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

Assignment 6 - Due date 03/06/23

Maggie O'Shea

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
#install.packages("sarima")
library(sarima)
library(readxl)
library(ggplot2)
library(forecast)
library(tseries)
library(Kendall)
```

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) model the ACF will decay exponentially with time. The PACF will show cut offs which will help identify the order of the model – in this case there will be a cutoff at lag 2, with high correlation bars up to lag 2.

- MA(1)

Answer: An MA(1) model will have an ACF that cuts off at lag 1, which because the ACF starts at 0 would be the second line. The PACF will decay exponentially (similar to the ACF for the AR model).

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

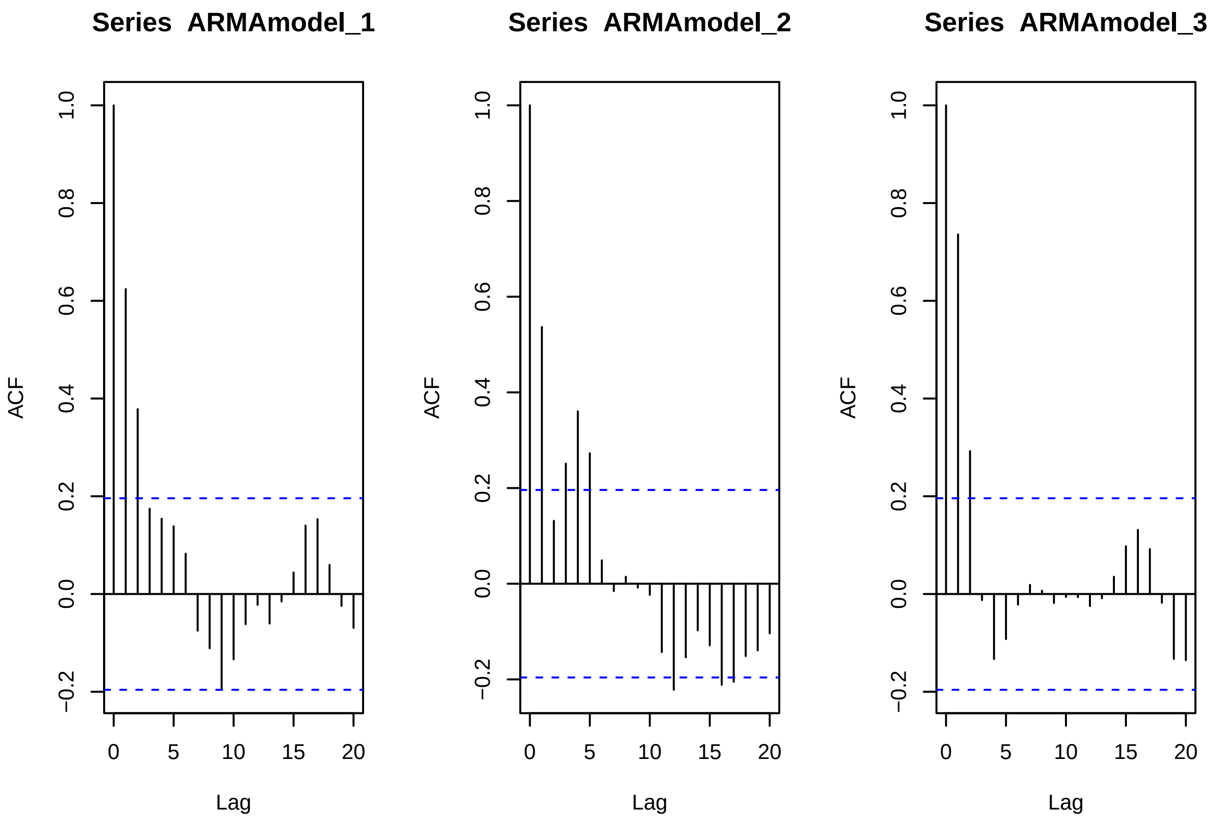
```
#ARMA(1,0)
ARMAmodel_1<- arima.sim(model=list(ar=0.6), n=100) #the AR coefficient is 0.6

