ENV 790.30 - Time Series Analysis for Energy Data | Spring 2021

Assignment 5 - Due date 03/12/21

Benjamin Culberson

Directions

You should open the .rmd file corresponding to this assignment on RStudio. The file is available on our class repository on Github. And to do so you will need to fork our repository and link it to your RStudio.

Once you have the project open the first thing you will do is change "Student Name" on line 3 with your name. Then you will start working through the assignment by **creating code and output** that answer each question. Be sure to use this assignment document. Your report should contain the answer to each question and any plots/tables you obtained (when applicable).

When you have completed the assignment, **Knit** the text and code into a single PDF file. Rename the pdf file such that it includes your first and last name (e.g., "LuanaLima_TSA_A05_Sp21.Rmd"). Submit this pdf using Sakai.

Questions

This assignment has general questions about ARIMA Models.

Packages needed for this assignment: "forecast","tseries". Do not forget to load them before running your script, since they are NOT default packages.\

```
#Load/install required package here
library(forecast)
```

```
## Registered S3 method overwritten by 'quantmod':
##   method            from
##   as.zoo.data.frame zoo
```

```
library(tseries)
library(stats)
library(sarima)
```

```
## Loading required package: FitAR

## Loading required package: lattice

## Loading required package: leaps

## Loading required package: ltsa

## Loading required package: bestglm
```

```
##
## Attaching package: 'FitAR'

## The following object is masked from 'package:forecast':
##
##     BoxCox

## Loading required package: stats4
```

Q1

Describe the important characteristics of the sample autocorrelation function (ACF) plot and the partial sample autocorrelation function (PACF) plot for the following models:

(a) AR(2)

> Answer: The ACF of an AR(2) model will decay over time. The PACF of this model will cut off at exaclty lag = 2 (identifying the order of the model).

(b) MA(1)

> Answer: The ACF of an MA(1) model cut off at lag = 1 (identifying the order of the model) and the PACF will decay over time.

Q2

Recall that the non-seasonal ARIMA is described by three parameters ARIMA(p, d, q) where p is the order of the autoregressive component, d is the number of times the series need to be differenced to obtain stationarity and q is the order of the moving average component. If we don't need to difference the series, we don't need to specify the "I" part and we can use the short version, i.e., the ARMA(p, q). Consider three models: ARMA(1,0), ARMA(0,1) and ARMA(1,1) with parameters $\phi = 0.6$ and $\theta = 0.9$. The ϕ refers to the AR coefficient and the θ refers to the MA coefficient. Use R to generate $n = 100$ observations from each of these three models

```r
sim1a<-arima.sim(list(order = c(1,0,0), ar = 0.6), n = 100)
sim2a<-arima.sim(list(order = c(0,0,1), ma = 0.9), n = 100)
sim3a<-arima.sim(list(order = c(1,0,1), ar = 0.6, ma = 0.9), n = 100)
```

(a) Plot the sample ACF for each of these models in one window to facilitate comparison (Hint: use command $par(mfrow = c(1, 3))$ that divides the plotting window in three columns).

(b) Plot the sample PACF for each of these models in one window to facilitate comparison.

(c) Look at the ACFs and PACFs. Imagine you had these plots for a data set and you were asked to identify the model, i.e., is it AR, MA or ARMA and the order of each component. Would you be identify them correctly? Explain your answer.

> Answer:

The first model shows somewhat of a slow decay in its ACF and a clear cut off in its PACF. I believe I would be able to identify it as an AR process with an order of (1,0) given its clear cutoff at lag = 1 in the PACF.

The second model seems to show a cutoff at lag = 1 in the ACF but its PACF does not have super obvious slow decay. I might have trouble correctly identifying this one. I would probably guess it's an

MA process based off the ACF and the somewhat slow decay in the PACF, but I would not be certain. I would therefore guess that it has an order of (0,1).

The last model possibly shows a gradual decay in the ACF and possibly a slow decay in the PACF. Neither of them shows a clear cutoff at any point so I would identify this as an ARMA model with an order of (1,1).

(d) Compare the ACF and PACF values R computed with the theoretical values you provided for the coefficients. Do they match? Explain your answer.

