The Application of the ARIMA-GARCH Hybrid Model for Forecasting the Apple Stock Price

Haoqing Jiang
Wuhan Britain-China School, Wuhan, Hubei, China

Abstract: Modeling and forecasting stock prices is a meaningful task and one of the methods for forecasting is the classic ARIMA models. However, when the data exhibits clustering effects and heteroscedasticity, the generalized auto regressive conditional heteroskedatic (GARCH) model must be used for modeling and forecasting. In this paper, as the object of data analysis, the combination of ARIMA model and GARCH model shows a very good ability to predict the stock price with a very good description of the clustering effect of volatility.

Keywords: Apple Stock Price; Heteroskedatic; Clustering Effects; ARIMA-GARCH Model; Forecasting

1. Introduction

Many residents will invest in the stock of the Apple. There are several reasons why Apple is widely chosen as an investment target. Firstly, Apple's stock price has generally risen. Secondly, trading is convenient. Apple's stock issuance is large and trading is very frequent, making it easy to sell its holdings. However, the overall rise of stocks does not guarantee that short-term holders will make a profit, as there will be fluctuations in stocks in the short term.

Various empirical studies have analyzed the use of the ARMI-GARCH model to measure and predict property price fluctuations for different products. For example, Ariyo et al. [1] describe a specific process by which the most appropriate ARIMA model can be found to predict stock prices. Kamruzzaman et al. [2] calculated stock returns based on the relative difference method, but unfortunately they were only limited to the ARIMA model and did not carry out heteroscedasticity modeling. In addition, Abbasi et al. [3] applied the linear combination of this sequence to the cement industry and proposed in their case analysis that the ARMA(1,1) model using second-order difference is the optimal model for predicting cement stock prices. Fang Yan [4] uses the ARMI-Garch model to model the media sector index and make short-term predictions, and finds that the model is also meaningful if it is extended to individual stocks. Feng Tieying [5] proposed an ARMI-GARch modeling and simulation method based on the combination of stochastic theory and time series analysis, which can effectively reduce the prediction error.

The structure of this paper is as follows: Section 1 briefly introduces Apple stock price and its investment value, and gives a brief overview of the ARIMA-GARCH model. Section 2 introduces the modeling methods of ARIMA and GARCH. In addition, the model order determination and parameter estimation are described in detail. Section 3 describes the sources of the relevant data sets and provides predictions of stock prices. Section 4 summarizes the conclusion and puts forward some suggestions for the future work.

2. Materials and Methods

2.1 The Arima Model

ARIMA is one of the more successful models for studying time-dependent series and forecasting. In this model, the variables are assumed to be linear combinations of occurred values and past errors. Based on the Box-Jenkins method, the fitting step of the ARIMA model requires two assumptions about the data: the stationarity of the autoregressive (AR) model and the moving average (MA).

The ARIMA(p, d, q) can be expressed as

$$\Phi(B)(1-B)^d x_t = \Theta(B)\varepsilon_t \quad (1)$$

Where $\Phi(B) = 1 - \sum_{i=1}^{p} \phi_i B^i$ and $\Theta(B) = 1 - \sum_{j=1}^{q} \theta_j B^j$ are polynomials in term of B of degree p and q. x_t denotes the stationary time series and ε_t is the random error at time period t, d is a non-negative integer representing the difference order. $\Delta^d x_t = \Delta^{d-1} x_t - \Delta^{d-1} x_{t-1}$, and B is the back shift operation, $\Delta x_t = x_t - x_{t-1} = (1-B)x_t$. $\phi_i, i = 1,2,\cdots,p$ and $\theta_j, j =$

$1,2,\cdots,q$ are the parameters of auto-regressive and moving average term.
The ARIMA(p,d,q) model essentially involves difference non-stationary data to obtain stationary data, and using the ARMA(p,q) model for modeling.

