

EMHW2

Problem 1

- a. Recall the autocorrelation function of AR(2) model satisfy the following recursive relations:

$$\rho_j = \phi_1 \rho_{j-1} + \phi_2 \rho_{j-2}$$

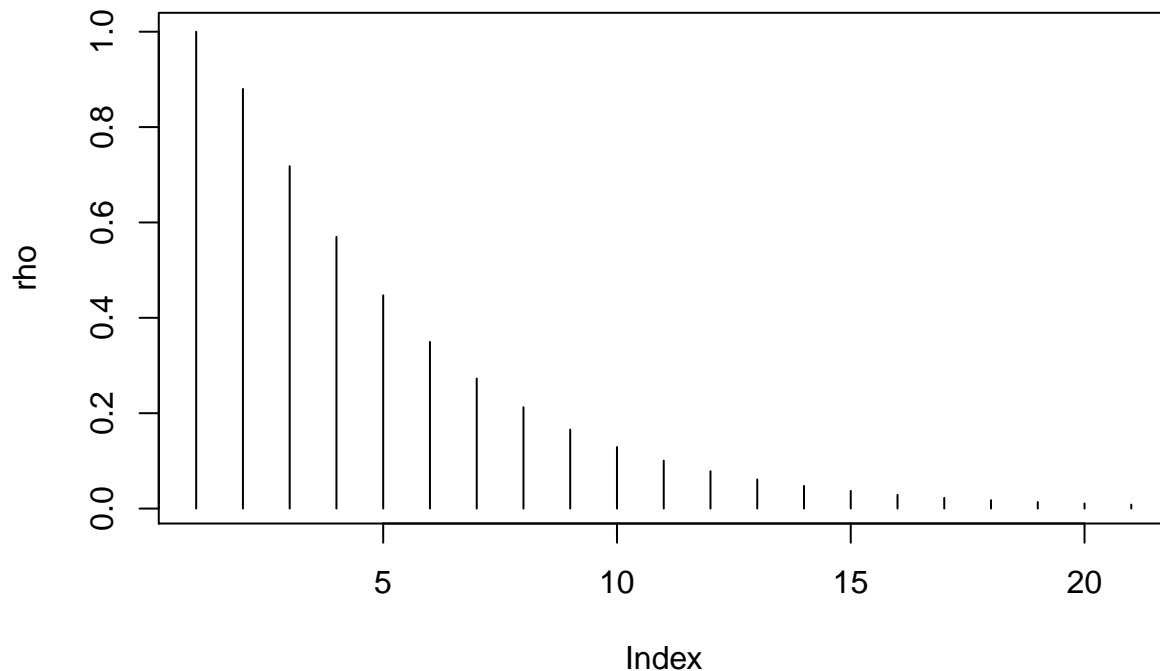
which also implies $\rho_1 = \phi_1 \rho_0 + \phi_2 \rho_1$ because we assume stationary.

Recall $\rho_0 = 1$ solve for $\rho_1 = \frac{\phi_1}{1-\phi_2} = \frac{1.1}{1-(-0.25)} = 0.88$

Now we have the initial two conditions $\rho_0 = 1, \rho_1 = 0.88$, and so we can solve for ρ_j using the recursive relation above.

```
#initialization
phi1=1.1
phi2=-0.25
rho=rep(0,21)
rho[1]=1.0
rho[2]=0.88
for (j in 3:21){
  rho[j]=phi1*rho[j-1]+phi2*rho[j-2]
}
plot(rho,type='h',main='Autocorrelation fuction of AR(2) model ')
```

Autocorrelation fuction of AR(2) model



- b. The characteristic polynomial should be: $1 - 1.1x + 0.25x^2 = 0$. Solve for the roots, we have

$$x_1, x_2 = \frac{\phi_1 + \sqrt{\phi_1^2 + 4\phi_2}}{-2\phi_2} = \frac{1.1 + \sqrt{1.1^2 + 4 * (-0.25)}}{-2 * (-0.25)} = 3.116515, 1.283485$$

Because $|x_1|, |x_2| > 1$, the process is stationary.

c. define the de-meaned process $X_t = R_t - \mu$. Because we assume stationary, we have:

$$\frac{\partial R_{t+6}}{\partial \epsilon_t} = \frac{\partial X_t}{\partial \epsilon_{t-6}}$$

, and now we can start our recursive scheme.

