Time Series Analysis Project

Chia-Chien Chou

1. Introduction

In this final project, I perform an analysis of the daily closing prices of NASDAQ-100 (NDX) stock index. According to Wikipedia, "NASDAQ-100 is a stock market index of 100 of the largest domestic and international non-financial companies listed on the NASDAQ." I adopt Time Series methods to analyze the 5-year (from January 1, 2005 to March 31, 2010) stock price of NASDAQ-100, and observe the relationship between stock prices of NASDAQ-100 and Google Inc. (Goog). The reason why I choose Google Inc. is Google Inc. is one of the components in the NASDAQ-100.

Furthermore, the analysis conducted in this project will be organized in two parts: standard analysis and specific analysis. In standard analysis, first, I will analyze the basic properties of the time series; next, the price process of NASDAQ-100 will be analyzed and modeled, also checked the model; finally, I will predict future values of the time series and a prediction interval. In specific analysis, I will divide the time series into subperiods, and fit models separately and compare each of the subperiods further. Last but not least, I want to explore the relationship of NASDAQ-100 and Google Inc..
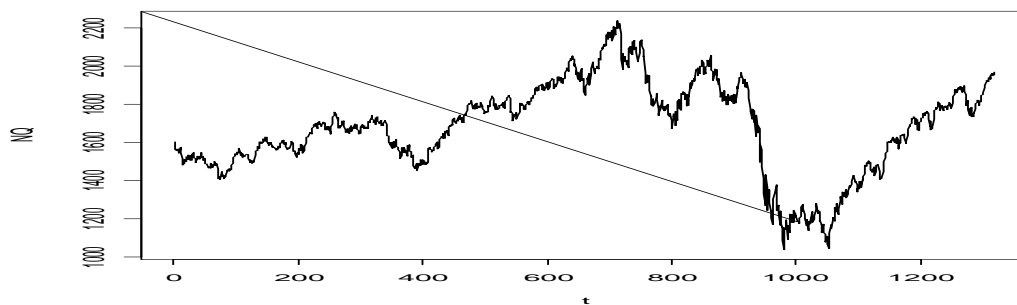
2. Standard Analysis

Figure 1: NDX Historical Price (from Jan 1, 2005 to Mar 31, 2010)
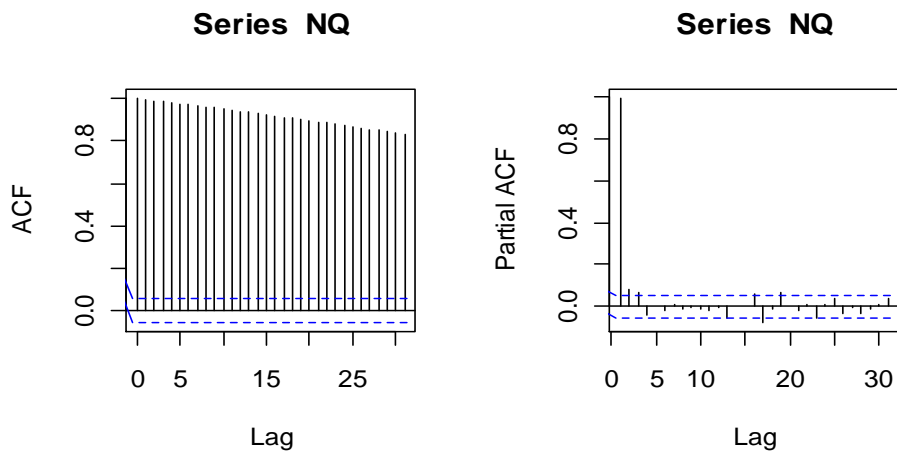
2.1 Basic Properties

Figure 2: ACF and PACF of $\{NDX_t\}$

From Figures 1 and 2, we can see that the ACF of the time series decline very slowly. This suggests that the time series behaves as random walk, which means it is not stationary.

2.2 Decomposition Model

Since from Figure 1 we can see that the time series has no seasonal pattern, we use the nonparametric method to decompose the time series into trend and random components. The Decomposition model is shown as followed:

$$NDX_t = m_t + Y_t$$

Also, From Figure 3, we can see that the time series have a stochastic trend.
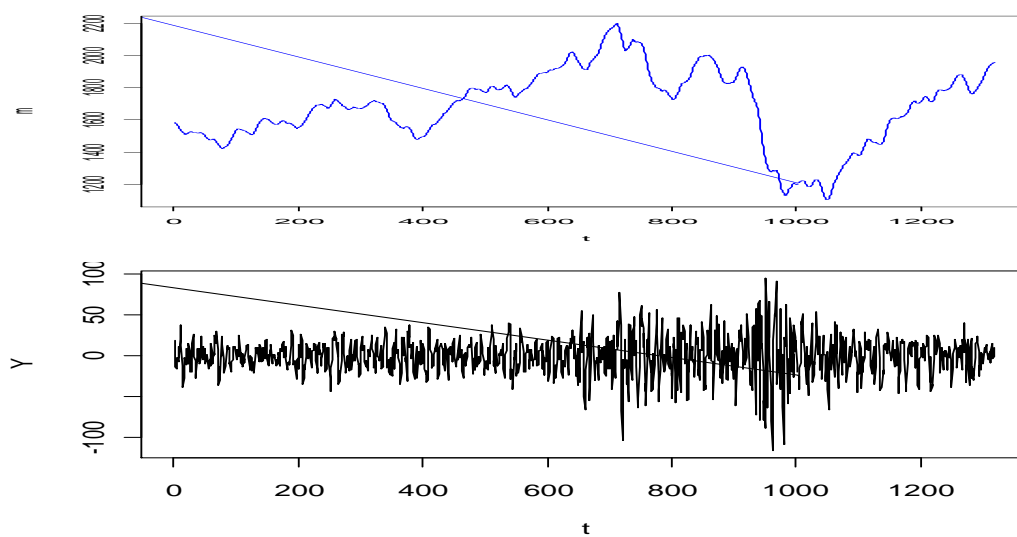

