Empirical analysis of stock prices based on ARIMA-GARCH

Pu Feng *

School of Economics & Management, Nanjing University of Science and Technology, Nanjing, China

* Corresponding Author Email: 1260917958@qq.com

Abstract. Although ARIMA time series model can grasp the dynamic law of financial time series well, it can not reflect the volatility characteristics of stock price. GARCH can reflect the volatility of financial time series well and is suitable for forecasting in financial field. Taking the closing price of Shanghai Dazhong Public Utilities (dzgy) from January 2, 2019 to December 31, 2021 and using R , ARIMA-GARCH is proposed and the predictive analysis is carried out. The results show that ARIMA-GARCH has a good effect in short-term stock prediction, and has certain reference value and practical significance.

Keywords: ARIMA-GARCH; forecast; stock price.

1. Introduction

With the improvement of per capita GDP in our country, more and more people choose to manage money. Securities investment, especially stock investment, has become the choice of most people. For the Stockholders in the face of many stocks, it is particularly important to choose the stock with high yield. As a kind of financial time series data, stock data can be used to make corresponding trend prediction, so that investors can have reference before making decisions and help they make judgments [1].

At present, the common analysis methods of securities investment include basic analysis method and technical analysis method. The analysis of basic factors is to determine the intrinsic value of securities through the analysis of the basic factors affecting the supply and demand relationship in the securities market, and to develop an analytical tool for investment in securities. The basic factor analysis is to find the securities with long-term investment value through macro investment environment analysis, industry analysis and company analysis. However, these two kinds of analysis can generally only predict the trend of stock price changes. Therefore, if we want to obtain a more accurate portfolio of expected returns and corresponding risks, we need to conduct further analysis prediction based on the results of these two kinds of analysis, and we need to use time series models such as ARIMA model [2]. Although ARIMA time series model can grasp the dynamic law of financial time series well, it cannot reflect the volatility characteristics of stock price. The GARCH model can overcome the assumption of constant difference in the traditional stock forecast, better reflect the volatility of financial time series, and is suitable for financial forecasting. Therefore, this paper combines ARIMA and GARCH model to construct a new ARMI-GARCH model, which is used in the prediction of dzgy closing price.

2. Model Introduction

2.1. General Form of the ARMA Model

A model with the following structure is called autoregressive integrated moving average model, abbreviated as ARIMA (p, d, and q) model:

$$Y_t - \varphi_0 - \varphi_1 Y_{t-1} - \ldots - \varphi_p Y_{t-p} = a_t - e_1 a_{t-1} - \ldots - e_q a_{t-q}. \quad (1)$$

Where, a_t, a_{t-1},... a_{t-q} is a stationary white noise with a mean of 0 and a variance of ξ^2 ,where p and q are respectively order of autoregressive model and moving average model [3].

The essence of ARIMA model is the combination of difference operation and ARMA model. This relationship is significant, indicating that if any non-stationary sequence can achieve post-difference stationarity through the difference of appropriate order, ARMA model can be fitted to post-difference sequence. The analysis method of ARMA model is very mature, which means that the analysis of differential stationary sequence will be very simple and very reliable. In particular, when d=0, the ARIMA (p, d, q) model is actually the ARMA (p, q) model. When p=0, the ARIMA (p, d, q) model can be abbreviated as the IMA (d, q) model. When q=0, the ARIMA (p, d, 0) model can be abbreviated as the ARI (p, d) model [4].

