ENV 790.30 - Time Series Analysis for Energy Data | Spring 2022

Assignment 5 - Due date 02/28/22

Colin Lee

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.\

```
library(forecast)
```

```
## Warning: package 'forecast' was built under R version 4.1.2

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

```
library(tseries)
library(sarima)
```

```
## Warning: package 'sarima' was built under R version 4.1.2

## 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:

a) AR(2)

   Answer: The ACF plot will decay exponentially over time. The PACF plot will reveal the order of the model, so in this case, we will see two large bars (or three including the bar at lag 0)

b) MA(1)

   Answer: It is like the opposite of AR where the ACF plot will reveal the order of the MA model, where we will expect to see one large bar or two including the bar at lag 0 and then the PACF plot will exponentially decay.

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

```
#arima.sim(n = 100, list(ar = 0.6, ma = 0.8))

arma10 <- arima.sim(model = list(order = c(1,0,0), ar = 0.6), n = 100)  ## ARMA(1,0)
arma01 <- arima.sim(model = list(order = c(0,0,1), ma = 0.9), n = 100)  ## ARMA(0,1)
arma11 <- arima.sim(model = list(order = c(1,0,1), ar = 0.6, ma = 0.9), n = 100)  ## ARMA(1,1)

arma10
```

```
## Time Series:
## Start = 1
## End = 100
## Frequency = 1
##   [1] -0.84835640 -1.58517491 -1.73013678 -0.70640575  0.48566496 -0.59091731
##   [7] -0.40375209 -0.64266441 -1.32260132 -1.99414101 -0.02149890 -0.69974887
##  [13] -0.43493352 -0.24885951 -0.70497241  0.38013895  0.07681615  0.13263465
##  [19]  0.16218875 -0.39371168  1.48868925 -0.12219313  0.68817206 -1.70078322
##  [25]  0.29301492  0.09942456  0.26859804 -0.45692964 -0.20640174 -0.26468210
##  [31] -1.26734955 -1.53575618 -1.01453507 -0.76529872 -2.14439374 -1.97420115
##  [37] -1.07685830 -1.85089512 -2.22776914 -1.71302974 -1.23182152  0.89202616
##  [43] -0.49646187 -0.68782213 -0.32891617  1.50161331  1.03278991 -1.69970544
##  [49] -0.61624123  0.59763673  1.19643429  1.79502038  1.68668203  1.31099123
##  [55]  2.22315748  0.98171470  0.07344629  0.96912913  1.00673921  0.37158348
##  [61]  0.65230969  0.54733521 -0.01258714 -0.05463148  0.24926787 -0.66164593
##  [67] -0.69153431 -0.50787573 -0.76484527 -0.03360395 -0.19871251 -1.05456242
##  [73]  1.02113194  1.41109225 -0.65357918 -1.02012385 -1.06409402 -0.59755307
##  [79] -2.08804415 -0.55704889 -0.09135711  2.10606878  1.24368912  1.37541122
```

```
## [85]  1.46022292 -0.11544034 -0.20056107 -0.47747293  0.89024754  0.01583474
## [91]  1.06183809  0.39170307  1.54050365  0.30509167  0.11279566 -0.87733694
## [97]  0.26217856 -0.30335697 -1.24436138 -1.10756708
```

```
arma01
```

