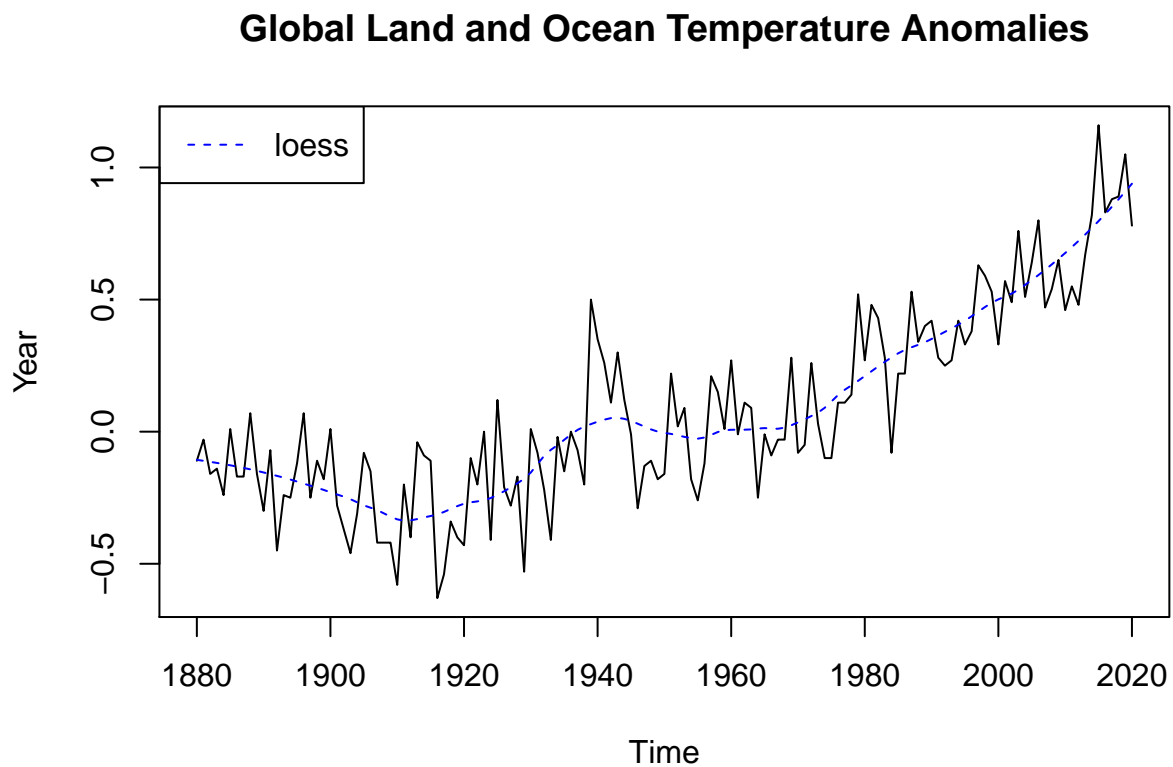


STA137HW4

Yirui Li, Chunqiu Li, Haoming Lei

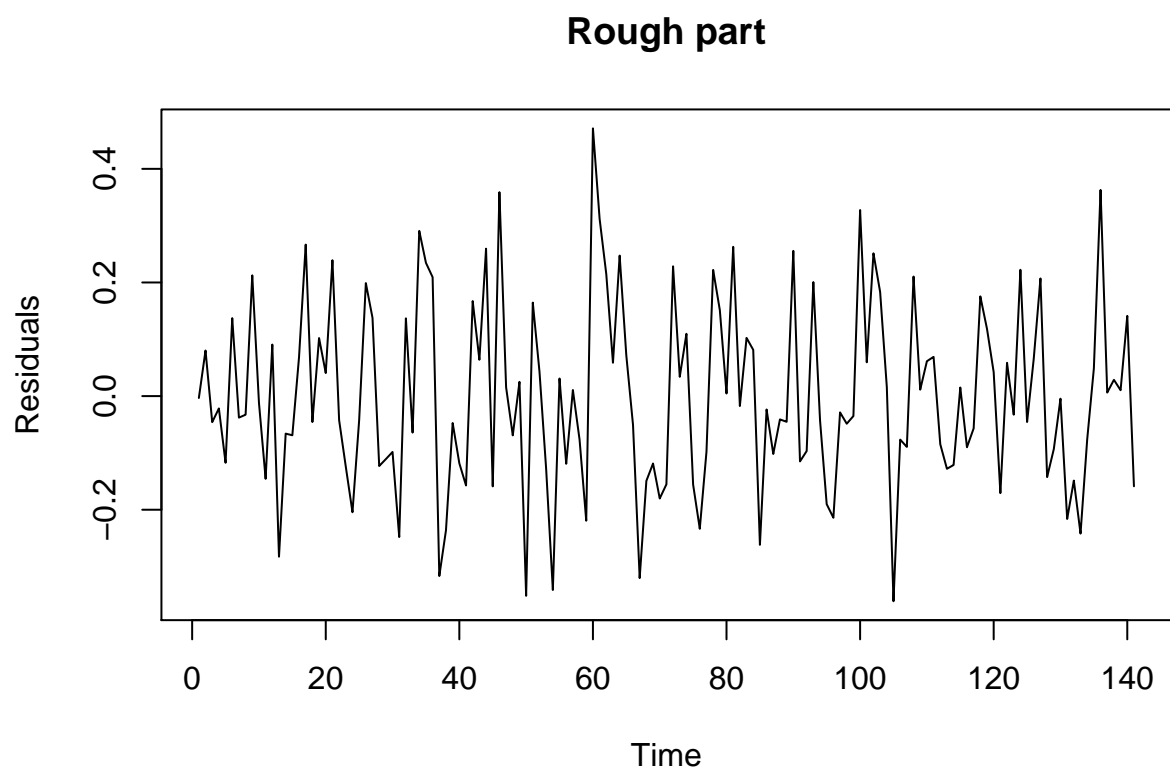
1a

```
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 Temperature Anomalies")
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")
```



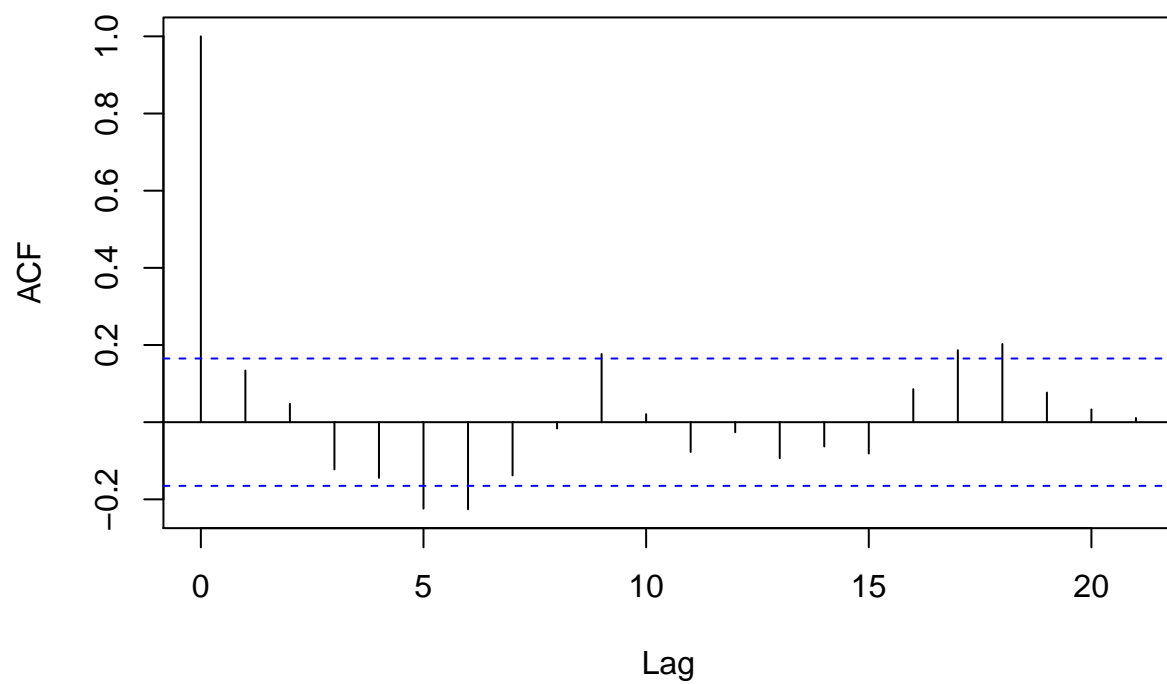
#The rough part of the trend is symmetrically distributed

1c.

