Prediction of China CITIC Bank's stock price based on ARIMA model

Shengnan Guo
Unit：School of Economics, Anhui University, Hefei, Anhui, China

IJASR 2021
VOLUME 4
ISSUE 3 MAY – JUNE **ISSN: 2581-7876**

Abstract: The stock price series is a set of random variables that depend on time changes, which can be approximated by the ARIMA model. This article uses the closing price of China CITIC Bank's stock from 2018 to 2020 to predict the stock price of China CITIC Bank. The results show that the error between the predicted price and the actual price of the stock in the next three working days will not exceed 0.5%. It can be seen that the model constructed in this article is relatively accurate and can well reflect the changing law of stock price sequence, which provides a useful reference for investors to allocate funds reasonably.

Keywords: ARIMA model; stock price prediction

1 Introduction

In 1984, China's first public offering of stocks heralded that my country's financial market ushered in reform and opening up. After entering the 21st century, stocks are an indispensable part of the market economy. With the continuous development of the stock market, people have become accustomed to buying Stocks are used as an important investment tool. The stock price forecast can give certain opinions to companies or investors when investing. For investors, being able to predict the future development trend of stocks is of profound significance to their investment and financial planning. With the prosperity of the stock market, stockholders continue to seek various methods to judge and choose the best investment portfolio in pursuit of maximizing their own returns.

In recent years, a number of social platforms around stock price predictions have emerged at home and abroad. Netizens are free to publish their own experience in predicting stock price trends, and scholars are also actively researching more accurate and effective stock price prediction algorithms.

2 Model introduction

ARIMA model (Auto Regressive Integrated Moving Average Model), also known as differential Auto Regressive Moving Average Model, Auto Regressive Modeland Moving Average Model. The specific formula of the Auto Regressive Model is as follows:

$$y_t = \mu + \sum_{i=1}^{p} \gamma_i \, y_{t-1} + \theta t \quad (1)$$

Among them, $y_t$represents the current value,$p$is the autoregressive order, $\gamma_i$is the autocorrelation coefficient, and the error termθt is independent and identically distributed. The autoregressive model describes the relationship between the current value of the sequence and the historical value, that is, the historical data of the independent variable is used for the future to make predictions, and the autoregressive model requires stationarity sequences as training data.

The structure of the moving average model MA is expressed by formula as follows:

$$y_t = \mu + \sum_{i=1}^{q} \varepsilon_i \, \theta_{t-1} + \theta t \quad (2)$$

For the Moving Average Model, it mainly deals with the error term in the autoregressive model. The error term is accumulated, and the moving average can better predict the random fluctuations in the error term. The input of the ARIMA model should be non-stationary data, and Irepresents the difference. Then the non-stationary data is differentiated and converted into stationary data. The three important parameters in the ARIMA model are p、q、d. p is the autoregressive order, q is the number of moving average terms, and d is the number of differences when non-stationary data is transformed into stationary data.

It can be understood from the relevant literature that the ARIMA model has the following properties when selecting the autocorrelation coefficient and the partial autocorrelation coefficient, as shown in Table 1.

Table 1 ARIMA model ordering properties

Model	ACF	PACF
AR（p）	Attenuation tends to 0	P-order post-censoring
MA（q）	q-order after censoring	Attenuation tends to 0
ARMA（p，q）	Attenuation tends to zero after q-th order	Attenuation tends to zero after p-th order

When estimating the parameters of the ARIMA model, the above properties can be combined. When p and q meet the above description in the ACF diagram and the PACF diagram, it can be considered that this parameter is more suitable for the model, and the truncation indicates that it falls within the confidence interval.

3 Model establishment and testing

3.1 Stationarity test

The stationarity test of the data is the prerequisite for establishing the ARIMA condition. If it is not met, the stationarity of the time series data needs to be adjusted and corrected. There are two commonly used test methods: one is the graph test method, that is, draw the autocorrelation graph of time series data, and judge the relationship between the autocorrelation coefficient and the confidence interval to judge the stationarity, when the autocorrelation coefficients fall into the selected confidence.The data is in a stationary state; the second is the hypothesis test method, which is the unit root test. This test can determine whether there is a unit root in the given sequence. If there is a unit root, then the given sequence is in an unstable state. If it does not exist, It cannot be stated directly, and further testing is needed. This article will use the unit root test method to determine whether there is a unit root in the closing price sequence of China CITIC Bank's stock, so as to determine whether the time sequence is stationary.

