Proceedings of ICEMGD 2025 Symposium: Innovating in Management and Economic Development
DOI: 10.54254/2754-1169/2025.LH23934

Comparison of ARIMA Model and Holt Exponential Smoothing Method in Predicting the Stock Price of China Merchants Bank

Xunheng Huang

College of Mathematics and Statistics, Wuhan University, Wuhan, China
2022302011087@whu.edu.cn

Abstract: In recent years, more and more people have invested their funds in the stock market, hoping that the stocks they invest in can bring high returns. Investors, financial industry professionals, and others hope to predict changes in stock prices through modeling or other methods. By providing prediction results, investors can grasp the trend of the stock price changes, which is beneficial for avoiding risks and making better investment decisions. This article is to use different prediction models to predict the close price of China Merchants Bank, compare the accuracy of different models, and then find the most suitable prediction model for the data set in this article. The specific idea is to use the closing price data of stocks from 2023 to 2024 to establish ARIMA models and Holt exponential smoothing prediction models, and use both models to predict the closing price of the stock in January 2025. Select the model with better predictive performance by comparing the error between the predicted and actual values of two models, as well as the minimum mean square error (MSE). The calculation results indicate that the ARIMA (4,1,4) model has better predictive performance. The calculation results indicate that the ARIMA (4,1,4) model has better predictive performance.

Keywords: China Merchants Bank (CMB), ARIMA model, Holt exponential smoothing method.

1. Introduction

As a prominent financial instrument in capital markets, equity securities exhibit inherent volatility that has garnered significant scholarly and practitioner interest. While equity investments inherently contain return-generating potential, they simultaneously embody systematic and unsystematic risks consistent with modern portfolio theory. This risk-return duality necessitates the development of robust quantitative forecasting methodologies. Therefore, accurate prediction of stock prices helps investors make correct decisions and improve their investment returns.

The stock price series is essentially a time series, so many scholars have used time series prediction methods to predict it and achieved good prediction results. Guo and Wang predicted the stock price of Maotai based on grey theory and ARIMA model. From the analysis of the results, the GM-ARIMA regression model can fit and predict the open and close prices of stocks well and be applied to price prediction [1]. Li established ARIMA-ARCH model and neural network model to predict the close price and daily return of Hengrui Pharmaceutical stock. The results indicate that the former is

DOI: 10.54254/2754-1169/2025.LH23934

effective in short-term forecasting, while the latter performs better in long-term forecasting [2]. Li and Cheng established ARIMA model to predict the daily close prices of Zijin Mining and Vanke A, and the results also showed the feasibility of their short-term predictions [3]. Huang constructed an ARIMA (1,1,1) model to predict the closing price of Shenxin Technology Company's stock. By comparing the predicted value with the actual value, the conclusion was drawn that the static prediction effect of the model was good, while the dynamic prediction had a significant difference from the actual value [4]. Li and Xin conducted empirical analysis on the stock data of a well-known cosmetics company in the past decade by establishing an ARIMA model. The results showed that the ARIMA model has a relatively accurate trend prediction function [5]. Liang and Wang established ARIMA model and the Holt exponential smoothing prediction model for the close prices of the Shanghai Stock Exchange Index and the treasury bond Index. Through comparison, the ARIMA model is considered to have a better prediction effect, and it is used to predict, and analyze the impact of COVID-19 epidemic on the stock market [6]. Sun established ARIMA (2,1,2) model and logistic regression model to predict the Shanghai Composite Index. The results showed that after p months, the ARIMA model's predicted values gradually tended towards the average, which is not conducive to long-term prediction and requires consideration of model updates [7]. Li established a multiple linear regression and time series prediction model for the daily data of stock 00001 on the Shanghai Stock Exchange in 2014, and concluded that the two fit well and have certain practical significance [8]. Zhu used ARIMA (4,1,5) - ARCH (1) model and BP neural network model to predict the closing price of China Construction Bank stocks, and found that the BP model had better prediction performance than traditional time series models [9].

The above research covers various time series models and neural network models. However, the development of neural network models is not yet mature, so this article chooses traditional time series models for modeling. Apply ARIMA model and Holt index smoothing method separately to predict and analyze the close price of China Merchants Bank, and compare and select the model with better performance.

