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

Assignment 6 - Due date 03/06/23

Angela Zeng

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(readxl)
library(forecast)
```

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

```
library(tseries)
library(ggplot2)
library(Kendall)
library(lubridate)
```

```
## Loading required package: timechange

##
## Attaching package: 'lubridate'

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

```r
library(tidyverse)
```

```
## -- Attaching packages --------------------------------------- tidyverse 1.3.2 --
## v tibble  3.1.8      v dplyr   1.0.10
## v tidyr   1.2.1      v stringr 1.5.0
## v readr   2.1.3      v forcats 0.5.2
## v purrr   1.0.1
## -- 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()
```

```r
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: The ACF will have slow decay (due to time dependence) and the PACF will cut off at the second lag, which is indicated by the first lag that is not at 1.0 on the y-axis.

- MA(1)

  Answer: The ACF will have a negative coefficient lag (not including the lag at 1.0 on the y-axis). The PACF will have slow decay and a cut off at lag 1.

Q2

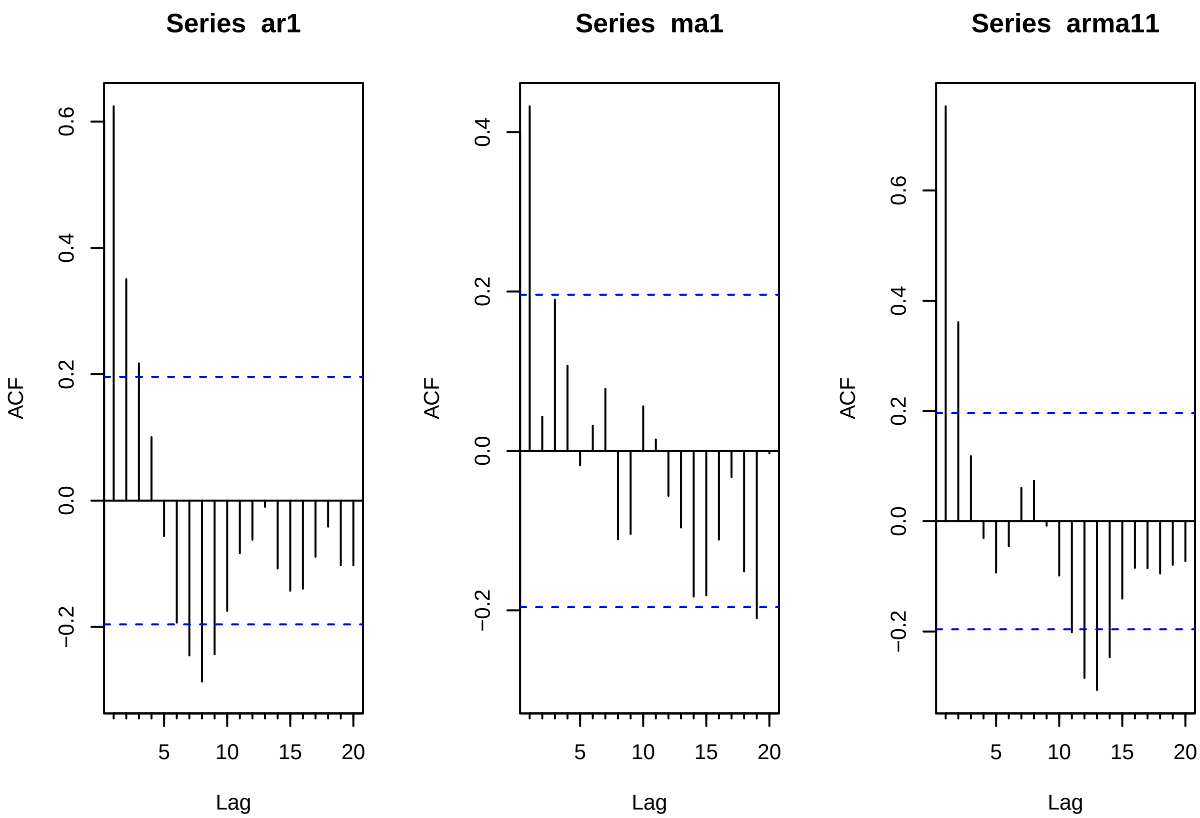
Recall that the non-seasonal ARIMA is described by three parameters ARIMA(p, d, q) where p is the order of the autoregressive component, d is the number of times the series need to be differenced to obtain stationarity and q is the order of the moving average component. If we don't need to difference 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

