

# TIMESERIES ANALYSIS TAKE HOME EXAM

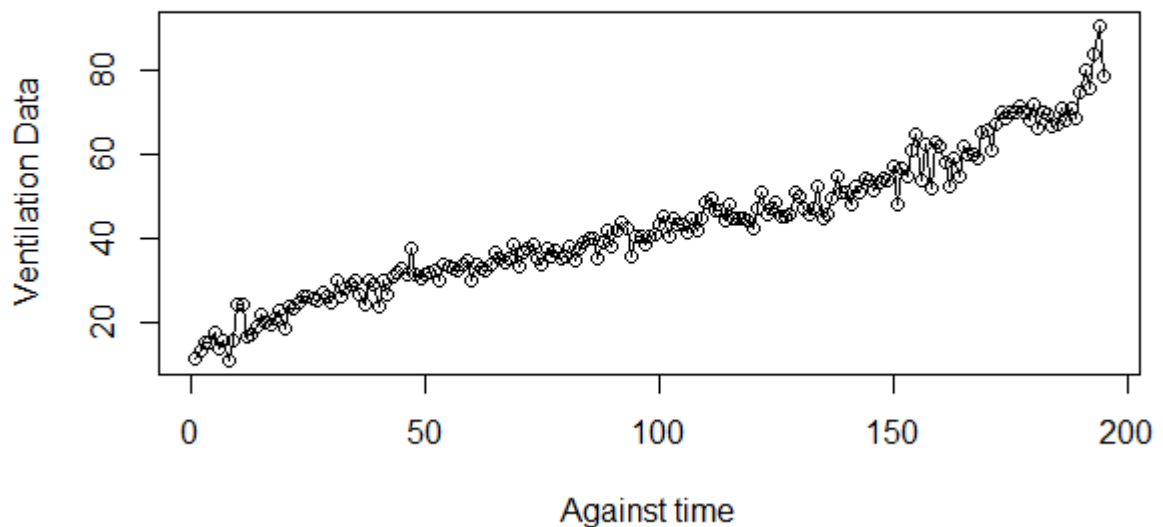
NAME – AKINDAINI BOLARINWA

STUDENT NUMBER – 422932

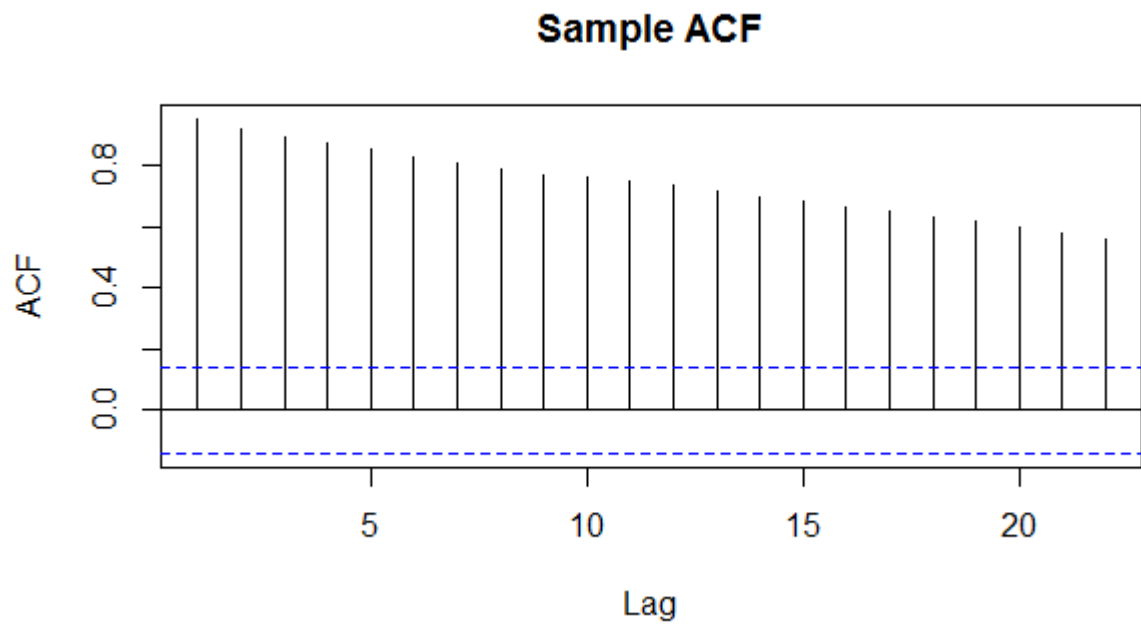
DEGREE PROGRAMME – MDP COMPUTATIONAL BIG DATA ANALYTICS

