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

Assignment 5 - Due date 03/12/21

Yash Doshi

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.

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)

AR is an autoregressive model, meaning we perform regression on previous observations. The generalize the idea of regression to represent the linear dependence between a dependent variable and an explanatory variable. If we have two previous observations, then we say that we are working with 2nd order autoregressive process, or AR 2 model. The order is related to how many previous observations you have on your regression, or how many previous observations do we need to have a good representation of Yt (our independent variable).

AR models show a slow decay on ACFs. Slow decay means how long does it take to reach 0 in ACF plot. In simpler terms, if the values go up, they will stay up for sometime. They will not drop from a positive number to a negative number all of a sudden. We can say that AR models have a relatively long memory. Since AR models decay exponentially with time, it means that there is some relationship between the current observation and previous observations.

The PACF plots will help us in identifying the order of the AR model. PACF will cut-off at the order of the AR model. For instance, if the order is 1, then anything after lag two will be insignificant. There is only one significant correlation with previous observations.

I would also like to point out that AR(2) means that p = 2. The 2 in brakets give us the order of the model. In this case it is 2, which means that the order of this AR model will be 2. This will also be reflected in the PACF plot for the AR model. The cut off will be at lag 2, so the previous two terms, yt-1 and yt-2, will be needed to represent the series.

MA(1)

MA models have short to no memory. Therefore, its plots will have high fluctuations. For instance, it can go from -3 to 6 to -10 in three consecutive time periods. MA components can have an order higher than 1.

For MA models, ACFs will identify the order of the model. Unlike AR models, there will be no slow decay for MA terms in the ACF. However, PACF of MA models will have a slow, exponential decay.

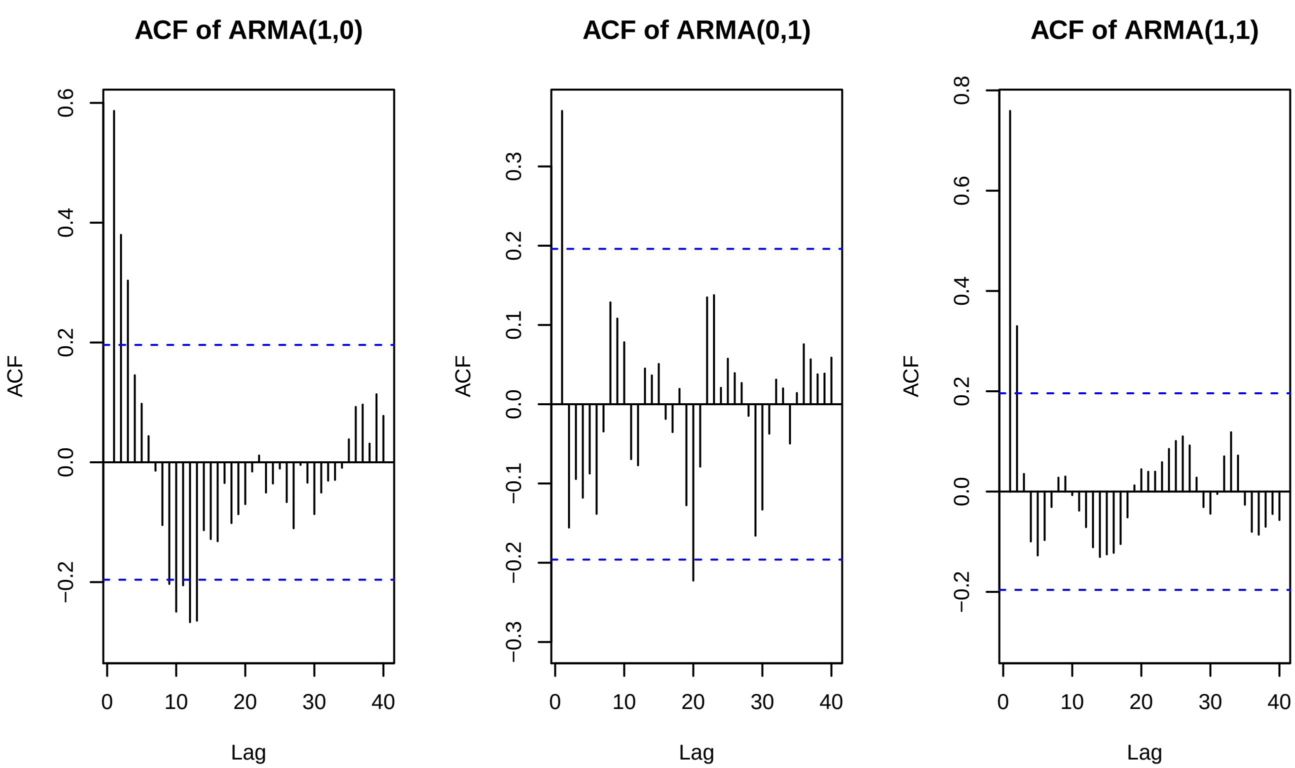
I would also like to point out that MA(1) means that q = 1. The 1 in brakets give us the order of the model. In this case it is 1, which means that the order of this MA model will be 1. This will also be reflected in the ACF plot for the MA model. There will be a cut off at lag 1. Meaning, yt-1 will be needed to represent the series.