Taking days as a unit length, select data with a time series length of 363, and draw an initial time series trend chart, as shown in Figure 1:

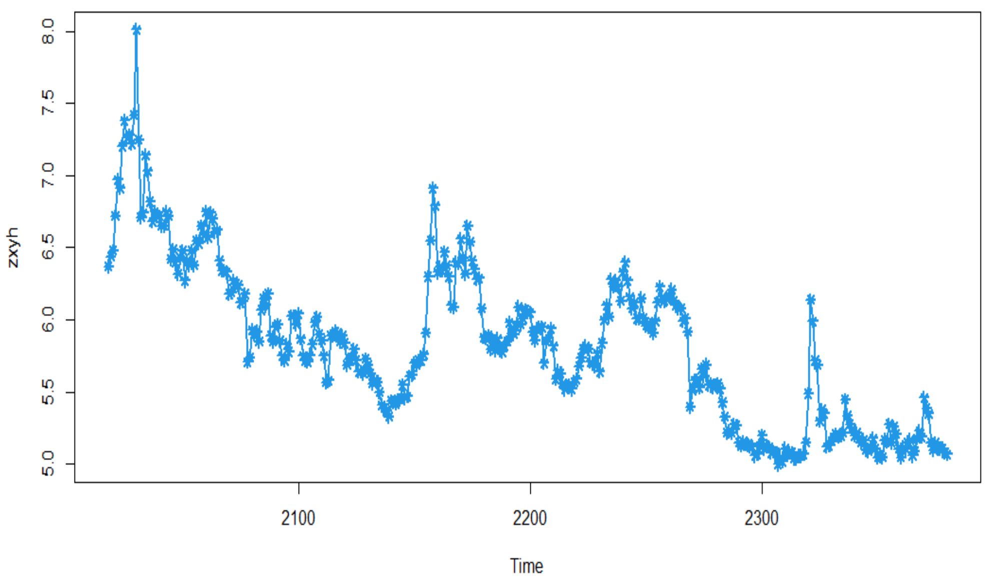

Figure 1 Initial sequence diagram of the closing price of China CITIC Bank stock

Observing Figure 1 we can find that the time sequence diagram of the closing price of China CITIC Bank stock shows up and down fluctuations, so it is preliminarily judged that the sequence is a non-stationary time series. Through the unit root test process, the P value calculated for the ADF test is 0.0382, which is greater than 0.01 of the significance level. Therefore, the null hypothesis is not rejected, so a judgment can be made that the sequence is in a non-stationary state, which is consistent with the initial judgment. This paper uses the ADF test method to test the stationarity of the data, and the test results are shown in Table 2.

Table 2 ADF inspection value

		t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic		-1.850502	0.08968
Test critical values	1%level	-3.191928	
	5%level	-2.888411	
	10%level	-2.581176	

It can be seen from the ADF test result that the ADF test statistic P value of variable Y is 0.08968>0.05, so the null hypothesis is accepted at the 5% significance level, that is, the original data is a non-stationary series, and the original data is first-order Difference, get the data after the first-order difference, make a sequence diagram of the data after the first-order difference, and check the stationarity again, the result is shown in the figure below

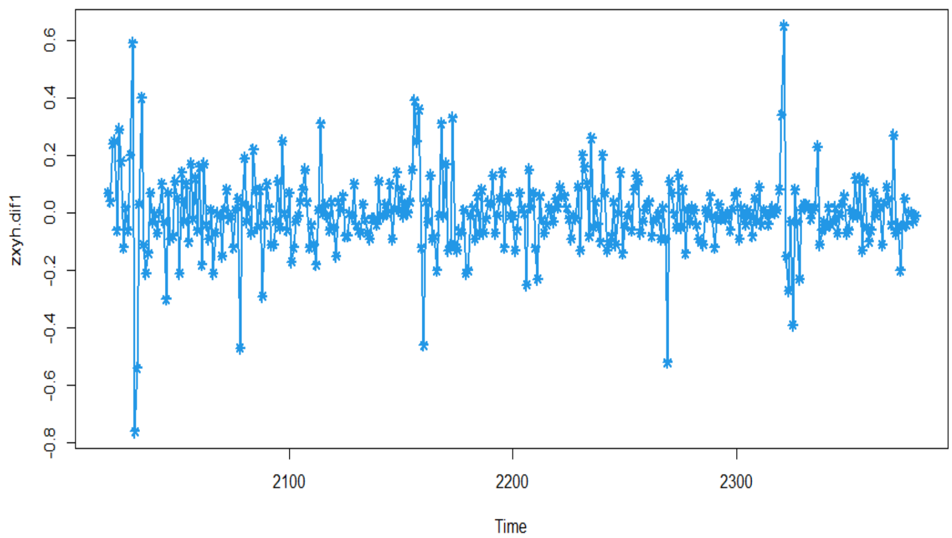

Figure 2 Timing diagram after first-order difference

Table 3 ADF test value after difference

		t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic		-1.850502	0.01
Test critical values	1%level	-3.191928	
	5%level	-2.888411	
	10%level	-2.581176	