We know $X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + \epsilon_t$, from which we can derive:

$$\frac{\partial X_{t-6}}{\partial \epsilon_{t-6}} = 1, \frac{\partial X_{t-5}}{\partial \epsilon_{t-6}} = \phi_1$$

Using chain rule of calculus, we can derive:

$$\frac{\partial X_{t-4}}{\partial \epsilon_{t-6}} = \phi_1 \frac{\partial X_{t-5}}{\partial \epsilon_{t-6}} + \phi_2 \frac{\partial X_{t-6}}{\partial \epsilon_{t-6}} = \phi_1^2 + \phi_2$$

Following the same process, we can derive:

$$\frac{\partial X_{t-3}}{\partial \epsilon_{t-6}} = \phi_1^3 + 2\phi_1\phi_2$$

$$\frac{\partial X_{t-2}}{\partial \epsilon_{t-6}} = \phi_1^4 + 3\phi_1^2\phi_2 + \phi_2^2$$

$$\frac{\partial X_{t-1}}{\partial \epsilon_{t-6}} = \phi_1^5 + 4\phi_1^3\phi_2 + 3\phi_1\phi_2^2$$

$$\frac{\partial X_t}{\partial \epsilon_{t-6}} = \phi_1^6 + 5\phi_1^4\phi_2 + 6\phi_1^2\phi_2^2 + \phi_2^3 = 0.379561$$

d. if we now assume $\phi_1 = 0.9, \phi_2 = 0.8$, the dynamic multiplier would be

$$\frac{\partial X_t}{\partial \epsilon_{t-6}} = \phi_1^6 + 5\phi_1^4\phi_2 + 6\phi_1^2\phi_2^2 + \phi_2^3 = 6.778241$$

In this case, the characteristic polynomial would be $1 - 0.9x - 0.8x^2 = 0$. Solve for the roots, we have

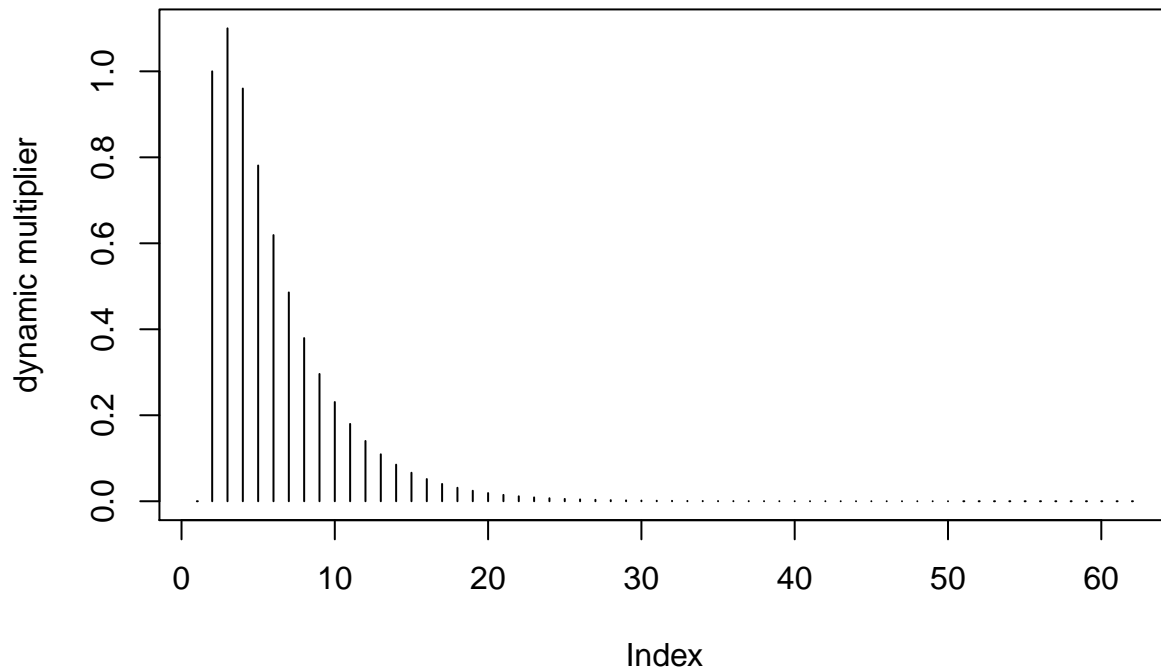
$$x_1, x_2 = \frac{\phi_1 + \sqrt{\phi_1^2 + 4\phi_2}}{-2\phi_2} = \frac{0.9 + \sqrt{0.9^2 + 4 * 0.8}}{-2 * 0.8} = -1.814062, 0.6890615$$