#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)
```

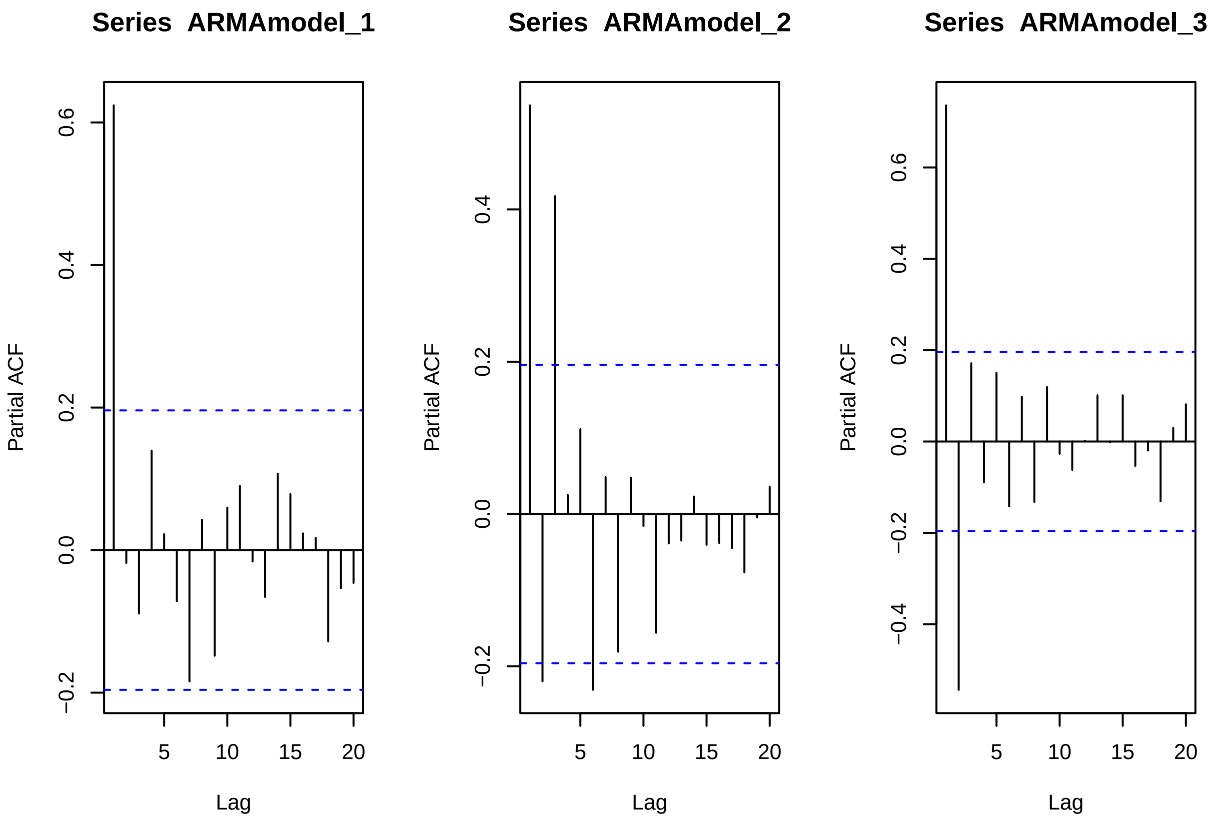
(a) Plot the sample ACF for each of these models in one window to facilitate comparison (Hint: use command $par(mfrow = c(1, 3))$ that divides the plotting window in three columns).

```
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: I think I would be able to identify the AR and MA model, but likely have trouble with the ARMA model. The ACF and PACF for the AR model match the descriptions of the expected plots well - the ACF has a slow decay and the PACF has a cutoff at lag 1, indicating order 1. Similarly the MA has the expected cut off at lag 1, and the PACF decays as well, making this clear to identify. The final model the ARMA(1,1) I think would be more difficult because it includes components of both. There is somewhat of a cutoff at lag 1 in the PACF but also it appears to decay after that. The ACF is even less clear as it appears to decay like an AR model.

