Part I
Question 1

A time series is given by the formula:

$$Xt = -0.022X_{t-1} - 0.405X_{t-2} + 0.493X_{t-3} + 0.550\varepsilon_{t-1} + 0.276\varepsilon_{t-2} + 0{:}490\varepsilon_{t-3} + 0.682\varepsilon_{t-4} + 0.222\varepsilon_{t-5} - 0.232\varepsilon_{t-6} + \varepsilon_t$$

where for all t, the innovations ϵ_t are distributed according to the standard normal distribution, that is, $\varepsilon_t \sim N(0,1)$.

This time series is a $ARMA(3,6)$ model with $\phi_1 = -0.022, \phi_2 = -0.405, \phi_3 = -0.493$, $\psi_1 = 0.550$, $\psi_2 = 0.276, \psi_3 = 0.490$, $\psi_4 = 0.682$, $\psi_5 = 0.222$ *and* $\psi_6 = -0.232$.

By using R command “arima.sim” (see appendix, section Part 1 Q1), we generate a realization from this time series of length 1000 and plot the realization. Then we used R command “acf” and “pacf” to find its autocorrelation and partial autocorrelation functions.

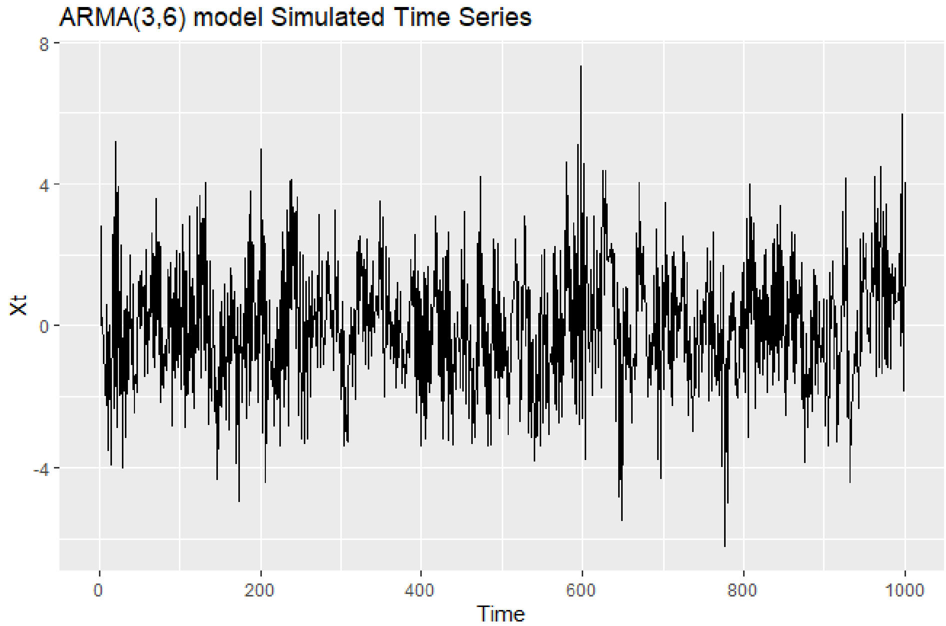

Figure 1.1: Plot of a simulated time series following the ARMA (3,6) model with length 1000