Figure 3: The trend component $\{m_t\}$ and the random component $\{Y_t\}$

2.3 Stationary

Since the original time series is not stationary, what we are going to do is to make the process stationary. First, we have to make a variance-stabilizing transformation by taking logarithms. Then, use the Rule of Thumb to determine how many times of differences we should do. From Table 1, we know that we have to take differencing once, i.e. make $\{NDX_t\}$ to be $\{\Delta \log NDX_t\}$. From Table 2, we can see that the result of Dickey-Fuller Test is that the time series $\{NDX_t\}$ is not stationary. However, after we take differencing once, from table 2 we can see that the series $\{\Delta \log NDX_t\}$ is stationary. Also, see Figure 4 and Figure 5, they show the time series $\{\Delta \log NDX_t\}$ is stationary.

```
> #rule of thumb
> sd(NQ)
[1] 235.7425
> sd(diff(NQ))
[1] 24.51142
> sd(diff(diff(NQ)))
[1] 36.3073
```

Table 1: Rule of Thumb Test

```
> adf.test(NQ)

        Augmented Dickey-Fuller Test

data:  NQ
Dickey-Fuller = -1.6087, Lag order = 10, p-value = 0.744
alternative hypothesis: stationary
> adf.test(diff(log(NQ)))

        Augmented Dickey-Fuller Test

data:  diff(log(NQ))
Dickey-Fuller = -10.7047, Lag order = 10, p-value = 0.01
alternative hypothesis: stationary
```
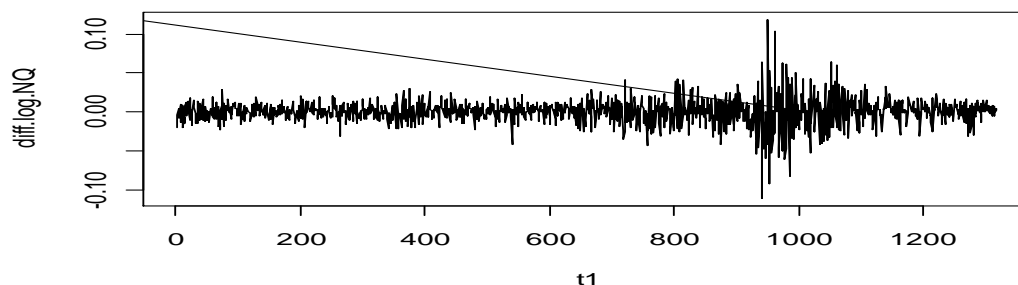
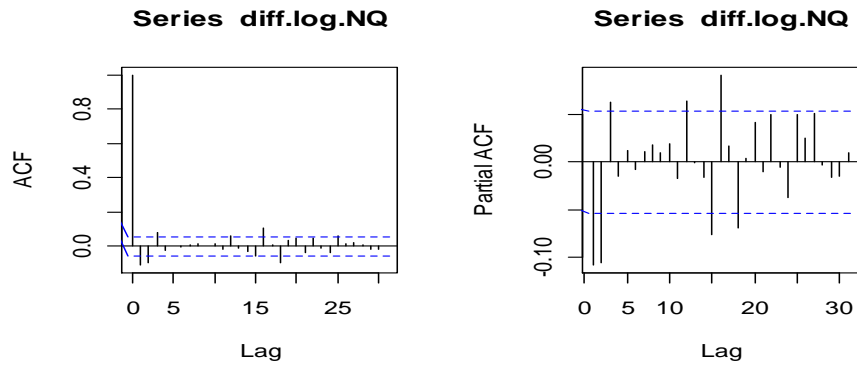Table 2: Dickey-Fuller Test

Figure 4: Plot of $\{\Delta \log NDX_t\}$

Figure 5: ACF and PACF of $\{\Delta \log \mathrm{NDX}_t\}$

2.4 Fit an ARIMA Model

After we do logarithm and difference to make the process a stationary model, we want to fit a model into ARIMA model. See Table 3, we find that the smallest AIC is -7201.201, the result is $\{\Delta \log \mathrm{NDX}_t\} \sim \mathrm{ARMA}(2,1)$. From Table 4, we can obtain the estimated coefficients of the model. The model is shown as following:

$$\Delta \log \mathrm{NDX}_t = -0.59 \Delta \log \mathrm{NDX}_{t-1} - 0.16 \Delta \log \mathrm{NDX}_{t-2} + Z_t + 0.48 Z_{t-1}$$