2.2. General Form of the GARCH Model

In 1985, Bollerslov proposed generalized autoregressive conditional heteroskedastic model, its structure is as follows:

$$\sigma_t^2 = \omega + \alpha_1 a_{t-1}^2 + \alpha_2 a_{t-2}^2 + ... + \alpha_q a_{t-r}^2 + \beta_1 \sigma_{t-1}^2 + \beta_2 \sigma_{t-2}^2 + ... + \beta_p \sigma_{t-s}^2. \quad (2)$$

Its constraint conditions are:

$$\sum_{i=1}^{t} \alpha_i + \sum_{i=1}^{s} \beta_i < 1, \alpha_0 > 0, \alpha_i \geq 0, i = 1,2,..., r, \beta_i \geq 0, i = 1,2,..., s. \quad (3)$$

Where, σ_t^2 represents the variance of the disturbed term a_t. GARCH(r,s) is a model with additional lag term, which is actually formed by adding the p-order autocorrelation of the heteroscedastic function on the basis of ARCH model, and it can effectively fit the heteroscedastic function with long-term memory.

A complete conditional heteroscedastic model is composed of three parts: mean model, conditional heteroscedastic model and distribution assumption. The GARCH model can be used to analyze many different types of financial data, such as macroeconomic data. This model is commonly used by financial institutions to estimate the earnings volatility of stocks, bonds, and market indices. They use the information obtained to help determine pricing and determine which assets can provide higher returns, as well as forecasts to help make asset allocation, hedging, risk management, and portfolio optimization decisions [5].

2.3. ARIMA-GARCH Model

The ARIMA-GARCH hybrid model is the combination of ARIMA model and GARCH model, combined with formula (1) and formula (2), considering the ARIMA (p,d,q)-GARCH(r,s) process, then

$$\begin{aligned} Y_t &= c + \sum_{i=1}^{p} \varphi_i Y_{t-i} + ... + \sum_{j=1}^{q} \theta_j a_{t-j} + a_t, \\ \sigma_t^2 &= \omega + \sum_{i=1}^{r} \alpha_i a_{t-i}^2 + \sum_{j=1}^{s} \beta_j \sigma_{t-j}^2, \\ a_t &= \sigma_t \varepsilon_t, \varepsilon_t \sim N(0,1). \end{aligned} \quad (4)$$

ARIMA-GARCH model can not only better grasp the dynamic law of financial time series, but also better reflect the volatility of financial time series, which is more suitable for financial field and more practical.

3. Experimental Procedures

(1) The ADF unit root test completes the stationarity test of stock price.

(2) If the stock sequence is non-stationary, the stock price sequence is converted to stationary sequence by differential processing.

(3) AIC (Akaike information criterion) was used to determine the order of the model and estimate the parameters.

(4) ARCH test was performed on the ARIMA model to determine whether there is heteroscedasticity. If there is heteroscedasticity, GARCH model should be established.

(5) The residual test of the ARMI-GARCH model is carried out to check whether the residual is consistent with the white noise. If not, the model needs to be further optimized.

(6) After the parameters of the ARIMA-GARCH model are determined, the closing price of dzgy is predicted.

4. Empirical analysis

4.1. Data Source

730 stock closing price data of dzgy from January 2, 2019 to December 31, 2021 were randomly selected to establish a time series model to be analyzed. Some data are shown in Table 1. The closing prices of 5 days on January 4, January 5, January 6, January 7 and January 10, 2022, are selected as test data to predict the closing price from finance-163.com.

Table 1. January 2, 2019 - December 31, 2021 dzgy partial stock data.

Date	Closing price	Date	Closing price	Date	Closing price
20190102	4.73	20200102	4.94	20210104	4. 13
20190103	4.48	20200103	5	20210105	4. 11
20190104	4 66	20200106	5 31	20210106	4 14
20190107	4.74	20200107	5.26	20210107	4.06
20190108	4.87	20200108	5.1	20210108	4.06
20190109	4.91	20200109	5.1	20210111	4
20190110	4.74	20200110	5.05	20210112	4.06
20190111	4 77	20200113	5 09	20210113	4 06
20190114	4.79	20200114	5.08	20210114	4.06
20190115	4.76	20200115	5.03	20210115	4.07
20190116	4.87	20200116	4.98	20210118	4.05
20190117	5. 17	20200117	4.96	20210119	4.06
20190118	4 97	20200120	5 01	20210120	4 04
20190121	5.03	20200121	4.93	20210121	4.04
20190122	4.99	20200122	4.91	20210122	4
20190123	5.02	20200123	4.72	20200125	3.88
20190124	4.89	20200203	4.25	20200126	3.84
20190125	4 63	20200204	4 13	20200127	3 89
20190128	4.64	20200205	4. 18	20200128	3.83
20190129	4.45	20200206	4.23	20200129	3.77
20190130	4.32	20200207	4.22	20210201	3.69
20190131	4. 15	20200210	4.27	20210202	3.69

