ENV 790.30 - Time Series Analysis for Energy Data | Spring 2023

Assignment 6 - Due date 03/06/23

John Rooney

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 file open on your local machine the first thing you will do is rename the file such that it includes your first and last name (e.g., "LuanaLima_TSA_A06_Sp23.Rmd"). Then change "Student Name" on line 4 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. Submit this pdf using Sakai.

R packages needed for this assignment: "xlsx" or "readxl", "ggplot2", "forecast","tseries", and "Kendall". Install these packages, if you haven't done yet. Do not forget to load them before running your script, since they are NOT default packages.

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(tseries)
```

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

```
library(forecast)
library(sarima)
```

```
## Loading required package: stats4
```

```
##
## Attaching package: 'sarima'
```

```
## The following object is masked from 'package:stats':
##
##     spectrum
```

Q1

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

- AR(2)

  Answer: For an AR(2) plot, I would expect to see 2 lags on the PACF plot. I would expect to see exponential decay on the ACF plot.

- MA(1)

  Answer: For an MA(1) I would expect to see exponential decay on the PACF plot and 1 lag on the ACF plot before lags become insignificant.

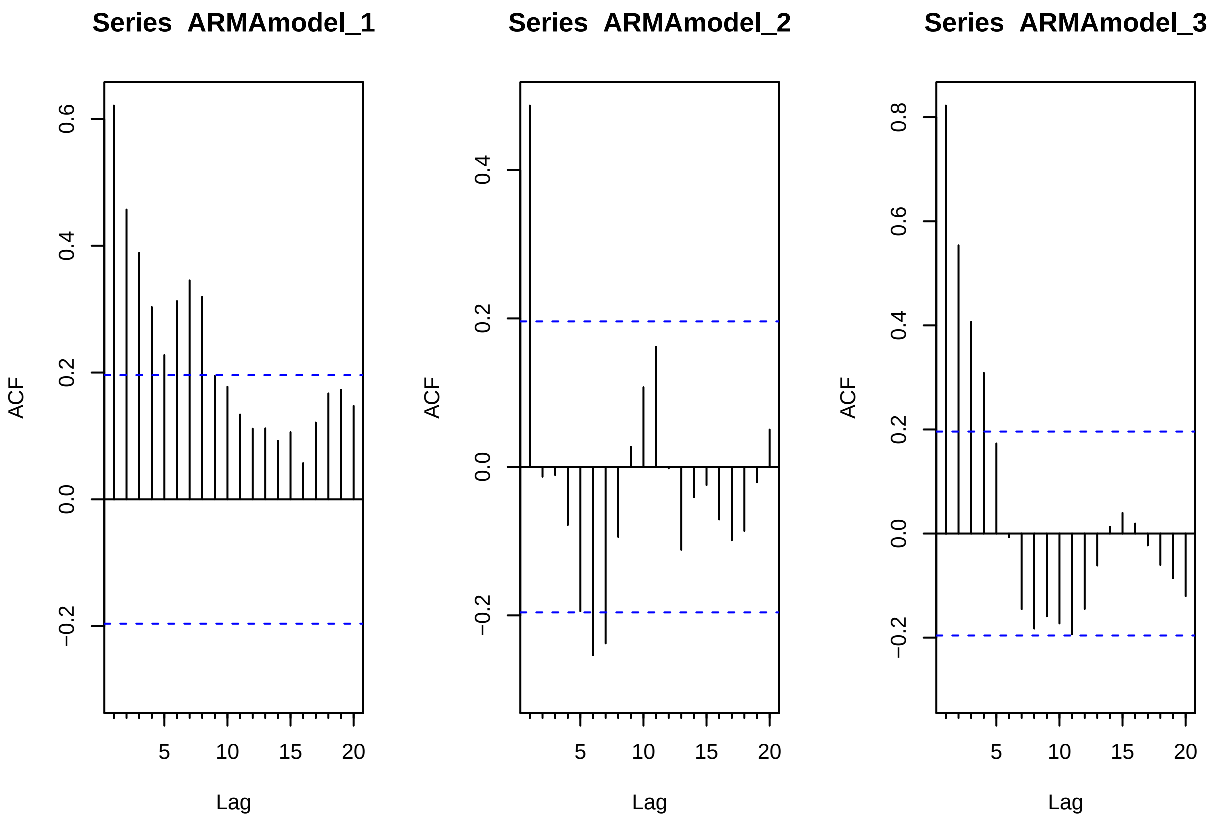
Q2

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
#ARMA(1,0)

ARMAmodel_1<- arima.sim(model=list(ar=0.6), n=100) #the AR coefficient is 0.6

#ARMA(0,1)

ARMAmodel_2<- arima.sim(model=list(ma=0.9), n=100) #the MA coefficient is 0.9

#ARMA(1,1)
ARMAmodel_3<- arima.sim(model=list(ar=0.6, ma=0.9), n=100)
```