2.2 Model Identification

Model order identification refers to determining the order of the model, that is, the values of p, d, and q. By studying the time series, the order of the difference can be determined, as well as the stationarity test, also known as the unit root test of the Augmented Dickey-fuller (ADF) test. the ACF and PACF plots is employed to determine the value of p and q.
If the time series plot has a linear or parabolic trend, d=1 or d=2 for difference can be used, respectively. We should avoid over differencing, because over differencing can cause the standard deviation to increase.
The Bayesian information criterion (BIC) and Akaike information criterion (AIC) are the important methods for determining p and q values, the smallest BIC or AIC for the most appropriate p,q. The formula for calculating BIC and AIC can be expressed as follows

$$\mathrm{BIC} = -\log(L) + n\log(m) \quad (2)$$

$$\mathrm{AIC} = -\log(L) + 2n \quad (3)$$

L is the maximum likelihood function when the error assumption is normal distribution and the ARMA model coefficients are determined. n and m are the number of parameters and the number of training samples of the model, respectively.

2.3 Estimation of Parameters

In this step, Sample autocorrelation function (ACF) and partial autocorrelation function (PACF) are important methods to judge the order of a model. In general, maximum likelihood estimation (MLE) is a commonly used parameter estimation method.

2.4 Diagnostic Testing of the Model

Box-Jenkins model diagnosis requires the assumption that the error term of a proper model should also be stationary. Only when the relevant conditions are met can the model be used for prediction. A normal Q-Q plot is usually used to test whether white noise satisfies the normality hypothesis.

2.5 The GARCH Model

Generalized Autoregressive Conditional Heteroskedasticity (GARCH) is a development of the ARCH model and is generally combined with the ARMA model.

$$\begin{cases} x_t = f(t, x_{t-1}, x_{t-2}, \cdots) + \varepsilon_t \\ \varepsilon_t = \sqrt{h_t} e_t \\ h_t = w + \sum_{i=1}^{r} \eta_i h_{t-i} + \sum_{j=1}^{s} \lambda_j \varepsilon_{t-j}^2 \end{cases} \quad (4)$$

Where $f(t, x_{t-1}, x_{t-2}, \cdots)$ is deterministic information fitting model for $\{x_t\}$. e_t follows the independent and identically norm distribution, with mean equates to 0 and variance equates to σ^2. h_t used as modeling the conditional variance of x_t, $w > 0$ and $\sum_{i=1}^{\max(r,s)}(\eta_i + \lambda_i) < 1$. Note that η_i and λ_i are the coefficients of the parameters ARCH and GARCH, respectively. ACF and PACF of the residuals help to specify the GARCH orders, r and s, respectively.

3. Modeling of Apple Stock Price

3.1 Data of Study

In our paper, a total of 1005 daily Apple stock price data series in USD is used from August 1 2019 to July 28 2023 of 5-day-per-week frequencies. The data is divided into two parts :(i) the first 95% of the sample is used for the trained model; (ii) The last 5% of the sample is used for testing to verify the merits of the model.

3.2 Model Identification

The first step of identification is to judge the trend of the sequence and determine whether the price movement is seasonal by drawing the sequence diagram of the sample, as shown in Figure 1. As can be seen from the figure, the stock price series has an increasing trend. Meanwhile, through the QQ-norm plotting (Figure 2), it can be found that the original sequence is not white noise.
That after performing first-order difference on the sequence and carried out with the Augmented Dickey-Fuller (ADF) test, according to the p-value (0.01) of the test, the sequence can be considered stationary.

3.3 Parameter Estimation and Selection of the Best Model

After observing the ACF diagram (Figure. 3a), it is found that if the delay is of order 1, the value of the phase relation does not exceed the lower limit of 5%, which can be preliminatively judged that the sequence has the characteristics of MA model and is of order 1, and the value of the phase bias relation does not exceed the lower limit of 5% (Figure. 3b). According to ACF and PACF, the tentative model can be ARIMA(0,1,1). Parameters estimation results of ARIMA(0,1,1) are shown in Table 1