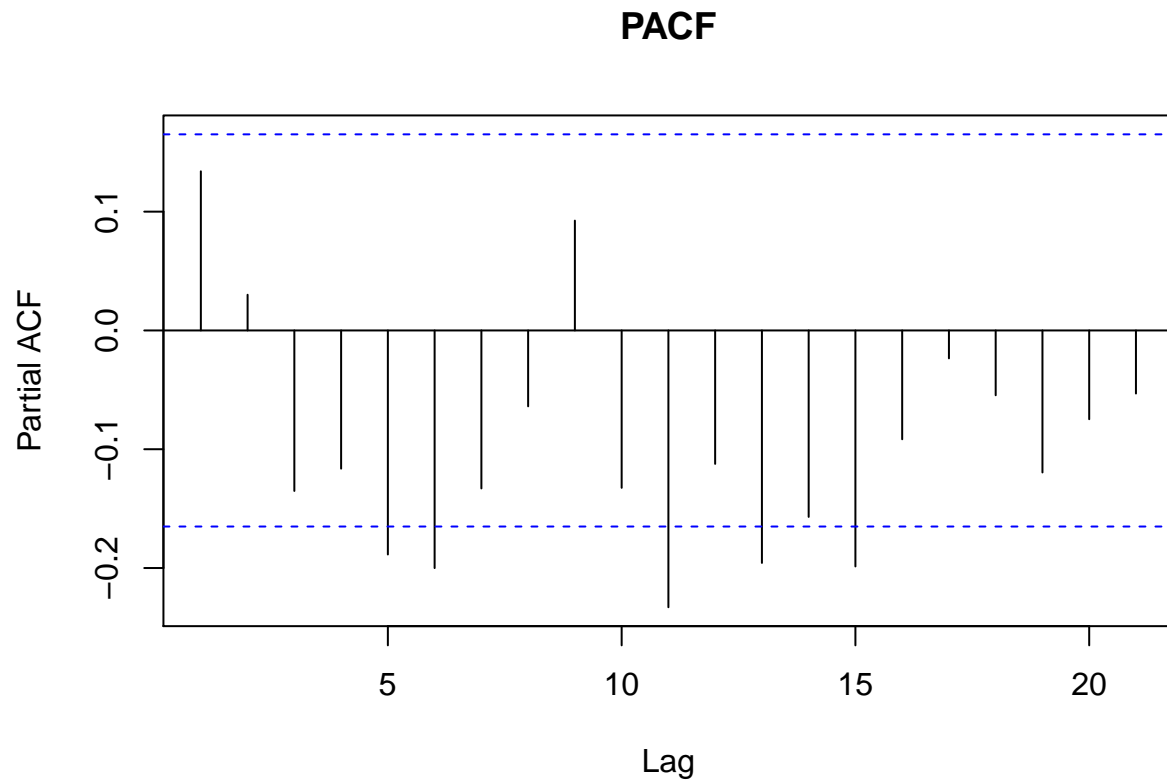
In the acf plot, there five $p(j)$ outside the bound, which show that this model is fairly fit to the data.

```
acf(rough)
```

Series 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

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?

##Also carry out the Ljung-Box test on the residuals at level $\alpha = 0.05$. Summarize your findings. p-value is larger than alpha level, so we cannot reject H_0 .

```
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
```

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

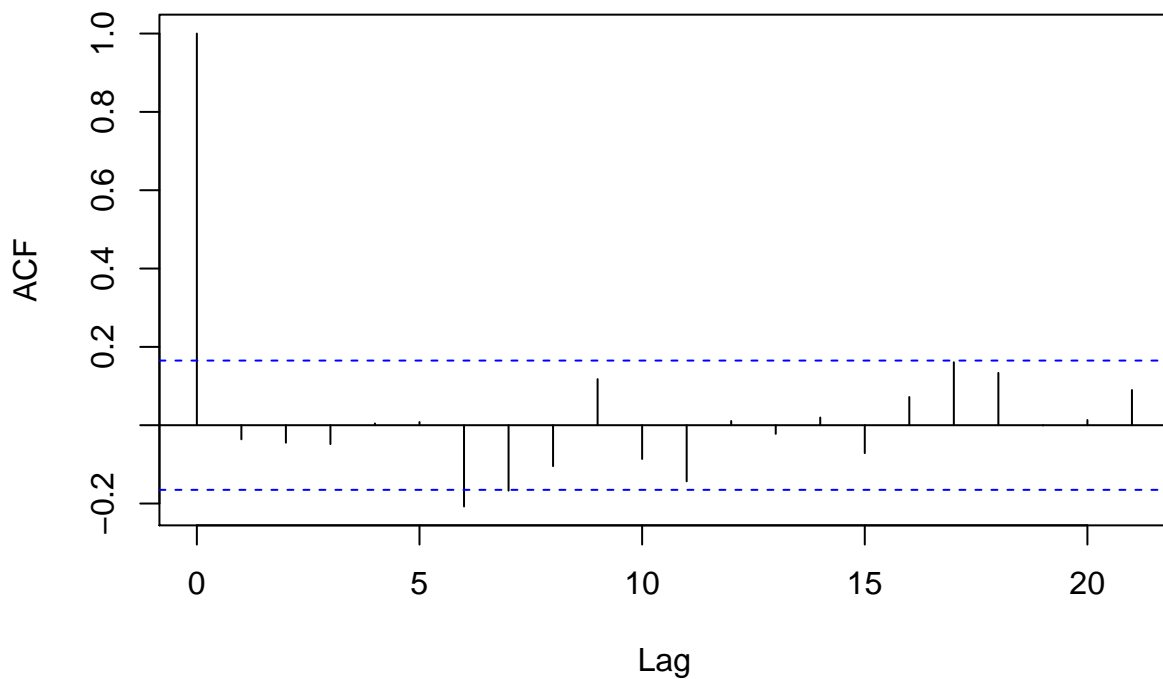
```
##          ar1          ar2          ar3          ar4          ar5
## ar1      6.832629e-03 -8.483368e-04 -3.491002e-04  8.677604e-04  8.009566e-04
## ar2     -8.483368e-04  6.864572e-03 -8.719928e-04 -4.255251e-04  8.088365e-04
## ar3     -3.491002e-04 -8.719928e-04  6.724257e-03 -9.141307e-04 -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
```

```
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
```

