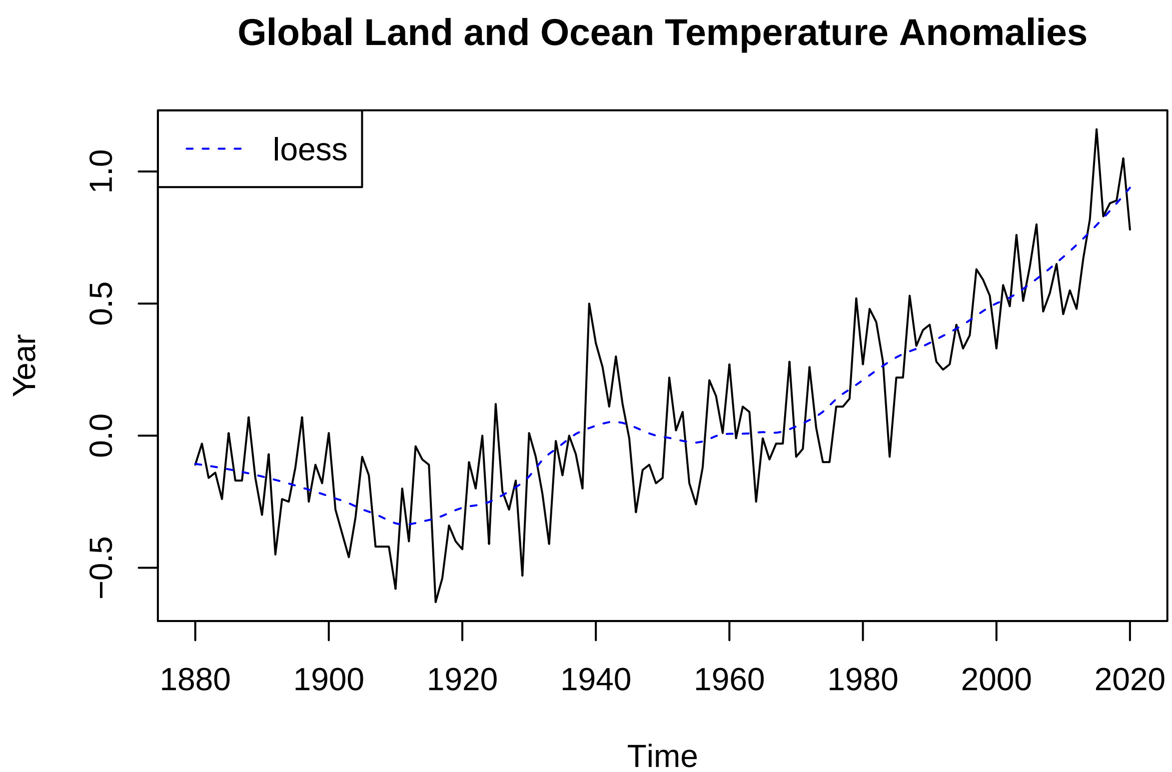
STA137HW4

Yirui Li,Chunqiu Li,Haoming Lei

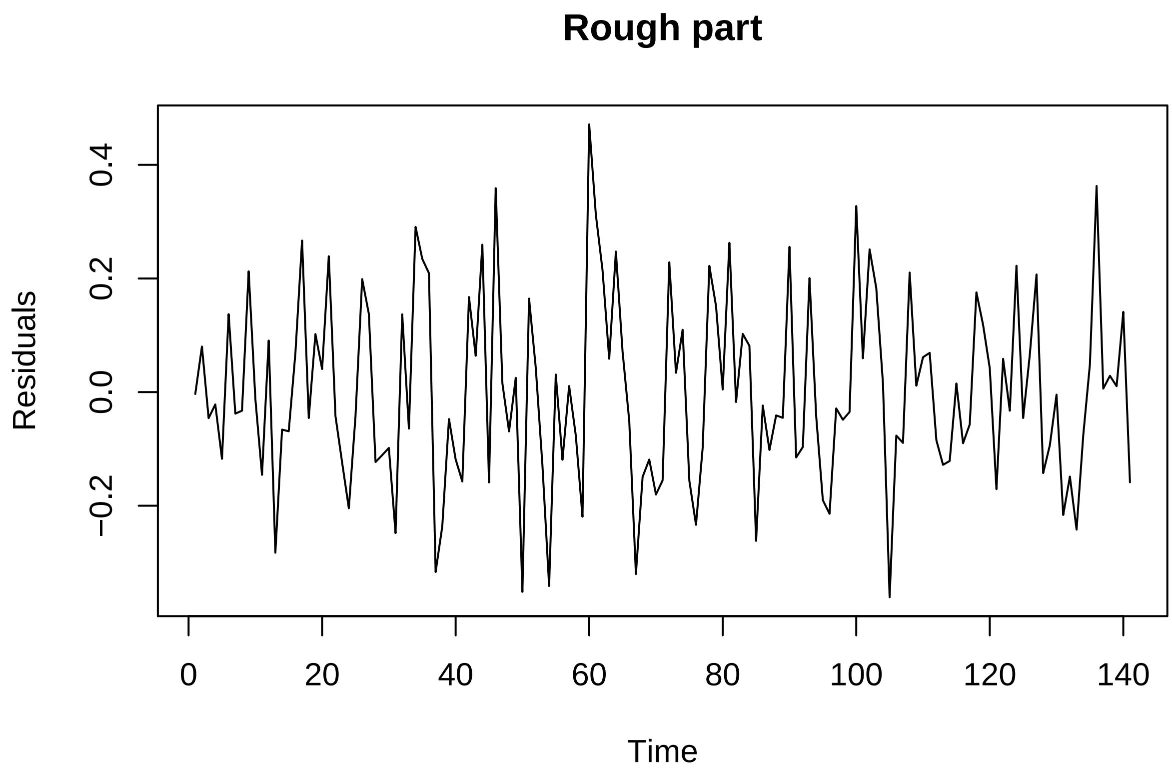
1a

```
data <- read.csv("C:/Users/Administrator/Desktop/sta137/GlobTempNASA_2020.csv")
newdata=data[-c(1,2,3,4),]
y<-newdata$December
tm<-newdata$Global.Land.and.Ocean.Temperature.Anomalies
trnd <- loess(y ~ tm, span = 0.3)
plot(tm, y, type = "l", lty = 1, xlab = "Time", ylab = "Year", main = "Global Land and Ocean Temperatur
points(tm, trnd$fitted,type = "l", lty = 2, col = "blue")
legend("topleft", "loess", lty = 2, col = "blue")
```



1b

```
rough = trnd$residuals
plot(rough, type = "l", xlab = "Time", ylab = "Residuals", main = "Rough part")
```

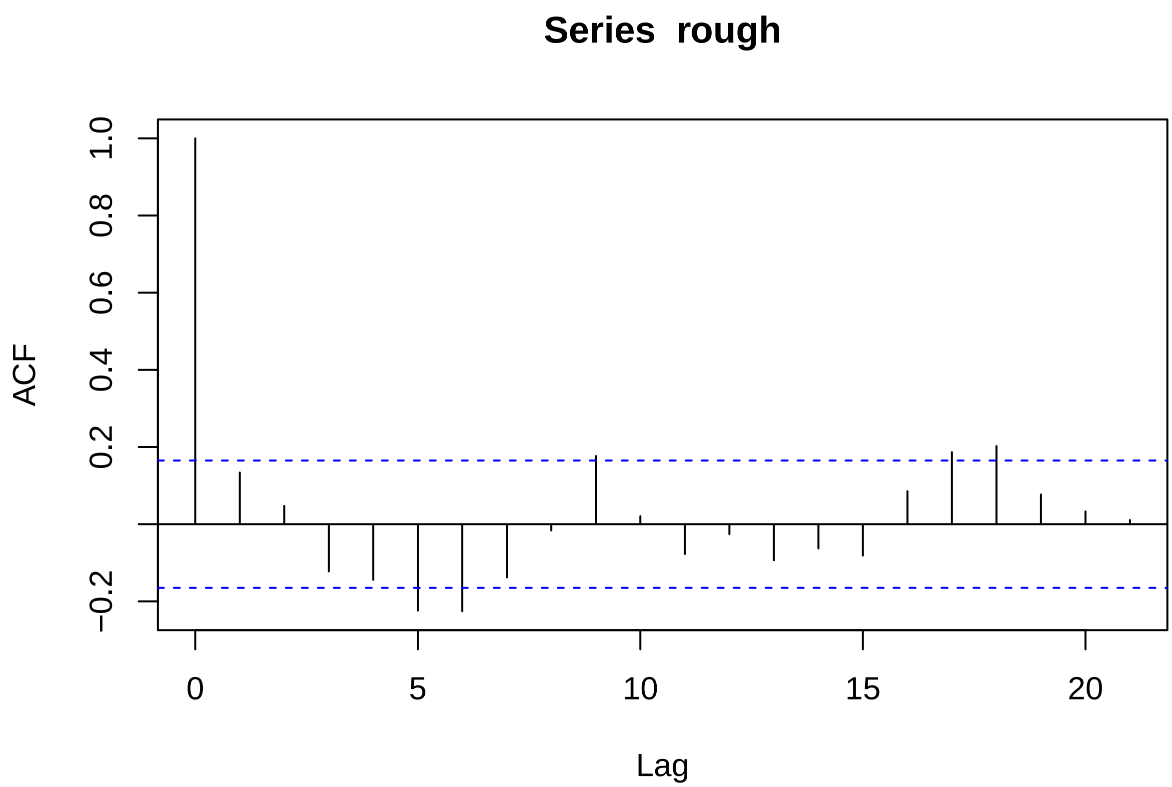



```
#The rough part of the trend is symmetrically distributed
```