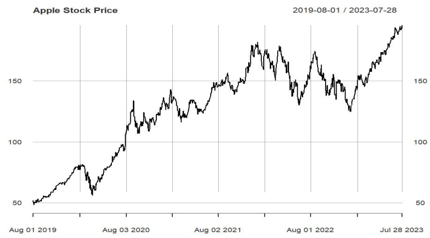


Figure 1. Plot of Apple Stock Price

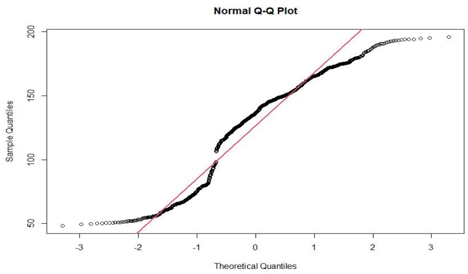


Figure 2. QQ Plot of Original Data

Table 1. Regression Result of ARIMA(0,1,1) model

Call: ARIMA (0,1,1)		
parameters	MA1	mean
coefficients	-0.0625	0.1431
S.E.	0.0327	0.0755
Sigma^2= 6.521: log-likelihood =-2364.89 AIC=4735.78 AICc=4735.8		

Because the squared residual in ARIMA model has conditional heteroscedasticity, GARCH model is needed to simulate and explain conditional heteroscedasticity. According to the above analysis, we combine ARIMA(0,1,1)-GARCH(1,1) model to fit the Apple stock price and the parameters estimation is shown in Table 2.

From the regression results, we can get the final estimation model:

$$x_t = x_{t-1} + \varepsilon_t + 0.065\varepsilon_{t-1} \quad (5)$$

$$\varepsilon_t = \sqrt{h_t} e_t \quad (6)$$

$$h_t = 5.521 + 0.1352\varepsilon_{t-1}^2 + 0.4602h_{t-1} \quad (7)$$

Table 2. Regression Results of ARIMA-GARCH Model

parameters	Estimate	Std.Error	t-value	Prob.(>|z|)
C	5.521	1.422	3.89	5.99e-05*
RESID(-1)^2	0.1352	0.03.182	4.913	2.37e-05*
GARCH(-1)	0.4602	0.03953	3.020	0.0211*

LM is conducted to test if the residual of the ARIMA-GARCH model is relevant. The results in Table 3 show that the residual sequence conforms to white noise, that is, the fitting error considers that H_0 of white noise cannot be rejected.

Table 3. Heteroscedasticity Test Result

F-statistic	0.1382	p-value	0.6521
R-squared	0.12844	p-value	0.7217

3.4 Forecasting

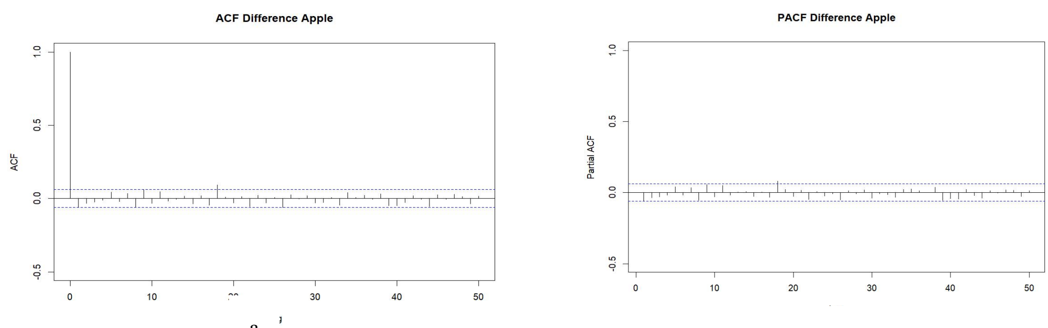
Figure 4, 5and 6 shows the results of the fitted forecast for Apple's daily stock price using ARIMA(0,1,1)-GARCH(1,1). It can be seen from the results that the mixed model fits well in predicting the price series, that is, the trend of the predicted price is closely related to the actual data, including the simulation part of 5% of the external sample.

4. Conclusion

Through the data analysis and modeling of stock prices, we can draw the following conclusions:

(1) In general, if the non-stationary sequence has short-term correlation, a satisfactory result can be obtained by using the ARMI-GARCH model.

