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

Assignment 5 - Due date 03/12/21

Marie McNamara

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(lubridate)
```

```
##
## Attaching package: 'lubridate'

## The following objects are masked from 'package:base':
##
##     date, intersect, setdiff, union
```

```
library(ggplot2)
library(forecast)
```

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

```
library(Kendall)
library(tseries)
library(outliers)
library(tidyverse)
```

```
## -- Attaching packages --------------------------------------- tidyverse 1.3.0 --

## v tibble  3.0.6     v dplyr   1.0.3
## v tidyr   1.1.2     v stringr 1.4.0
```

```
## v readr   1.4.0     v forcats 0.5.1
## v purrr   0.3.4
```

```
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x lubridate::as.difftime() masks base::as.difftime()
## x lubridate::date()        masks base::date()
## x dplyr::filter()          masks stats::filter()
## x lubridate::intersect()   masks base::intersect()
## x dplyr::lag()             masks stats::lag()
## x lubridate::setdiff()     masks base::setdiff()
## x lubridate::union()       masks base::union()
```

```
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:

AR(2) > Answer: Is an autoregressive process in which the current value is based on the two preceding values. The 2 refers to the order of the autoregressive.

MA(1) > Answer: is a moving average process in which the current value is based on the preceding value. The order of this moving average is 1.

