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

Assignment 6 - Due date 03/06/23

Kelsey Husted

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(gridExtra)
require(scales)
library(readxl)
library(openxlsx)
library(forecast)
library(tseries)
library(ggplot2)
library(Kendall)
library(lubridate)
library(tidyverse)
```

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: When looking at the ACF plot, the lags are geometric with a gradual decrease for an AR(2) model. The PACF will display two significant initial lags in which any following lags are cut off right after which represents the order (i.e., p=2).

- MA(1)

  Answer: The ACF plot will provide the order of the model (i.e.,q) and should display two significant lags with the first lag at 1 representing zero and then the adjacent, significant lag respresenting an order of one. The PACF plot will display lags that are geometrically decreasing or diminishing.

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

```
#ARIMA(1,0)
arma10 <- arima.sim(model = list(ar = 0.6),n = 100)
#ARIMA(0,1)
arma01 <- arima.sim(model = list(ma = 0.9),n = 100)
#ARIMA(1,1)
arma11 <- 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).

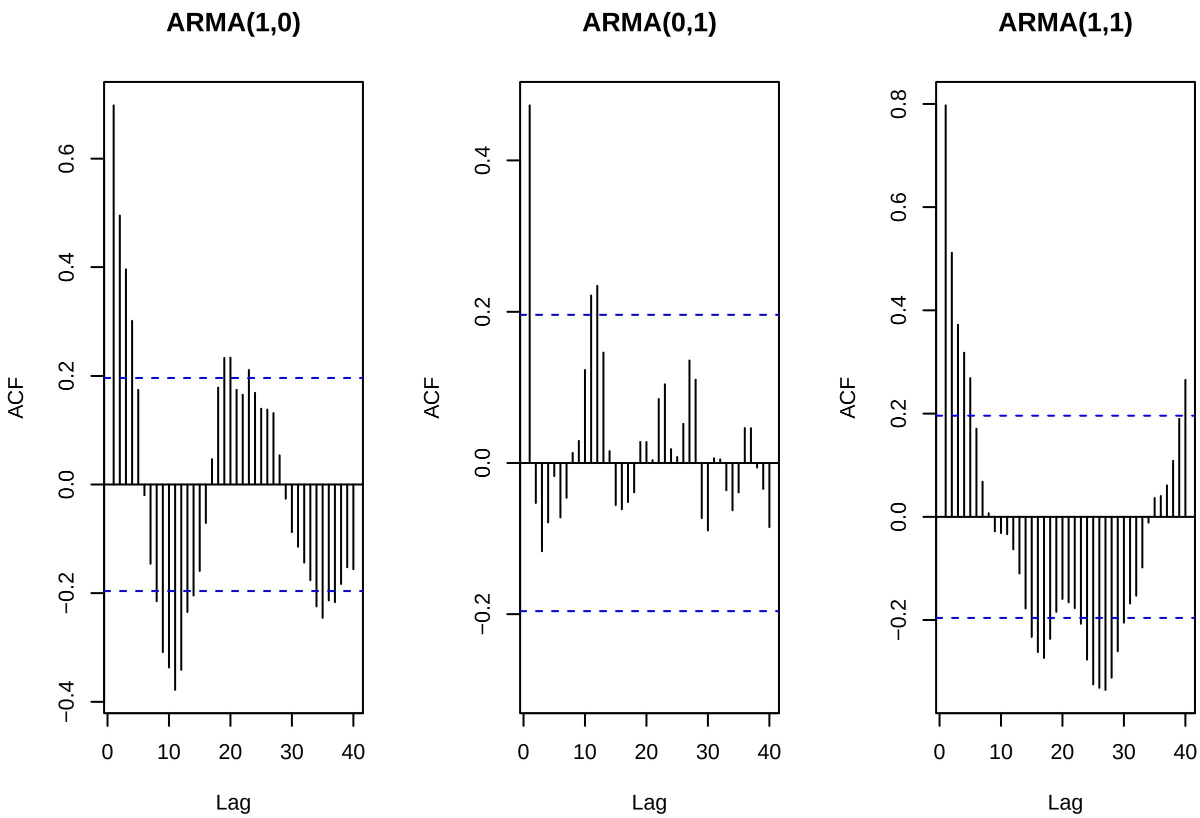
```
par(mfrow=c(1,3))
Acf(arma10, lag.max = 40,main = "ARMA(1,0)")
print(Acf(arma10, lag.max = 5, plot=FALSE))
```

```
##
## Autocorrelations of series 'arma10', by lag
##
##     0     1     2     3     4     5
## 1.000 0.698 0.495 0.396 0.301 0.174
```

```
Acf(arma01, lag.max = 40,main = "ARMA(0,1)")
print(Acf(arma01, lag.max = 5, plot=FALSE))
```

```
##
## Autocorrelations of series 'arma01', by lag
##
##      0      1      2      3      4      5
##  1.000  0.473 -0.053 -0.117 -0.079 -0.017
```

```
Acf(arma11, lag.max = 40,main = "ARMA(1,1)")
```




```
print(Acf(arma11, lag.max = 5,  plot=FALSE))
```

```
##
## Autocorrelations of series 'arma11', by lag
##
##     0     1     2     3     4     5
## 1.000 0.797 0.512 0.372 0.318 0.269
```

