ISSN: 2688-9323

Short-term Forecast of Stock Price based on ARMA Model

-- Take SAIC Stock as an Example

Xuanjing Ti, Zhiqing Zhang

School of Statistics and Applied Mathematics, Anhui University of Finance and Economics, Bengbu 233000, China

Abstract

Time series is a series formed by sorting the statistical values in the order of time. It can be used to analyze the historical data of a statistic in the past and predict the future data. It has a wide range of applications in the fields of economy and finance. In this paper, R software is used to select the closing price of SAIC stock (600104) from January 25, 2022 to January 13, 2023 as the time series analysis data, and the stock price is fitted and predicted by constructing ARMA model. The empirical results show that the ARMA model is effective in the short-term prediction of stock price, but in the long run, the stock price is affected by many factors, so it is no longer suitable to use ARMA model to forecast.

Keywords

Time Series; ARMA Model; Share Price; Short-term Forecast.

1. Introduction

With the continuous improvement of the financial market system, the stock has played an increasingly important role in the development of the national economy. People's traditional investment concept and risk tolerance have also changed, and more and more people began to pay attention to and participate in the stock market investment. However, the high return does not make the shareholders feel comfortable. Along with the high return, there are also high risks, which is because the price change trend of the stock market is affected by many factors [1]. On the surface, the stock market lacks certain regularity, and its investors also have certain particularity in structure - the psychological state of investors can have a direct impact on the trading behavior of the stock market and lead to stock price fluctuations. For investors, the more accurate the stock price prediction is, the more risks can be avoided and the greater returns can be obtained. As for national economy, the stock market is related to the property safety of state shareholders, for effective control and management of risks, can promote endogenous impetus of the national economy and form a stock market with perfect allocation of resources, so as to be adapted with the economic development of our country [3].

At present, the stock market mostly adopts non-random analysis methods, such as fundamental analysis. But these methods can not accurately reflect the changes in stock prices. With the continuous progress of empirical research methods, mathematical theory research and data mining have been developed rapidly. As one of the important tools of econometrics research, time series analysis is widely used to predict the trend of stock price and trading volume. Since stock prices are obviously uncertain, the formation of stock prices is a random process [4]. In other words, the stock price at any fixed moment is regarded as a random variable, and time series models such as ARMA, GARCH and ARCH can be used for financial prediction [2]. Among them, ARMA model is the most commonly used model to fit stationary time series at present. In forecasting, ARMA model not only considers the dependence of financial market indicators in time, but also considers the interference of random fluctuations, which can accurately predict its short-term change trend [5].

In recent years, ARMA model has also been widely concerned by many industries. It is mainly used in the economic field, such as predicting the sales volume and market size with seasonal changes, predicting the stock market, predicting the trend of GDP, and predicting the deterioration trend of mechanical properties, etc., and has produced satisfactory results.

Based on this, this paper collects the daily closing price of SAIC from 2022.1.25 to 2023.1.13, constructs an ARMA model to fit and forecast the stock price, and finally draws a conclusion.

2. Model Introduction

2.1. AR(p) Model

The AR part represents the linear combination of the current value of the time series with p values in the past.

If a linear process excluding the mean and deterministic components can be expressed as:

$$X_t = \emptyset_1 X_{t-1} + \emptyset_2 X_{t-2} + \cdots + \emptyset_3 X_{t-3} + \emptyset_p X_{t-p} + u_t \tag{1}$$

Among them:

$\emptyset_i$, i=1,...p is an autoregressive parameter;

u_t is a white noise process;

Then x_t is called the P-order autoregressive process, represented by AR(p). Is the weighted sum and addition of its p lag variables.

2.2. MA(q) Model

The MA part represents the linear combination of the current value of the time series with q lag errors in the past.

If a linear stochastic process excluding the mean and deterministic components can be expressed as follows:

$$X_t = u_t + \theta_1 u_{t-1} + \cdots + \theta_q u_{t-q} \tag{2}$$

Among them:

θ_1,...θ_q is the moving average parameter;

u_t is a white noise process;

The above formula is called the Q-order moving average process and is denoted as MA(q). It is called a "moving average" because $x_t$is constructed by the weighted sum of q+1 u_t and u_t lag terms. "Move" refers to the change in t, and "average" refers to the weighted sum.

