Quantitative Forecasting Method of Stock Price Based on Time Series Model

Lu Xu

Hebei University,Bao Ding,China

327849639@qq.com

Abstract. Nowadays, our economy is developing rapidly, and the financial market has gradually become a very important part of our economic development, in which the stock market is an important part of the financial market and closely related to our economy. For investors, a key problem in making decisions is how to accurately analyze the stock market by knowing the price fluctuation in time; For the managers of the stock market, it is also a very arduous task to create a relatively stable trading environment by mastering the real-time dynamics of the stock market. Because of this, it is of great significance for us to better understand the characteristics of stock market fluctuations and find out the laws. Taking the Shanghai-Shenzhen 300 Index, which reflects the overall trend of A-share market from 2009 to 2018, as an example, this paper uses stationarity test, pure randomness test and other test methods to fit ARIMA model, ARCH model and AR-GARCH model, compare their advantages and disadvantages in the stock price trend, and then make a short-term forecast of the stock price with the fitting model that has passed the test. Finally, it is found that the AR-GARCH model has a good fitting effect on the original sequence, and a more accurate prediction result is obtained.

Keywords: stock price , forecast,fitting , Shanghai and Shenzhen 300 index

1 Data cleaning and Data Preprocessing

1.1 Stationarity Test

This thesis selects the closing price of the CSI 300 Index every working day from January 2009 to December 2018 as the research object, draws its time series diagram, and makes a preliminary analysis.

Y. Jiao et al. (eds.), *Proceedings of the 3rd International Conference on Internet Finance and Digital Economy (ICIFDE 2023)*, Atlantis Highlights in Economics, Business and Management 1,
https://doi.org/10.2991/978-94-6463-270-5_6

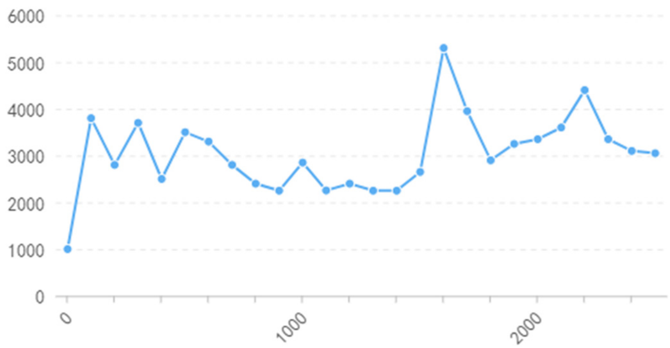

Fig. 1. the timing diagram of the original sequence

From the timing diagram of the original sequence (fig.1), we can see that the sequence is not obvious periodicity, and we can't determine whether it is stable by direct observation. At this point, we use auto-correlation diagram to determine whether the sequence is stationary.

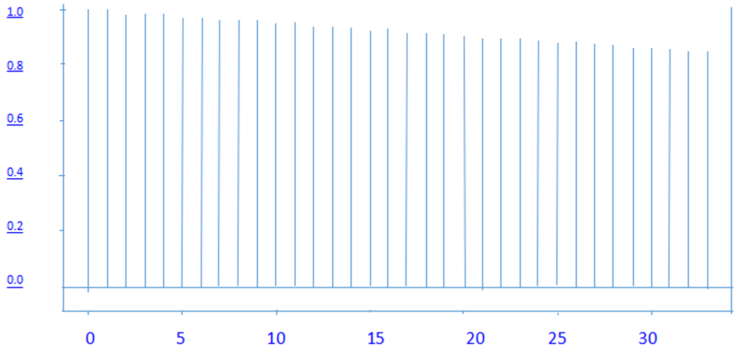

Fig. 2. the ACF diagram of the original sequence

According to the ACF diagram of the original sequence (fig.2), it can be seen that with the increase of the delay period k, ACF is always positive above the zero axis and decays to zero very slowly, so it can be judged that the original sequence is a non-stationary sequence.