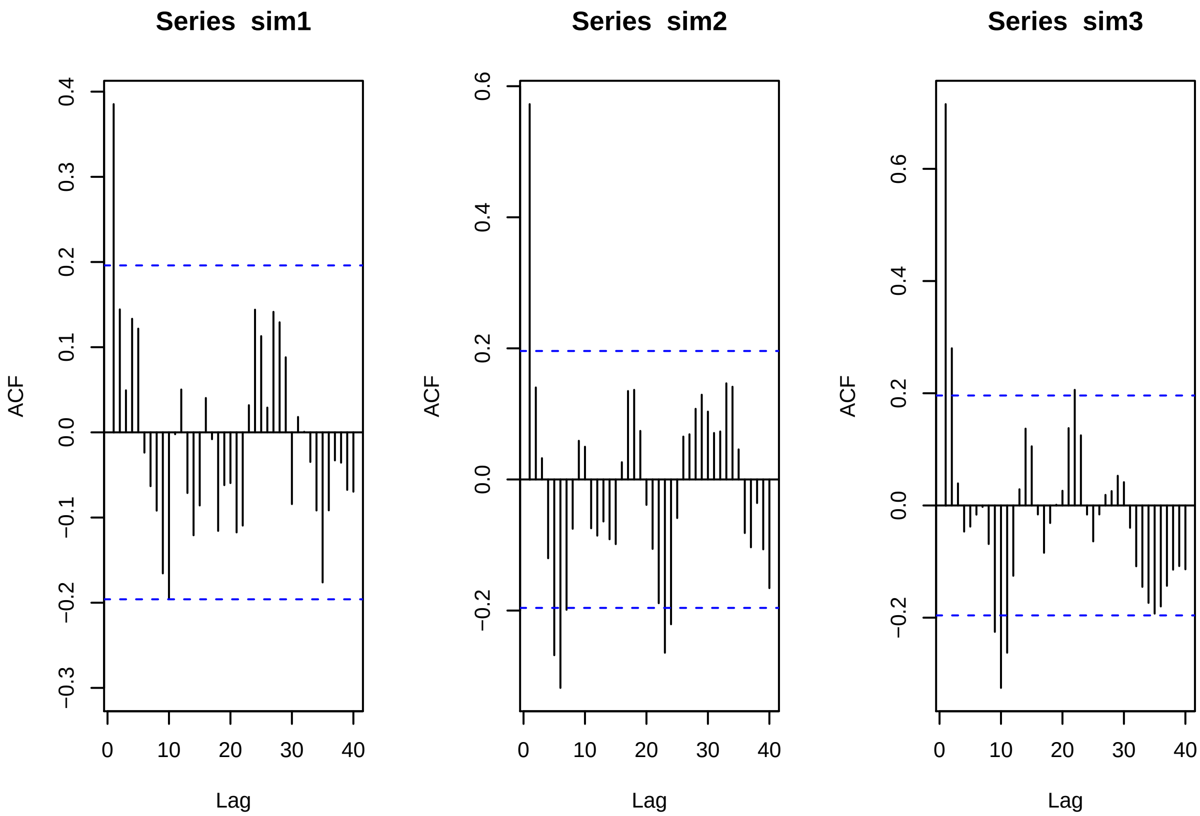
Q2

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

```
sim1 <- arima.sim(list(order = c(1,0,0), ar = 0.6), n = 100)

sim2 <- arima.sim(list(order = c(0,0,1), ma =0.9), n =100)

sim3 <- arima.sim(list(order = c(1,0,1), ar = 0.6, ma =0.9), n = 100)
```

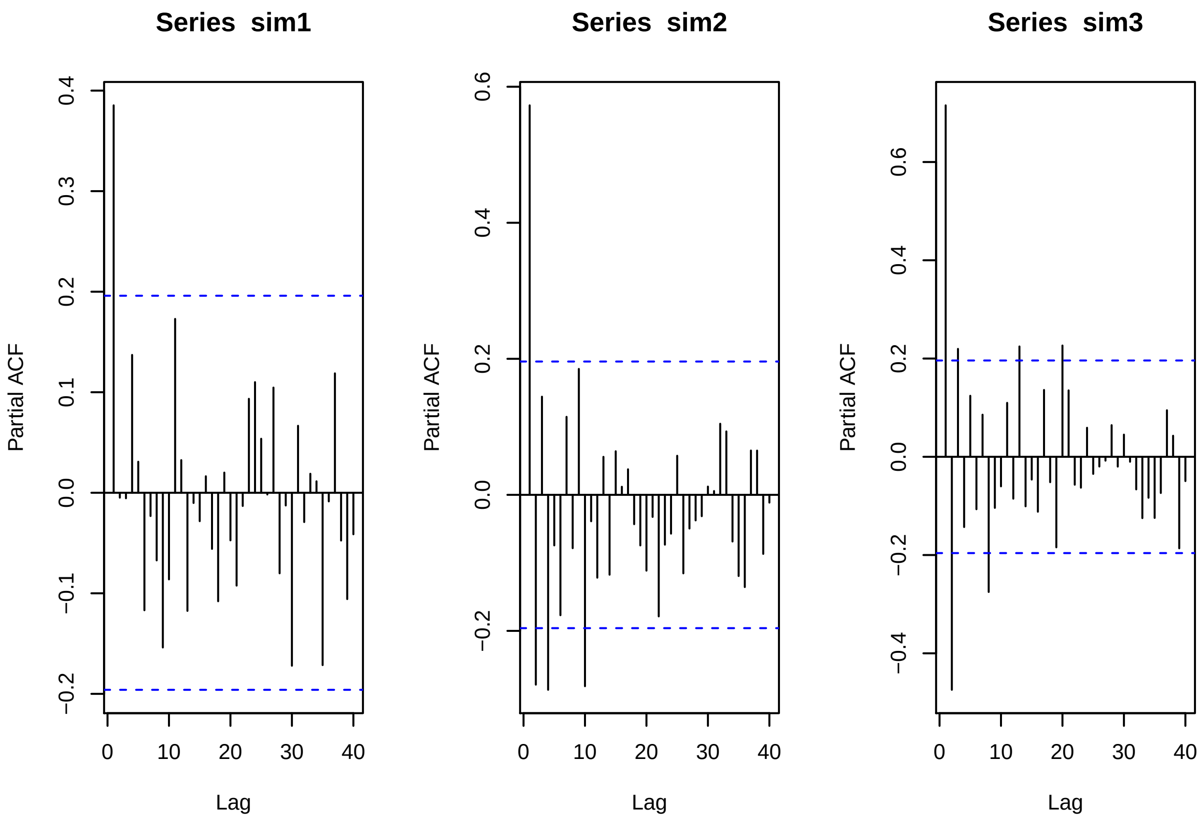
Plot the sample ACF for each of these models in one window to facilitate comparison (Hint: use command $par(mfrow = c(1, 3))$ that divides the plotting window in three columns).

```
par(mfrow=c(1,3))
acf1 = Acf(sim1, lag.max = 40, plot = TRUE)
acf2 = Acf(sim2, lag.max = 40, plot = TRUE)
acf3 = Acf(sim3, lag.max = 40, plot = TRUE)
```