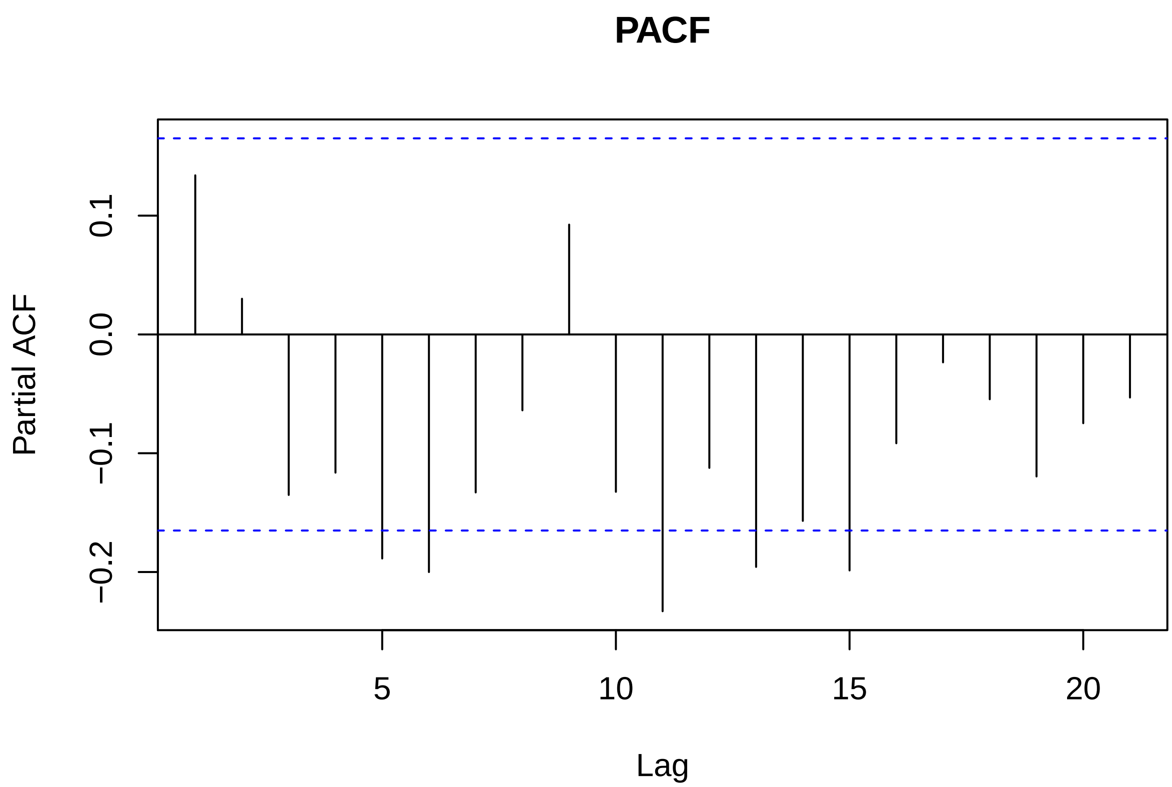
1c.

In the acf plot, there five p(j) outside the bound, which show that this model is fairly fit to the data.

```
acf(rough)
```




```r
#From the acf graph, there are six significance in the model
pacf(rough,main="PACF")
```




```
#From the pacf graph, there are five significance in the model,which mean we should use the ar(1) model
```

1d We see AR(6) gives the lowest BIC and lowest number of parameters.

```
aic_table=rep(0,6)
for (i in 0:5){##you can check for higher orders
   aic_table[i+1]=arima(rough,order=c(i,0,0))$aic
}
aic_table
```

```
## [1] -104.9375 -105.4893 -103.6089 -104.2036 -104.0697 -107.0694
```

1e ##Does this plot suggest that the model selected in part (d) a reasonable one? Yes, from the plot the autocorrelation values between each other is small,therefore it suggested the AR(5) model is accurate

##Also carry out the Ljung-Box test on the residuals at level a = 0:05. Summarize your findings. p-value is larger than alpha level, so we cannot reject Ho.

```
mod_ar5 <- arima(rough,order=c(5,0,0))
mod_ar5$coef
```

```
##           ar1           ar2           ar3           ar4           ar5
##  0.0959195897  0.0310941415 -0.1088444527 -0.0906333840 -0.1851227647
##     intercept
## -0.0002618578
```

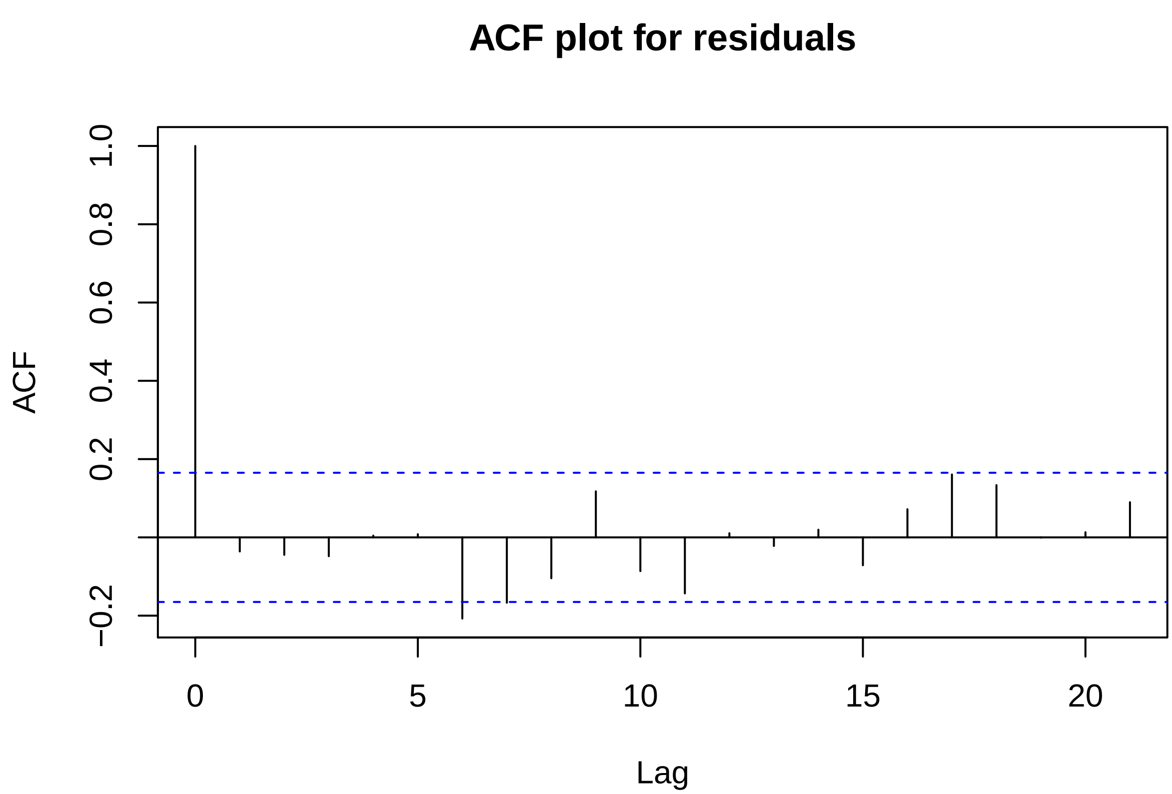
```r
mod_ar5$var.coef##variance covariance matrix, so select diagonals to get variance
```

```
##                     ar1           ar2           ar3           ar4           ar5
## ar1        6.832629e-03 -8.483368e-04 -3.491002e-04  8.677604e-04  8.009566e-04
## ar2       -8.483368e-04  6.864572e-03 -8.719928e-04 -4.255251e-04  8.088365e-04
## ar3       -3.491002e-04 -8.719928e-04  6.724257e-03 -9.141307e-04 -3.538982e-04
## ar4        8.677604e-04 -4.255251e-04 -9.141307e-04  6.765658e-03 -8.547344e-04
## ar5        8.009566e-04  8.088365e-04 -3.538982e-04 -8.547344e-04  6.716116e-03
## intercept -7.035091e-06  2.472104e-06  8.928982e-07  4.510128e-07 -3.828198e-06
##               intercept
## ar1       -7.035091e-06
## ar2        2.472104e-06
## ar3        8.928982e-07
## ar4        4.510128e-07
## ar5       -3.828198e-06
## intercept  1.128922e-04
```

```r
sqrt(diag(mod_ar5$var.coef))##std.error = sqrt(variance)
```

```
##        ar1        ar2        ar3        ar4        ar5  intercept
## 0.08265972 0.08285271 0.08200157 0.08225362 0.08195192 0.01062507
```

```r
acf(mod_ar5$residuals,main="ACF plot for residuals")
```