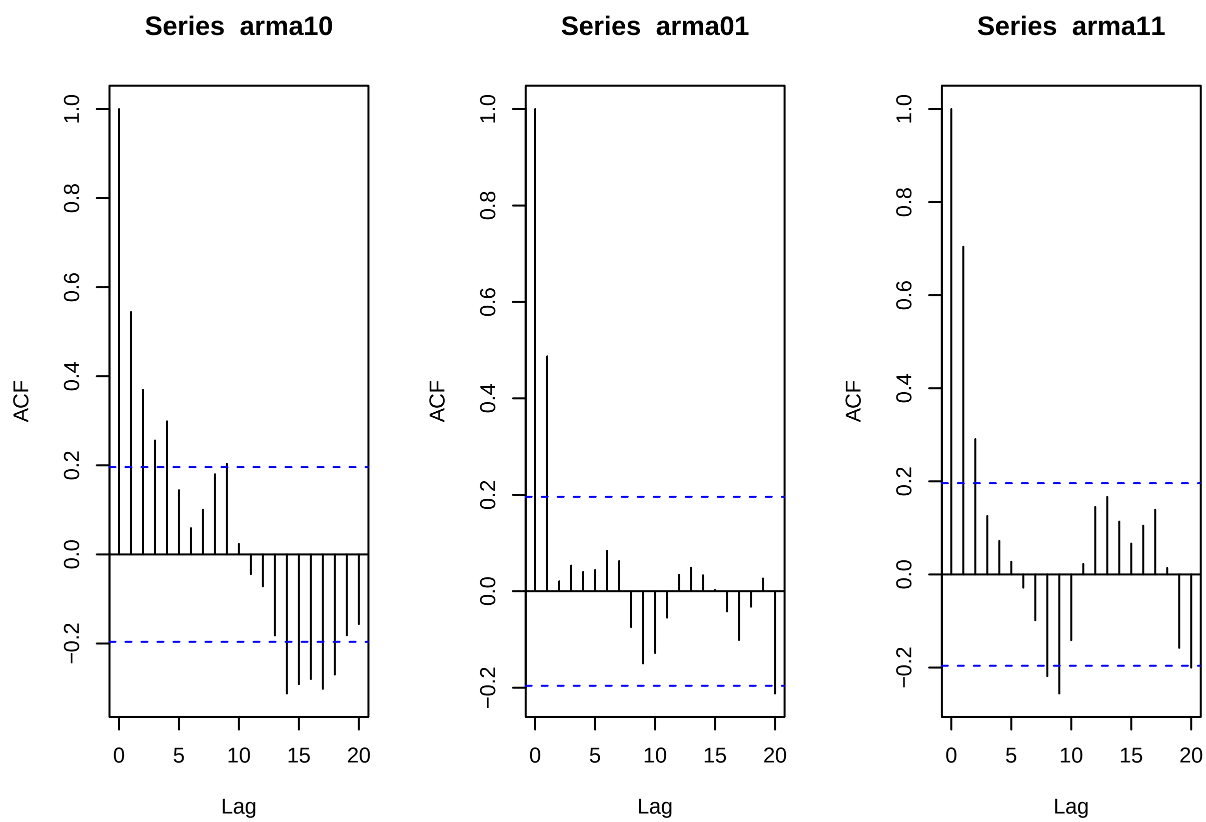
```
## Time Series:
## Start = 1
## End = 100
## Frequency = 1
##   [1] -0.70507706 -0.07183255 -0.07604969 -0.22045865  1.17670710  0.82713800
##   [7]  0.58982754  0.60861858  0.48739644  0.78408158  0.63369294  1.83298106
##  [13]  2.73050875  1.62150481  0.25822127  0.85282229 -0.59439324 -1.32047599
##  [19]  1.41257411 -0.58177436 -2.48211711 -1.34429646  0.57049608  2.64888016
##  [25]  0.40492054 -2.45105572 -1.57954242 -0.84658406 -0.83894827 -0.49457166
##  [31]  0.19868357 -1.71787879 -3.76463969 -2.62664416  0.32045950 -0.88929756
##  [37] -2.55012696  0.21756204  0.62167933 -1.25735304 -0.21518441  1.76280776
##  [43]  1.82756876  0.80119693  0.43760639  0.45743559  1.70863914  1.76831706
##  [49]  0.84016503 -0.93907395 -0.98459694 -0.12247543 -1.05959857  0.56285829
##  [55] -0.05695723 -0.86362415  2.19357216  0.69963472 -0.90740107 -0.31263990
##  [61]  0.30042625 -1.05176393 -1.48010441 -1.72410853 -0.89701653  2.41119164
##  [67] -0.25731482 -3.75991935 -2.59222554 -1.62099900  0.30303304  1.84423273
##  [73]  2.21446350  0.81175376 -1.84221060 -1.44227931 -2.22265459 -2.56213887
##  [79] -1.04476863 -1.41902958 -1.36152926 -2.39546137 -1.23664399 -0.53692064
##  [85]  0.50072447  2.12284742  1.20407008 -0.51809563  1.47716781  1.14803992
##  [91] -1.09727789  0.49953639 -0.30015232 -0.97575070 -0.62493489 -0.71907076
##  [97] -0.76565357 -0.65724174  2.11117932  0.84859735
```

```
arma11
```

```
## Time Series:
## Start = 1
## End = 100
## Frequency = 1
##   [1] -1.39190432  0.26629125  1.08940117 -0.29236533 -0.58701857  0.12456001
##   [7] -0.22387638  0.77148621  2.97696632  2.92158671  1.77206615 -0.28307404
##  [13] -0.82987799 -0.90685264 -1.80259695 -1.04674088 -0.86285412 -3.55126859
##  [19] -2.29924815  0.49031454 -0.89322300 -2.23081720 -1.12748008  0.25162760
##  [25] -0.65841085 -0.82866686 -1.27796023 -2.24193598 -0.81411118 -0.84051083
##  [31] -1.74261047 -2.20757190 -3.69530321 -5.25864258 -6.49697944 -5.30563480
##  [37] -3.53193007 -2.83156024 -1.30320993 -0.04578095 -0.17167696  0.12194399
##  [43]  1.14149139  1.68008649  1.59622571 -0.93300389 -3.34749636 -2.55106004
##  [49] -0.76326551 -1.34597041 -1.87359073 -2.29134106 -1.63466466 -1.10628635
##  [55]  0.42589166  3.17067934  2.21325617  1.65801143  0.02705677 -0.37589998
##  [61]  0.93911807  0.74280753  0.39965589 -1.05052519 -0.59173286  0.90305192
##  [67]  0.25691958 -2.04771156 -1.87460712  1.03751942  0.84013272 -0.87820192
##  [73] -0.55725404 -1.63348804 -3.97998090 -3.90464246 -2.45637730 -1.62582483
##  [79] -1.08492081 -1.68903311 -1.41936191  0.80908898  2.13899421  1.83131610
##  [85] -2.15285713 -3.58921145 -1.71700655 -0.14367604 -0.50576738 -2.34528087
##  [91] -2.05066181 -2.98822214 -2.66185513 -0.30205219  2.27712191  2.24002846
##  [97] -1.61979247 -2.28104644 -1.36326707 -2.37818517
```