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

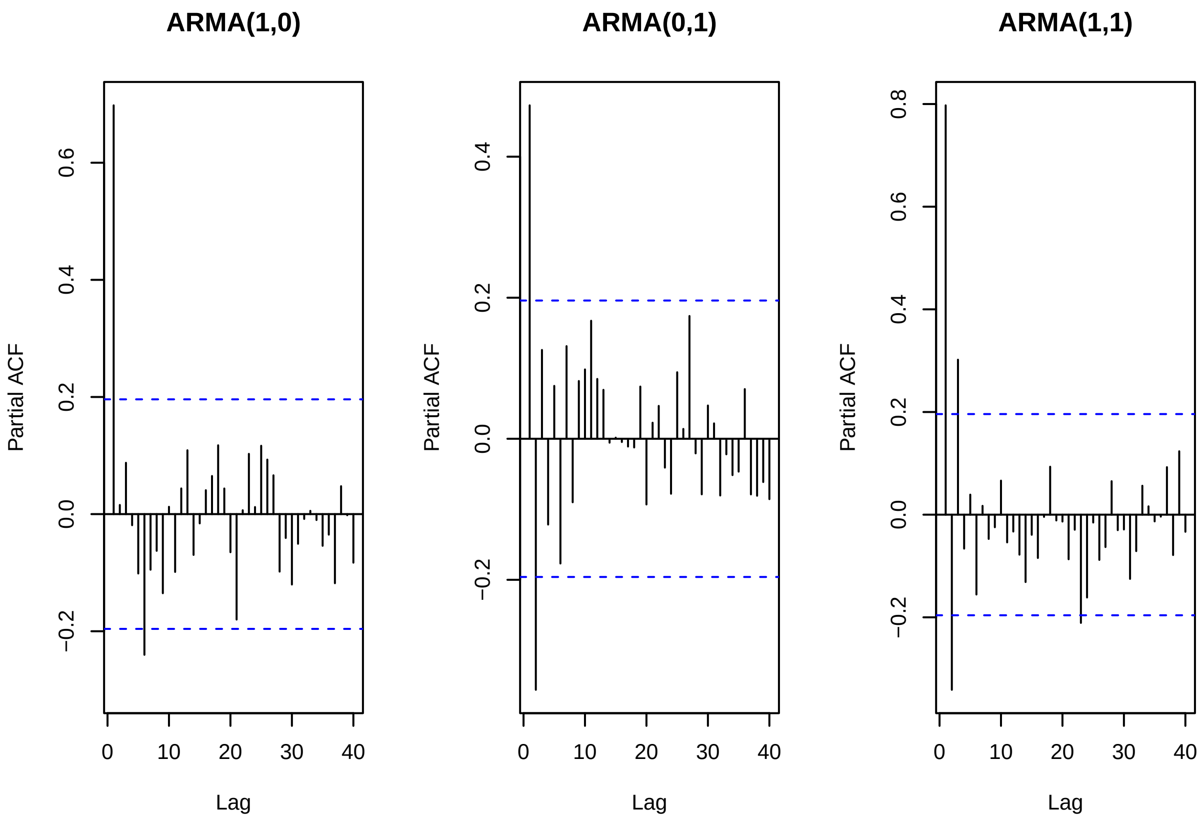
```
par(mfrow=c(1,3))
Pacf(arma10, lag.max = 40,main = "ARMA(1,0)")
print(Pacf(arma10, lag.max = 5, plot=FALSE))
```

```
##
## Partial autocorrelations of series 'arma10', by lag
##
##      1      2      3      4      5
##  0.698  0.016  0.088 -0.019 -0.101
```

```
Pacf(arma01, lag.max = 40,main = "ARMA(0,1)")
print(Pacf(arma01, lag.max = 5, plot=FALSE))
```

```
##
## Partial autocorrelations of series 'arma01', by lag
##
##      1      2      3      4      5
##  0.473 -0.356  0.126 -0.122  0.075
```

```
Pacf(arma11, lag.max = 40,main = "ARMA(1,1)")
```




```
print(Pacf(arma11, lag.max = 5, plot=FALSE))
```

```
##
## Partial autocorrelations of series 'arma11', by lag
##
##      1      2      3      4      5
##  0.797 -0.341  0.302 -0.066  0.039
```

