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

Assignment 6 - Due date 03/16/22

Sarah Mansfield

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_A06_Sp22.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(forecast)
library(tseries)
library(sarima)
```

Q1

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

(a) AR(2)

> Answer: For an AR(2) model, the current value depends on $p = 2$ previous values and the ACF plot will exhibit slow decay over time. The PACF plot decides the order of the process, therefore we would expect to see the plot cut off after lag 2.

(b) MA(1)

> Answer: For a MA(1) model, the current deviation depends on $q = 1$ previous deviations and the PACF plot will exhibit slow decay over time. The ACF plot decides the order of the process, therefore we would expect to see the plot cut off after lag 1.

Q2

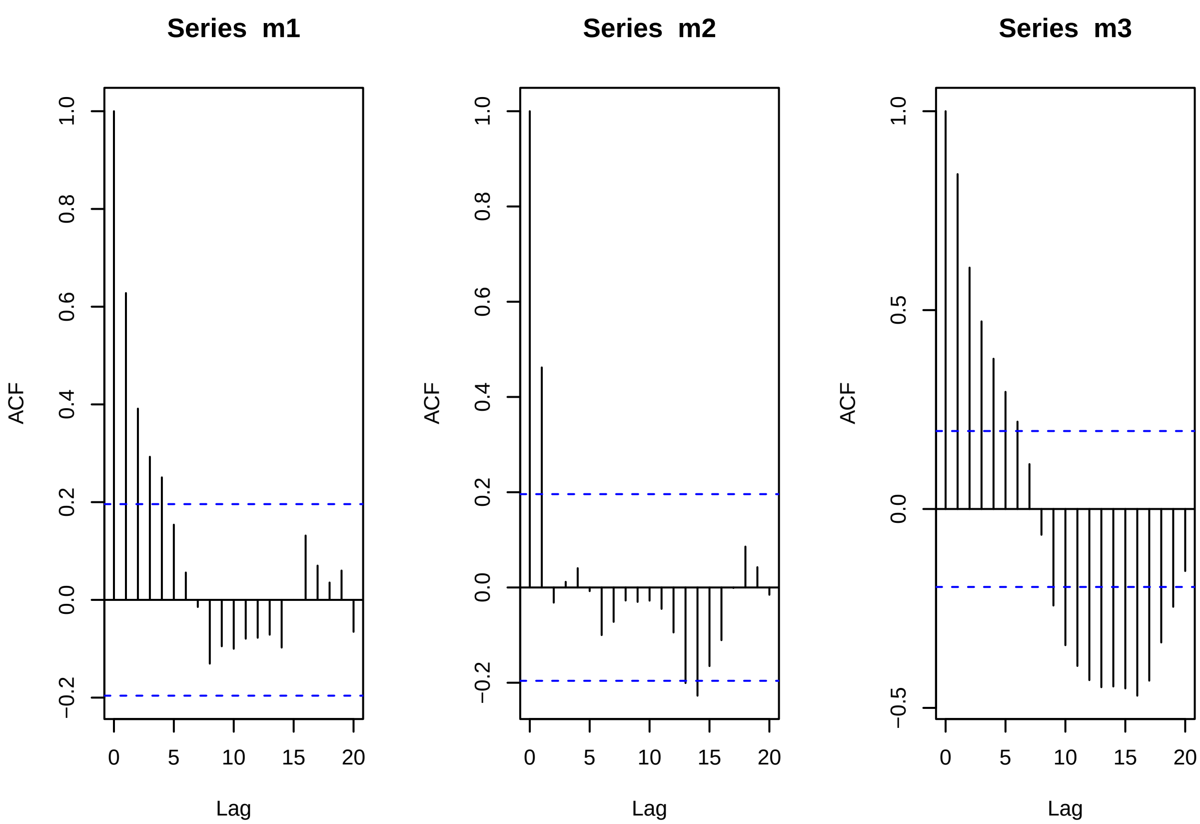
Recall that the non-seasonal ARIMA is described by three parameters ARIMA(p, d, q) where p is the order of the autoregressive component, d is the number of times the series need to be differenced to obtain stationarity and q is the order of the moving average component. If we don't need to difference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