```r
#ARMA(1,0)
ar1<- arima.sim(model=list(ar = 0.6), n = 100)

#ARMA(0,1)
ma1<- arima.sim(model=list(ma = 0.9), n = 100)
```

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

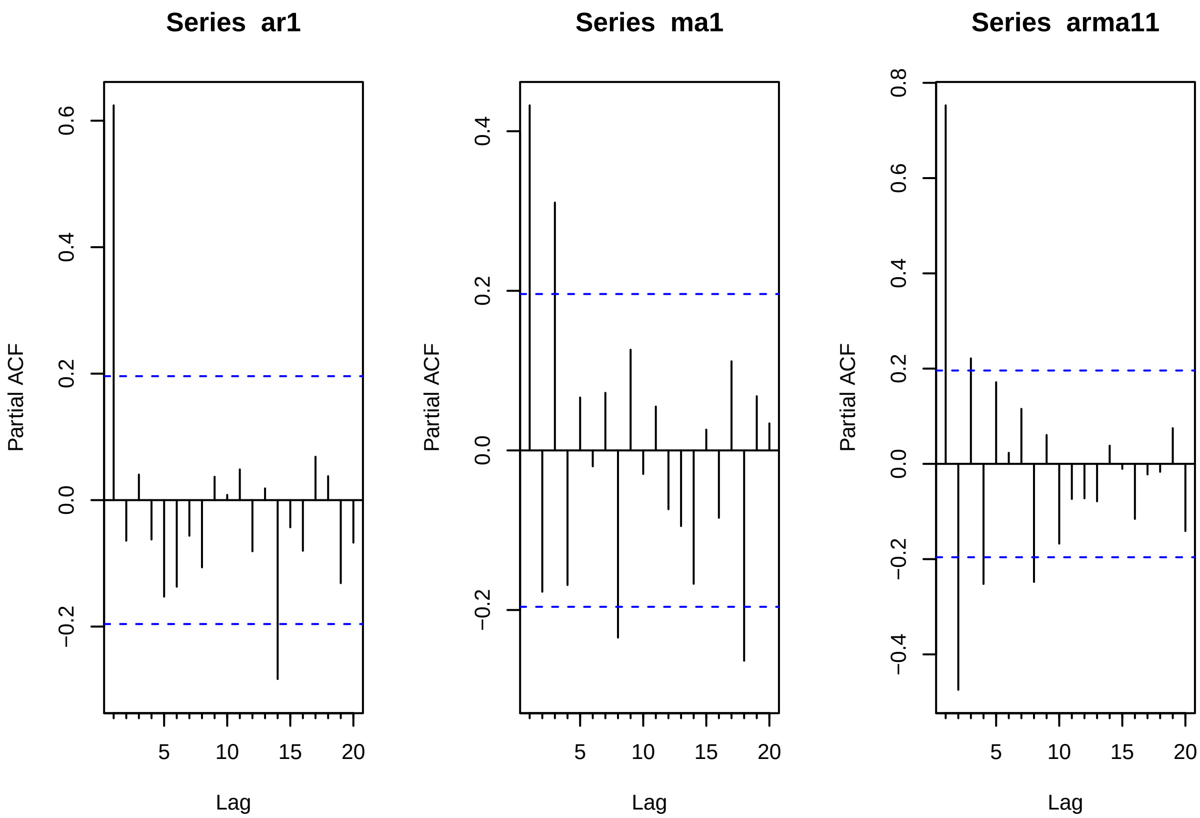
(a) Plot the sample ACF for each of these models in one window to facilitate comparison (Hint: use command $par(mfrow = c(1, 3))$ that divides the plotting window in three columns).