, where Z is the model residuals

```
> aic
          [,1]       [,2]       [,3]       [,4]
[1,] -7171.242 -7188.208 -7194.106 -7200.827
[2,] -7184.777 -7189.524 -7199.302 -7199.415
[3,] -7197.770 -7201.201 -7199.353 -7197.479
[4,] -7200.942 -7199.395 -7197.484 -7195.496
```

Table 3: AIC

```
> arima(diff.log.NQ,order=c(2,0,1),method="ML")

Call:
arima(x = diff.log.NQ, order = c(2, 0, 1), method = "ML")

Coefficients:
         ar1      ar2     ma1  intercept
      -0.5937  -0.1602  0.4811       2e-04
s.e.   0.1528   0.0278  0.1538       4e-04

sigma^2 estimated as 0.0002473:  log likelihood = 3605.6,  aic = -7201.2
```

Table 4: Result and coefficients of ARIMA model

4
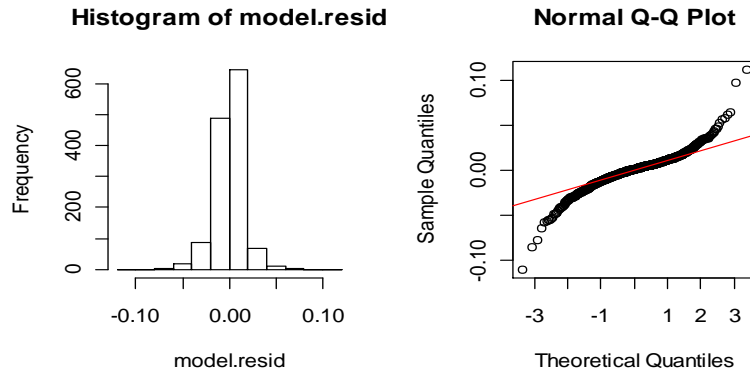
Figure 6: Histogram and QQ-plot of Residuals

```
> Box.test(model.resid, type=c("Ljung-Box"))

        Box-Ljung test

data:  model.resid
X-squared = 2e-04, df = 1, p-value = 0.9888

> skewness(model.resid)
[1] -0.2090489
> kurtosis(model.resid)
[1] 10.14522
```

Table 5: Skewness, Kurtosis, and L-B test of Residuals

As Table 5 implied, we know that the residuals are uncorrelated by Ljung-Box test. From Figure 6, we can see that the residuals are not normally distributed that means this model is not fitted, but in QQ-plot, the residuals seems quadratic. Thus, we do the ACF of squared residuals.
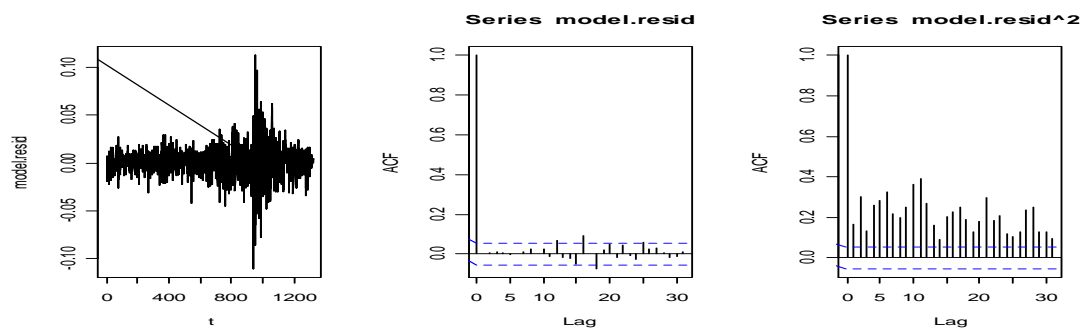

Figure 7: plot and ACF of residuals, and ACF of resid^2

See Figure 7 of the ACF of squared residuals, those autocorrelations are not large, but they are significant. They are all positive, which is uncommon in most economic time series yet is an implication of the GARCH(1,1) model. So, the ARIMA model is not complete, and we need to fit GARCH model.

5

2.5 Fit GARCH Model

We take the model residuals $\{Z_t\}$ to fit the GARCH(1,1) model, the result is in Table 6. The model is as following:

$$Z_t = \sigma_t e_t$$
$$\sigma_t^2 = 1.79e^{-06} + 0.0678Z_{t-1}^2 + 0.923\sigma_{t-1}^2$$

