ARIMA-TGARCH Modeling Analysis of Tesla Stock Data

Yuliang Han
Wuhan Britain-China School, Wuhan, China

Abstract: This paper employs the ARMI-TGARCH model to analyse Tesla stock data. Firstly, the correlation of the data is examined to validate the necessity of data analysis. Subsequently, the ARIMA model is utilized for parameter estimation and analysis. The presence of heteroscedasticity is preliminarily identified through an analysis of residual sequences and their squared values. The need for modelling heteroscedasticity is confirmed via PQ and LM tests. Finally, parameters are estimated using the TGARCH(1,1) model, and the validity of the model is verified through interval estimation of the data. The methodology employed in this paper demonstrates scientific rigor and effectiveness.

Keywords: ARIMA Model, GARCH Model, TGARCH Model, ADF Test, Heteroscedasticity

1. Introduction

GARCH model, also knownas generalized ARCH model, is an extension of ARCH model developed by Bollerslev in 1986 [1]. GARCH model can better capture the long-term dependence and volatility changes of time series data, and provide more flexible parameters to describe the dynamic changes of volatility.

Studying Tesla's stock is also important. First, Tesla is a pioneer and leader in the electric vehicle industry. Its performance and stock movements can provide valuable insights into the overall growth and potential of the electric vehicle market.

Second, Tesla's business model is highly innovative, including not only car manufacturing, but also advances in battery technology, autonomous driving and energy storage. Studying its stock can assess the market's perception and acceptance of these cutting-edge technologies and their potential to disrupt traditional automotive and energy markets. In addition, Tesla's stock price is often influenced by a variety of factors, and analyzing these influences can give us a broader understanding of the complex dynamics affecting the company's financial performance and stock valuation. Finally, as Tesla moves forward, its stock performance could have implications for the stock market as a whole and the economy as a whole. The insights gained from studying Tesla stock help in strategic investment decisions, industry analysis, and policy making.

2. GARCH Model for Data Analysis

The GARCH model can be expressed as:

$$\begin{cases} x_t = f(x_{t-1}, x_{t-2}, \cdots, x_1) + a_t \\ a_t = z_t \sigma_t \\ \sigma_t^2 = w + \sum_{i=1}^{p} \alpha_i \varepsilon_{t-i}^2 + \sum_{j=1}^{q} \beta_j \sigma_{t-j}^2 \end{cases} \quad s.t. \quad \sum_{i=1}^{p} \alpha_i + \sum_{j=1}^{q} \beta_j = 1 \tag{1}$$

$f(x_{t-1}, x_{t-2}, \cdots, x_1)$ means to fit the model of data certainty information, mainly to obtain the mean estimate of the series, such as using regression model or ARIMA model. $E(a_t) = 0$ and $\rho_k \approx 0, \forall k > 0$. $z_t \sim N(0,1)$, σ_t^2 represents the conditional variance of the time series at time t . w is constant term, α and β are coefficients of the GARCH model. ε_{t-i}^2 is square of the error term (residual), reflects the influence of past residuals on current conditional variance. σ_{t-j}^2 represents the influence of past volatility on current conditional variance.

In empirical research, the GARCH(1,1) model is most commonly used, and its expression is as follows:

$$\sigma_t^2 = w + \alpha \varepsilon_{t-1}^2 + \beta \sigma_{t-1}^2, \tag{2}$$

In order to ensure the stationarity of the GARCH process, the constraint of the model is:

$$\alpha + \beta = 1. \tag{3}$$

The limitation of GARCH model is that it cannot explain the negative correlation between stock returns and return changes, that is, conditional variance responds symmetrically to positive price changes and negative price changes. However, empirical studies have found that the market does not respond symmetrically to good news and bad news, and people are more willing to accept good news and more sensitive to bad news. At the same time, in order to ensure that the GARCH model is non-negative, all coefficients are assumed to be greater than zero. These constraints imply that any increase in the lag term will result in conditional variance, thus excluding some random fluctuation behavior, which may lead to the oscillation phenomenon when estimating the GARCH model.