```
par(mfrow=c(1,3))
Acf(ar1)
Acf(ma1)
Acf(arma11)
```



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

```
par(mfrow=c(1,3))
Pacf(ar1)
Pacf(ma1)
Pacf(arma11)
```



(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.

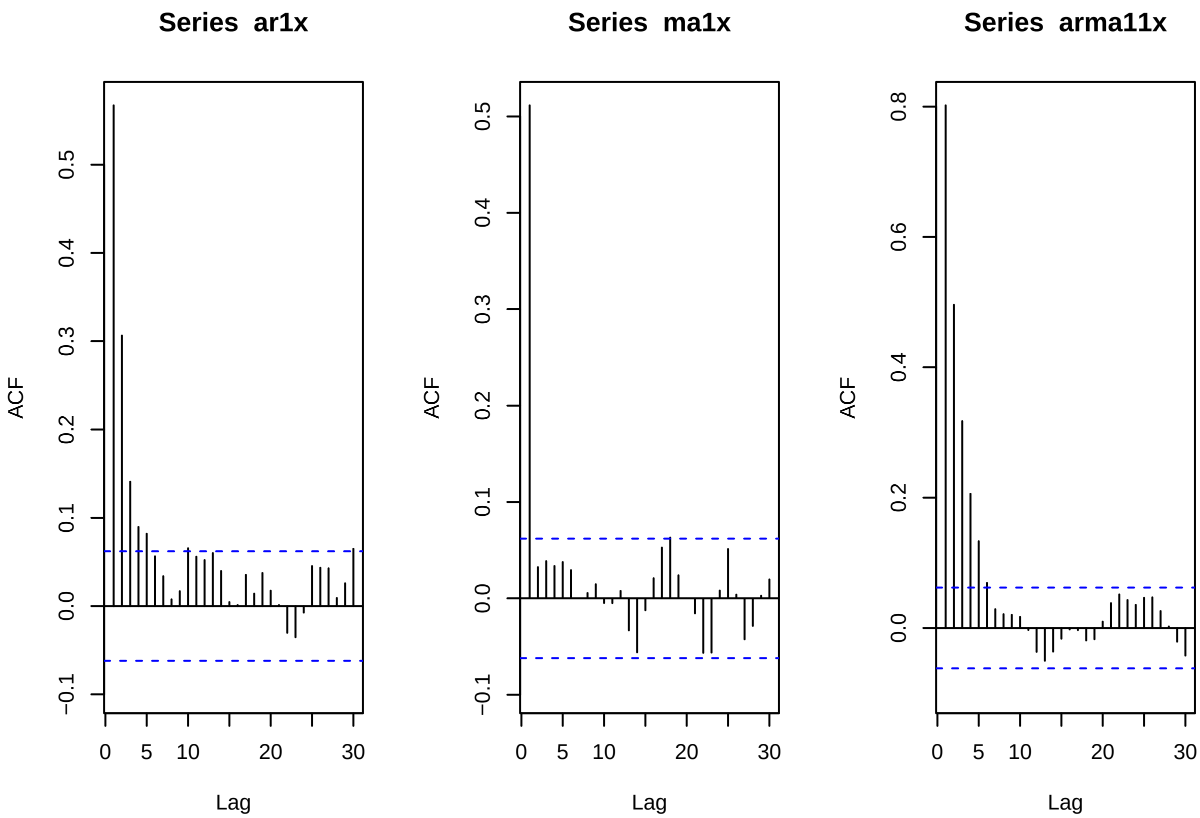
Answer: Yes, the ARMA(1,0) has slow decay in the ACF plot and the a cut off at the first lag in the PACF plot. The ARMA(0,1) does not have slow decay and