, where $\{e_t\}$ is GARCH(1,1) model residual.

```
Call:
garch(x = model.resid, order = c(1, 1))

Model:
GARCH(1,1)

Residuals:
    Min       1Q   Median       3Q      Max
-5.1584  -0.5877   0.1030   0.6486   3.1301

Coefficient(s):
     Estimate  Std. Error  t value  Pr(>|t|)
a0  1.787e-06   6.867e-07    2.603   0.00925 **
a1  6.777e-02   1.048e-02    6.467  9.97e-11 ***
b1  9.231e-01   1.301e-02   70.939   < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Diagnostic Tests:
        Jarque Bera Test

data:  Residuals
X-squared = 115.3919, df = 2, p-value < 2.2e-16


        Box-Ljung test

data:  Squared.Residuals
X-squared = 2.363, df = 1, p-value = 0.1242
```
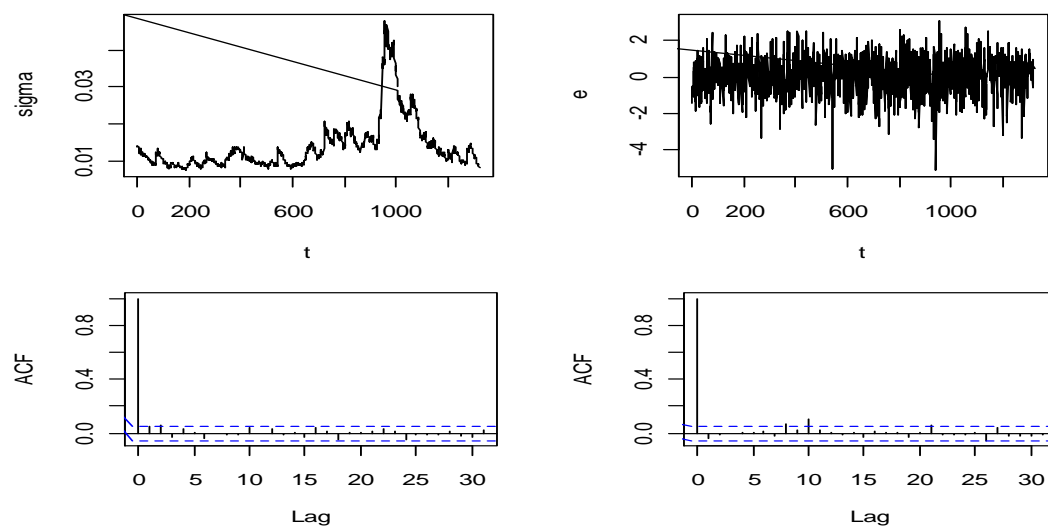
Table 6: Result of fitting GARCH(1,1)

Figure 8: plot of sigma and residuals {e}, and ACF of residual {e} and {e^2}

6

In Figure 8, we can see the residuals and squared residuals are stationary. Finally, we get a complete stationary model that is

$\{\Delta \log \text{NDX}_t\} \sim \text{ARMA}(2,1) - \text{GARCH}(1,1)$.

$$\Delta \log \text{NDX}_t = -0.59 \Delta \log \text{NDX}_{t-1} - 0.16 \Delta \log \text{NDX}_{t-2} + Z_t + 0.48 Z_{t-1}$$

$$Z_t = \sigma_t e_t$$

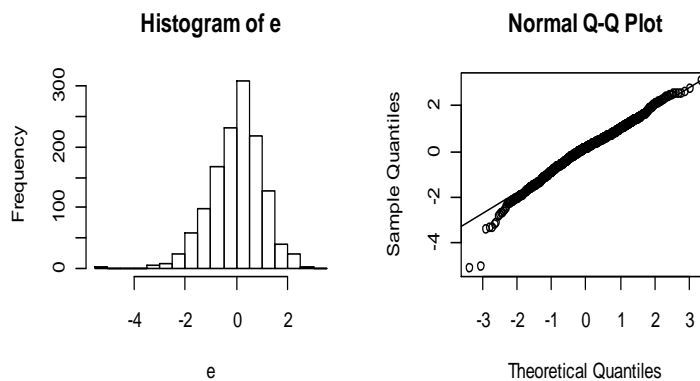$$\sigma_t^2 = 1.79e^{-06} + 0.0678 Z_{t-1}^2 + 0.923 \sigma_{t-1}^2$$

2.6 Model Checking

Figure 9: Histogram and QQ-plot of residuals{e}

```
> mean(e[2:n])
[1] 0.02521744
> sd(e[2:n])
[1] 0.9992923
> skewness(e[2:n])
[1] -0.4363721
> kurtosis(e[2:n])
[1] 4.156695

> shapiro.test(e)

        Shapiro-Wilk normality test

data:  e
W = 0.9865, p-value = 9.965e-10
```

Table 7: Result of skewness, kutosis, and Shapiro Test

Since $\alpha_1 + \beta \approx 0.99$, which means stationary, the model we get is a stationary and appropriate model. However, according to the Shapiro-Wilk test in Table 7 and Figure 9, we know that residual {e} seems not normally distributed. And from the skewness and the kurtosis in Table 7, {e} is more likely a t-distribution. So, maybe we need to fit tGARCH.
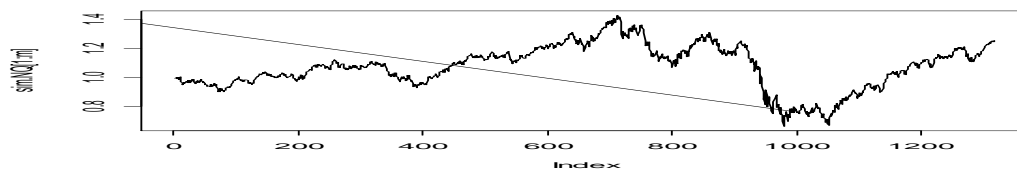
2.7 Simulate the time series
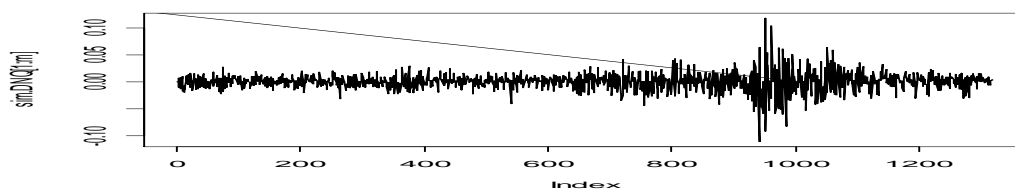
Figure 10: Simulation of $\{\Delta \log NDX_t\}$ and $\{NDX_t\}$

Substituting the model residuals {e} into the fitted stationary model, and then to simulate the original data. We can see that the simulation plot as in Figure 10 is similar to the original plot as in Figure 1, and the only difference between two figures are the scales of y-axis. So, I think the model I estimated is fitted.