2.3. ARMA(p,q) Model

The autoregressive moving average model, also known as ARMA model, was proposed by the American statistician Box and the British statistician Jenkins. It is an important method to study time series by modeling stationary series and "mixing" based on the autoregressive model and moving average model. It has good predictive performance and explanatory ability.

The random process composed of autoregressive and moving average is called the autoregressive moving average process and is denoted as ARMA(p,q), where p and q respectively represent the maximum order of the autoregressive and moving average parts [10]. The general expression for ARMA(p,q) is [6]:

$$X_t = \emptyset_1 X_{t-1} + \emptyset_2 X_{t-2} + \cdots + \emptyset_p X_{t-p} + u_t + \theta_1 u_{t-1} + \theta_2 u_{t-2} \cdots + \theta_q u_{t-q} \tag{3}$$

ISSN: 2688-9323

To wit:

$$\emptyset(L)x_t = \theta(L)u_t \tag{4}$$

Where and represent the characteristic polynomials of order p and q of L respectively.

3. Modeling steps

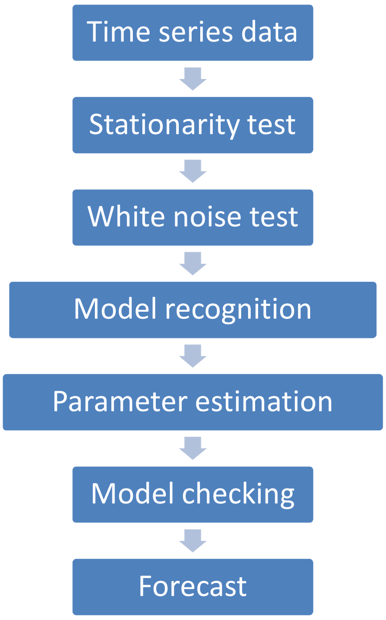

Figure 1. Modeling flow chart

3.1. Stationarity Test

Stationarity requires the fitting curve of the sample time series to maintain its current form in inertia for a period of time in the future, that is to say, a time series is said to be stationary if there is no systematic change in the mean (no trend), no systematic change in the variance, and strict elimination of periodic changes. When we find that the analyzed time series is not stationary, we need to carry out the difference operation on the time series, and then analyze the stationarity after the difference until the series is found to be stationary. In the process of actual operation, the phenomenon of excessive difference should be prevented.

There are several methods to test the stationarity of sequences: time series graph test, autocorrelation graph test and unit root test. Among them, the time series graph test is to see whether the sequence value of a time series graph always fluctuates randomly around a constant, and the range of fluctuations is bounded. If it meets the requirement, the sequence can be considered stable, and vice versa. The autocorrelation test is to see whether the autocorrelation coefficient tends to zero quickly with the increase of the number of delay periods k and fluctuates randomly around zero. If so, it can be considered stable, and vice versa. The unit root test is to check whether the sequence has unit roots. If there are unit roots, it is considered that the sequence is not stable, and vice versa.

3.2. White Noise Test

The test method often used is called Ljung-Box test (LB test), which is a test method used to test whether it is white noise. Based on the autocorrelation coefficient and the number of samples, this method constructs a chi-square distribution statistic for testing [9].

ISSN: 2688-9323

3.3. Model Recognition

If we want to use ARMA model to model time series, we must first determine the order of AR and MA parts of the model (p,q), so we need to find a measurement tool to determine the best order. The determination method is shown in Table 1.

Table 1. Methods for determining order

Model	ACF	PACF
AR(p)	Trailing	p order back truncation
MA(q)	q order back truncated	Trailing
ARMA(p,q)	Trailing	Trailing

Method of model estimation and establishment:

Determine the values of p and q. When we build the autoregressive model, we need to determine the values of p and q in order to get the optimal model structure. The order here can be roughly determined by the autocorrelation coefficient ACF and the partial autocorrelation coefficient PACF, that is, the correlation graph and the partial autocorrelation graph.