Because $|x_2| < 1$, this process is not stationary.

e.

```
x= rep(0,63) #initialization
x[3]=1 # inject the shock
for (j in 4:63){
  x[j]=1.1*x[j-1]-0.25*x[j-2] # set all further shocks to be zeros
}
plot(x[2:63],type='h',main='Impulse-Response plot for a one standard deviation positive shock',
      ylab = 'dynamic multiplier')
```

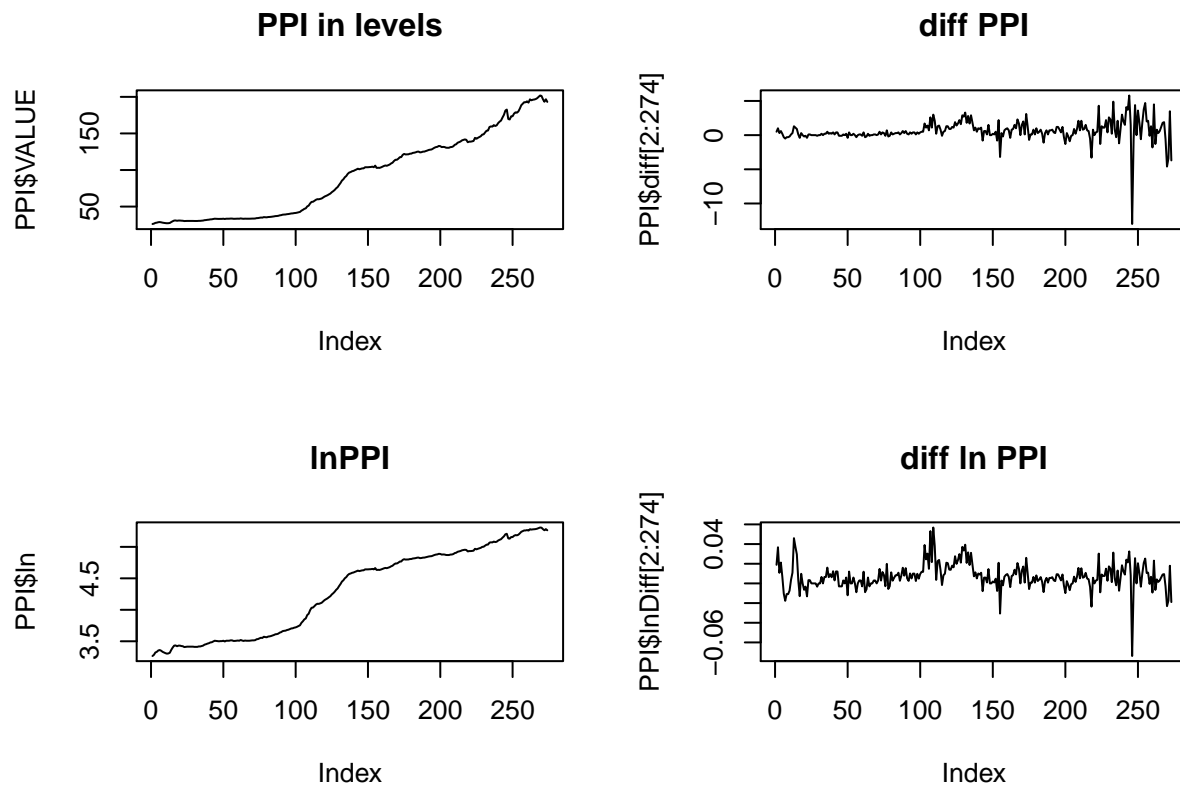
Impulse–Response plot for a one standard deviation positive shock



Problem 2

1.

```
#import the data
library(readxl)
PPI=read_excel("PPIFGS.xls", col_types = c("date", "numeric"))
#compute diff PPI, log(PPI),diff(log(PPI))
PPI['diff']=0
PPI$lnDiff=0
PPI$diff[2:274]=diff(PPI$VALUE)
PPI$ln=log(PPI$VALUE)
PPI$lnDiff[2:274]=diff(PPI$ln)
#plot the four subplot
par(mfrow=c(2,2))
plot(PPI$VALUE,type='l',main='PPI in levels')
plot(PPI$diff[2:274],type='l',main='diff PPI')
plot(PPI$ln,type='l',main='lnPPI')
plot(PPI$lnDiff[2:274],type='l',main='diff ln PPI')
```



