STAT 463 – Extra Credit

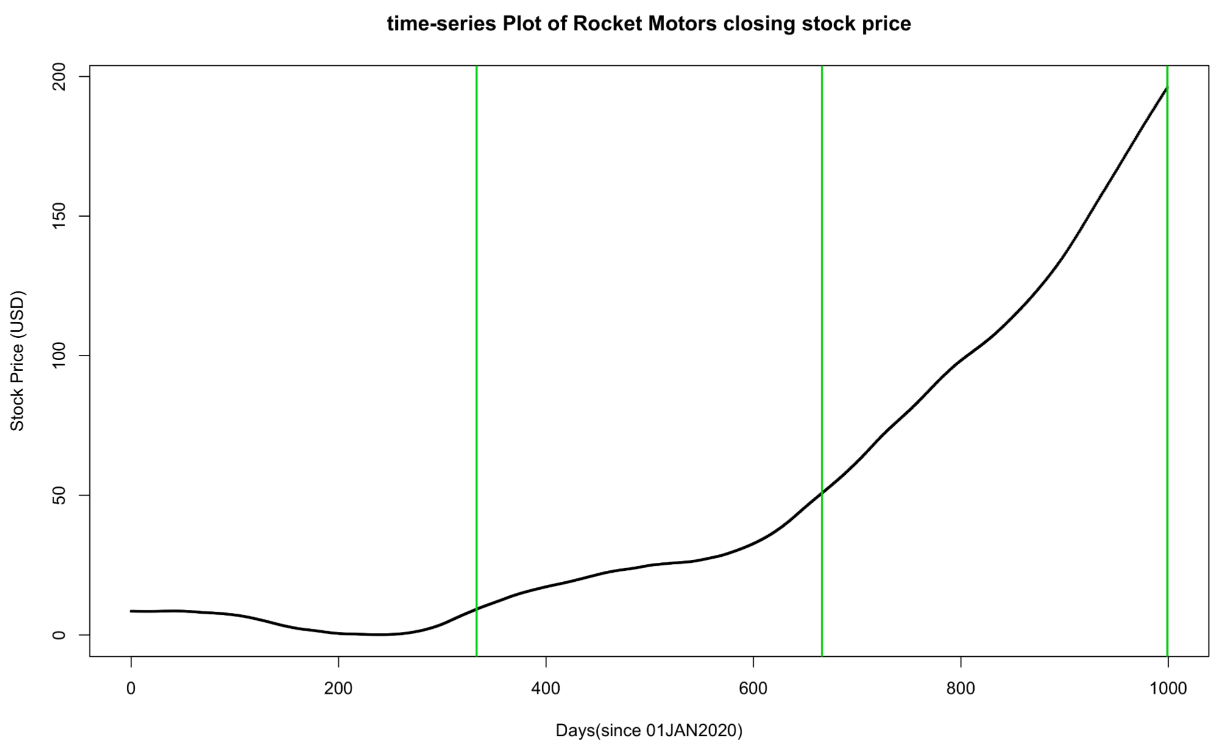
Due Date: Saturday, April 30th, 2023

Data Description and Background

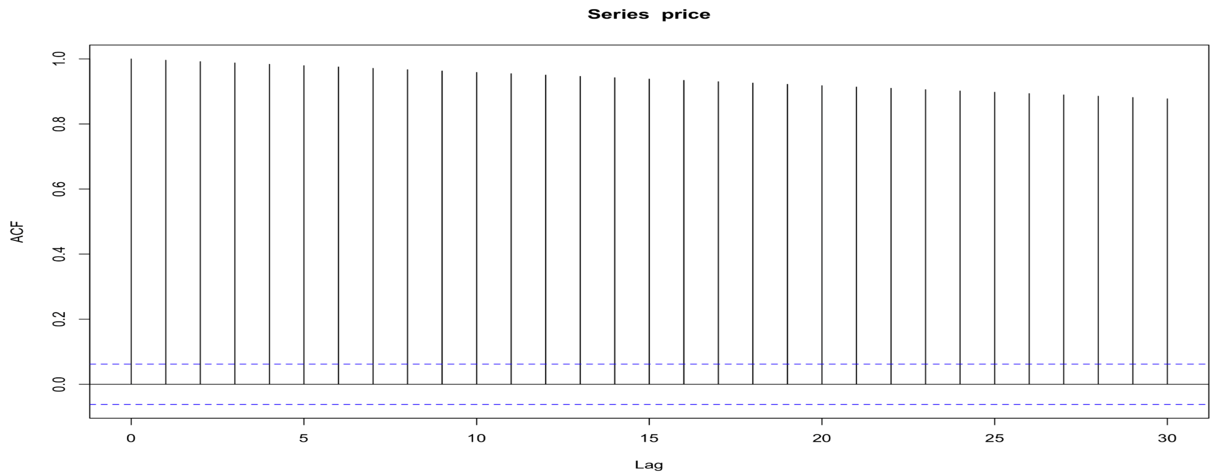
The dataset to be analyzed contains the closing stock price for Rocket Motors from 01JAN2020 to 26JAN2022.