The test results show that the corresponding P value is 0.01, which is much less than 0.05. Therefore, the null hypothesis is rejected at the 5% significance level, that is, the data after a difference is a stationary series, which meets the stationarity condition for establishing the ARIMA model. Then, the Box-Ljung white noise test is performed on the stationary sequence after the first-order difference. The test result shows that the p value is 0.02018, which is less than 0.05. Therefore, the null hypothesis is rejected, that is, the data after the first-order difference is not a white noise sequence.

3.2 Establishment of ARIMA model

3.2.1 Model ordering and parameter estimation

After the first-order difference, the time series of China CITIC Bank stock price reaches a stationary state, and then the ARIMA model is trained. Because the initial sequence has been first-ordered, the value of d is 1. The auto.arima function in the R software selects the model by selecting the smallest AIC and BIC. We can combine the autocorrelation graph and partial autocorrelation graph after the first-order difference to determine the order of the model.

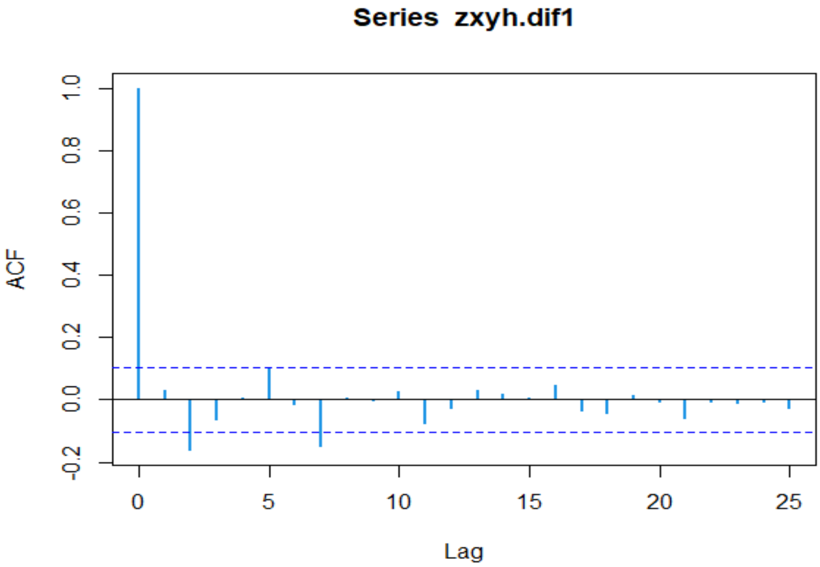

Figure 3 Autocorrelation diagram after sequence difference

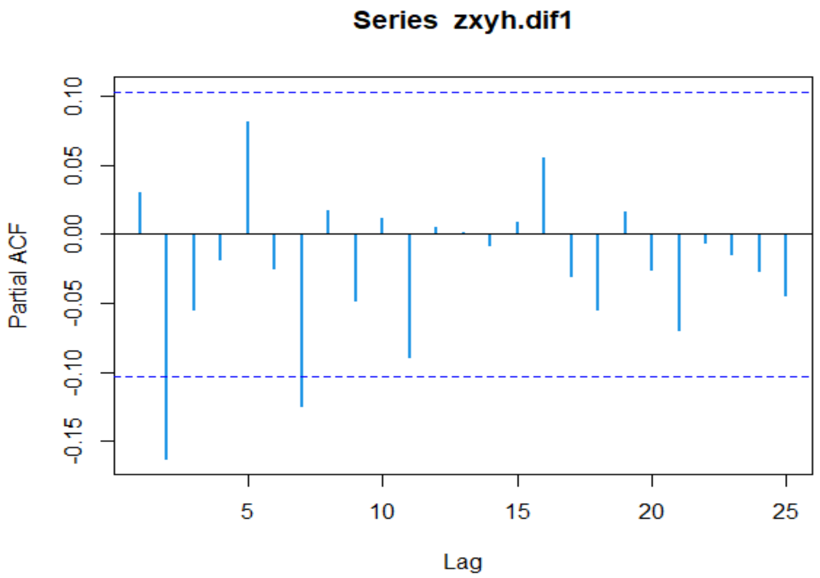

Figure 4 Partial correlation diagram after sequence difference

Table 4 AIC comparison

Model	AIC
ARIMA（1， 1， 1）	-423.59
ARIMA（1， 1， 2）	-428.9
ARIMA（1， 1， 3）	-429.23

According to the best model output by the software, combined with the first-order difference autocorrelation graph and partial autocorrelation graph, we finally determine the model to be ARIMA (1, 1, 3). The parameters of the model are estimated below, and the specific results are shown in the table below.