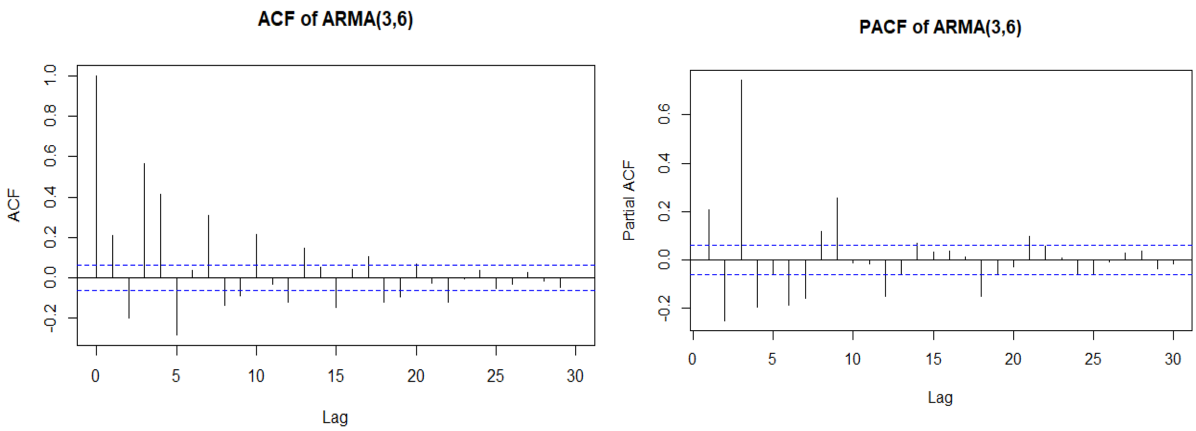

Figure 1.2: Plot of autocorrelation function (ACF) and partial autocorrelation function (PACF) of the $ARMA(3,6)$ model

Question 2

We will check the causal stationarity and invertibility of the time series given by $ARMA(3{,}6)$ model from Question 1. Recall that, the time series is a $ARMA(3{,}6)$ model with $\phi_1 = -0.022, \phi_2 = -0.405, \phi_3 = -0.493, \psi_1 = 0.550,\ \psi_2 = 0.276, \psi_3 = 0.490,\ \psi_4 = 0.682,\ \psi_5 = 0.222\ and\ \psi_6 = -0.232$.

A $ARMA(p,q)$ model is causal (stationary) and invertible if and only if all roots of pth order polynomial $\phi(z)$ and qth order polinomial $\psi(z)$ lie outside the unit circle, i.e., $|z_k| > 1, \forall\ 1 \le k \le p, q$.

Setting $\phi(z) = 0$, we obtain the characteristic equation of an $ARMA(3{,}0)$ model:

$$\phi(z) = 1 - (-0.022)z - (-0.405)z^2 - (-0.493)z^3 = 0$$

By using R command "polyroot(a)" and "Mod(z)" (see appendix, section Part 1 Q2) to find the roots of the equation and calculate the modulus of a complex number. From the R output, we get 3 roots for $\phi(z)$ which are

$$\boldsymbol{z_1 = 1.118630, z_2 = 1.118630 \text{ and } z_3 = 1.620988.}$$

Setting $\psi(z) = 0$, we obtain the characteristic equation of an $ARMA(0{,}6)$ model:

$$\psi(z) = 1 + 0.550_1 z + 0.276z^1 + 0.490z^3 + 0.682z^4 + 0.222z^5 + (-0.232)z^6 = 0$$

By using same R command as above, from the R outputs, we get 6 roots for $\psi(z)$ which are

$$\boldsymbol{z_1 = 1.027907, z_2 = 1.139698, z_3 = 1.139698, z_4 = 1.027907,}$$

$$\boldsymbol{z_5 = 1.218771, z_6 = 2.576926.}$$

From above we can see that, all the roots of $\phi(z)$ and $\psi(z)$ lie outside the unit circle. Hence, **the time series given by $ARMA(3,6)$ model is causal stationary and invertible**.

Question 3a

The formula of the $MA(3)$ (2.1 Moving Average Models (MA models)) model with normal innovations,

$$\boldsymbol{X_t = \mu + \psi_1 \varepsilon_{t-1} + \psi_2 \varepsilon_{t-2} + \psi_3 \varepsilon_{t-3} + \varepsilon_t}$$

where for all t, the innovations ϵ_t are distributed according to the standard normal distribution, that is, $\varepsilon_t \sim N(0, \sigma_\epsilon^2)$ for some $\sigma_\epsilon^2 > 0$.
The R command "arima" (see appendix, section Part 1 Q3) was used to fit the $MA(3)$ model to the observed time series, then find the estimated coefficient and estimated innovation variance of this $MA(3)$ model.
From the R output**, the estimated coefficient of this MA (3) model is $\boldsymbol{\psi_1 = -0.08882102, \psi_2 = 0.81335449, \psi_3 = 0.17357350}$.** The **intercept term** in the R output is the **estimated mean of the model, $\boldsymbol{\mu = -0.11715686}$.** (Some Time Series Issue)