## QUESTION1:

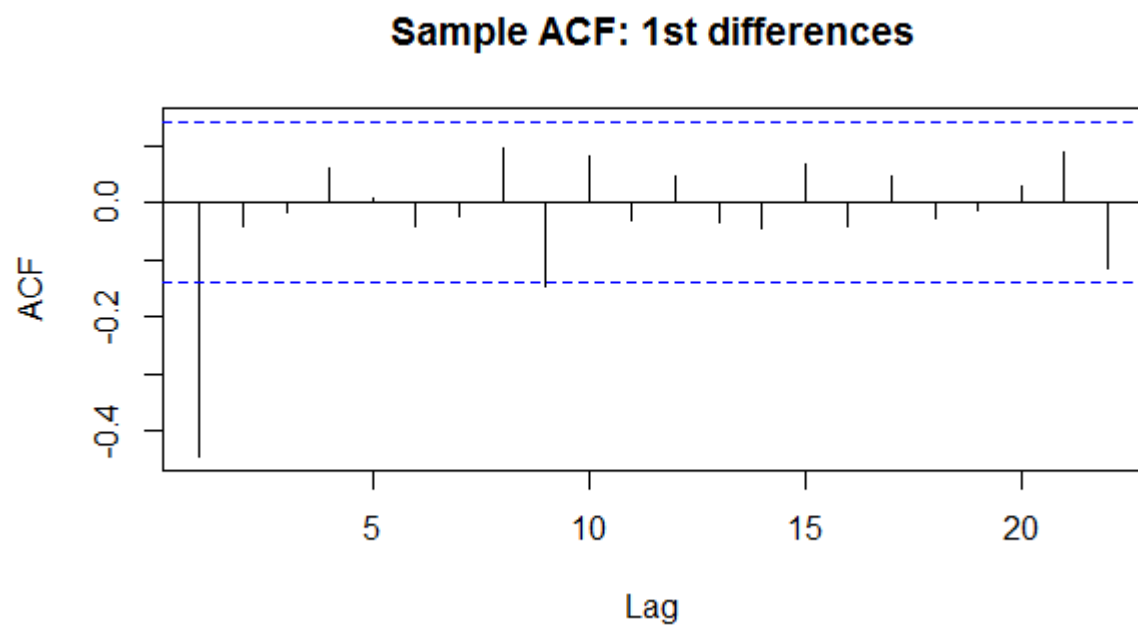
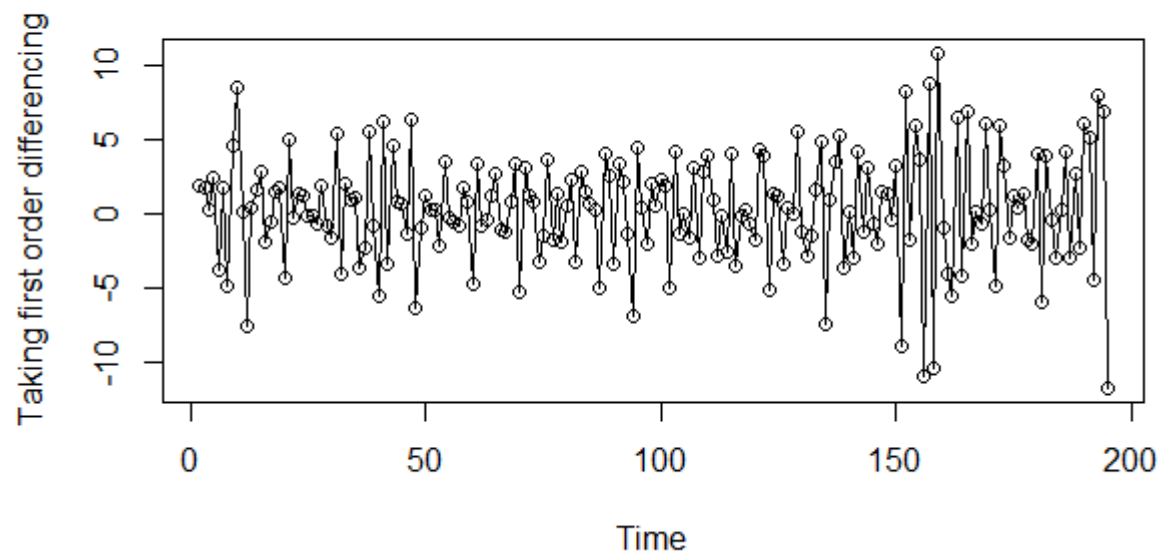
- A)  $Y_t = 15.511 + 0.279t$ ;  $t = 1, 2, \dots, 195$  The data from the linear regression graph shows a clear sign of non-stationary pattern with a linear upward/ increasing trend. Observing the autocorrelation plot further confirms this. The autocorrelation shows a sign of slow decay.



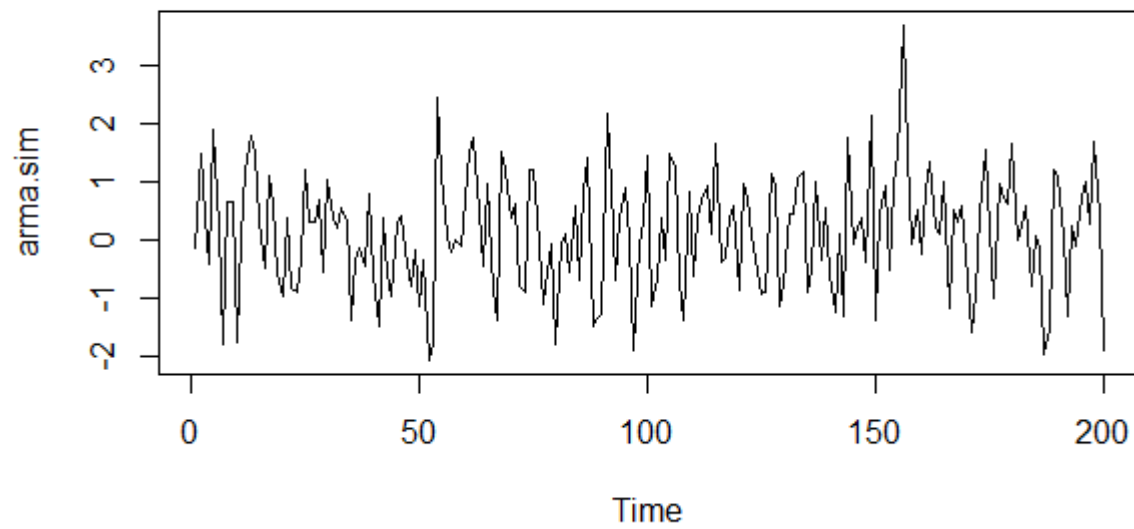
- B) From the ACF plot, i observed a substantial amount of structure left in the residuals. Positive autocorrelations are strong until time lag20. There are still noticeable structures, such as non-normality, dependency, and strong autocorrelations, left in standardized residuals. So they do not resemble a stationary white noise process.