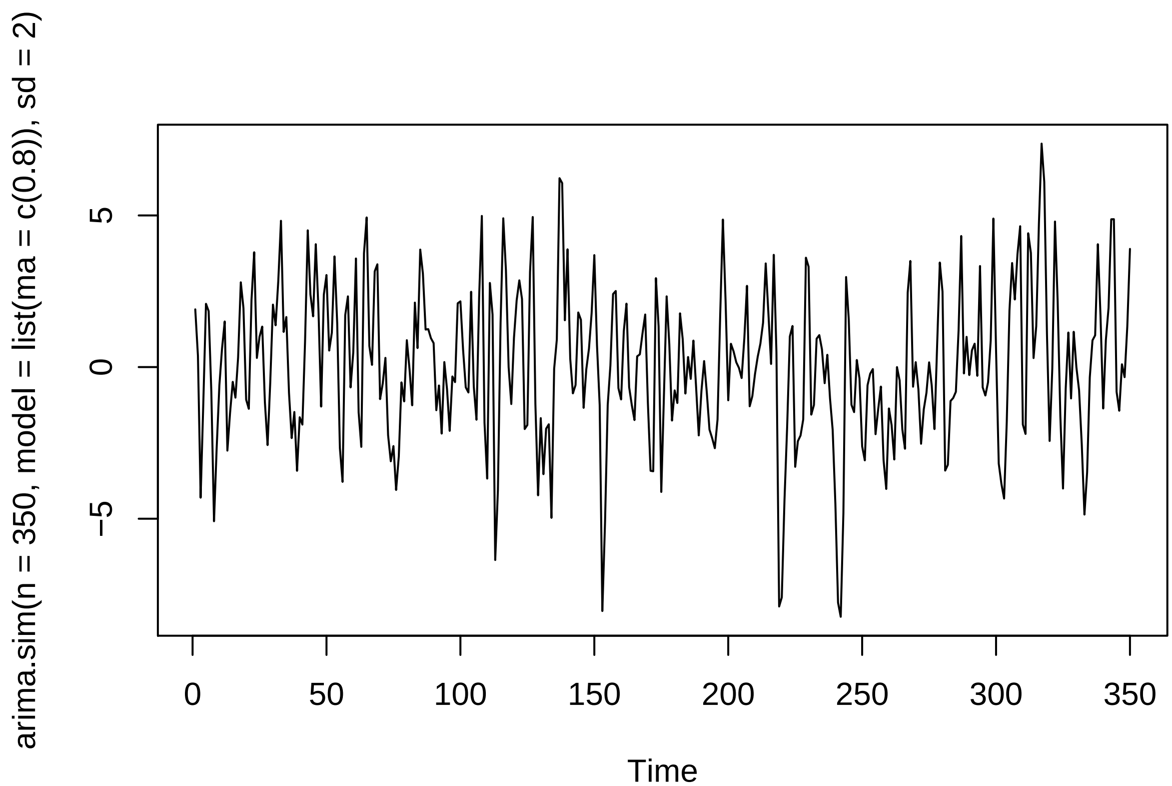


```r
Box.test(mod_ar5$residuals, lag=10, type='Ljung-Box')
```

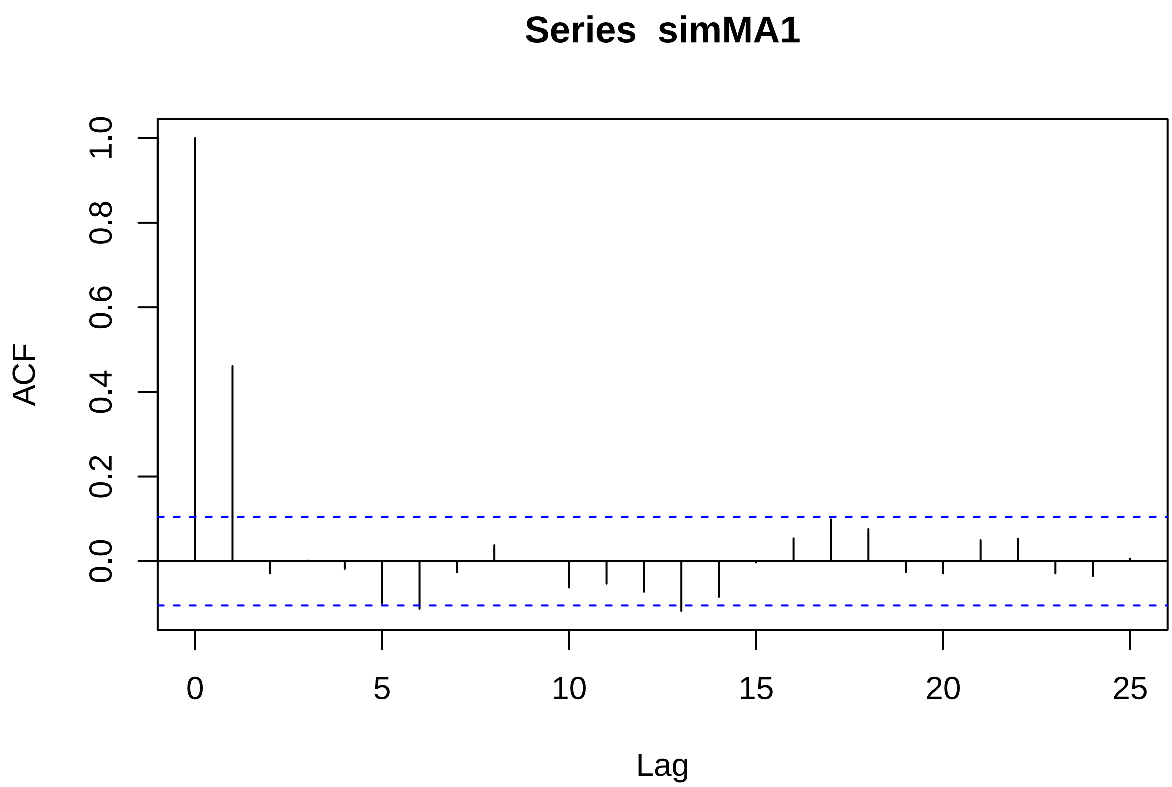
```
##
##  Box-Ljung test
##
## data:  mod_ar5$residuals
## X-squared = 16.367, df = 10, p-value = 0.08959
```

2a As expected, for the MA(1) case, the ACF plot shows that only p^(1) is outside the +-1.96/sqrt(n) bars and the rest are inside, which shows that MA(1) is not a very appropriate model for this.

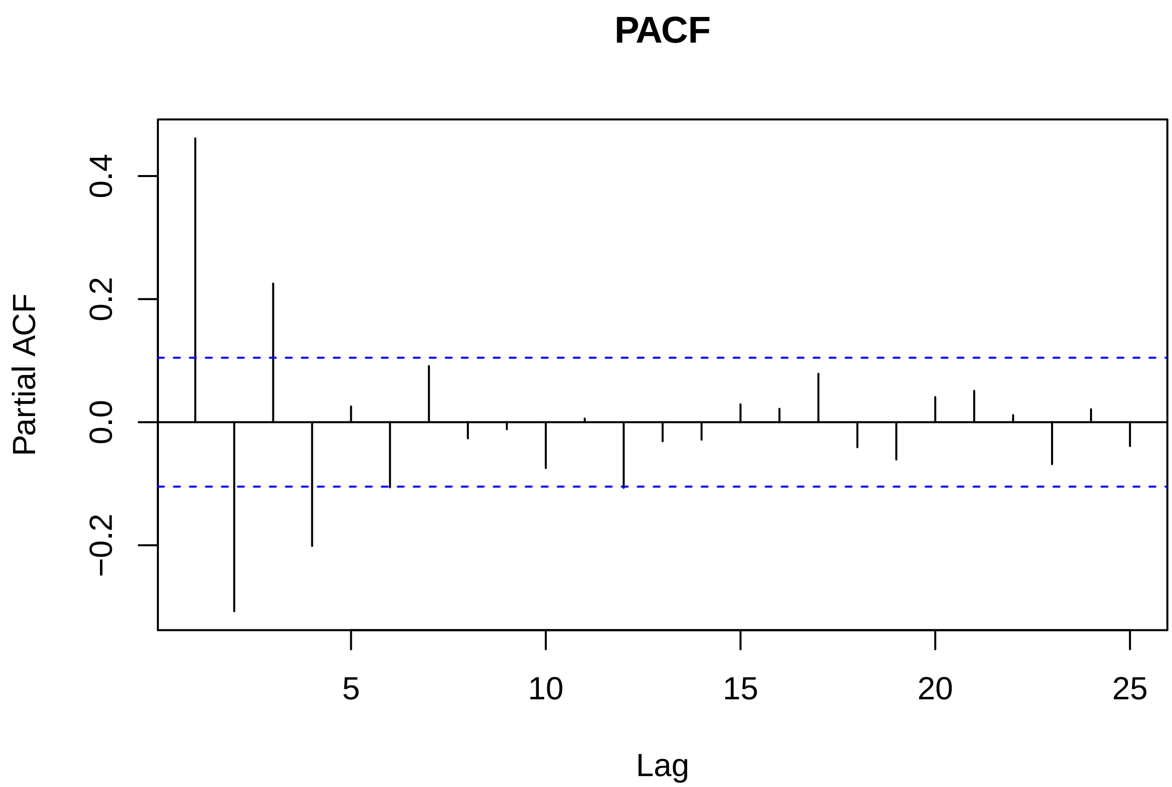
```r
simMA1 <- arima.sim(n=350, model=list(ma=c(0.8)),sd = 2)
ts.plot(arima.sim(n=350, model=list(ma=c(0.8)),sd = 2))
```




```r
acf(simMA1)
```

