<u>Time Series Final Project</u>

Netflix Stock Price

Department of Statistics

Students: Tai, Yan-Ni, Jiang, Jyun-You, Ruan, Zi-Qing

ID: B54076101, H24096134, H23096053

A. Introduction:

Netflix is a subscription-based streaming service company that allows its members to watch TV shows and movies on an internet-connected device. The company is one of the world's biggest entertainment companies and has over 233 million paid memberships in over 190 countries.

Stocks are a type of security that gives stockholders a share of ownership in a company, which are also called "equities". Companies issue stocks to get money for expanding their market or region, pay off their debts, or make new products. Stocks usually fluctuate daily, and many people regard stocks as an investment, even earning lots of fortune through buying stocks.

In Taiwan, about 4.1 million people have subscribed to streaming services, and about 210 thousand people subscribed to Netflix, which holds 21% of the market share. Due to the many users in Taiwan and around the world, we are interested in the stock price of Netflix company (NFLX). Therefore, in the project, we aim to understand the pattern of the close price of Netflix and make predictions about the company.

B. Methodology:

➢ ARIMA (Auto-Regressive Integrated Moving Average) Model

ARIMA models are a general class of models using differencing and combining an auto-regressive (AR) and a moving average (MA) model. The model is usually written as ARIMA(p,d,q), and the equation is as follows:

$$\left(1-\sum_{i=1}^{p}\phi_i\beta^i\right)(1-\beta)^d y_t=\left(1+\sum_{i=1}^{q}\theta_i\beta^i\right)e_t$$

$$e_t \sim N(0,\sigma^2)$$

- y_t: the closing price of Netflix Company
- p: the number of autoregressive terms
- q: the number of lagged forecast errors in the prediction equation
- d: the number of nonseasonal differences needed for stationarity
- L: the lag operator

➢ GARCH Model

We use the GARCH model to predict the series with the variance error serially autocorrelated.

$$y_t = \mu_t + e_t$$

$$e_t = \sigma_t \varepsilon_t, \qquad \varepsilon_t \sim a\ distribution(mean = 0, variance = 1)$$

$$\sigma_t{}^2 = \alpha_0 + \sum_{i=0}^{m} \alpha_i e_{t-i}^2 + \sum_{j=0}^{s} \beta_j \sigma_{t-j}^2$$

- y_t: the closing price of Netflix Company
- μ_t: any function, such as arma(p,q)
- α_0: refer to the intercept (mean)
- α_i: autoregressive coefficient on lagged e_t^2
- β_j: regression coefficient on lagged σ_t^2
- m: the number of lagged e_t^2
- s: the number of lagged σ_t^2

➢ Holt's Linear Trend Model

$$\widehat{y_{t+h|\,t}} = \ell_t + hb_t$$

$$\ell_t = \alpha y_t + (1-\alpha)(\ell_{t-1} + b_t - 1)$$

$$b_t = \beta^* (\ell_t - \ell_{t-1}) + (1 - \beta^*) b_{t-1}$$

- y_t: the closing price of Netflix Company
- Smoothing parameter: α, β
- Initial value: b_0 and ℓ_0

C. Data Analysis:

The time length of the Netflix stock price data is from 2018-01-02 to 2022-12-30, 1259 days in total. Also, there are 6 variables in the data, which are Open, High, Low, Close, Adjusted Price, and Volume. In this project, we choose closing price as the indicator in the following analysis.

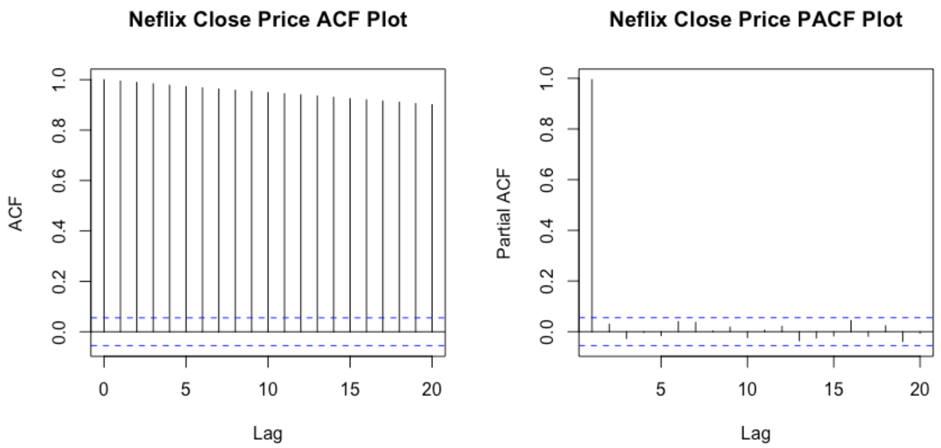
The following plot is the time series plot of Netflix's stock closing price. By inspecting the pattern of the data, we thought that the closing price of the company presents a random pattern. In addition, we found out that the closing price had a significant drop in the first half of 2022. Moreover, the lowest point (blue point) is 166.37 on 2022-05-11, and the highest point (red point) is 691.69 on 2021-11-17.

