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

Assignment 5 - Due date 03/12/21

Keyang Xue

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(forecast)
library(tseries)
library(astsa)
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: the ACF will decay exponentially with time while the PACF will change sharply (cut off) after lag 2 and determine the order of the AR model.

(b) MA(1)

> Answer: the ACF will cut off after lag 1 while the PACF will decay exponentially with time.

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)
```

```
sim4<-arima.sim(list(order = c(1,0,0), ar = 0.6),n = 1000)
sim5<-arima.sim(list(order = c(0,0,1), ma = 0.9),n = 1000)
sim6<-arima.sim(list(order = c(1,0,1), ar = 0.6, ma = 0.9),n = 1000)
```

(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).

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

(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.

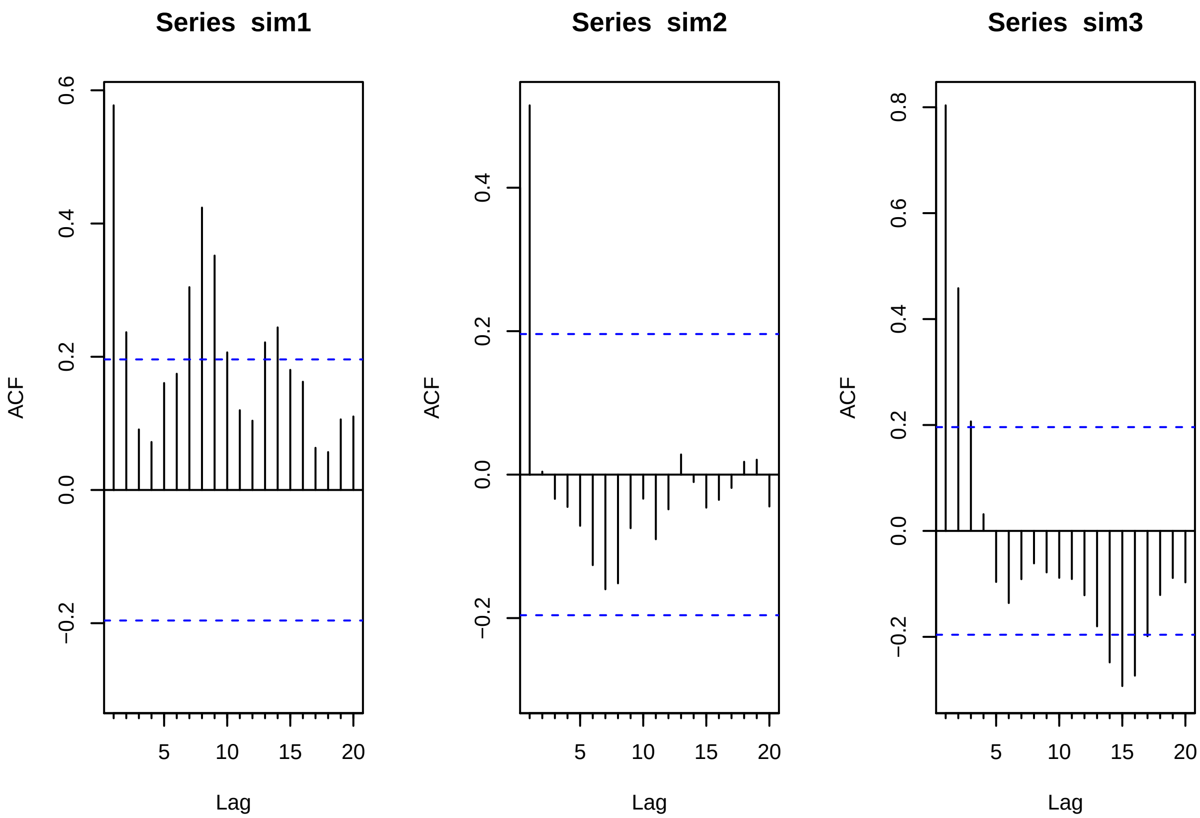
> Answer:since the simulation generates random values for each model everytime, it is hard to get an ACF plot that decays exponentially and a PACF plot that cuts off after lag 1 for ARMA(1,0), an ACF plot that cuts off after lag 1 and a PACF plot that decays exponentially for ARMA(0,1), and an ACF plot and a PACF plot that both decay exponentially for ARMA(1,1).

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