1) Create a time-series plot of Rocket Motors closing stock price. Comment on anything of note. – 2 pts

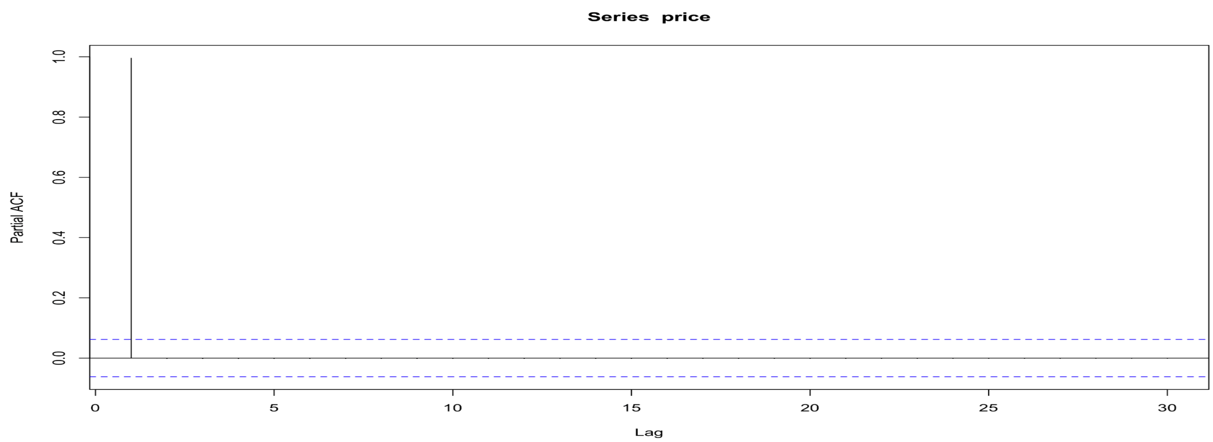
```
rm = Rocket_Motors.1
x = seq(0,999)
price = rm$Price_in_USD
plot(x, price[1:1000], type = "o", pch = 0.5, cex = 0.2, xlab = "Days(since 01JAN2020)", ylab = "Stock Price (USD)", main = "time-series Plot of Rocket Motors closing stock price")
abline(v = c(999*(1/3), 999*(2/3), 999), col = c(3, 3, 3), lwd = c(2, 2, 2))
```



>acf(price,type="correlation")



>acf(price,type="partial")



>adf.test(price)

Augmented Dickey-Fuller Test

data: price

Dickey-Fuller = 0.65879, Lag order = 9, p-value = 0.99

alternative hypothesis: stationary

From the time series plot we are able to see there is an upward trend in data. The ACF plot and Partial ACF plot show that there are significant autocorrelations in the data. To remove autocorrelation we should difference the data.