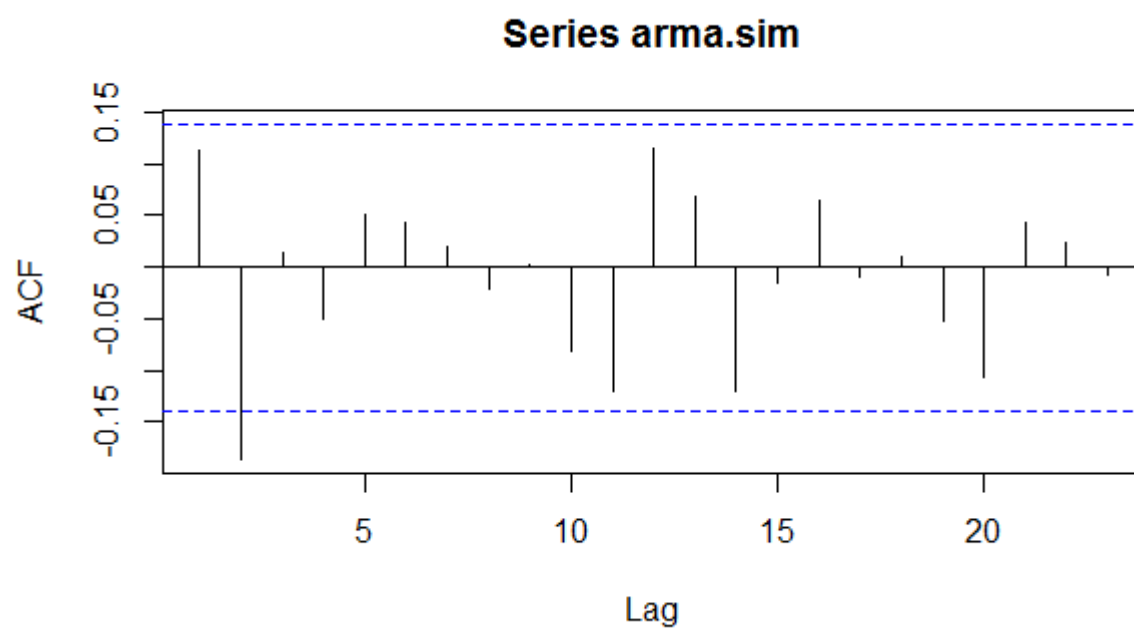
- C) Using the first order differencing we get a zero mean model with obvious look of an MA(1). And from the ACF plot of the residuals of the first differencing, it looks stationary.



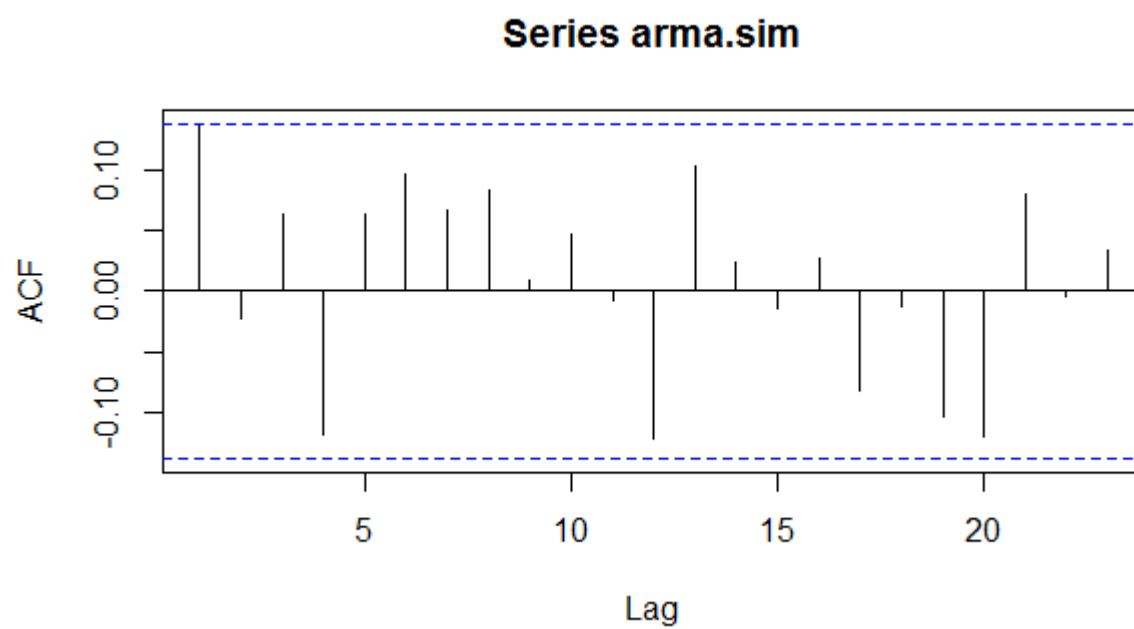
QUESTION2: The time series plot is as shown below. The data were generated using the R command. ( See R code for details)



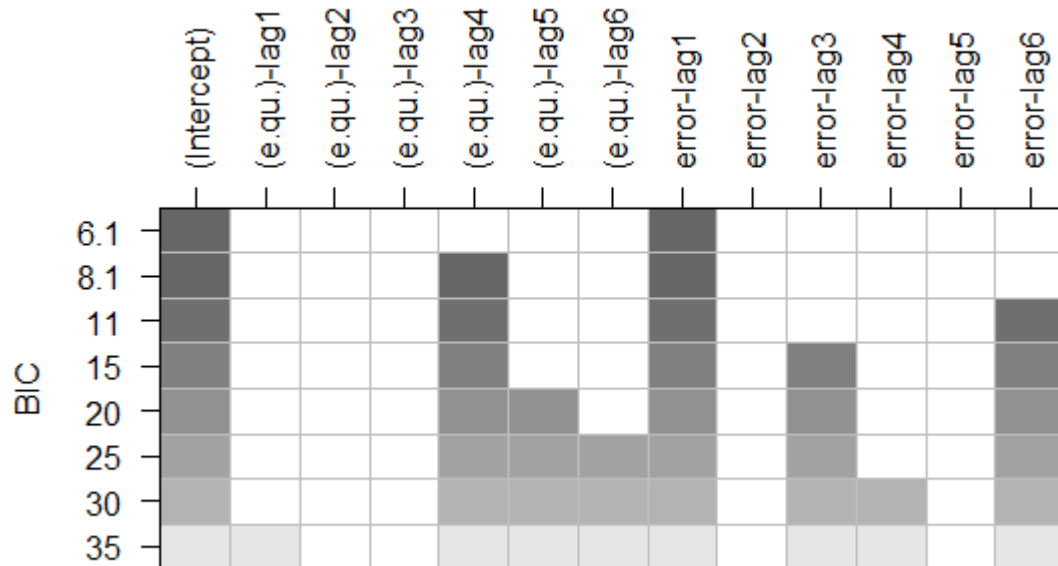
- C) AUTO-CORRELATION AND PARTIAL- AUTOCORRELATION PLOT PLOT.  
The ACF and PACF plot shows a sign of stationarity.



PACF PLOT:



c) The BIC shows MA(1)/ ARMA(0,1) which doesn't match with the known model of ARMA(1,1). Obviously, BIC fails to identify the correct model as the best model.