2. Theoretical analysis

2.1. Theoretical analysis ARIMA (p, d, q) model

ARIMA model, fully known as autoregressive integrated moving average model, was proposed by Box and Jenkins in the 1970s [1]. Also known as B-J model, it was a very famous time series prediction algorithm at that time, and it was also one of the traditional random time series models. The specific method of the model is to convert the non-stationary time series data into a stationary time series, and only regress the lag value of the dependent variable and the present value and lag value of the random error term [2]. A model with the following structure is called autoregressive integrated moving average model, abbreviated as ARIMA (p, d, q) model:

$$\nabla^d x_t - \lambda_1 \nabla^d x_{t-1} - \cdots - \lambda_p \nabla^d x_{t-p} = \theta_t - \mu_1 \theta_{t-1} - \cdots - \mu_q \theta_{t-q} \tag{1}$$

Where p is the order of the autoregressive process, d is the order of the difference, and q is the order of the moving average process [3].

After obtaining the observation value sequence, establish an ARIMA model through the following steps [4]:

(1) Stationary tests are performed on time series data. If the sequence is non-stationary, proceed to step (2); If the sequence is stationary, perform a white noise test. If the test indicates it is white noise, the process end; If it is not white noise, proceed to step (3).

(2) When the sequence is non-stationary, the d-order difference can be used to transform the non-stationary sequence into a stationary sequence.

DOI: 10.54254/2754-1169/2025.LH23934

(3) Determine the values of p and q based on the AIC criteria, and fit the ARIMA (p, d, q) model.

(4) The effectiveness of the model is demonstrated through residual sequence white noise testing. If the fitted model fails the test, proceed to step (4), adjust the values of p and q, and fit the model again.

(5) Using the ARIMA (p, d, q) fitting model, predict the values for the next few periods of the time series.

The modeling process of ARIMA model is shown in the figure 1:

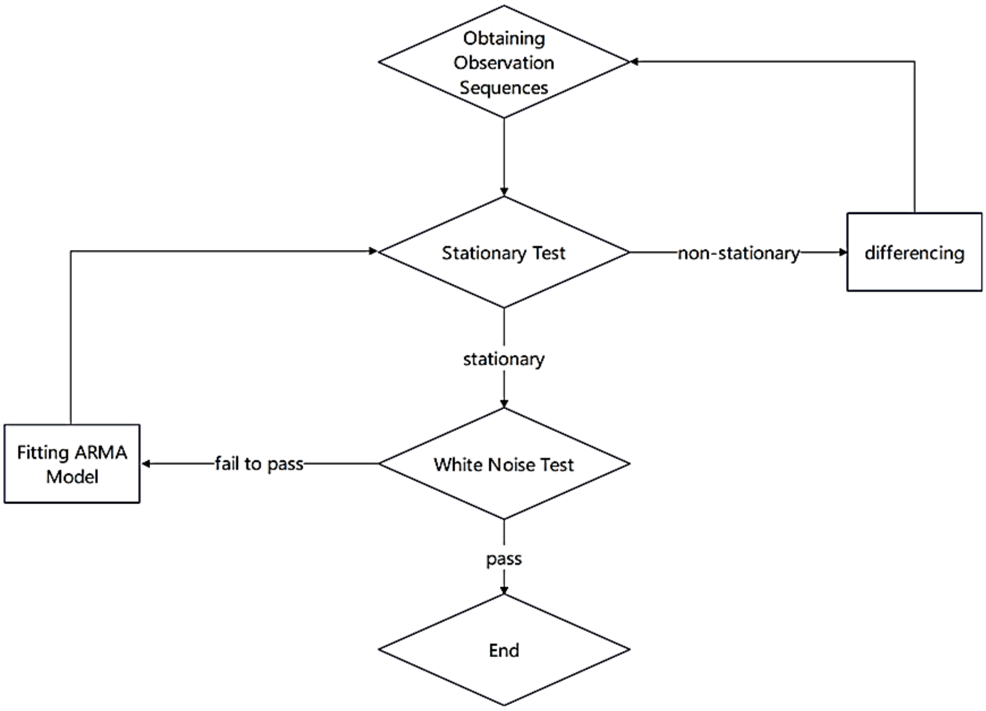

Figure 1: Modeling flow chart of ARIMA

ARIMA model has a simple structure and can be used as a reference for the trend prediction of time series data. However, the effect of ARIMA model in practical application depends on the quality of data, the characteristics of data, especially the stationary of time series data and the rationality of three parameters p, d and q [5].