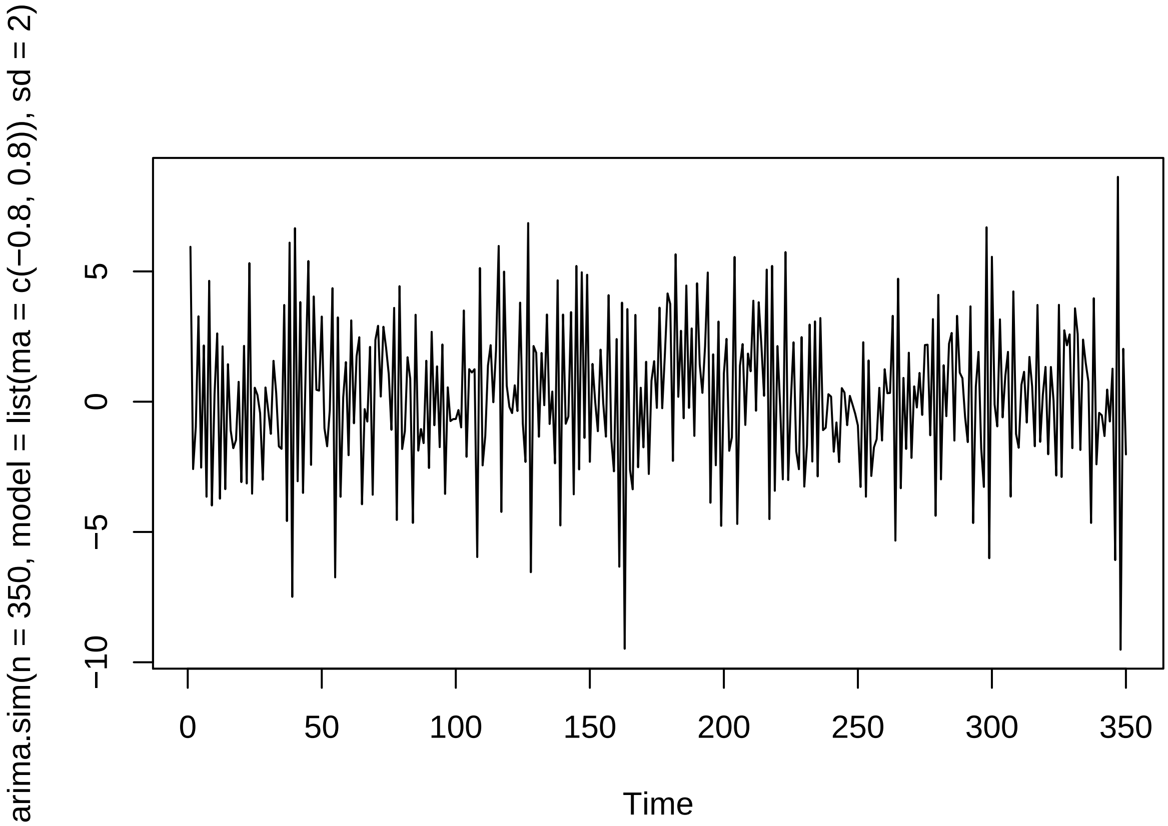



```r
pacf(simMA1,main="PACF")
```

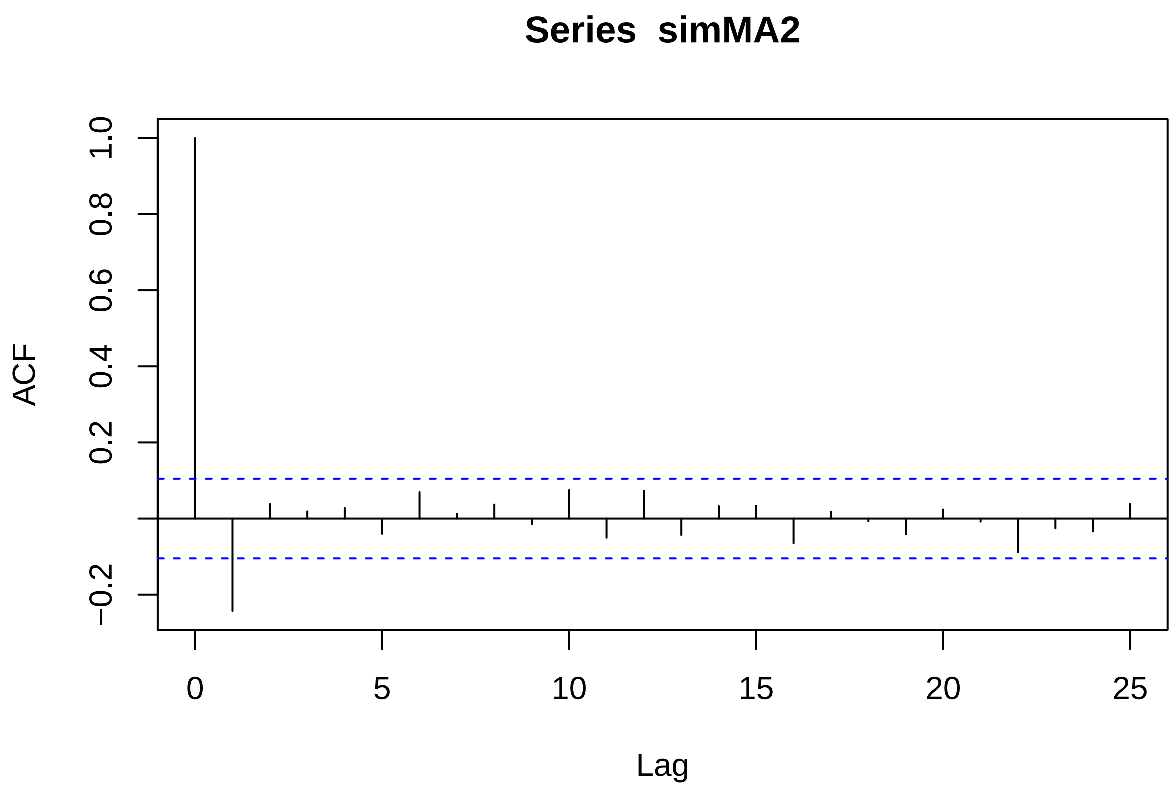


2b As expected, for the MA(2) case, the ACF plot shows that only p^(1) is outside the +-1.96/sqrt(n) bars and the rest are inside, which shows that MA(2) is an appropriate model.

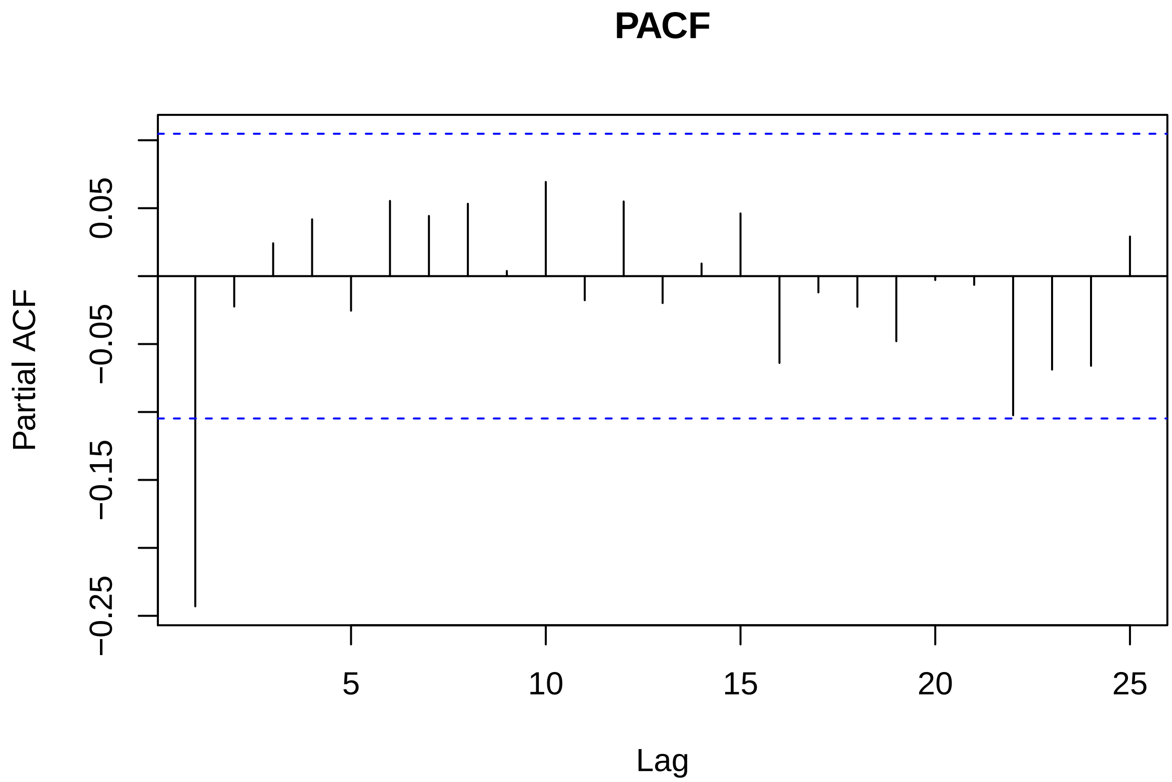
```
simMA2 <- arima.sim(n=350, model=list(ma=c(-08,0.8)),sd = 2)
ts.plot(arima.sim(n=350, model=list(ma=c(-0.8,0.8)),sd = 2))
```


arima.sim(n = 350, model = list(ma = c(−0.8, 0.8)), sd = 2)
−10
−5
0
5
0
50
100
150
200
250
300
350
Time

```r
acf(simMA2)
```

