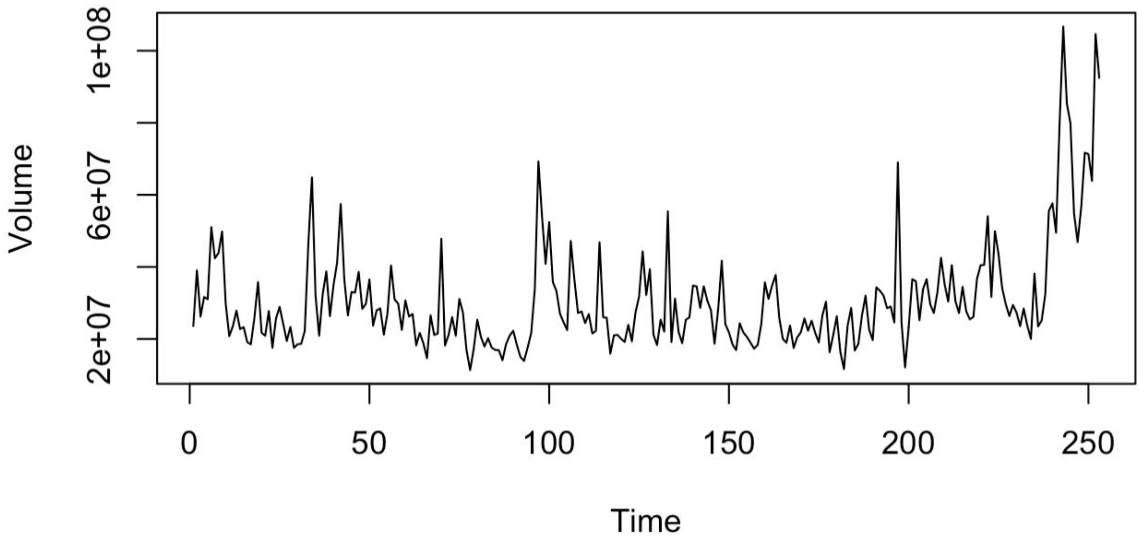
Analysis

[PLACE VOLUME TIME SERIES PLOT HERE]



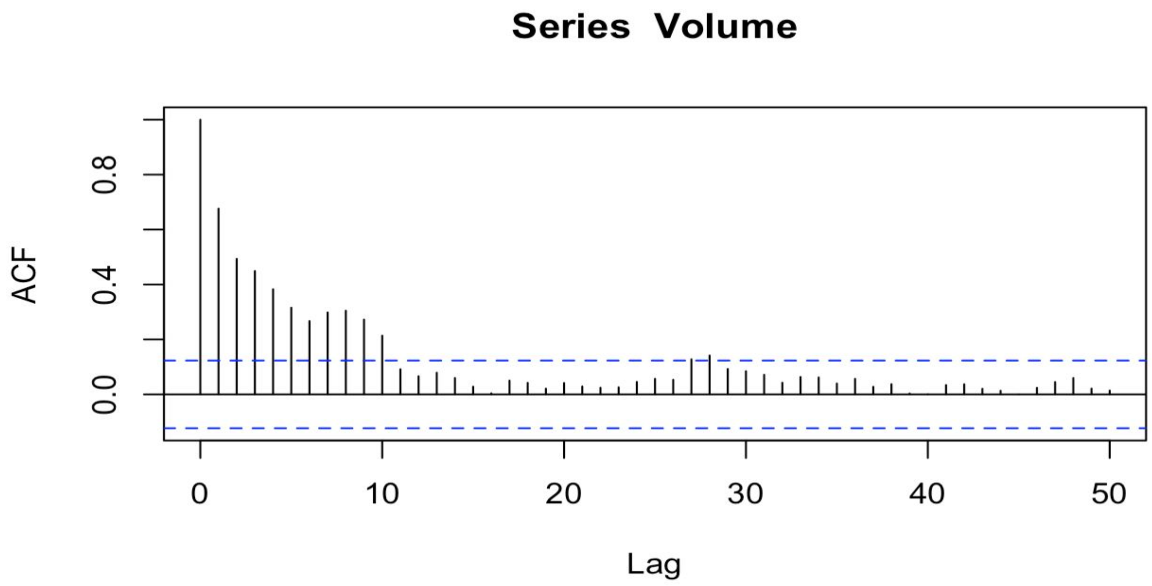
Time interval representation: at 0 it means the date Mar.13th 2019, at 250 it means the date Mar.10th 2020.