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

ACF plot for residuals

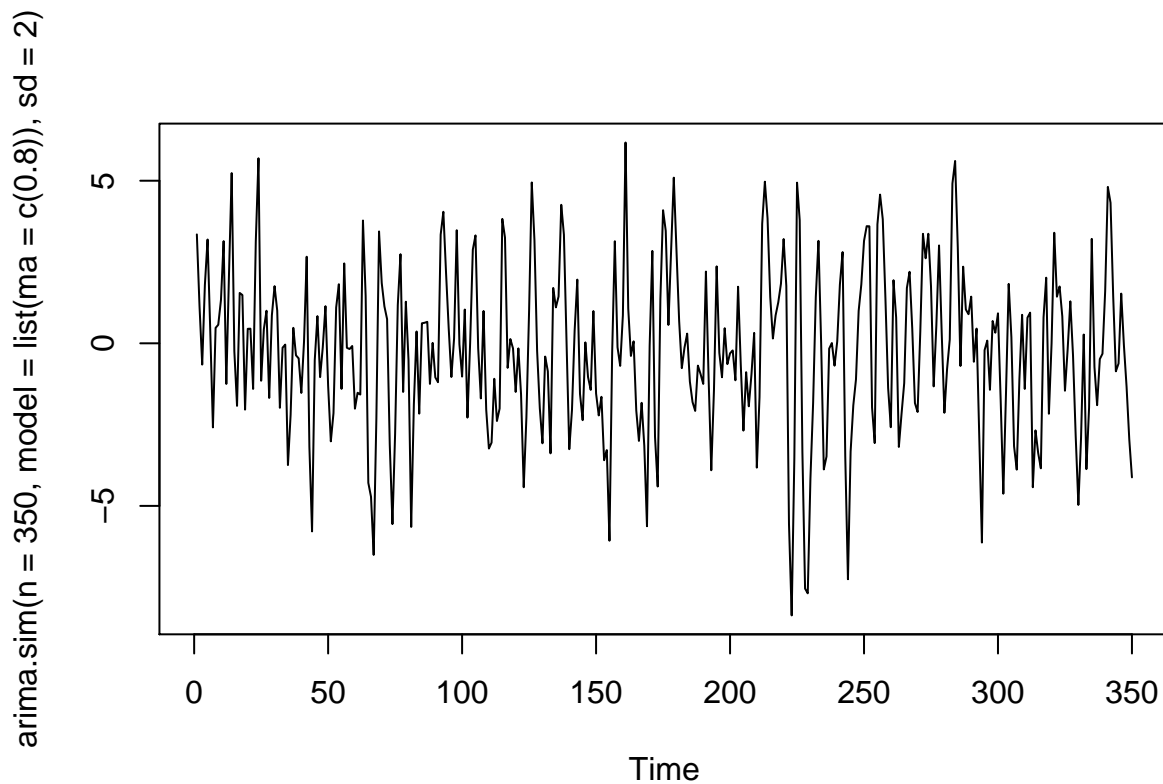


```
Box.test(mod_ar5$residuals, lag=10, type='Ljung-Box')
```

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

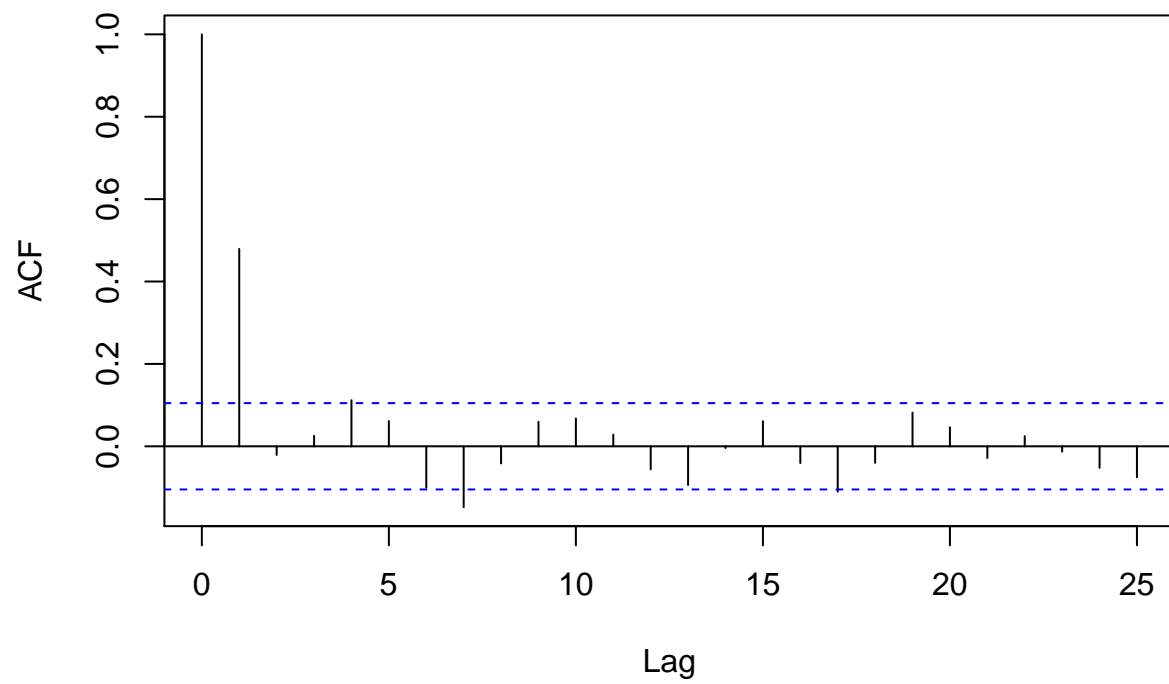
2a As expected, for the MA(1) case, the ACF plot shows that only $\hat{p}(1)$ is outside the $\pm 1.96/\sqrt{n}$ bars and the rest are inside, which shows that MA(1) is not a very appropriate model for this.

```
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))
```

