The Time Series Analysis of Tesla

—— BOSI TU， 3774297802

1 Introduction

In the first half of 2020, most industries have been hit hard by COVID-19. However, the new energy industry has become a bull sector in the downward cycle of the stock market. CATL, as a leading company in new energy batteries, has maintained a strong stock price trend. As a result, as a downstream industry of new energy batteries, the trend of new energy vehicles is particularly eye-catching.

In this paper, we will focus on the data of daily earnings ratio of Tesla (TSLA), the world's leading new energy vehicle company. We will use the knowledge in Time Series Analysis to dig out the potential information hidden in the data set and provide prediction of the future rate of the future earnings rates. As for the pandemic impact, we would conduct an intervention study to verify the significance of the impact on the stock price of Tesla. After that, we will analyze the data using different methods in order to find the most appropriate model. What's more, we will formulate a GARCH model and see how it works because of the unsatisfying results given by former models. Prediction results based on model is demonstrated in the last part.

2 Data Selected

We choose the following data—the daily earnings ratio of Tesla (TSLA) data from 11th April 2019 to 26th April 2023, four years data.

3 Data Preprocessing

At first, we will plot all the data of daily earnings ratio in order to make a rough prediction on various aspects of properties of the data set. The plot is given in figure 1.

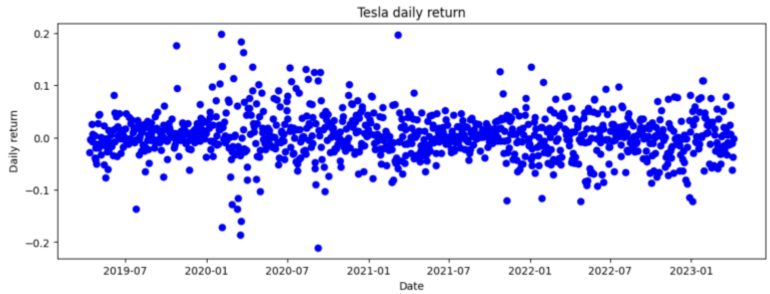

Figure 1: The daily earnings ratio

Figure 1 showed that the variance of the DER (abbreviation of daily earnings ratio) is sometimes relatively large, and sometimes rather small, which states that the variance is not a constant for sure which means the corresponding model would be complicated possibly. Based on this discovery, we will focus on heteroscedastic models in the rest part of the article.

3.1 Difference Analysis

In order to make further discoveries in the data, we will plot the 1st difference and 2nd difference of DER itself in figure 2.

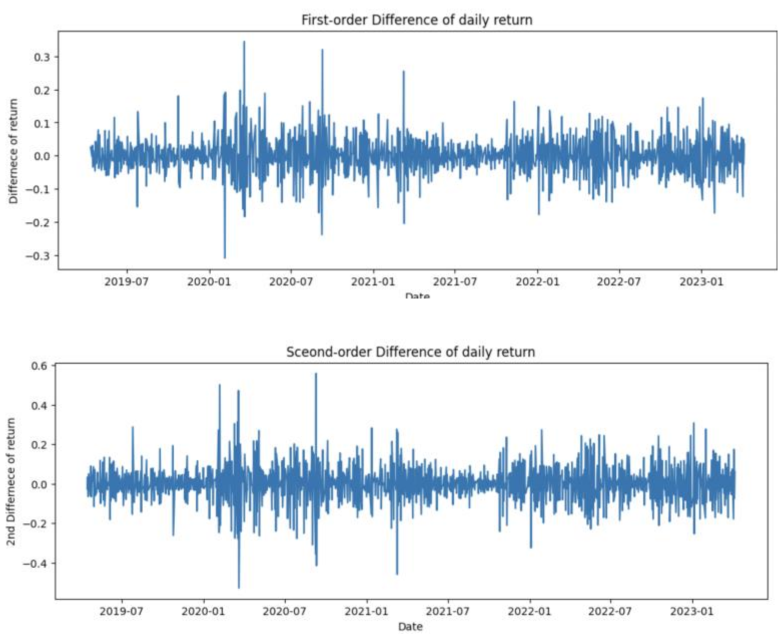

Figure 2: The 1st and 2nd order difference of DER

As we can see, the 1st and 2nd order difference of DER seems to give us a model which has variance verges to a constant. What's more, the 1st and 2nd difference have almost the same properties form the graphs. They both have heteroscedasticity which indicates us to use the GARCH model probably. Furthermore, we can plot the ACF and PACF of DER to uncover more properties in figure 3.