2.2. Theoretical analysis ARIMA (p, d, q) model

The Holt exponential smoothing prediction model is a time series prediction model suitable for predicting sequences with linear trends but no seasonal variations. The Holt exponential smoothing prediction model includes two parameters: the smoothing coefficient α and the trend coefficient β, both of which have values ranging from 0 to 1. The smoothing coefficient controls the weight of historical data, while the trend coefficient controls the speed of trend change. By continuously adjusting these two parameters, the optimal prediction results can be obtained [10]. The prediction formula for the Holt two parameter exponential smoothing prediction model is

$$x_t = a_0 + bt + \varepsilon_t = a(t-1) + b(t) \tag{2}$$

$$\hat{a}(t) = \alpha x_t + (1-\alpha)\left[\hat{a}(t-1) + \hat{b}(t-1)\right] \tag{3}$$

$$\hat{b}(t) = (1-\beta)\hat{b}(t-1) + \beta[\hat{a}(t) - \hat{a}(t-1)] \tag{4}$$

DOI: 10.54254/2754-1169/2025.LH23934

Using Holt two parameter exponential smoothing method, the predicted value of the forward K period is [6]:

$$\hat{x}_{t+k} = \hat{a}(t) + \hat{b}(t)k, \forall k \geq 1 \tag{5}$$

3. ARIMA (p, d, q) model for predicting the stock price of CMB

3.1. Theoretical analysis ARIMA (p, d, q) model

Considering the length of the time series used for modeling, if the data is too short, there may not be sufficient information for prediction. If it is too long, there will be too much old information, resulting in excessive noise. ARIMA model belongs to predictive time series analysis. It uses the historical information of time series to predict the future value of the series. It is a data driven prediction method, so the data used for modeling cannot be too long [7]. So, this article primarily comprising the daily close prices of CMB stocks from January 3, 2023, to December 31, 2024, as the training set, totaling 484 close price data points. Additionally, close prices from January 2, 2025, to January 27, 2025—spanning 18 trading days—were collected as the validation set. The data collected in this study were exported from the Investing.com (https://cn.investing.com/). Since the historical close price data of CMB only includes trading days, weekends and holiday periods were treated as missing values and subsequently removed. The next step involves fitting time series models to the collected data for predictive analysis, aiming to identify a model with higher predictive accuracy.

3.2. ARIMA (p, d, q) model

3.2.1. Stationary test

(1) Time series plot analysis

Before establishing a predictive model, it is necessary to preprocess the data by checking whether the sequence has stationary and randomness. Only through tests of stationary and randomness can further research and analysis of the data be conducted [8]. So, using R to draw time series plot of daily close price of CMB stock. The horizontal axis represents sequentially ordered dates after excluding weekends and holidays, while the vertical axis denotes the daily close price (CNY) of CMB stocks.