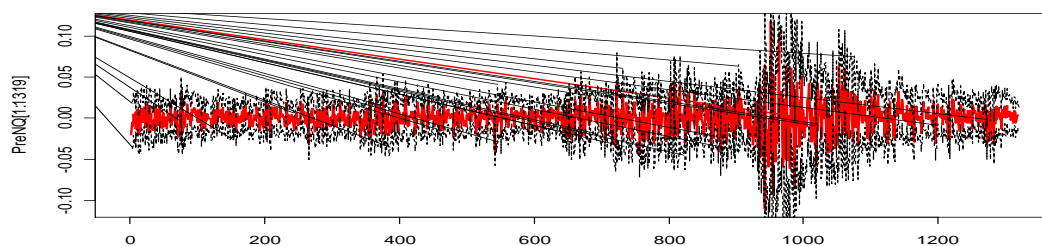2.8.1 Prediction(1): predict future values and prediction interval
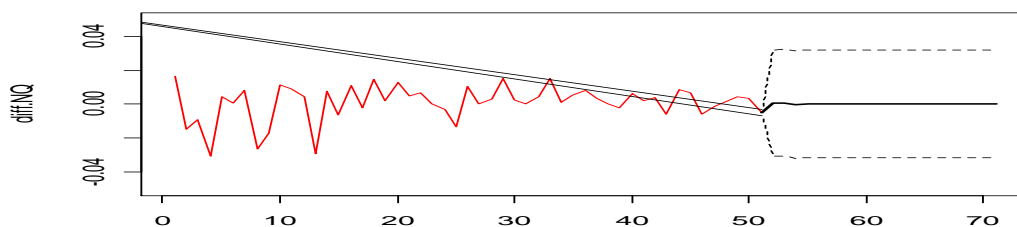
According to the plot, the values we predicted are all in the prediction interval.

2.8.2 Prediction(2): one-step ahead prediction & prediction interval
    Broad line - prediction of return model (ARMA(2,1)~GARCH(1,1))
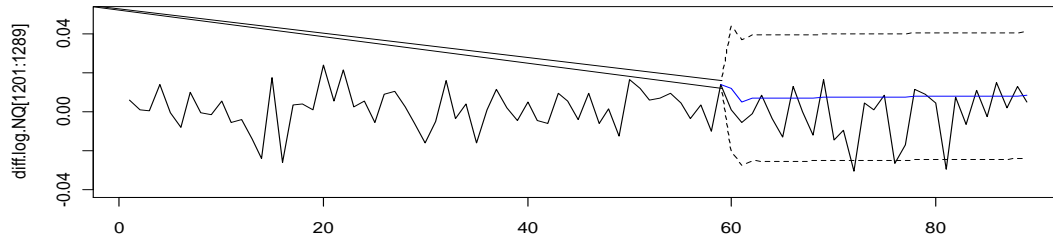   Dotted line - prediction interval (assumption − GARCH variance)

We can see the above plot is similar to the original time series $\{\Delta \log NDX_t\}$ as Figure 4.

2.8.3 Prediction(3):
    In-sample is the data from Jan 1,2005 − Dec 31,2009
    Out-of-sample is 30 observations (use in-sample part to predict 30 values.)

8

According to the plots, we can see there are three predicted values lies outside the prediction interval.

3. Specific Analysis

3.1 Subperiods

We divide the time series dataset into three periods: (1) Jan 2005-Dec 2007, (2) Jan 2008-Dec 2008, and (3) Jan 2009-Mar 2010.
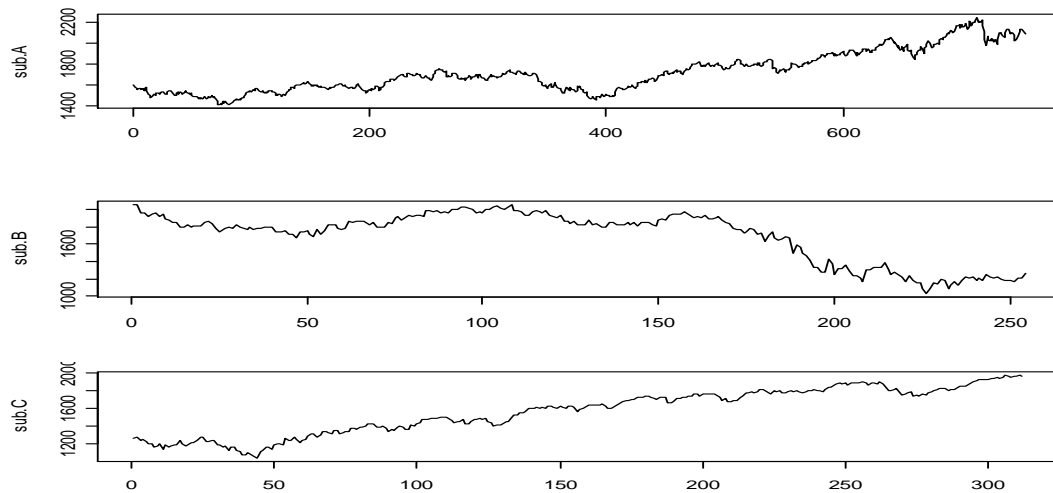