Figure C.1: Time series plot of Netflix closing price

➢ <u>ARIMA Model</u>

To check whether the time series data is stationary or not, we first plot the ACF and PACF plot and conduct the Augmented Dickey-Fuller (ADF) test.

Table C.1: ADF test for Netflix closing price

ADF Test
H_0: *the series is nonstationary vs* H_1: *the series is stationary*

p-value = 0.7735 > 0.05, do not reject H_0 => the series is nonstationary

By the table above, the ACF plot decreases slowly, and the series did not reject the null hypothesis of the ADF test. Thus, we conclude that the series is nonstationary and needs to be differenced.

Then, we differentiate the time series data and plot the first-order differenced series. We also conduct the ADF test and draw the ACF and PACF plot using the closing price data after first-order differentiation to check the stationarity.

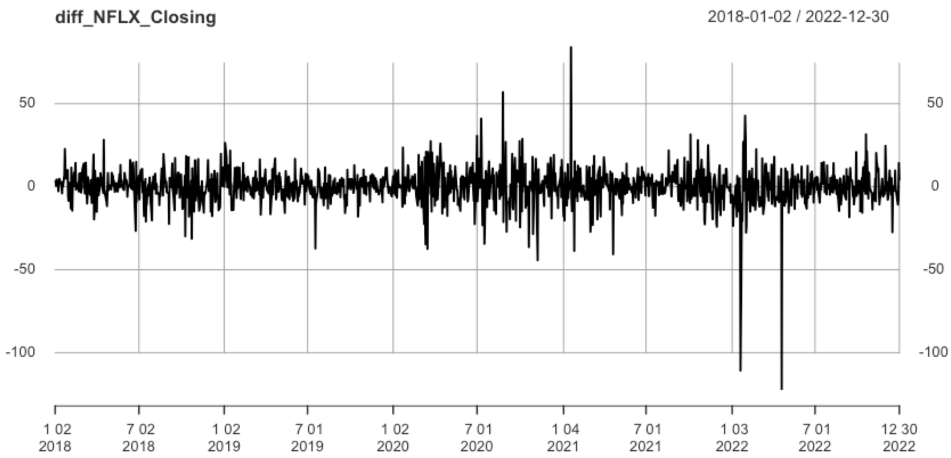
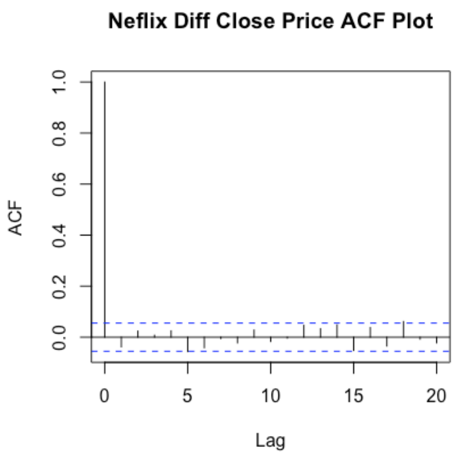
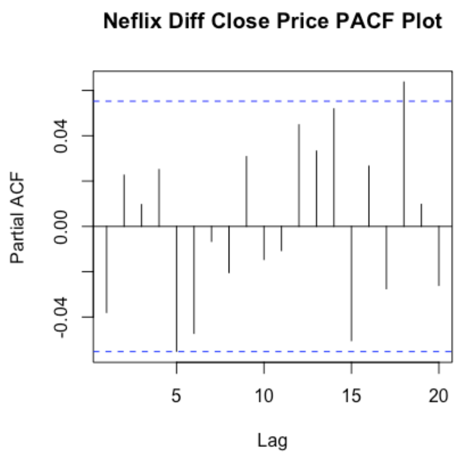


Figure C.2: Time series plot of Netflix closing price after differentiation

Table C.2: ADF test for Netflix closing price after differentiation

ADF Test
H_0: *the series is nonstationary vs* H_1: *the series is stationary*
 
p-value = 0.01 < 0.05, reject H_0 => the series is stationary

Through the ACF plot, all lags are below the blue dotted lines, and in the PACF plot, we find that lag 18 is significant. However, we will not consider lag 18. Moreover, by the ADF test, we can conclude that the time series data is stationary after differentiation and use this series to build the ARIMA model.

