series forecasting model for the Bovespa stock index. Applied Mathematics. 5, 3383 (2014).
4. Y. X. Wu, X. Wen. Short-term stock price forecasting based on ARIMA model. Statistics and Decision Making. 23, 83-86 (2016).
5. X. N. Li. Application of multiple linear regression and time series models in stock forecasting. Science and Technology Entrepreneurship Monthly. 32, 153-155 (2019).
6. G. S. Zhang, X. D. Zhang. A hybrid forecasting model for stock prices based on time correlation. Economic Issues. 9, 23-28 (2015).
7. S. Oh, H. Baek, J. Ahn. Predictive value of video-sharing behavior: sharing of movie trailers and box-office revenue. Internet Research. 27, 691-708 (2017).
8. X. Ding, Y. Zhang, T. Liu, et al. Deep Learning for Event-Driven Stock Prediction. Proceedings of the IJCAI 2015 paper exchange. 1-31 (2015).
9. X. X. Xu. Research on the law and prediction of stock price change based on ARIMA model. Journal of Economic Research. 19, 77-78 (2016).
10. N. Zhang. Research on stock trend forecasting based on time series and application of R language. Modern Business. 23, 112-113 (2016).