If stationary models have a negative autocorrelation at lag 1, MA terms work the best.

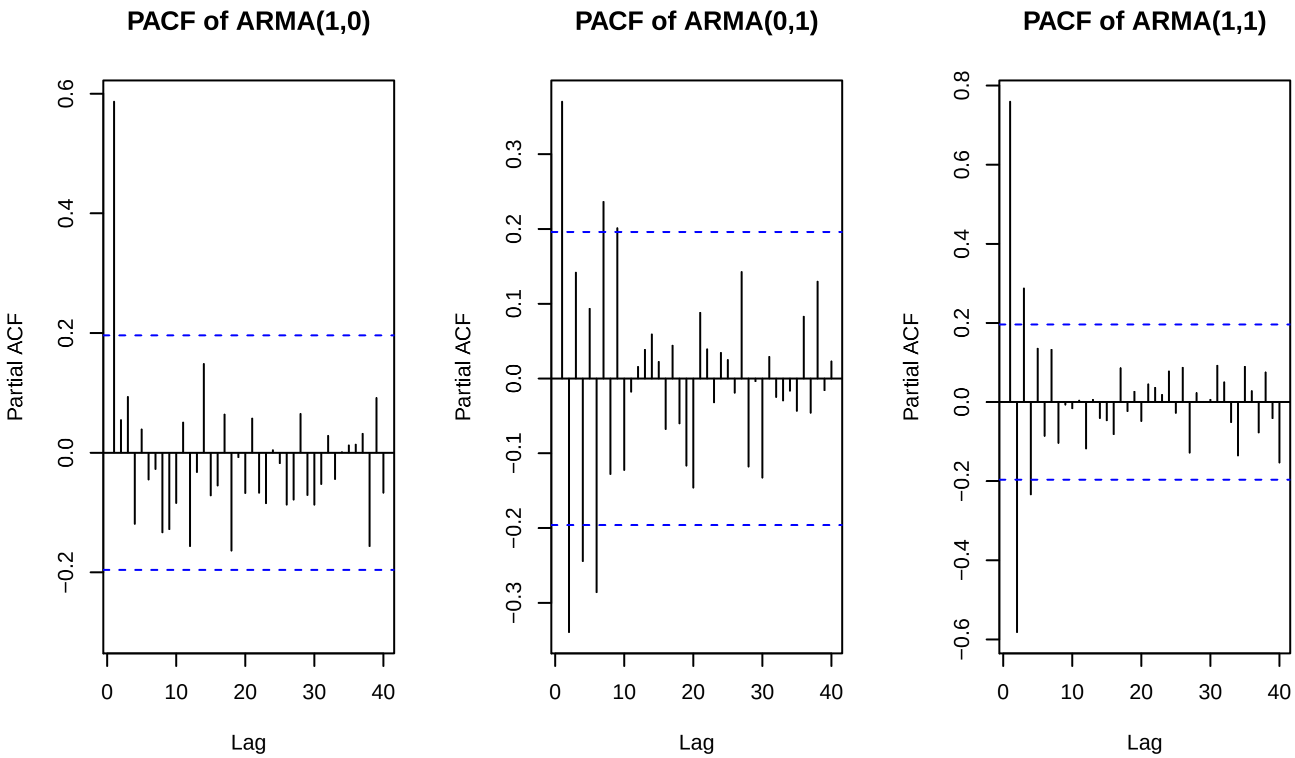
Q2

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

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).



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



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.

Model 1 - By looking at the ACF and PACF plots of model 1 (left plot), we can see that the PACF is positive at lag 1 and then it drops to negative (within the limits) at lag 2. It means that it is PACF where the cut-off is happening. Furthermore, we can also notice a slow decay in ACF plot. This proves that the model is an AR model. Since this is an AR model, we can look at the PACF plot to know the order. In this model we can see that the cut-off is at lag 1. It means that the order is 1. AR(1)

Model 2- Model 2 is completely opposite of Model 1. In this model, we can see a clear cut-off in the ACF plot. Furthermore, we can also see a slow decay in the PACF plot of this model. Therefore, this makes it clear that this is an MA model. For the order, we can look at the ACF plot. It can be seen from the plot that after lag 1 all the other lines are within the blue lines. It means that the order of this model is 1. MA(1).

Model 3- For this model we will not be able to correctly identify the model. Since it is given in the question that this is an ARMA model with order (1,1), we are able to tell. Without that it would be difficult to tell that this is an ARMA model. The ACF and PACF of the ARMA process is the result of superimposing the AR and MA properties. Therefore, it is difficult to say that this is an ARMA model. If we only looked at the plot, we would have said that this is an MA model with an order of 2. This is because the cut off is at 2 in the ACF plot, and the PACF plot shows a slow decay.