To build an ARIMA model, we have to consider both parameters p and q. According to the ACF and PACF plots, p=0 and q=0 are chosen. Then, we split the data into training data and testing data to train and evaluate the data. The training data is from 2018-01-02 to 2022-10-06 with 1200 observations; the testing data is from 2022-10-07 to 2022-12-30 with 59 observations.

The ARIMA(0,1,0) is shown as follows:

$$y_t = y_{t-1} + W_t, \qquad W_t \sim N(0, 125.5)$$

The model is a random walk model, and the AIC of the model is 9198.94.

After fitting the model, we check the residual to confirm that the model residual has followed the assumptions.

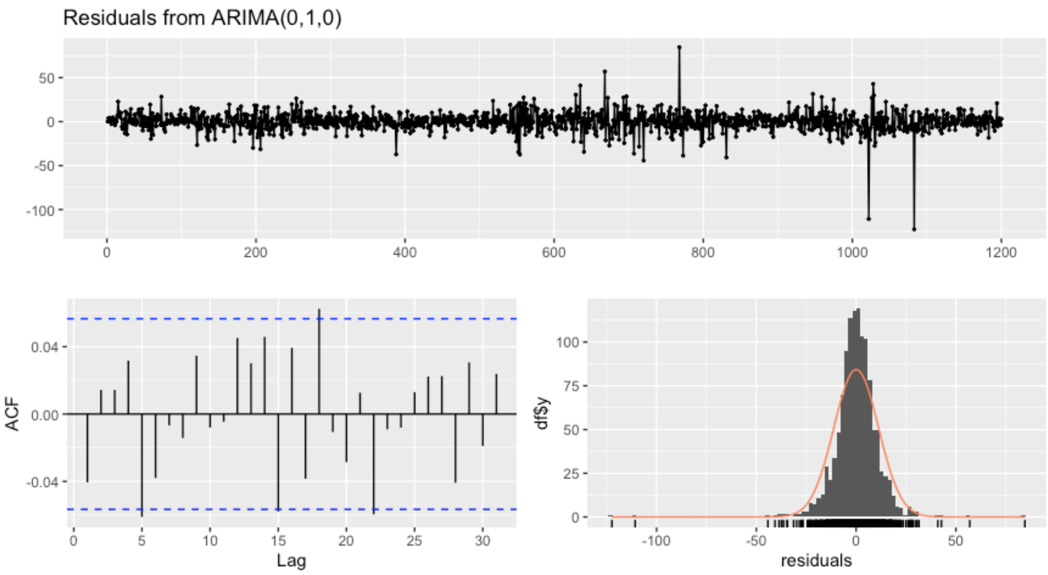


Figure C.3: Residual plots for ARIMA model

By the plots above, we can observe that the residuals of the ARIMA model are mostly around 0 and are normally distributed, which satisfies the residual assumptions. To further check, we conduct several tests to confirm.

Table C.3: Residual tests for ARIMA model

Ljung-Box test	$H_0: residuals\ are\ independent$ $H_1: residuals\ are\ dependent$	p-value = 0.175 Do not reject H_0
t-test	$H_0: residuals'\ expected\ value\ is\ 0$ $H_1: residuals'\ expected\ value\ is\ not\ 0$	p-value = 0.8125 Do not reject H_0

The tests above show that the residuals of the data are independent and truly have 0 as the expected value.

Then we try to predict the test data to evaluate the model. The RMSE of the model is 11.1643. The following plot shows the result of the prediction and we can see that the predicted results were not good, so other models should be considered.