(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: I don't know if I would personally be able to identify each model and the order. All PACF and ACF plots look a little geometric so maybe the models would be an ARMA. However, the PACF plots seem to have more of a cutoff than the ACF plots. Furthermore, I would probably struggle identifying each of the models and their order. I think a higher number of observations (n) would really help distinguish the plots with respect to each model and order.

(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 values provided included phi (ar) = 0.6 and theta (ma)=0.9. The computed values don't really match the theoretical values that were provided as seen in the ACF and PACF plots above. I'm not too sure why these coefficients aren't matching; maybe more observations are needed to reconcile these value differences.

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

```
arma10_1000obs <- arima.sim(n = 1000, model = list(ar = 0.6))
arma01_1000obs <- arima.sim(n = 1000, model = list(ma = 0.9))
arma11_1000obs <- arima.sim(n = 1000, model = list(ar = 0.6, ma = 0.9))
```

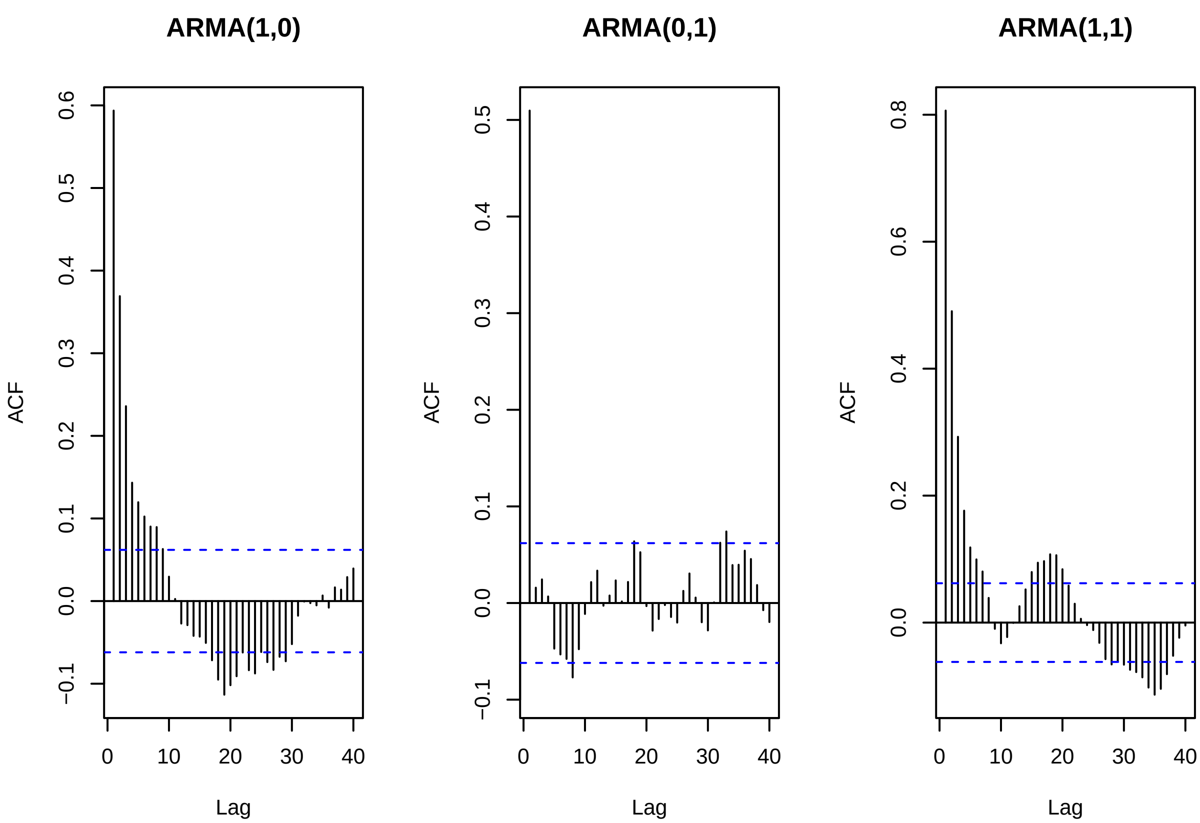
```r
par(mfrow=c(1,3))
Acf(arma10_1000obs, lag.max = 40,main = "ARMA(1,0)")
print(Acf(arma10_1000obs, lag.max = 5, plot=FALSE))
```

```
##
## Autocorrelations of series 'arma10_1000obs', by lag
##
##     0     1     2     3     4     5
## 1.000 0.594 0.369 0.236 0.143 0.120
```

```r
Acf(arma01_1000obs, lag.max = 40,main = "ARMA(0,1)")
print(Acf(arma01_1000obs, lag.max = 5, plot=FALSE))
```

```
##
## Autocorrelations of series 'arma01_1000obs', by lag
##
##      0      1      2      3      4      5
##  1.000  0.510  0.016  0.024  0.007 -0.047
```

```r
Acf(arma11_1000obs, lag.max = 40,main = "ARMA(1,1)")
```




```r
print(Acf(arma11_1000obs, lag.max = 5, plot=FALSE))
```

```
##
## Autocorrelations of series 'arma11_1000obs', by lag
##
##     0     1     2     3     4     5
## 1.000 0.807 0.491 0.293 0.176 0.118
```