```
set.seed(70)

m1 <- arima.sim(model = list(ar = 0.6), n = 100)
m2 <- arima.sim(model = list(ma = 0.9), n = 100)
m3 <- arima.sim(model = list(ar = 0.6, ma = 0.9), n = 100)
```

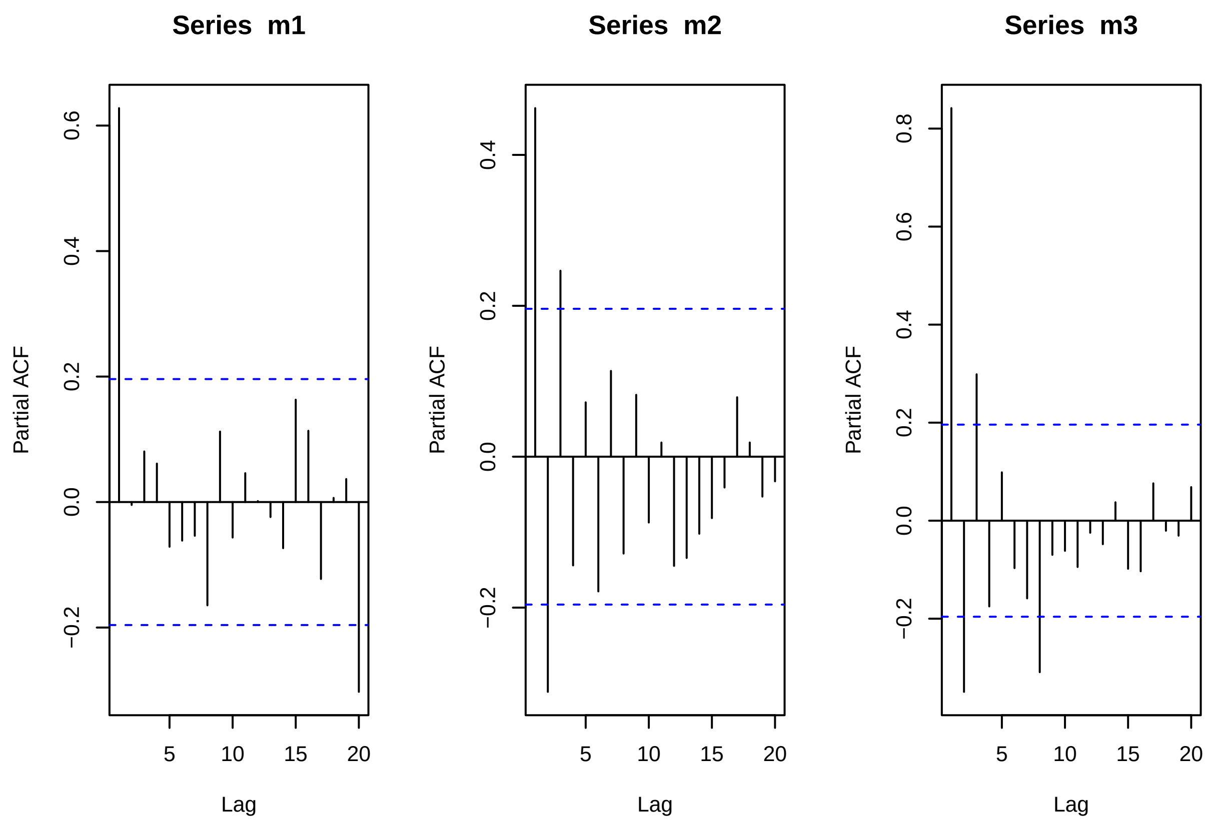
(a) Plot the sample ACF for each of these models in one window to facilitate comparison (Hint: use command $par(mfrow = c(1, 3))$ that divides the plotting window in three columns).