3. TGARCH Model for Data Analysis

GARCH model has been widely studied, but GARCH model does not consider the lever effect, regardless of the previous sequence values increase or decrease, which think about the future at the same sequence swings. However, in reality, many sequences, such as stock volatility, have obvious leverage effects. When asset prices rise or fall sharply, volatility reacts differently, and when asset prices fall sharply, volatility is generally larger.

TGARCH model is an extension of the GARCH model, which captures the non-linear changes of market volatility by introducing threshold effect. Zakoian used a piece smart linear function based on GARCH model to describe the asymmetry of market volatility in order to describe the impact of a negative shock on the current volatility [2]. This nonlinear feature enables the TGARCH model to accurately describe the volatility changes in the market, thus improving the effect of risk management. At the same time, the TGARCH model allows conditional variance to have different responses to positive and negative price changes, which is called asymmetric effect, which can describe the different reactions of the market to good and bad news.

According to the study of Glisten [3], the form of TGARCH model is as follows:

$$\begin{cases} x_t = f(x_{t-1}, x_{t-2}, \cdots, x_1) + a_t \\ a_t = z_t \sigma_t \\ \sigma_t^2 = w + \sum_{i=1}^{p} \alpha_i \varepsilon_{t-i}^2 + \sum_{j=1}^{q} \beta_j \sigma_{t-j}^2 + \gamma \sigma_{t-1}^2 d_{t-1} \end{cases} \quad (4)$$

The first order form is as follows:

$$\sigma_t^2 = w + \alpha \varepsilon_{t-1}^2 + \beta \sigma_{t-1}^2 + \gamma \sigma_{t-1}^2 d_{t-1}, \quad (5)$$

Where $f(x_{t-1}, x_{t-2}, \cdots, x_1)$ represents the modeling of the direction of the data, $z_t \sim N(0,1)$, σ_t^2 represents the conditional variance of the time series at time t . w is constant term, α and β are coefficients of the TGARCH model. ε_{t-i}^2 represents the square of the error term (residual), reflects the influence of the past residual on the current conditional variance, σ_{t-j}^2 represents the past conditional variance, reflects the influence of the past volatility on the current conditional variance.

Compared with the GARCH model, $\gamma \sigma_{t-1}^2 d_{t-1}$ is the TGARCH term, A threshold d_{t-1} is set to describe the impact of information, When $d_{t-1} = 1$ is positive news impact, $d_{t-1} = 0$ is negative information impact. γ is a parameter that reflects the influence of positive and negative information on market asymmetry.

4. Tesla Stock Analysis

4.1 Data Source and Description

Taking the closing price data of Tesla stocks as the research object, the data is derived from https://cn.investing.com/equities/tesla-motors-historical-data. A total of 639 data span from January 3, 2022 to July 18, 2024. We take the time as the horizontal coordinate and the closing price of the stock as the vertical coordinate to get the time series diagram. We can find that the price of the stock changes greatly over time, with a maximum value of \$399.93 per share and a minimum value of \$108.1 per share. As can be seen from the figure, the change of the closing price fluctuates around 250, a large fluctuation is often accompanied by another large fluctuation, and a small fluctuation is also accompanied by another small fluctuation, which conforms to the characteristics of high frequency time series, indicating that the data also has a volatility aggregation effect.

QQ graph is used to test whether the data is white noise. If the stock data is white noise, there is no correlation between the data and there is no need for further modeling. The QQ graph from which it can be seen that more than 50% of the data is not on the straight line, indicating that the original data is not white noise. Combining Shapiro-Wilk normality test, we get the test statistic is 0.9822, and p-value is $5.212e^{-7}$, which indicates that there is a correlation between stock closing price data.

According to the calculation of the stock closing price, the skewness of the sequence data is 0.3474, the peak is -0.2717, the Jarque-Bera statistic is 14.736, and its corresponding p value is 0.0006, indicating that the distribution of the data is not normal, and the skewness is significantly greater than 0. This shows that there is a right rear tail phenomenon in the series, and it can be seen that the distribution of the series has obvious "peak and thick tail" and asymmetric phenomenon, which accords with the common characteristics of financial time series.