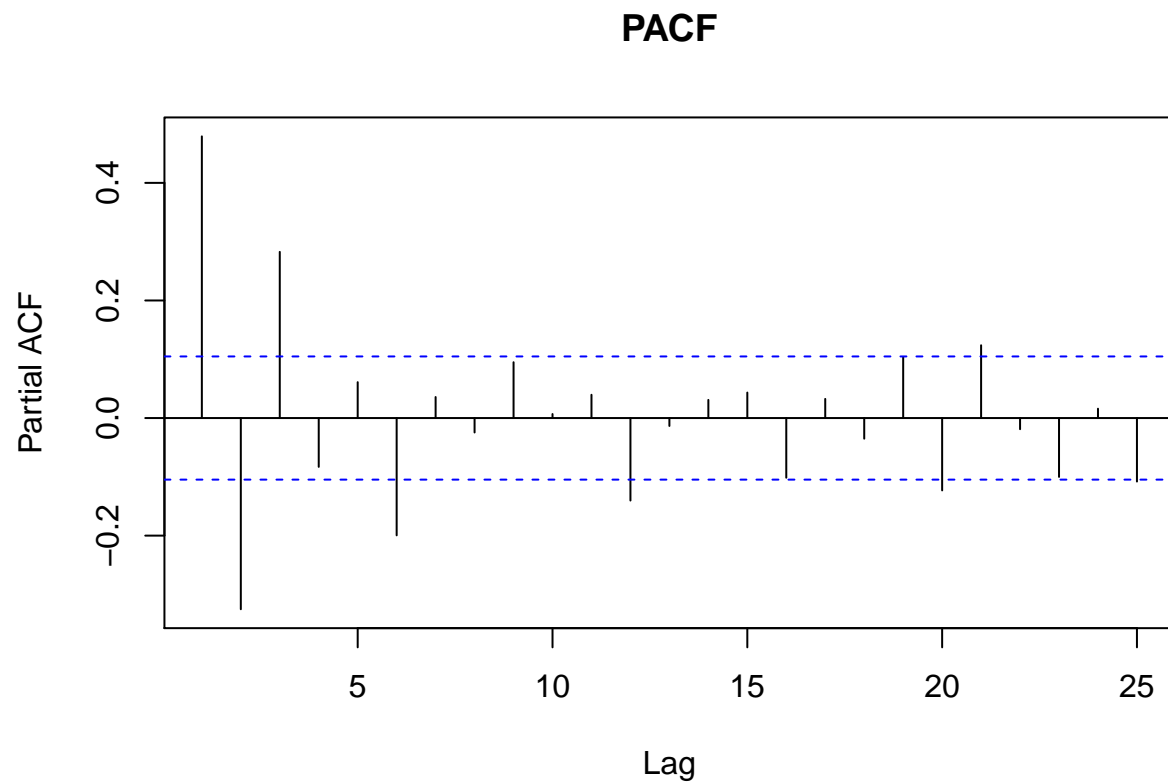


```
acf(simMA1)
```

Series simMA1



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

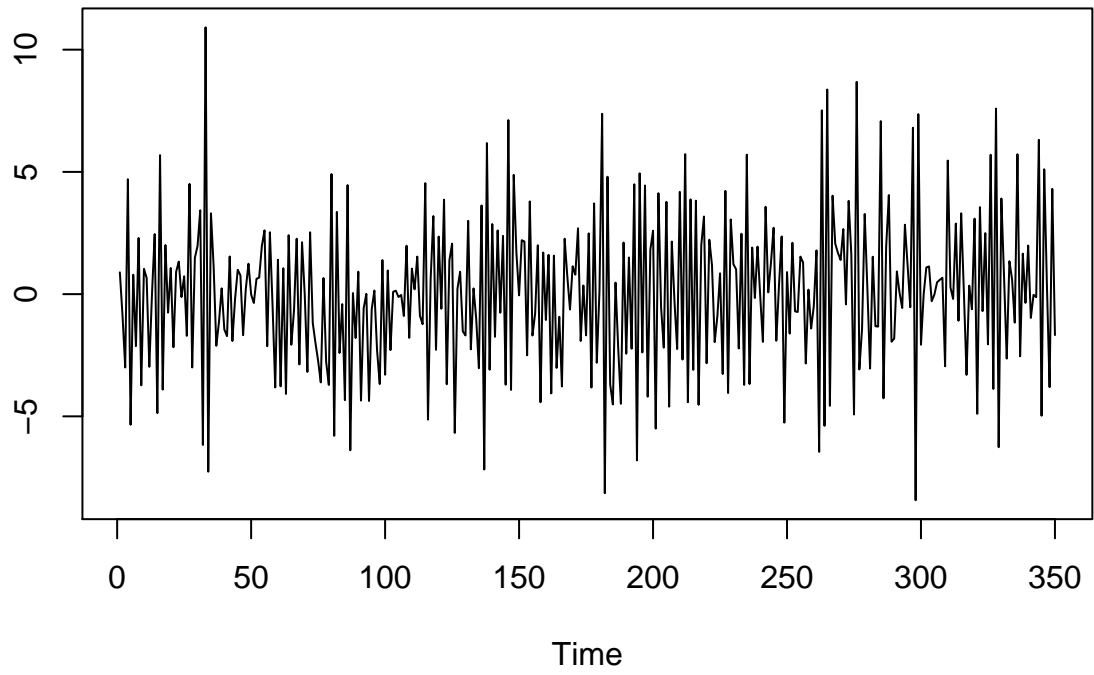


2b As expected, for the MA(2) case, the ACF plot shows that only $p^{\wedge}(1)$ is outside the $\pm 1.96/\sqrt{n}$ bars and the rest are inside, which shows that MA(2) is an appropriate model.

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

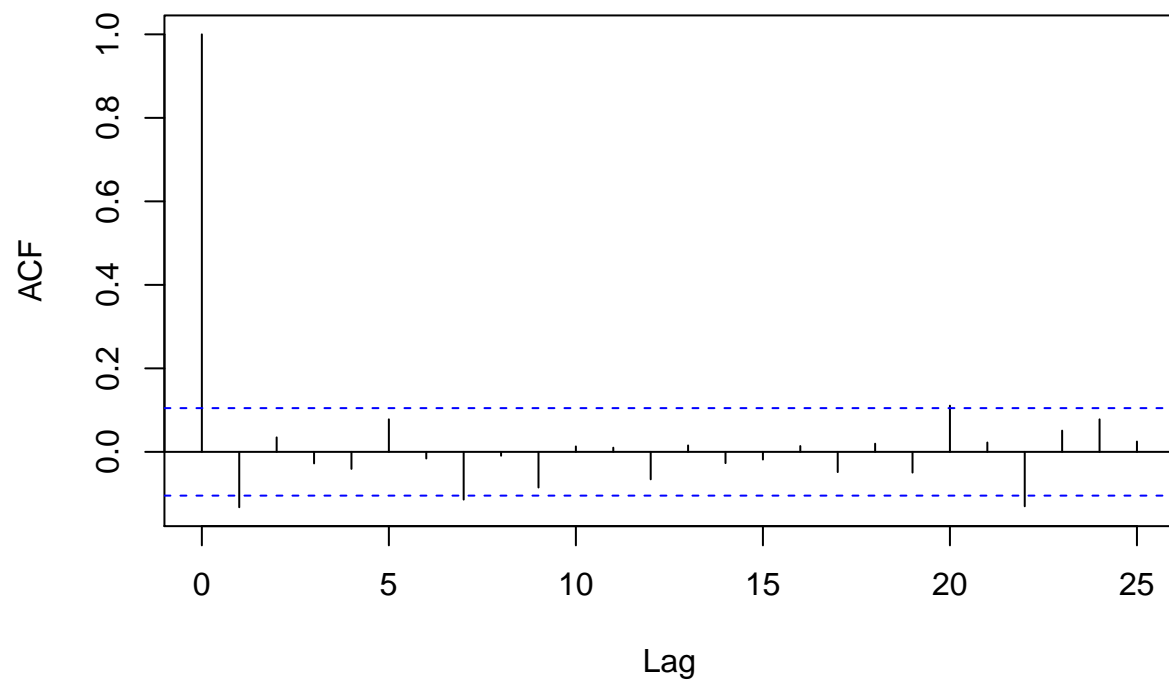


```
arima.sim(n = 350, model = list(ma = c(-0.8, 0.8)), sd = 2)
```