The estimated innovation variance, $\hat{\sigma}_\epsilon^2$ = 37953.47

Question 3b

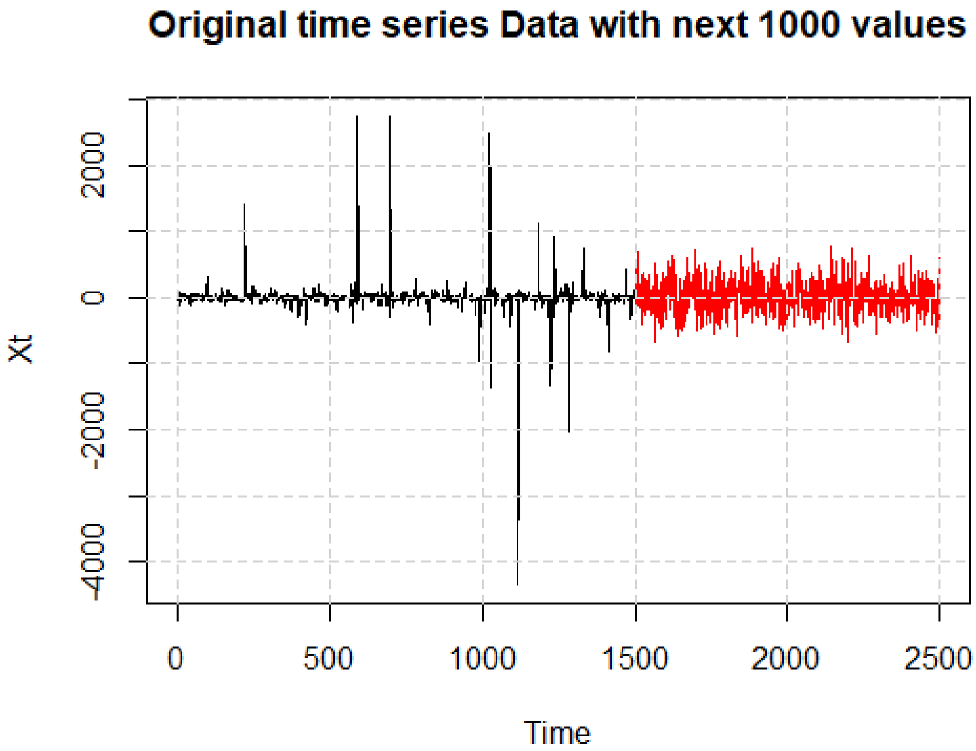

Figure 1.3: Plot of original data with simulation of next 1000 values using the model fitted in 3(a)

In figure 1.3, despite the extreme values in the original time series, the original time series only fluctuates a little around 0. On the other hand, the simulated time series with the $MA(3)$ model have larger fluctuations around 0. The reason is that the extreme values in original time series made the estimated innovation variance, $\hat{\sigma}_\epsilon^2$ becomes large. Hence, the simulated time series fluctuates larger than the original time series around 0.

To investigate further, we compare the autocorrelation function (ACF) for both time series.