In the process of order determination, if the ACF and PACF of the sequence are not very clear, we can use other models to determine the order, including AIC and BIC information criteria. AIC is an indicator for model selection, and the fitting degree and simplicity of the model are considered at the same time. BIC is an improvement of AIC. In general, when multiple models are established, we generally choose the model with smaller AIC or BIC as the optimal model, because it can better fit the time series under the condition that the model remains equally simple.

3.4. Model Test

The estimated results are diagnosed and tested to find out whether the selected model is suitable. There are many criteria for testing, among which the residual sequence of the model must pass the Q test. It means that any two data sets can be compared to determine whether to obey the same distribution, calculate the quantile of each distribution, one data set corresponds to the x axis, the other corresponds to the y axis, by doing a 45 degree reference line, observe whether the data point falls near this reference line. If it falls near the reference line, the two data can be considered to come from the same distribution, and vice versa.

4. Empirical Analysis and Prediction

4.1. Data Source

This paper selects 236 samples through Choice financial terminal, including the closing price of SAIC Motor (600104) every day from 2022.1.25 to 2023.1.13. R software is used to model and analyze this time series, and forecast the closing price of its stock in the next three days.

4.2. Stationarity Test

Two methods can be used to test the stationarity of time series: one is the visual method, which is evaluated by the naked eye; One is a statistical method that passes the unit root test.

As shown in Figure 2, a time sequence diagram is drawn for the original data. It can be intuitively seen from the time sequence diagram that it has different mean values in different time periods, so it can be preliminarily judged that the series is unstable.

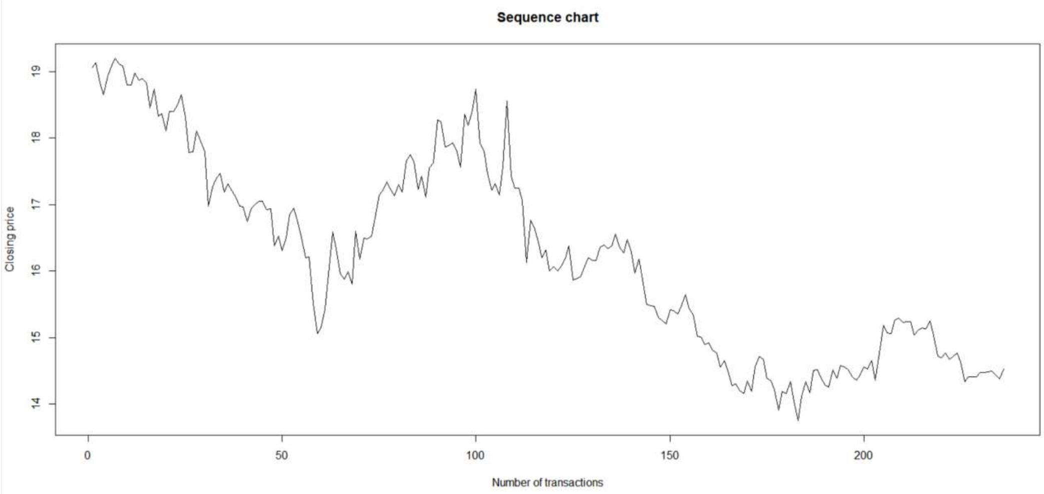

Figure 2. Closing time sequence diagram of SAIC

In order to further confirm the stationarity of the original data, ADF test was performed on it, and the result was shown in Table 2. It can be seen that the test statistic value was -1.81, which was greater than the critical value under the significance level of 1%, 5% and 10%, and the P-value was greater than 0.05. Therefore, the original hypothesis was accepted and the series was considered non-stationary. (The null assumption of the unit root test is that there is a unit root).

Table 2. ADF test of raw data

The Dickey-Fuller value	-1.814284240890034
The P value	0.3733933788460362
Dickey-Fuller value of 1% significance level	-3.458854867412691
Dickey-Fuller value of 5% significance level	-2.8740800599399323
Dickey-Fuller value of 10% significance level	-2.573453223097503

Since the original sequence is not stationary, if ARMA model needs to be established for analysis and prediction, this unstable feature must be eliminated, so the difference on the basis of the original sequence is tried. The next difference operation is performed on the data, and several differencing orders are judged to be a stationary sequence. By executing the command ndiffs (), d is 1.