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

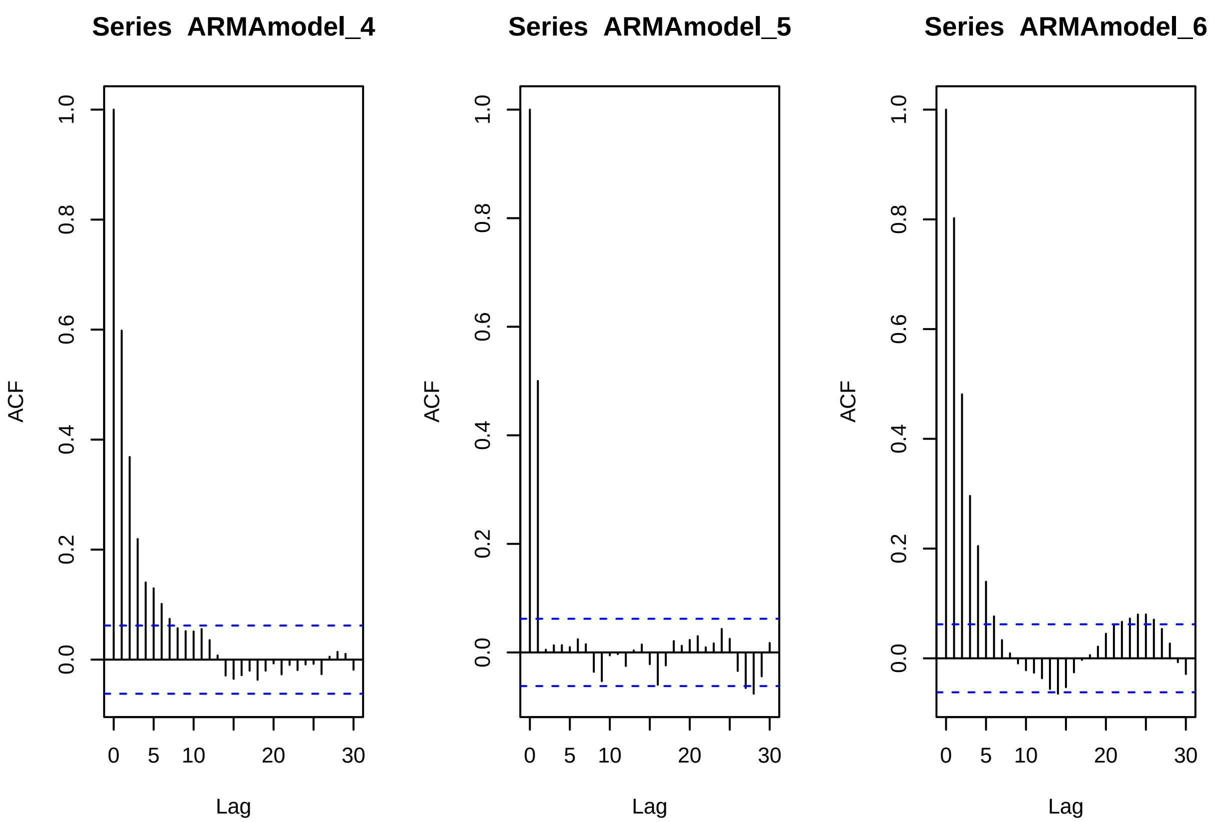
Answer: The phi is 0.6 for the AR model and the PACF at lag 1 shows a value >0.6 not matching what I expected. Because the model is of order 1, I expected the phi of 0.6 to be represented at lag 1 in the PACF. In addition, the MA and ARMA models do not show values at the lags that match the input phi and theta but this is more consistent with what I expected because the coefficients for the MA model are not displayed on the ACF/PACF.