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

```
par(mfrow=c(1,3))
pacf1 = pacf(sim1, lag.max = 40, plot = TRUE)
pacf2 = pacf(sim2, lag.max = 40, plot = TRUE)
pacf3 = pacf(sim3, lag.max = 40, plot = TRUE)
```



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 is an AR order 1. You can tell from the PACF cut off at lag 1. You can also see the slow decay in the ACF.The second model is a MA order 1 you can see this from th ACF cut off at lag 1. The third model is a ARMA you can see some slow decay decay in the ACF. But there are two significant cut offs in the PACF. It appears that the AR and MA are superimposed

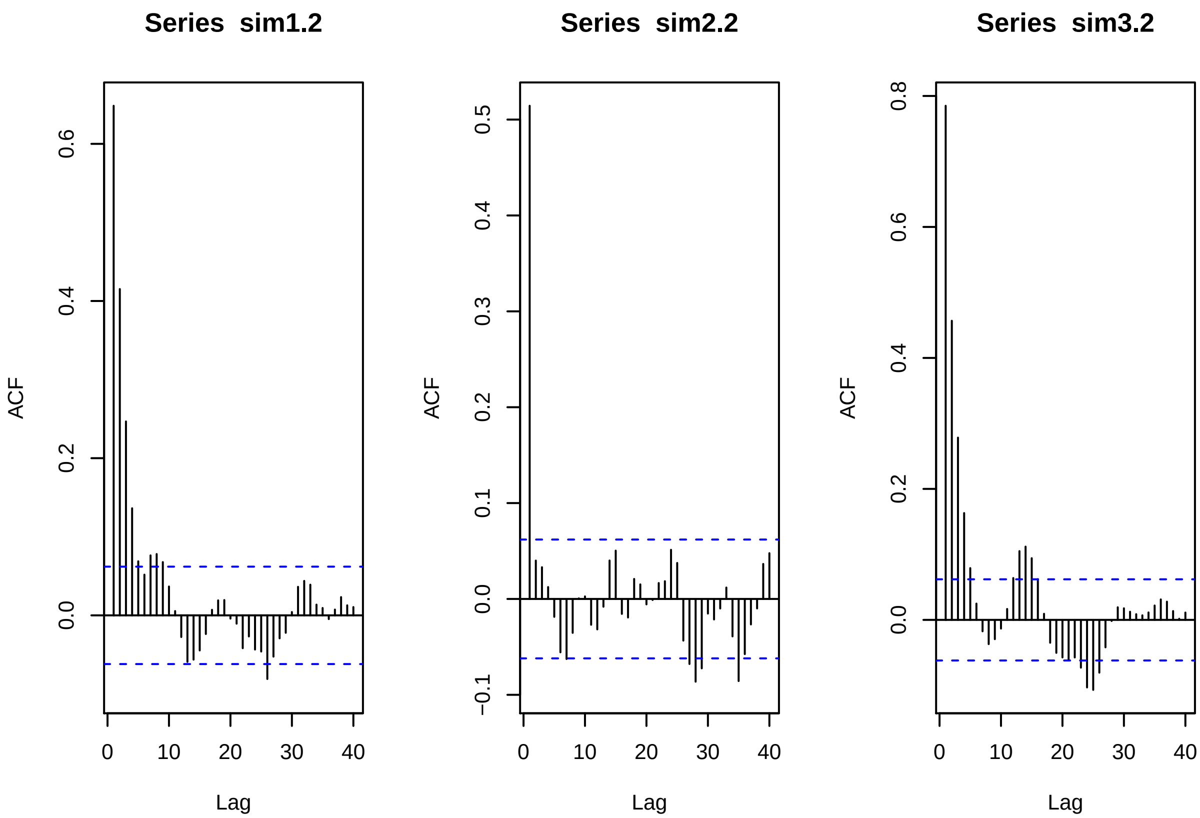
Compare the ACF and PACF values R computed with the theoretical values you provided for the coefficients. Do they match? Explain your answer.