3.2 ACF & PACF Analysis

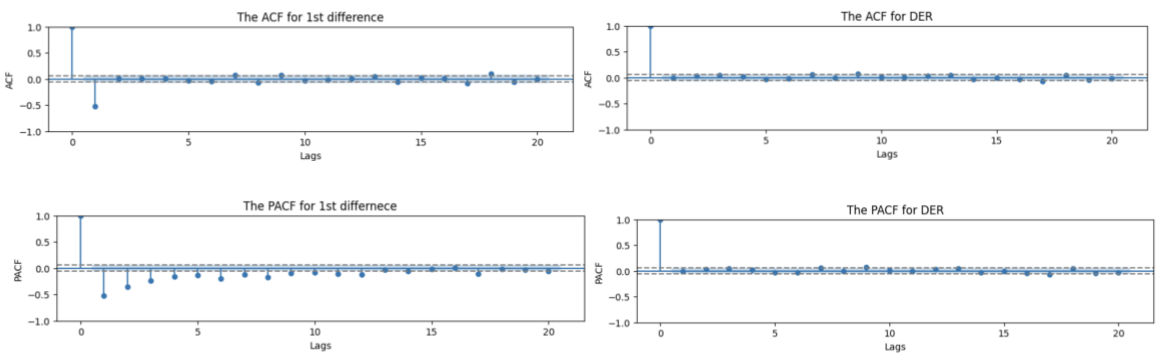

Figure 3: The ACF and PACF of DER and its 1st order difference

Form the figure of ACF and PACF of DER and its 1st order difference, we can find that the corresponding parameters of the model without difference are prominent even when lags are large enough which states that we cannot use such model. We should use 1st difference in order to avoid overfitting. Therefore, we choose the 1st order difference model to fit the previous data.

4 Model Formulation

4.1 ARIMA Model

Next, we will determine the orders of ARIMA model including the corresponding parameters. In order to achieve this target, we would use the ACF and PACF plots to determine the possible order of AR terms and MA terms, then Akaike's Information Criterion (AIC) to determine the most suitable model for the DER.

Form the ACF and PACF we have above in the ACF & PACF analysis, we can determine the order of MA term is 1 by ACF graph since only lag 1 is above the 0.05 significance line. But form PACF graph, we cannot determine the order of AR term since the graph can be regarded as an infinite tail or decreasing to zero after lag 1 or 2. Thus, we need AIC to determine the final order of AR term. We will test three models including ARIMA (2,1,1), ARIMA (1,1,1) and ARIMA (0,1,1).

We should distinguish these three models by p-value and their AIC numbers hoping to find the one that fits best. The results are shown in the following figure 4:

	ARIMA (2,1,1)	ARIMA (0,1,1)	ARIMA (1,1,1)
P-value	0.2146	0.4535	0.09713
AIC	-3457.947	-3456.743	-3319.432

Figure 4: AIC and P-value of DER

According to AIC, we have the result that three models have nearly the same. We know that if P-value of the model is smaller than 0.05, then H0 assumption can be rejected. After taken P-values into consideration, we can choose model ARIMA (1,1,1). After we choose the model, we can forecast using ARIMA (1,1,1) model, then we get the coefficients are:

```
                               SARIMAX Results
==============================================================================
Dep. Variable:                      y   No. Observations:                 1004
Model:                 ARIMA(1, 1, 1)   Log Likelihood                1536.724
Date:                Mon, 24 Apr 2023   AIC                          -3067.447
Time:                        14:59:55   BIC                          -3052.715
Sample:                             0   HQIC                         -3061.849
                               - 1004
Covariance Type:                  opg
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
ar.L1         -0.5193      0.022    -23.285      0.000      -0.563      -0.476
ma.L1         -0.9998      0.421     -2.375      0.018      -1.825      -0.175
sigma2         0.0027      0.001      2.366      0.018       0.000       0.005
===================================================================================
Ljung-Box (L1) (Q):                  33.40   Jarque-Bera (JB):               549.71
Prob(Q):                              0.00   Prob(JB):                         0.00
Heteroskedasticity (H):               0.86   Skew:                             0.10
Prob(H) (two-sided):                  0.18   Kurtosis:                         6.62
===================================================================================
```