> Answer:

The theoretical value $\phi = 0.6$ appears to loosely match the PACF of the AR model. The value of the AR model's PACF at lag $= 1$ sits at around 0.5, so it's not a perfect match but it's close. The MA model's coefficient ($\theta = 0.9$) cannot be matched on the either the ACF or the PACF graphs so I cannot tell if it comes close to the theoretical coefficient. The PACF of the ARMA model has a value at lag $= 1$ of somewhat close to 0.75 so it does not appear to match the theoretical value of the coefficient.

(e) Increase number of observations to $n = 1000$ and repeat parts (a)-(d).

(f) Look at the ACFs and PACFs. Imagine you had these plots for a data set and you were asked to identify the model, i.e., is it AR, MA or ARMA and the order of each component. Would you be identify them correctly? Explain your answer.

> Answer:

The first model shows a clear slow decay in its ACF and a clear cut off in its PACF at lag $= 0$. I believe I would be able to identify it as an AR process with an order of (1,0) given its clear cutoff at lag $= 0$ in the PACF and its obvious decay in the ACF.

The second model shows a clear cutoff at lag $= 0$ in the ACF and its PACF has an obvious slow decay. I would therefore be able to identify this model as an MA process with an order of (0,1).

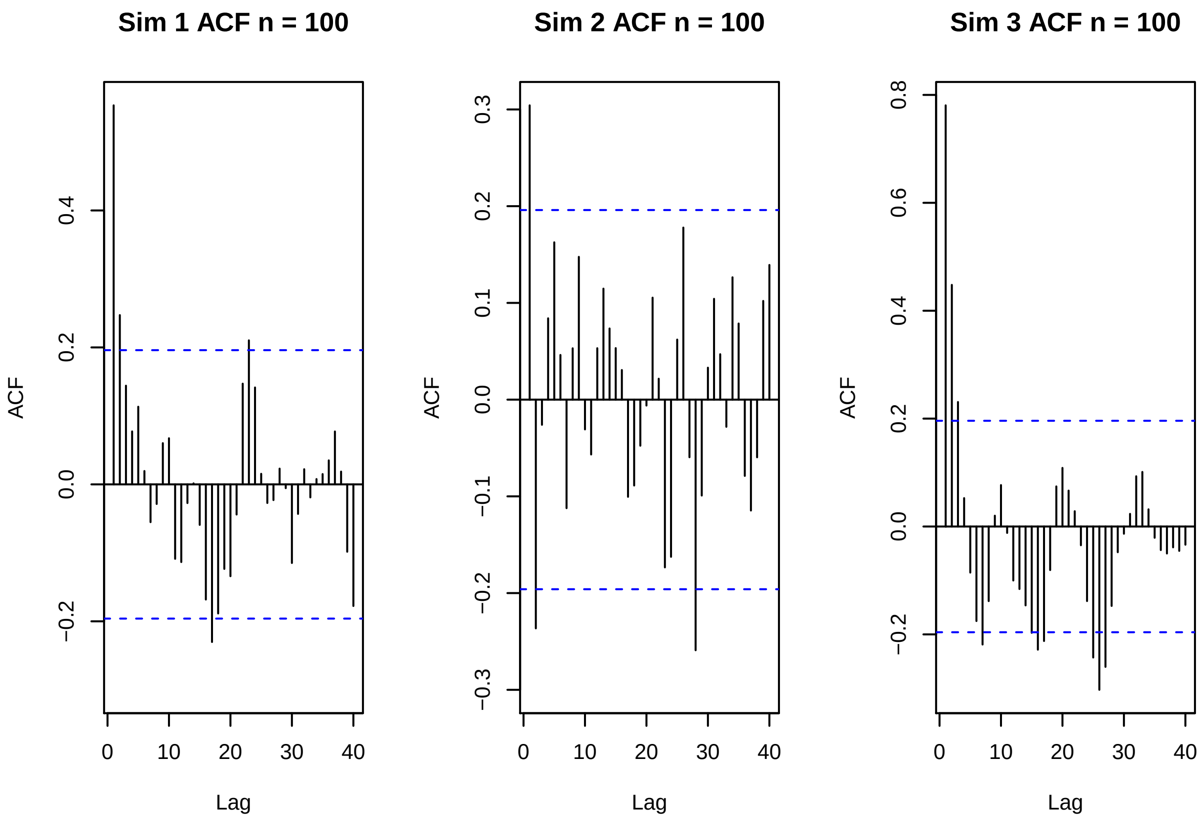
The last model shows an obvious gradual decay in the ACF and an obvious gradual decay in the PACF. Based on these plots, I would be able to determine that this is an ARMA model, and I would put the order at (1,1).