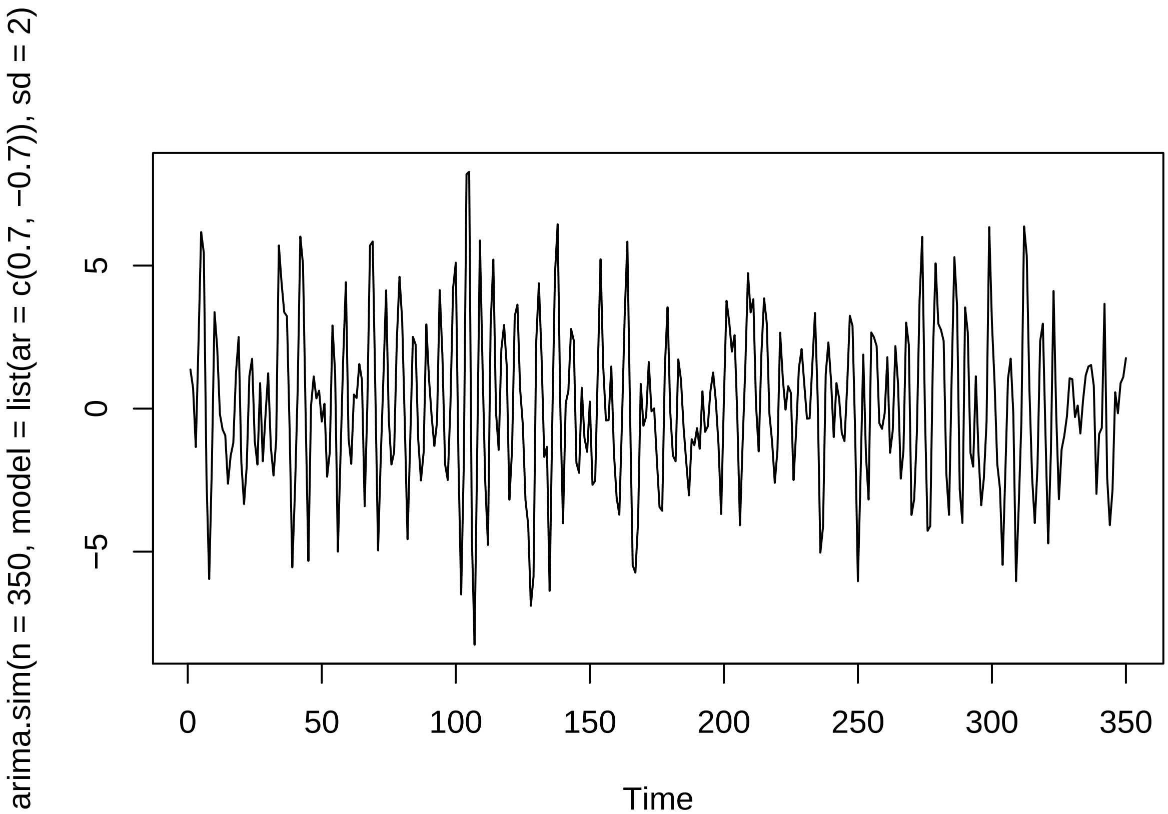



```r
pacf(simMA2,main="PACF")
```

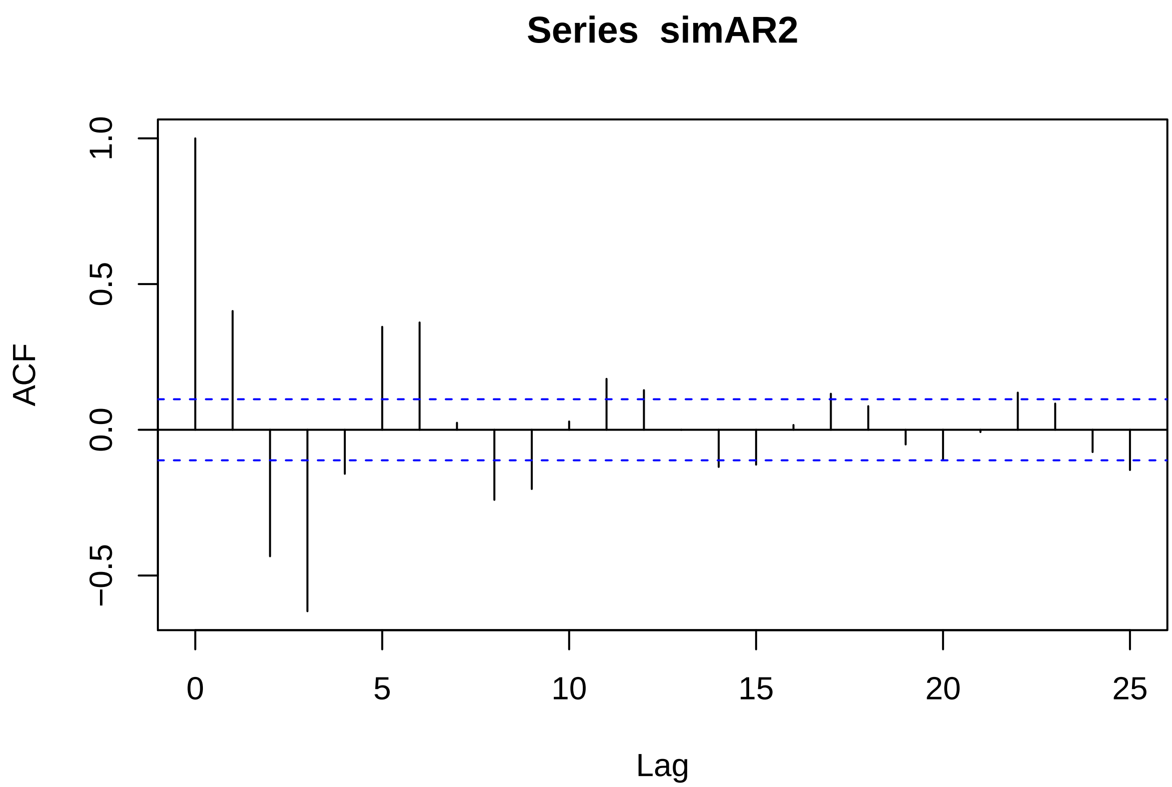


2c In this case, most p(j) are outside the +-1.96/sqrt(n).

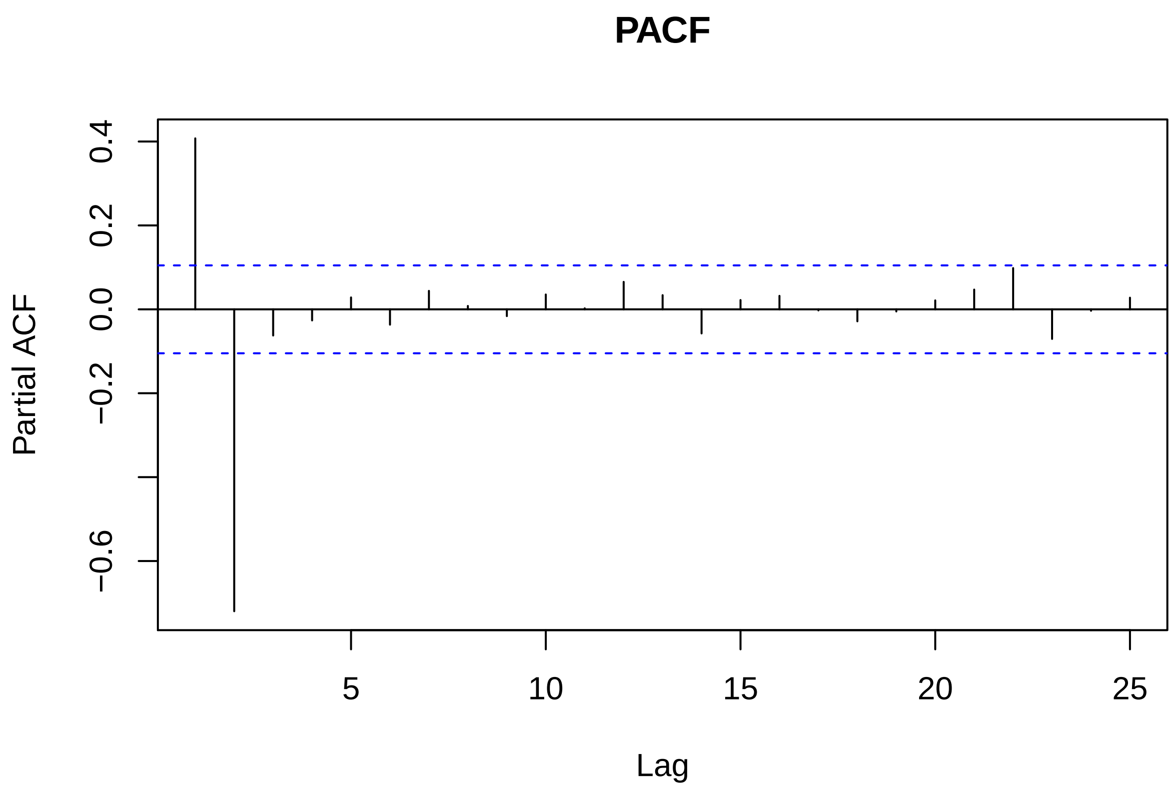
```
simAR2 <- arima.sim(n=350, model=list(ar=c(0.7,-0.7)),sd = 2)
ts.plot(arima.sim(n=350, model=list(ar=c(0.7,-0.7)),sd = 2))
```


arima.sim(n = 350, model = list(ar = c(0.7, −0.7)), sd = 2)
−5
0
5
0
50
100
150
200
250
300
350
Time

```
acf(simAR2)
```

