WEP Warwick Evans Publishing

International Journal of Computer Science and Information Technology

ISSN: 3005-9682 (Print), ISSN: 3005-7140 (Online) | Volume 3, Number 3, Year 2024
DOI: https://doi.org/10.62051/ijcsit.v3n3.49
Journal homepage: https://wepub.org/index.php/IJCSIT/index

Research on the Return Rate of Shenzhen Composite Index in my country——Based on ARMA and GARCH Models

Qianxia Peng

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

ABSTRACT

Stock price forecasting plays an important role in judging the trend of my country's stock market. The research object is the daily closing price of Shenzhen Composite Index from January 2, 2015 to December 30, 2020. First, the ARMA (2,4) model is established for the logarithmic return sequence after the first-order difference. Secondly, through analysis, it is concluded that the residual sequence has conditional heteroskedasticity. First, we can try to study the GARCH (1,1) model in depth. Next, a comprehensive ARMA (2,4)-GARCH (1,1) model is constructed, and relevant diagnostic tests are performed. The research results show that the logarithmic return sequence of Shenzhen Composite Index has volatility clustering, and the short-term prediction effect is better than the long-term prediction effect.

KEYWORDS

Shenzhen Composite Index; Logarithmic return sequence; Short-term prediction; ARMA-GARCH model

1. INTRODUCTION

The Chinese stock market is the most important and indispensable component of the international financial market, and is a major factor in promoting national economic development and global integration. Stock prices fluctuate every second, and the fluctuations of stock prices rising continuously are the manifestation of market power. It is of great significance to study the changes of stock prices rising continuously. In the past decade, the scale of stock market purchases has expanded rapidly over time, the market transparency has become higher, investors' trading ideas and products have become more homogeneous, and it has become more difficult to predict the overall price of the stocks they buy. In order to better predict the market trend of the total price of purchased stocks, it is indispensable to explore more reasonable and effective prediction methods.

The autoregressive conditional heteroskedasticity model proposed by Engel [1] can describe and analyze the singularity of financial accounting ROE in detail, and is a practical tool. Bollerslev [2] proposed that the general ARCH model is widely used in calculations, and the GARCH model can more effectively describe the dynamic external characteristics of different probability distributions under different conditions. Wan Wei [3] used GARCH and other models to study stocks purchased in the Shanghai and Shenzhen stock markets, and used dialectics to conduct a comparative study of actual yields, analyzing the existing yield rules and summarizing them. Huang Xuan et al. [4] constructed an ARMA-GARCH model to conduct an empirical analysis of the implied volatility of the CSI 500 Index and concluded that the ARMA-GARCH model can more effectively predict its future short-term implied volatility. Li Xiongying [5] fitted and predicted the daily a+b+c actual yield series of the four major state-owned banks with big data, and believed that the ARMA-GARCH model was not as good as the ARMA model in fitting effect. This paper constructs an ARMA-GARCH model to predict the yield of the Shenzhen Composite Index for reference by investors.

2. RESEARCH METHODS AND MODEL ESTABLISHMENT

2.1. Data Source and Processing

The data in this article comes from the official website of the Shenzhen Stock Exchange. The closing prices of the Shenzhen Composite Index from January 5, 2015 to December 30, 2020 are selected, totaling 1,342 data. In order to ensure that the instability caused by time series changes is eliminated as much as possible, the Shenzhen Composite Index is logarithmized in this article, and the logarithmic index yield can be obtained: Rt＝Ln(Pt)-Ln(Pt-1). Among them, Rt is the index yield on the tth day, and Pt is the index closing price of the Shenzhen Composite Index on the tth day.

2.2. ARMA Model

The ARMA model is a high-precision time series prediction model. By processing and analyzing historical time series data, and referring to and combining this dependent relationship to expand this law, it can discover the inherent law of a phenomenon changing over time. Predict how the phenomenon will change in the future. Generally speaking, the specific form of the ARMA model is as follows:

$$X_t = \varphi_1 X_{t-1} + \varphi_2 X_{t-2} + \cdots + \varphi_p X_{t-p} + a_t - \theta_1 a_{t-1} - \cdots - \theta_q a_{t-q} \tag{1}$$

In this model, the parameter p represents the order of the autoregressive part, q represents the order of the moving average part, and at represents the white noise sequence.