```
par(mfrow=c(1,3))
acf(m1)
acf(m2)
acf(m3)
```



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

```r
par(mfrow=c(1,3))
pacf(m1)
pacf(m2)
pacf(m3)
```



(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: We know that for an AR model, its ACF plot will show slow decay and the PACF plot decides the order based on where the lag cuts off. Since we see slow decay in the first ACF plot and see that the corresponding PACF plot cuts off after lag 1, we can identify the first model as an AR model with $p = 1$. Similarly, we know that for an MA model, its PACF plot will show slow decay and the ACF plot decides the order based on where the lag cuts off. Since the second pair of plots showcases this pattern with the ACF plot cutting off at lag 1, we can identify the second model as an MA model with $q = 1$. Lastly, for ARMA models we know that both the ACF and PACF plots will tend to be more vague on whether they are exhibiting slow decay or cutting off, so we can assume that the third model is an ARMA model. However, it is difficult to identify the order of the components since although the PACF plot seems to cut off after lag one (implying the AR component $p = 1$), the ACF plot does not have aclear cutoff so we cannot determine the order of the MA component.

(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 - we can see that at lag one for the AR model, both plots record a value a little bit more than 0.6, which matches up with the AR coefficient we set at the beginning.

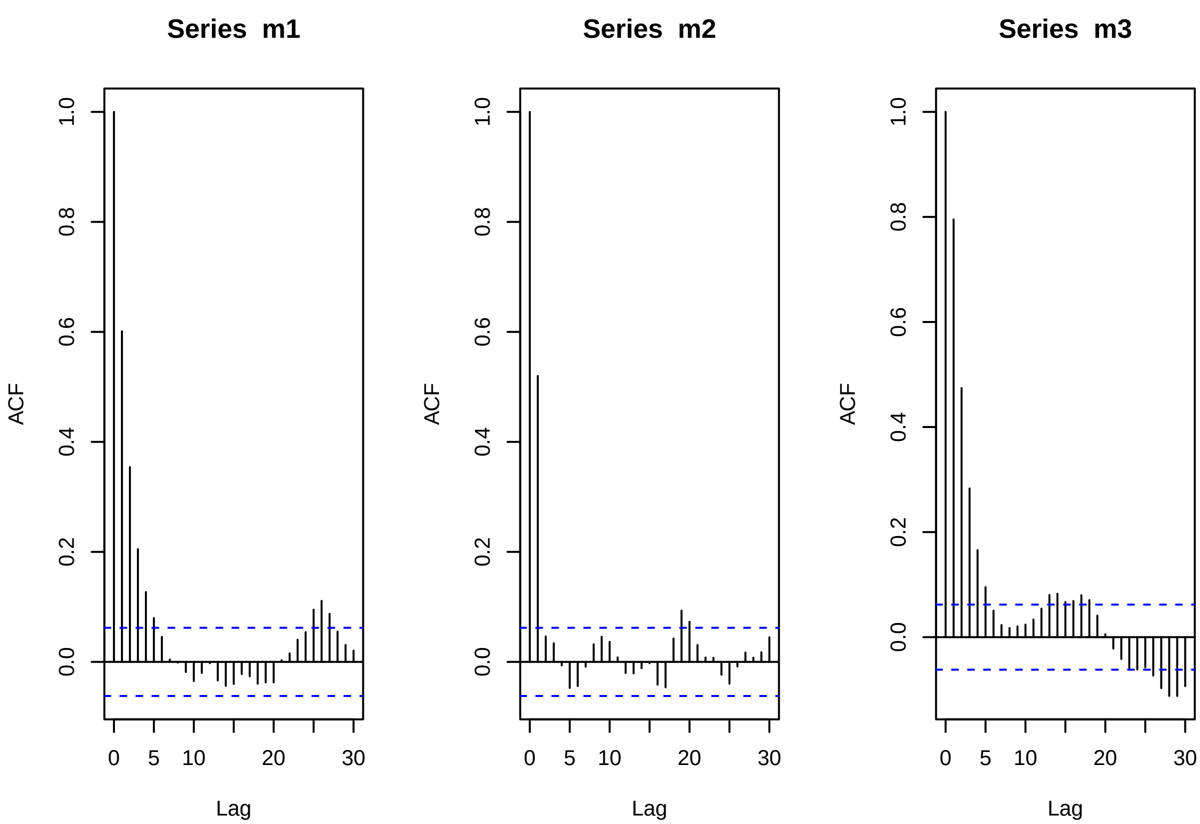
(e) Increase number of observations to $n = 1000$ and repeat parts (a)-(d).