cuts off at the first lag in the ACF plot; the PACF plot has a negative lag outside of the bounds. The ARMA(1,1) has slow decay in the ACF plot, cuts off at the first lag in the ACF plot, and has a negative lag outside of the bounds in the PACF plot.

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

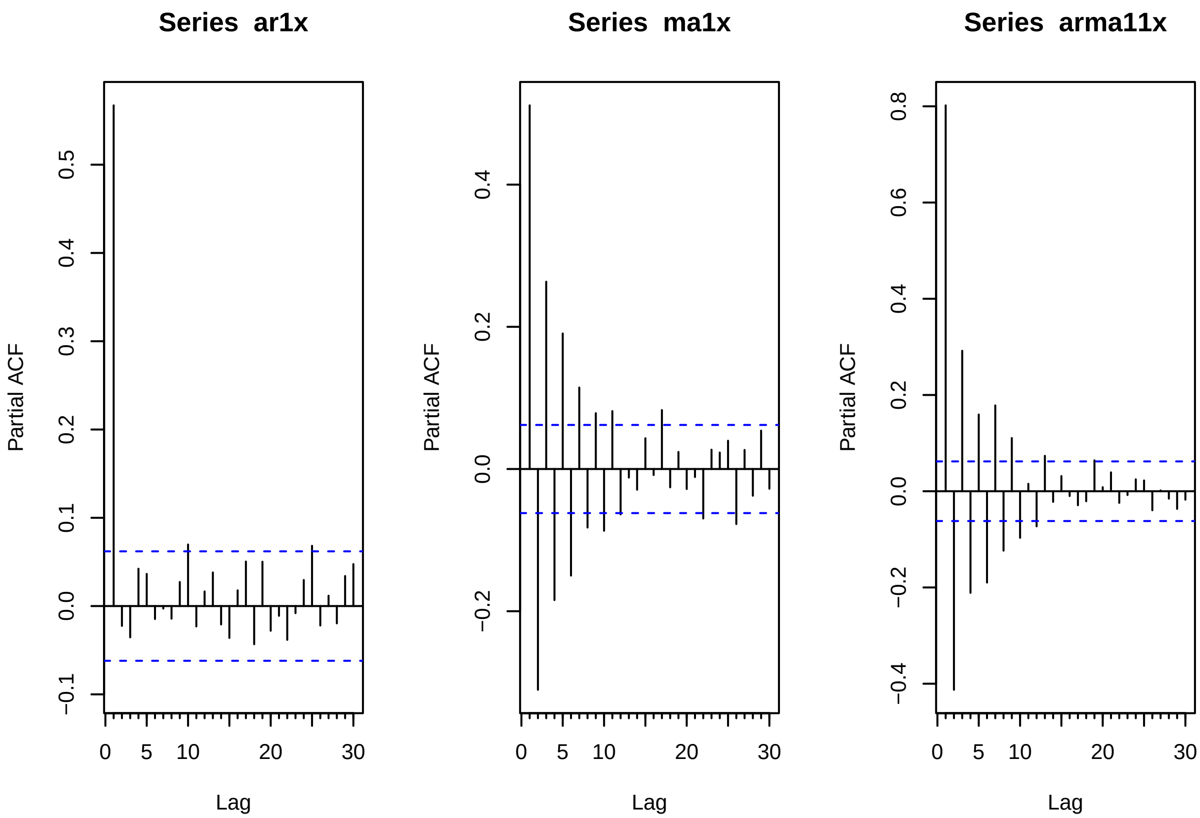
Answer: Yes they match.

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