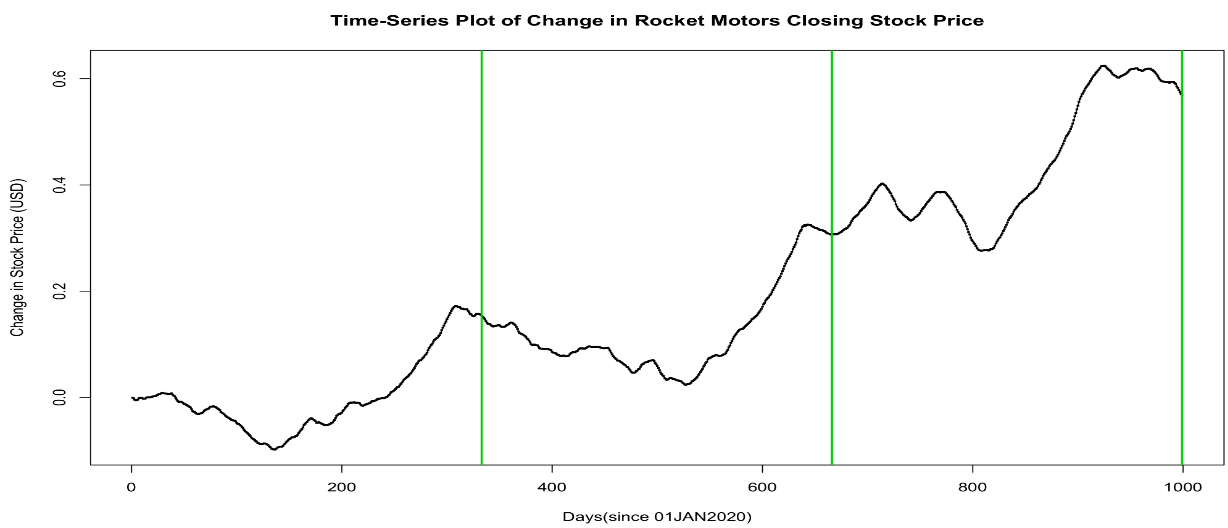
The Augmented Dickey-Fuller Test is a stationary Unit Root test where the null hypothesis is that the data is non stationary. Since the pval of 0.99 is greater than the alpha value of 0.05 we fail to reject the null hypothesis and we can conclude that the trend is non-stationary. If we are to fit an ARIMA model to the data, it should be differenced.

2) It is needed to fit an ARIMA(p,d,q) model to this data. What are the appropriate AR and MA orders as well as the degree of differencing needed (i.e. what are p, q and d)? Explain your rationale for arriving at the model, providing any graphs and outputs that were used to inform your decision. – 4 pts

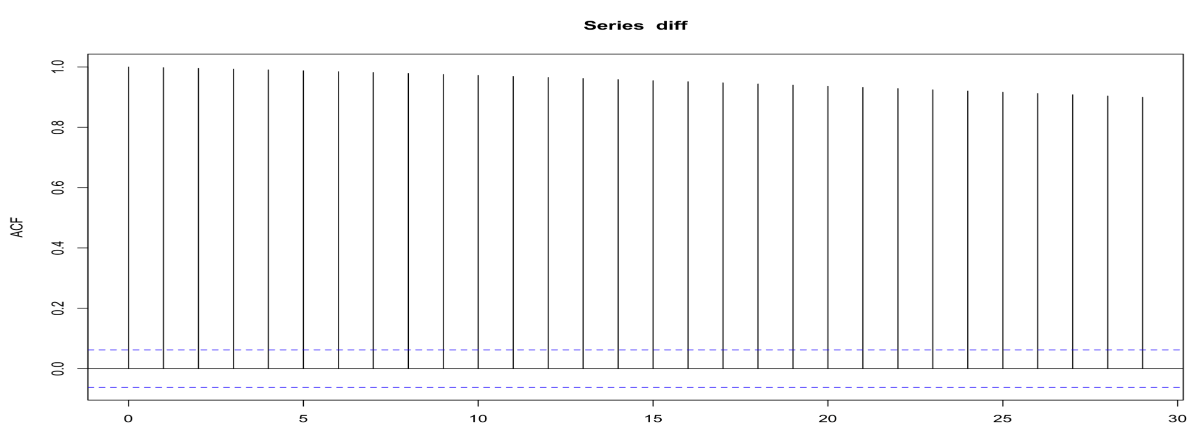
In order to fit an ARIMA(p,q,d) model, the time series plot needs to satisfy stationarity assumptions. We must employ differencing by subtracting each observation from the previous observation for increasing degree order until the series becomes stationary according to the adf test.