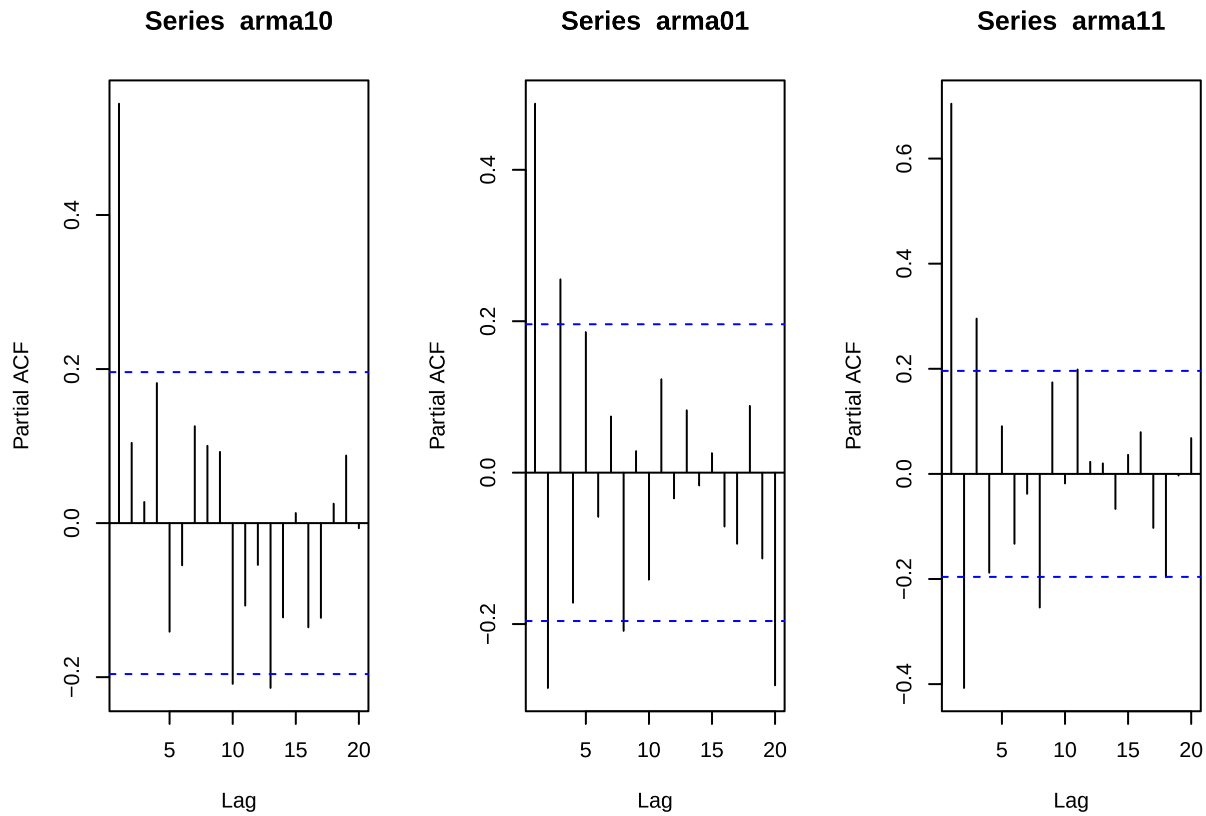
a) Plot the sample ACF for each of these models in one window to facilitate comparison (Hint: use command $par(mfrow = c(1, 3))$ that divides the plotting window in three columns).

```r
par(mfrow=c(1,3))
acf(arma10)
acf(arma01)
acf(arma11)
```