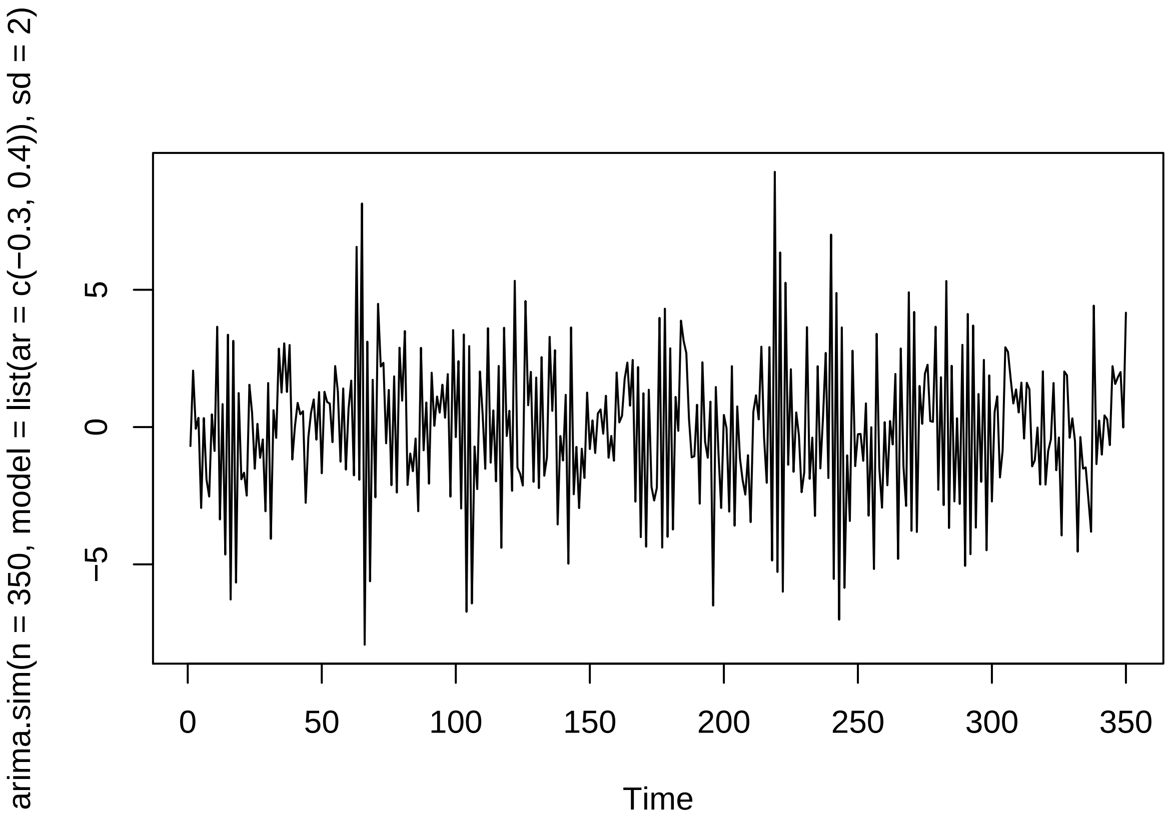



```
pacf(simAR2,main="PACF")
```

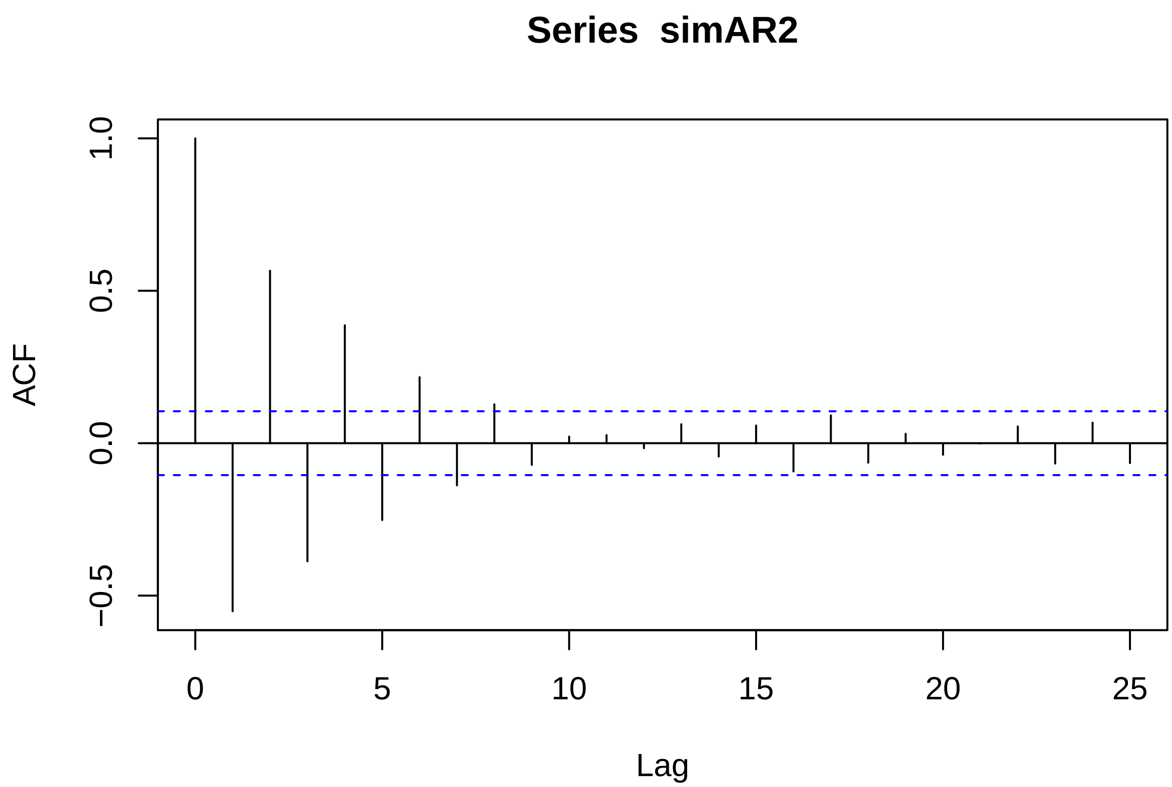


2d In this case, some of p(j) are outside the +-1.96/sqrt(n).

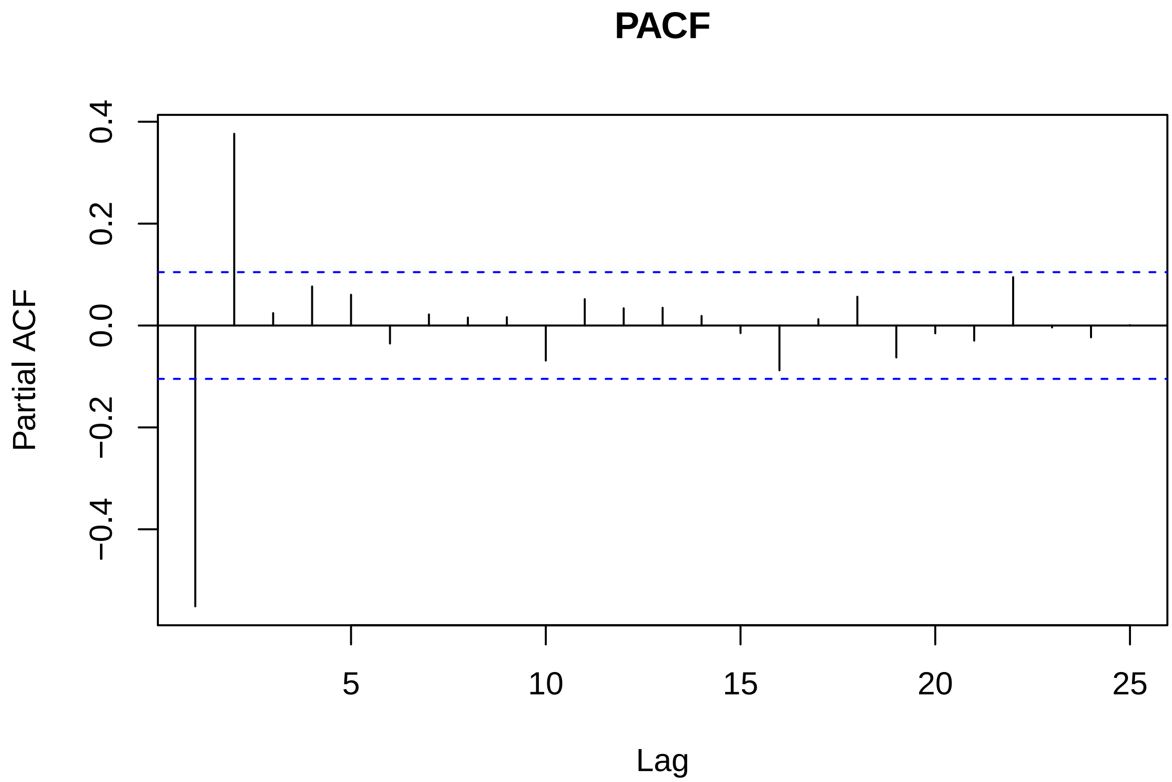
```
simAR2 <- arima.sim(n=350, model=list(ar=c(-0.3,0.4)),sd = 2)
ts.plot(arima.sim(n=350, model=list(ar=c(-0.3,0.4)),sd = 2))
```


arima.sim(n = 350, model = list(ar = c(−0.3, 0.4)), sd = 2)
−5
0
5
0
50
100
150
200
250
300
350
Time

```
acf(simAR2)
```