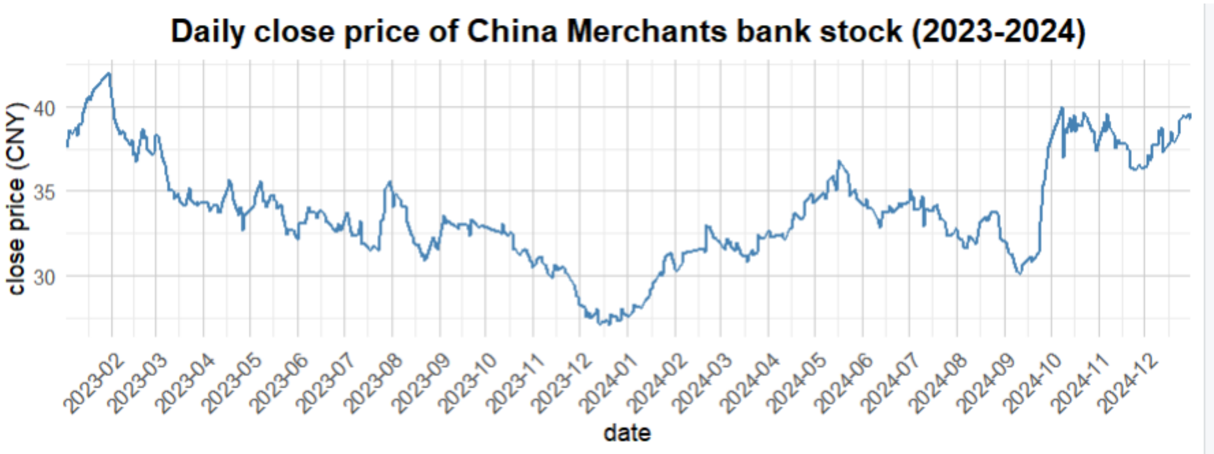

Figure 2: Daily close price of China merchants bank stock

DOI: 10.54254/2754-1169/2025.LH23934

The figure 2 illustrates that since February, 2023, close price experienced a general downward trend until Mid-December, 2023, followed by a general increase to October, 2024. Then it gradually stabilized.

The time series plot of stationary series should exhibit random fluctuations around a constant level because stationary time series are characterized by constant mean and variance. However, the presented chronograph clearly deviates from this pattern, displaying discernible trend components. These visual observations preliminarily suggest the non-stationary nature of the series

(2) ADF test

Although visual inspection of time series plot allows preliminary identification of non-stationary, such graphical methods carry significant subjectivity. To substantiate these empirical observations, formal unit root testing—specifically the augmented Dickey-Fuller (ADF) procedure—will be implemented for rigorous statistical verification. The table 1 shows that in three types of ADF tests, at the significance level of 0.05, the p-values exceed 0.05, leading to a failure to reject the null hypothesis. So, the close price of CMB stock is a non-stationary series, as the analysis of time series plot.

Table 1: P-value of three types ADF tests

Lag	No drift no end	With drift no end	With drift with end
1	0.546	0.281	0.551
2	0.539	0.300	0.574
3	0.545	0.274	0.544

(3) Differencing

To set ARIMA model, differencing transformations must be applied to the original dataset to induce stationary in the processed series. After first-order differencing, this article use R to draw time series plot.

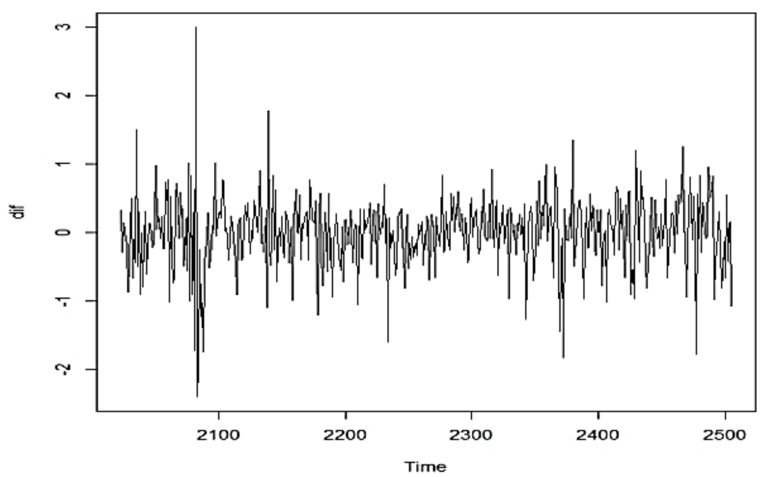

Figure 3: Time series plot of first-order differencing

After first-order differencing, sequence fluctuate around 0 (see figure 3). That means the first-order differenced series exhibits no statistically discernible trend component, satisfying the stationary prerequisite for time series modeling. But to receive a more accurate result, the ADF test should be conducted to the sequence.

DOI: 10.54254/2754-1169/2025.LH23934

Table 2: P-value of three types ADF tests of first- order differencing series

Lag	No drift no end	With drift no end	With drift with end
1	0.01	0.01	0.01
2	0.01	0.01	0.01
3	0.01	0.01	0.01

The table 2 illustrates that p-values are smaller than significance level 0.05. That means the sequence after differencing is transforming into a stationary sequence.

(4) White noise test

Furthermore, the white noise test should be performed on the differenced series. If there is no correlation among the sequence values, it implies that the sequence lacks memory, which means past behavior has no influence on future developments. The definition of such the sequence is white noise sequence. From a statistical analysis perspective, white noise sequences possess no analytical significance. Therefore, further analysis of the sequence is only meaningful when the differential sequence does not meet the conditions of a white noise sequence. So, to ascertain whether continued analysis of the first-order differencing stationary sequence is warranted, a white noise test needs to be conducted on it. Ljung-Box is selected here to implement the above process.