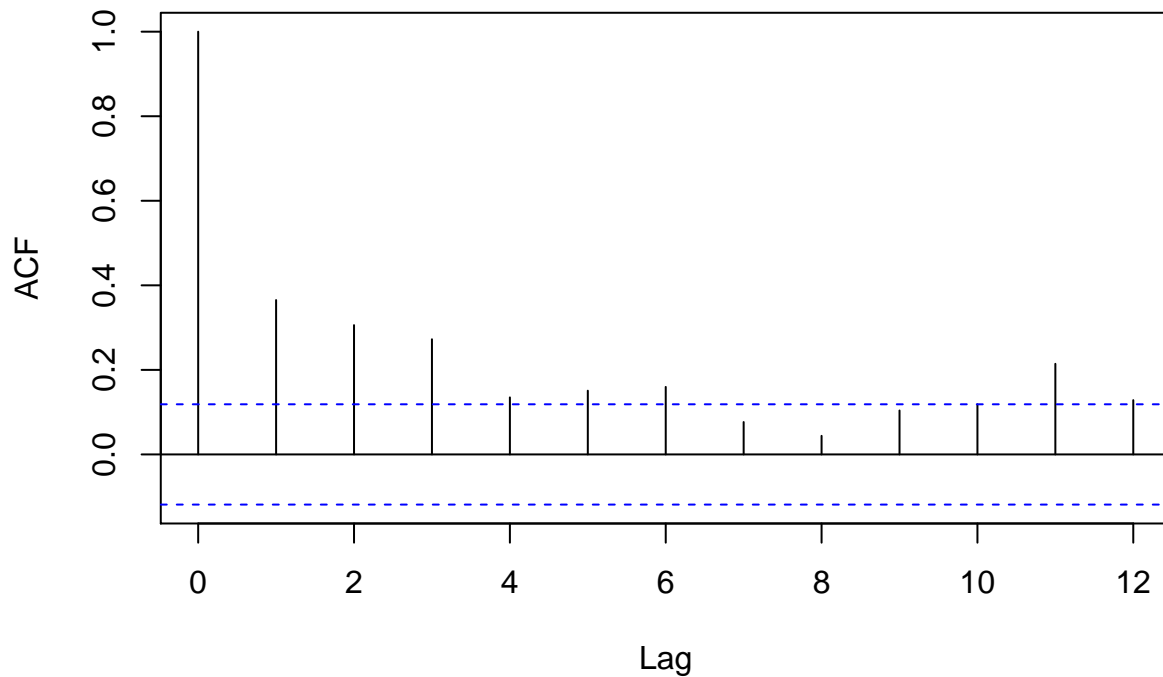
2. The $\Delta \log PPI$ series looks covariance-stationary, because it has no up- or downward trend and its variance does not change over time.

```
Y_t=PPI$lnDiff[2:274]
```