The first plot seen here is the raw time series plot of the daily volume of AAPL stocks traded. Visually, we can see that for the majority of the time, the daily volume has a constant mean aside from a few high spikes. Near the end of the plot, we see a massive increase in the daily volume (most likely due to COVID19). From the plot, we may say that there is some kind of cyclical effect and does not have a trend until near the end.

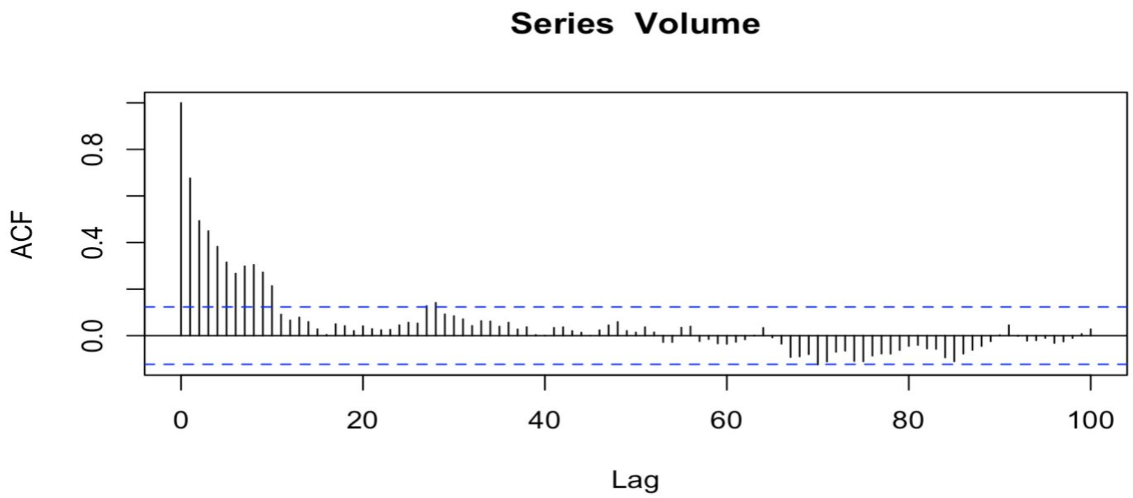
[PLACE ACF PLOT OF VOLUME TIME SERIES HERE]

ACF plot

Choosing lag.max=50



Choosing lag.max=100

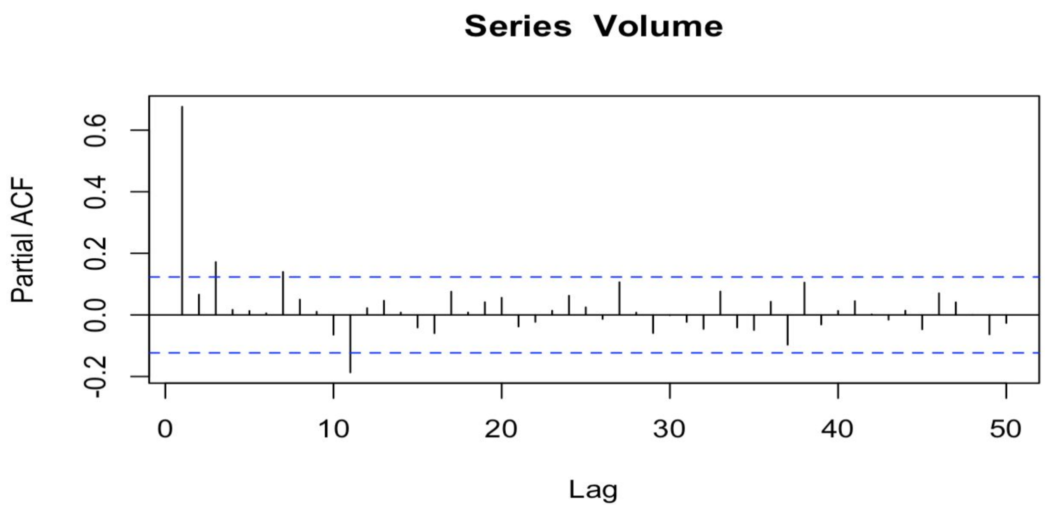


Plotting the ACF against the lag for the daily volume, we see that the ACF tails off at a moderate rate and is significant from lag 0 to lag 10.