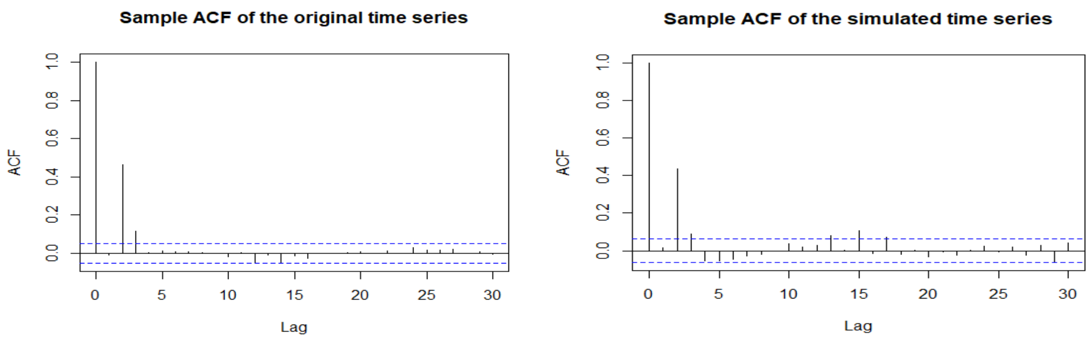

Figure 1.4: Plots of Sample ACF of the original time series and the simulated time series

In figure 1.4, both of the plots of sample ACF are very similar to each other.

Hence, we can conclude that the $MA(3)$ model (from Question 3a) is adequate for this time series.

Part II

Question 1

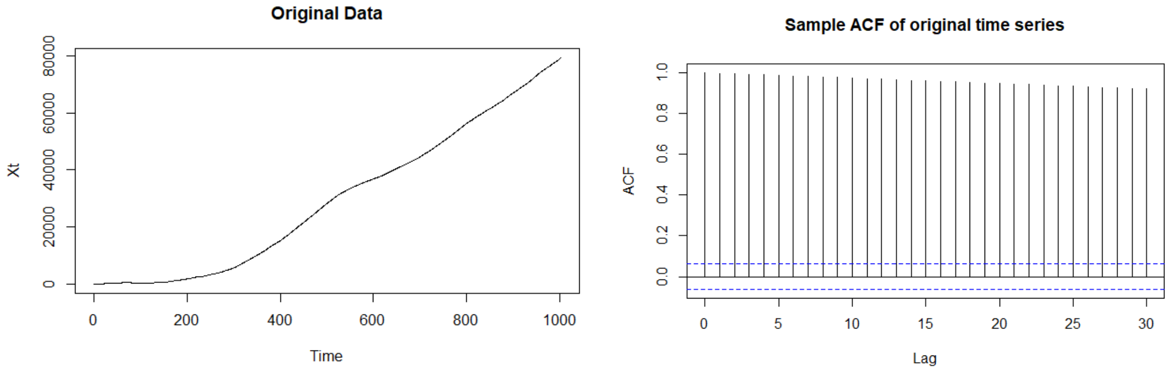

Figure 2.1: Plot of original time series and its sample autocorrelation function (ACF)

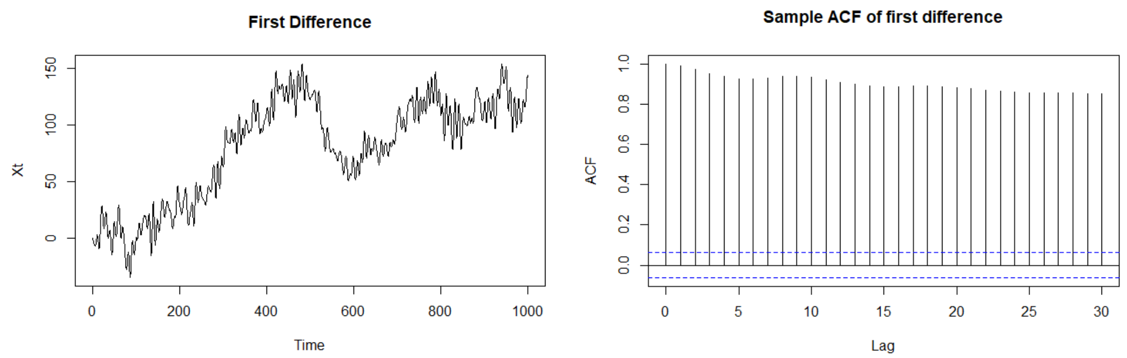

Figure 2.2: Plot of first difference time series and its sample autocorrelation function (ACF)