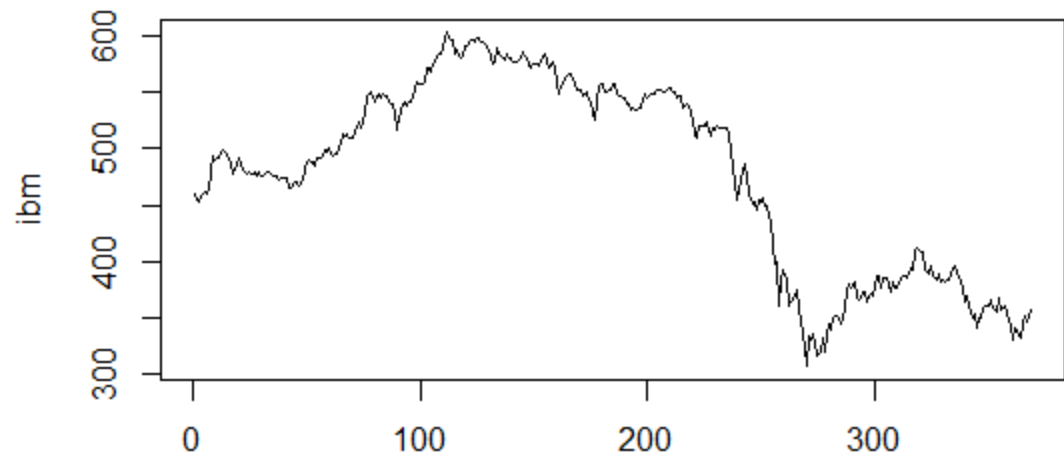


In my opinion, I think there's a sign of under fitting from the BIC. A further confirmation will be to use the AIC to investigate further.

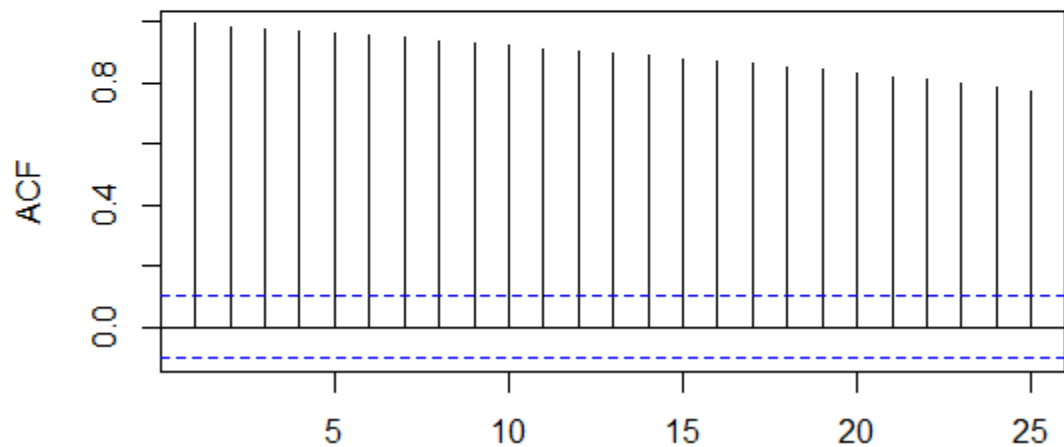
### QUESTION 3

a) Augmented Dickey-Fuller Test data: ibm

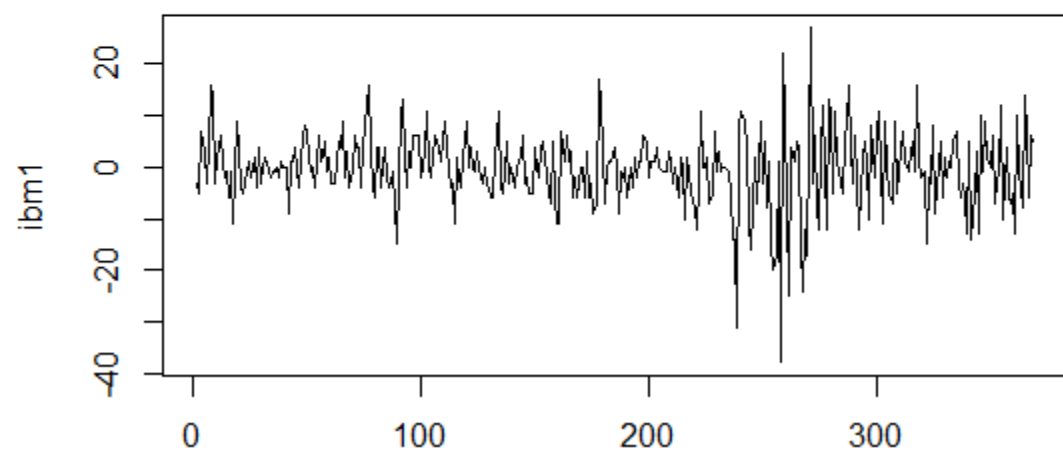
Dickey-Fuller = -1.8863, Lag order = 1, p-value = 0.625, alternative hypothesis: stationary  
The initial plot shows an upward and alternating pattern. Also changes in variation and noisy values. The initial ACF plot also shows that the data is not stationary. The p value from the Augmented dickey – Fuller test is relatively large which suggest we may not reject the null hypothesis. The factors stated above suggests that the IBM data is not stationary, there fore taking the first difference will be a good option.