Table 3: Results of Ljung-Box test

Lag	X-squared	P-value
6	10.346	0.1108
12	20.695	0.05503

Table 3 shows that in the p-value corresponding to the 12th-order lagged LB statistic is significantly less than 0.05, which means the null hypothesis that the sequence is a white noise sequence is rejected. Consequently, the first-order differenced stationary sequence is determined to be a non-white noise sequence. Hence, an ARIMA (p, d, q) model will be established for this sequence.

3.2.2. Model order determination

Through the above tests, it can be concluded that the first-order differenced sequence is a stationary non-white noise sequence. Therefore, the author can draw the ACF and PACF diagram of the sequence, and determine the order of the sequence by ARIMA (p, d, q) model by comparing the tailing behavior and cut-off behavior of its auto-correlation coefficient and partial auto-correlation coefficient.

Table 4: ARMA model order specification matrix

Model	ACF	PACF
AR(p)	Tailing behavior	Order-p cut-off
MA(q)	Order-q cut-off	Tailing behavior
ARMA(p,q)	Tailing behavior	Tailing behavior

The order determination basis can refer to the table 4, and the ACF and PACF diagrams of the sequence can be drawn in R.

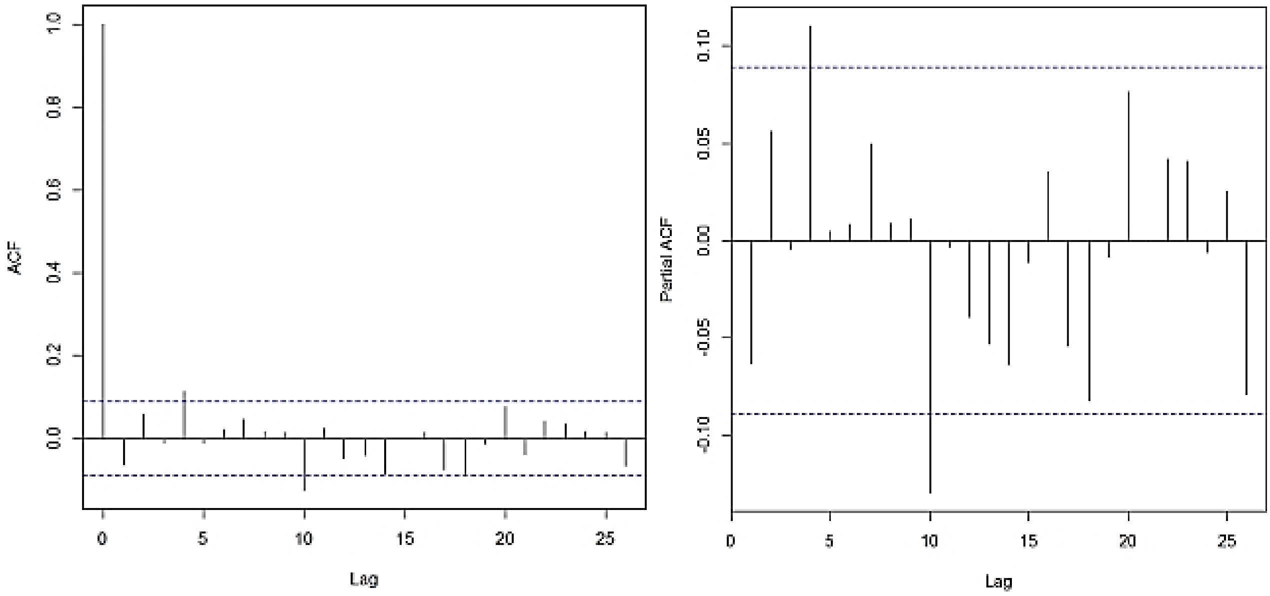

Figure 4: ACF&PACF plot of sequence

The figure 4 shows auto-correlation coefficient and partial auto-correlation coefficient are trailing at each delay orders. Consequently, direct determination of either the AR order p or MA order q from these plots proves infeasible. Besides, Judgments based on ACF and PACF plots often have biases and potential errors. Therefore, the author selects the AIC criterion as the model order - determination criterion. The AIC criterion, proposed by Akaike, a Japanese statistician in 1973. It represents a weighted function that balances fitting accuracy against the number of parameters. The AIC function is a weighted function of fitting accuracy and the number of parameters:

$$AIC = -2ln(L) + 2C \tag{6}$$

In function (6), L represents the maximum likelihood function value of the model, while C denotes the number of unknown parameters within the model. The model that minimizes the AIC function value is considered the optimal model. Thus, by gradually comparing the AIC function values of ARIMA models with different orders, this paper choose ARIMA (4, 1, 4) as the final fitting model.