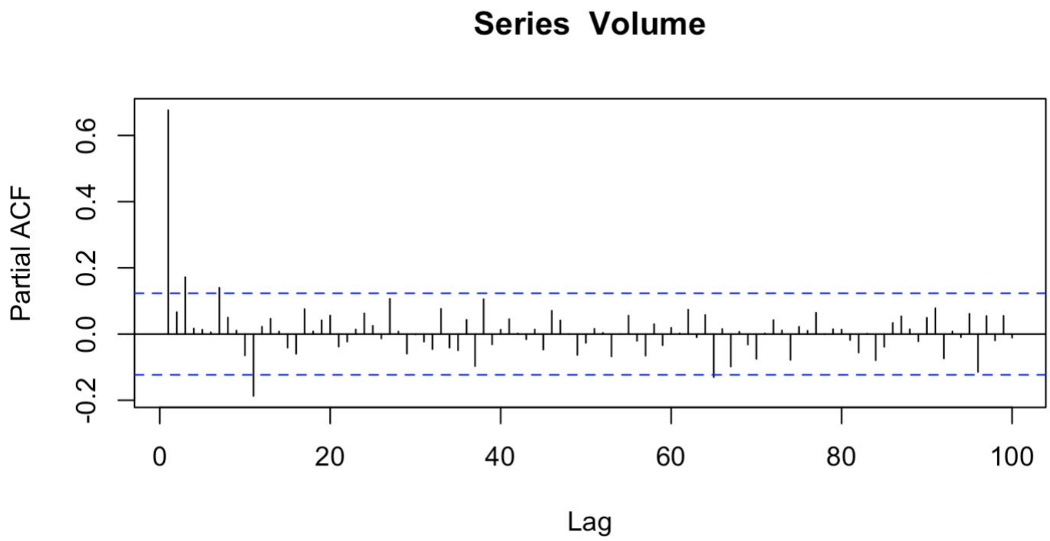
[PLACE PACF PLOT OF VOLUME TIME SERIES HERE]

PACF Plot:

Choosing lag.max =50



Choosing lag.max =100



After plotting the PACF against the lag for the daily volume, we see that the PACF is significant at lag 1, 3, 7 and 11. We may think that the PACF significance at lags 3 and 7 happen simply due to chance since they seem to be marginally significant.

More analysis:

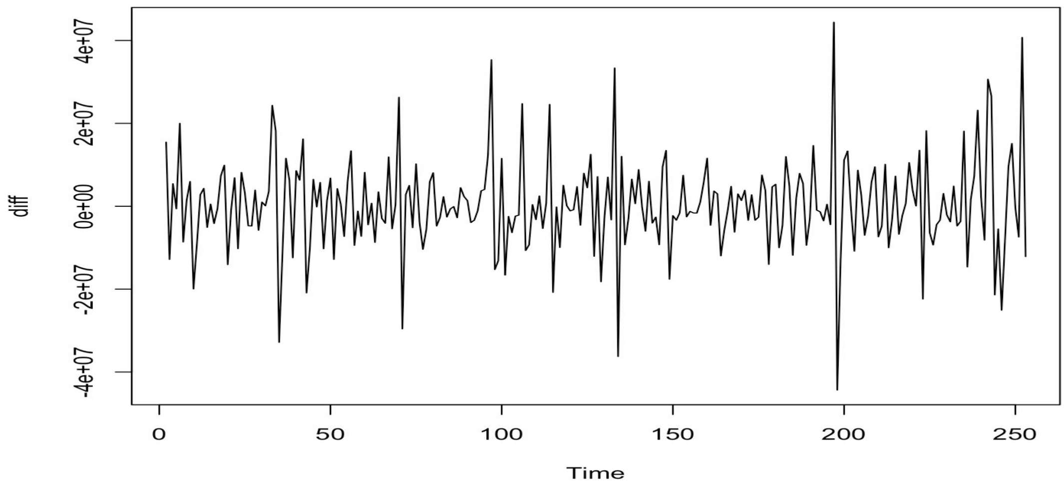
1.Use Box-Pierce test to decide whether the series appears to be a realization from a white noise process.

```
> Box.test(Volume, lag = 50, type = "Box-Pierce", fitdf = 0)

        Box-Pierce test

data:  Volume
X-squared = 416.43, df = 50, p-value < 2.2e-16
```