1.2 Pure Randomness Test

At this point, the pure randomness of the sequence is tested again. The results show that the P value of LB test statistics is far less than 0.05 under each delay, so we can make sure that the original sequence is non-white noise sequence, which is of significance for further research and future prediction.

2 Stock Price Modeling

2.1 ARIMA model

The deterministic information extraction function of difference operation is very powerful. Generally speaking, stationary series will be displayed after the difference of non-stationary time series with random trend. This non-stationary sequence is called differential stationary sequence. For this difference stationary sequence, we need to use ARIMA model to fit [1-5].

The sequence diagram after first-order difference processing will show that the sequence has been fluctuating around the zero axis, and there is no obvious law. At this time, the unit root test is carried out, and the p value is less than 0.05, so we have to reject the original hypothesis, and the sequence can be considered stable at this time. Then, we test the white noise of the sequence after the first-order difference, and find that the P value of the LB test statistic is less than 0.05 at each delay, so the sequence after the first-order difference is a stationary non-white noise sequence and can be fitted by ARIMA model.

It can be considered that both self-correlation coefficient and partial self-correlation coefficient are trailing, and the order cannot be determined accurately. At this time, the auto.arima function provided by R software is used to automatically identify the model order, and finally the ARIMA(3,1,2) model is fitted to the original sequence.

At this time, the model is tested for significance to test the effectiveness of the model. Because the P value of LB statistics is greater than 0.05 under each delay, it can be considered that the residual sequence of this fitting model belongs to white noise sequence, that is, the fitting model is obviously effective.

According to the output fitting results, the fitting model is obtained as follows:

$$(1-B)x_t = \frac{1-0.1425B+0.9055B^2}{1-0.1921B+0.9612B^2-0.0738B^3}\varepsilon_{t}, \quad \varepsilon_t \sim N(0,\ 2475) \tag{1}$$

2.2 ARCH model

When we use ARIMA model to fit nonstationary series, { ε }t usually defaults to zero-mean white noise series directly. But before that, we ignored any homogeneity test of the other party. When dealing with financial time data, ignoring the existence of heteroscedasticity will cause the variance of residual to be seriously underestimated, making the significance test of parameters meaningless, and then affecting the fitting accuracy of the model [6-8].

In the actual processing of financial data, self-correlation conditional heteroscedasticity model, referred to as ARCH model, is widely used.

$$h_t = E(\varepsilon_t^2) = \omega + \sum_{j=1}^{q} \lambda_j\, \varepsilon_t^2 - \mathrm{j} \tag{2}$$

The variance of the columns is homogeneous, and the residual sequence after squared processing shows obvious heteroskedasticity characteristics which require further processing. The specific form of the heteroskedasticity function is unknown, so to fit the conditional heteroskedasticity model.

Next, ARCH test is performed on the residual sequence. Portmanteau Q test and LM test are usually used. After the test, it is found that the P values of Portmanteau Q test statistics and LM test statistics are both less than 0.05 under each delay, so the original hypothesis is rejected, and the variance of the sequence is considered to be non-homogeneous and self-correlation, so ARCH model can be used to extract the relevant information contained in the residual square sequence.

The fit model obtained is ARCH(3) model:

$$E\left({\varepsilon_t}^2 | \varepsilon_{t-1,}\varepsilon_{t-2\ldots}\right)=1107+0.1186\varepsilon_{t}^{2}{}_{-1}+0.2386{\varepsilon_t}^2{}_{-2}+0.205{\varepsilon_t}^2{}_{-3} \quad (3)$$

ARCH model fits the range of 95% confidence interval, because it considers the fluctuation characteristics of the first-order differential sequence, so ARCH model better fits the fluctuation characteristics of the cluster effect of the first-order differential sequence.

2.3 AR-GARCH model

Because the residual sequence {e}t has been found to have self-correlation in previous tests, it needs to be fitted by auto-regressive model first.