- 3.

```
acf(Y_t,lag.max = 12)
```

Series Y_t

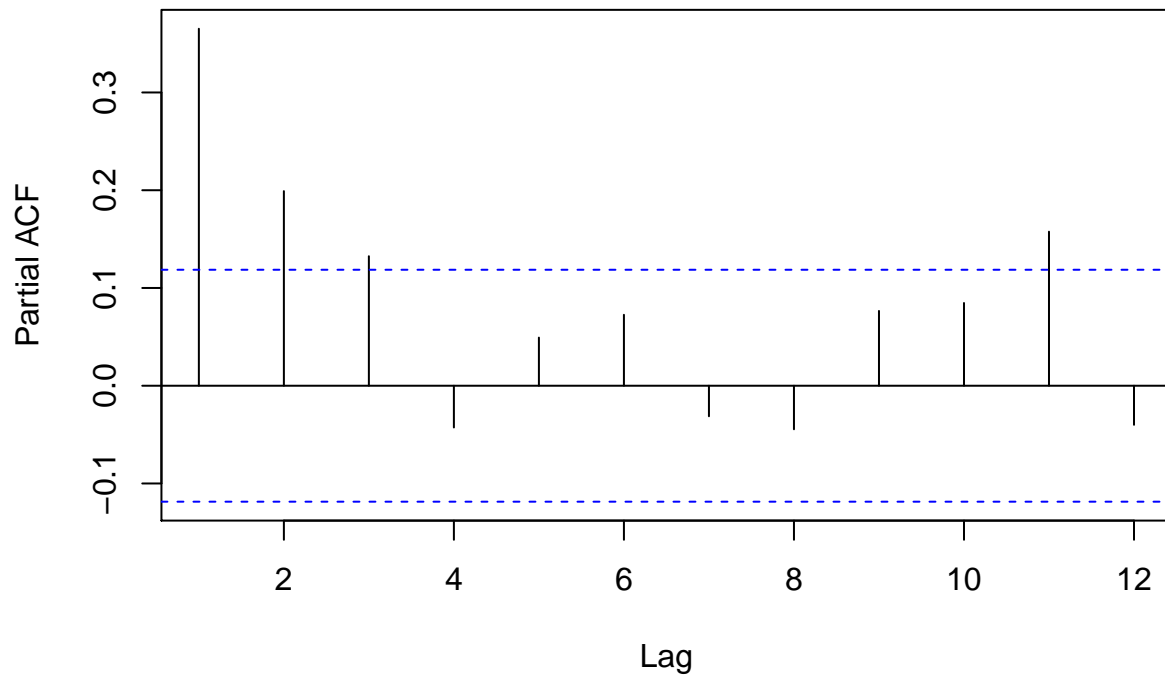


We observed some seasonality from the serie. In particular, the lag 4, 8 , 12 quarter's PPI is much less correlated with current quarter's PPI.

4.

```
pacf(Y_t, lag.max = 12)
```

Series Y_t



From

the pacf graph, we observed strong seasonality in the serie and a possible periodicity of 3 years, since the partial autocorrelation of lag 11 quarter is statistically significant.

5. From observing the plot of acf and pacf, we decided to select AR(3) and AR(11) as our candidate models. From the plot, we can observe that the autocorrelation and partial autocorrelation die out quickly after three lags and are not statistically significant except lag 11. Therefore, we propose AR (3) and AR(11) model.

a. For AR(3) model

```
#fit the AR(3) Model
AR_3=arima(x=Y_t,order=c(3,0,0))
print(AR_3)

##
## Call:
## arima(x = Y_t, order = c(3, 0, 0))
##
## Coefficients:
##          ar1          ar2          ar3  intercept
##          0.2687  0.1607  0.1402          0.0073
## s.e.    0.0602  0.0628  0.0614          0.0016
##
## sigma^2 estimated as 0.0001389:  log likelihood = 824.87,  aic = -1639.75
```

To check stationarity, we have so solve the characteristic polynomial and check whether the modulus of the roots are all greater than 1

```
##check for stationary
library(polynom)

## Warning: package 'polynom' was built under R version 3.5.2
Char_poly=polynomial(as.vector(AR_3$coef))
Char_roots=solve(Char_poly)
print(Mod(Char_roots))

## [1] 18.040015  1.426272  1.426272
```

As all the modulus of the roots are greater than 1, the fitted AR(3) Model is stationary.

For AR(11) Model

```
AR_11=arima(x=Y_t,order=c(11,0,0))
print(AR_11)

##
## Call:
## arima(x = Y_t, order = c(11, 0, 0))
##
## Coefficients:
##          ar1          ar2          ar3          ar4          ar5          ar6          ar7          ar8
##          0.2547  0.1697  0.1436 -0.0680  0.0318  0.0756 -0.0427 -0.1133
## s.e.    0.0599  0.0630  0.0639  0.0649  0.0647  0.0647  0.0650  0.0649
##          ar9          ar10          ar11  intercept
##          0.0262  0.0473  0.1715          0.0071
## s.e.    0.0645  0.0636  0.0618          0.0022
##
## sigma^2 estimated as 0.0001312:  log likelihood = 832.35,  aic = -1638.71
```