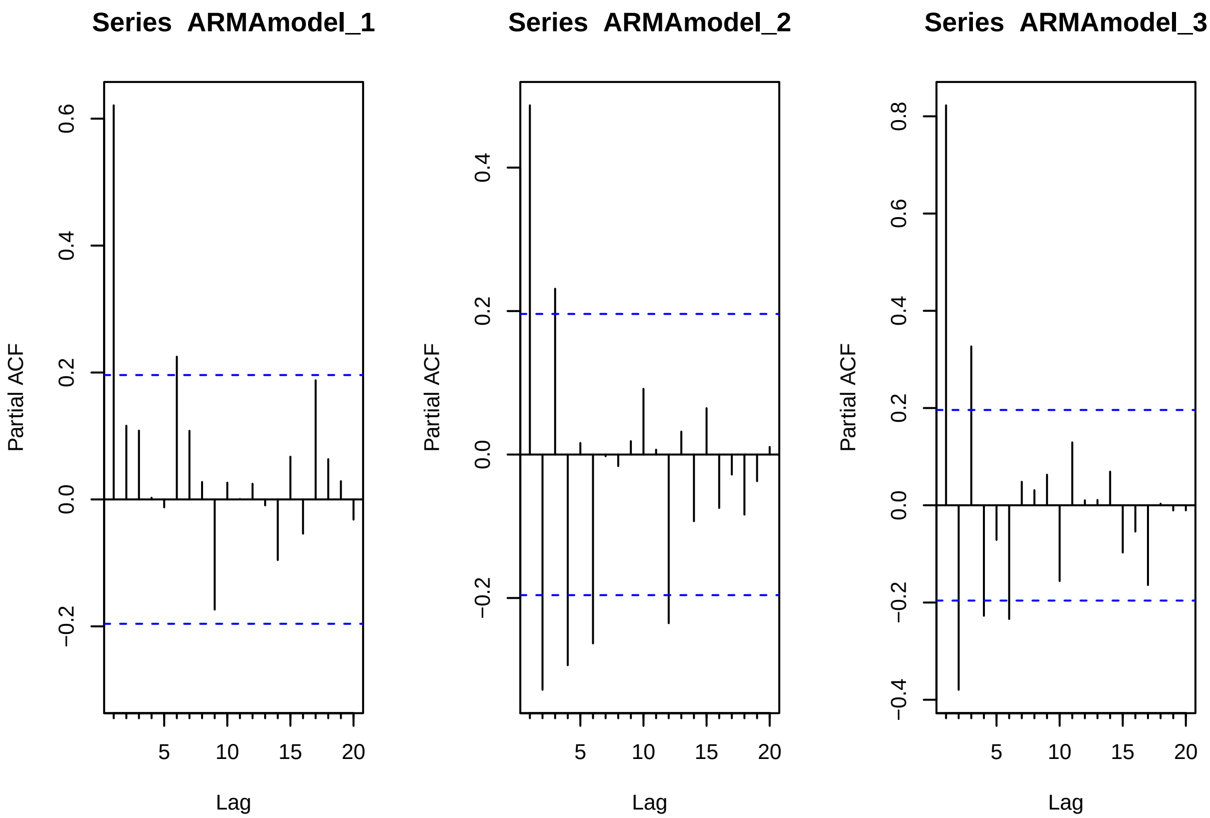
(a) Plot the sample ACF for each of these models in one window to facilitate comparison (Hint: use command $par(mfrow = c(1, 3))$ that divides the plotting window in three columns).

```r
par(mfrow=c(1,3))
Acf(ARMAmodel_1)
Acf(ARMAmodel_2)
Acf(ARMAmodel_3)
```