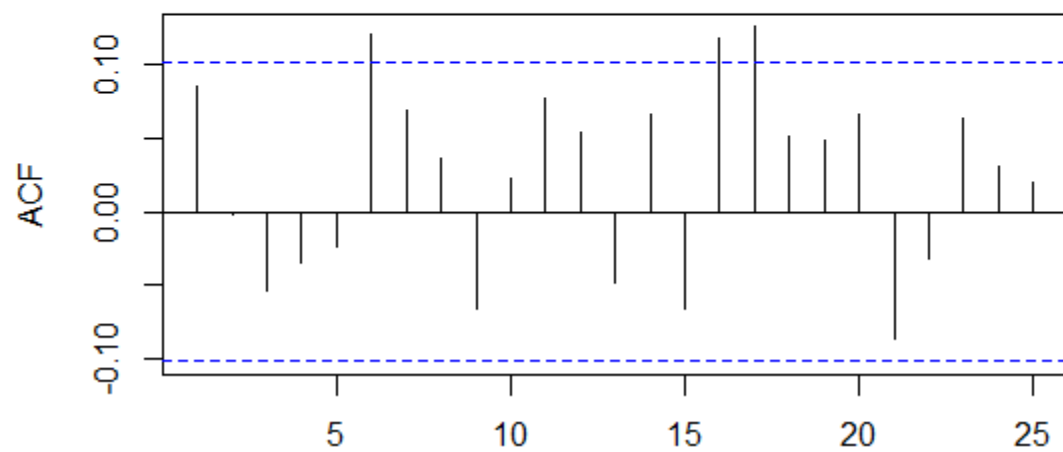
**Sample ACF**



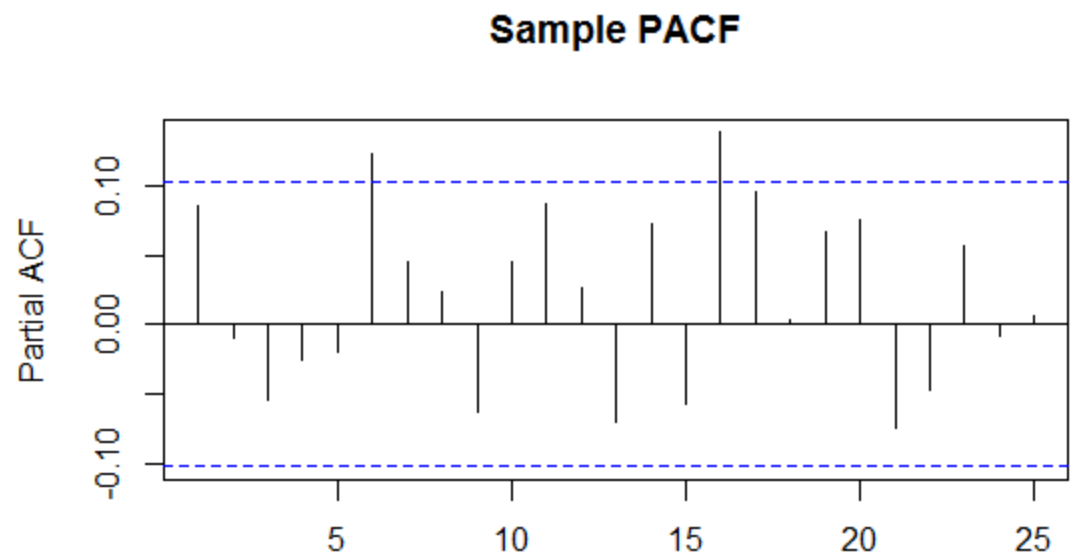
- b) B) Cyclic periodic variation pattern noticed after the first differencing. The data looks a bit stationary after the first differencing. The time series plot looks like a mean zero white noise process except there is large unusual oscillation between 200 and 300. The sample ACF and PACF plots do not suggest really large estimations of autocorrelations and partial autocorrelations at early time lags.



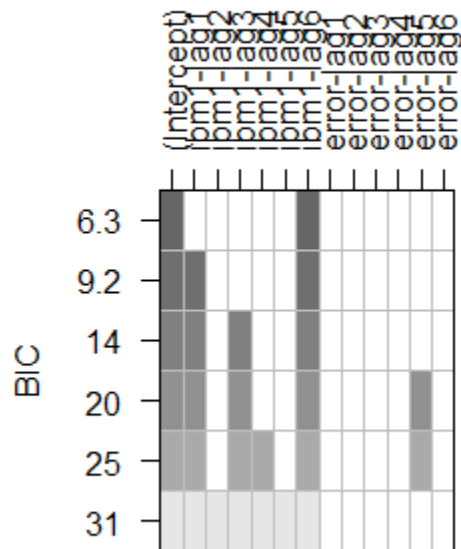
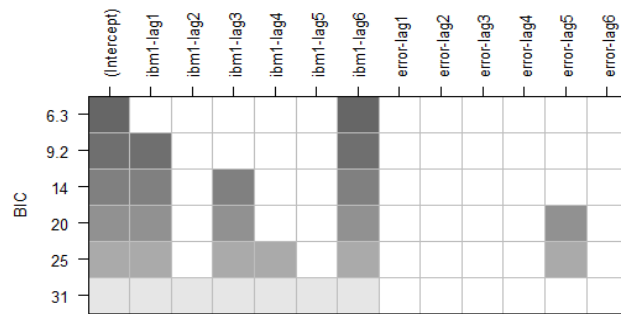
**Series ibm1**







- C) The BIC output shows that the best model (smallest BIC) contains lag 6 of the time series. The 'armasubsets' function shows AR(6) as the best in terms of the BIC. The second best model by the BIC occurs at lag 1 and 6 of AR.



D) ARIMA(6,1,0): THE "1" in the middle indicates the data was differenced once.

```
arima(x = ibm1, order = c(6, 1, 0), method = "ML")
```

```
arima(x = ibm1, order = c(6, 1, 0), method = "ML")
```

Coefficients:

	ar1	ar2	ar3	ar4	ar5	ar6
	-0.8007	-0.6781	-0.5998	-0.4982	-0.4093	-0.1662
s.e.	0.0516	0.0631	0.0673	0.0670	0.0632	0.0521

sigma^2 estimated as 56.88: log likelihood = -1262.93, aic = 2537.86

ARIMA(1,1,0)

arima(x = ibm1, order = c(1, 1, 0), method = "ML")

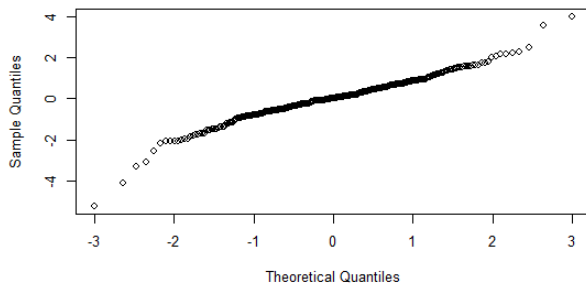
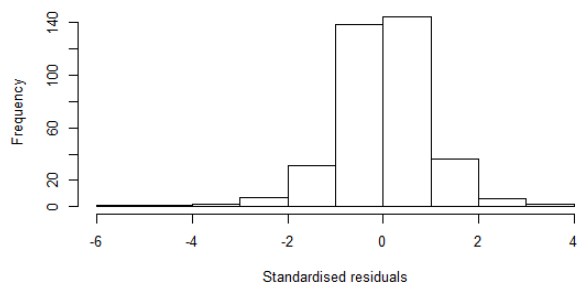
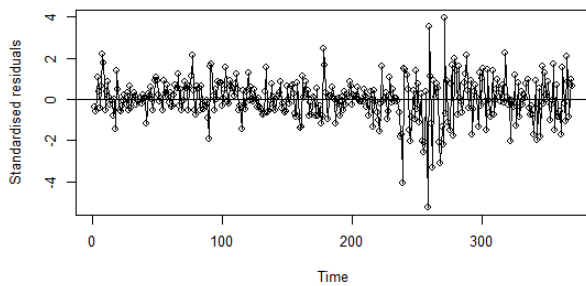
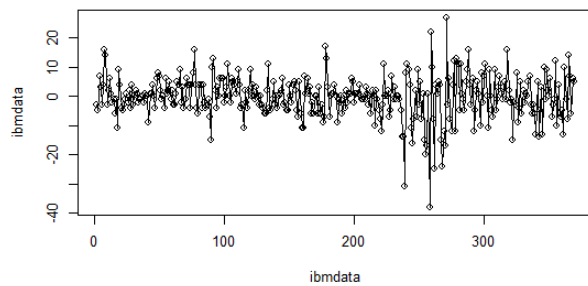
Coefficients:

	ar1
	-0.4520
s.e.	0.0465

sigma^2 estimated as 76.48: log likelihood = -1316.71, aic = 2635.41

The AIC also confirms the result of the BIC that AR(6) is a better model.

Diagnostic of ARIMA(6,1,0)



From the above diagnostic we see some signs that the first difference data is white noise with mean zero (Although not so convincing). Although the autocorrelation seems to look good from the graph, of great concern is the fact I observed the presence of large significance value at long lags. Also, I observed that all of the modied Ljung-Box thest  $p$  value are mostly larger than 0.05 at early  $K$ . However, they are close to  $\alpha = 0.05$  and turn to significant at the end. This further shows that the model selected by the BIC and AIC may need before investigations before it can be judged as the best model to fit the data (My opinion).

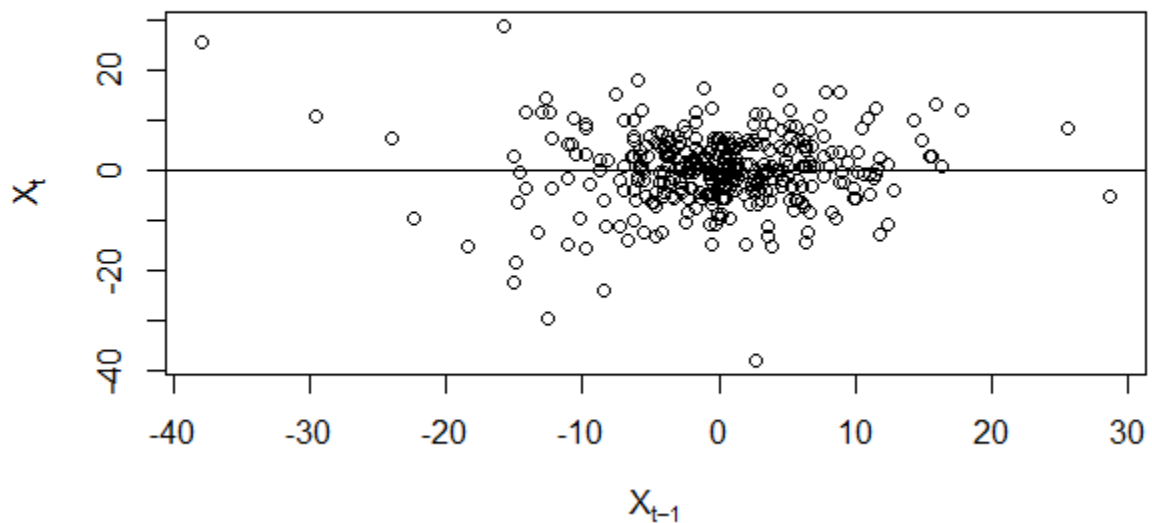
#### Box-Ljung test

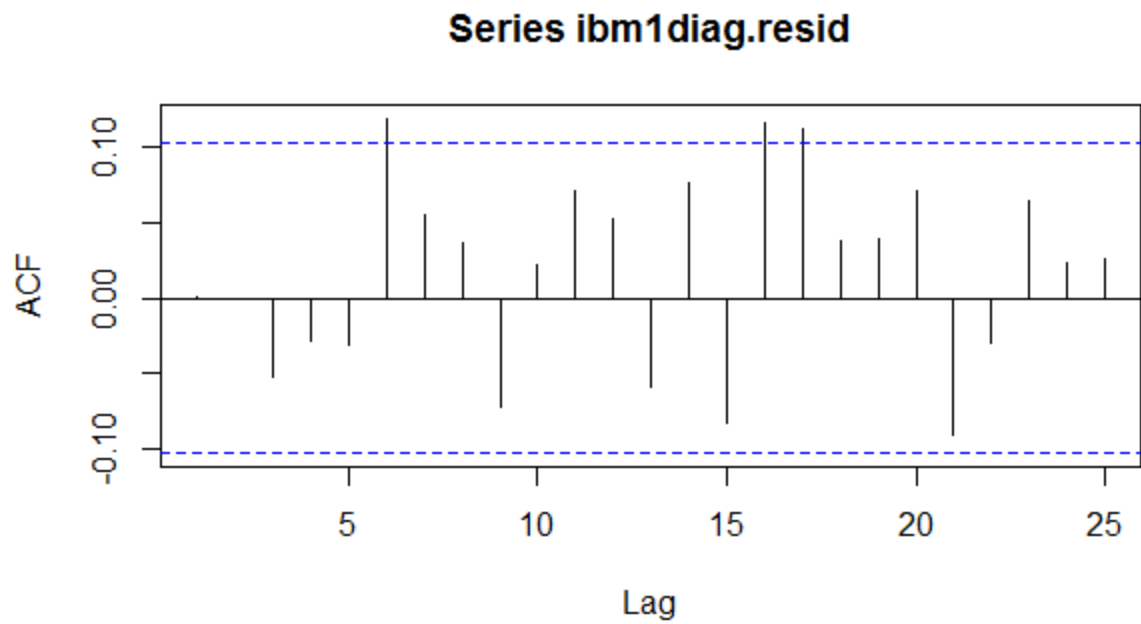
```
data: residuals(ibm1diag)
X-squared = 15.803, df = 4, p-value = 0.003295
```