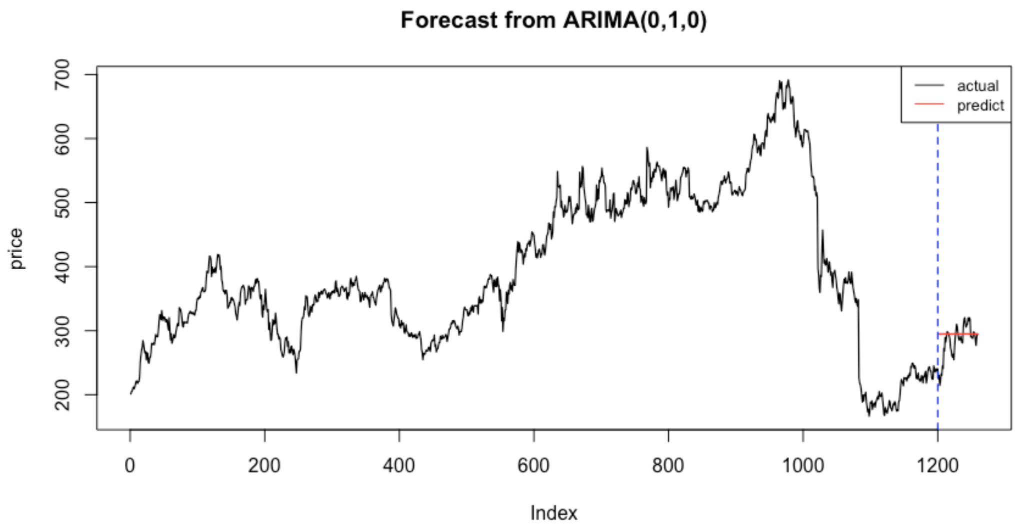


Figure C.4: Prediction plots for ARIMA model

- <u>GARCH Model</u>
  - **Descriptive statistics of raw data**

Minimum	Maximum	Median	Variance	Skewness	kurtosis
166.3699	691.69	362.32	14278.0065	0.381356	-0.664377

Table C.4: Descriptive statistics of raw data

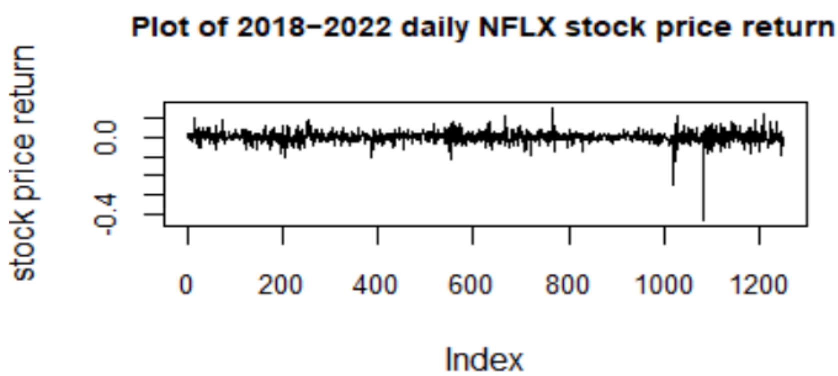
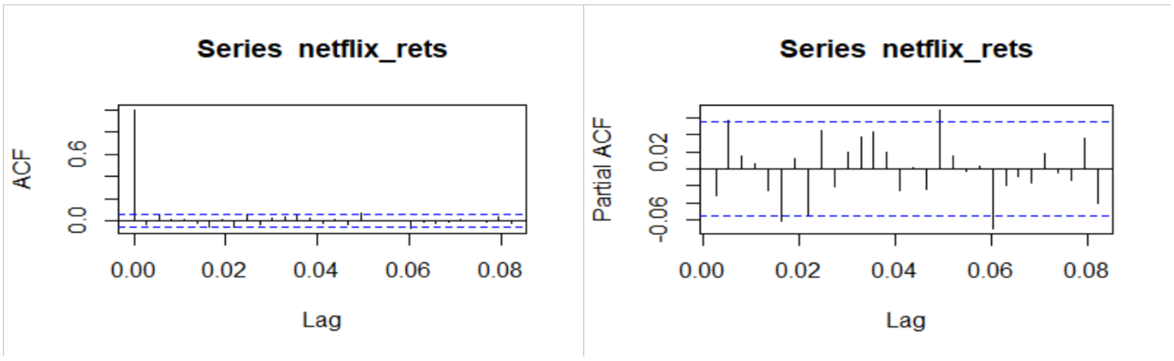


Figure C.4: Netflix stock price return plot

=>From the stock price return plot, we observed volatility clustering.

Table C.5: Ljung box test for Netflix stock return

Stock price return ACF 、PACF



Ljung-Box test

H_0: returns are independent vs H_1: returns are dependent

p-value = 0.27 > 0.05, do not reject H_0, which implied returns are independent.