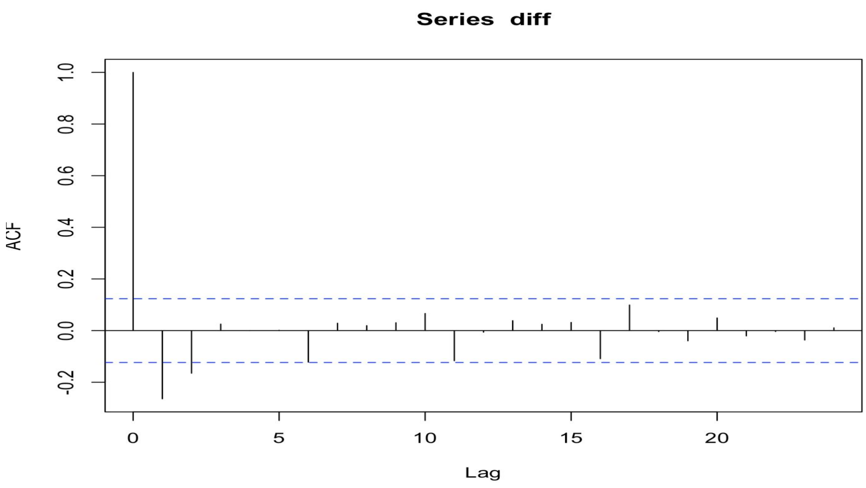
Conclusion: Since p-value is 2.2e-16 which is super small,then we may conclude the series doesn't appear to be a realization from a white noise process.

2. Apply the first difference operator to the time series that might reasonably be expected to remove the non-stationary component.

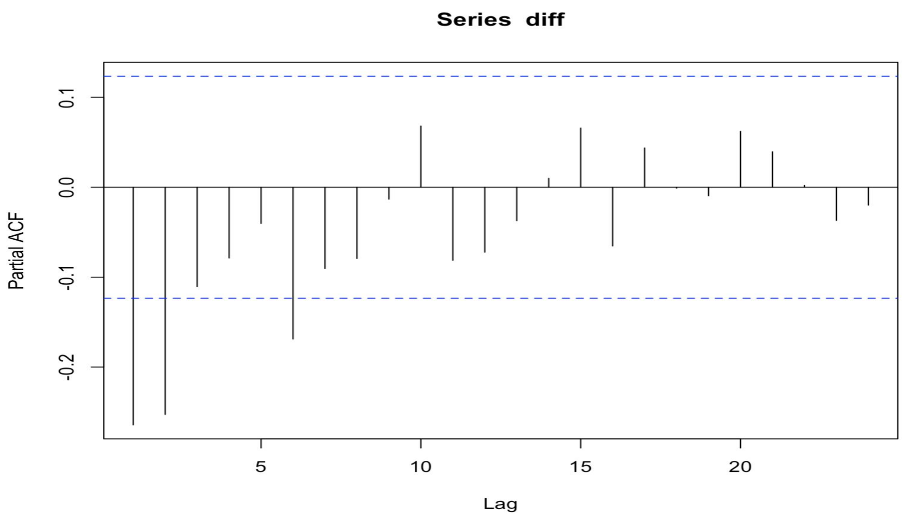
Plot :



ACF plot:



PACF plot:



After applying the first difference, we can see that there is non-seasonal and no trend.
The acf is significant at lag=1,2 comparing to at other lags, and all the other values of acf are very small.
At lag=1,2,6, the pacf has significant values.
Thus the new series is stationary and it is from a white noise process.