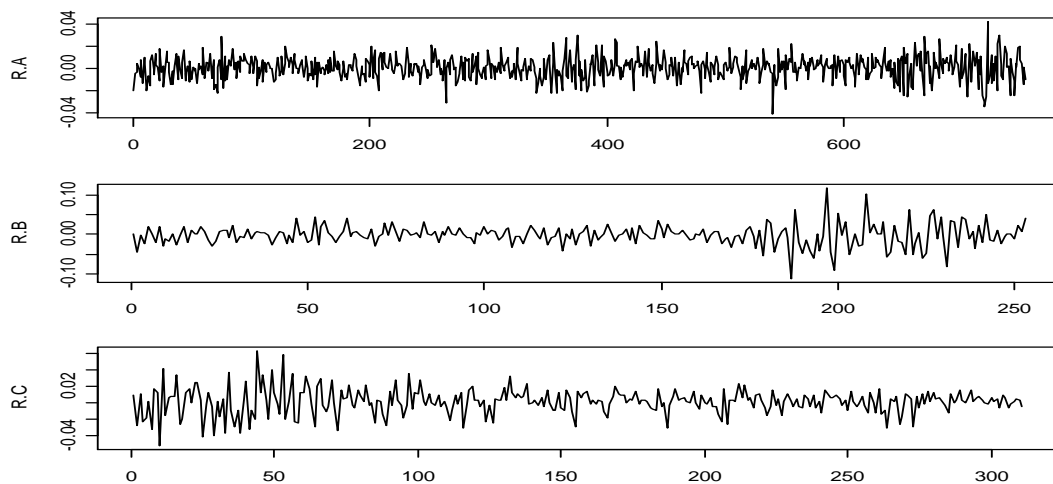
Figure 11: plot of 3 periods of original data

Figure 12: plot of 3 periods of returns data

9

From Figure 11 and Figure 12, we can see that three time series seem very different. Then, we fit models separately for the three subperiods. The result is shown as Table 8: subperiod(1) cannot fit any ARIMA model, subperiod(2) fits AR(2) model, and subperiod(3) fits ARMA(2,2) model. Since they are different models, and see the standard errors of the estimators in Table 8, we can conclusion that the models of subperiods differ significantly.

```
> arima(R.B, order=c(2,0,0))

Call:
arima(x = R.B, order = c(2, 0, 0))

Coefficients:
          ar1       ar2   intercept
      -0.1953   -0.1839     -0.0019
s.e.   0.0621    0.0623      0.0012

sigma^2 estimated as 0.0006704:  log likelihood = 565.38,  aic = -1122.76

> arima(R.C, order=c(2,0,2))

Call:
arima(x = R.C, order = c(2, 0, 2))

Coefficients:
          ar1       ar2       ma1       ma2   intercept
       1.4230   -0.8815   -1.5196    1.0000      0.0014
s.e.   0.0301    0.0286    0.0141    0.0162      0.0009

sigma^2 estimated as 0.0002248:  log likelihood = 862.56,  aic = -1713.13
```

Table 8: Result of fitting ARIMA model to each period
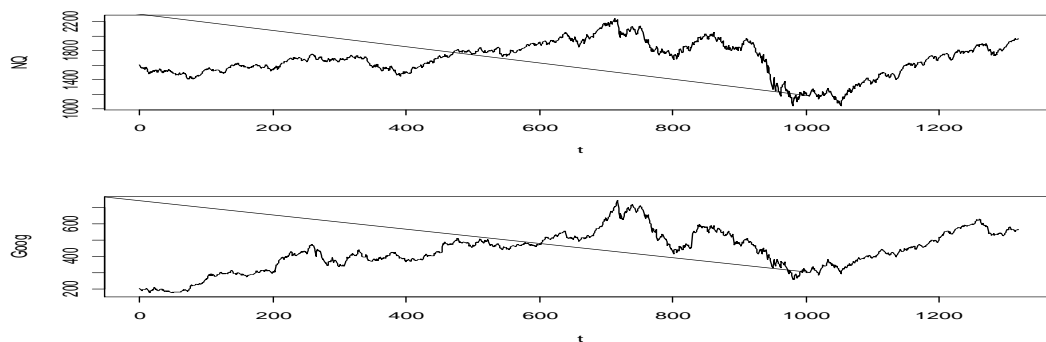
3.2 Multivariate Time Series: Fitting VARMA model
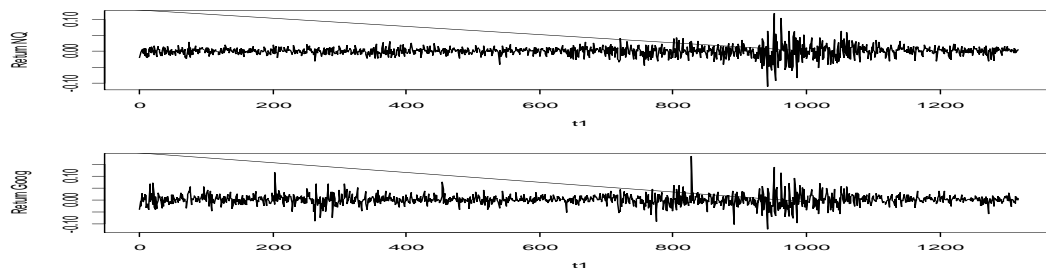