Stationary Check

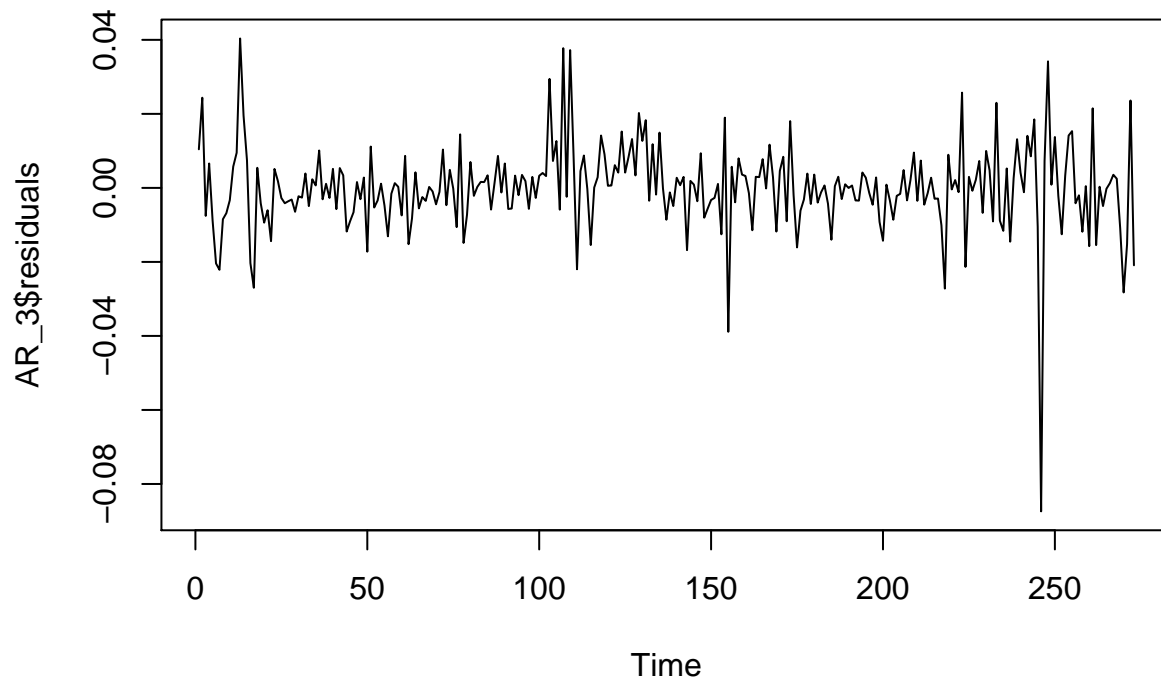
```
##check for stationary
Char_poly=polynomial(as.vector(AR_11$coef))
Char_roots=solve(Char_poly)
print(Mod(Char_roots))
```

```
## [1] 23.9499852  1.0361635  1.0361635  0.9319346  0.9319346  1.1147328
## [7]  1.1147328  1.0188528  1.0188528  1.1176002  1.1176002
```

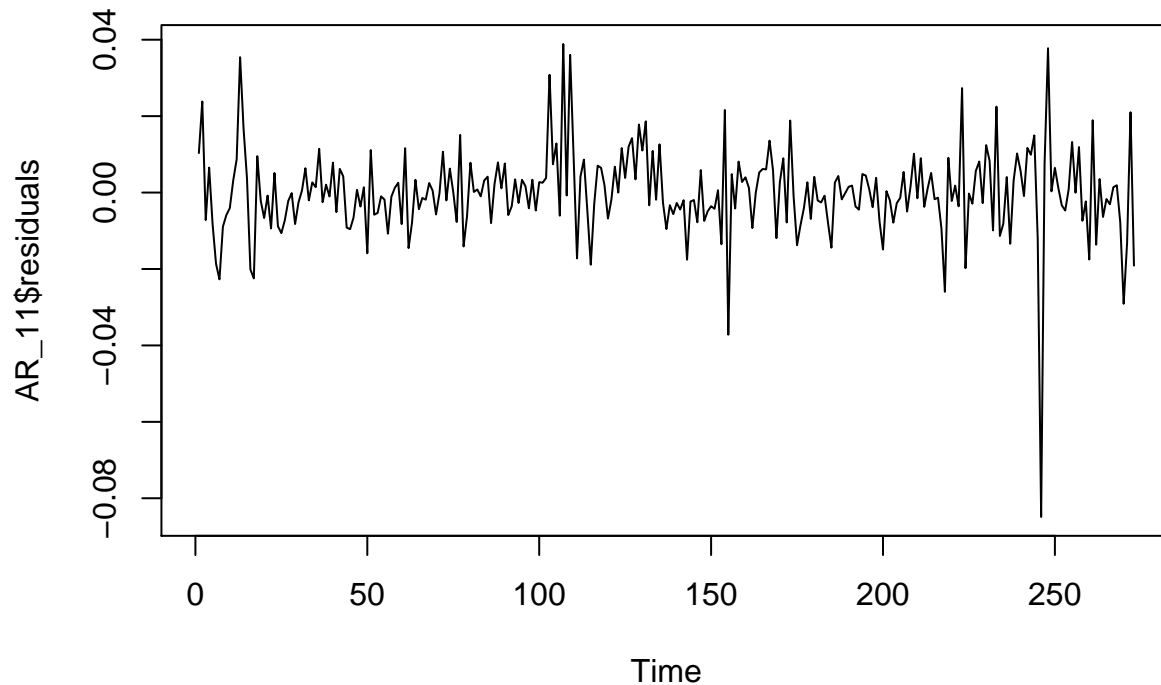
There are roots whose modulus is smaller than 1, so the fitted AR(11) Model is NOT stationary.