The sequence diagram after first-order difference is shown in Figure 3. It can be seen that the sequence has no obvious periodicity, and it is considered that the sequence after first-order difference is a stationary sequence.

The ADF test [7] was further performed on the data after first-order difference, and the result was shown in Table 3. It can be seen that the test statistic value is -9.538045149848188, which is less than the critical value under the significance level of 1%, 5% and 10%, and the P-value is less than 0.05. Therefore, the null hypothesis is rejected, and the series is considered stable after difference and can be used to establish the model. (The null assumption of the unit root test is that there is a unit root).

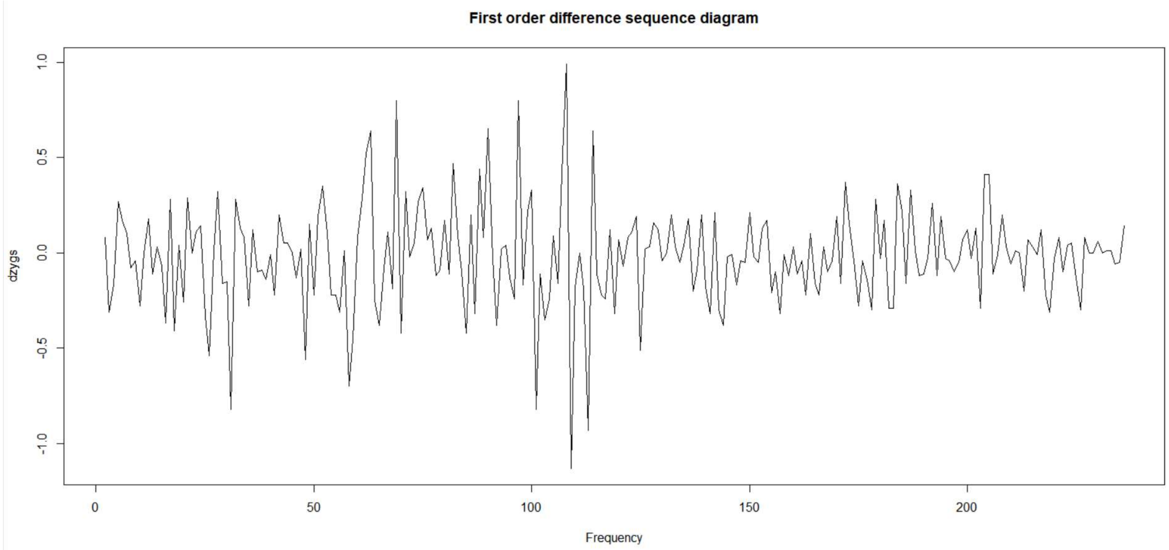

Figure 3. Time sequence diagram after difference

Table 3. ADF test after difference

The Dickey-Fuller value	-9.538045149848188
The P value	2.7743994979180224e-16
Dickey-Fuller value of 1% significance level	-3.458854867412691
Dickey-Fuller value of 5% significance level	-2.8740800599399323
Dickey-Fuller value of 10% significance level	-2.573453223097503

The autocorrelation and partial autocorrelation graphs after first-order difference are shown in Figure 4 and Figure 5. It can be seen that the autocorrelation coefficient decays rapidly and tends to zero, and fluctuates randomly around zero, so it is considered that the series after difference is stable.