Under the conditional heteroscedasticity test, the statistics of Portmanteau Q test are less than 0.05 significance level under each delay, which indicates that ARIMA(3,1,2) has heteroscedasticity, so it is fitted with GARCH(1,1) model, and the 95% confidence interval diagram of fluctuation is drawn according to the fitting results of this model[9-10].

Combining the horizontal model and the fluctuation model, the complete AR-GARCH fitting model is finally obtained as follows (fig.3):

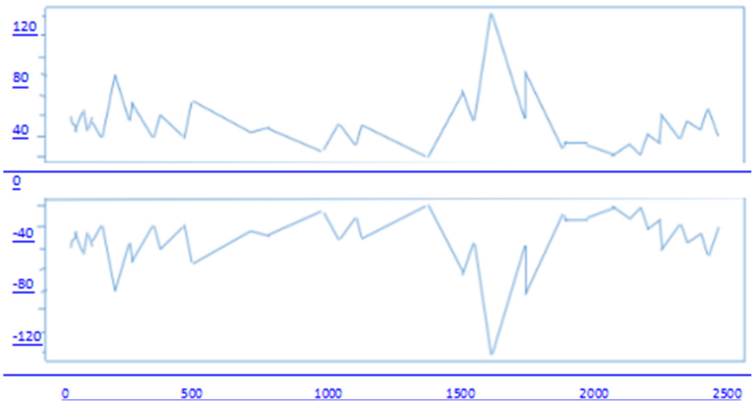

Fig. 3. the complete AR-GARCH fitting model

$$\begin{cases} \nabla x_t = \frac{1-0.1425B+0.9055B^2}{1-0.1921B+0.9612B^2-0.0738B^3}\varepsilon_t + \upsilon_t, \upsilon_t \sim N(0.2475) \\ \upsilon_t = \sqrt{h_t}\, e_t \\ h_t = 0.938859h_{t-1} + 0.058688\upsilon_t^2 - 1 \end{cases} \tag{4}$$

Figure 4 below gives the fluctuation confidence interval of residual sequence fitted by AR-GARCH model.

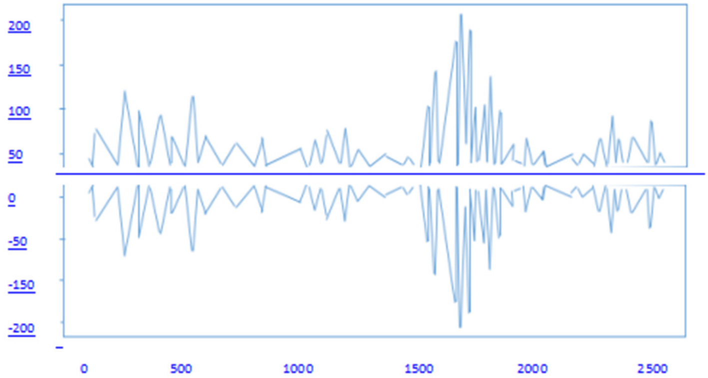

Fig. 4. Confidence Interval of Residual Sequence Fluctuation

After comparison, it is found that the 95% confidence interval drawn by the fitting result of AR-GARCH model is more consistent with the real fluctuation of the original sequence than the 95% confidence interval fitted by ARCH model, which shows that the fitting effect of AR-GARCH model is better and the prediction of sequence fluctuation will be more accurate.

2.4 Model Comparison

In this chapter, ARIMA(3,1,2), ARCH(3) and AR-GARCH(1,1) models are used to fit the Shanghai and Shenzhen 300 stock prices,all three models have passed the significance test and obtained good fitting results. In order to get a relatively optimal fitting model, the minimum information criterion (AIC) should be used to compare the three models. We use the following formula to calculate the AIC value of the fitting model:

AIC=-2ln (maximum likelihood function value of the model) +2 (number of unknown parameters in the model)

Table 1. AIC of Three Fitting Models

Fitting model	AIC
ARIMA(3,1,2)	25896.29
ARCH(3)	25319.1
AR-GARCH(1,1)	24944.96

According to the AIC minimum principle (Table 1) , it is concluded that AR-GARCH(1,1) is the relatively optimal model, followed by ARCH(3) and ARIMA(3,1,2).