3.2.3. Model significance test

Using the ARIMA (4, 1, 4) model to fit the sequence. After fitting, the significance test should be conducted on the model. The significance test of a model primarily aims to assess its validity, as the effectiveness of a model hinges largely on the adequacy of the information it extracts. An excellent fitting model ought to capture nearly all the pertinent sample information within the observed value sequence, implying that the fitting residuals will contain no further pertinent information, and consequently, the residual sequence should constitute a white noise sequence.

DOI: 10.54254/2754-1169/2025.LH23934

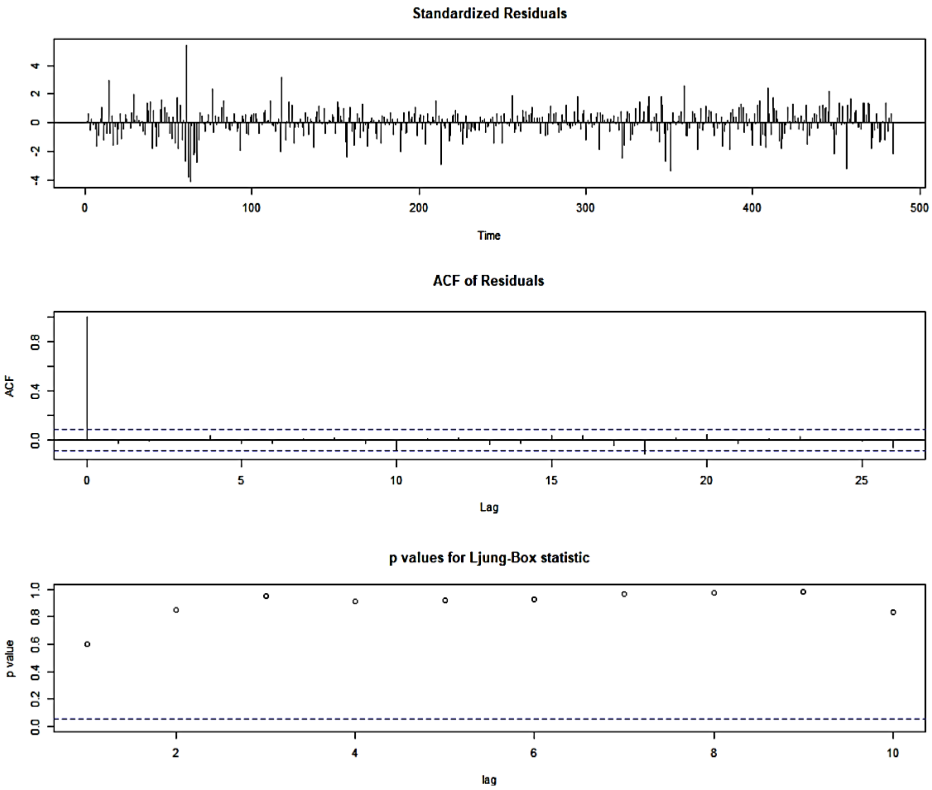

Figure 5: Model significance test chart

It is obvious that in figure 5, the p - values corresponding to all lag orders of the Ljung - Box test are greater than twice the standard deviation 0.05, which proves that the residual sequence of the fitting model can be regarded as a white - noise sequence. Therefore, the model is significantly effective.

3.2.4. Prediction of model

The so-called prediction is to estimate the future value of a sequence at a certain time by using the observed sample values of the sequence. This article uses the ARIMA (4,1,4) model to fit the sequence and plots the fitted and true values on one image.