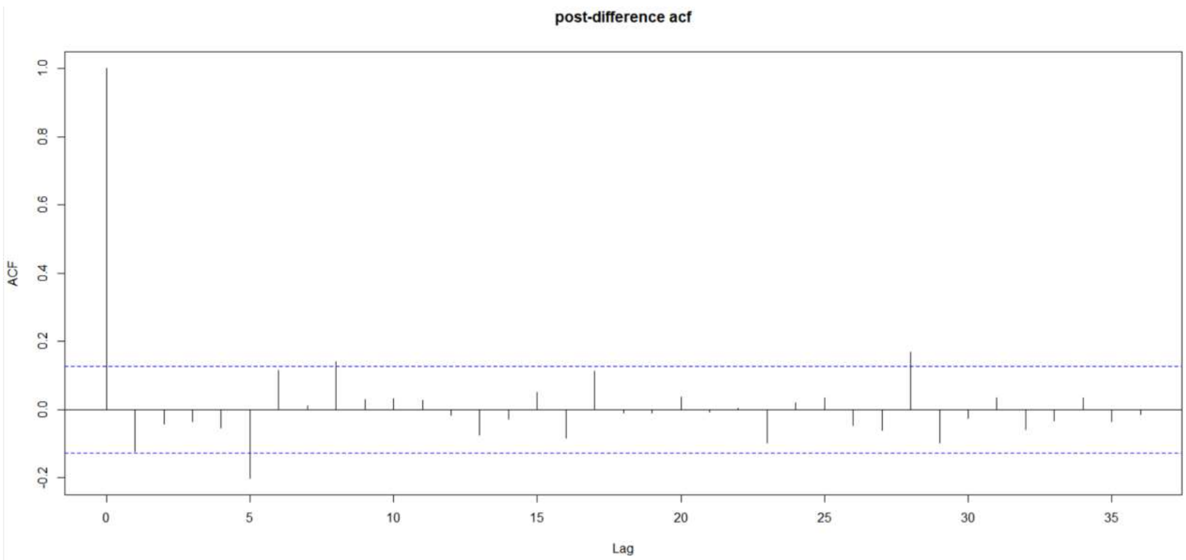

Figure 4. ACF diagram after difference

ISSN: 2688-9323

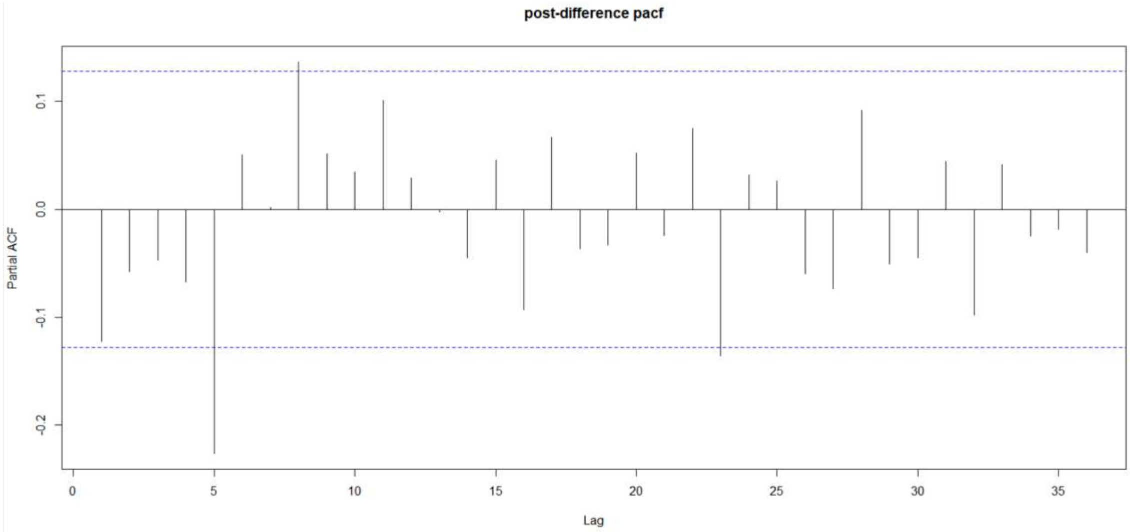

Figure 5. PACF diagram after difference

4.3. White Noise Test

Table 4. Test results of white noise

X-squared	3.4417
df	1
p-value	0.03357

After completing the parameter estimation operation, the residual sequence of the model needs to be tested. This is because if a residual sequence is white noise, it means that the useful information in the residual sequence has been extracted; otherwise, it means that there is still some unextracted information in the residual sequence, and the model needs to be further improved [8]. The white noise test was conducted on the data after difference, and the result was shown in Table 4. It can be seen that the p value is less than 0.05, the null hypothesis that the sequence is white noise can be rejected, and the sequence after difference is considered non-white noise, which can be analyzed and predicted in the next step.