#### Box-Pierce test

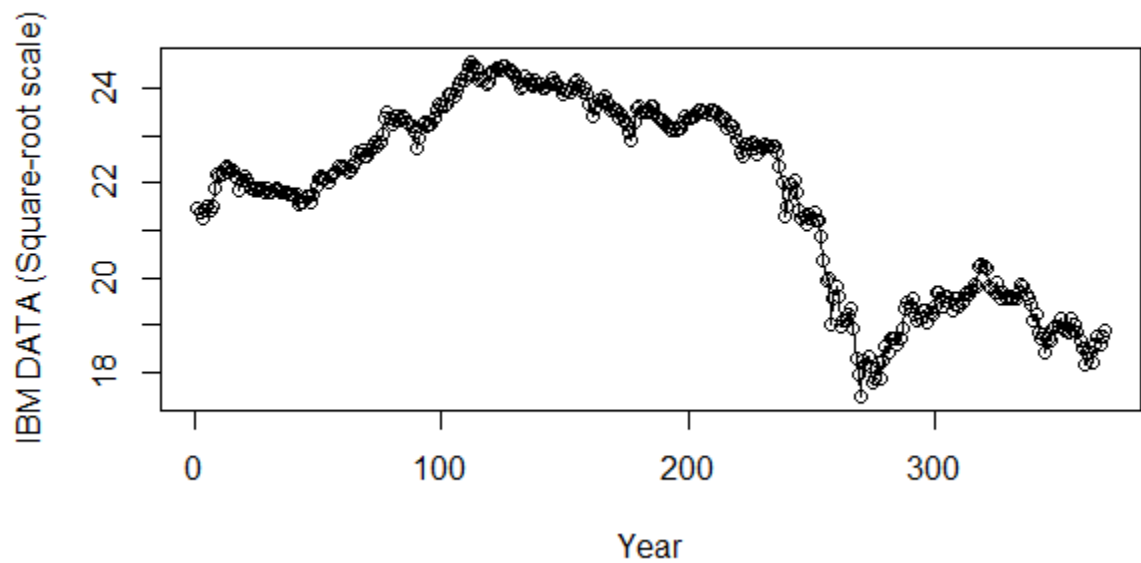
```
data: residuals(ibm1diag)
X-squared = 4.2752, df = 4, p-value = 0.37
```

#### Lag 6 scatterplot

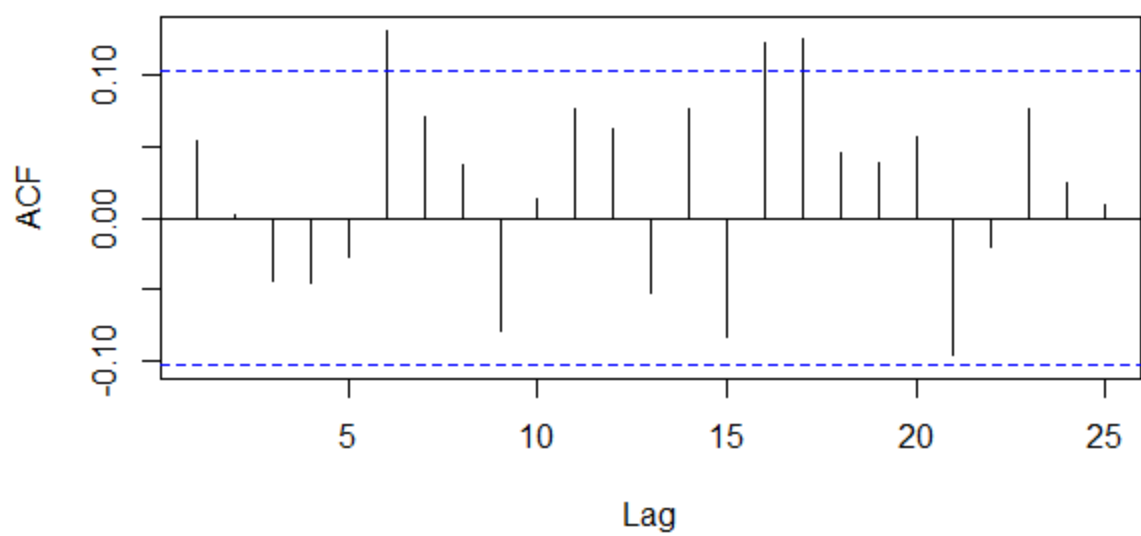




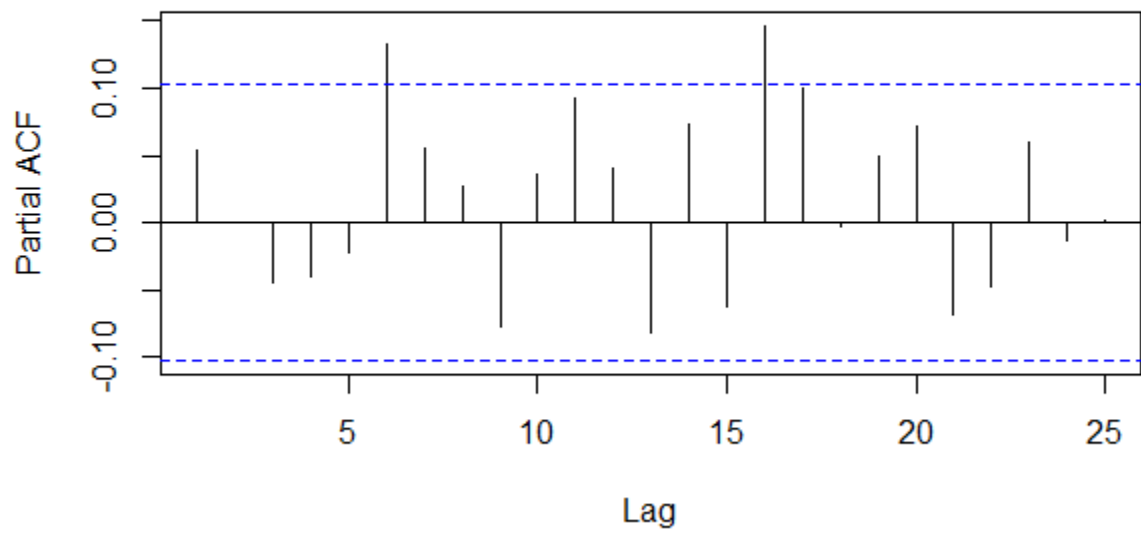
- E) From the transformed data,  $mle = 2 > 0$ , so we choose  $\lambda$  to be 0.5 and use squareroot transformation. The AIC and BIC both still support ARIMA(6,1,0). Again the AcF and PACF suggests white- noise.