```
set.seed(70)

m1 <- arima.sim(model = list(ar = 0.6), n = 1000)
m2 <- arima.sim(model = list(ma = 0.9), n = 1000)
m3 <- arima.sim(model = list(ar = 0.6, ma = 0.9), n = 1000)
```

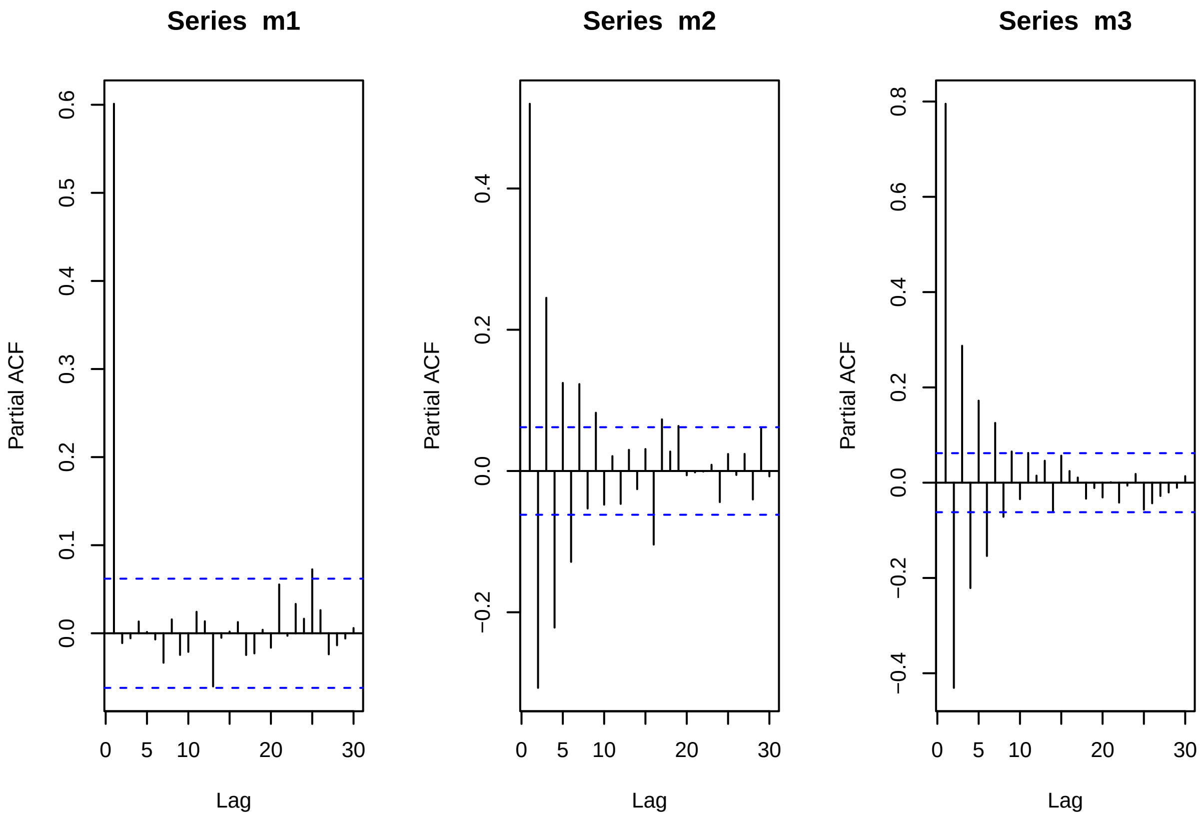
(i)

```
par(mfrow=c(1,3))
acf(m1)
acf(m2)
acf(m3)
```



(ii)