Figure 5: Coefficients and standard error

According to the results and 2-sigma theorem, we get the actual ARIMA model can be written as:

$$r_t = -0.5193 r_{t-1} + e_t - 0.9998 e_{t-1}. \tag{1}$$

4.2 Intervention Study

Now, we have the most suited ARIMA model for the DER of Tesla. Since the COVID-19 had huge impact on various industries, I am curious about the impact of the pandemic on the DER of Tesla. We use the intervention study on the pandemic period (3/1/2020- 6/30/2020) and see the significance of the COVID:

```
                            OLS Regression Results
==============================================================================
Dep. Variable:           Daily return   R-squared:                       0.001
Model:                            OLS   Adj. R-squared:                 -0.001
Method:                 Least Squares   F-statistic:                    0.4945
Date:                Sun, 23 Apr 2023   Prob (F-statistic):              0.610
Time:                        21:10:15   Log-Likelihood:                 1738.0
No. Observations:                1005   AIC:                            -3470.
Df Residuals:                    1002   BIC:                            -3455.
Df Model:                           2
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
const          0.0028      0.001      1.999      0.046    5.25e-05       0.006
Drift         -0.0067      0.032     -0.212      0.832      -0.069       0.055
dummy          0.0048      0.005      0.976      0.329      -0.005       0.014
==============================================================================
Omnibus:                       79.603   Durbin-Watson:                   1.999
Prob(Omnibus):                  0.000   Jarque-Bera (JB):              406.252
Skew:                           0.076   Prob(JB):                     6.07e-89
Kurtosis:                       6.111   Cond. No.                         23.4
==============================================================================
```

Figure 6: The outcome of intervention study

We created dummy variables for the data during pandemic dates and ran a linear regression. The coefficient of the dummy variables had a t-statistics 0.329 which is less than 1.96 that indicates the **insignificance of the impact of COVID on the daily return of Tesla**.

4.3 Residual analysis for ARIMA model

Then we test our model by residual analysis. The residual figure and QQ plot are given as follow:

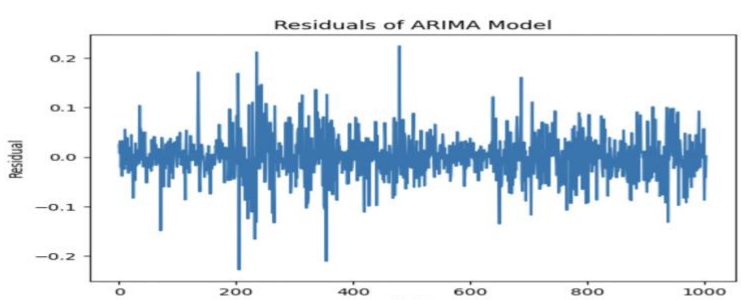

Figure 7: Residual

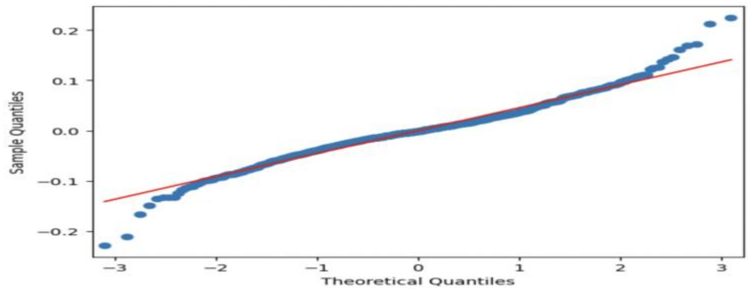

Figure 8: QQ-plot

Figure gives the quantile-quantile plot and the residuals from the ARIMA (1,1,1), the matching line of the QQ-plot is not nearly a straight line through the origin. The large deviation of the values at both ends of the QQ plot indicates that the distribution of the residuals has a fat tail. Then, we shall do residual test in the following process.