First order difference:

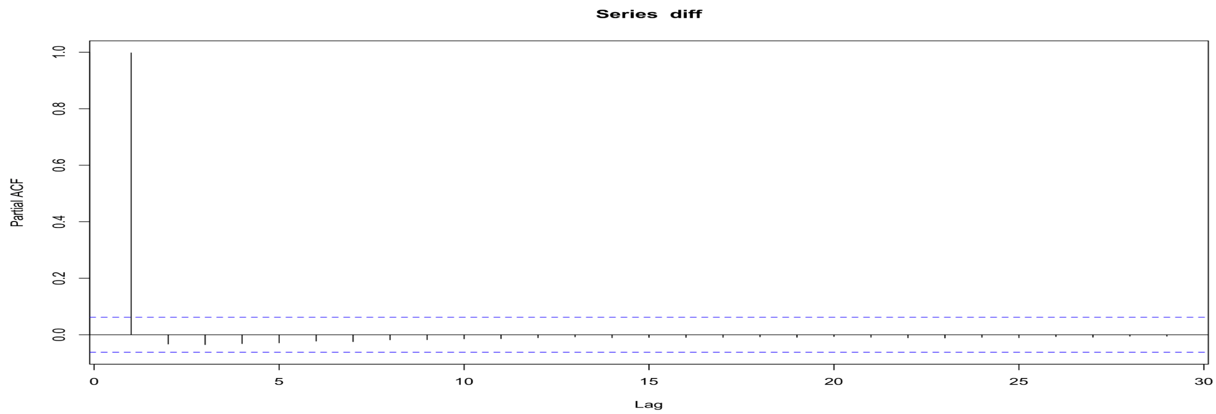
```
diff = (price[2:length(price)])-(price[1:(length(price)-1)])
plot(diff, type = "o", pch = 0.5, cex = 0.2, xlab = "Days(since 01JAN2020)", ylab = "Change in Stock Price (USD)", main = "Time-Series Plot of Change in Rocket Motors Closing Stock Price")
abline(v = c(999*(1/3), 999*(2/3), 999), col = c(3, 3, 3), lwd = c(2, 2, 2))
```




```
>acf(diff,type="correlation")
```



>acf(diff,type="partial")



>adf.test(diff)