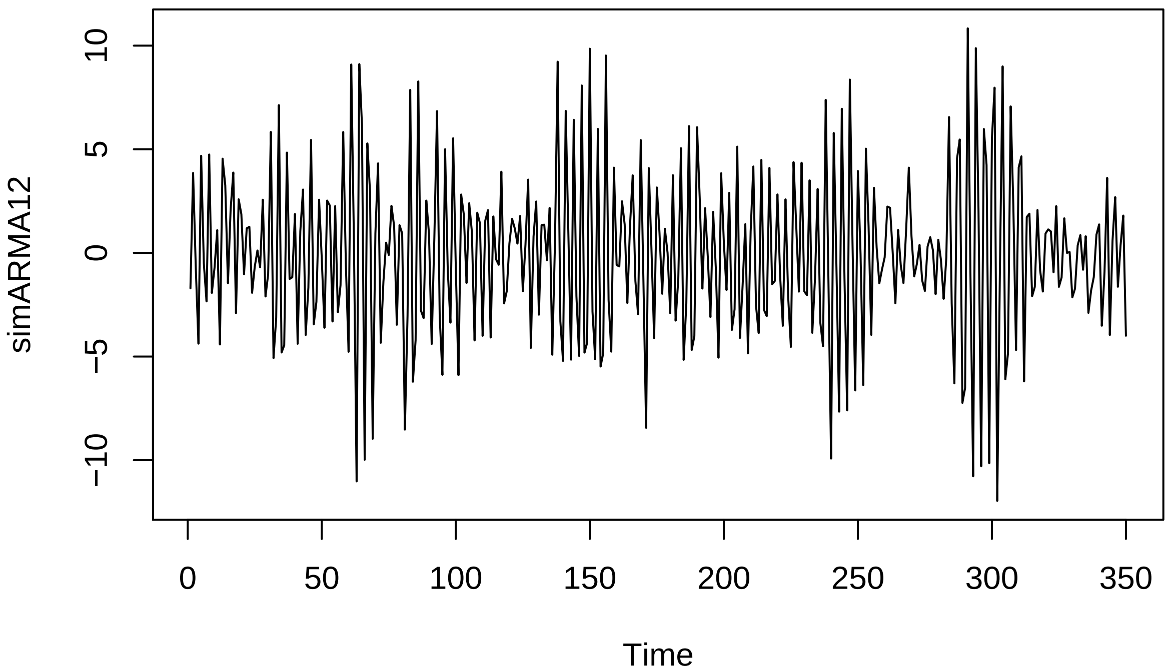

```
pacf(simAR2,main="PACF")
```



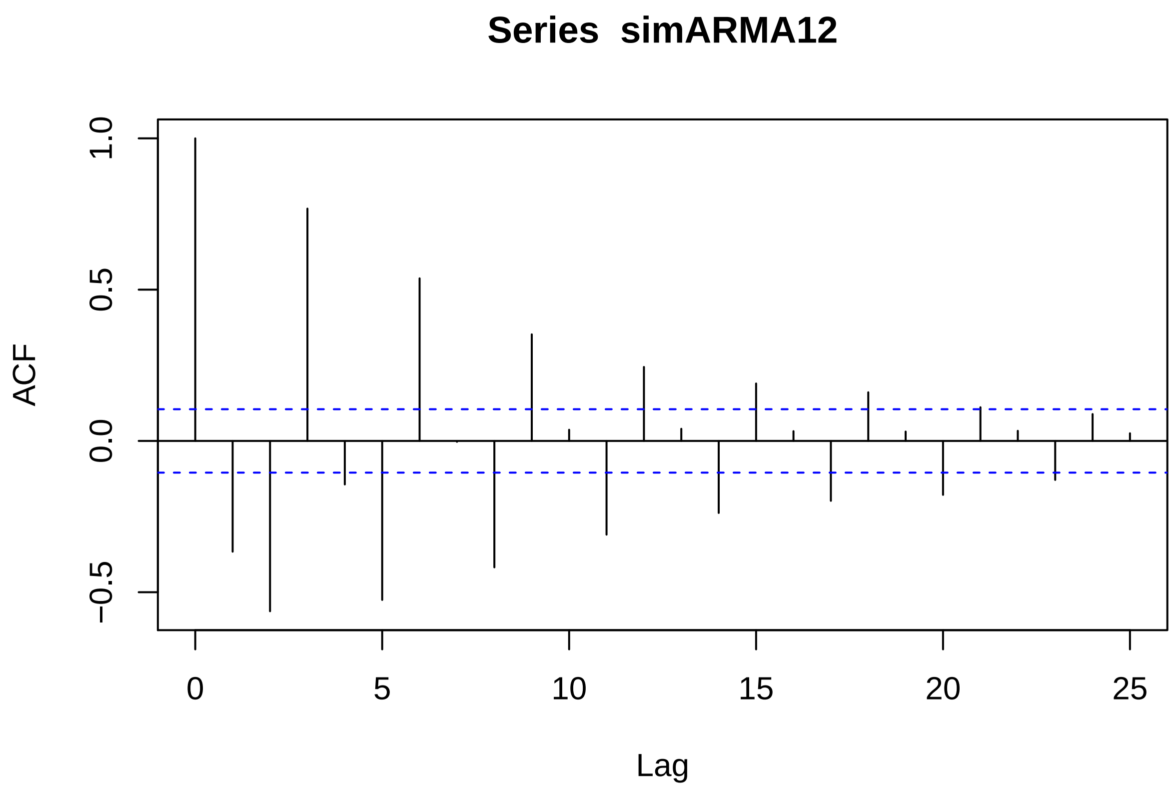
2e.

In this case, most of the p(j) are outside the +-1.96/sqrt(n).

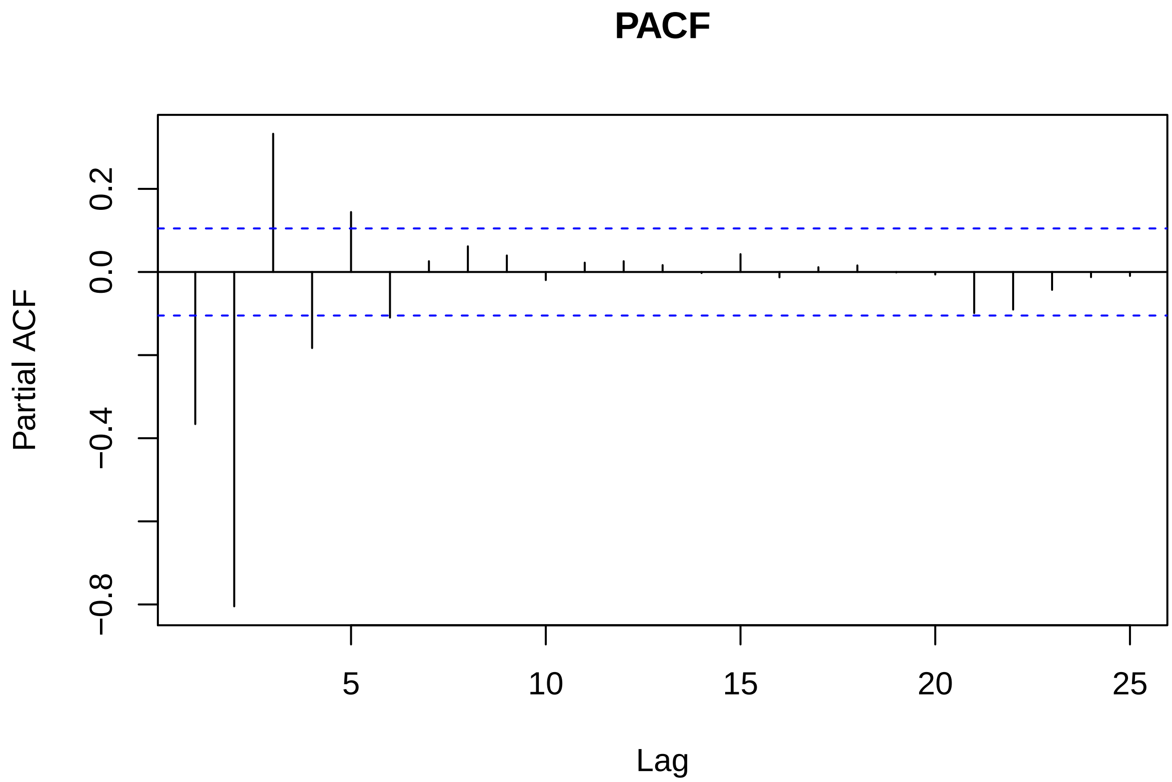
```
#alternatively
simARMA12 = arima.sim(model=list(ar = c(-0.8, -0.8), ma=0.6), n = 350,sd = 2)
ts.plot(simARMA12)
```




```
acf(simARMA12)
```




```
pacf(simARMA12,main="PACF")
```



3.