Figure 13: Plot of {NDX} and {Goog}

Figure 14: plot of $\{\Delta \log \text{NDX}_t\}$ and $\{\Delta \log \text{Goog}_t\}$

```
VAR Estimation Results:
===========================
Endogenous variables: diff.log.NQ, diff.log.Goog
Deterministic variables: const
Sample size: 1316
Log Likelihood: 7121.293
Roots of the characteristic polynomial:
0.4802 0.4802 0.3631 0.3589 0.3589 0.169
Call:
VAR(y = cbind(diff.log.NQ, diff.log.Goog), p = 3)


Estimation results for equation diff.log.NQ:
============================================
diff.log.NQ = diff.log.NQ.l1 + diff.log.Goog.l1 + diff.log.NQ.l2

              + diff.log.Goog.l2 + diff.log.NQ.l3 + diff.log.Goog.l3 + const


                  Estimate Std. Error t value Pr(>|t|)
diff.log.NQ.l1   -0.1220391  0.0375654  -3.249  0.00119 **
diff.log.Goog.l1  0.0080707  0.0260459   0.310  0.75671
diff.log.NQ.l2   -0.0809684  0.0374666  -2.161  0.03087 *
diff.log.Goog.l2 -0.0192650  0.0259517  -0.742  0.45801
diff.log.NQ.l3    0.1080473  0.0372468   2.901  0.00378 **
diff.log.Goog.l3 -0.0464851  0.0258732  -1.797  0.07262 .
const             0.0002378  0.0004352   0.546  0.58487
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1


Residual standard error: 0.01575 on 1309 degrees of freedom
Multiple R-Squared: 0.02977,    Adjusted R-squared: 0.02532
F-statistic: 6.693 on 6 and 1309 DF,  p-value: 5.417e-07


Estimation results for equation diff.log.Goog:
==============================================
diff.log.Goog = diff.log.NQ.l1 + diff.log.Goog.l1 + diff.log.NQ.l2 + diff.log.Goog.l2

                + diff.log.NQ.l3 + diff.log.Goog.l3 + const

                  Estimate Std. Error t value Pr(>|t|)
diff.log.NQ.l1   -0.0208644  0.0541365  -0.385  0.70000
diff.log.Goog.l1 -0.0068067  0.0375354  -0.181  0.85613
diff.log.NQ.l2    0.0057211  0.0539941   0.106  0.91563
diff.log.Goog.l2 -0.0105736  0.0373997  -0.283  0.77744
diff.log.NQ.l3    0.1616956  0.0536773   3.012  0.00264 **
diff.log.Goog.l3 -0.0864441  0.0372866  -2.318  0.02058 *
const             0.0008964  0.0006272   1.429  0.15315
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1


Residual standard error: 0.0227 on 1309 degrees of freedom
Multiple R-Squared: 0.007658,   Adjusted R-squared: 0.003109
F-statistic: 1.684 on 6 and 1309 DF,  p-value: 0.1214


Covariance matrix of residuals:
           diff.log.NQ diff.log.Goog
diff.log.NQ   0.0002481     0.0002423
diff.log.Goog 0.0002423     0.0005152

Correlation matrix of residuals:
           diff.log.NQ diff.log.Goog
diff.log.NQ    1.0000        0.6778
diff.log.Goog  0.6778        1.0000
```

Table 9: Results of VAR estimation

According to the results, which are shown in Table 9, some parameters are significantly different from 0 at the level 5%. And we get the estimated model is:

$$\Delta \log \mathrm{NDX}_t = -0.12 \, \Delta \log \mathrm{NDX}_{t-1} - 0.08 \, \Delta \log \mathrm{NDX}_{t-2} + 0.11 \, \Delta \log \mathrm{NDX}_{t-3}$$

$$\Delta \log \mathrm{Goog}_t = 0.162 \, \Delta \log \mathrm{NDX}_{t-3} - 0.086 \, \Delta \log \mathrm{Goog}_{t-3}$$

Variance matrix of $\mathbf{Z}_t = (Z_{1t}, Z_{2t})^T$:

$$\mathbf{\Sigma} = \begin{bmatrix} 2.481e-04 & 2.423e-04 \\ 2.423e-04 & 5.152e-04 \end{bmatrix}$$
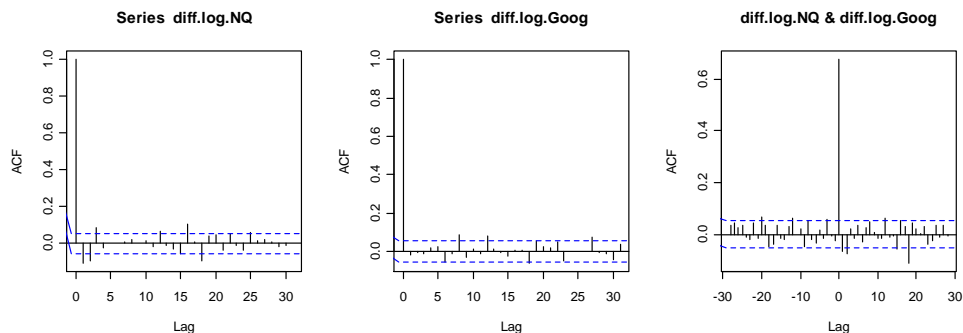