```
#ARMA(1,0)
ar1x<- arima.sim(model=list(ar = 0.6), n = 1000)

#ARMA(0,1)
ma1x<- arima.sim(model=list(ma = 0.9), n = 1000)

#ARMA(1,1)
arma11x<- arima.sim(model = list(ar = 0.6, ma = 0.9), n = 1000)

par(mfrow=c(1,3))
Acf(ar1x)
Acf(ma1x)
Acf(arma11x)
```




```
par(mfrow=c(1,3))
Pacf(ar1x)
Pacf(ma1x)
Pacf(arma11x)
```



Answer: The AR(1) has slow decay again in the ACF plot and a very obvious cut off at the first lag in the PACF plot. The MA(1) has slow decay and a cut off at the first laf in the PACF plot. The ARMA(1,1) has a slow decay in both the ACF and the PACF plot.

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.

Answer: ARIMA(1,0,1)(1,0,0)[12]

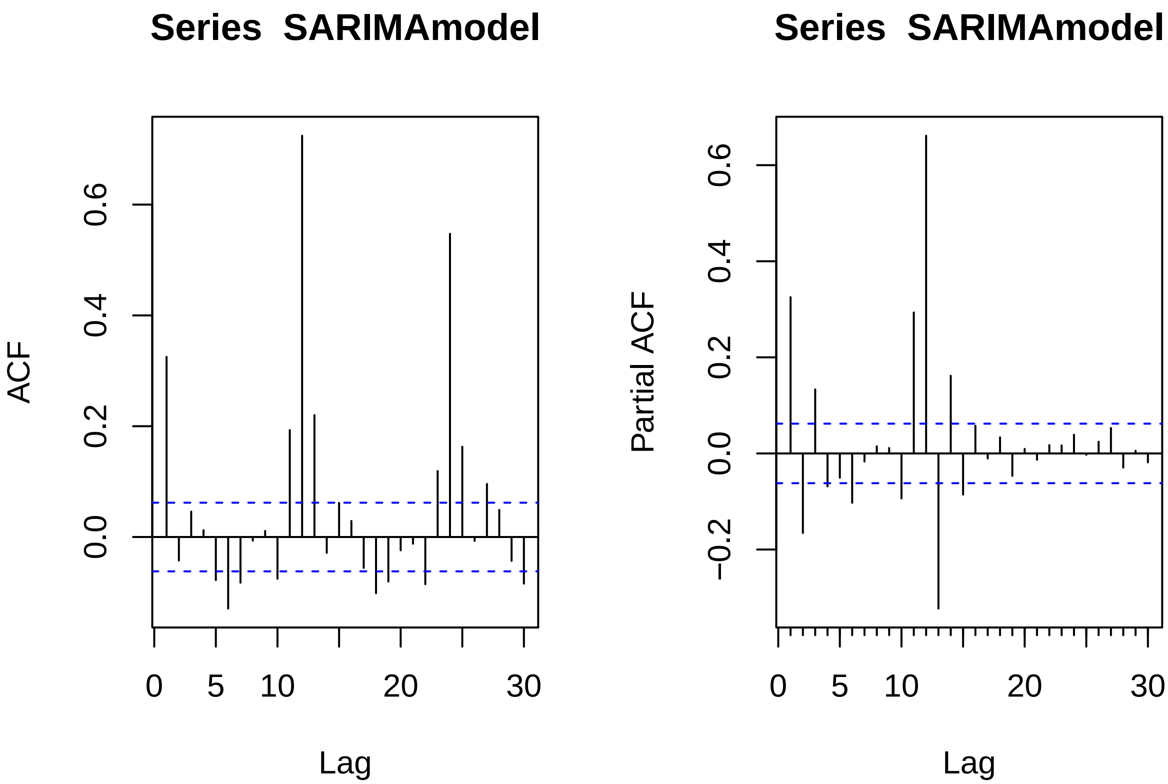
(b) Also from the equation what are the values of the parameters, i.e., model coefficients.

Answer: phi= 0.7, theta= -0.1, sar=-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.

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

par(mfrow=c(1,2))
Acf(SARIMAmodel)
Pacf(SARIMAmodel)
```



The plots represents the model well. There are more pronounced lags at every 12 months in the Acf plot. There is also a large positive spike at lag 12 in the Pacf plot. That shows a P of 1.