Verifying it using Box-Pierce Test:

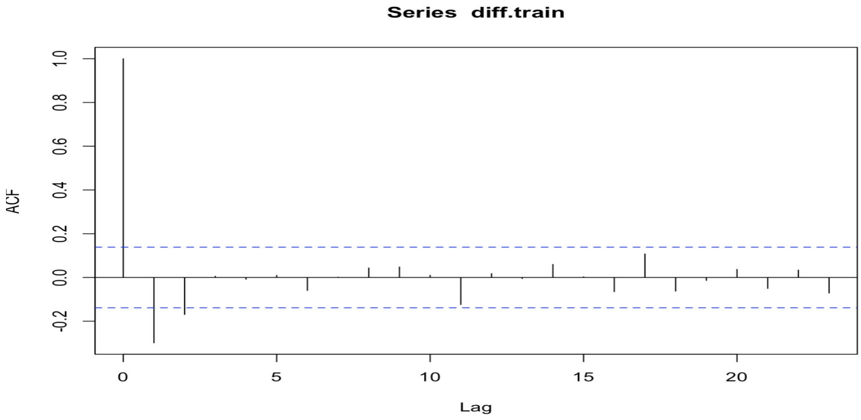
```
> Box.test(diff, lag = 50, type = "Box-Pierce")

        Box-Pierce test

data:  diff
X-squared = 65.196, df = 50, p-value =
0.07305
```

Conclusion: The p-value is 0:5833 which is greater than 0.05, thus the series appearing to be a realization from a white noise process.

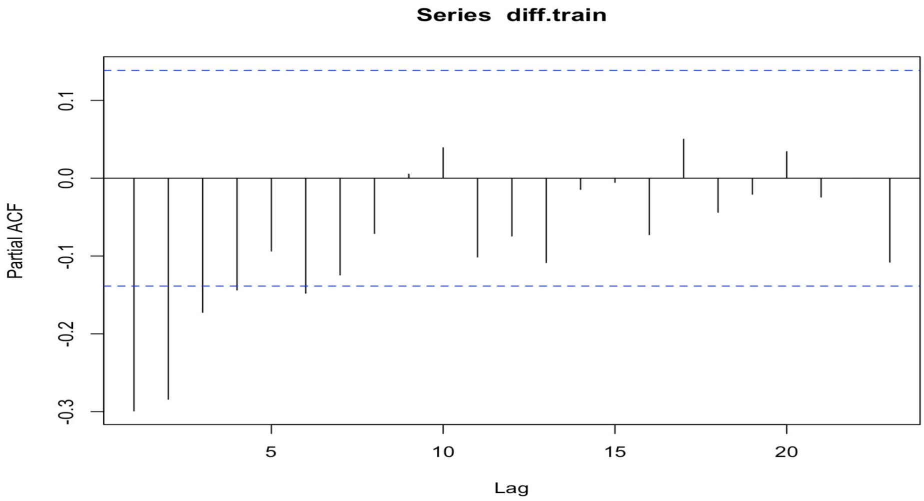
3. Try different ARMA models to find the least test error using training and test sets

```
##split data into 2 parts, training and test sets to train our model
length(diff)
#if we use the last 12 data as our test set
diff.train <- as.ts(diff[1:200])
diff.test <- as.ts(diff[241:252])
length(diff.test)
plot(diff.train)
acf(diff.train)
pacf(diff.train)
```

ACF plot of diff.train:



PACF plot of diff.train:



Therefore, we may choose the following models to continue our analysis.

AR(2) Model:

```
##use ar(2) to be our model
fit<-arima(diff.train, order = c(2,0,0))
fit
acf(fit$residuals)
tsdiag(fit)

predict(fit, n.ahead= 12)

##test error
foremodel<- predict(fit, n.ahead= 12)
foremodel
error <- sum((diff[241:252] - foremodel$pred)^2)
error
```

The error is 5.729422e+15 which is very small.

ARMA(2,3) Model:

```
##use arma(2,3) to be our model
fit<-arima(diff.train, order = c(2,0,3))
fit
acf(fit$residuals)
tsdiag(fit)

predict(fit, n.ahead= 12)

##test error
foremodel<- predict(fit, n.ahead= 12)
foremodel
error <- sum((diff[241:252] - foremodel$pred)^2)
error
```

The error is 5.632291e+15 which is also very small.

Predict the next value (or more values) using the above two models.

```r
#prediction
length(Volume)
pred01<-Volume[253]+foremodel$pred[1]
pred01
pred02<-pred01+foremodel$pred[2]
pred02
```

.

The predicted values of the next two day are 85478945, 82561363.