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

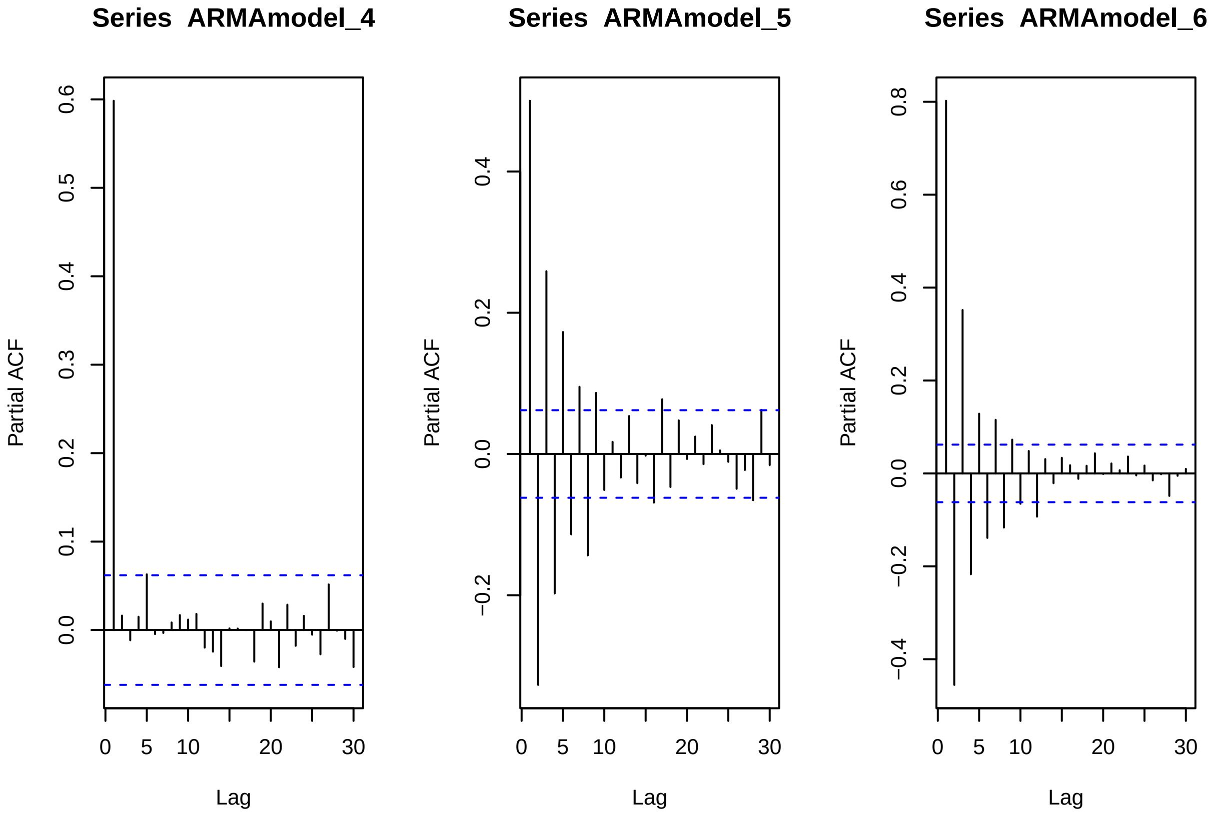
```
#ARMA(1,0)
ARMAmodel_4<- arima.sim(model=list(ar=0.6), n=1000) #the AR coefficient is 0.6
#ARMA(0,1)
```

```r
ARMAmodel_5<- arima.sim(model=list(ma=0.9), n=1000) #the MA coefficient is 0.9
#ARMA(1,1)
ARMAmodel_6<- arima.sim(model=list(ar=0.6, ma=0.9), n=1000)

#ACFs
par(mfrow=c(1,3))
acf(ARMAmodel_4)
acf(ARMAmodel_5)
acf(ARMAmodel_6)
```




```r
#PACFs
par(mfrow=c(1,3))
pacf(ARMAmodel_4)
pacf(ARMAmodel_5)
pacf(ARMAmodel_6)
```



(e-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: Again, I think I would have an easier time identifying the MA and AR because they more clearly show the patterns I expect. The AR ACF has a slow decay and the MA ACF has a cutoff after lag 1. The AR PACF has a very clear cutoff at lag 1, and the MA PACF has a very obvious slow decay. The ARMA(1,1) would be more difficult to identify, and I may mis-identify this model as an MA model because it shows a similar decaying PACF plot, though it also has a clearly decaying ACF which would indicate an AR. Because of these different characteristics, perhaps I would be able to identify it is an ARMA model because it has qualities of both AR and MA.

(e-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: In this case, the AR phi is 0.6 and the first lag in the PACF appears to be close to that value. I expect phi at lag 1 to be visible in the PACF which appears to be true in this case. The theta for the MA model I did not expect to find in the ACF/PACF and do not see the 0.9 represented in either of these.