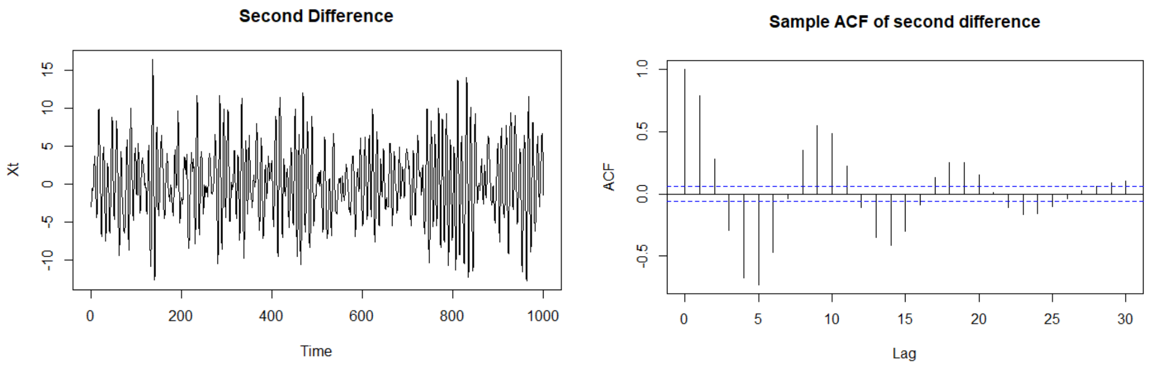

Figure 2.3: Plot of second difference time series and its sample autocorrelation function (ACF)

In figure 2.1, the time series plot of the original data clearly shows a trend and the corresponding sample ACF tails off relatively slowly. Hence, the original time series is not stationary. Then in figure 2.2, the time series plot of first differenced data has a little trend and the corresponding sample ACF still tails off slowly, indicating that these first differences are not stationary either. However, when we difference again, in figure 2.3, the second difference of time series show no obvious trend in the time series plot, and the ACF tails off sinusoidally and no long decays slowly. We can conclude that these second differences are stationary.

Since we had to difference twice to reach stationary data, we should select **the value of d equals to 2**.

Question 2

Now we want to find the values of p and q by comparing Akaike information criterion (AIC) and Bayesian information criterion (BIC) for $p = 0,1,2,3 \; and \; q = 0,1,2,3$ with $ARIMA(p, 2, q)$.

Mohammed, Naugler and Far (2015) stated that Akaike information criterion (AIC) is a fined technique based on in-sample fit to estimate the likelihood of a model to predict the future values. On the other hand, Bayesian information criterion (BIC) (Stone, 1979) is criteria for model selection that measures the trade-off between model fit and complexity of the model. A lower AIC or BIC value indicates a better fit.

The AIC formula is

$$AIC(p, q) = ln\hat{\sigma}_{p,q} + \frac{2(p+q)}{n}$$

And the BIC formula is

$$BIC(p, q) = ln\hat{\sigma}_{p,q} + \frac{\ln(n)\,(p+q)}{n}$$

Where $\hat{\sigma}_{p,q}$ is an estimator of innovation variance, σ_ϵ^2 and n is the number of data.

After implement these formulas in R (see appendix, section Part2 Q2), we get the values of AIC and BIC for q=0,1,2,3

	$q = 0$	$q = 1$	$q = 2$	$q = 3$
$p = 0$	3.15136589	1.97073205	1.13648211	0.70266561
$p = 1$	2.17060164	1.19494024	0.71848923	0.52284402
$p = 2$	0.19335378	0.04323358	0.04519213	0.04571831
$p = 3$	0.06253018	0.04518311	0.04661401	0.04730467

Table 2.1: Values of BIC of $ARMA(p, 2, q)$ where $p = 0,1,2,3 \; and \; q = 0, 1, 2,3$

After implement this formula in R, we get the values of AIC and BIC for q=0,1,2,3