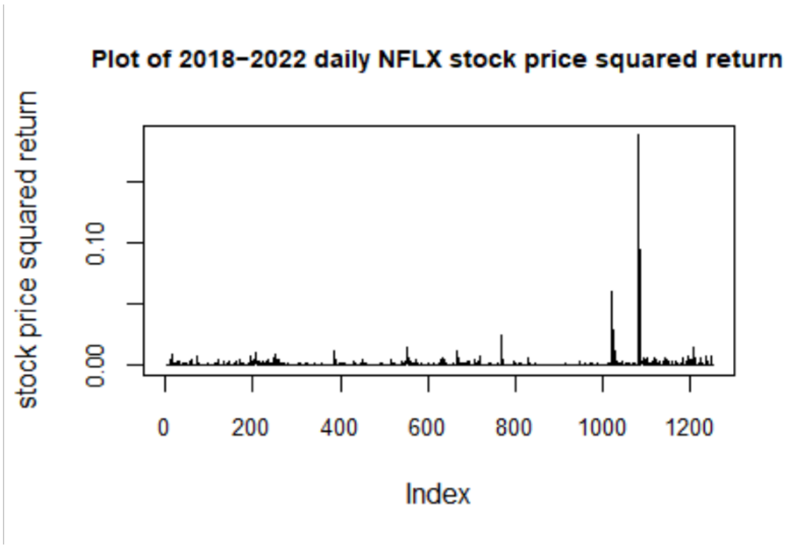
- **Descriptive statistics of return**

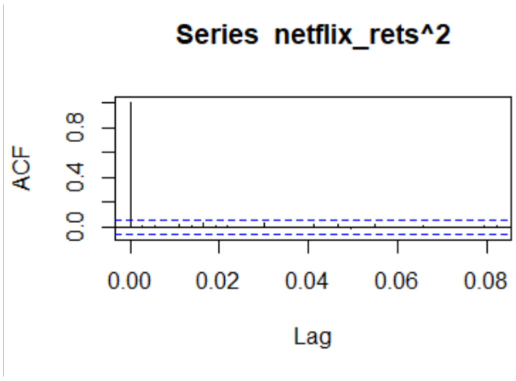
Table C.4: Descriptive statistics of return

Minimum	Maximum	Median	Variance	Skewness	kurtosis
-0.432578	0.155758	0.000448	0.000959	-2.429998	34.179283

Table C.5: Ljung-Box tests for squared return

Stock price squared return ACF 、PACF





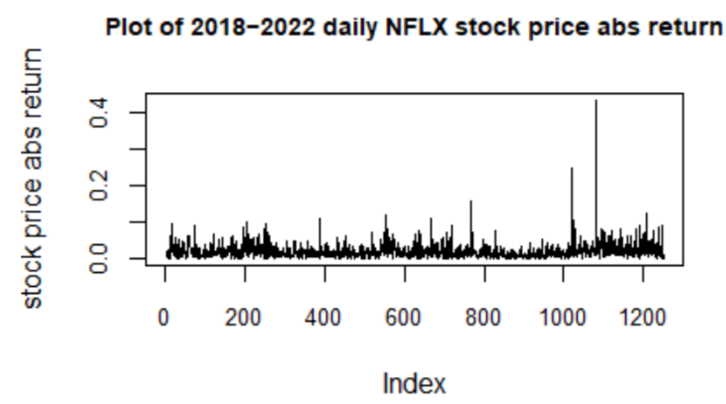
Ljung-Box test

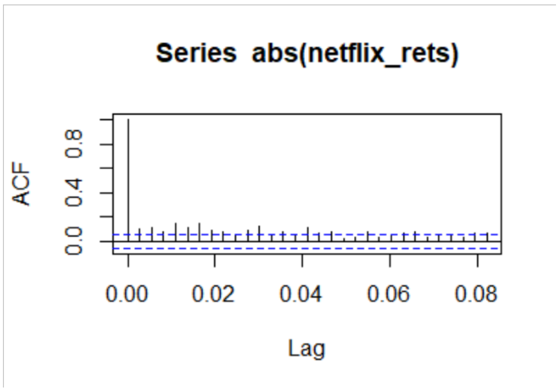
H_0: returns are independent vs H_1: returns are dependent

p-value = 0.9692 > 0.05, do not reject H_0, which implied squared returns are independent. This is not common on a return series, usually there is strong autocorrelation among the values of squared returns.