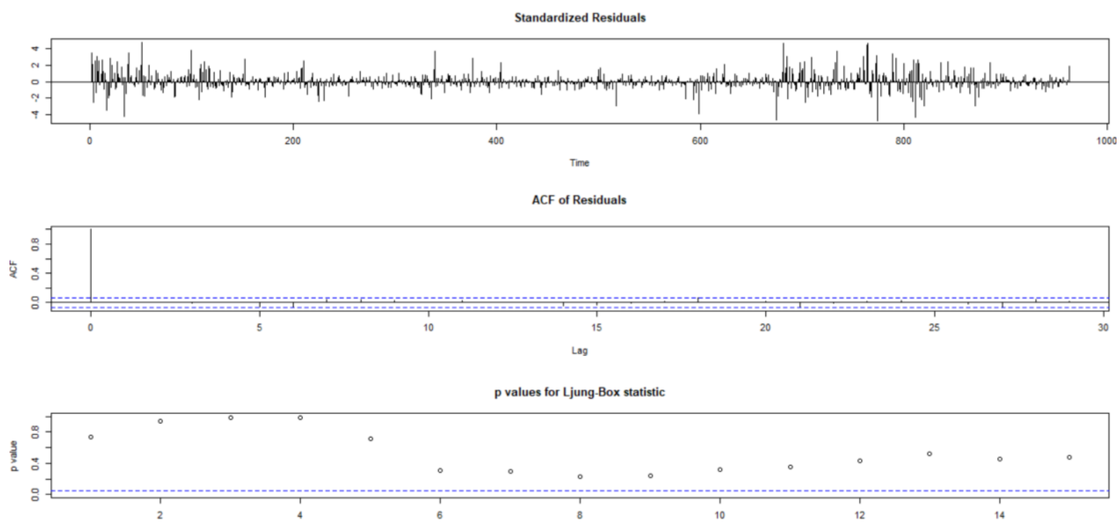

Figure 9: Residual Test

The above figure is a sequence plot of the standard residuals, the sample ACF of residual and P-values for the Ljung_Box test statistic for a whole range of values of K from 1 to 15. The horizontal dashed line at 5% helps judge the size of the P-values. The extreme value at 6,21,27 implies that the original hypothesis should not be rejected.

From now on, we can see that, using the ARIMA model cannot simulate the time series well.

4.3 GRACH Model

We may look back into the graph of DER to have a better understanding of the clues hidden. The DER seems to have a constant mean which is around 0, but the variance of it seems to be fluctuating that indicates the presence of heteroscedasticity.

GARCH Model is really convenient when we are trying to solve finical time series. It is not only an advanced method to simulate heteroscedasticity sequence but also a method to subscribe the volatility of financial data.

We will first use ARCH-LM test to verify the existence of the GARCH effect in the DER data:

```
ARCH-LM Test
H0: Residuals are homoskedastic.
ARCH-LM Test
H1: Residuals are conditionally heteroskedastic.
Statistic: 152.2943
P-value: 0.0000
Distributed: chi2(22)
```

Figure 10: Result of the ARCH-LM test

The zero hypothesis is that the time series does not exist GARCH effect. The original hypothesis is rejected below the 1% significant level. After testing, we see that the DER series exists GARCH effect. The next step is to fit GARCH model.

We tried four GARCH models, GARCH (1,1), GARCH (2,1), GARCH (1,2) and GARCH (2,2). The AICs of the four models are given in the following table:

	GARCH (1,1)	GARCH (2,2)	GARCH (1,2)	GARCH (2,1)
AIC	-2936.42	-2932.85	-2935.44	-2934.42

The last three models all have insignificant coefficients and their AIC are all larger than that of GARCH (1,1). Then we approximate the GARCH (1,1) model parameters and get the following results:

```
<bound method ARCHModelResult.summary of                    Zero Mean - GARCH Model Results
==========================================================================
Dep. Variable:                      y   R-squared:                   0.000
Mean Model:                 Zero Mean   Adj. R-squared:              0.001
Vol Model:                      GARCH   Log-Likelihood:            1471.21
Distribution:                  Normal   AIC:                      -2936.42
Method:            Maximum Likelihood   BIC:                      -2921.69
                                        No. Observations:             1004
Date:                Mon, Apr 24 2023   Df Residuals:                 1004
Time:                        18:45:51   Df Model:                        0
                              Volatility Model
========================================================================
                 coef    std err          t      P>|t|      95.0% Conf. Int.
------------------------------------------------------------------------
omega      1.3165e-03  3.536e-04      3.723  1.970e-04 [6.234e-04,2.010e-03]
alpha[1]       0.3904  6.428e-02      6.074  1.250e-09     [  0.264,  0.516]
beta[1]        0.2655      0.115      2.306  2.109e-02  [3.988e-02,  0.491]
========================================================================
```