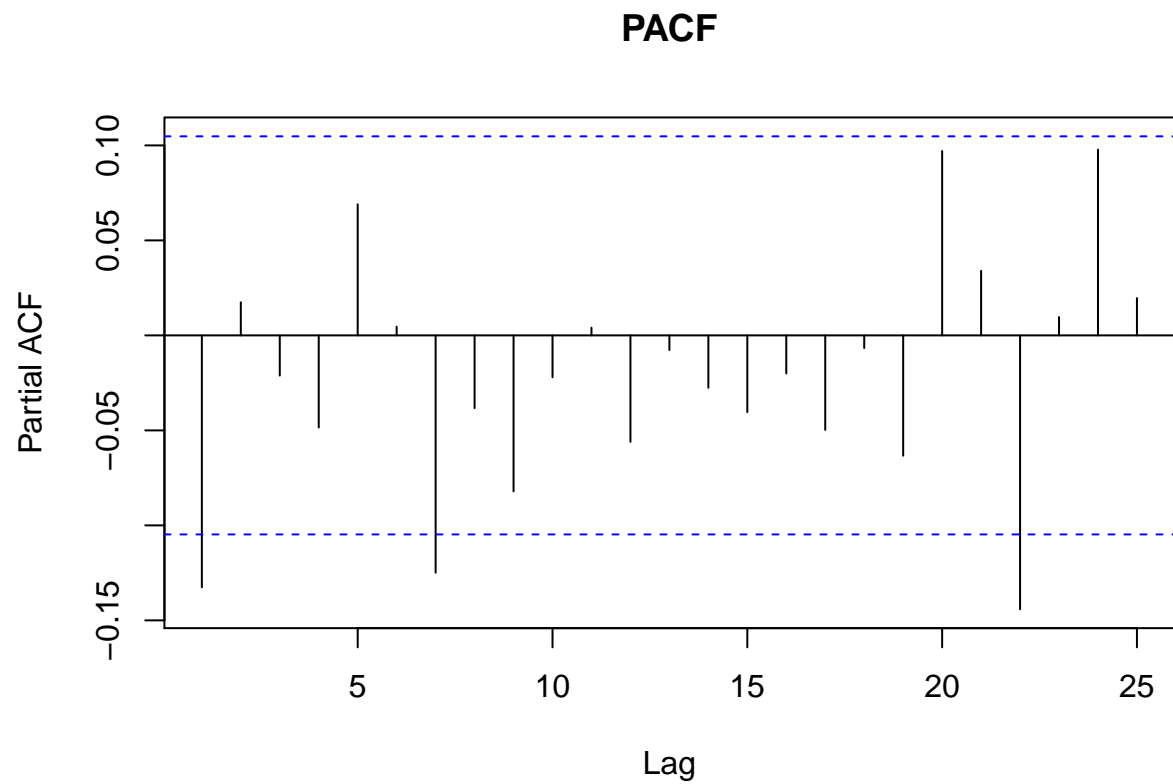


```
acf(simMA2)
```

Series simMA2



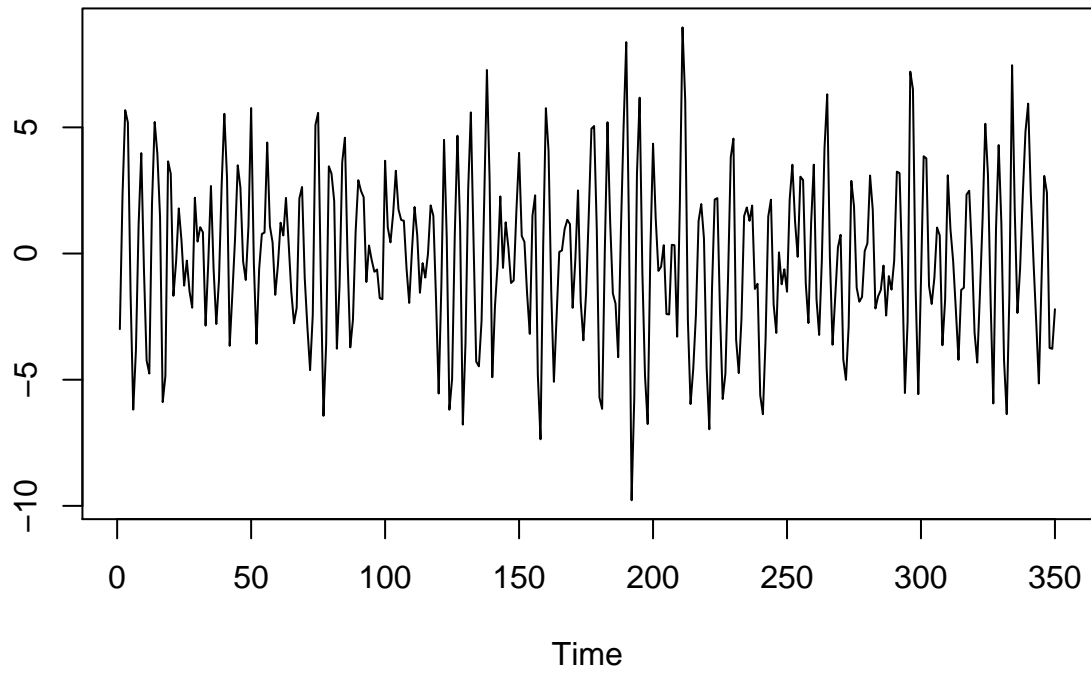
```
pacf(simMA2,main="PACF")
```



2c In this case, most $p(j)$ are outside the $\pm 1.96/\sqrt{n}$.

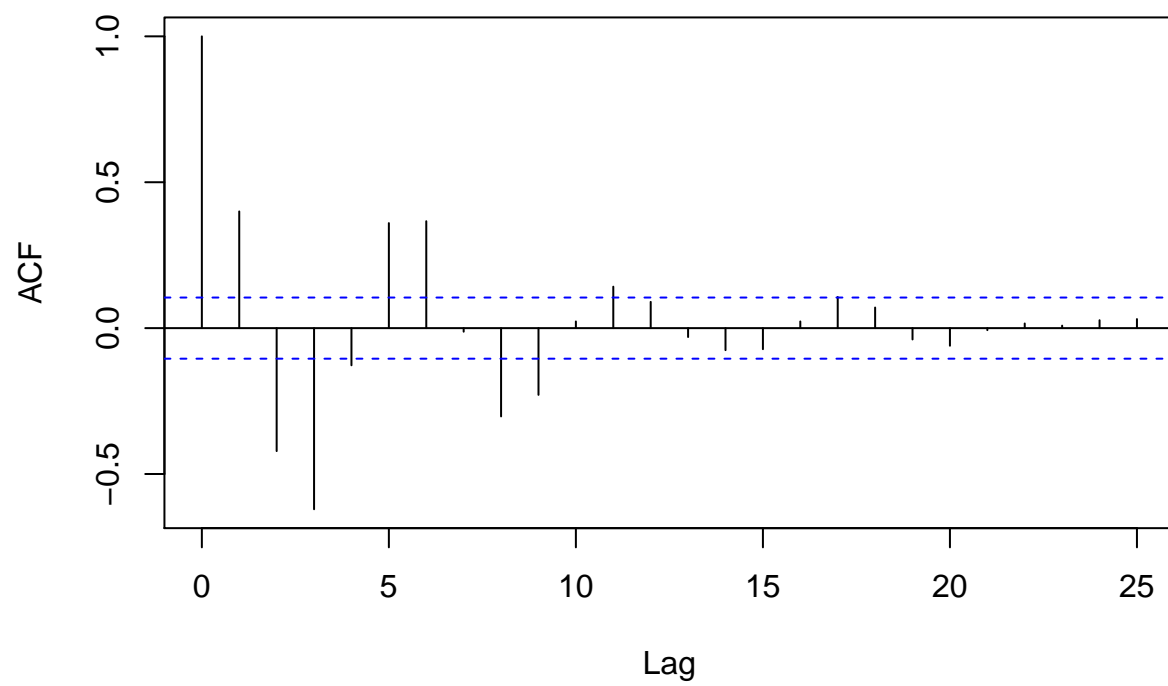
```
simAR2 <- arima.sim(n=350, model=list(ar=c(0.7,-0.7)),sd = 2)
ts.plot(arima.sim(n=350, model=list(ar=c(0.7,-0.7)),sd = 2))
```