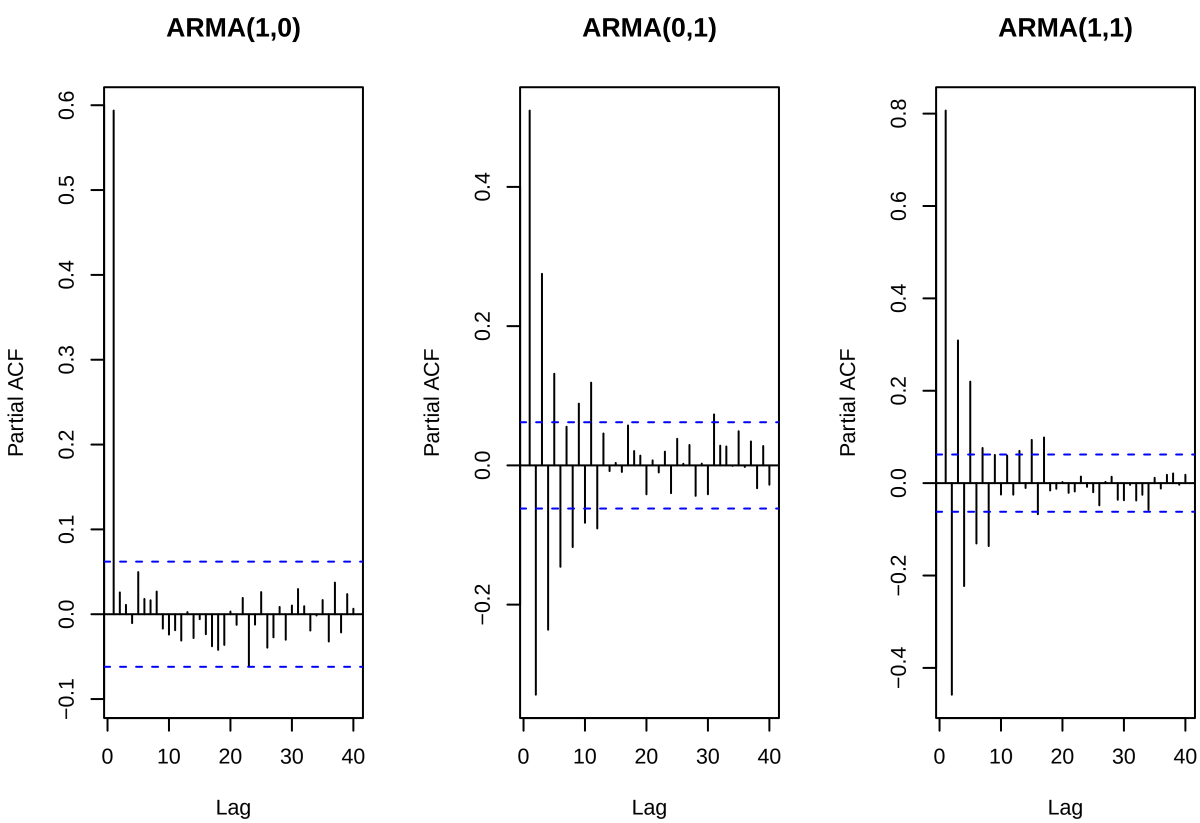
```r
par(mfrow=c(1,3))
Pacf(arma10_1000obs, lag.max = 40,main = "ARMA(1,0)")
print(Pacf(arma10_1000obs, lag.max = 5, plot=FALSE))
```

```
##
## Partial autocorrelations of series 'arma10_1000obs', by lag
##
##      1      2      3      4      5
##  0.594  0.026  0.011 -0.010  0.050
```

```r
Pacf(arma01_1000obs, lag.max = 40,main = "ARMA(0,1)")
print(Pacf(arma01_1000obs, lag.max = 5, plot=FALSE))
```

```
##
## Partial autocorrelations of series 'arma01_1000obs', by lag
##
##      1      2      3      4      5
##  0.510 -0.329  0.275 -0.236  0.132
```

```r
Pacf(arma11_1000obs, lag.max = 40,main = "ARMA(1,1)")
```




```r
print(Pacf(arma11_1000obs, lag.max = 5, plot=FALSE))
```

```
##
## Partial autocorrelations of series 'arma11_1000obs', by lag
##
##      1      2      3      4      5
##  0.807 -0.458  0.309 -0.223  0.220
```

(c.2) 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: With more observations (i.e., n = 1000), I can identify the AR in ARMA(1,0) with a geometric ACF and a cut off PACF. I also notice the geometric decrease on the PACF and ACF cut off for the ARMA(0,1) model which is what I would expect to observe. The ARMA(1,1) displays a geometric decay with both ACF and PACF plots which is expected.

(d.2) 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 values provided included phi (ar) = 0.6 and theta (ma)=0.9. The computed values still don't really match the theoretical values that were provided as seen in the ACF and PACF plots above. I'm not too sure why these coefficients aren't matching; I originally thought thta more observations are needed to reconcile these value differences.

Q3

Consider the ARIMA model $y_t = 0.7 * y_{t-1} - 0.25 * y_{t-12} + a_t - 0.1 * a_{t-1}$ #if no constant term that means the series has been differentiated

(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.