Figure11: Model result for GARCH (1,1)

It is clear that all the parameters in the GARCH (1,1) model are significant, so we use the GARCH (1,1) model to fit the data.

One step further, we plot the GARCH conditional volatility. The volatility is a crucial index in derivative pricing, we can use the fitted model to forecast the implied volatility and put it into the BS model to price options based on the corresponding asset.

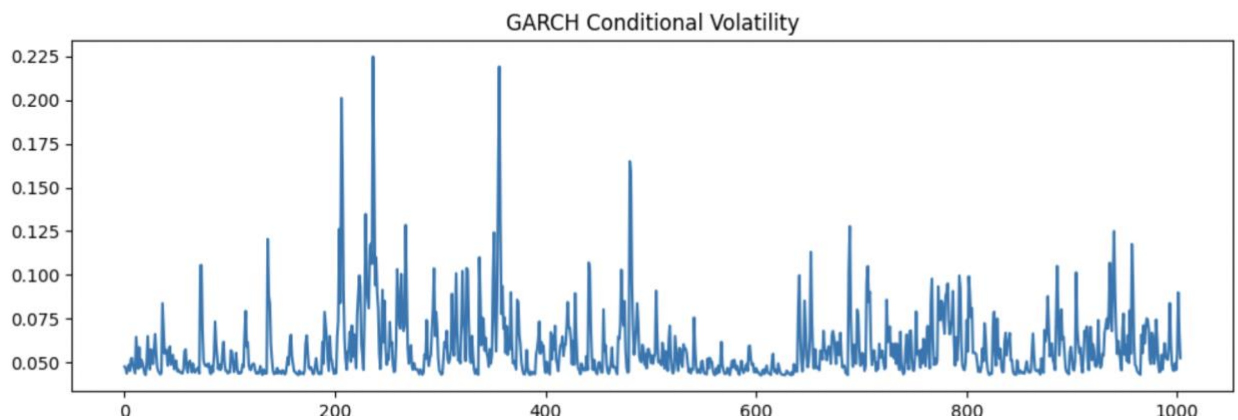

Figure12: GARCH model conditional volatility

5 Option pricing

Black-Scholes model is a widely used pricing model in derivatives. The most important index of the model in the implied volatility of the asset. From the research we did above, we constructed a reliable model to predict the implied volatility. Now, we will try to price the option based on the model we constructed.

5.1 European call option

We attempt to price a European call option with an expiration date of one month using the implied volatility predicted by our previously constructed GARCH model. The date is 4/19/2023.

```
res = model.fit()
forecast = res.forecast(horizon=30)
print(forecast)
```

```
1005    0.015892
dtype: float64
```

Figure13: The implied volatility of the Tesla

The spot price is $180.59, the implied volatility of in 30 days is 15.89%, time to maturity is 1/12, the risk-free interest rate is 3.43% and the strike price is $185. According to the Black-Scholes model, the price of the European call option is $1.56.

By the Investing.com (https://cn.investing.com/equities/tesla-motors-options), the call option price of the Tesla with the strike price of $185 and time to maturity of 30 days is $1.185. The difference between my expected price and the actual price is $0.375 which is also known as the absolute error. The relative error is 32%.

5.2 Model evaluation

According to the above results of the option pricing, I concluded the following reasons that could lead to the difference:

1. The trading cost: I ignored the trading cost which is a crucial influence towards pricing in the real world.

2. Data selection: I used the data with too long a period which should be shorten since the stock has a strong momentum effect.

6 Conclusion

The GARCH model has a high accuracy in predicting the implied volatility only if the data was selected and processed carefully and correctly. The model can be used in pricing model but in the real market the model needs to consider more transaction factors that will affect the price.