4.4. Model Recognition

Arima model parameters were automatically selected through the auto arima () function, and the optimal model was automatically fitted as ARIMA (0,1,1). The fitting results and detailed information of the model were shown in Table 5.

Table 5. Model fitting results and details

Best model:ARIMA(0,1,1)
Sigma^2=0.06987
AIC=45.06
BIC=52.01

4.5. Model Test

The model is tested by checking the normality of the residuals, and if the model is correct, the quantile-quantile plot should have a straight line through numerous points. As shown in Figure 6, the residual QQ diagram is drawn. It can be seen that except for a few points that deviate from the straight line, the rest points fall roughly near the reference line, so it can be considered that the residual is approximately subject to normal distribution, thus verifying the accuracy of the model.

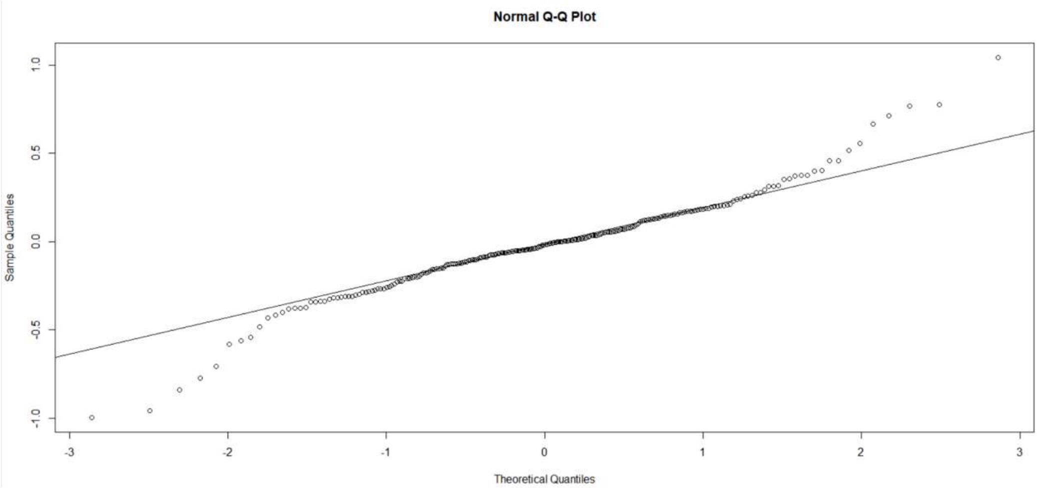

Figure 6. Residual QQ diagram

After parameter estimation, white noise detection is also needed for residual to confirm whether the established ARIMA model is suitable. Through the operation of residual white noise test, the result is shown in Table 6, it can be seen that p is greater than 0.05, so at the significance level of 0.05, the null hypothesis is not rejected, that is, the residual sequence is considered as white noise sequence, and the model is significant.

Table 6. Residual white noise test results

X-squared	6.6711e-05
df	1
p-value	0.9935

4.6. Model Prediction

Since the established ARIMA (0,1,1) model has passed the white noise detection, it can predict the future closing price of SAIC stock. The forecast timing diagram is shown in Figure 7. It can be seen that the predicted value is basically horizontal, indicating that the effect is not particularly good. It can be seen that although the ARMA model can predict the stock turnover in the short term, the result is still uncertain.