The unit root test (ADF test) was further conducted to judge the stationarity of the time series, and the results were shown in Table 1. As can be seen from the unit root test results in Table 1, the ADF test statistic of the closing price data is -1.0033, and the null hypothesis is accepted at all three levels. Common methods for processing stock data to make it become a stable time series include difference and logarithm. In this paper, d-order difference method is adopted, and 1-order difference is denoted as $x_t - x_{t-1}$.

Table 1. ADF Test.

	t- statistic	p-value
Value of test-statistic	1.0033	0.3161
Critical value at 1%	-2.58	
Critical value at 5%	-1.95	
Critical value at 10%	-1.62	

4.2 ARIMA model

Auto-regression moving average model (ARMA model) is an important method to study time series. It is a mixture of auto-regression model (AR model) and moving average model (MA model). The ARIMA model is a modeling approach that utilizes the ARMA model based on data differencing, which can be expressed as flowing

$$(1-\phi_1 B-\cdots-\phi_p B^p)(1-B)^d x_t = \mu + (1+\theta_1 B+\cdots+\theta_q B^q)\varepsilon_t \tag{6}$$

B is the backward operator, μ is the mean of the sequence. ϕ_i and θ_j are autoregressive coefficient and moving average coefficient respectively.

Table 2. Model Parameter Estimation Results.

Variable	Coefficient	Std.Error
μ	-0.2504	0.4082
AR(1)	0.1923	0.1912
AR(2)	0.6084	0.1619
MA(1)	-0.2130	0.1991
MA(2)	-0.5486	0.1670

In order to determine the order and parameters in the ARIMA(p,d,q) model, the fitted model is determined according to the sum of squares of the error and AIC criterion. The estimation results of model parameters are shown in Table 2. The AIC value of the model is 4565.42, and the log-likelihood function is -2276.71, The expression we get for the model is:

$$(1-0.1923B-0.6084B^2)(1-B)x_t = -0.2504 + (1-0.213B-0.5486B^2)\varepsilon_t \tag{7}$$

4.3 Analysis of Variance

4.3.1Phillips-Perron (PP) test

Phillips-Perron test was used to verify the stationarity of the time series [4]. The null hypothesis of the test is that the sequence has a unit root, that is, non-stationarity. The test statistics of PP test are:

$$Z(t) = (\sqrt{\hat{\sigma}}\ /\ \hat{\lambda})t - \frac{\hat{\lambda}^2-\hat{\sigma}}{2\hat{\lambda}}\cdot\frac{n\hat{\sigma}_p}{s_n} \tag{8}$$

(1) $\hat{\sigma}^2$ is the unconditional variance sample estimate of σ^2, *i.t.* $\hat{\sigma}^2 = n^{-1}\sum_{t=1}^{n}\hat{\varepsilon}_t^2$.

(2) $\hat{\lambda}^2 = \hat{\sigma}^2 + 2n^{-1}\sum_{j=1}^{q}[1-\frac{j}{q+1}]\sum_{t=j+1}^{n}\hat{\varepsilon}_t\hat{\varepsilon}_{t-j}$, $s_n^2 = \frac{1}{n-k}\sum_{i=1}^{n}\hat{\varepsilon}_i^2$,

where $\hat{\varepsilon}_i$ is the OLS residual, k is the number of covariates in the regression, q is the number of Newey-West lags to use in calculating $\hat{\lambda}$, and $\hat{\sigma}_\rho$ is the OLS standard error of $\hat{\rho}$.

The critical values, which have the same distribution as the Dickey Fuller statistic, which means that the heteroscedasticity series only needs to be modified based on the original statistic, $Z(t)$ not only take into account the effects of autocorrelation errors, but also continue to use the critical table of the statistic for testing without fitting a new critical table.

Based on the fitting error acquired through formula (7), we computed the pp-test value of the squared errors to be 0.01. Hence, we contend that the variance signified by the sum of squared errors is dissimilar, and further modeling is requisite.