Table 5 Model parameters

Variable	AIC
AR（1）	0.9410
MA（1）	-0.9249
MA（2）	-0.1832
MA（3）	0.1311

3.2.2 Model checking

The Ljung-Box test is a test for the existence of lagging correlation in the time series. Its main function is to judge whether the overall sequence is correlated. When the statistics are in the rejection domain, it means that the null hypothesis is rejected and the sequence is correlated. Otherwise, the sequence is a white noise sequence.

Perform LB test on the residual of the obtained ARIMA model. If the residual of the model is a white noise sequence, the model is well trained and can be used for subsequent prediction. Observing the output results of the software, the calculated p-value is 0.985, which is obviously greater than 0.05. Therefore, the null hypothesis is not rejected. The residual data obtained after the model training is a white noise sequence, indicating that the model is good.

Plot the Q-Q (Quantile-Quantile Plot) chart. It can be seen from Figure 6 that the distribution of the residual sequence is similar to a straight line, indicating that the selected sequence approximately obeys the normal distribution. It can be seen from the normal QQ graph that the model fits well.

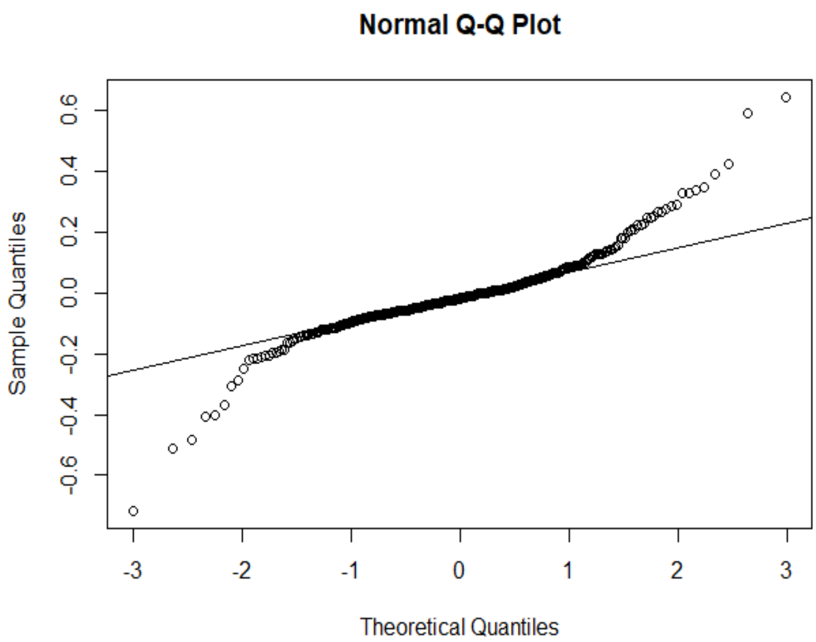

Figure 5 Q-Q plot of residual normal test

The residual test passed, and then the significance of the parameters was tested. According to the software running results, the p-value of 4.538195e-21 was far less than 0.05, and the parameter test was significant, and the model was successfully fitted.

4 Model prediction

After the series of numbers used have passed the test, predictions are made according to the ARIMA model. Based on the above analysis conclusion, it is determined that ARIMA(1,1,3) is used to predict the future price based on the 2018-2020 China CITIC Bank stock closing price data, which confirms that the built model ARIMA(1,1,3) has good credibility. According to the obtained prediction (Figure 8), the real price is within the confidence interval of the prediction. This result proves that the established model has a good degree of fit and can more accurately predict the closing price of CITIC Bank stock. It can be seen from Table 6 that the error between the predicted value and the actual value of CITIC Bank's stock in the next three working days will not exceed 0.5%. It can be seen that the prediction accuracy of the model is very high, and the predicted value is very close to the true value. At the same time, it is verified that the model constructed in this paper is relatively accurate and can well reflect the changing law of the stock price sequence of China CITIC Bank.