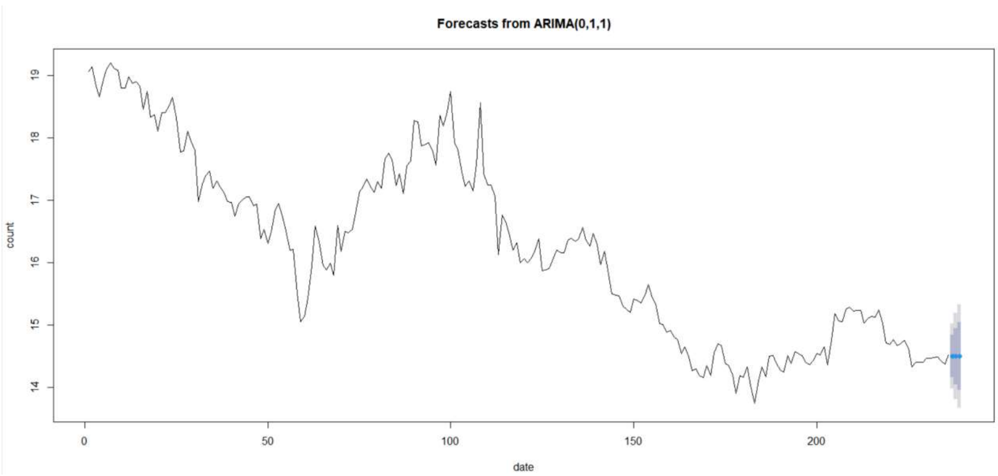

Figure 7. Forecast time series diagram

We compared the real value with the predicted value, and the result is shown in Table 7.

Table 7. Comparison between predicted and true values

Trading day	Predicted value	Actual closing price	Error(%)
2023.1.16	14.503	14.70	1.3%
2023.1.17	14.503	14.73	1.3%
2023.1.18	14.503	14.72	1.3%

As can be seen from the comparison results in Table 7, the error is controlled within the range of 5%, and the prediction effect is good.

5. Conclusion

This paper selects the closing price of SAIC stock as the time series data, constructs the ARMA model to fit, and predicts its closing price in the next three days.

To sum up, ARMA model is indeed an effective tool for analyzing and forecasting time series, which can be applied in the field of economic and financial investment and provide scientific reference value for short-term investment activities of small and medium investors. However, the time series of stock prices is non-stationary, so a lot of important information may be lost after the difference. Moreover, the long-term trend of stock prices is greatly affected by external shocks, such as changes in international trends, changes in domestic policies and other uncertain factors will interfere with the change of stock prices, while the short-term forecast is less affected by disturbing factors. Therefore, the ARMA model is only suitable for short-term prediction, and cannot accurately fit the long-term trend of stock prices. To analyze long-term changes in predictors, more precise and detailed models are needed.

References

[1] L.Q. Pan,H.F He and L.J QIAO:Empirical analysis of Spring Festival effect in Shenzhen Stock market based on ARMA model,Financial Economics,(2017) No.4,p.129-131.

[2] J. Lin,J.M. Zhu and F.Y. Chen:Analysis and prediction of trading volume of Qingdao Haier based on ARMA model,Journal of Natural Science,Harbin Norma University,Vol.32 (2016) No.6,p.6-10.

[3] Q. Zhong:Stock Price analysis and prediction based on ARMA-GARCH model, China Market,(2017) No.1,p.68-69,75.

[4] C.C. Guo,J.J GuoX. Liu and D.F Zhang:Empirical research on Yunnan characteristic stocks based on ARMA (p,q) model,Journal of Economic Research,(2015) No.14,p.233-234 + 247.

[5] Z.X. MA:Application of ARMA Model in short-term stock forecast,Guangzhou Quality Supervision Review,(2020)No.5,p.152-153. (in Chinese).

[6] X.D. ZHANG:Fundamentals of Econometrics(Nankai University Press,China 2014).

[7] K. Meng, L. Li:Empirical analysis of stock price prediction based on ARMA model, Journal of Hebei North University: Natural Science Edition,Vol.32 (2016) No.5,p.55-60. (in Chinese).

[8] C.M. Xu,H. Fang:Empirical research on stock price prediction using ARMA Model, Economic Research Guide,(2019) No.31,p.77-82.

[9] L.Y. Lin,X.F. Chen and D.F. Zhang:Application of ARMA model in stock,Economic Research Guide, Vol.376 (2018) No.26,p.151-153. (in Chinese).

[10] X.Z. Wu,M. Liu:Applied Time Series Analysis: Accompanied by R software(China Machine Press, China 2014).