> Answer: ARIMA(1,0,1)(1,0,0)__12 p = 1 since there is one ar lag unit: 0.7yt-1 d = 0 since there is no constant which means the series has been differentiated q = 1 since there is one ma lag unit: 0.1at-1 P = 1 since there is one SAR term: 0.25yt-12 D = 0 since there is no constant Q = 0 since there is no SMA term

(b) Also from the equation what are the values of the parameters, i.e., model coefficients. > Answer: phi(ar) = 0.7; theta(ma) = 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?

> Answer: When I initially look at the plots, I can tell that ACF lags decay slowly, while the PACF cuts off so I think this indicates an AR process. I believe that the PACF plot indicates a order of 2 (i.e., p = 2)which cuts off around lag 12.

```
#install.packages("sarima")
require(sarima)
```

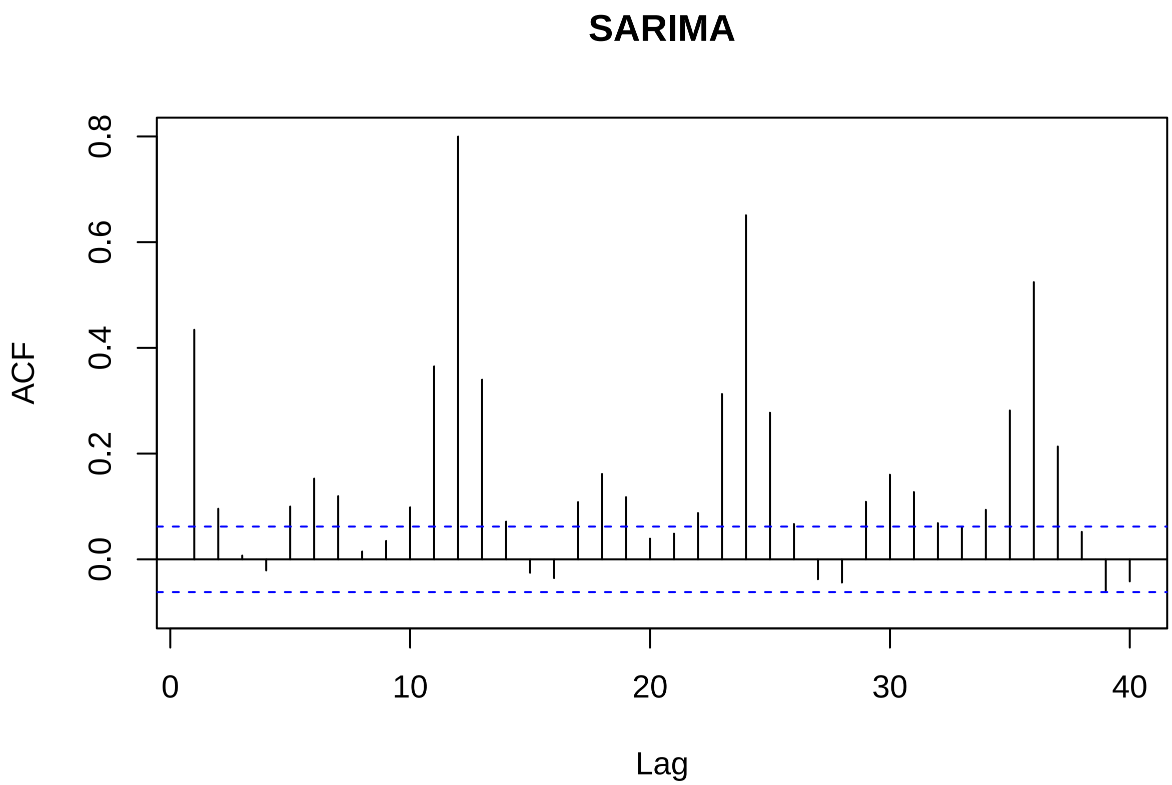
```
## Loading required package: sarima

## Loading required package: stats4

##
## Attaching package: 'sarima'

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

```r
SARIMAmodel_4<- sim_sarima(model=list(ma=0.5,sar=0.8, nseasons=12), n=1000)
Acf(SARIMAmodel_4, lag.max = 40,main = "SARIMA")
```




```r
Pacf(SARIMAmodel_4, lag.max = 40,main = "SARIMA")
```