

Research on Forecasting the Shanghai and Shenzhen 300 Index Based on the ARIMA-GARCH Model

Changyong Xie^{1, a}

¹College of Management, Shenzhen University, Shenzhen, 518055, China

^a2020041132@email.szu.edu.cn

Abstract: This paper aims to forecast the closing prices of the Shanghai and Shenzhen 300 Index using time series models. By combining the ARIMA and GARCH models, the study aims to improve the accuracy of short-term predictions for the index and enhance risk management capabilities. The research begins by collecting and preprocessing historical data of the Shanghai and Shenzhen 300 Index. Then, an ARIMA-GARCH model is established, and the model parameters are estimated and tested. Finally, the model is used to forecast the closing prices of the index, and the accuracy of the predictions is evaluated. The results demonstrate that the constructed ARIMA-GARCH model can provide accurate predictions for the closing prices of the Shanghai and Shenzhen 300 Index. This research is expected to enhance prediction accuracy, improve risk management, increase practical value, and serve as a reference for similar studies.

Keywords: ARIMA model, GARCH model, Shanghai and Shenzhen 300 Index, short-term forecasting, stationarity test.

1. Introduction

Accurate short-term stock market predictions and risk assessment are vital in today's dynamic financial landscape. This study focuses on the practical application and effectiveness of the ARIMA-GARCH model in forecasting the Shanghai and Shenzhen 300 Index.

By combining the strengths of the ARIMA and GARCH models, this approach enhances the capture of market volatility and nonlinear characteristics. The study aims to enrich time series analysis methods in finance and provide a theoretical foundation for future research. In practical terms, accurate short-term forecasts of the index can assist investors in making informed decisions, optimizing portfolios, and mitigating risks. The volatility forecasts provided by the ARIMA-GARCH model enable effective risk assessment and control. Financial institutions can leverage these predictions to enhance asset allocation, improve investment decisions, regulate markets, protect investors, and promote overall financial market development.

2. Literature Review

Financial market forecasting has gained significant attention in both academic and practical domains, with the ARIMA-GARCH model being widely applied and studied, particularly in stock price forecasting.

In China, researchers have made notable advancements in utilizing time series analysis for financial data modeling. Zha Zhenghong introduced the ARIMA model to analyze the Shanghai Stock Composite Index in 1999[1]. Hui Xiaofeng et al. applied the GARCH model to forecast the RMB-USD exchange rate[2]. Liu Yan et al. explored the integration of ARIMA and GARCH models by employing the ARIMA-GARCH model for spot electricity price prediction[3]. These domestic studies signify the progress in applying time series analysis to financial data forecasting in China.

Internationally, the ARIMA-GARCH model has been extensively studied and applied in stock price forecasting. Engle introduced the ARCH and GARCH models[4], while

Bollerslev further developed the GARCH model[5]. Ding, Granger, and Engle proposed the ARIMA-GARCH model for modeling stock market returns[6], and Mohammadi et al. applied the ARIMA-GARCH model to forecast crude oil spot prices[7]. The model has also been used in other domains such as tourism forecasting by Chhorn [8] and network traffic prediction by Zhou[9]. These international studies highlight the effectiveness of the ARIMA-GARCH model in different markets.

Overall, the ARIMA-GARCH model has shown promise in financial forecasting, particularly in stock price prediction. The domestic research in China has contributed to the advancement of time series analysis for financial data modeling, while international studies have showcased the model's effectiveness in various markets. However, continuous improvement and development of more accurate and reliable prediction models are necessary to meet the evolving demands of the financial market.

3. Theoretical Basis

3.1. Principle of the ARIMA Model

The ARIMA model, developed by Box and Jenkins, is a stochastic time series model known as the Box-Jenkins model. It uses autocorrelation and partial autocorrelation functions to analyze data randomness and make future trend predictions. Represented as ARIMA(p, d, q), the model adjusts parameters (p, d, q) to capture dataset characteristics and randomness. The ARIMA(p, d, q) model can be represented by the equation:

$$dX_t = c + \sum_{i=1}^p \phi_i dX_{t-i} + \varepsilon_t + \sum_{j=1}^q \theta_j \varepsilon_{t-j} \quad (1)$$

Where c is a constant, p, d, q represents the autoregressive order, differencing order, and moving average order, respectively. X_t represents the time series at time t, ϕ_i and θ_j are the stationary polynomial operators of order p and q, respectively, and ε_t denotes the residual term.