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

```
par(mfrow=c(1,3))
Pacf(ARMAmodel_1)
Pacf(ARMAmodel_2)
Pacf(ARMAmodel_3)
```



(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: For the first ARMA(1,0) plot, I think I would have likely struggled to identify the slow decay from the ACF plot but would have noticed the quick jump in the PACF and been able to identify the order. For the second ARMA(0,1) plot, I think I would have looked at the ACF and guessed a second order rather than first due to the second line being above the significance line. I would have hopefully picked up the decay in the PACF. For the third ARMA (1,1) plot, I would've guessed higher orders.

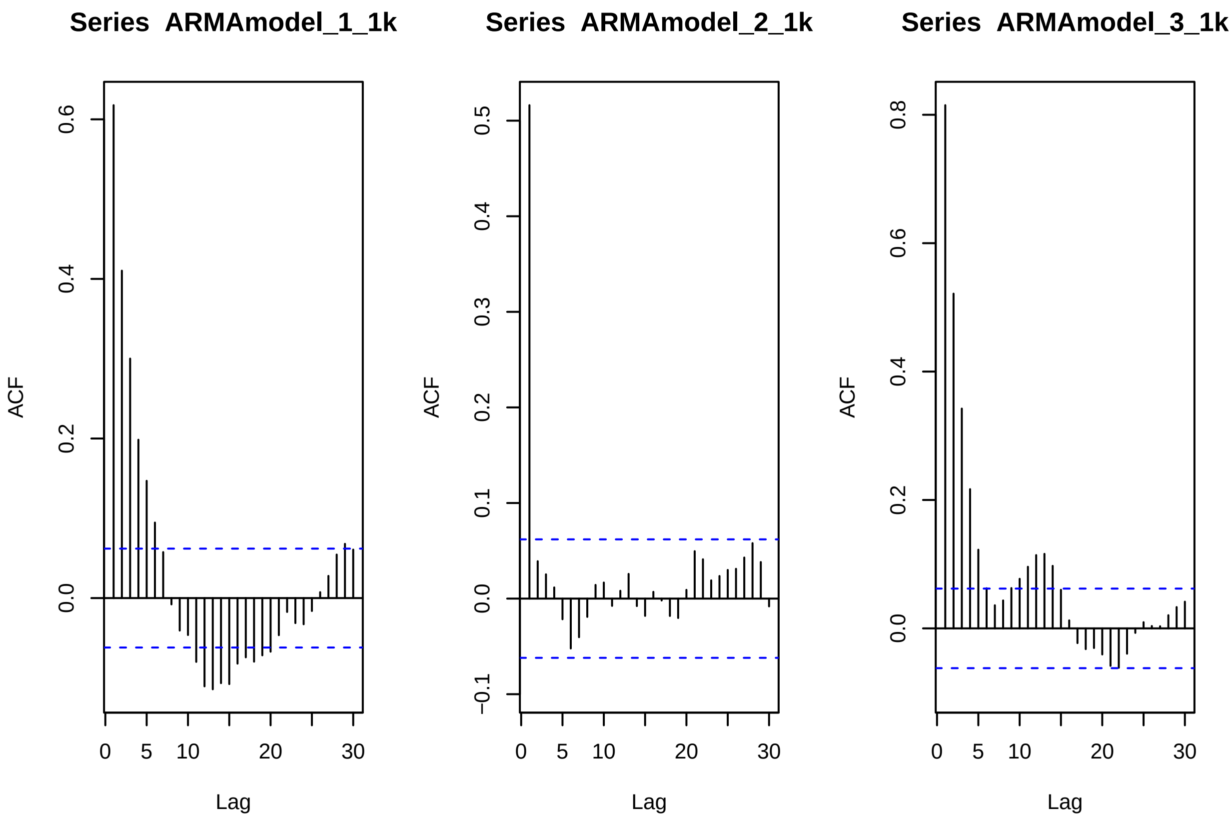
(d) Compare the ACF and PACF values R computed with the theoretical values you provided for the coefficients. Do they match? Explain your answer.