> Answer: The ACF values of R range from +0.4 to 0.8. Phi is an auto correlation. The phi (which is 0.6) is within the range of computed R values on the ACF plots for lag 1. The PACF values of R do not match theta. This is because theta represents the relationship among yt and a_t-1, and is not autocorrelation or autoregressive.

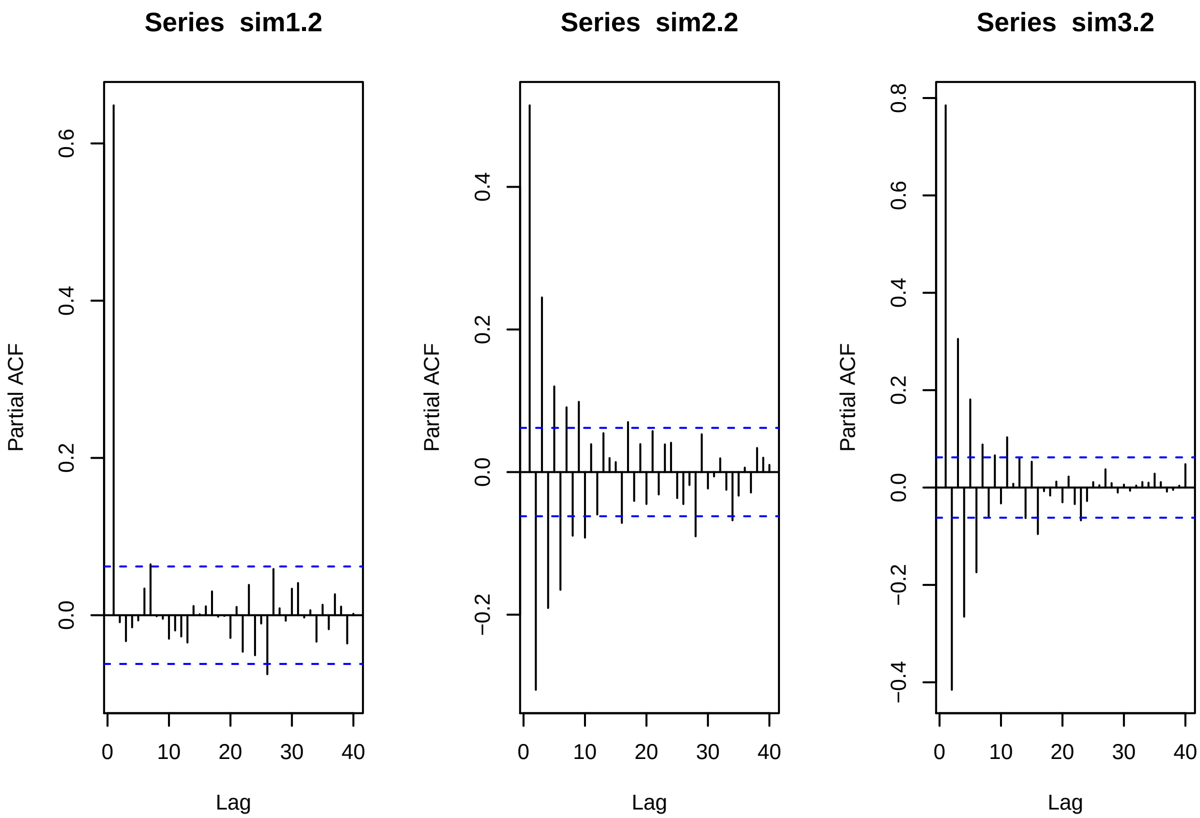
Increase number of observations to $n = 1000$ and repeat parts (a)-(d).

```
sim1.2 <- arima.sim(list(order = c(1,0,0), ar = 0.6), n = 1000)


sim2.2 <- arima.sim(list(order = c(0,0,1), ma =0.9), n =1000)

sim3.2 <- arima.sim(list(order = c(1,0,1), ar = 0.6, ma =0.9), n = 1000)

par(mfrow=c(1,3))
pacf1.2 = Acf(sim1.2, lag.max = 40, plot = TRUE)
pacf2.2 = Acf(sim2.2, lag.max = 40, plot = TRUE)
```

```r
pacf3.2 = Acf(sim3.2, lag.max = 40, plot = TRUE)
```




```r
par(mfrow=c(1,3))
pacf1.2 = Pacf(sim1.2, lag.max = 40, plot = TRUE)
pacf2.2 = Pacf(sim2.2, lag.max = 40, plot = TRUE)
pacf3.2 = Pacf(sim3.2, lag.max = 40, plot = TRUE)
```