4.2. Stationarity Test

First of all, it can be seen from Figure 1 that the sequence of dzgy closing prices from January 2, 2019 to December 31, 2021 has an obvious decreasing trend year by year. Among them, the closing price has small peaks in April 2019, June 2019, June 2020 and August 2021 respectively, and then it gradually declines, which shows that the series has more peaks and is not a stable series. Because it is very subjective to judge the stationarity of the sequence by the time series diagram. It is not certain whether the sequence is stationary or non-stationary. Therefore, the unit root test in Figure 3 is

combined, and the test results show that if the structure of the sequence considers the three types in the table, the P-value is significantly greater than the significance level (α=0.05). That is, the null hypothesis cannot be rejected and the sequence is non-stationary.

The first-order difference was performed on the sequence, and the time sequence diagram after the difference (see Figure 2) showed that the sequence after the difference basically fluctuated around the value 0, and there was no obvious trend feature. In order to further determine the stability of the differential sequence, ADF test is performed on the differential sequence. The test results show that the P-values measured by all the ADF test series of this series are lower than the significance level (α=0.05), then the null hypothesis is rejected and the time series is a stationary time series. Then the pure randomness test of the sequence after first difference is carried out. The test results show that the p-values of LB statistics at each delay order are all smaller than the significance level, which indicates that the differential sequence is not a white noise sequence. Therefore, it can be confirmed that the sequence after the first difference is a stationary non-white noise sequence, which meets the requirements.

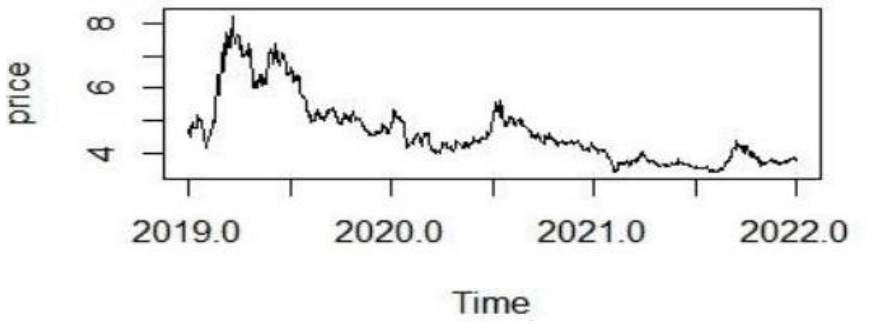


Fig 1. Time series of dzgy closing prices.

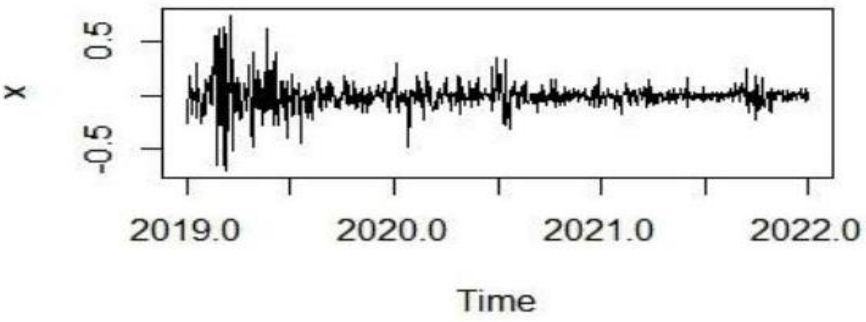


Fig 2. Dzgy 1 order difference closing time series diagram.