Augmented Dickey-Fuller Test

data: diff

Dickey-Fuller = -2.7407, Lag

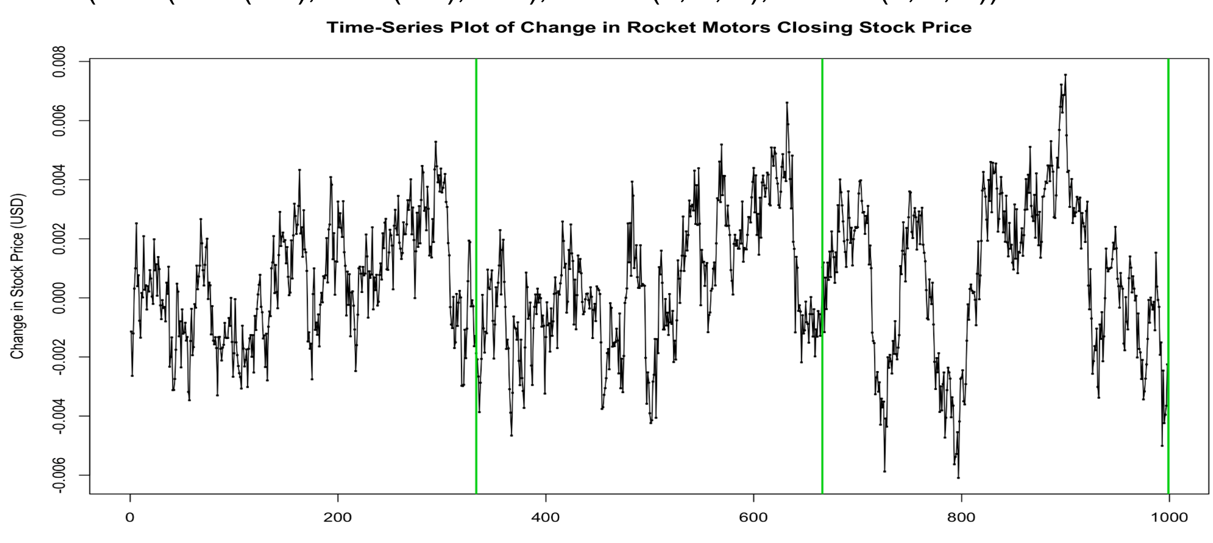
order = 9, p-value = 0.2648

alternative hypothesis: stationary

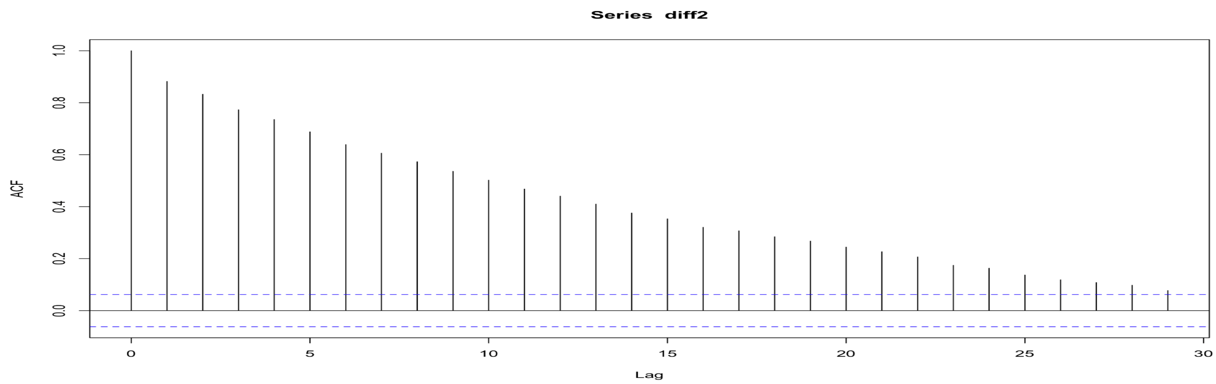
The first order difference has ACF plots that show significant auto correlations and the adf test returns a pval of 0.2648 which is greater than the alpha value of 0.05 we fail to reject the null hypothesis and we can conclude that the trend is non-stationary.

We continue to second order difference:

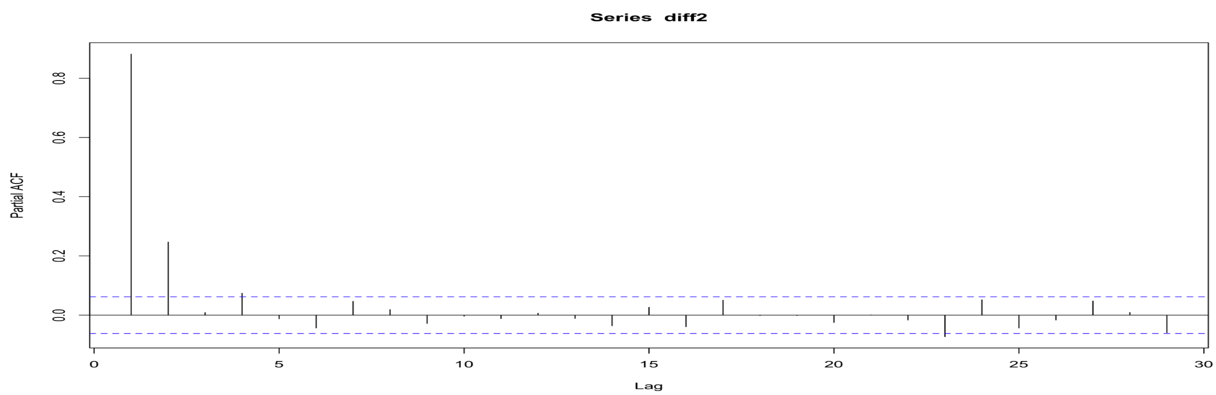
```
diff2 = (diff[2:length(diff)])-(diff[1:(length(diff)-1)])
plot(diff2, type = "o", pch = 0.5, cex = 0.2, xlab = "Days(since 01JAN2020)",
    ylab = "Change in Stock Price (USD)", main = "Time-Series Plot of Change in Rocket Motors Closing Stock Price")
abline(v = c(999*(1/3), 999*(2/3), 999), col = c(3, 3, 3), lwd = c(2, 2, 2))
```