2.3. GARCH Model

Stock prices often have the characteristics of volatility clustering, "peaked and thick tails", and show heteroscedasticity. The GARCH model can be used to fit the conditional variance of the random error term. In general, the specific form of the GARCH (p, q) model is:

$$\sigma_t^2 = \alpha_0 + \alpha_1\mu_{t-1}^2 + \alpha_2\mu_{t-2}^2 + \cdots + \alpha_q\mu_{t-q}^2 + \beta_1\sigma_{t-1}^2 + \cdots + \beta_p\sigma_{t-p}^2 \tag{2}$$

Where p represents the order of the autoregressive term in the GARCH (p, q) model, and q represents the order of the ARCH term.

3. ANALYSIS OF EMPIRICAL RESULTS

3.1. Descriptive Statistics of Samples

First, the Eviews software is used to conduct a descriptive analysis of the logarithmic return time series of the Shenzhen Composite Index, and the logarithmic return time series chart of the Shenzhen Composite Index is obtained, as shown in Figure 1.

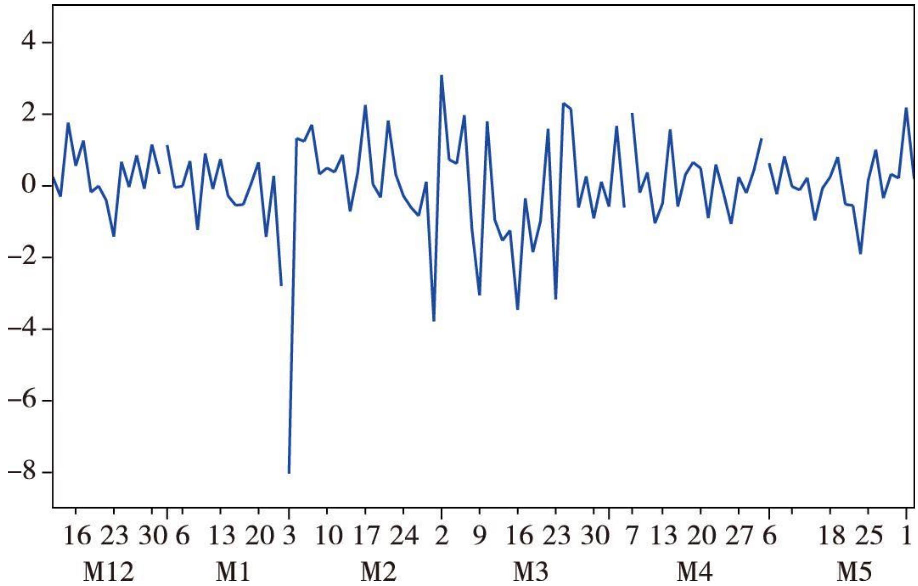
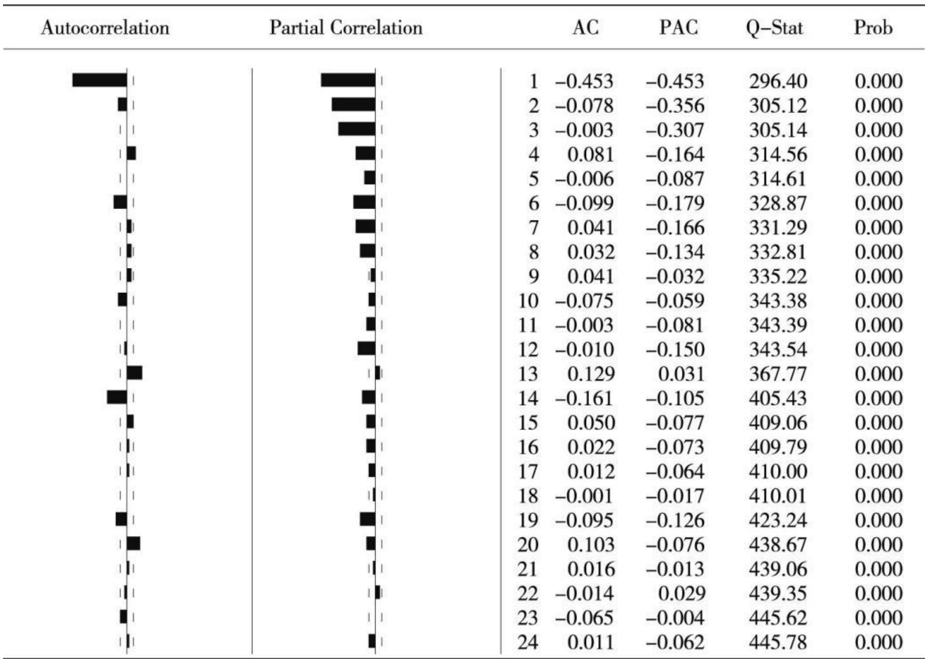