(model 1)There is clear slow decay in simulation 1 indicating an AR process. The AR order is 1, you can tell from the PACF cut off at lag 1.

(model 2)The ACF cut off at lag 1 in simulated 2 is very apparent even more so than when n=100, again signalling a MA process.

(model 3) The third model is a ARMA you can see some decay in the ACF. The PACF plot seems to oscillate between positive and negative even more so than when the n=100 scenario was modeled.This oscillation indicates that the AR and MA more clearly superimposed.

(The R coefficients)T he R coefficients in the ACF plot are even closer to the phi value of 0.6. The range of R coefficients are 0.5 to 0.8. Again th R coefficients for the PACF are not similar to the 0.9 modeled value as theta. Theta represents the relationship among yt and a_t-1, and is not autocorrelation.

Q3Consider the ARIMA model $y_t = 0.7 * y_{t-1} - 0.25 * y_{t-12} + a_t - 0.1 * a_{t-1}$

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. p=1 d=0 q=1 P=1 D=0 q=0 (1,0,1)(1,0,0)

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

(AR) phi=0.7 (MA) theta=-0.1 SAR= -.25

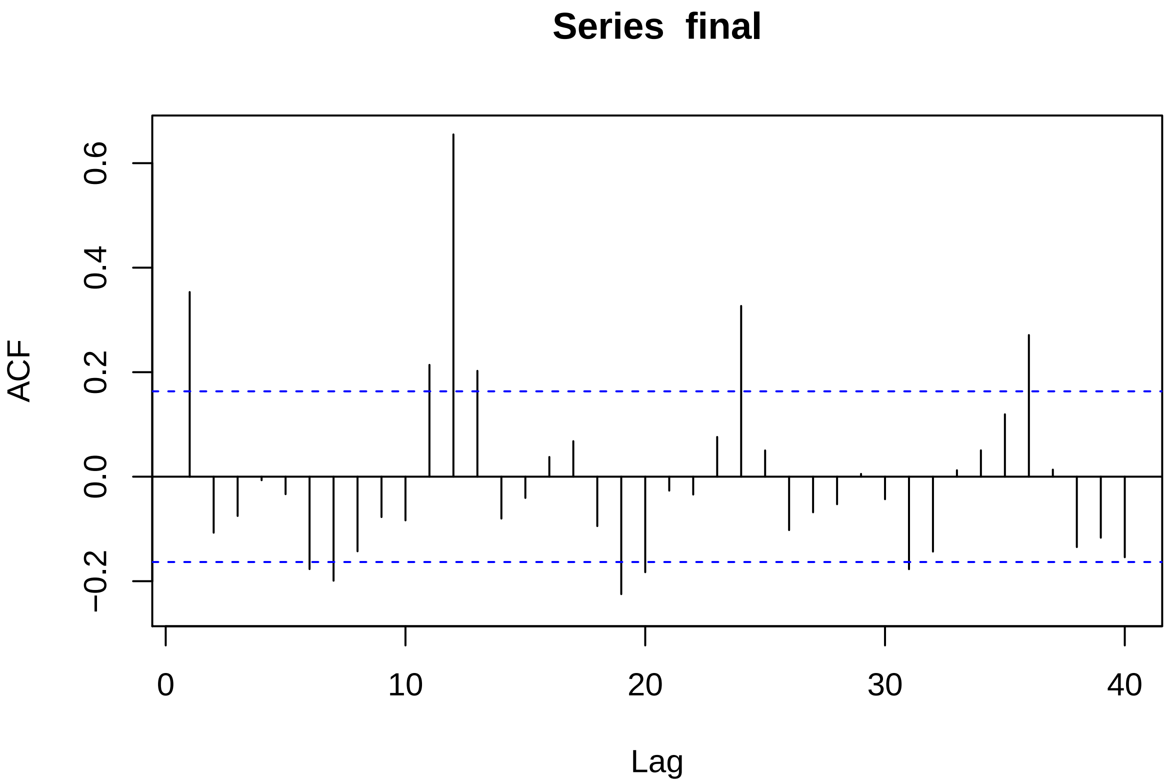
Q4

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.