>acf(diff2,type="correlation")



>acf(diff2,type="partial")



>adf.test(diff2)

Augmented Dickey-Fuller Test

data: diff2

Dickey-Fuller = -4.8479, Lag order = 9, p-value = 0.01

alternative hypothesis: stationary

The adf test returns a pval of 0.01 which is less than the alpha value of 0.05 we reject the null hypothesis and we can conclude that the trend is stationary.

Now that the stationary assumption has been fulfilled we can continue to fit an ARIMA model. The ARIMA model is composed of p (the order of the autoregressive process), d (the degree of differencing), q (the order of the moving average process)

We already completed differencing to the degree of 2 to assume stationarity, so the d parameter will be 2. Therefore we start our search for the best model with ARIMA(0,2,0) model and proceed by testing various orders of the AR (lagged values of the variable being predicted) and MA (lagged forecast errors, which represent the difference between the predicted value and the actual value) components and choosing the model with the lowest AIC (Akaike Information Criterion).

```
> AR1 = arima(x = diff2, order = c(1, 0, 0), include.mean = FALSE, method ="ML")

Coefficients:
         ar1
      0.8895
s.e.  0.0144

sigma^2 estimated as 1.142e-06:  log likelihood = 5410.98,  aic = -10817.97

> AR2 = arima(x = diff2, order = c(2, 0, 0), include.mean = FALSE, method ="ML")

Coefficients:
         ar1     ar2
      0.6592  0.2590
s.e.  0.0306  0.0306

sigma^2 estimated as 1.065e-06:  log likelihood = 5445.49,  aic = -10884.99

> AR3 = arima(x = diff2, order = c(3, 0, 0), include.mean = FALSE, method ="ML")

Coefficients:
         ar1     ar2     ar3
      0.6553  0.2491  0.0151
s.e.  0.0316  0.0371  0.0317

sigma^2 estimated as 1.065e-06:  log likelihood = 5445.6,  aic = -10883.21

> MA1 = arima(x = diff2, order = c(0, 0, 1), include.mean = FALSE, method ="ML")

Coefficients:
         ma1
      0.6522
s.e.  0.0177

sigma^2 estimated as 2.758e-06:  log likelihood = 4971.33,  aic = -9938.67

> MA2 = arima(x = diff2, order = c(0, 0, 2), include.mean = FALSE, method ="ML")

Coefficients:
         ma1     ma2
```

```
     0.7976  0.5612
s.e.  0.0287  0.0219

sigma^2 estimated as 1.832e-06:  log likelihood = 5175.13,  aic = -10344.26

> MA3 = arima(x = diff2, order = c(0, 0, 3), include.mean = FALSE, method ="ML")

Coefficients:
        ma1     ma2     ma3
     0.8566  0.7081  0.3224
s.e.  0.0322  0.0292  0.0257

sigma^2 estimated as 1.588e-06:  log likelihood = 5246.56,  aic = -10485.12
```

<u>**Based of the lowest AIC values, the most appropriate are AR(2) and MA(3)**</u>

<u>Now we fit an ARIMA model with all the appropriate pdq values:</u>

```
> ARIMA = arima(x = diff2, order = c(2, 2, 3), include.mean = FALSE, method ="ML")
Coefficients:
         ar1     ar2      ma1      ma2     ma3
     -0.3208  0.2390  -1.0010  -0.3225  0.3235
s.e.   0.3054  0.1211   0.3032   0.3360  0.0953
sigma^2 estimated as 1.087e-06:  log likelihood = 5421.38,  aic = -10830.76
```

<u>With the library(forecast) we can run the function auto.arima on the original non stationary price series</u> and the function will difference the data and output the best model by using a variation of the Hyndman-Khandakar algorithm (Hyndman & Khandakar, 2008), which combines unit root tests, minimisation of the AICc and MLE to obtain an ARIMA model.