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

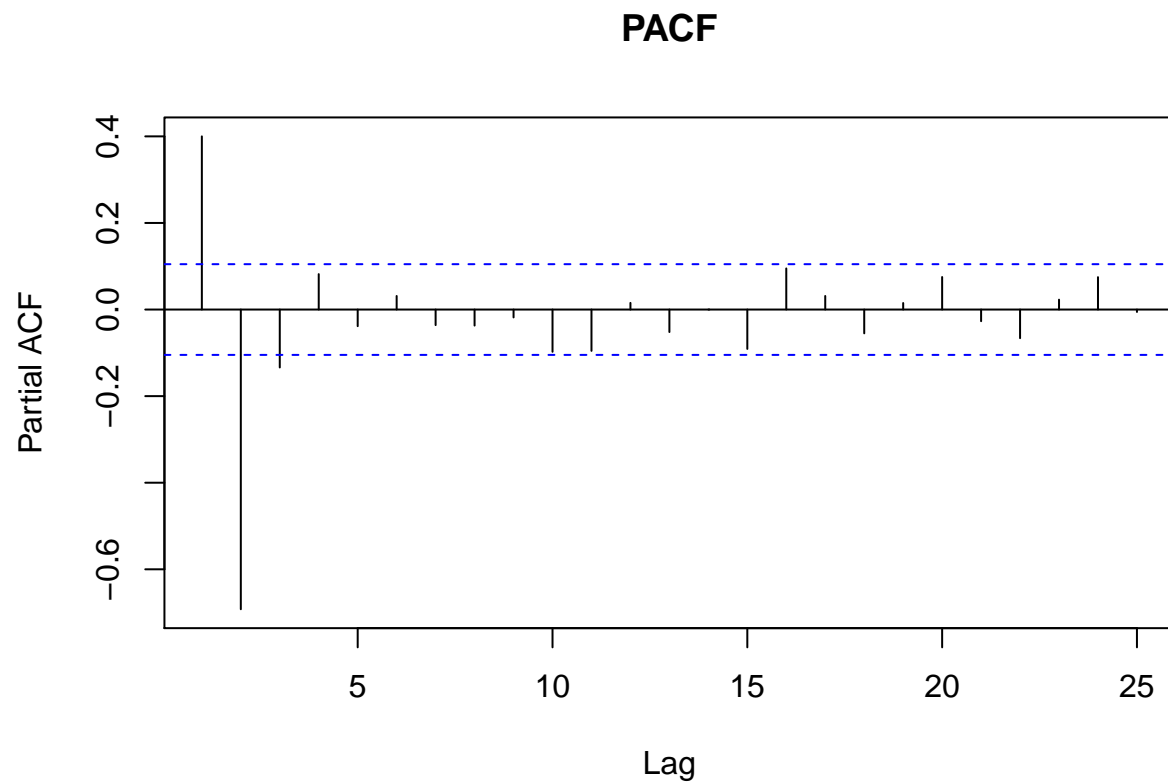


```
acf(simAR2)
```

Series simAR2



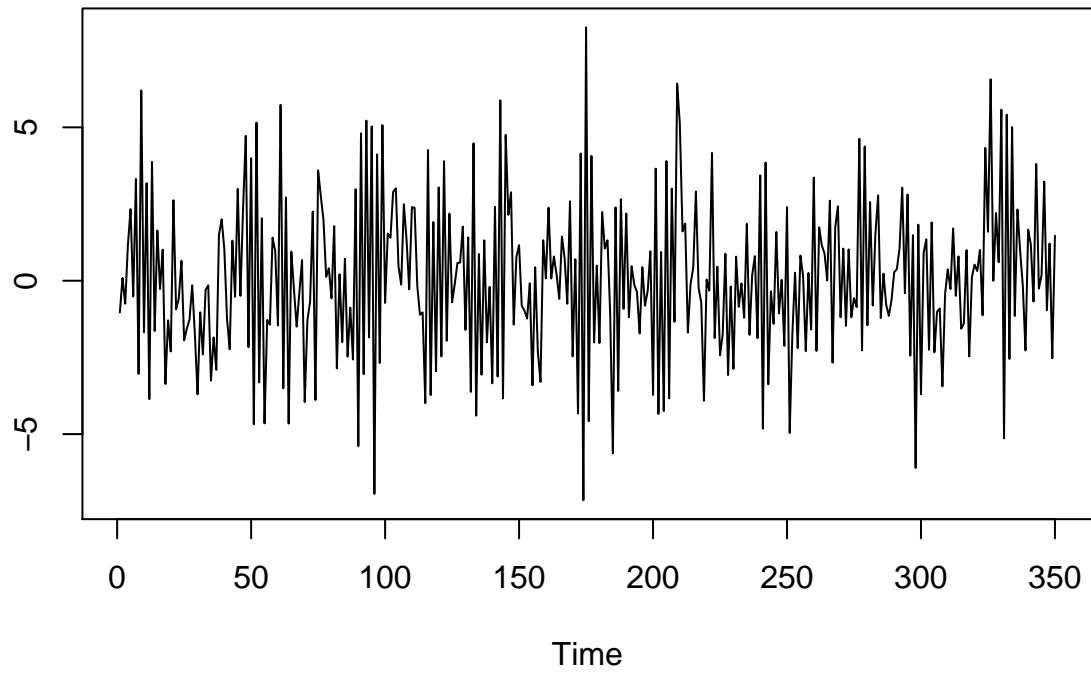
```
pacf(simAR2,main="PACF")
```



2d In this case, some of $p(j)$ are outside the $\pm 1.96/\sqrt{n}$.

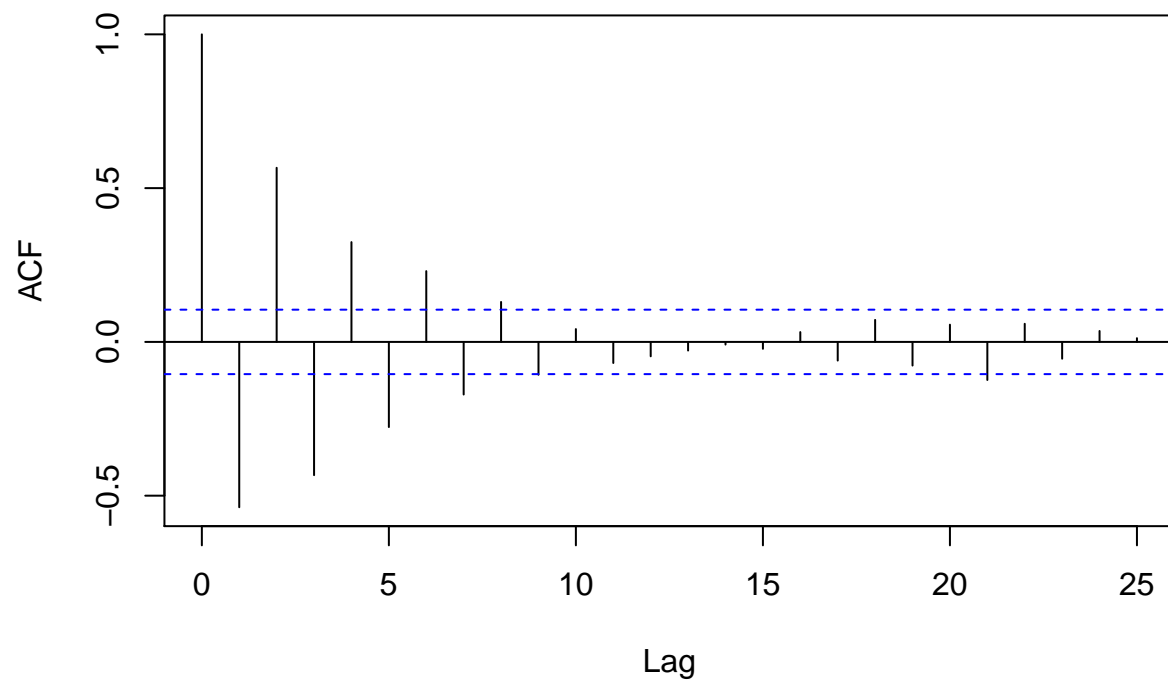
```
simAR2 <- arima.sim(n=350, model=list(ar=c(-0.3,0.4)),sd = 2)
ts.plot(arima.sim(n=350, model=list(ar=c(-0.3,0.4)),sd = 2))
```