(g) Compare the ACF and PACF values R computed with the theoretical values you provided for the coefficients. Do they match? Explain your answer.

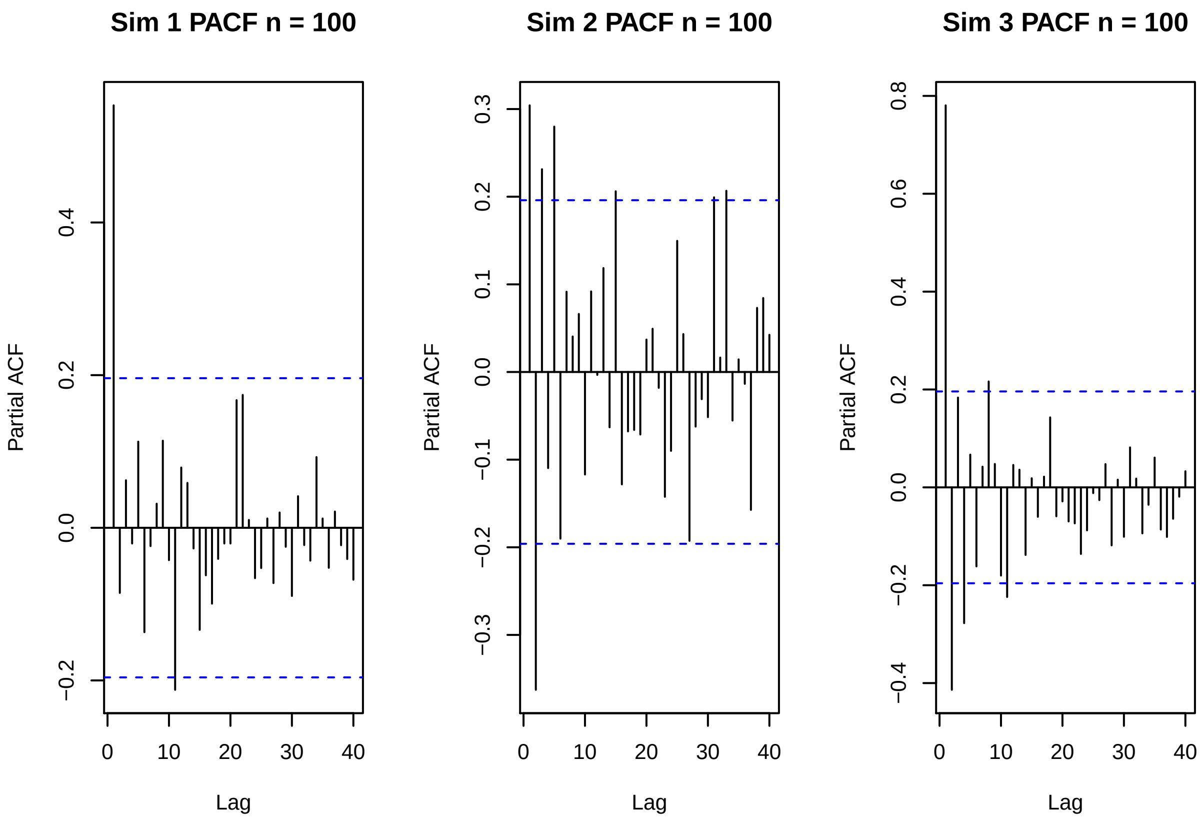
> Answer:

The theoretical value $\phi = 0.6$ appears to somewhat closely match the PACF of the AR model. The value of the AR model's PACF at lag $= 1$ sits at around 0.6, so it's a reasonable match to the theoretical value. The MA model's coefficient ($\theta = 0.9$) cannot be matched on the either the ACF or the PACF graphs so I cannot tell if it comes close to the theoretical coefficient. The PACF of the ARMA model has a value at lag $= 1$ of roughly 0.8 so it does not match the theoretical coefficient very well.

```
par(mfrow=c(1,3))  #place plot side by side
  Acf(sim1a,lag.max=40,main=paste("Sim 1 ACF n = 100"))
  Acf(sim2a,lag.max=40,main=paste("Sim 2 ACF n = 100"))
  Acf(sim3a,lag.max=40,main=paste("Sim 3 ACF n = 100"))
```

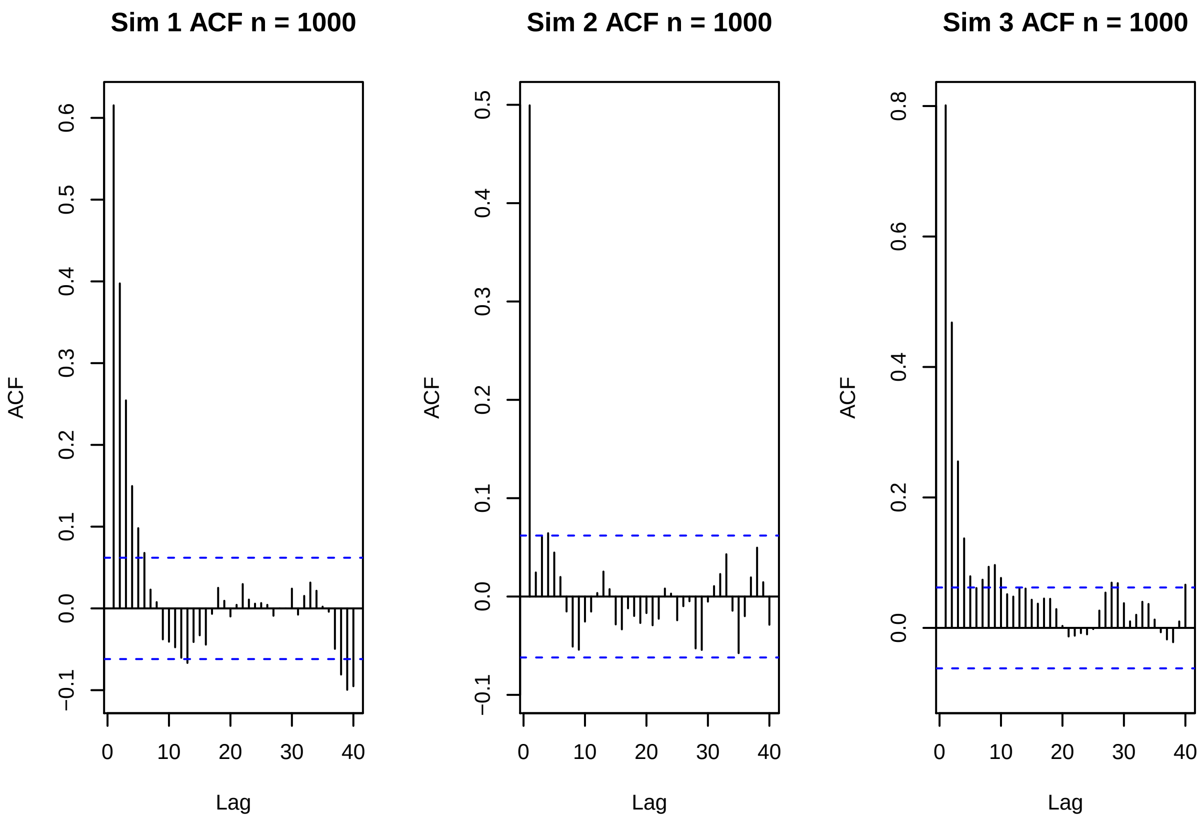



```r
par(mfrow=c(1,3))  #place plot side by side
  pacf(sim1a,lag.max=40,main=paste("Sim 1 PACF n = 100"))
  pacf(sim2a,lag.max=40,main=paste("Sim 2 PACF n = 100"))
  pacf(sim3a,lag.max=40,main=paste("Sim 3 PACF n = 100"))
```