3.2. Principle of the GARCH Model

In 1986, Bollerslev developed the GARCH model by incorporating lag order from the ARCH model. The GARCH model effectively addresses the issue of "fat-tailed" distribution in time series data, resulting in enhanced forecasting accuracy. It is widely used for analyzing financial data and making decisions related to asset allocation, hedging, risk management, and portfolio optimization. The GARCH(m, n) model can be represented as follows:

$$\sigma_{t=\alpha_0}^2 + \sum_{i=1}^u \alpha_i \sigma_{t-i}^2 + \sum_{j=1}^v \rho_j \varepsilon_{t-j}^2 \quad (2)$$

Subject to the following constraints:

$$\alpha_0 > 0, \alpha_i \geq 0, \beta_i \geq 0, i = 1, 2, \dots, u; j = 1, 2, \dots, v \quad (3)$$

3.3. Principle of the ARIMA-GARCH Model

The ARIMA-GARCH hybrid model combines the ARIMA model and the GARCH model and can be represented as follows:

$$dX_t = \phi_0 + \sum_{i=1}^p \phi_i dX_{t-i} + u_t + \sum_{j=1}^q \theta_j u_{t-j} \quad (4)$$

$$u_t = \sigma^t \varepsilon^t \quad (5)$$

$$\sigma_{t=\alpha_0}^2 + \sum_{i=1}^u \alpha_i \sigma_{t-i}^2 + \sum_{j=1}^v \rho_j \varepsilon_{t-j}^2 \quad (6)$$

Where $0 < \sum_{i=1}^u \alpha_i + \sum_{j=1}^v \beta_j < 1$, ε_t represents the random error term, and u and v denote the orders of the GARCH model and the ARCH model, respectively. σ_t^2 represents the conditional variance of ε^t , reflecting the volatility characteristics of the time series.

4. Empirical Study

4.1. Research Method

This study utilizes the ARIMA-GARCH model to forecast short-term trends in the CSI 300 Index. The ARIMA model captures and predicts the trend component, while the GARCH model models and predicts the volatility component. By considering both trend and volatility characteristics, this combined model improves forecasting accuracy.

4.2. Experimental Steps

1. Data Collection and Preprocessing: Historical data of the CSI 300 Index is collected from official websites and undergoes cleaning and preprocessing, including handling missing values, outlier treatment, data smoothing, and normalization.

2. Building the ARIMA-GARCH Model for Stock Price Forecasting: The correlation between time series models is tested to examine the existence of autocorrelation among the data.

3. Stationarity Testing of Stock Price Time Series using ADF Unit Root Test.

4. Non-stationary series are differenced to transform the original stock price series into a stationary series.

5. Model order is determined and parameter estimation is conducted using the AIC criterion.

6. ARCH test is performed on the selected ARIMA model

to determine the presence of heteroscedasticity. If heteroscedasticity is detected, a GARCH model needs to be established.

7. Residuals of the developed ARIMA-GARCH model are tested for a white noise process. If the criteria are not met, further model improvements are needed.

8. Using the established ARIMA-GARCH model, the closing prices of the CSI 300 Index are forecasted, and the relative errors between predicted and actual values are calculated. Accuracy is evaluated and compared with the ARIMA model.

9. Using the finalized model, the closing prices of the CSI 300 Index are predicted, and the relative errors between the predicted values and the actual values are calculated.

4.3. Data Source

This article is based on the closing price data of the CSI 300 Index (000300) from February 9, 2015, to April 28, 2023. A stock price forecasting model is established using this data. For testing purposes, the closing prices for a total of 20 days from May 4, 2023, to May 31, 2023, are selected as the test data for predicting the closing prices.

4.4. Data Preprocessing

The results of the Durbin-Watson test are shown in Table 1.

Table 1. Durbin-Watson test results

Statistic	DW Statistic	p-value
Durbin-Watson	1.9167	0.02838

The DW statistic is 1.9167, and the p-value is 0.02838. Based on the range of the DW statistic, a value close to 2 implies the possible presence of positive autocorrelation. Since the p-value is less than the significance level (usually 0.05), we can reject the null hypothesis of zero autocorrelation and accept the alternative hypothesis of positive autocorrelation. Overall, based on the results of the DW test, it suggests that this time series data may exhibit positive autocorrelation, indicating that the current observations may be influenced by lagged observations.

4.5. Stationarity Test

After the stationarity test, the time series was evaluated using the ADF test. The ADF test is a commonly used method to determine the stationarity of a time series. If a non-stationary time series has unit roots, differencing can be used to eliminate the unit roots and obtain a stationary series. The results of the ADF test are shown in Table 2.

Table 2. Augmented Dickey-Fuller test results

Test Statistics	Results
Dickey-Fuller Statistics	-2.2951
Lag order	12
p-value	0.4534

The Dickey-Fuller statistic value is -2.2951 (p-value = 0.4534) with a lag order of 12. The ADF test was conducted assuming the series is stationary. Based on the test results and the p-value, the following conclusions can be made: The Dickey-Fuller statistic value is smaller than the critical value ($-2.2951 < -2.885$), indicating insufficient evidence to reject the null hypothesis of unit roots in the time series. Moreover,

the p-value (0.4534) is greater than the significance level (usually 0.05), indicating that the series is non-stationary. To address this, differencing was applied to the series. The `adfTest` function confirmed that the first-order differenced series has a p-value of 0.01, which is lower than the significance level of 0.05, indicating that the differenced series is stationary.