4.3. Model Selection and Parameter Estimation

Dzgy closing price sequence with known first-order difference is a stationary non-white noise sequence. For the results of the differential autocorrelation coefficient (ACF) and partial autocorrelation coefficient (PACF) (see Figure 3), there are two interpretations. One is that both autocorrelates and partial autocorrelates show an untruncated nature, so you can try using the ARIMA (1, 1, and 1) model. Second, the autocorrelation is truncated at the lag of 1, and the partial autocorrelation shows a trailing property, so you can also try to use the ARIMA (1, 1, and 0) model.

Alternatively, you can try using the auto.arima function, which selects ARIMA (2, 1, and 0) (see Figure 4). The AIC criterion is used to determine the order of the model. The smaller the AIC value, the higher the fit degree of the model.

AIC values of ARIMA(1,1,1), ARIMA(1,1,0), ARIMA(2,1,0) and other models were calculated respectively, and the results were shown in Table 2. It is clearly shown that the AIC values of ARIMA(1,1,1), ARIMA(1,1,0), ARIMA(2,1,0), ARIMA(1,1,1), ARIMA(2,1,1), and ARIMA(2,1,1) are -908.05, -909.62,-907.66, -907.04, -905.77, -905.53 respectively.

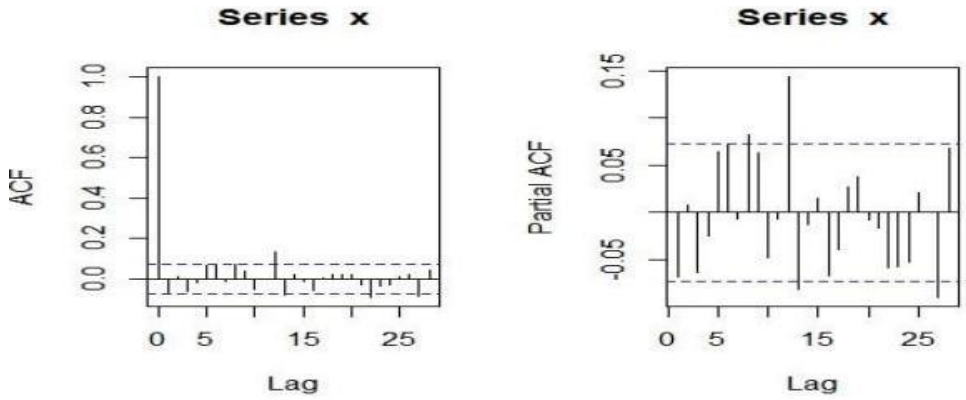


Fig 3. Dzgy closing price differential sequence autocorrelation and partial autocorrelation diagram.