Figure 1. Shenzhen Composite Index Logarithmic Return Time Series

From the descriptive analysis results, the sample data has an average value of 0.013770, a standard deviation of 1.50687, a skewness of -31.2225 <0, and a kurtosis of 9.97098 >3 under normal distribution. It can be found that the sequence has a relatively obvious left skewness and does not conform to the standard normal distribution. From the logarithmic return figure 1, it can be seen that the time series is roughly in a relatively stable state. There are multiple abnormal peaks in the sequence. Small fluctuations are followed by smaller fluctuations, and large fluctuations are accompanied by larger fluctuations. There is an obvious fluctuation clustering phenomenon, indicating that the time series fluctuations show obvious conditional heteroskedasticity.

3.2. ARMA Model Establishment

Before building the ARMA model, in order to ensure the "consistency" requirement of statistical inference under large samples, it is necessary to test the stability of the sequence to determine whether the time series has a random trend or a certain trend. Since there may be high-order lag correlations in the time series, it is necessary to first use the ADF test to determine whether the sequence has a unit root. The ADF test results show that the t value of the ADF test statistic is -36.08224. The critical values of the t statistic at the test levels of 1%, 5%, and 10% are -3.434683, -2.863341, and -2.567777, respectively. The values of the t statistic are all less than the critical values at the test levels of 1%, 5%, and 10%. The null hypothesis is rejected. There is no unit root in the time series, that is, the logarithmic return series of the Shenzhen Composite Index after the first-order difference is stable, and the ARMA (p, q) model can be considered.



Autocorrelation	Partial Correlation		AC	PAC	Q–Stat	Prob
		1	–0.453	–0.453	296.40	0.000
		2	–0.078	–0.356	305.12	0.000
		3	–0.003	–0.307	305.14	0.000
		4	0.081	–0.164	314.56	0.000
		5	–0.006	–0.087	314.61	0.000
		6	–0.099	–0.179	328.87	0.000
		7	0.041	–0.166	331.29	0.000
		8	0.032	–0.134	332.81	0.000
		9	0.041	–0.032	335.22	0.000
		10	–0.075	–0.059	343.38	0.000
		11	–0.003	–0.081	343.39	0.000
		12	–0.010	–0.150	343.54	0.000
		13	0.129	0.031	367.77	0.000
		14	–0.161	–0.105	405.43	0.000
		15	0.050	–0.077	409.06	0.000
		16	0.022	–0.073	409.79	0.000
		17	0.012	–0.064	410.00	0.000
		18	–0.001	–0.017	410.01	0.000
		19	–0.095	–0.126	423.24	0.000
		20	0.103	–0.076	438.67	0.000
		21	0.016	–0.013	439.06	0.000
		22	–0.014	0.029	439.35	0.000
		23	–0.065	–0.004	445.62	0.000
		24	0.011	–0.062	445.78	0.000