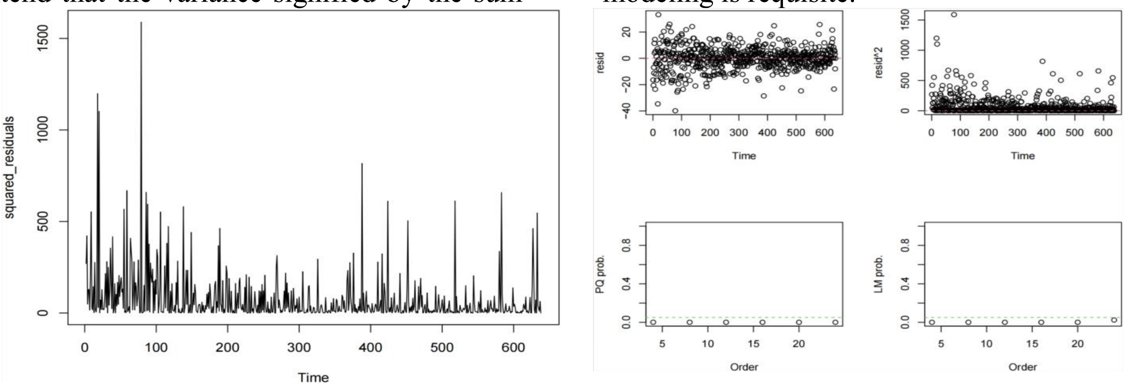

Figure 1. (a) Time Sequence Diagram of Residuals Squared, (b) ARCH Diagnostic Diagram.

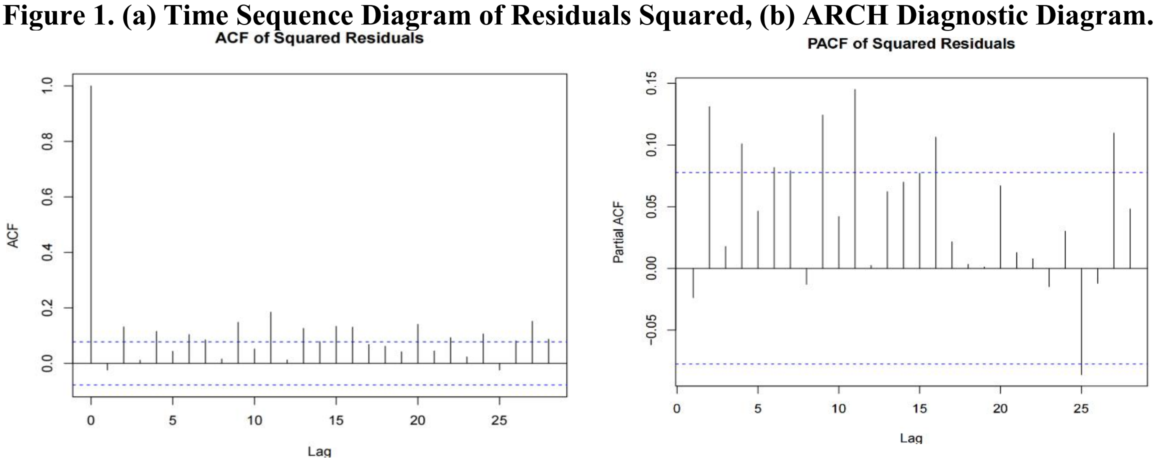

Figure 2. (a) Autocorrelation Function Diagram of the Squared Residuals, (b) Partial Autocorrelation Function Graph of Residual Squared.

4.3.2 ARCH test

By observing the residual plot of the ARIMA(2,1,2) model, it can be found that the residual is not completely stable near 0, which can be better reflected by the residual square plot, so heteroscedasticity is inferred.

The P-Q test shows that for all tests, the lag order (4,8,12,16,20,24), the P-value is much less than 0.05, which means that there is enough evidence to accept the null hypothesis, that is, the residual may have significant autocorrelation; The LM test shows that the P-value for any order of lag is much less than 0.05, indicating that at the significance level, there is sufficient evidence to reject the null hypothesis that the residuals are heteroscedasticity. Although the P-value is larger when the lag order is 4 and 8, it can be considered that there is a certain degree of heteroscedasticity. Therefore, using the ARIMA model to fit a merge is not the most appropriate.