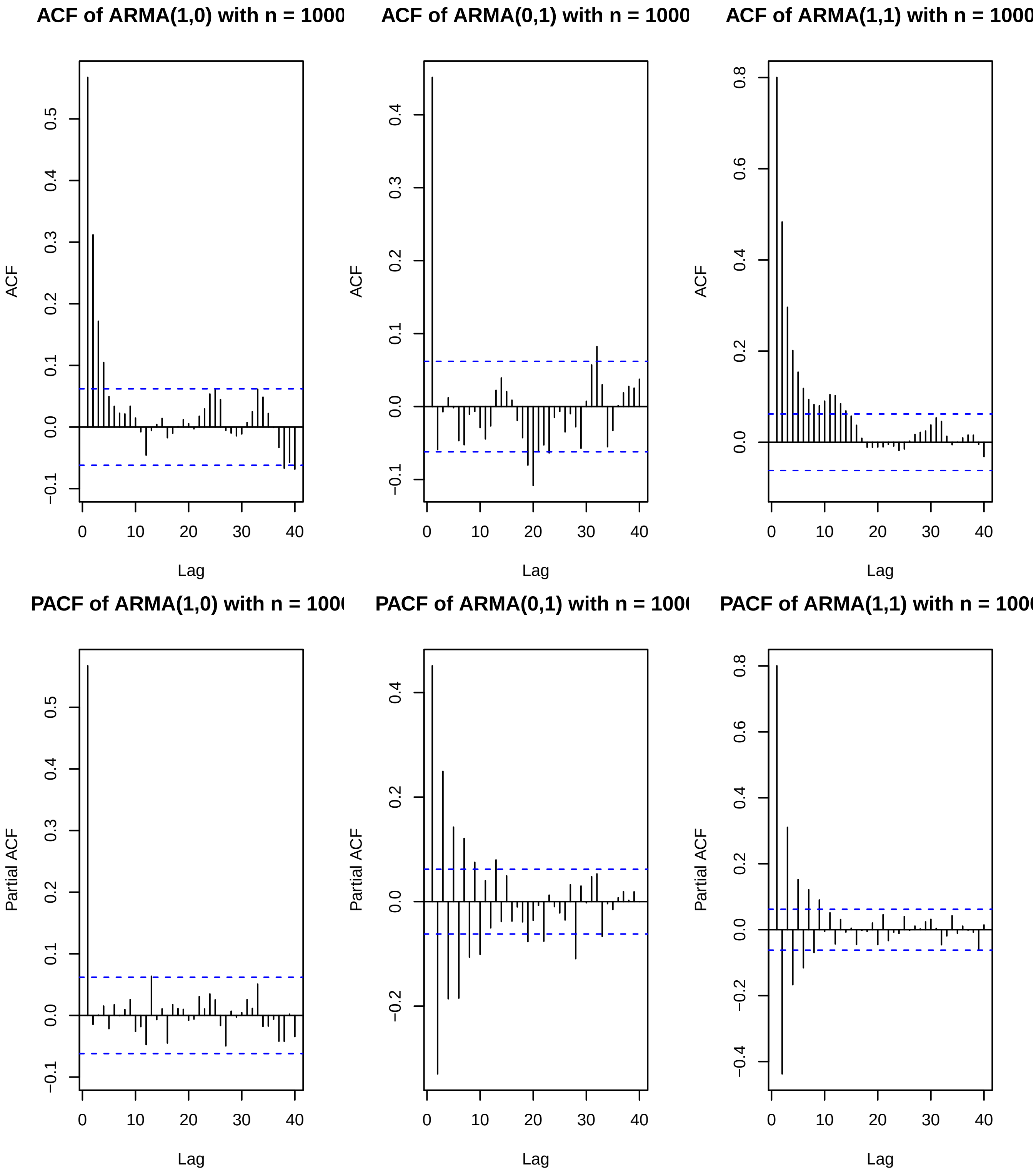
Compare the ACF and PACF values R computed with the theoretical values you provided for the coefficients. Do they match? Explain your answer.

Model 1- Since we know that model 1 is an AR model, we know that we will be looking at Phi in the PACF plot. The theoretical value of Phi is 0.6. In the plots we can see that it is 0.58, which is very close to 0.6. Therefore, we can conclude that the plotted values are almost matching the theoretical values.

Model 2- In this MA model, we can see that our theoretical value (theta = 0.9) does not match with the ACF plot. The reason why we are considering ACF is because we are looking at the MA model, and for MA models, the order is identified by the ACF plots. The reason why the theoretical value and the value obtained from the graph do not match is because this is a moving average model. It is not an autocorrelation model. It uses a moving average method. Therefore, the values will not match.

Model 3- This is a combination of autoregressive and moving average models. Therefore, the ARMA model itself is difficult to identify. In this case, by looking at the plots it is clear that the values do not match the theoretical values. ARMA models also have a moving average property, because of whichs the plotted ACF and PACF values will not match the theoretical values.

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



Comparison of all three models

Model 1 - By looking at the ACF and PACF plots of model 1 (left plot), we can see that the PACF is positive at lag 1 and then it drops within the blue line. It means that it is PACF where the cut-off is happening. Furthermore, we can also notice a slow decay in ACF plot. This proves that the model is an AR model. Since this is an AR model, we can look at the PACF plot to know the order. In this model we can see that the cut-off is at lag 1. It means that the order is 1. AR(1)

Model 2- Model 2 is completely opposite of Model 1. In this model, we can see a clear cut-off in the ACF plot. Furthermore, we can also see a slow decay in the PACF plot of this model. Therefore, this makes it clear that this is an MA model. For the order, we can look at the ACF plot. It can be seen from the plot that after lag 1 all the other lines are within the blue lines. It means that the order of this model is 1. MA(1).