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

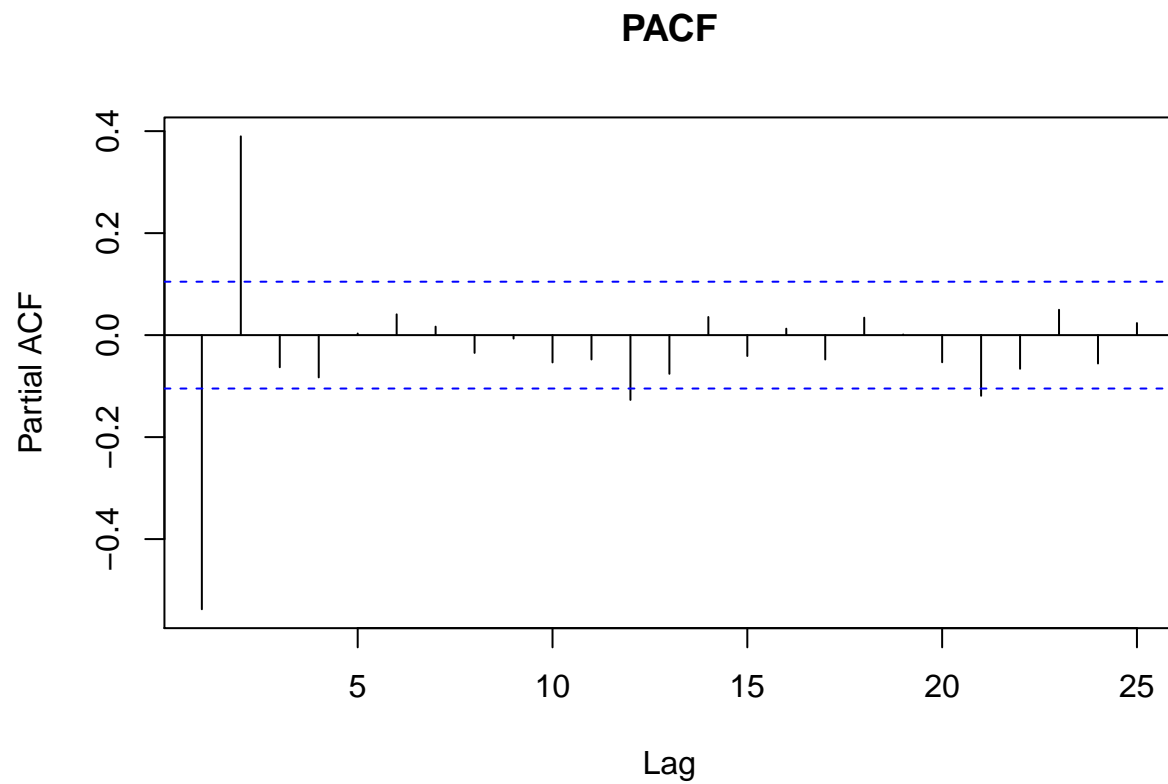


```
acf(simAR2)
```

Series simAR2



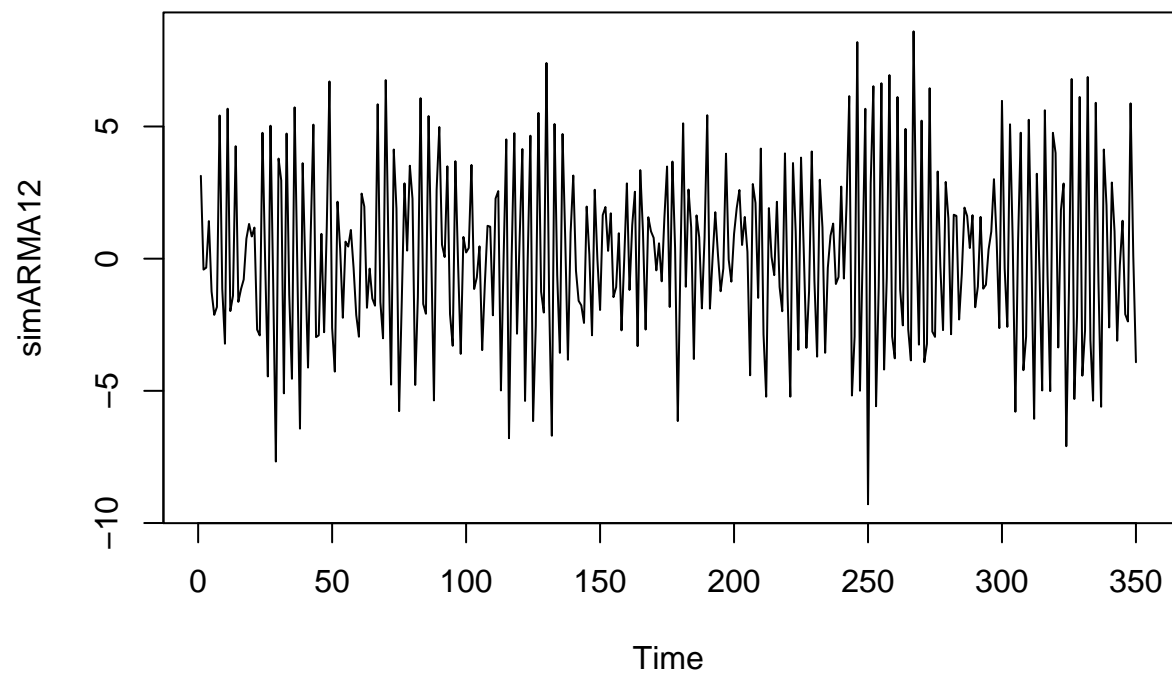
```
pacf(simAR2,main="PACF")
```

2e.

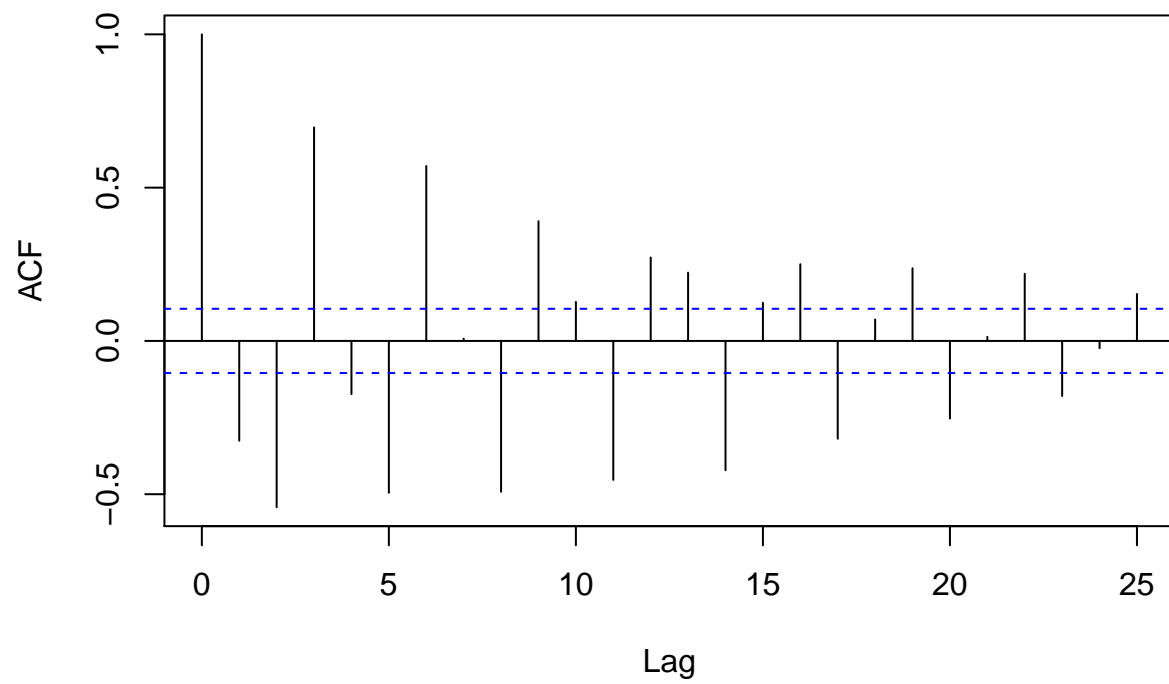
In this case, most of the $p(j)$ are outside the $\pm 1.96/\sqrt{n}$.