3 Model Prediction

3.1 AR-GARCH model prediction

Select the relative optimal model, use the "rugarch" package to predict with the fitted AR-GARCH model, and convert the prediction value of the residual sequence into the prediction value of the original sequence.

Since the real value of the closing price of the Shanghai and Shenzhen 300 Index in the next 10 days has been obtained at this time, the quality of the model prediction can be judged by comparing the error between the real value and the predicted value. Next, the predicted values obtained by the three models are compared with the real values:

Among them, the relative errors of ARIMA(3,1,2) model and ARCH(3) model are very close, and it can be considered that the prediction effects of the two models are similar. The relative error of GARCH(1,1) model is obviously smaller than the first two models, so it can be considered that GARCH(1,1) model has the highest prediction accuracy, that is, the prediction result will be closer to the real value. This result is also consistent with the result of comparing the advantages and disadvantages of the model in the previous chapter (Table 2) .

Table 2. Comparison of Relative Errors of Three Models' Prediction

Period	True Value	ARIMA True Error	ARCH True Error	GARCH True Error
T+1	2969.5353	1.4949%	1.4005%	1.5230%
T+2	2964.8421	1.6072%	1.5610%	1.7904%
T+3	3035.8741	0.8339%	0.8153%	0.6349%
T+4	3054.303	1.3914%	1.4138%	1.3008%
T+5	3047.7035	1.1125%	1.2003%	1.0118%
T+6	3078.4759	2.1321%	2.1879%	1.8760%
T+7	3072.6864	2.0131%	2.0036%	1.6962%
T+8	3094.7782	2.6904%	2.7031%	2.4793%
T+9	3067.7845	1.7693%	1.8470%	1.5877%
T+10	3127.9904	3.6735%	3.7362%	3.3510%

3.2 Horizontal Forecast

In order to test the long-term prediction effect of the model, the AR-GARCH(1,1) model with the best prediction effect among the above models is selected to predict the closing price in the next 50 days (about January-February 2019).

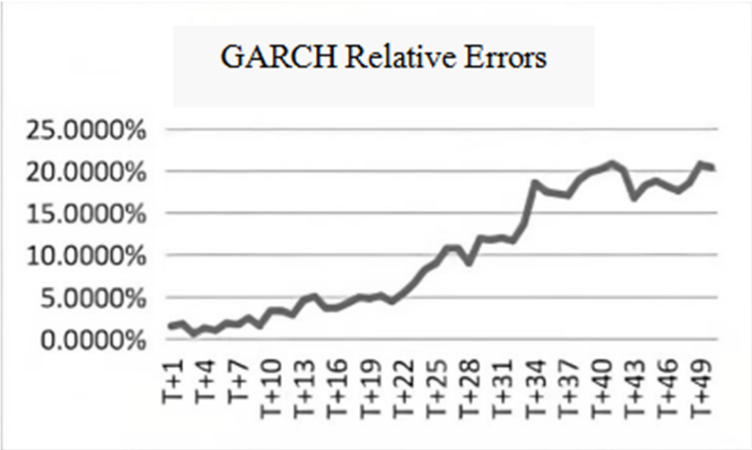

Fig. 5. Relative Error of Predictive Value of AR-GARCH Model in the Next 50 Days

As shown in Figure 5, as the number of periods increases, the relative error between the predicted value and the actual value of the AR-GARCH (1, 1) model gradually increases, by the 50th period, the relative error has exceeded 20%, indicating that the prediction accuracy of the model is getting lower and lower, and it loses its ability to accurately predict the closing price.

4 Conclusion

This paper mainly discusses the application of sum auto-regressive moving average model and conditional heteroscedasticity model in stock price time series analysis. By fitting the Shanghai and Shenzhen 300 stock prices with three models, the significance models have been successfully established, which once again proves the predictability of stock prices:

(1) In the case studied in this paper, the heteroscedasticity function of residual sequence has long-term self-correlation, so it will produce high-order moving average when using ARCH model for fitting, and GARCH model can overcome this problem and improve the fitting accuracy. Therefore, for the long-term self-correlation series of heteroscedasticity function such as stock price time series, GARCH model fitting can obtain relatively optimal fitting effect, so as to get more accurate prediction.