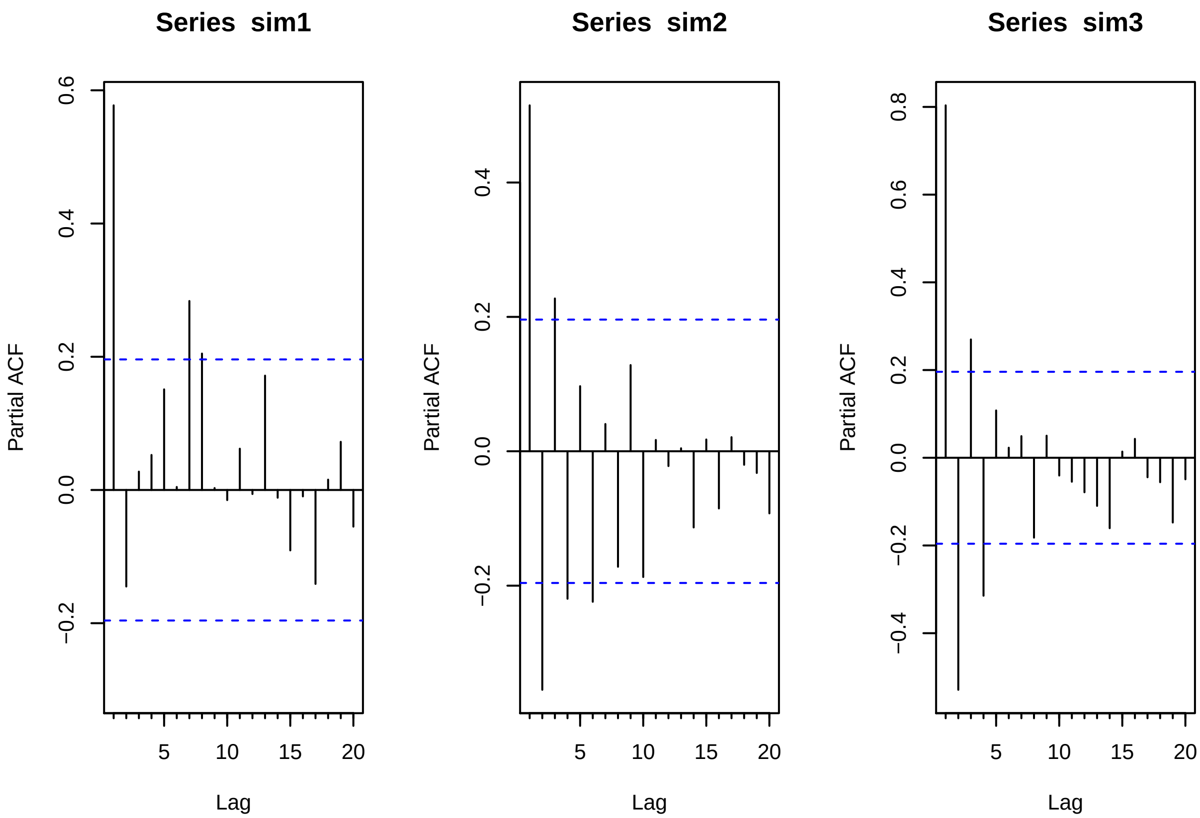
> Answer: the PACF matches the phi=0.6 while the theta=0.9 does not match the ACF because the phi represents the autocorrelation while the theta represents the relationship among yt and at-1.

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

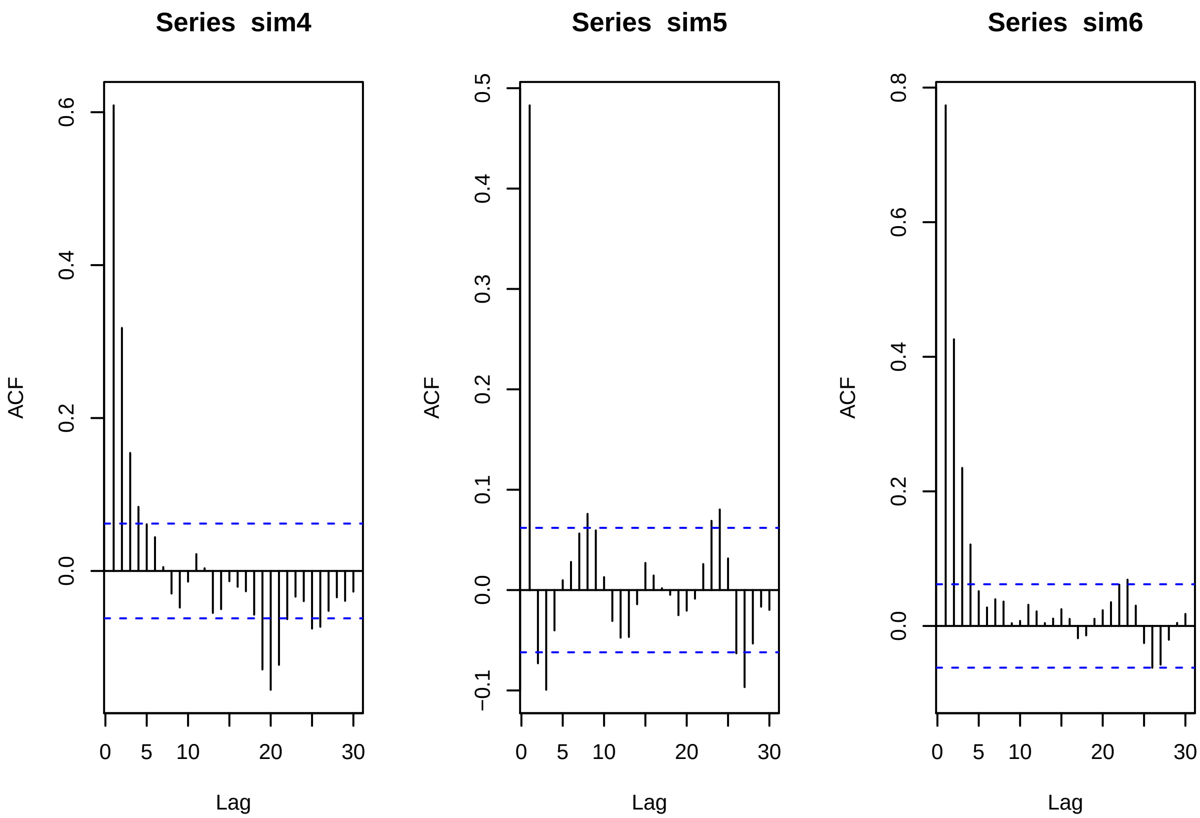
```
par(mfrow=c(1,3))
sim1acf <- Acf(sim1)
sim2acf <- Acf(sim2)
sim3acf <- Acf(sim3)
```



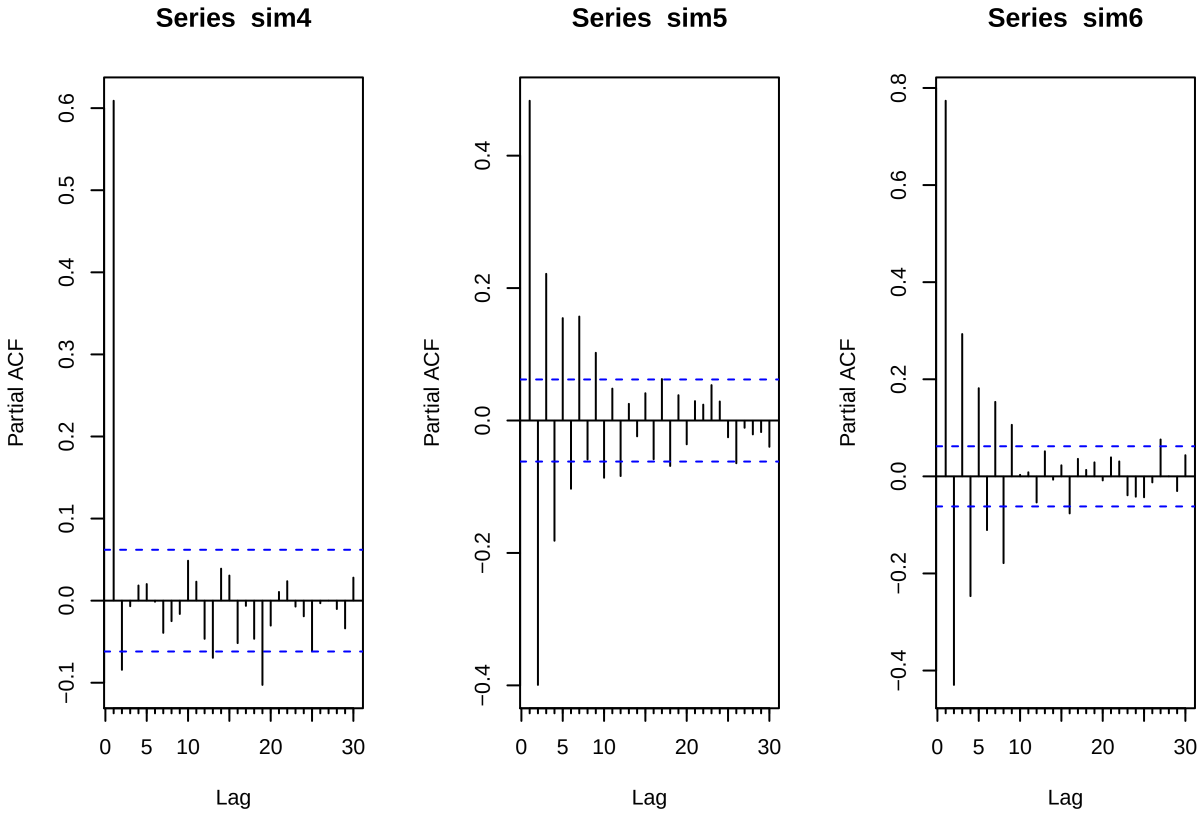