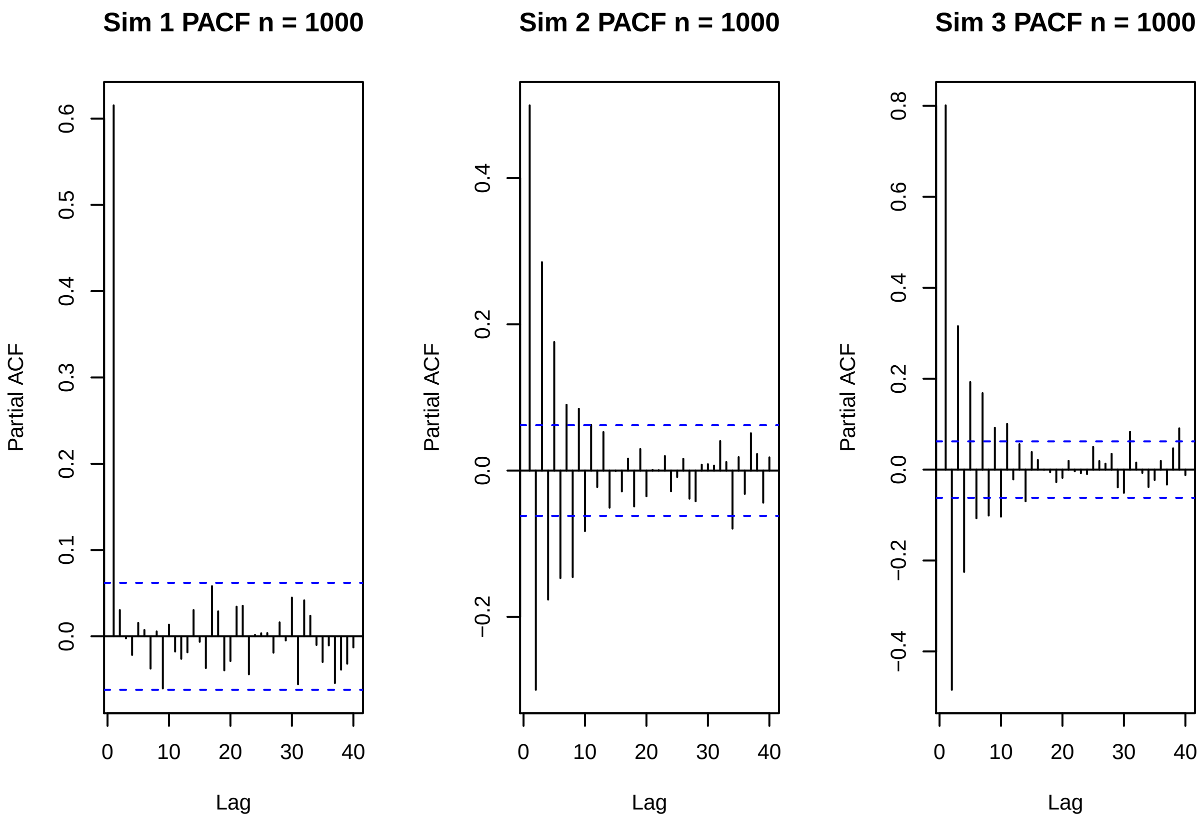



```
sim1<-arima.sim(list(order = c(1,0,0), ar = 0.6), n = 1000)
sim2<-arima.sim(list(order = c(0,0,1), ma = 0.9), n = 1000)
sim3<-arima.sim(list(order = c(1,0,1), ar = 0.6, ma = 0.9), n = 1000)

par(mfrow=c(1,3))   #place plot side by side
  Acf(sim1,lag.max=40,main=paste("Sim 1 ACF n = 1000"))
  Acf(sim2,lag.max=40,main=paste("Sim 2 ACF n = 1000"))
  Acf(sim3,lag.max=40,main=paste("Sim 3 ACF n = 1000"))
```




```r
par(mfrow=c(1,3))  #place plot side by side
  pacf(sim1,lag.max=40,main=paste("Sim 1 PACF n = 1000"))
  pacf(sim2,lag.max=40,main=paste("Sim 2 PACF n = 1000"))
  pacf(sim3,lag.max=40,main=paste("Sim 3 PACF n = 1000"))
```



##

Q3

Consider the ARIMA model $y_t = 0.7 * y_{t-1} - 0.25 * y_{t-12} + a_t - 0.1 * a_{t-1}$

(a) Identify the model using the notation ARIMA$(p, d, q)(P, D, Q)_s$, i.e., identify the integers p, d, q, P, D, Q, s (if possible) from the equation.