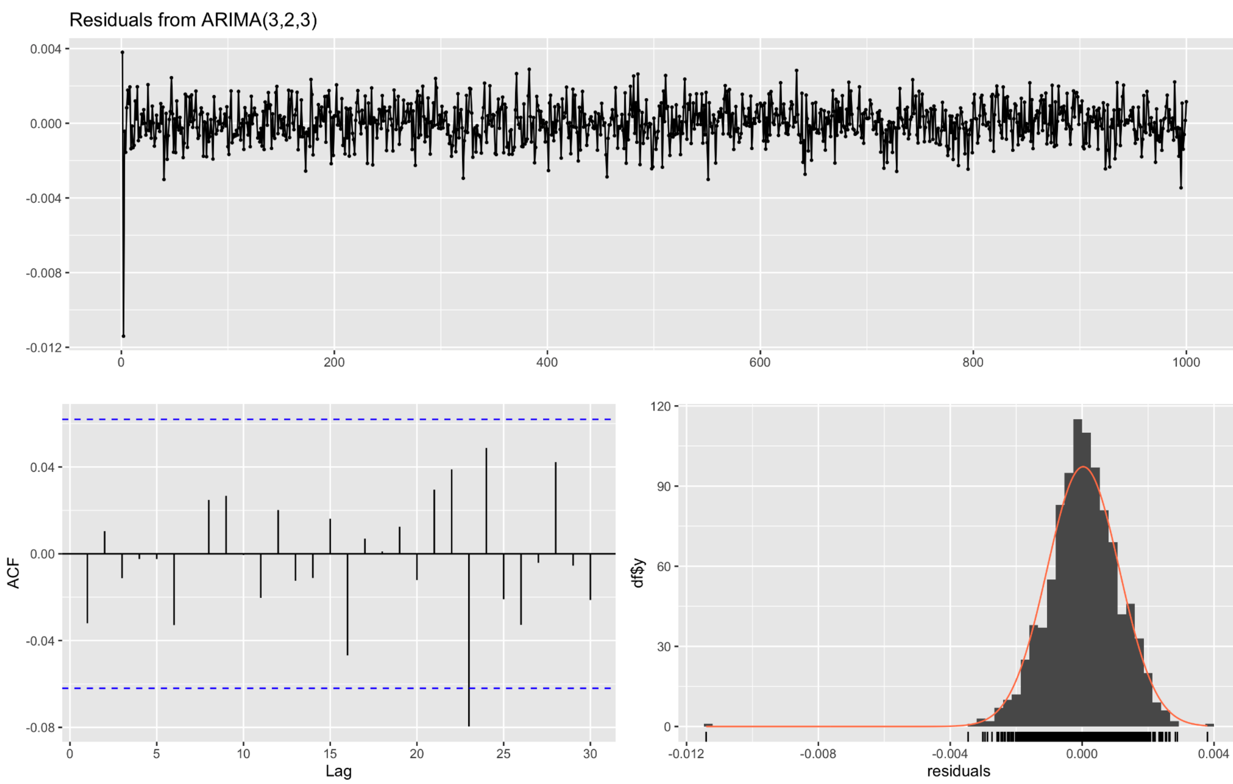
```
> auto = auto.arima(price)

Series: price
ARIMA(3,2,3)
Coefficients:
         ar1     ar2     ar3     ma1     ma2      ma3
     -0.0505  0.3119  0.5983  0.7074  0.3866  -0.2218
s.e.   0.2754  0.1366  0.2778  0.2753  0.2245   0.0703

sigma^2 = 1.208e-06:  log likelihood = 5450.05
AIC=-10886.1   AICc=-10885.99   BIC=-10851.76
```

The auto.arima model using the price series has a smaller AIC value that the ARIMA(2,2,3) model using the diff2 series.