	$q = 0$	$q = 1$	$q = 2$	$q = 3$
$p = 0$	3.15136589	1.97563981	1.14629762	0.71738888
$p = 1$	2.17550939	1.20475575	0.73321249	0.54247504
$p = 2$	0.20316929	0.05795685	0.06482315	0.07025708
$p = 3$	0.07725345	0.06481414	0.07115279	0.07675120

Table 2.2: Values of BIC of $ARMA(p, 2, q)$ where $p = 0,1,2,3 \; and \; q = 0, 1, 2,3$

In table 2.1 and table 2.2, we can see that the time series given by $\boldsymbol{ARIMA(2, 2, 1)}$ model has the lowest AIC and BIC (highlight in yellow) among all the others.

Hence, we will select $\boldsymbol{p = 2}$ and $\boldsymbol{q = 1}$ for $ARIMA(p, d, q)$ model.

Question 3

Using R command "arima" (see appendix, section Part 2 Q3) to fit the $\boldsymbol{ARIMA(2,2,1)}$ model to the data. then find the estimated coefficient and estimated innovation variance.
From the R output, we get the values of estimates of the coefficients**,**

$$\boldsymbol{\phi_1 = 1.475, \phi_2 = -0.888, \psi_1 = 0.429}$$

The estimated innovation variance, $\boldsymbol{\hat{\sigma}^2_\epsilon}$ **= 1.0380**

The time series is given by the formula

$$\boldsymbol{X_t = 1.475X_{t-1} - 0.888X_{t-2} + 0.429\varepsilon_{t-1} + \varepsilon_t}$$

where for all t, the innovations ϵ_t are distributed according to the standard normal distribution, that is, $\varepsilon_t \sim N(0, \sigma^2_\epsilon)$ for some $\sigma^2_\epsilon > 0$.