```
par(mfrow=c(1,3))
pacf(m1)
pacf(m2)
pacf(m3)
```



(iii) 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: We see essentially the same patterns as the previous part, though cutoffs seem more significant in comparison. Again, we know that for an AR model, its ACF plot will show slow decay and the PACF plot decides the order based on where the lag cuts off. Since we see slow decay in the first ACF plot and see that the corresponding PACF plot cuts off after lag 1, we can identify the first model as an AR model with $p = 1$. Similarly, we know that for an MA model, its PACF plot will show slow decay and the ACF plot decides the order based on where the lag cuts off. Since the second pair of plots showcases this pattern with the ACF plot cutting off at lag 1, we can identify the second model as an MA model with $q = 1$. Lastly, for ARMA models we know that both the ACF and PACF plots will tend to be more vague on whether they are exhibiting slow decay or cutting off, so we can assume that the third model is an ARMA model. However, it is difficult to identify the order of the components since although the PACF plot seems to cut off after lag one (implying the AR component $p = 1$), the ACF plot does not have aclear cutoff so we cannot determine the order of the MA component.

(iv) 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 - we can see that at lag one for the AR model, both plots record a value right around 0.6, which matches up with the AR coefficient we set at the beginning.

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: $\phi_1 = 0.7$ (AR term), $\theta_1 = 0.1$ (MA term), $\phi_{12} = -0.25$ (SAR term)

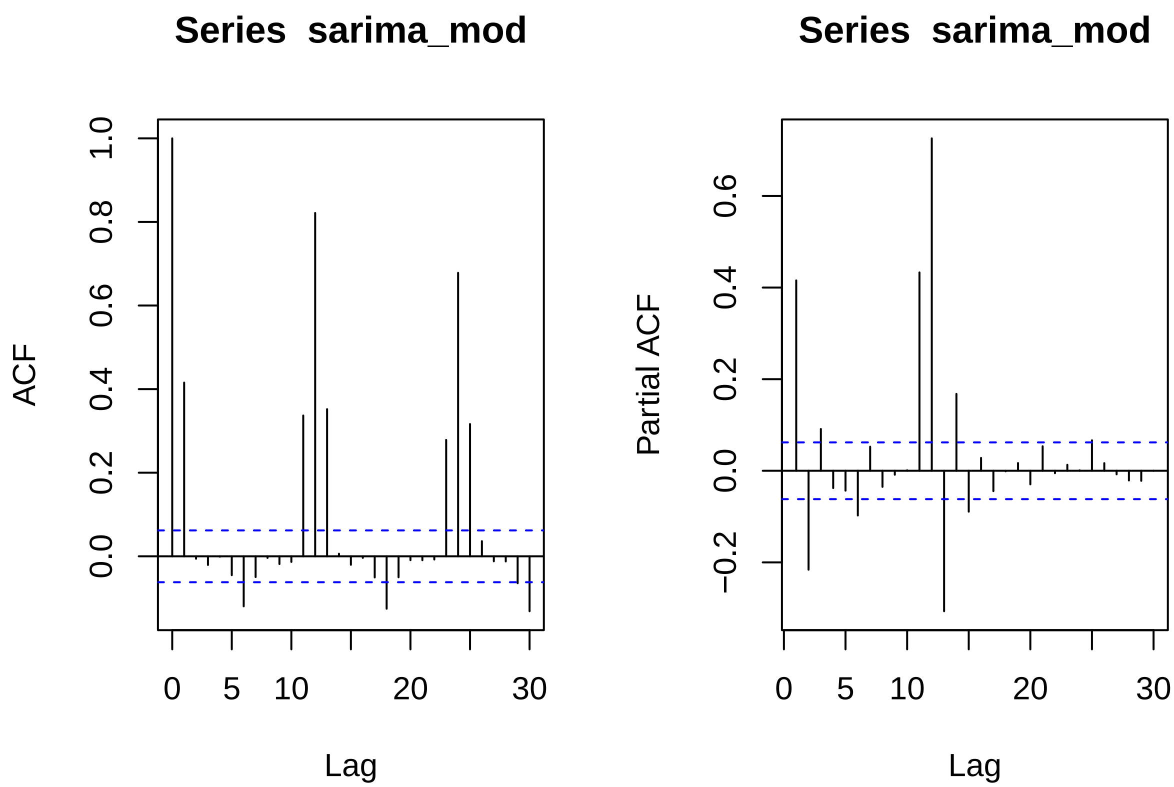
Q4

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
set.seed(70)

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

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



Answer: From both plots it's easy to identify the order of the seasonal component as $P = 1$ since we see positive spikes in ACF at lags that are a multiple of 12, as well as a single positive spike in PACF at lag 12. To determine the non-seasonal component, we can look at the behavior of non-seasonal lags and see that the ACF plot cuts off at lag 1, which indicates an MA process with $q = 1$. However, looking at the PACF plot it's difficult to tell what the order of the AR process might be since we see several spikes in the plot and a particular cutoff isn't clear.