Table C.6: Ljung-Box tests for absolute return

Stock price abs return ACF





Ljung-Box test

H_0: returns are independent vs H_1: returns are dependent

p-value <0.05, reject H_0, which implied absolute values of returns are not independent.

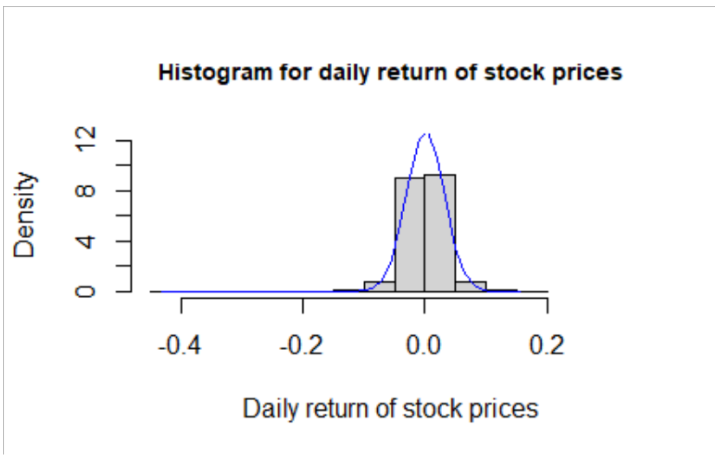


Figure C.5: Histogram of Netflix stock price return

Also, the return has a very large kurtosis (= 34.1793). High kurtosis of the return distribution curve implies that there have been many price fluctuations in the past away from the average returns.

From the above analyzing the feature of return, then, we further use the GARCH models to fit the return series, the candidates are shown as follows.

Table C.7: GARCH fitted model

Model1	arma(0,1) + garch(1,1) with cond.dist = "norm"	AIC = -4.1747
Model2	arma(1,1) + garch(1,1) with cond.dist = "std"	AIC = -4.4963

Since Model 2 has a better result with lower AIC, we will consider this model. The model summary and the tests for the residual of the model are shown as follows:

Table C.8: Fitted result of GARCH model 2

Model 2 : arma(1,1) + garch(1,1) with cond.dist = "std"

```
Error Analysis:
        Estimate  Std. Error  t value Pr(>|t|)
mu      1.449e-03  9.510e-04    1.524 0.127590
ar1    -5.623e-01  2.537e-01   -2.217 0.026650 *
ma1     5.220e-01  2.583e-01    2.021 0.043324 *
omega   1.253e-05  5.673e-06    2.208 0.027224 *
alpha1  5.268e-02  1.483e-02    3.552 0.000382 ***
beta1   9.347e-01  1.626e-02   57.495  < 2e-16 ***
shape   3.935e+00  4.302e-01    9.145  < 2e-16 ***
---

Log Likelihood:
 2819.451    normalized:  2.253758
```

```
Standardised Residuals Tests:
                                Statistic p-Value
 Jarque-Bera Test   R    Chi^2  88439.37  0
 Shapiro-Wilk Test  R    W      0.8252569 0
 Ljung-Box Test     R    Q(10)  6.998733  0.7255646
 Ljung-Box Test     R    Q(15)  12.14885  0.6677292
 Ljung-Box Test     R    Q(20)  19.39197  0.4964942
 Ljung-Box Test     R^2  Q(10)  0.3343718 0.9999991
 Ljung-Box Test     R^2  Q(15)  0.6738024 1
 Ljung-Box Test     R^2  Q(20)  1.148499  1
 LM Arch Test       R    TR^2   0.5566862 0.9999995

Information Criterion Statistics:
      AIC       BIC       SIC      HQIC
-4.496325 -4.467611 -4.496387 -4.485531
```

The arma(1,1) parameter estimators are

$\widehat{\varphi_0} = 0.001449, \widehat{\varphi_1} = 0.5623, \widehat{\theta_1} = 0.522$.

The GARCH(1,1) parameter estimators are

$\widehat{\alpha_0} = 0.00001253, \widehat{\alpha_1} = 0.05268, \widehat{\beta_1} = 0.9347.$

=> Just $\widehat{\varphi_0}$ is not significant, all the other parameter estimators are significant.