Table 6 Error table of predicted value and actual value

Date	Actual value	Predictive value	Error
2020-12-25	5.07	5.079787	0.19%
2020-12-28	5.07	5.087365	0.34%
2020-12-29	5.07	5.092448	0.44%

This model can well predict the changes in the closing price of China CITIC Bank stock. However, since the price series will change with the changes of various influencing factors in the actual analysis, all prices in the series are generally random. When using the past real price data to predict, there will inevitably be certain errors. Therefore, it is more suitable for short-term forecasting. The established model should be supplemented and revised by continuously updating the data to make the prediction more accurate.

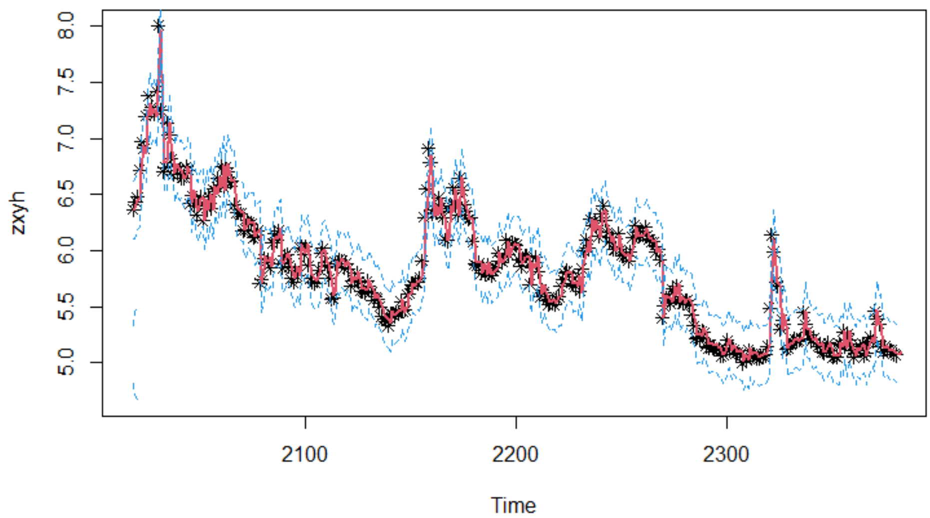

Figure 6 Model prediction results

According to the relevant value of the software operation result, the forecast situation can be expressed scientifically and reasonably from the aspect of econometrics. Among them, the coefficient of sill inequality is 0.017358, and its value is obviously less than 0.05; the value of the partial correlation coefficient is 0.00012 , This value represents that the prediction has a very high degree of fit, and the value of the variance ratio is 0.01698. The above three can effectively prove that the results of the study have a high degree of accuracy. The value of the covariance ratio in the correlation value is 0.875932, which is very close to 1. This result strongly proves that the model has a high degree of credibility. But when the value of the covariance ratio is high, it means that the deviation is extremely small. It also verifies that the forecast results are very similar to the real price data.

5 Conclusions

The method of studying stocks in this paper is different from the traditional simple linear analysis method. It uses the ARIMA model in time series analysis to conduct empirical analysis and short-term forecast of the closing price of CITIC Bank through model construction and fitting. First, the sequence composed of 363 samples is tested for stationarity and white noise by observing the time series diagram and unit root test. Secondly, estimate the model coefficient and order, and judge the fit of the model through residual test. Finally, establish a model and make short-term predictions through sequence analysis. Through research and the error results of the short-term prediction of the model, it can be seen that the ARIMA model has a good predictive effect on the short-term change of the stock price sequence. Therefore, investors can have a better judgment on the stock trend based on the predicted results. Investors and managers provide beneficial assistance in investment decisions.

References

1.Chen Xiaoling. Stock price prediction based on ARIMA model and neural network model[J]. Economic Mathematics, 2017, 34(04): 30-34.
2.Yan Yu, Wu Haitao. Short-term forecast of Nasdaq index based on ARIMA model[J]. Information and Computer (Theoretical Edition), 2020, 32(20): 155-158.
3.Song Gang, Zhang Yunfeng, Bao Fangxun, Qin Chao. Stock prediction model based on particle swarm optimization LSTM[J]. Journal of Beijing University of Aeronautics and Astronautics, 2019, 45(12): 2533-2542.
4.Yang Jin, Chen Liang. Application of combination model based on wavelet neural network and ARIMA in stock forecasting [J]. Economic Mathematics, 2018, 35(02): 62-67.
5.Wu Yuxia, Wen Xin. Short-term stock price prediction based on ARIMA model[J]. Statistics and Decision, 2016(23): 83-86.