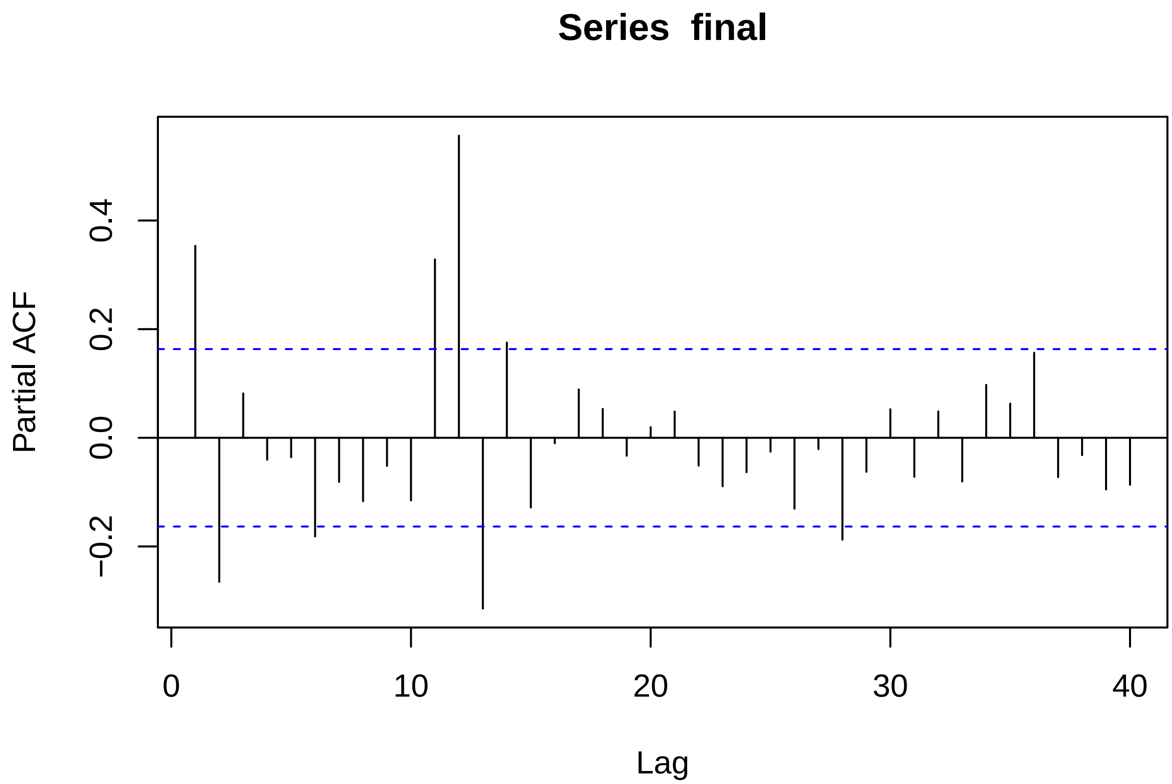
Yes, the ACF cuts off at lag 12, 24 signalling an MA process. The plots have multiple cut offs, with spikes at lags 12 and 24 leading me to concluded this is a SAR process.

The PACF gives you the value of P, there is just one single spike at the seasonal lag which leads me to concluded this is SAR process.

```
final <- sim_sarima(n=144, model = list(ma=0.5, sar=0.8, iorder=0, siorder=0, nseasons=12))

pacf_final = Acf(final, lag.max = 40, plot = TRUE)
```




```
acf_final = Pacf(final, lag.max = 40, plot = TRUE)
```


Series final
Partial ACF
0.4
0.2
0.0
−0.2
0
10
20
30
40
Lag