(2) The method presented in this paper is suitable for the modeling of stock prices affected by random effects, if there are significant policy changes, the sample data will change, the model is no longer suitable. Therefore, it is necessary to pay attention to the degree of market volatility..



a b

Figure 3. Plot of ACF and PACF

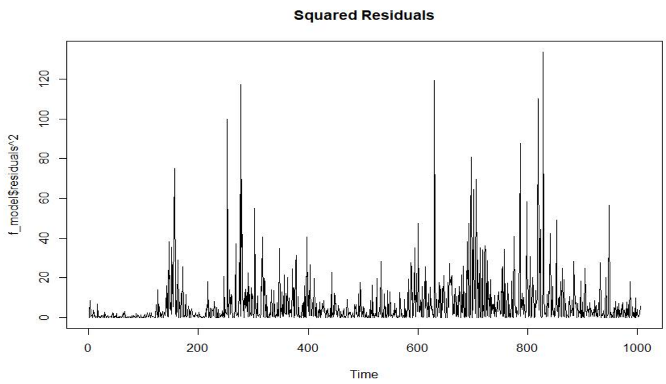


Figure 4. Sequence Diagram of Residents Squared

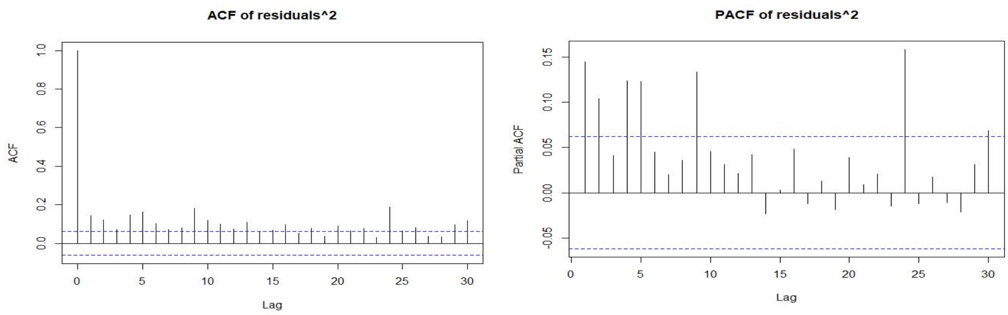


Figure 5. ACF and PACF of the Squared Residual

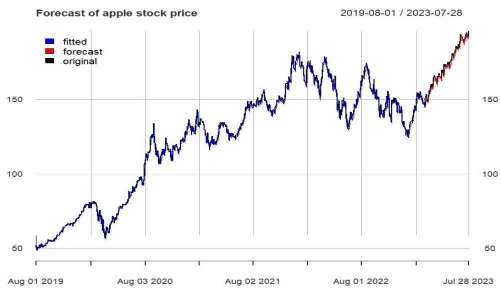


Figure 6. The Actual Data and Forecast Data

Overall, the combination of ARIMA model and GARCH is effective for analyzing Apple stock data, and it is relatively accurate in predicting stock price.

References

[1] Ariyo, A.; Adewumi, A.; Ayo, C. Stock Price Prediction Using the ARIMA Model. In Proceedings of the UKSim-AMSS 16th International Conference on Computer Modelling and Simulation, Cambridge, UK, 26–28 March 2014; pp. 105–111.

[2] Kamruzzaman, M.; Khudri, M.M.; Rahman, M.M. Modeling and Predicting Stock Market Returns: A Case Study on Dhaka Stock Exchange of Bangladesh. Dhaka Univ. J. Sci. 2017, 65, 97–101.

[3] Abbasi, U.; Almarashi, A.; Khan, K.; Hanif, S. Forecasting cement stock prices using ARIMA model: A case study of flying cement. Sci. Int. 2017, 29, 1939-1043.

[4] Fang Yan, Geng Xueyang, Qin Shanshan. Study on the price forecast of media sector index in Shanghai and Shenzhen stock markets -Analysis based on ARIMA-GARCH model [J]. Price Theory and Practice, 2018, 102-105.

[5] Feng Tieying, Luo Tianheng. Stochastic prediction of fertility rate based on ARIMA-GARCH model [J]. statistics and decision, 2015, 24, 21-24.