Question 4

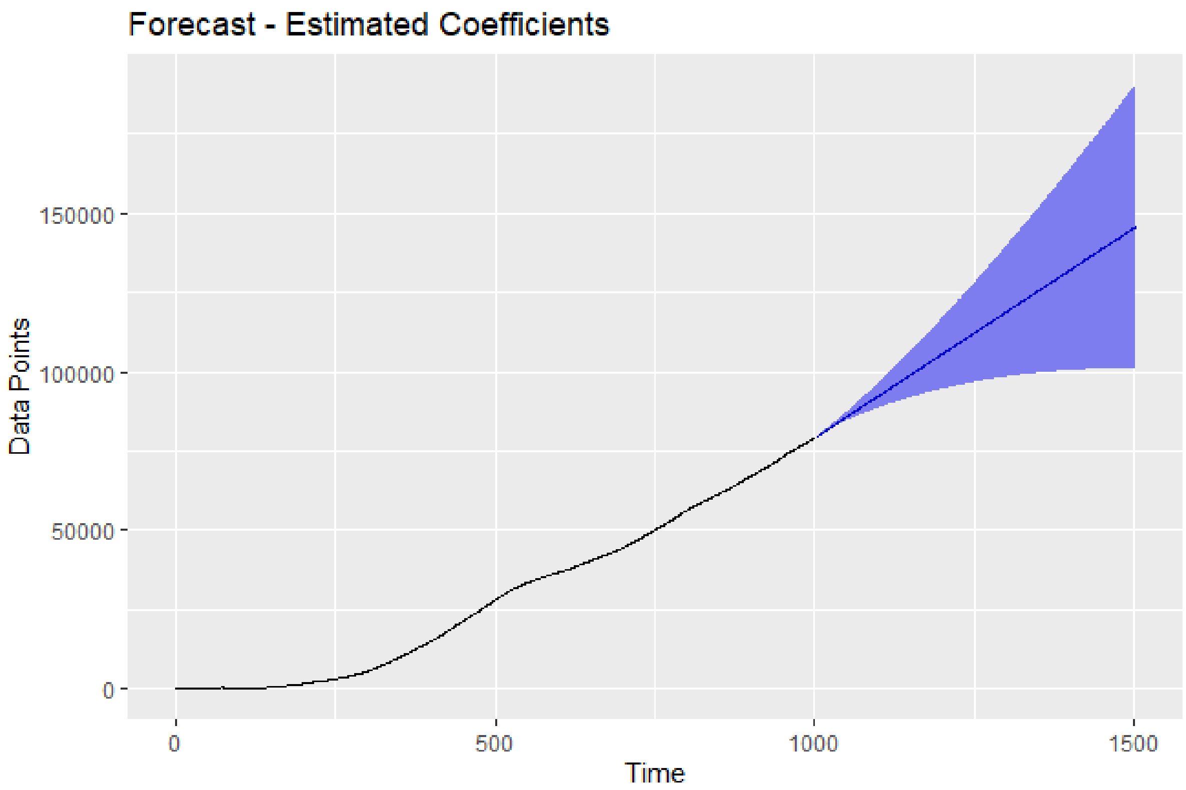

Figure 2.4: A realisation of the original time series and the corresponding forecast intervals for next 500 values

Using R command "forecast" (see appendix, section Part 2 Q4) to find the 95% forecast interval for the next 500 values. From the R output, in Figure 2.4, the forecast time series clearly shows an increasing trend with the 95%-forecasting intervals (shown in blue area). The forecast intervals are getting wider While the future time series values are getting further from the original time series data, the past observations will contain less information about future observation. Hence, the uncertainty increases and the 95% forecast intervals become wider and wider.

References

Mohammed, E., Naugler, C. and Far, B., 2015. Emerging Business Intelligence Framework for a Clinical Laboratory Through Big Data Analytics. *Emerging Trends in Computational Biology, Bioinformatics, and Systems Biology*, pp.577-602.

Some Time Series Issue, [Online]. Available at: https://www.stat.pitt.edu/stoffer/tsa4/Rissues.htm [Accessed 24 March 2021]

2.1 Moving Average Models (MA models) [Online]. Available at: https://online.stat.psu.edu/stat510/lesson/2/2.1 [Accessed 24 March 2021]