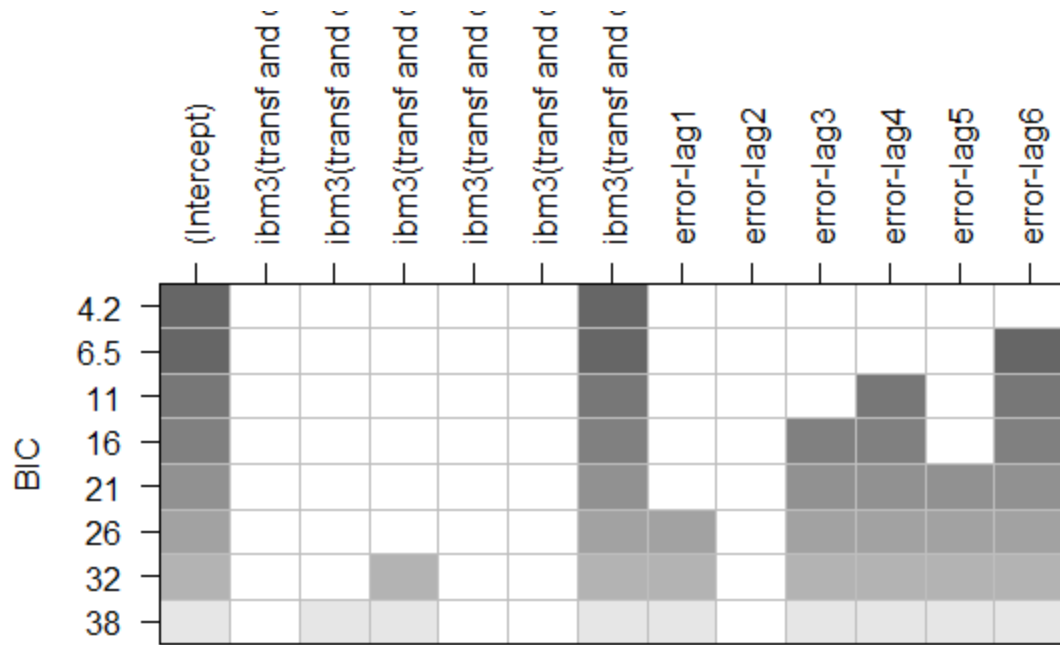


**Series ibm3**

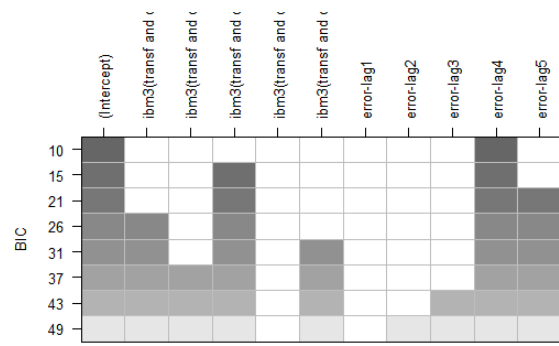
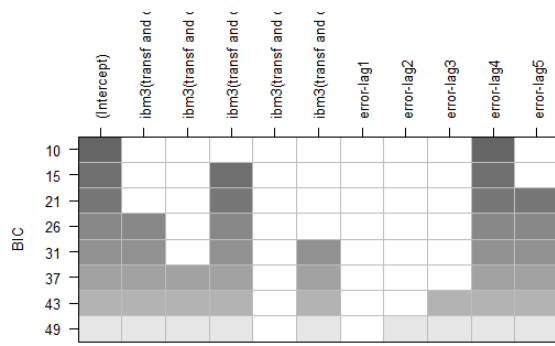


**Sample PACF**



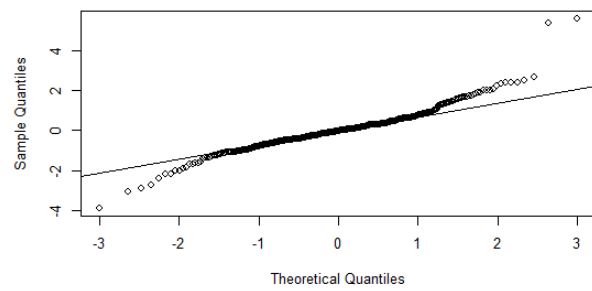
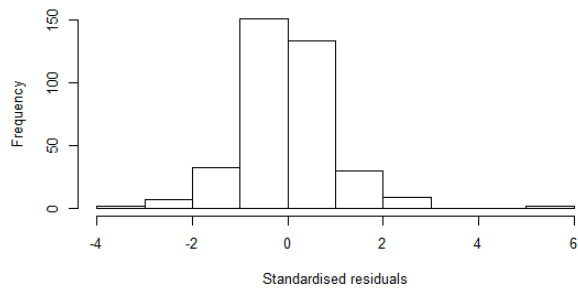
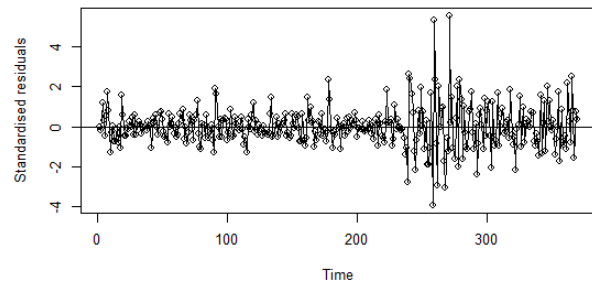
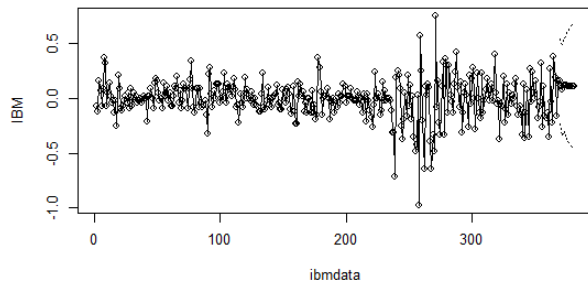


I tried using 'armasubsets' with 5 by 5 for the parsimony. The BIC with the 5 by 5 gave AR(4)/ARIMA(4,1,0) and ARMA(3,4)/ARIMA(3,1,4). The candidate with the lower AIC is AR(4)



Diagnostic graphs of AR(4)/ARIMA(4,1,0)





F)Based on the above, either of AR(6) or AR(4) may be a likely candidate to fit the data properly.