```
par(mfrow=c(1,3))
sim1pacf <- Acf(sim1,type = "partial")
sim2pacf <- Acf(sim2,type = "partial")
sim3pacf <- Acf(sim3,type = "partial")
```




```
par(mfrow=c(1,3))
sim4acf <- Acf(sim4)
sim5acf <- Acf(sim5)
sim6acf <- Acf(sim6)
```




```r
sim4pacf <- Acf(sim4, type = "partial")
sim5pacf <- Acf(sim5, type = "partial")
sim6pacf <- Acf(sim6, type = "partial")
```



(c)since the simulation generates random values for each model everytime, it is hard to get an ACF plot that decays exponentially and a PACF plot that cuts off after lag 1 for ARMA(1,0), an ACF plot that cuts off after lag 1 and a PACF plot that decays exponentially for ARMA(0,1), and an ACF plot and a PACF plot that both decay exponentially for ARMA(1,1). (d)the PACF matches the phi=0.6 better than the previous case when n=100. The theta=0.9 still does not match the ACF because the phi represents the autocorrelation while the theta represents the relationship among yt and at-1.

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.

$p = 1, d = 0, q = 1, P = 1, D = 0, Q = 0$

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

$\phi_1 = 0.7$, $\theta_1 = 0.1$,$\phi 12 = -0.25$

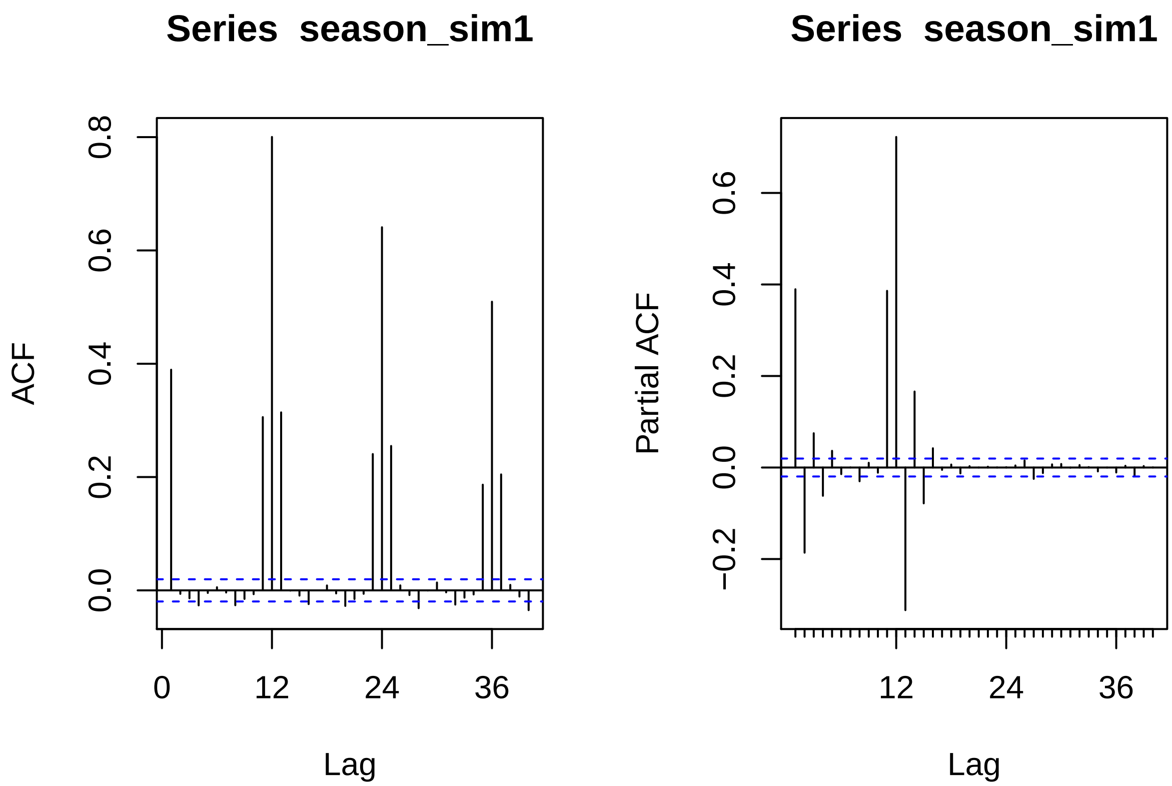
Q4

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.

```
par(mfrow=c(1,2))
season_sim1<-sarima.sim(ar = NULL, d = 0, ma = 0.5, sar = 0.8, D = 0, sma = NULL,S = 12, n = 10000)

season_sim1acf <- Acf(season_sim1)

season_sim1pacf <- Acf(season_sim1, type = "partial")
```



The order of non-seasonal component would not be identified through these graphs since the ACF for non-seasonal lags shows a cut-off after lag 1, indicating this is a MA process while the PACF also shows a cut-off after lag 1, indicating an AR process. The order of seasonal components are identifiable through these plots - there are multiple spikes at seasonal lags in the ACF plot and one single spike at seasonal lags in the PACF plot, indicating there is a SAR process.