- **Regression coefficients:**

$$y_t = 0.001449 + 0.5623\,\varepsilon_{t-1} + 0.522\sigma_{t-1} + e_t$$
$$e_t = \sigma_t \varepsilon_t \;,\quad \varepsilon_t \sim \text{a distribution(mean=0, variance=1)}$$
$$\sigma_t^2 = 0.00001253 + 0.05268e_{t-1}^2 + 0.9347\sigma_{t-1}^2$$

- **Persistence** = 0.05268+0.9347 = 0.98738 ≈ 1

Jarque-Bera tests

H_0: residuals have the skewness and kurtosis matching a normal distribution.

H_1: residuals have some non − normal skewness and kurtosis.

=>From Jarque-Bera statistic tests, the residuals have some non-normal skewness and kurtosis.

Shapiro-Wilk test

H_0: residuals are normally distributed vs H_1 : residuals are not normally distributed

=>From Shapiro-Wilk statistic tests, the residuals appear to be not normal.

Ljung-Box test

H_0: residuals are independent (white noise) vs H_1: residuals are dependent

=> From Q-statistic (Ljung-Box statistic), we conclude that the squared residuals appear to be an uncorrelated sequence.

The following plot shows the prediction of model 2, the black line is the original data and the red line is the predicted result. The RMSE of the model is 9.825

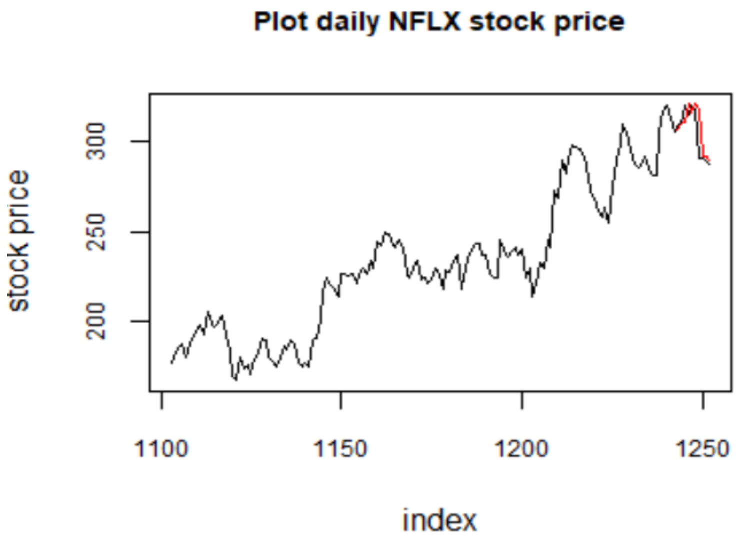


Figure C.6: Prediction plots for GARCH model

- <u>Holt's Linear Trend Model</u>

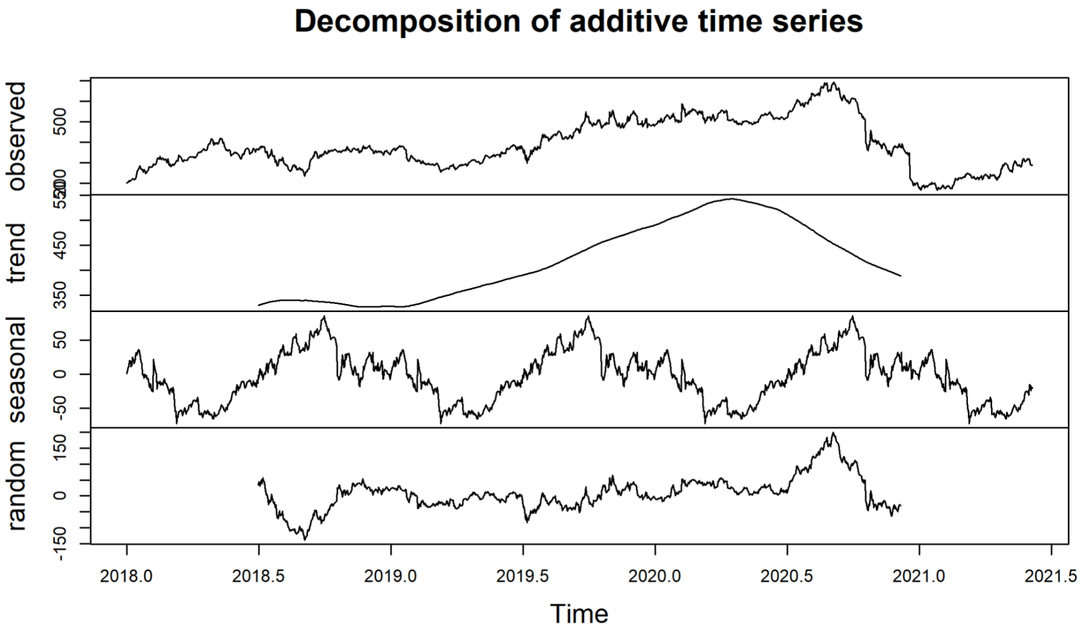


Figure C.7: Decomposition plot

Since this series only has both trend and seasonal patterns, we determine to use Holt's linear trend model.