b. plot the residual of AR(3) Model

```
plot(AR_3$residuals)
```



```
plot(AR_11$residuals)
```



c. perform Box-Ljung Test on the residuals and report Q-stats of 8 and 12 quarters For the AR(3) model residuals of 8 quarter

```
Box.test(AR_3$residuals,lag = 8,type = 'Ljung-Box')
```

```
##
## Box-Ljung test
##
## data: AR_3$residuals
## X-squared = 5.3458, df = 8, p-value = 0.7201
```

For the AR(3) model residuals of 12 quarter

```
Box.test(AR_3$residuals,lag = 12,type = 'Ljung-Box')
```

```
##
## Box-Ljung test
##
## data: AR_3$residuals
## X-squared = 13.829, df = 12, p-value = 0.3118
```

For the AR(11) model residuals of 8 quarter

```
Box.test(AR_11$residuals,lag = 8,type = 'Ljung-Box')
```

```
##
## Box-Ljung test
##
## data: AR_11$residuals
## X-squared = 0.16598, df = 8, p-value = 1
```

For the AR(11) model residuals of 11 quarter

```
Box.test(AR_11$residuals,lag = 11,type = 'Ljung-Box')
```

```
##
```



```
## Box-Ljung test
##
## data: AR_11$residuals
## X-squared = 0.33232, df = 11, p-value = 1
```

AIC and BIC for AR(3) model:

```
print(c(AIC(AR_3),BIC(AR_3)))
```

```
## [1] -1639.746 -1621.699
```

AIC and BIC for AR(11) model:

```
print(c(AIC(AR_11),BIC(AR_11)))
```

```
## [1] -1638.706 -1591.783
```

From the tests above, we concluded that AR(3) model is superior to AR(11) model because AR(11) model has many statistically insignificant coefficient, is unstationary, and has greater AIC and BIC.

6.

```
Y_t_train=Y_t[1:234]
Y_t_test=Y_t[235:length(Y_t)]
SPE=data.frame(matrix(data = 0,ncol = 3,nrow = length(Y_t_test)))
colnames(SPE)=c('AR3','AR11','RandomWalk')
RW_sr=rep(0,length(Y_t_test))
drift=mean(diff(Y_t))
RW_sr[1]=Y_t[234]+rnorm(1,mean = 0,sd = 1)+drift #initialize the first term of the random walk model
for (t in 2:length(Y_t_test)){
  RW_sr[t]=RW_sr[t-1]+rnorm(1,mean = 0,sd = 1)+drift
}
for (t in 1:(length(Y_t_test))){
  #fit the model
  AR_3_insam=arima(x=Y_t[1:233+t],order = c(3,0,0))
  AR_11_insam=arima(x=Y_t[1:233+t],order = c(11,0,0))
  #perform prediction
  SPE$AR3[t]=(as.numeric(predict(AR_3_insam,n.ahead = 1)$pred)-Y_t_test[t])^2
  SPE$AR11[t]=(as.numeric(predict(AR_11_insam,n.ahead = 1)$pred)-Y_t_test[t])^2
  SPE$RandomWalk[t]=(RW_sr[t]-Y_t_test[t])^2
}
colMeans(SPE)
```

```
##          AR3          AR11  RandomWalk
## 4.167217e-04 4.081279e-04 5.798395e+01
```

From the MSPE above, we can see that AR3 and AR11 have a much smaller out-of-sample prediction error than the random walk model. For these two models, they have almost identical out-of sample prediction errors, so we finally prefer AR3 over AR11 because of its stationarity, and better performance on AIC and BIC metrics.