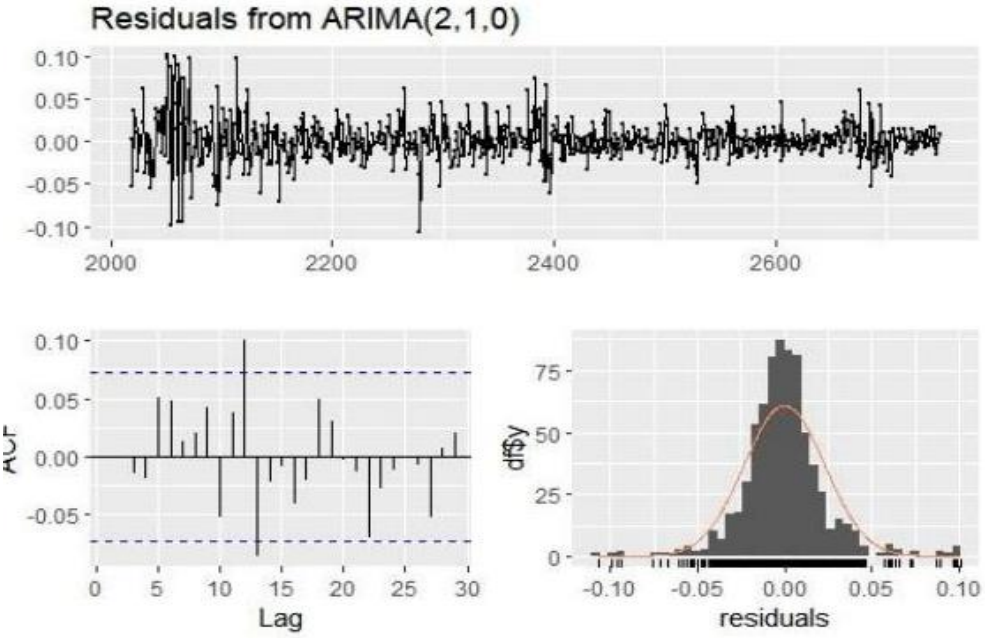


Fig 4. Auto. arima function automatic order ARIMA (2, 1, 0).

Table 2. AIC values of models of different order.

Model	AIC value
ARIMA(1,1,1)	-908.05
ARIMA(1,1,0)	-909.62
ARIMA(2,1,0)	-907.66
ARIMA(1,1,2)	-907.04
ARIMA(2,1,1)	-905.77
ARIMA(2,1,2)	-905.53

As can be seen from Table 2, the model with the smallest AIC value is ARIMA (1, 1, and 0).

The ARIMA (1, 1, 0) model is predicted, and the forecast function is called to make a 5-stage forecast. The prediction results is shown in Figure 5. In Figure 5, the dotted line is the observed value, the solid line is the fitting value, the dark shaded part is the predicted value confidence interval with 80% confidence level, and the light shaded part is the predicted value confidence interval with 95% confidence interval.

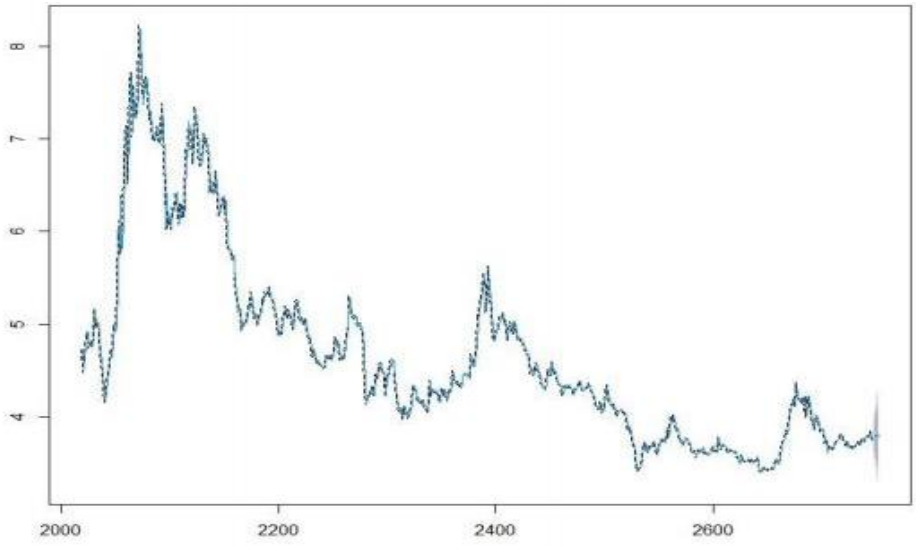


Fig 5. ARIMA (1, 1, 0) model 5 prediction.

In order to compare the prediction effect of ARIMA (1, 1, and 0) model, the results were compared with the real data, and the results were shown in Table 3.