4.6. Model Selection

The `auto.arima` function is used to automatically select the

order of the ARIMA model based on the time series data and provides suggested model orders. According to the output, the suggested model order is ARIMA(2,1,3), where 2 represents the autoregressive (AR) order, 1 represents the differencing order, and 3 represents the moving average (MA) order.

Furthermore, the AIC criterion is used to determine the model order. A lower AIC value indicates a better fit of the model. The AIC values for models such as ARIMA(0,1,1), ARIMA(1,1,0), and ARIMA(1,1,1) are calculated and the results are as follows:

Table 3. AIC values of models with different orders

Model	AIC
ARIMA(0,1,1)	22158.142652579
ARIMA(0,1,2)	22154.0783469131
ARIMA(1,1,0)	22158.5709645317
ARIMA(1,1,1)	22157.1845689969
ARIMA(1,1,2)	22156.0656794051
ARIMA(2,1,0)	22153.46743471
ARIMA(2,1,1)	22155.4528326974
ARIMA(2,1,2)	22113.6034106281
ARIMA(2,1,3)	22108.3001928576

Based on the table results, the ARIMA(2,1,3) model has the lowest AIC value. Two white noise tests were conducted on the residuals of this model: one for the residual sequence and another for the squared residual sequence. The results indicate that both p-values are greater than 0.05, suggesting that the residuals conform to the assumption of white noise. This implies that there is no significant autocorrelation or heteroscedasticity in the residual sequence, aligning with the model's basic assumptions.

Furthermore, an ARCH test was performed on the standardized residuals of the ARIMA(2,1,3) model. The results in Table 4 reveal a p-value significantly smaller than 2.2×10^{-16} , rejecting the hypothesis that the squared residual sequence is white noise. This indicates the need to incorporate a conditional heteroscedasticity model, such as the GARCH model. Considering the potential instability introduced by higher order GARCH models, this study selects GARCH(1,1) to capture the residual dynamics.

Table 4. ARCH effect test

X-squared	P	Whether there is an ARCH effect
72.868	P-value $\ll 2.2 \times 10^{-16}$	Yes

4.7. Model Parameter Estimation

For the ARIMA(2,1,3)-GARCH(1,1) model, the estimated

values of the parameters are as follows:

Table 5. Parameter estimation results of ARIMA(2,1,3)-GARCH(1,1) model

Parameter	Estimate
Mu	3675.933
Ar1	0.02322135
Ar2	0.9711691
Ma1	0.9743554
Omega	17.21428
Alpha1	0.08904022
Beta1	0.9099598

4.8. Residual Test

The LM test was conducted on the obtained ARIMA-GARCH model to examine the independence of residuals. The results are shown in Table 6.

of no autocorrelation in the residual sequence. Hence, we can conclude that the ARIMA model in this case is adequate and the residuals do not show significant autocorrelation over time.

Table 6. LM Test results of residual sequences

LM Statistics	P
0.002427	0.9607

4.9. Forecast Results

Using the established ARIMA(2,1,3)-GARCH(1,1) model, the closing prices of the CSI 300 index are predicted for the next 20 days. The results are shown in Table 7.

Based on the p-values, we cannot reject the null hypothesis

Table 7. Forecast results of CSI 300 Index closing price using ARIMA(2,1,3)-GARCH(1,1)

Number	Date	Actual Value	Predicted Value	Relative Error
1	2023-5-4	4030.247	4028.013	-0.000554388
2	2023-5-5	4016.877	4027.08	0.002539897
3	2023-5-8	4062.662	4026.016	-0.009020266
4	2023-5-9	4027.875	4025.085	-0.000692679
5	2023-5-10	3996.869	4024.031	0.006795614
6	2023-5-11	3990.664	4023.102	0.008128622
7	2023-5-12	3937.762	4022.056	0.021406639
8	2023-5-15	3998.887	4021.131	0.005562384
9	2023-5-16	3978.205	4020.093	0.010529359
10	2023-5-17	3960.165	4019.17	0.014899646
11	2023-5-18	3956.07	4018.141	0.015690124
12	2023-5-19	3944.54	4017.221	0.018425585
13	2023-5-22	3969.328	4016.2	0.011808704
14	2023-5-23	3913.188	4015.283	0.026089825
15	2023-5-24	3859.093	4014.27	0.040210676
16	2023-5-25	3850.5	4013.355	0.042294743
17	2023-5-26	3850.951	4012.351	0.041911723
18	2023-5-29	3833.942	4011.439	0.046296422
19	2023-5-30	3837.753	4010.443	0.044997598
20	2023-5-31	3798.543	4009.534	0.055545185

To calculate the accuracy of the data, we can calculate the Mean Absolute Percentage Error (MAPE) based on the given values. Using the given data, we can calculate the MAPE is approximately 2.0682%.