Based on the equation above, this model is ARIMA(1,0,1)(1,0,0). p = 1, d = 0, q = 1, P = 1, D = 0, Q = 0, s = 12. Because we only use t - 1 and t - 12, we neither seasonally nor non-seasonally difference the equation.

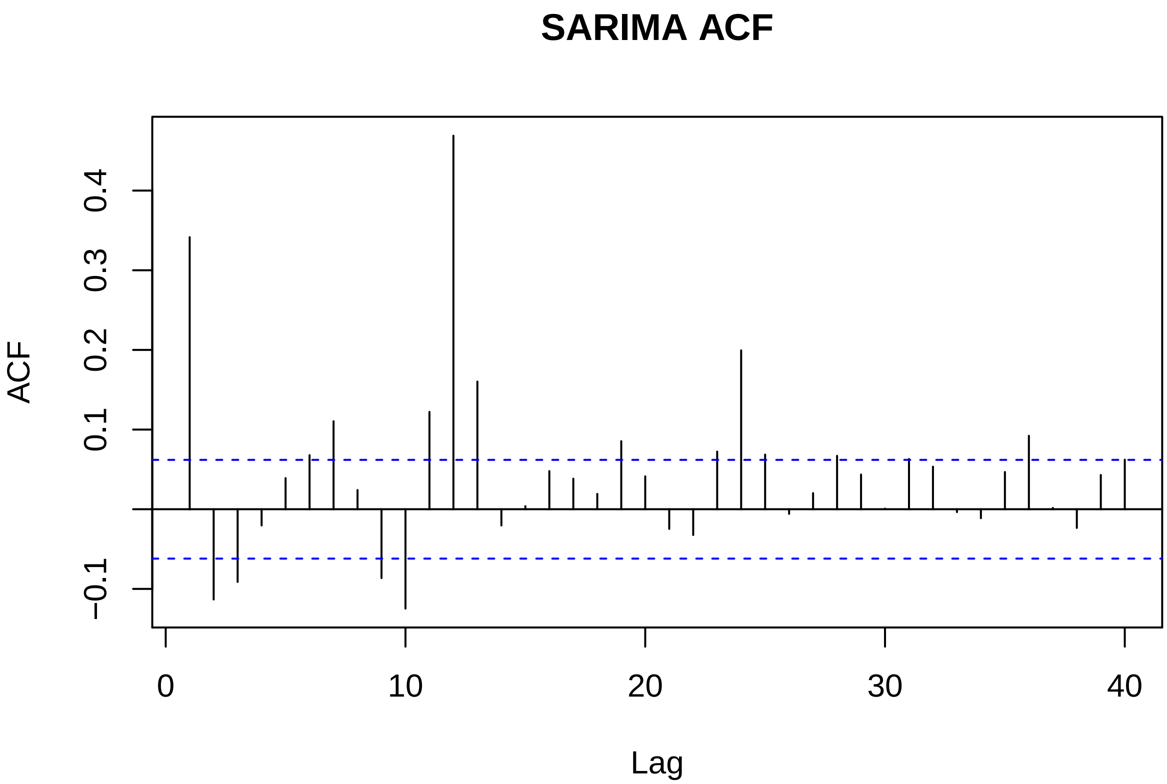
(b) Also from the equation what are the values of the parameters, i.e., model coefficients.