Answer: They don't. They line up for the first plot but not the other two. I'm guessing this has somewhat to do with the relatively low number of observations.

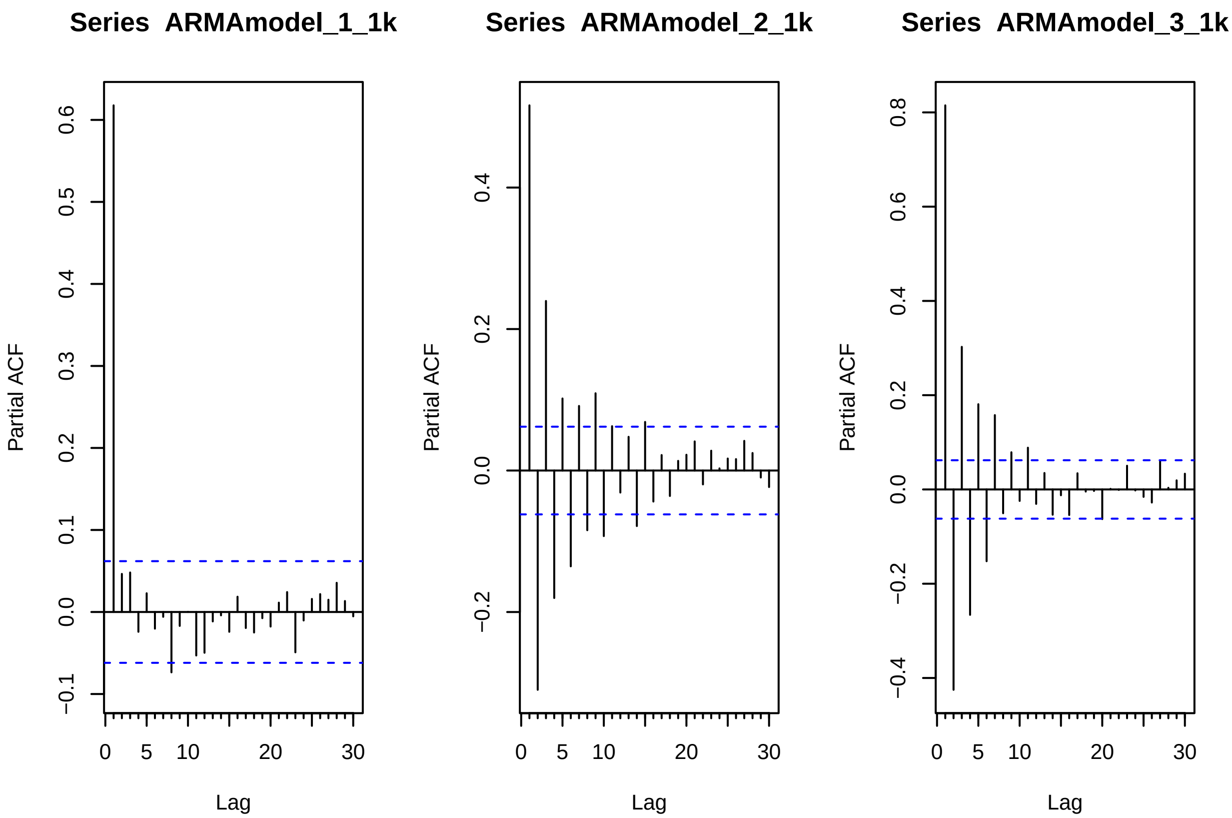
(e) Increase number of observations to $n = 1000$ and repeat parts (a)-(d).

```
#Observations to 1000
#ARMA(1,0)
ARMAmodel_1_1k<- arima.sim(model=list(ar=0.6), n=1000) #the AR coefficient is 0.6

#ARMA(0,1)

ARMAmodel_2_1k<- arima.sim(model=list(ma=0.9), n=1000) #the MA coefficient is 0.9
```

```r
#ARMA(1,1)
ARMAmodel_3_1k<- arima.sim(model=list(ar=0.6, ma=0.9), n=1000)

#new ACFs
par(mfrow=c(1,3))
Acf(ARMAmodel_1_1k)
Acf(ARMAmodel_2_1k)
Acf(ARMAmodel_3_1k)
```




```r
#new PACFs
par(mfrow=c(1,3))
Pacf(ARMAmodel_1_1k)
Pacf(ARMAmodel_2_1k)
Pacf(ARMAmodel_3_1k)
```