The Holt's linear trend model fitting result is

Forecast equation: $\widehat{y_{t+h|t}} = \ell_t + hb_t$

Level equation: ℓ_t=9587y_t+(1-0.9587)(ℓ_t-1+b_t-1)

Trend equation: $b_t = 0.0027(\ell_t - \ell_{t-1}) + (1 - 0.0027)b_{t-1}$

Smoothing parameter : $\alpha = 0.9587$, $\beta^* = 0.0027$

Initial value : $\ell_0 = 199.3877$, $b_0 = 3.3101$

The AIC of this model is 14982.18

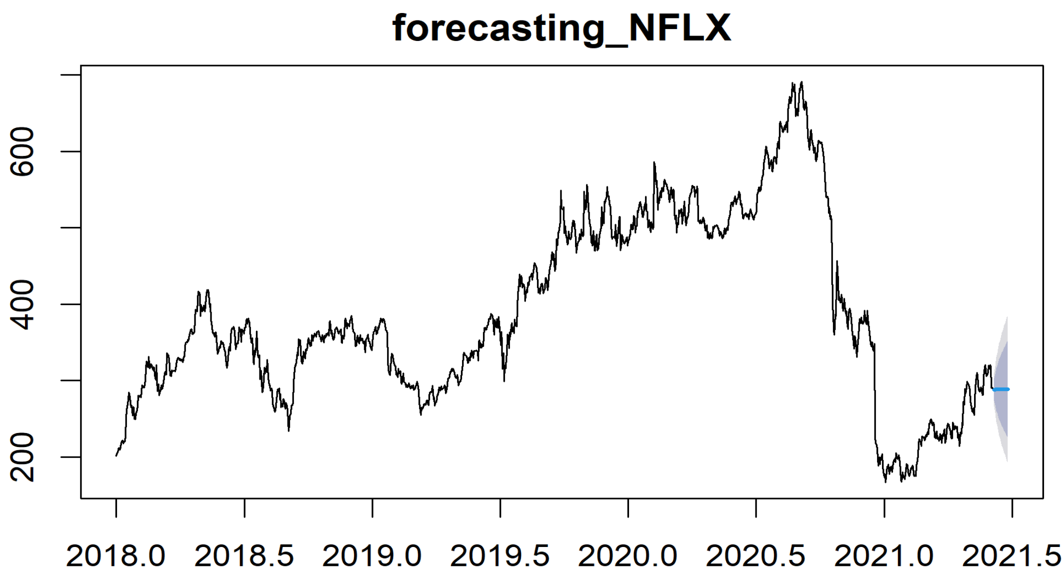


Figure C.8: Prediction plots for Holt's model

The predicted results are presented in the figure below. The blue line is the predicted value, the dark gray lines are an 80% confidence interval, and the light gray lines are a 95% confidence interval. We can see that the predicted trend is down, and our actual data is down, too.

D. Conclusion:

Table C.9: Comparison table

Model	**ARIMA(0,1,0) Model**	**GARCH Model**	**Holt's Model**
AIC	9198.94	-4.4963	14982.18

We fit different models by using the closing price of Netflix stock price, and the model AIC is shown above. In the table, the GARCH model has the lowest AIC. However, we think that the model with negative AIC may contain over-fitting, thus the GARCH model will not be taken into comparison and will choose the ARIMA model as our final model.

To conclude, the random walk model indicates that the pattern of the stock price is random, which means it cannot be predicted. Due to the uncertainty and numerous factors that could possibly affect the stock price, we believe that the Netflix stock price is mostly unpredictable.

E. Reference:

https://ir.netflix.net/ir-overview/profile/default.aspx
https://reurl.cc/EXvOlv
https://www.investor.gov/introduction-investing/investing-basics/investment-products/stocks
https://www.investopedia.com/terms/k/kurtosis.asp