4.4 TARCH Modeling

Since the low-order TGARCH model has a better fitting effect, the TGARCH(1,1) model may be considered.

$$x_t = \mu + a_t, a_t = z_t\sigma_t, z_t \sim N(0,1) \tag{9}$$

$$\sigma_t^2 = w + \alpha\varepsilon_{t-1}^2 + \beta\sigma_{t-1}^2 + \gamma\sigma_{t\text{-}1}^2 d_{t-1}$$

The calculated parameters are estimated as

Table 3. Parameter Estimation of TGARCH.

	Estimate	Std.Error	t-value	P-value
μ	-0.105699	0.316555	-0.33391	0.738451
w	0.536288	0.486153	1.10313	0.269972
α	0.032583	0.014584	2.23412	0.025475
β	0.965218	0.015951	60.51170	0.00000
γ	-0.012214	0.014299	-0.85420	0.392997

The AIC value of this model is 7.0777, and the BIC value is 7.1126, which is obviously lower than the ARIMA (2,1,2) model fitted before, so the TGARCH model is relatively better at this time.

Nyblom test is used to check whether the model parameters meet the stationarity condition. The statistic 1.1576 of the joint test does not exceed the critical value 1.88 of 1%, so the result is not significant, and it can be considered that the model parameters meet the stationarity condition.

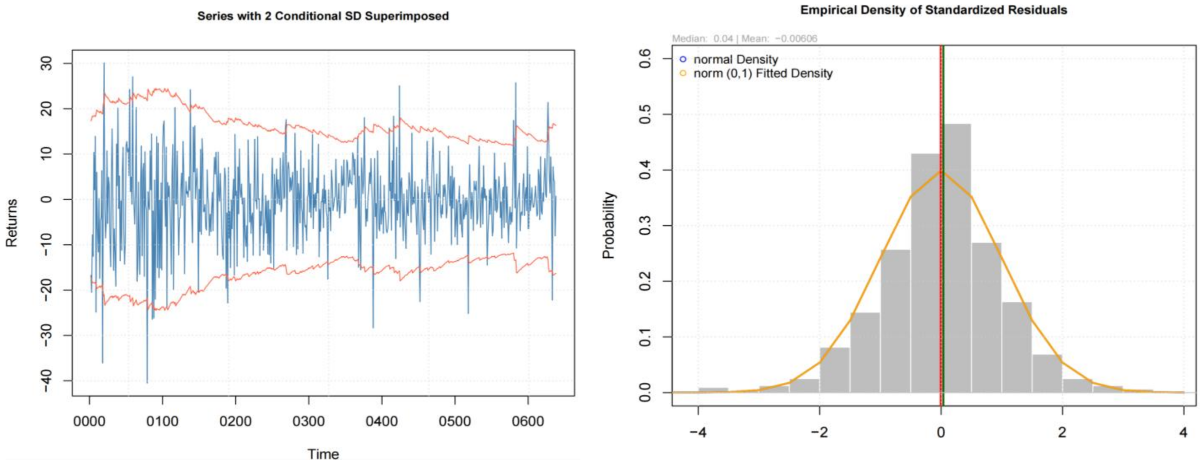

Figure 3. (a)Interval Estimation, (b) Empirical Density Function Diagram 0f Residuals.

In the 99% confidence level curve of the model 3 (a), it can be found that most of the data are within the 99% confidence interval except for a few points. In the figure 3(b), the normalized residual distribution is approximately normal, and the thick tail phenomenon is not significant.

5. Conclusions

Through the analysis of Tesla stock data, we give a modeling method of skewable heteroscedasticity data, and this method performs well in accuracy and interval estimation. The model error is within the acceptable range, and our model is valid..

References

[1] Bollerslev,T.(1986) Generalized autoregressive conditional heteroskedasticity, Journal of Econometrics, 31, no. 3, 307–327.

[2] Zakoian J.M. (1994) Threshold heteroskedastic models, Journal of Economic Dynamics and Control, 18, no. 5, 931–955.

[3] Glosten, L. R. Jagannathan R., and Runkle D. E.(1993) On the relation between the expected value and the volatility of the nominal excess return on stocks, Federal Reserve Bank of Minneapolis, Staff Report 157.

[4] Phillips P.C. B. and Perron P. (1988) Testing for a unit root in time series regression, Biometrika, 75, no. 2, 335–346.