(2) The result of short-term prediction is not much different from the actual value, and the relative error of prediction can basically be controlled within 2%, which shows that the fitting effect is very good, and the model can be used to make short-term prediction of the recent stock market to help investors make better investment decisions. However, when the model is used for long-term prediction, the results are not satisfactory, the prediction accuracy drops obviously, and the ability of accurate prediction is lost.

(3) Time series is an effective tool for financial market analysis, which has great practical value and development prospects. One application is to explore the law of stock price fluctuation through time series, so as to help investors make investment decisions, avoid risks and obtain maximum benefits. However, it is very complicated to study the price fluctuation in the stock market. In this paper, we only analyze and predict the trend of the Shanghai and Shenzhen 300 Index based on historical data, without considering other factors, such as return on investment and risk premium. How to deal with the original data and choose the optimal model needs further research and improvement.

References

1. A. A. Garibo-Morante and F. Ornelas Tellez, "Univariate and Multivariate Time Series Modeling using a Harmonic Decomposition Methodology," in IEEE Latin America Transactions, vol. 20, no. 3, pp. 372-378, March 2022
2. A. Samé and G. Govaert, "Online Time Series Segmentation Using Temporal Mixture Models and Bayesian Model Selection," 2012 11th International Conference on Machine Learning and Applications, Boca Raton, FL, USA, 2012, pp. 602-605
3. N. Zeng, Z. Wang, Y. Li, M. Du, J. Cao and X. Liu, "Time Series Modeling of Nano-Gold Immunochromatographic Assay via Expectation Maximization Algorithm," in IEEE Transactions on Biomedical Engineering, vol. 60, no. 12, pp. 3418-3424, Dec. 2013
4. B. Khayyat, F. Harrou and Y. Sun, "Predicting COVID-19 Spread using Simple Time-Series Statistical Models," 2021 International Conference on ICT for Smart Society (ICISS), Bandung, Indonesia, 2021, pp. 1-5
5. T. N. A. B. T. M. Busu, S. A. Kamarudin, N. A. Ahad and N. A. M. G. Mamat, "Prediction of FTSE Bursa Malaysia KLCI Stock Market using LSTM Recurrent Neural Network," 2022 IEEE International Conference on Computing (ICOCO), Kota Kinabalu, Malaysia, 2022, pp. 415-418
6. L. Lindsay, D. Kerr, S. Coleman and B. Gardiner, "Transparent Models for Stock Market Price Forecasting," 2022 IEEE Symposium Series on Computational Intelligence (SSCI), Singapore, Singapore, 2022, pp. 860-866
7. H. Xianping, H. Dengshi, X. Kai and Z. Hu, "Research on contagion effect among China's bond market and stock market," 2013 6th International Conference on Information Management, Innovation Management and Industrial Engineering, Xi'an, China, 2013, pp. 648-651
8. C. Fonseka and L. Liyanage, "A Data mining algorithm to analyse stock market data using lagged correlation," 2008 4th International Conference on Information and Automation for Sustainability, Colombo, Sri Lanka, 2008, pp. 163-166
9. L. Mathanprasad and M. Gunasekaran, "Analysing the Trend of Stock Marketand Evaluate the performance of Market Prediction using Machine Learning Approach," 2022 International Conference on Advances in Computing, Communication and Applied Informatics (ACCAI), Chennai, India, 2022, pp. 1-9
10. N. Sharma and B. R. Mohan, "Stock Market Prediction Using Historical Stock Prices And Dependence On Other Companies In Automotive Sector," 2022 IEEE 10th Region 10 Humanitarian Technology Conference (R10-HTC), Hyderabad, India, 2022, pp. 425-431