Model 3- For this model we will not be able to correctly identify the model. Since it is given in the question that this is an ARMA model with order (1,1), we are able to tell. Without that it would be difficult to tell that this is an ARMA model. The ACF and PACF of the ARMA process is the result of superimposing the AR and MA properties. Therefore, it is difficult to say that this is an ARMA model. If we only looked at the plot, we would have said that this is an MA model with an order of 2. This is because the cut off is at 2 in the ACF plot, and the PACF plot shows a slow decay.

Compare the ACF and PACF values R computed with the theoretical values provided for the coefficients

Model 1- Since we know that model 1 is an AR model, we know that we will be looking at Phi in the PACF plot. The theoretical value of Phi is 0.6. In the plots we can see that it looks like it is at 0.6. Therefore, we can conclude that the plotted values are matching the theoretical values.

Model 2- In this MA model, we can see that our theoretical value (theta = 0.9) does not match with the ACF plot. The reason why we are considering ACF is because we are looking at the MA model, and for MA models, the order is identified by the ACF plots. The reason why the theoretical value and the value obtained from the graph do not match is because this is a moving average model. It is not an autocorrelation model. It uses a moving average method. Therefore, the values will not match.

Model 3- This is a combination of autoregressive and moving average models. Therefore, the ARMA model itself is difficult to identify. In this case, by looking at the plots it is clear that the values do not match the theoretical values. ARMA models also have a moving average property, because of whichs the plotted ACF and PACF values will not match the theoretical values.

Q3

Consider 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

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

p = 0.7

d = 0

q = 0.1

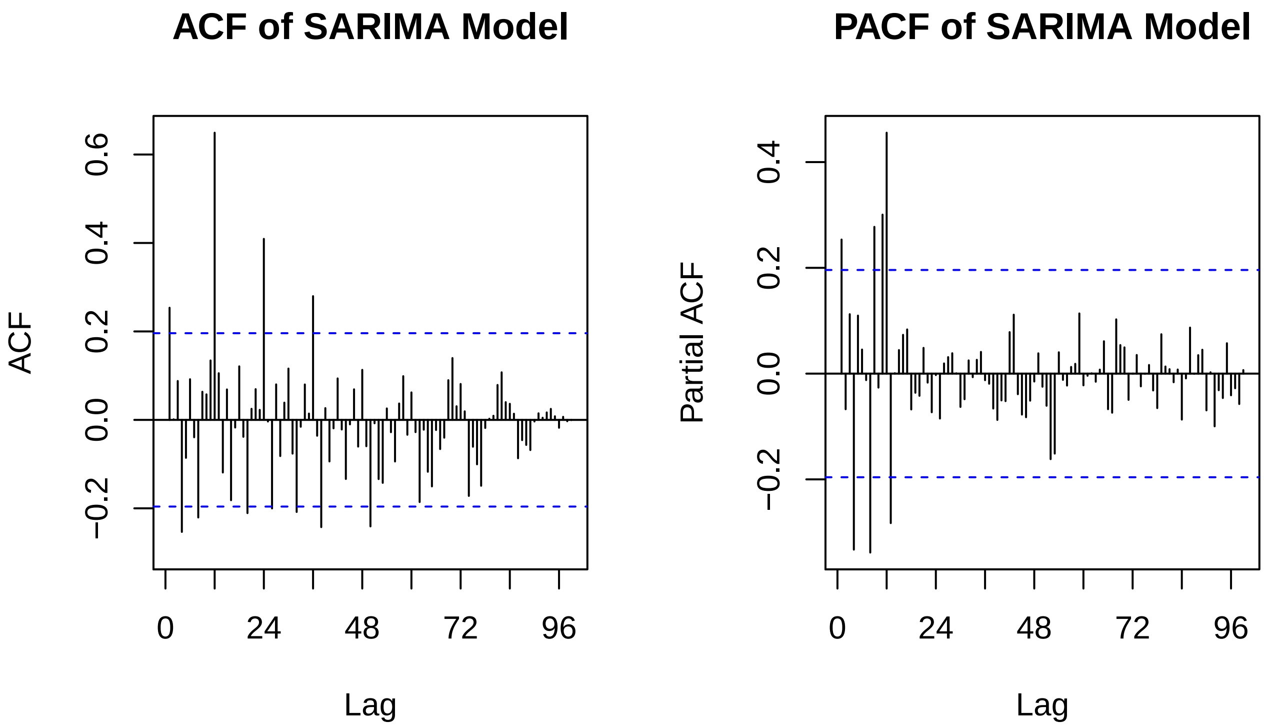
P = -0.25

D = 0

Q = 0

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.