Figure 2. Logarithmic return autocorrelation and partial autocorrelation

Table 1. ARMA model coefficients

Each model	ARMA(1,4)	ARMA(2,4)
AR(1)	2.65E-05	0.134664
AR(2)	0	-0.943689
MA(1)	0.051444	-0.084857
MA(2)	-0.027258	0.928522
MA(3)	0.012816	0.077569
MA(4)	0.07186	0.043442

Next, the autocorrelation and partial autocorrelation graphs of the logarithmic return time series of the Shenzhen Stock Exchange Composite Index are drawn, as shown in Figure 4. It can be seen that the autocorrelation function of the series shows statistically significant differences at the lag 1st and 14th orders. sharp column, while the performance at other levels is not statistically significant. The partial autocorrelation function significantly exceeds the 95% confidence area at lags 1, 2, 3, and 6. It is initially determined that the moving average process of the ARMA model is fourth order, while the autoregressive process should be low order.

Combined with the AIC information criterion to select the best model, two models, ARMA (2, 4) and ARMA (1, 4), were estimated and compared respectively. The results are shown in Table 1. Comparing the two models, it can be seen that the values of AIC and SC corresponding to ARMA (2, 4) are smaller than those corresponding to ARMA (1, 4) and the goodness of fit is higher. Finally, the ARMA (2, 4) model is constructed:

$$X_t = 0.1347X_{t-1} - 0.9737X_{t-2} + a_t - 0.08486a_{t-1} + 0.9285a_{t-2} + 0.07757a_{t-3} + 0.004344a_{t-4} + 0.02118 \tag{3}$$

3.3. GARCH Model Establishment

On the basis of establishing the ARMA model, it is necessary to further determine whether the time series has conditional heteroskedasticity. The article generates the residual sequence based on the establishment of the mean model, and draws the autocorrelation and partial autocorrelation plots of the squared residuals. It can be seen from Table 2 that most of the residual squared autocorrelation function values exceed the 95% confidence area and are statistically significantly different from 0, and the probability values corresponding to the Q statistics are all less than 0.05, that is, the residual squared sequence exists There is an ARCH effect in the autocorrelated residual sequence. Since there is a higher order ARCH effect, it is necessary to determine the appropriate model order based on the AIC criterion. By setting the order of the GARCH model multiple times and fitting it multiple times, we can calculate each The AIC values corresponding to the model are shown in Table 2.

Table 2. AIC values of each GARCH model

GARCH model	(1,1)	(1,2)	(2,1)	(2,2)
AIC value	3.1125	3.1134	3.1168	3.1133

From Table 2, it can be concluded that the AIC value corresponding to the GARCH (1,1) model is the smallest. Combined with the AIC optimal criterion, the GARCH (1,1) model is finally constructed, and the estimation results of the parameters of the GARCH (1,1) model are obtained, as shown in Figure 4. From Figure 7, it can be seen that the yield volatility model is GARCH (1,1):

$$\sigma_t^2 = 0.006224 + 0.067589\mu_{t-1}^2 + 0.932037\sigma_{t-1}^2 \tag{4}$$

3.4. Establishment and Analysis of the ARMA-GARCH Comprehensive Model

When establishing the ARMA model to predict the changes in the logarithmic returns of stocks, the ARCH effect of the residual square term in the model is ignored. It is also incomplete to establish a model based only on volatility, and the establishment of the GARCH model needs to be based on the ARMA mean equation. Therefore, the article finally establishes an ARMA-GARCH comprehensive model to estimate and predict the logarithmic returns of the Shenzhen Stock Exchange Index. The model form is as follows:

Table 3. Results of Data Processing and Analysis

Variable	Coefficient	Std.error	z-Statistic	Prob.
C	0.022050	0.024502	0.899921	0.3682
Variance Equation				
C	0.006224	0.001680	3.705591	0.0002
RESID(-1)^2	0.067589	0.006077	11.12226	0.0000
GARCH(-1)	0.932037	0.004952	188.2162	0.0000

Compared with the autocorrelation and partial autocorrelation functions of the ARMA (2, 4) model, the autocorrelation and partial autocorrelation functions of the ARMA-GARCH model's residual square series are significantly reduced, and the ARMA-GARCH model corresponds to At the 5% significance level, the probability value p of the q statistic is much greater than 0.05. The q statistic is not significant. There is no serial correlation in the residual square sequence. The residual sequence of this model effectively eliminates ARCH effect, so the ARMA (2, 4) GARCH (1, 1) model is established.