Table 3. ARIMA (1, 1, and 0) predicts 5 closing price results

Date	True value	Predicted value	Relative error
4	3.85	3.787959	0.016114545
5	3.82	3.788098	0.008351309
6	3 84	3 788088	0 01351875
7	3.86	3.788089	0.018629793
10	3.89	3.788089	0.026198201

Next, for the ARIMA (1, 1, and 0) model, ARCH test is performed on its normalized residuals to determine whether the model has heteroscedasticity.

The results show that both the Q test and the LM test show significant variance heterogeneity in the 24-order delay, which indicates the existence of a long-term correlation in the residual squared sequence. In this case, it is usually possible to use the high-order ARCH model or the low-order GARCH model to extract the correlation contained in the residual square sequence. Therefore, this paper chooses to try to fit the GARCH (1, 1) model.

4.4. Model Parameter Estimation

The ARIMA (1, 1, 0)-GARCH (1, 1) model was used to fit the data and estimate the parameter model. The results are shown in Table 4.

Table 4. ARIMA (1, 1, 0)-GARCH (1, 1) model parameter estimation results.

Parameter	Estimated value	Deviation	T	P
mu	-0.001576	0.002523	-0.62486	0.042063
ar1	-0.039773	0.040416	-0.98409	0.035073
omega	0.000112	0.000048	2.33268	0.019665
alpha1	0.123186	0.023506	5.24053	0.000000
beta1	0.872214	0.021251	41.04274	0.000000

The fitting model can be obtained from the fitting parameters in Table 4 as follows:

$$\begin{cases} y_t = -0.001576 - 0.39773 y_{t-1} + a_t \\ a_t = \sigma_t \varepsilon_t \\ \sigma_t = 0.000112 + 0.123186 a_{t-1}^2 + 0.872214 \sigma_{t-1}^2 \end{cases} \quad (5)$$

4.5. Residual Test

Using the fitting value of GARCH (1, 1), the standardized residual sequence is transformed by heteroscedasticity normalization, and the standardized residual square sequence is obtained. The white noise test is carried out on the two sequences respectively. The standardized residual sequence test results show that the P-value of LB statistic with delay of order 1 is equal to 0.8569.After the second order, the P-values are all greater than 0.05, so it can be considered that the current mean value model is sufficient to extract information.