(c with n=1000) > Answer: Compared with part c at 100 observations, I find the first two plots much easier to interpret. I think I would misinterpret the order of the third plot.

(d with n=1000) > Answer: I believe the do match for the first two, but on the third I would guess different orders.

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.

ARIMA (1,0,1)(1,0,0)

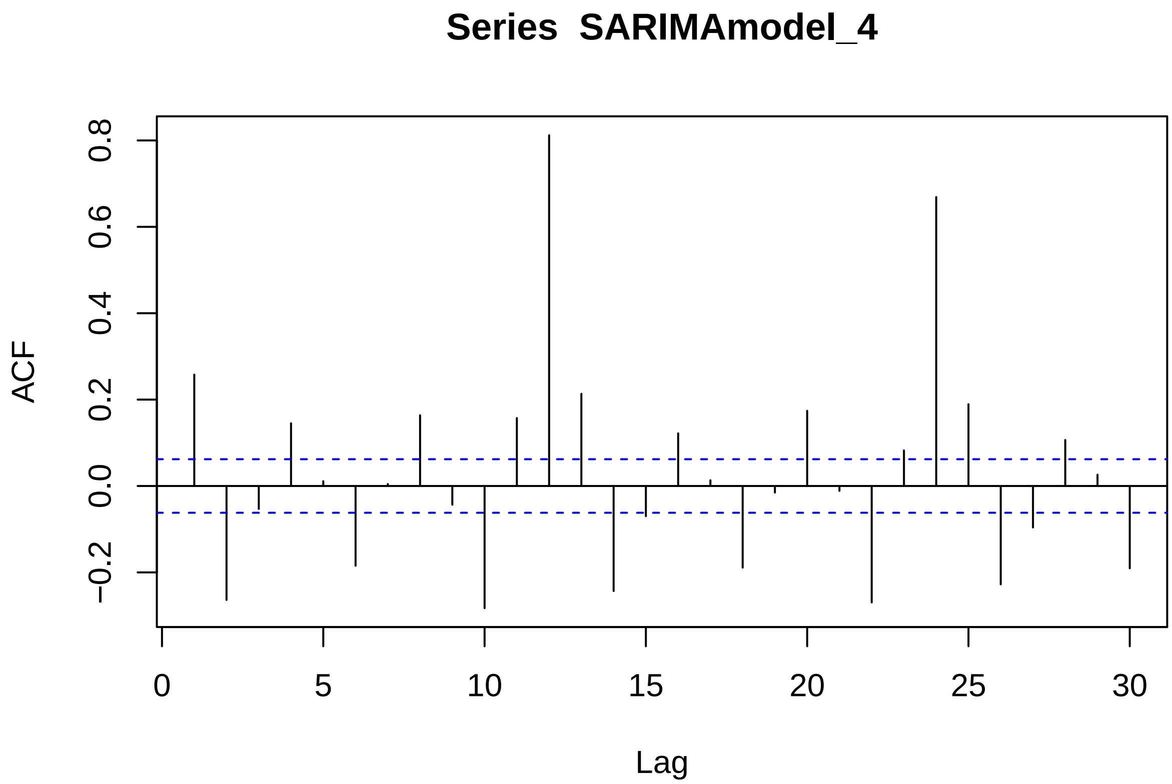
(b) Also from the equation what are the values of the parameters, i.e., model coefficients.