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(1,2,1)(1, 0, 0). I think that p=1 because there is one autoregressive term (0.7 * y_{t-1}), q = 1 because there is one MA term (+a_t-0.1 * a_{t-1}). d=2 because there is no constant in the expression, and if d>1 then no constant is allowed. Finally, P=1 because there is only one seasonal term (-0.25*y_{t-12}) and it is a seasonal autoregressive term.

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

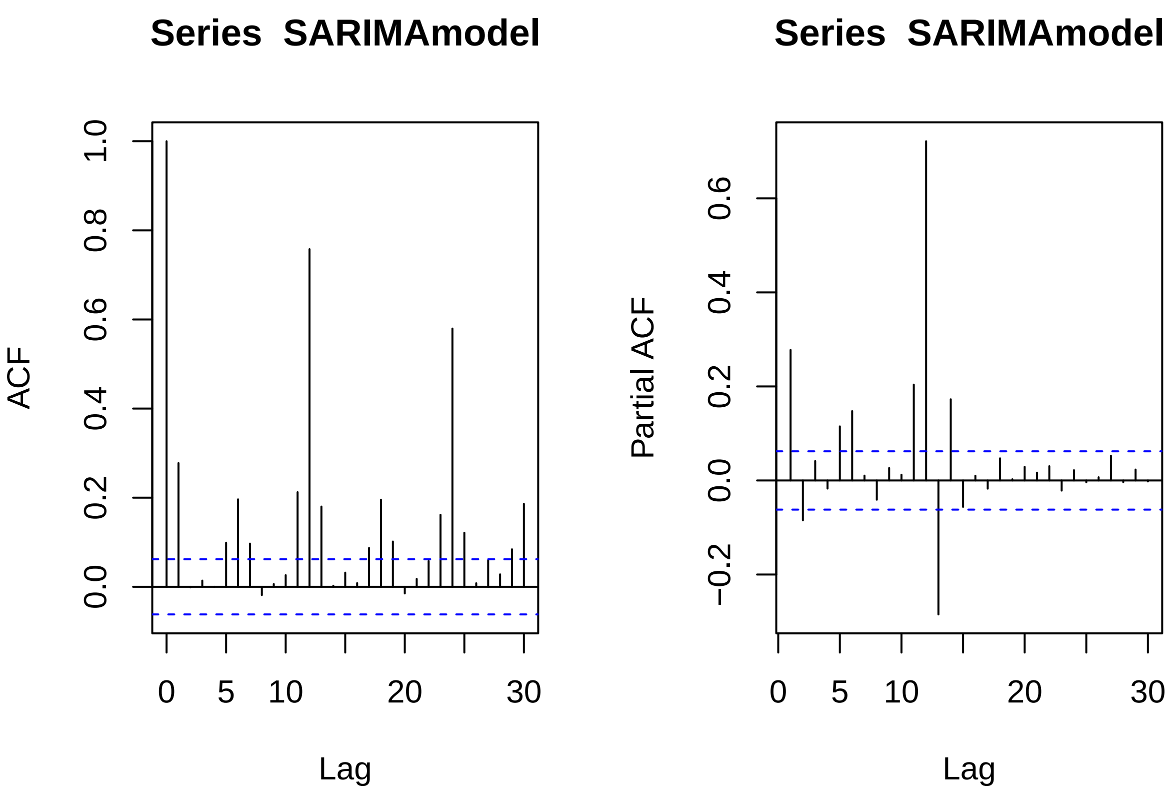
- MA term = 0.1
- AR term = 0.7
- SAR term = -0.25

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.

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

par(mfrow=c(1,2))
acf(SARIMAmodel)
pacf(SARIMAmodel)
```



> Answer: Looking at the non-seasonal component, I see a cut off at lag 1 rather than a slow decay which

suggests an MA process, with an order of 1. The PACF has somewhat of a decay which is expected for the MA process, validating this finding from the ACF that it is an MA model. Now examining the plots for the seasonal component, I notice the ACF shows spikes at seasonal lags 12 and 24, and the PACF has a single spike at lag 12. The multiple spikes in the ACF suggests that its an SAR process, and the PACF single spike also suggests it is a SAR process.