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

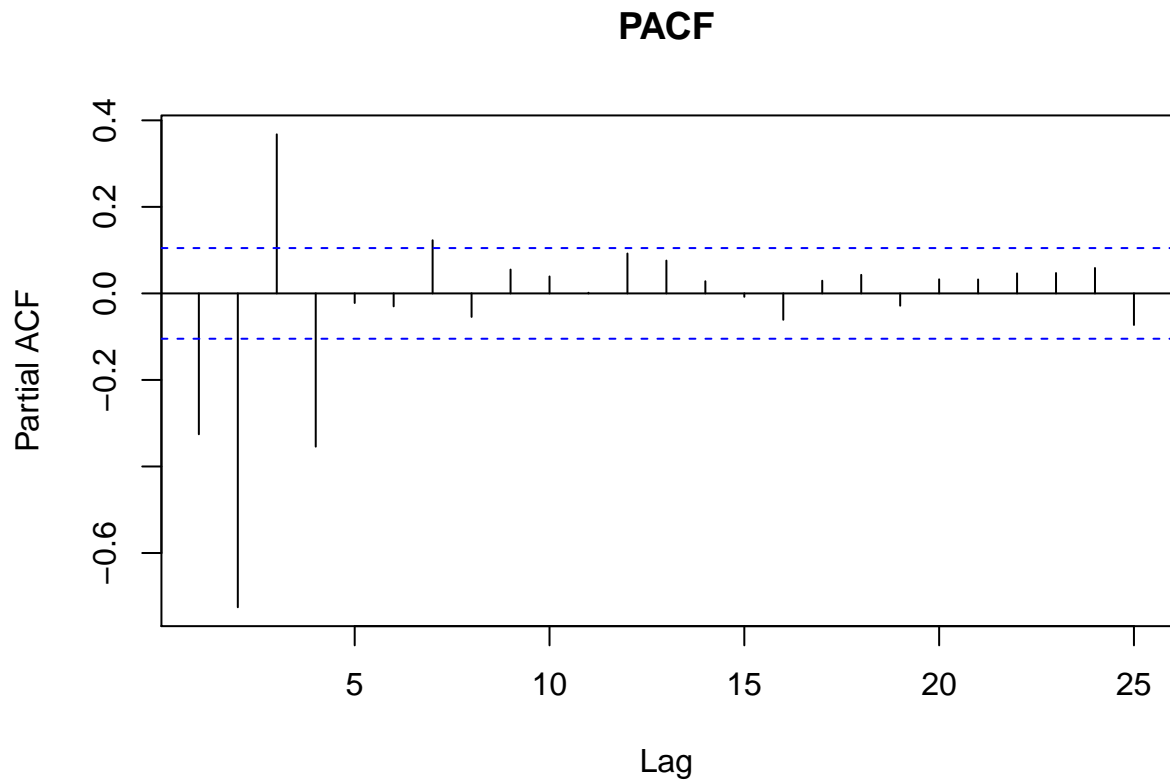


```
acf(simARMA12)
```

Series simARMA12



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



- 3.
- a). If X_t follows an AR(1) model, then.
- $$X_t - \mu = \phi(X_{t-1} - \mu) + \varepsilon_t \quad X_t = \mu - \phi X_{t-1} - \mu\phi + \varepsilon_t \quad X_t^f = (1-\phi)\mu - \phi X_{t-1}$$
- $$X_t - X_t^f = \varepsilon_t \quad \text{So } X_t^f \text{ is the fitted value when regressing } X_t \text{ on } X_{t-1}$$
- b) $X_{t+2} - \mu = \phi(X_{t+1} - \mu) + \varepsilon_{t+1} \quad X_{t+2} = \mu + \phi(X_{t+1} - \mu) + \varepsilon_{t+1}$
- $$X_{t+2}^b = \mu + \phi(X_{t+1} - \mu) \quad \text{So the fitted value } X_{t+2}^b \text{ is a linear function of } X_{t+1}$$

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

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

$$= \mu + \phi X_{t-1} - \phi \mu$$

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

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

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

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

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

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

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

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

$$\begin{aligned} \text{therefore, } E(X_t - X_t^{(1)})^2 &= (\gamma(0) + \mu^2) + \phi^2 (\gamma(0) + \mu^2) + \mu^2(1-\phi)^2 - 2\phi(\gamma(1) + \mu^2) + 2\phi \mu(1-\phi) \mu - 2\mu^2(1-\phi) \\ &= \gamma(0)(1+\phi^2) - 2\phi \gamma(1) + \mu^2(1+\phi^2 + 1 - 2\phi + 2\phi - 2\phi^2 - 2 + 2\phi) \\ &= (1+\phi^2)\gamma(0) - 2\phi \gamma(1) + \mu^2 \times 0 \\ &= (1+\phi^2)\gamma(0) - 2\phi \gamma(1) \end{aligned}$$

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

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

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

$$E(X_{t+2}^2) = \gamma(0) + \mu^2$$

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

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

$$\begin{aligned} \text{Therefore: } E(X_{t+2} - X_{t+2}^{(2)})^2 &= (\gamma(0) + \mu^2) + \phi^2 (\gamma(0) + \mu^2) + \mu^2(1-\phi)^2 - 2\phi(\gamma(1) + \mu^2) + \mu^2(1-\phi)^2 + 2\phi \mu(1-\phi) \mu - 2\phi \mu(1-\phi) \mu \\ &= \gamma(0) + \phi^2 \gamma(0) - 2\phi \gamma(1) + \mu^2 \times 0 \\ &= \gamma(0)(1+\phi^2) - 2\phi \gamma(1) \end{aligned}$$

e. Let $X_t - \mu = Z_t$

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

$$\begin{aligned} \text{Cov}(X_t - X_t^{(f)}, X_{t+2} - X_{t+2}^{(b)}) &= \text{Cov}(Z_t - \phi Z_{t-1}, Z_{t+2} - \phi Z_{t+1}) \\ &= \gamma(2) - \phi \gamma(1) - \phi \gamma(1) + \phi^2 \gamma(0) \\ &= \gamma(2) - 2\phi \gamma(1) + \phi^2 \gamma(0) \end{aligned}$$

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

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

$$\begin{aligned} \therefore \text{PACF}(1) &= \frac{\text{Cov}(X_t - X_t^{(f)}, X_{t+2} - X_{t+2}^{(b)})}{\sqrt{E(X_t - X_t^{(f)})^2 \cdot E(X_{t+2} - X_{t+2}^{(b)})^2}} \\ &= \frac{\gamma(2) - 2\phi \gamma(1) + \phi^2 \gamma(0)}{(1 + \phi^2) \gamma(0) - 2\phi \gamma(1)} \end{aligned}$$

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

```
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$Global.Land.and.Ocean.Temperature.Anomalies
trnd <- loess(y ~ tm, span = 0.3)
plot(tm, y, type = "l", lty = 1, xlab = "Time", ylab = "Year", main = "Global Land and Ocean Temperature")
points(tm, trnd$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
```

```

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(-0.8,0.8)),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")

```