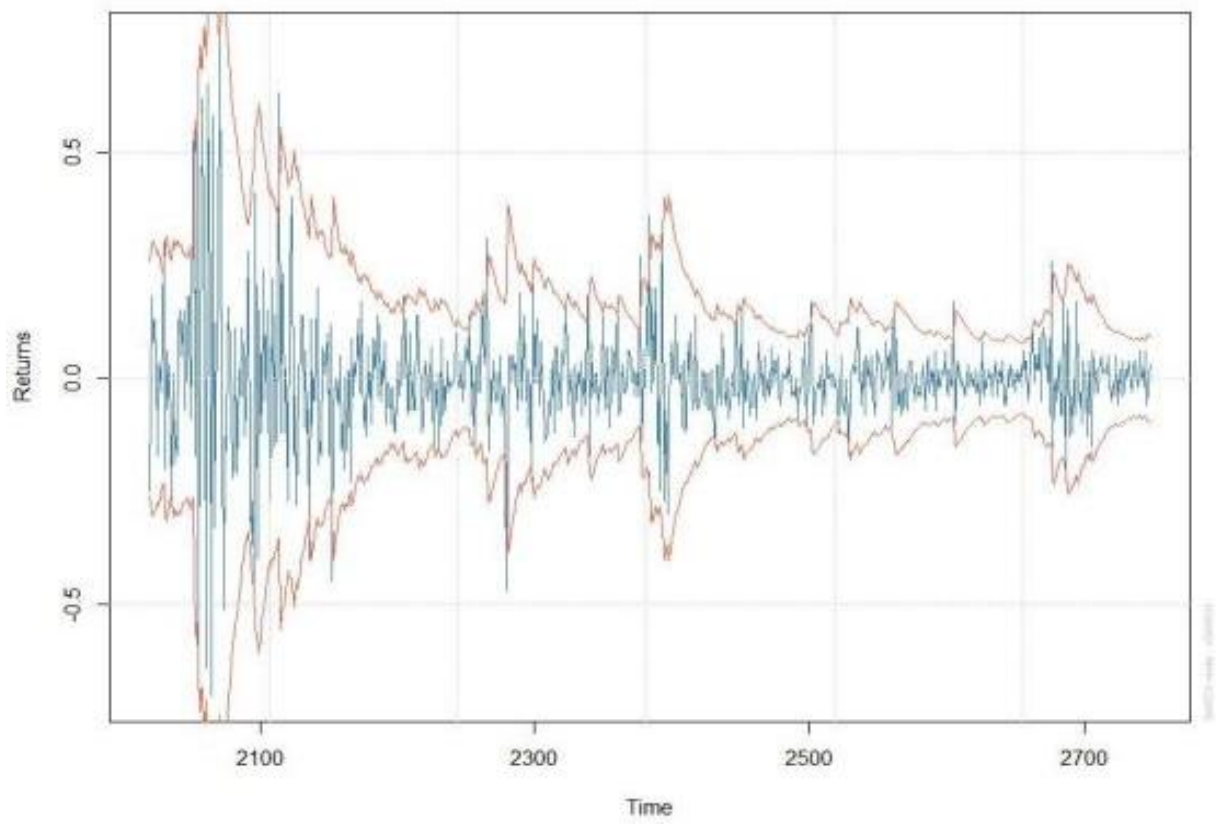


Fig 6. Comparison of confidence intervals under the two variance assumptions.

The test results of the standardized residual square sequence show that the p-values of LB test statistics are all greater than 0.05, which indicates that the GARCH (1, 1) model is very sufficient in extracting fluctuation information. This GARCH fitting model is remarkably valid.

Figure 6 describes the 95% confidence intervals of the residual sequences obtained under the assumption of homogeneity of variance and the 95% confidence intervals of the residual sequences obtained based on the GARCH (1, 1) model under the assumption of non-homogeneity of squares. It can be clearly seen from the figure that the conditional heteroscedastic confidence interval is wider when the series fluctuates greatly, while the conditional heteroscedastic confidence interval is narrower when the series fluctuates slightly. This shows that the conditional heteroscedasticity model is usually more accurate in fitting and predicting the risk of series fluctuation.

4.6. Prediction of Results

The ARIMA (1, 1, 0)-GARCH (1, 1) model was used to predict the five-day closing price of dzgy.

One step of sigma is expected to be 0.04525. Because a normal distribution is assumed, VaR with a confidence level of 95% can be calculated using the 95% quantile of the standard normal distribution (enter qnorm (0.95)). So for the next period, that is, 95% of T+1 VaR is qnorm (0.95)×0.04525=0.07442963. And so on, the prediction of five phases can be obtained, as shown in Figure 7. As can be seen from Figure 7, the trained model can well predict the closing price of dzgy, indicating that the ARIMA-GARCH model can be applied to stock price prediction.