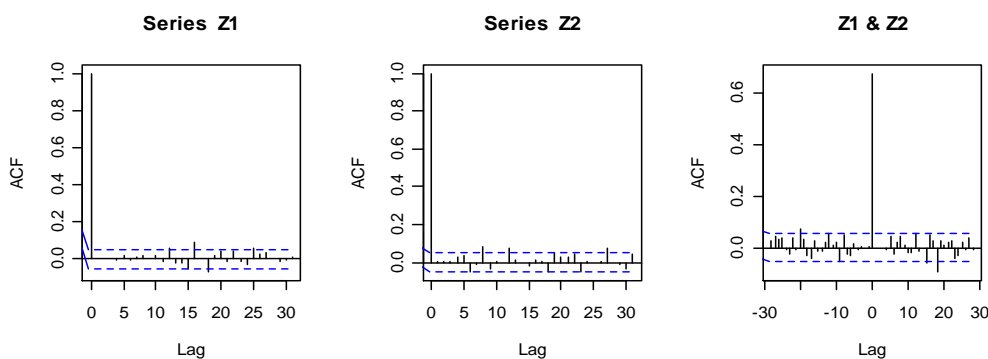Figure 15: ACF of $\{\Delta \log \mathrm{NDX}_t\}$ and $\{\Delta \log \mathrm{Goog}_t\}$

Figure 16: ACF of Z1, ACF of Z2, ACF of Z1 and Z2

We obtain the roots from Table 9, and all roots are greater than one. So, the model we get is stationary and causal. Also, compare Figure 15 to Figure 16, most of the autocorrelation and cross-correlation has been captured by the model and is not presented in the residuals. Therefore, the model above we obtained is a fitted model.

3.3 Transfer Function model

-Step (1): do Granger test, then define which one granger-causes another one.

```
> granger.test(cbind(diff.log.NQ,diff.log.Goog),p=3)
                          F-statistic    p-value
diff.log.Goog -> diff.log.NQ   1.283050  0.27869863
diff.log.NQ -> diff.log.Goog   3.176842  0.02333539
```

Table 9: Result of Granger Test

By the result of Granger test which is shown in Table 9, we know that $\{\Delta \log \mathrm{NDX}_t\}$

granger-causes $\{\Delta \log \text{Goog}_t\}$. So, I select the time series $\{\Delta \log \text{NDX}_t\}$ as the input and the time series $\{\Delta \log \text{Goog}_t\}$ as the output.

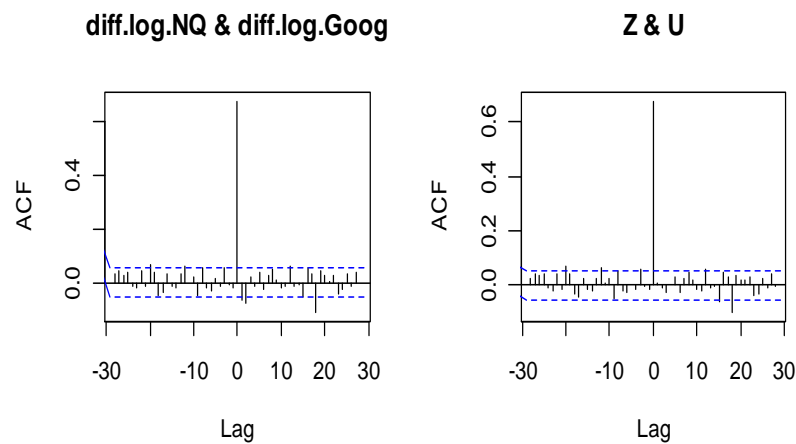-Step (2): identification of transfer function models

Figue 17: ACF of $\{\Delta \log \text{NDX}_t\}$&$\{\Delta \log \text{Goog}_t\}$ and ACF of $\{Z\}$&$\{U\}$

We get residuals-Z by fitting an ARMA model to the input series, and get filtered time series-U by filtering the output. Next, plot the ACF of {Z} and {U}. Comparing two plots in Figure 17, most of the cross-correlation has been captured by the model and is not presented in the resideuals. Since $\text{corr}(U_t, Z_{t+h}) \approx 0$ for all h > 0, the time series can be considered a causal transfer function model.

According to the results of transfer function model and multivariate time series model VAR(3), we obtain the same result that is two processes have one-way effect relationship.

4. Conclusion

In this analysis, we use several time series analysis methods to analyze a 5-year stock price of NASDAQ-100. We transform the data by taking log and difference to make the time series stationary, and take the data to apply the ARMA model first, and the GARCH model later. Although we can fit the time series as ARMA(2,1)-GARCH(1,1) model, the residuals are not normally distributed that against the assumptions of ARMA-GARCH model. This suggests that a different class of model will need to be used. Next, we try to investigate the relationship of NASDAQ-100 and Google. We fit a model that can show two processes are related.