3.5. Model-related Diagnostic Tests

After completing the establishment of the above model, it is necessary to verify whether the constructed ARMA-ARCH optimization model still has heteroskedasticity. If it exists, the model is incorrect. If it does not exist, the model is correct. Therefore, this article conducts an autocorrelation test on the residual square sequence in the Shenzhen Stock Exchange Composite Index logarithmic return model and finds that the sample autocorrelation function of the residual sequence of the ARMA (2, 4)-GARCH (1, 1) model is the corresponding p values are all much greater than 0.9, much greater than the test level, so the null hypothesis cannot be rejected. There is no autocorrelation in the residual sequence. That is, it can be considered that the ARMA-ARCH optimization model constructed in the article does not have heteroskedasticity and the model is reasonable.

3.6. Prediction Results

The average relative error of the ARMA-GARCH model is obtained through the EExcel table. First, the relative error of each data is calculated, which is numerically equal to the ratio of the absolute value of the difference between the actual value and the predicted value and the actual value. Then the average value is taken to obtain the model. The average relative error is finally squared to obtain an MSE of 0.037.

4. CONCLUSION

Forecasting stock returns is an important research direction in financial research and has received increasing attention in recent years. The article first uses the ARMA model to fit the time series formed by 1342 selected logarithmic return data of the Shenzhen Stock Exchange Index to establish the ARMA (2, 4) model. Since the random disturbance terms in the model still have obvious conditional heteroskedasticity , by testing the graphical test method, the autocorrelation function test method to test whether the ARCH effect exists, and estimating the model parameters to further establish a GARCH (1, 1) model to reflect the fluctuation of returns. Finally, it is considered that building a model based only on volatility is not comprehensive. , and the need for the GARCH model is based on the ARMA mean equation, so the ARMA (2, 4)-GARCH (1, 1) comprehensive optimization model was finally established, and the model was used to predict the January 2, 2015 As of December 30, 2020, the average relative error MSE value of the logarithmic return forecast of the Shenzhen Stock Exchange Index closing price is 0.0372. Although the model can predict changes in stock prices or rates of return to a certain extent, because stock prices in the stock market are always fluctuating, there are many factors that affect the stock market, such as changes in the international environment, government policies and regulations, and changes in the industry. And the economic cycle, etc., the effect of its forecast will definitely be affected. Therefore, the short-term prediction effect is better than the long-term prediction. The prediction model of the stock market must be constantly updated. At the same time, the influence of some objective environmental factors must also be considered when making predictions.

REFERENCES

[1] Engle R F. Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation [J]. Econometrica: Journal of the econometric society, 1982: 987-1007.

[2] Bollerslev T. Generalized autoregressive conditional heteroskedasticity [J]. Journal of econometrics, 1986, 31(3): 307-327.

[3] Wan Wei, Jiang Xiaogan. Research on the volatility of my country's Shanghai and Shenzhen stock markets - based on the GARCH family model [J]. Value Engineering, 2007 (10): 14-18.

[4] Huang Xuan, Zhang Qinglong. Analysis and prediction of volatility of Shanghai and Shenzhen 300 Index based on ARMA-GARCH model [J]. China Prices, 2018, No.350(06): 44-46.

[5] Li Xiongying, Chen Xiaoling, Zeng Kaihua. Research on stock yield prediction of the four major banks based on three types of models [J]. Economic Mathematics, 2018, 35(04): 21-27. DOI: 10.16339/j.cnki.hdjjsx.2018.04.023.