It is clear from the above plots that they are not well representing the model that I have simulated. From the ACF we can see that the cut-off is at 1. It means that the model is MA. If we now look at the PACF, we find that it also has a lag of 1. This becomes confusing if someone wants to identify the model and the order based on just ACF and PACF. Furthermore, had it not been mentioned in the question that differencing is 0, we would not have known to look at the same plot for P and Q (SAR and SMA). When we look at the seasonal lags, we find that both ACF and PACF have a cutoff at lag 12, and both of them have a single spike. We know from the question that SAR = 1 (meaning PACF gives single vale of P). This is, however, not clear from the plot. Therefore, it can be concluded that the plot is not well representative of the model that I have simulated.

APPENDIX

```
knitr::opts_chunk$set(echo = F, eval = T, comment = NA, message = F,warning = F,
                      fig.width = 7, fig.height = 4.2)

#Load/install required package here
library(forecast)
```

```r
library(tseries)
library(ggplot2)

#N = 100
set.seed(123)
model1 = arima.sim(model = list(order = c(1,0,0), ar = 0.6), n = 100)
model2 = arima.sim(model = list(ma = 0.9), n = 100)
model3 = arima.sim(model = list(order = c(1,0,1), ar = 0.6, ma = 0.9), n = 100)
par(mfrow = c(1,3))
acf1 = Acf(model1, lag.max = 40, main = "ACF of ARMA(1,0)")
acf2 = Acf(model2, lag.max = 40, main = "ACF of ARMA(0,1)")
acf3 = Acf(model3, lag.max = 40, main = "ACF of ARMA(1,1)")
par(mfrow = c(1,3))
pacf1 = Pacf(model1, lag.max = 40, main = "PACF of ARMA(1,0)")
pacf2 = Pacf(model2, lag.max = 40, main = "PACF of ARMA(0,1)")
pacf3 = Pacf(model3, lag.max = 40, main = "PACF of ARMA(1,1)")

#N = 1000
#Step 1
set.seed(123)
model4 = arima.sim(model = list(order = c(1,0,0), ar = 0.6), n = 1000)
model5 = arima.sim(model = list(ma = 0.9), n = 1000)
model6 = arima.sim(model = list(order = c(1,0,1), ar = 0.6, ma = 0.9), n = 1000)
#Step 2
par(mfrow = c(1,3))
acf4 = Acf(model4, lag.max = 40, main = "ACF of ARMA(1,0) with n = 1000")
acf5 = Acf(model5, lag.max = 40, main = "ACF of ARMA(0,1) with n = 1000")
acf6 = Acf(model6, lag.max = 40, main = "ACF of ARMA(1,1) with n = 1000")
#Step 3
par(mfrow = c(1,3))
pacf4 = Pacf(model4, lag.max = 40, main = "PACF of ARMA(1,0) with n = 1000")
pacf5 = Pacf(model5, lag.max = 40, main = "PACF of ARMA(0,1) with n = 1000")
pacf6 = Pacf(model6, lag.max = 40, main = "PACF of ARMA(1,1) with n = 1000")

#SARIMA model
library(astsa)
set.seed(123)

model = sarima.sim(ma = 0.5, sar = 0.8, S = 12, n = 100)

par(mfrow = c(1,2))
model_acf = Acf(model, lag.max = 100, main = "ACF of SARIMA Model")
model_pacf = Pacf(model, lag.max = 100, main = "PACF of SARIMA Model")
```