Appendix

```
###Part 1###

setwd("C:/Users/USER/Desktop/Study/Y3/Semester_2/Time Series/Assignment")

###Part 1 Q1###

library(ggplot2)

library(forecast)

n=1000

#Simulate the time series data with the given ARMA(3,6) model

model <- list(order=c(3,0,6),ar=c(-0.022,-0.405,0.493),ma=c(0.550,0.276,0.490,0.682,0.222,-0.232))

generated_data <- arima.sim(model,1000)

#Plot the time series given by ARMA(3,6) model and its ACF and PACF

autoplot(generated_data)+ ggtitle("ARMA(3,6) model Simulated Time Series")+

  xlab('Time')+ ylab('Xt')

model_acf <-  acf(generated_data, lag.max = 30, plot = FALSE)

model_pacf <- pacf(generated_data, lag.max = 30, plot = FALSE)

plot(model_acf,main="ACF of ARMA(3,6)")

plot(model_pacf,main="PACF of ARMA(3,6)")


###Part 1 Q2###

#Solve the characteristic equation of the ARMA(3,6) model

AR_root <- polyroot(c(1,0.022,0.405,-0.493))

MA_root <- polyroot(c(1,0.550,0.276,0.490,0.682,0.222,-0.232))

#to calculate the modulus of a complex number

Mod(AR_root)

Mod(MA_root)
```

```
###Part 1 Q3a###

library(gridExtra)

Data1 <- read.table("data1.txt",header = TRUE) #read data1

estimated_model <- arima(Data1, order=c(0,0,3)) #find the estimated MA(3) model for data1

coef_1 <- as.array(estimated_model$coef) #estimated coefficients of the MA(3) model

sigma_sq_epq_1 <-  estimated_model$sigma2 #estimated innovation variance of the MA(3) model


###Part 1 Q3b###


simulate_n <- 1000 #the number of simulation

simulation<-simulate(estimated_model,nsim=simulate_n)  #simulate  time  series  with  the  estimated model

#plot the original data with the realization

plot(Data1[[1]],type='l',xlim=c(0,2500),xlab="Time",ylab="Xt",main="Original time series Data with next 1000 values")

lines(simulation,col="red")

grid(lty=2)


Data1_acf <-  acf(Data1, lag.max = 30, plot = FALSE)

simulation_acf <- acf(simulation, lag.max = 30, plot = FALSE)

plot(Data1_acf, main="Sample ACF of the original time series")

plot(simulation_acf,main="Sample ACF of the simulated time series")

###Part 2###

###Part 2 Q1###

Data2 <- read.table("data2.txt",header = TRUE) #read data2

Original <- Data2[[1]] #original time series

FirstDiff <- diff(Data2[[1]]) #first differenced time series

SecondDiff <- diff(diff(Data2[[1]])) #second differenced time series

plot(Original,type='l',ylab="Xt",xlab="Time",main="Original Data")

plot(FirstDiff,type='l',ylab="Xt",xlab="Time",main="First Difference")

plot(SecondDiff,type='l',ylab="Xt",xlab="Time",main="Second Difference")

Ori_sample_acf <- acf(Original,plot = FALSE) #sample ACF of original time series
```

```
}

FD_sample_acf <- acf(FirstDiff,plot = FALSE) #sample ACF of first differenced time series
SD_sample_acf <- acf(SecondDiff,plot = FALSE) #sample ACF of second differenced time series
plot(Ori_sample_acf, main="Sample ACF of original time series")
plot(FD_sample_acf, main="Sample ACF of first difference")
plot(SD_sample_acf, main="Sample ACF of second difference")

###Part 2 Q2###
AICmodelP2 <- matrix(nrow=4,ncol=4) #Create empty matrix for AIC values
BICmodelP2 <- matrix(nrow=4,ncol=4) #Create empty matrix for BIC values
rownames(AICmodelP2) <- c("p=0","p=1","p=2","p=3")
colnames(AICmodelP2) <- c("q=0","q=1","q=2","q=3")
rownames(BICmodelP2) <- c("p=0","p=1","p=2","p=3")
colnames(BICmodelP2) <- c("q=0","q=1","q=2","q=3")

AICfunc <- function(p,q){ #AIC formula
  x <- arima(Original,order=c(p,2,q))
  AICvalue <- log(x$sigma2)+ 2*(p+q)/length(SecondDiff)
  return(AICvalue)
}
BICfunc <- function(p,q){ #BIC formula
  x <- arima(Original,order=c(p,2,q))
  BICvalue <- log(x$sigma2)+ (log(length(SecondDiff))*(p+q)/length(SecondDiff))
  return(BICvalue)
}

for (i in 1:4){
  for(j in 1:4){
    AICmodelP2[i,j] <- AICfunc(i-1,j-1) #fit the AIC values to corresponding p and q


  }
}
for (i in 1:4){
```

```
  for(j in 1:4){
    BICmodelP2[i,j] <- BICfunc(i-1,j-1) #fit the BIC values to corresponding p and q


  }
}


###Part 2 Q3###
modelP2_1 <- arima(Original,order=c(2,2,1)) #find the estimated ARIMA(2,2,1) model  for data2
coef_2 <- as.array(modelP2_1$coef) #estimated coefficients of the ARIMA(2,2,1) model
sigma_sq_epq_2 <- modelP2_1$sigma2 #estimated innovation variance of the ARIMA(2,2,1) model


###Part 2 Q4###
forecast_nP2 <- 500 #number of forecasting values
forecast_P2 <- forecast(modelP2_1, h=forecast_nP2,level=95) #generate 95% forecast interval for next 500 values of the time series
#plot the forecast intervals along with the data
est_plotP2 <- autoplot(forecast_P2, legend=TRUE) + ggtitle("Forecast - Estimated Coefficients")+
  xlab('Time')+ ylab('Data Points')
```