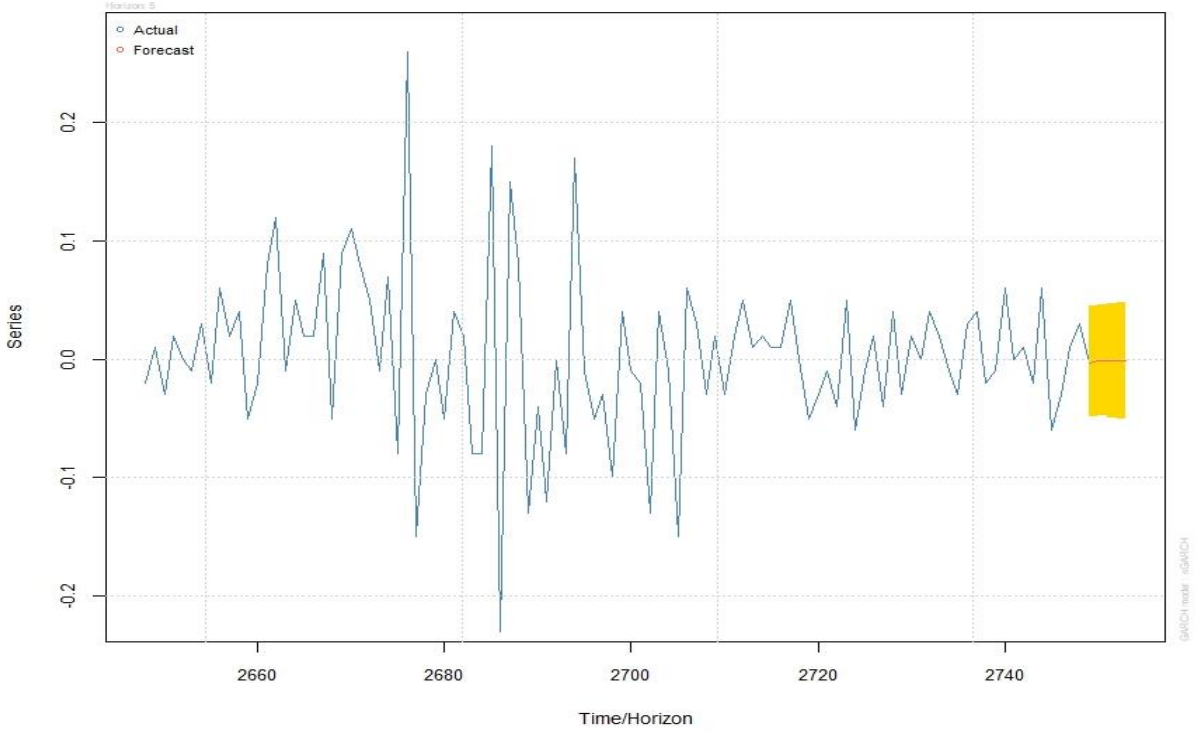


Fig 7. 5 phase forecast chart.

5. Summary

This paper uses the time series model ARIMA-GARCH to conduct modeling, fitting and empirical analysis on 730 sets of closing data from January 2, 2019 to December 31, 2021, and forecasts the stock prices in the next five periods. This indicates that the ARIMA-GARCH model can better solve and fit the fluctuations of the original time series, and can provide certain reference value for stock decision-making in short-term prediction. However, if the model continues to be used for long-term prediction analysis, the prediction error may accumulate, and the prediction will not be so accurate [6]. This paper only conducts modeling and analysis of dzgy's stock closing price. The establishment of this model is greatly influenced by stock data, and the model is sensitive to the influence of the sample data studied, so it lacks universality. By analyzing the closing price of dzgy, this paper uses the ARIMA-GARCH model to model the closing price and estimate the short-term stock price data, hoping to provide some reference for investors' investment plan.

References

[1] Wan Rui. Stock price volatility prediction based on GARCH model. Science & Technology Information, 2022, 20 (06): 129-132.

[2] Xu Shuya, Liang Xiaoying. Research on Stock Price prediction based on ARIMA-GARCH model. Journal of Henan Institute of Education (Natural Science Edition), 2019, 28(04): 20-24.

[3] Wang Yan. Applied Time Series Analysis. Beijing: Renmin University of China, 2008.

[4] Ding Lei, Guo Wanshan. Research on Gold Price prediction based on ARIMA-GARCH mixed model. Journal of Xuchang University, 2019, 38(06): 124-129.

[5] Yang Yuyuan, Zhang Mei. Empirical Analysis of Stock Price based on ARIMA Model. Science & Technology Information, 2021, 12 (29): 121-123.

[6] Sun Kaiying, Equity Return Modeling and Prediction Using Hybrid ARIMA-GARCH Model. International Journal of Financial Research, 2017, 8(3): 154-155