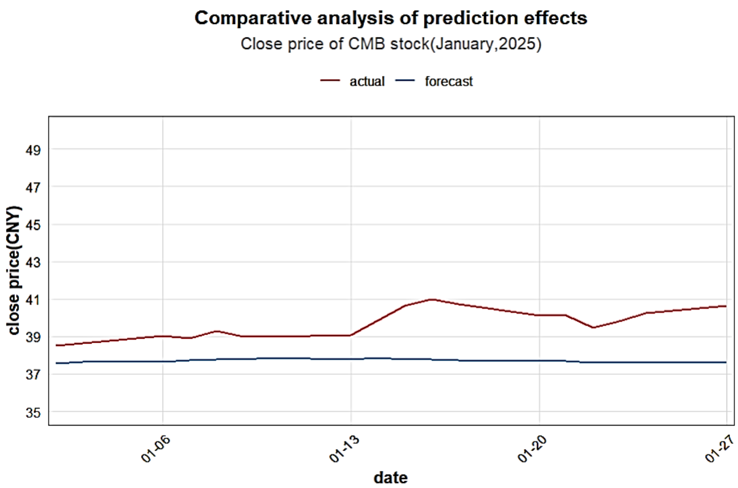

Figure 6: Comparative effect analysis of ARIMA model prediction

Based on the results presented in the figure 6, the model's fitted data initially succeeded in predicting the trend of stock close price movements. At the beginning of the prediction, the degree of outset between the actual value and the forecast value is less than 5%, but gradually the prediction error is increasing. As time progressed, the prediction error of the model increased progressively, suggesting that the model's long-term prediction performance is not notable.

4. Holt exponential smoothing method for predicting the close price of CMB

Through the time series analysis, this paper found that the series showed an overall upward trend since January 2024, but there was no obvious seasonality. Therefore, Holt two parameter exponential smoothing method is used to fit the series data, where R calculates the smoothing coefficient based on the optimal fitting principle. The result is that the value of α is 0.9454607, while that of β is 0.03536216.

Then drawing plot the fitted and true values on one image, and list the actual values and predicted values in a table format, and calculate the daily error.

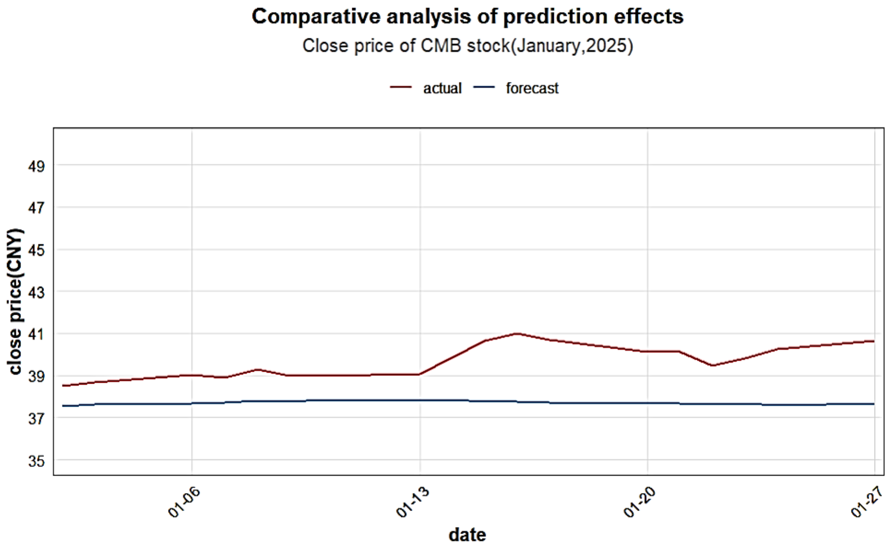

Figure 7: Comparative effect analysis of holt prediction

It is not difficult to see from the figure 7 that the model also shows good initial prediction effect and poor long-term prediction effect.

5. Further discussion

In order to compare the prediction effect of the two models, this article introduce the mean square error MSE as the comparison standard. The smaller the mean square error, the better the prediction effect of the model. The calculation formula is as follows:

$$MSE = \sum_{i=1}^{n}(x_i - \hat{x}_i)^2/N \tag{7}$$

In this formula, x_i is the actual value, $\hat{x}_i$ is the forecast value [9]. The table 5 gives the predicted values of the two models:

DOI: 10.54254/2754-1169/2025.LH23934

Table 5: Predicted values of two models

Date	Actual	Holt	ARIMA (4,1,4)
2025-01-02	38.51	37.63078	37.54039
2025-01-03	38.66	37.62207	37.63939
2025-01-06	39.05	37.61337	37.66283
2025-01-07	38.90	37.60467	37.71680
2025-01-08	39.28	37.59596	37.78785
2025-01-09	39.00	37.58726	37.77435
2025-01-10	38.98	37.57855	37.85739
2025-01-13	39.07	37.56985	37.79130
2025-01-14	39.85	37.56114	37.84781
2025-01-15	40.64	37.55244	37.76940
2025-01-16	40.99	37.54374	37.77556
2025-01-17	40.71	37.53503	37.72700
2025-01-20	40.12	37.52633	37.68357
2025-01-21	40.14	37.51762	37.68702
2025-01-22	39.45	37.50892	37.61755
2025-01-23	39.83	37.50022	37.66586
2025-01-24	40.26	37.49151	37.60440
2025-01-27	40.65	37.48281	37.66783
MSE		5.10349	3.96651

As evidenced by the comparative data presented in Table 5, both the ARIMA (4,1,4) model and Holt's exponential smoothing method demonstrate comparable predictive accuracy. However, quantitative analysis reveals that the ARIMA (4,1,4) configuration exhibits a statistically lower MSE compared to that of Holt method. This empirical comparison suggests that the ARIMA (4,1,4) model demonstrates superior forecasting performance for the temporal patterns inherent in this particular time series data set, as indicated by its enhanced error minimization capability.

6. Conclusion

This paper mainly models and forecasts the close price of CMB, and selects ARIMA (4,1,4) and Holt exponential smoothing method to model it respectively. In the third part, using ARIMA model to fit its close price, convert the original sequence into a stationary non-white noise sequence through the steps of stationary test and white noise test, and select the best p, d and q value for fitting based on AIC criteria. In the fourth part, using Holt two parameter exponential smoothing method, and then use R studio to give the smoothing coefficient, fit the series and predict it through continuous iteration and recursion. Finally, by comparing the MSE of the two models, it is concluded that ARIMA (4,1,4) has a better fitting effect on the close price.

The above two models have shown good accuracy in short-term prediction, but the following problems need to be solved in the follow-up study. Firstly, as time goes on, the error between the predicted value and the error value is getting larger and larger. This shows that the time series does not have good accuracy in the long-term prediction. Secondly, because the factors affecting the stock price are relatively complex and may be affected by national policies, the company's operating

DOI: 10.54254/2754-1169/2025.LH23934

conditions and some technical indicators in the stock, the deficiency of this paper is that only the technical indicators and the impact between before and after the stock price are considered. If more indicators are added, the prediction may be more accurate. Furthermore, in this paper, only two years data is selected as the training set of the model. From the perspective of experience, generally 5-10 years data is used as the training set to make the model have better prediction accuracy, but at the same time, increasing the length of time series may also make the model over fit and other problems, so the follow-up research can explore the most appropriate length of the training set. Finally, financial time series often exhibit properties such as heteroscedasticity and volatility clustering. Therefore, it is advisable to construct more complex time series models, such as using ARCH and GARCH models to address the heteroscedasticity of the series.

References

[1] *Guo, G. W., &Wang, S. H. (2023). Stock price prediction based on grey theory and ARIMA model. Journal of Henan University of Education (Natural Science Edition), 32 (02), 22-27*

[2] *Li, T. T. (2021). Stock price prediction based on time series analysis and neural network model. Northeastern University of Finance and Economics.*

[3] *Li, Y. J., & Cheng, Z. M. (2011). The application of time series model in stock price forecasting. Market modernization, (33), 61-63*

[4] *Huang, M. T. (2023). Empirical Study on Stock Price Prediction Based on ARIMA Model Neijiang Technology, 44 (03), 61-62*

[5] *Li, L. L., & Xin, H. (2024). The application of time series model ARIMA in data analysis. Fujian computer, 40 (04), 25-29.*

[6] *Liang, S. J., & Wang, Y. (2024). Research on the impact of the new crown epidemic on China's stock and bond market -- An Empirical Analysis Based on ARIMA model. China Securities and futures, (04), 59-66+89.*

[7] *Sun, X. Q. (2016). Application of time series model in stock price forecasting. Yunnan University.*

[8] *Li, X. N. (2019). The application of multiple linear regression and time series models in stock prediction. Technology Entrepreneurship Monthly, 32 (02), 153-155*

[9] *Zhu, X. G. (2020). Research on stock price forecasting based on ARIMA-ARCH model and BP neural network. Suzhou University.*

[10] *Wang, Y. (2015). Time series analysis: Based on R. China Renmin University Press.*