3.

a). If x_t follows an AR(1) model, then.

$x_t-\mu=\phi(x_{t-1}-\mu)+\varepsilon_t$ $x_t=\mu-\phi x_{t-1}-\mu\phi+\varepsilon_t$ $x_t^f=(1-\phi)\mu-\phi x_{t-1}$

$x_t-x_t^f=\varepsilon_t$ So x_t^f is the fitted value when regressing x_t on x_{t-1}

b) $x_{t-2}-\mu=\phi(x_{t-1}-\mu)+\varepsilon_t$ $x_{t-2}=\mu+\phi(x_{t-1}-\mu)+\varepsilon_t$

$x_{t-2}^b=\mu+\phi(x_{t-1}-\mu)$ So the fitted value x_{t-2}^b is a linear function of x_{t-1}

c. we have $E(X_t)=\mu$, $Var(X_t)=\gamma(0)$.

$$X_t^{(f)} = \mu + \phi(X_{t-1}-\mu)$$
$$= \mu + \phi X_{t-1} - \phi\mu$$
$$= \mu(1-\phi) + \phi X_{t-1}$$

$$(X_t - X_t^{(f)})^2 = (X_t - \phi X_{t-1} - \mu(1-\phi))^2$$
$$= X_t^2 + \phi^2 X_{t-1}^2 + \mu^2(1-\phi)^2 - 2\phi X_t X_{t-1} + 2\phi\mu(1-\phi)X_{t-1} - 2\mu(1-\phi)X_t$$

$$\therefore E(X_t - X_t^{(f)})^2 = E(X_t^2) + \phi^2 E(X_{t-1}^2) + \mu^2(1-\phi)^2 - 2\phi E(X_t X_{t-1}) + 2\phi\mu(1-\phi)E(X_{t-1}) - 2\mu(1-\phi)E(X_t)$$

$\because E(X_t)=\mu$, $Var(X_t)=\gamma(0)$

$\therefore E(X^2) = \gamma(0) + \mu^2$

$\gamma(1) = E(X_t X_{t-1}) - \mu^2$

$E(X_t X_{t-1}) = \gamma(1) + \mu^2$

Therefore, $E(X_t - X_t^{(f)})^2 = (\gamma(0)+\mu^2) + \phi^2(\gamma(0)+\mu^2) + \mu^2(1-\phi)^2 - 2\phi(\gamma(1)+\mu^2) + 2\phi\mu^2(1-\phi) - 2\mu^2(1-\phi)$

$$= \gamma(0)(1+\phi^2) - 2\phi\gamma(1) + \mu^2(1+\phi^2+(1-\phi)^2 - 2\phi + 2\phi - 2\phi^2 - 2 + 2\phi)$$
$$= (1+\phi^2)\gamma(0) - 2\phi\gamma(1) + \mu^2 \times 0$$
$$= (1+\phi^2)\gamma(0) - 2\phi\gamma(1)$$

d. As problem b proved, $X_{t-2}^{(b)} = \mu + \phi(X_{t-1}-\mu)$

$$E(X_{t-2} - X_{t-2}^{(b)})^2 = E(X_{t-2}^2) + \phi^2 E(X_{t-1}^2) + \mu^2(1-\phi)^2 - 2\phi E(X_{t-1}X_{t-2}) + 2\phi\mu(1-\phi)E(X_{t-1}) - 2\mu(1-\phi)(X_{t-2})$$

$\because E(X_{t-2})=\mu$, $Var(X_{t-2})=\gamma(0)$

$E(X^2) = \gamma(0) + \mu^2$

$\gamma(1) = E(X_{t-1}X_{t-2}) - \mu^2$

$E(X_{t-1}X_{t-2}) = \gamma(1) + \mu^2$

Therefore: $E(X_{t-2} - X_{t-2}^{(b)})^2 = (\gamma(0)+\mu^2) + \phi^2(\gamma(0)+\mu^2) + \mu^2(1-\phi)^2 - 2\phi(\gamma(1) + \mu^2(1+\phi^2+(1-\phi)^2 - 2\phi + 2\phi - 2\phi^2 - 2 + 2\phi)$

$$= \gamma(0) + \phi^2(\gamma(0)) - 2\phi\gamma(1) + \mu^2 \times 0$$
$$= \gamma(0)(1+\phi^2) - 2\phi\gamma(1)$$

e. set $X_t - \mu = Z_t$

$X_t - \hat{X}_t = X_t - \mu - \phi(X_{t-1} - \mu) = Z_t - \phi Z_{t-1}$

$cov(X_t - X_t^{(f)}, X_{t+2} - X_{t+2}^{(b)}) = cov(Z_t - \phi Z_{t+1}, Z_{t+2} - \phi Z_{t+1})$

$= \gamma(2) - \phi\gamma(1) - \phi\gamma(1) + \phi^2\gamma(0)$

$= \gamma(2) - 2\phi\gamma(1) + \phi^2\gamma(0)$

f. $PACF(1) = \dfrac{cov(X_t - X_t^{(f)}, X_{t+2} - X_{t+2}^{(b)})}{\sqrt{Var(X_t - X_t^{(f)})\, Var(X_{t+2} - X_{t+2}^{(b)})}}$

Because $E(X_t - X_t^{(f)}) = E(X_{t+2}, X_{t+2}^{(b)}) = 0$, so $Var(X_t - X_t^{(f)}) = E(X_t - X_t^{(f)})^2$

$Var(X_{t+2} - X_{t+2}^{(b)}) = E(X_{t+2} - X_{t+2}^{(b)})^2$

$\therefore PACF(1) = \dfrac{cov(X_t - X_t^{(f)}, X_{t+2} - X_{t+2}^{(b)})}{\sqrt{E(X_t - X_t^{(f)})^2 \cdot E(X_{t+2} - X_{t+2}^{(b)})^2}}$

$= \dfrac{\gamma(2) - 2\phi\gamma(1) + \phi^2\gamma(0)}{(1+\phi^2)\gamma(0) - 2\phi\gamma(1)}$

(devided by $\gamma(0)$) $= \dfrac{\rho(2) - 2\phi\rho(1) + \phi^2}{1 + \phi^2 - 2\phi\rho(1)}$

```
knitr::opts_chunk$set(echo = TRUE)
data <- read.csv("C:/Users/Administrator/Desktop/sta137/GlobTempNASA_2020.csv")
newdata=data[-c(1,2,3,4),]
y<-newdata$December
tm<-newdata$Global.Land.and.Ocean.Temperature.Anomalies
trnd <- loess(y ~ tm, span = 0.3)
plot(tm, y, type = "l", lty = 1, xlab = "Time", ylab = "Year", main = "Global Land and Ocean Temperatur
points(tm, trnd$fitted,type = "l", lty = 2, col = "blue")
legend("topleft", "loess", lty = 2, col = "blue")
rough = trnd$residuals
plot(rough, type = "l", xlab = "Time", ylab = "Residuals", main = "Rough part")
#The rough part of the trend is symmetrically distributed
acf(rough)
#From the acf graph, there are six significance in the model
pacf(rough,main="PACF")
#From the pacf graph, there are five significance in the model,which mean we should use the ar(1) model
aic_table=rep(0,6)
for (i in 0:5){##you can check for higher orders
  aic_table[i+1]=arima(rough,order=c(i,0,0))$aic
}
aic_table
mod_ar5 <- arima(rough,order=c(5,0,0))
mod_ar5$coef
```

```r
mod_ar5$var.coef##variance covariance matrix, so select diagonals to get variance
sqrt(diag(mod_ar5$var.coef))##std.error = sqrt(variance)
acf(mod_ar5$residuals,main="ACF plot for residuals")
Box.test(mod_ar5$residuals, lag=10, type='Ljung-Box')

simMA1 <- arima.sim(n=350, model=list(ma=c(0.8)),sd = 2)
ts.plot(arima.sim(n=350, model=list(ma=c(0.8)),sd = 2))
acf(simMA1)
pacf(simMA1,main="PACF")
simMA2 <- arima.sim(n=350, model=list(ma=c(-08,0.8)),sd = 2)
ts.plot(arima.sim(n=350, model=list(ma=c(-0.8,0.8)),sd = 2))
acf(simMA2)
pacf(simMA2,main="PACF")
simAR2 <- arima.sim(n=350, model=list(ar=c(0.7,-0.7)),sd = 2)
ts.plot(arima.sim(n=350, model=list(ar=c(0.7,-0.7)),sd = 2))
acf(simAR2)
pacf(simAR2,main="PACF")
simAR2 <- arima.sim(n=350, model=list(ar=c(-0.3,0.4)),sd = 2)
ts.plot(arima.sim(n=350, model=list(ar=c(-0.3,0.4)),sd = 2))
acf(simAR2)
pacf(simAR2,main="PACF")
#alternatively
simARMA12 = arima.sim(model=list(ar = c(-0.8, -0.8), ma=0.6), n = 350,sd = 2)
ts.plot(simARMA12)
acf(simARMA12)
pacf(simARMA12,main="PACF")
```