<u>Ideally if this is a good model, the residuals will be white noise, with no autocorrelation:</u>
>checkresiduals(auto)

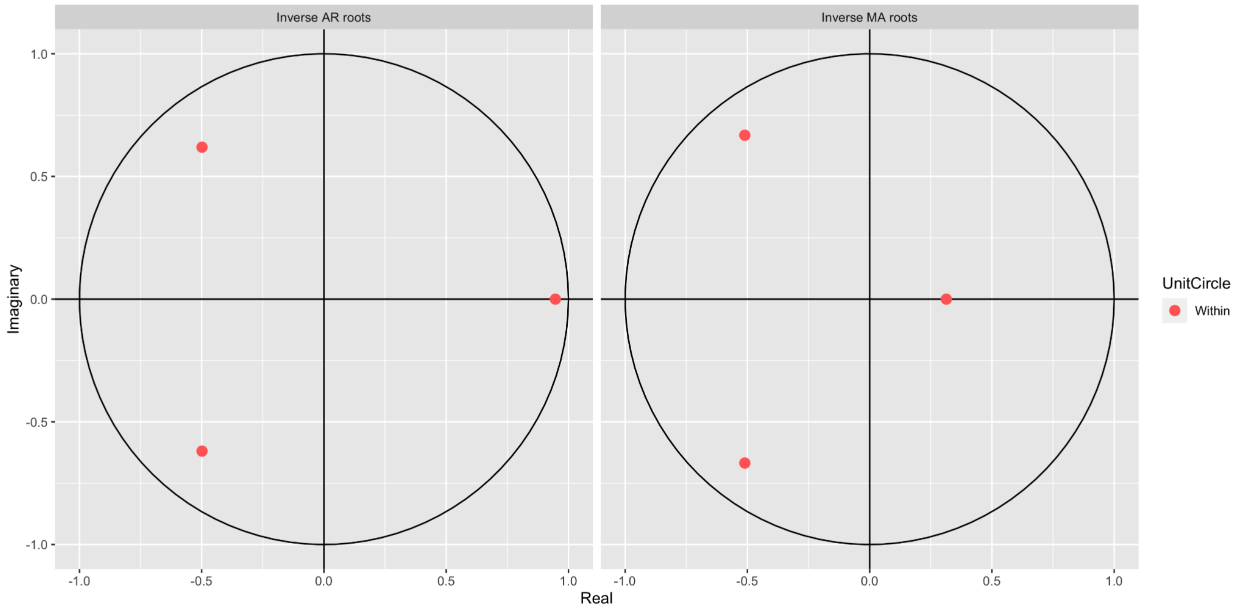
Ljung-Box test
data: Residuals from ARIMA(3,2,3)
Q* = 3.705, df = 4, p-value = 0.4474
Model df: 6.   Total lags used: 10

The null hypothesis of the Ljung-Box test is that there is no autocorrelation. Looking at the p-value above of 0.4475, we can see that it is greater than a= 0.05. Therefore, we cannot reject the null hypothesis, and conclude the residuals are indeed not correlated.



The characteristic polynomial of the AR part is denoted by φ(B), where B is the backward shift operator and the characteristic polynomial of the MA part is denoted by θ(B). The roots of these polynomials, values of b that satisfy θ(B) = 0 and φ(B) = 0, must lie within the unit circle when fitting an ARIMA model for forecasting. All the red dots plotted, the roots of φ(B) and θ(B), are inside the unit circle, which ensures the model auto.ARIMA(3,2,3) is both stationary and invertible.

>autoplot(auto)



3) Explicitly write out the model equations for the ARIMA(p, d, q) model you arrived at in part 2. – 1pt

formula:

$$y'_t = c + \phi_1 y'_{t-1} + \ldots + \phi_p y'_{t-p} + \theta_1 \epsilon_{t-1} + \ldots + \phi_q \epsilon_{t-q} + \epsilon_t$$

$y'_t = -0.0505y'_{t-1} + 0.3119y'_{t-2} + 0.5983y'_{t-2} + 0.7074e_{t-1} + 0.3866e_{t-2} - 0.2218e_{t-3} + e_t$