From the equation, the model coefficients are $\phi1 = 0.7$, $\phi12 = -0.25$, $\theta1 = -0.1$, $\theta12 = 0$

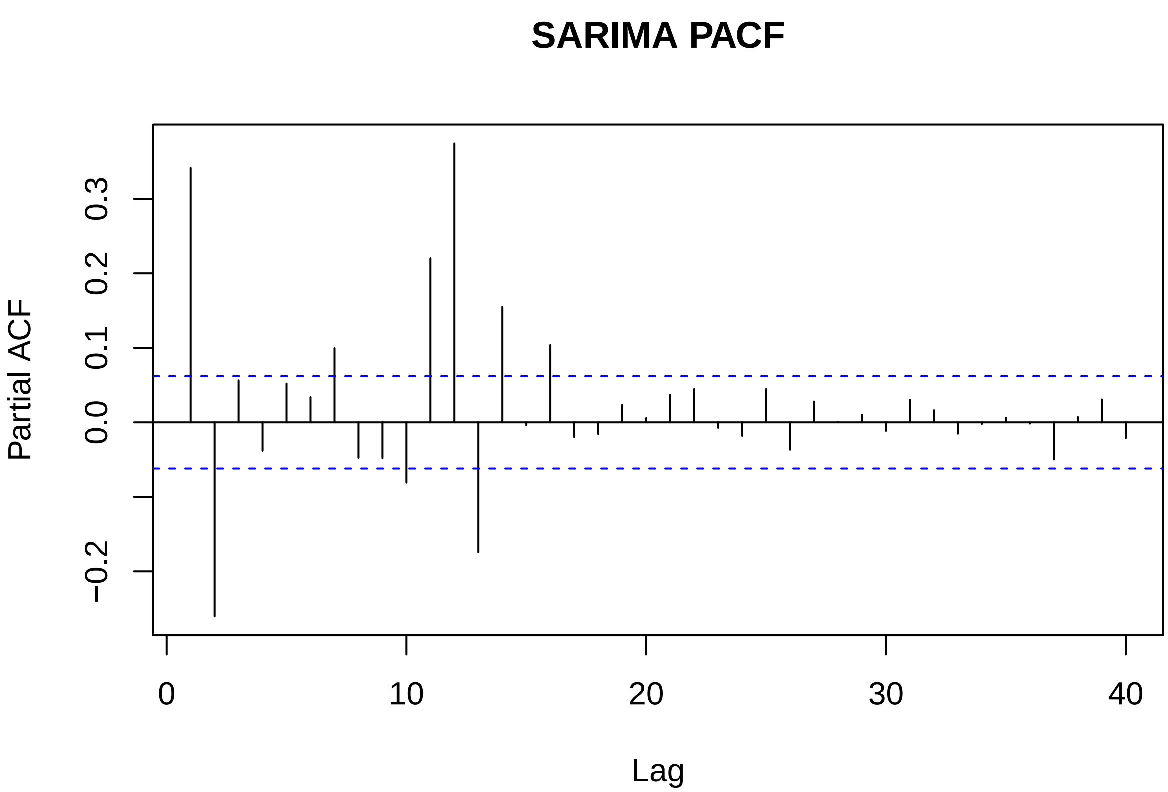
Q4

Plot the ACF and PACF of a seasonal ARIMA$(0, 1) \times (1, 0)_{12}$ model with $\phi = 0.8$ and $\theta = 0.5$ using R. The 12 after the bracket tells you that $s = 12$, i.e., the seasonal lag is 12, suggesting monthly data whose behavior is repeated every 12 months. You can generate as many observations as you like. Note the Integrated part was omitted. It means the series do not need differencing, therefore $d = D = 0$. Plot ACF and PACF for the simulated data. Comment if the plots are well representing the model you simulated, i.e., would you be able to identify the order of both non-seasonal and seasonal components from the plots? Explain.

```
sarima_model<-sim_sarima(n=1000, model = list(ar=0, ma=0.5, sar=0.5, sma=0, nseasons=12))
Acf(sarima_model,lag.max=40,main=paste("SARIMA ACF"))
```




```r
Pacf(sarima_model,lag.max=40,main=paste("SARIMA PACF"))
```



I would be able to tell that this has an order of 1 for the non-seasonal moving average component from the ACF which cuts off at lag = 1. I would also be able to tell that the seasonal autoregressive component has an order of one from the ACF which has a spike at lag = 12, 24, and possibly 36. Furthermore, the lack of a slow decay in the ACF tells me that the non-seasonal AR component has an order of zero and the lack of spikes at the lag = 12, 24, 36 spots in the PACF tell me this model has a seasonal MA component of zero.