On the other hand, using only the ARIMA(2,1,3) model to predict the closing prices of the CSI 300 Index for 20 days yields the results shown in Table 8.

Table 8. Forecast results of CSI 300 Index closing price using ARIMA(2,1,3)

Number	Date	Actual Value	Predicted Value	Relative Error
1	2023-05-04	4030.247	4028.121	-0.0005274776
2	2023-05-05	4016.877	4029.128	0.0030498446
3	2023-05-08	4062.662	4033.084	-0.0072804594
4	2023-05-09	4027.875	4032.607	0.0011746892
5	2023-05-10	3996.869	4028.766	0.0079802605
6	2023-05-11	3990.664	4028.751	0.0095440335
7	2023-05-12	3937.762	4032.422	0.0240389666
8	2023-05-15	3998.887	4032.887	0.0085022153
9	2023-05-16	3978.205	4029.433	0.0128771312
10	2023-05-17	3960.165	4028.565	0.0172720629
11	2023-05-18	3956.070	4031.761	0.0191327790
12	2023-05-19	3944.540	4032.983	0.0224215086
13	2023-05-22	3969.328	4030.077	0.0153047655
14	2023-05-23	3913.188	4028.552	0.0294808265
15	2023-05-24	3859.093	4031.143	0.0445829530
16	2023-05-25	3850.500	4032.920	0.0473756466
17	2023-05-26	3850.951	4030.660	0.0466660910
18	2023-05-29	3833.942	4028.684	0.0507943765
19	2023-05-30	3837.753	4030.602	0.0502505778
20	2023-05-31	3798.543	4032.727	0.0616508618

The MAPE for the given data is approximately 2.4833%.

By comparing the results, we can see that the ARIMA-GARCH model achieves higher accuracy in predicting the closing prices of the CSI 300 index compared to using the ARIMA model alone. This indicates that the trained model is capable of effectively forecasting the closing prices of the CSI 300 index, highlighting the potential application of the ARIMA-GARCH model in financial data prediction.

5. Conclusion

This study utilized ARIMA-GARCH modeling on the CSI 300 index time series data and performed model evaluation and prediction analysis. The Durbin-Watson test revealed positive autocorrelation, while the ADF test indicated possible non-stationarity of the series. The ARIMA(2,1,3) model was selected for forecasting after model selection. Both the white noise test and LM test confirmed that the residual sequence of the model met the assumptions and

required no further adjustments. By predicting the test data, we obtained forecasted closing prices, evaluated their accuracy, and compared them with the ARIMA model. The comprehensive findings demonstrate the ARIMA-GARCH model's accuracy and feasibility in short-term forecasting, providing investors with valuable reference and decision-making guidance.

References

- [1] Zha, Z. (1999). Statistical analysis and forecasting of the Shanghai Stock Composite Index. *Journal of Shanghai Maritime University*, 04:82-89.
- [2] Hui, X., Liu, H., Hu, W., et al. (2003). Forecasting RMB-USD exchange rate based on time series GARCH model. *Financial Research*, 05, 99-105.
- [3] Liu, Y., Xing, W., Ding, L., et al. (2012). Spot electricity price forecasting based on ARIMA-GARCH model. *Energy Technology and Economics*, 24(02), 59-63.
- [4] Engle, R. F. (1982). Autoregressive conditional heteroskedasticity with estimates of the variance of United Kingdom inflation. *Econometrica*, 50(4), 987-1007.
- [5] Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of econometrics*, 31(3), 307-327.
- [6] Ding, Z., Granger, C. W., & Engle, R. F. (1993). A long memory property of stock market returns and a new model. *Journal of empirical finance*, 1(1), 83-106.
- [7] Mohammadi, H., & Su, L. (2010). International evidence on crude oil price dynamics: Applications of ARIMA-GARCH models. *Energy Economics*, 32(5), 1001-1008.
- [8] Chhorn, T. (2018). Tourism demand and exogenous exchange rate in Cambodia: A stochastic seasonal arimax approach. *Theoretical and Practical Research in Economic Fields (TPREF)*, 9(17), 5-16.
- [9] Zhou, B., He, D., & Sun, Z. (2006). Traffic modeling and prediction using ARIMA/GARCH model. In: *Modeling and Simulation Tools for Emerging Telecommunication Networks*, Springer. pp.101-121.