AR term (phi) = 0.7 MA term (theta) = 0.1 SAR term = 0.25

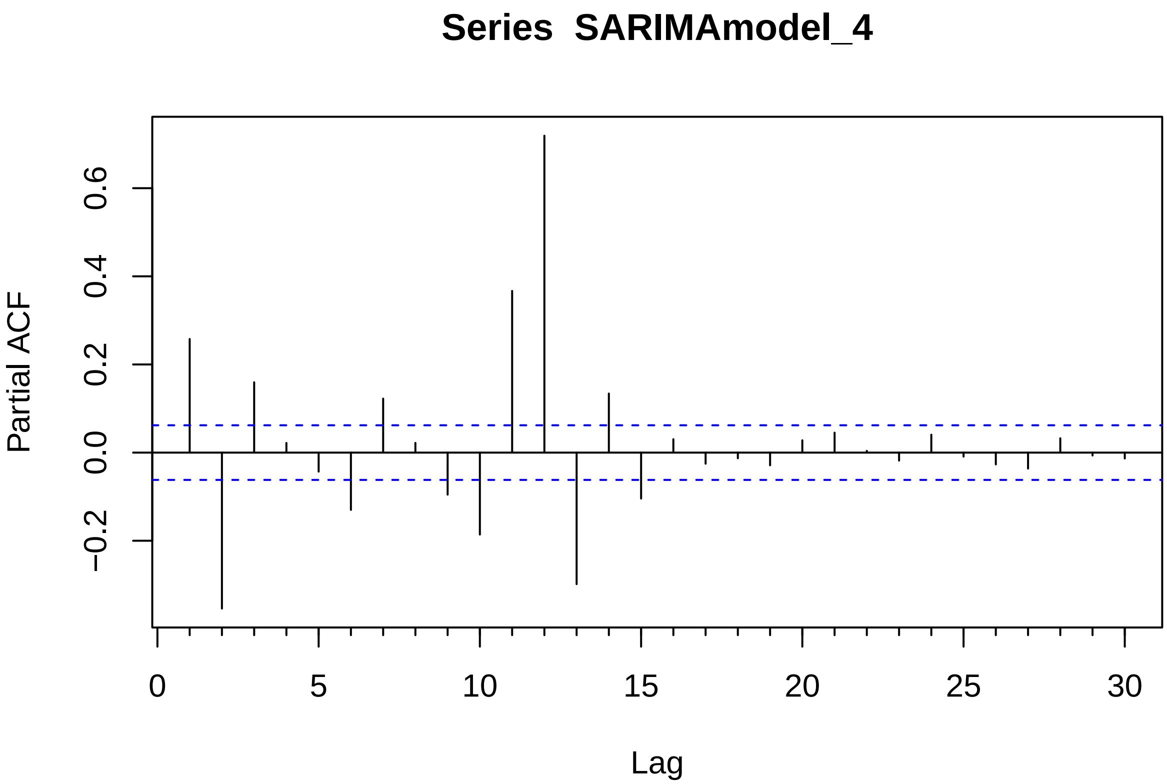
Q4

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.

```r
SARIMAmodel_4<- sim_sarima(model=list(ma=0.5,sar=0.8, nseasons=12), n=1000)

Acf(SARIMAmodel_4)
```




```r
Pacf(SARIMAmodel_4)
```



> Answer: I would be able to identify the non-seasonal components, ie the MA order being 1 from the ACF plot. I would not be able to identify the seasonal component from the PACF.