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

```r
par(mfrow=c(1,3))
pacf(arma10)
pacf(arma01)
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: I would have been able to identify the AR and MA plots. For ARMA(1,0), the ACF plot has an exponential decay while the PACF has a single high bar cutoff (that is not lag 0) and does not have a strong exponential decay trend. For ARMA(0,1), the ACF plot does not have a strong exponential decay trend and has 1 large bar that is above the threshold and not lag 0. On the other hand, the PACF has some semblance of exponential decay. It is very difficult to identify ARMA models, but I suppose the indicator here would be that we could see that there is no strong exponential decay in either ACF or PACF, but there still remain bars above the the thresholds, indicating cutoff bars and that the data is not white noise. However, I would not have been able to identify the order of the ARMA model since I see 2-3 bars for ACF and 2 bars for PACF making me think it could be an ARMA(2,2) or AMRA(2,3) when it is not.

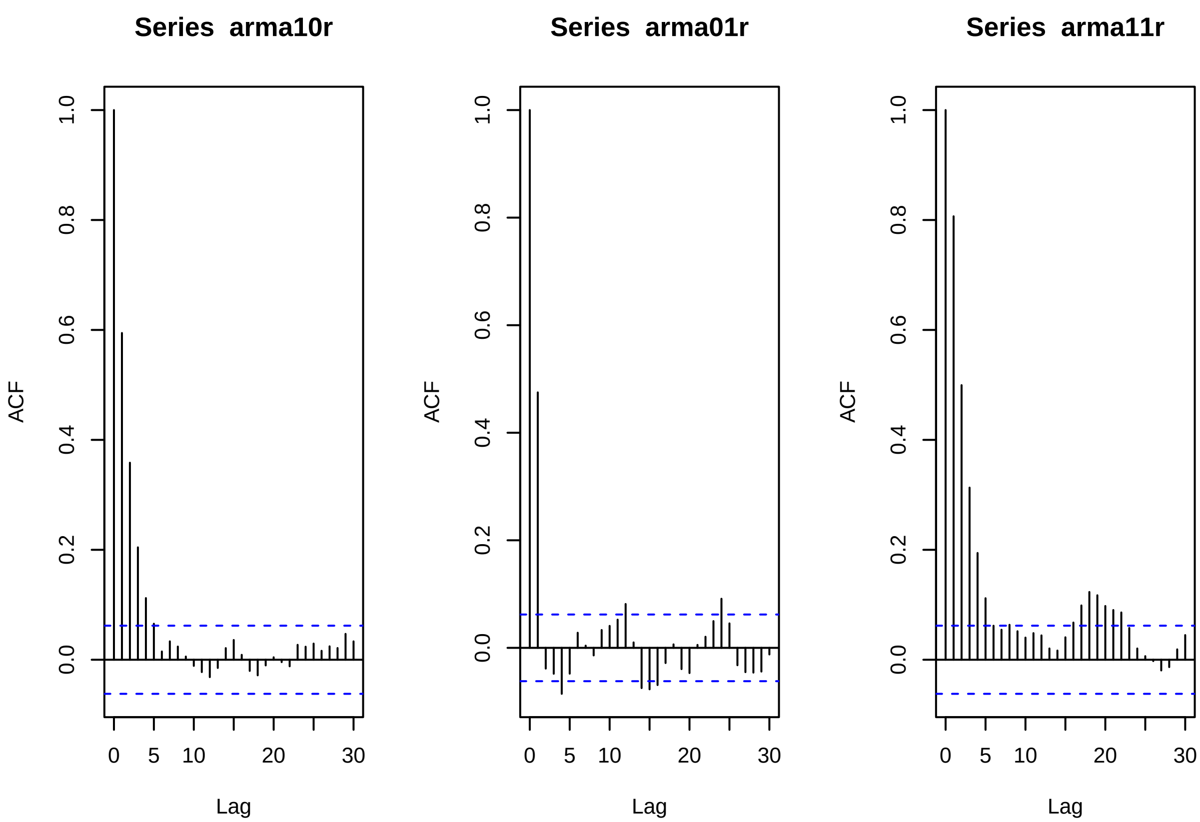
d) Compare the ACF and PACF values R computed with the theoretical values you provided for the coefficients. Do they match? Explain your answer.

Answer: For the AR model, the PACF first coefficient is 0.6, which is very close to the theoretical phi value of 0.6. For the MA model, the ACF first coefficient is 0.5, which is very far off from the input theta value of 0.9. The reason for it being off is because for the MA model, you cannot determine the error term of the previous term from the ACF plot. For the ARMA model, the AR coefficient (PACF) is 0.8, which is above the theoretical phi value of 0.6, and the MA coefficient (ACF) is 0.8, which is not too far off of the theoretical theta of 0.9.

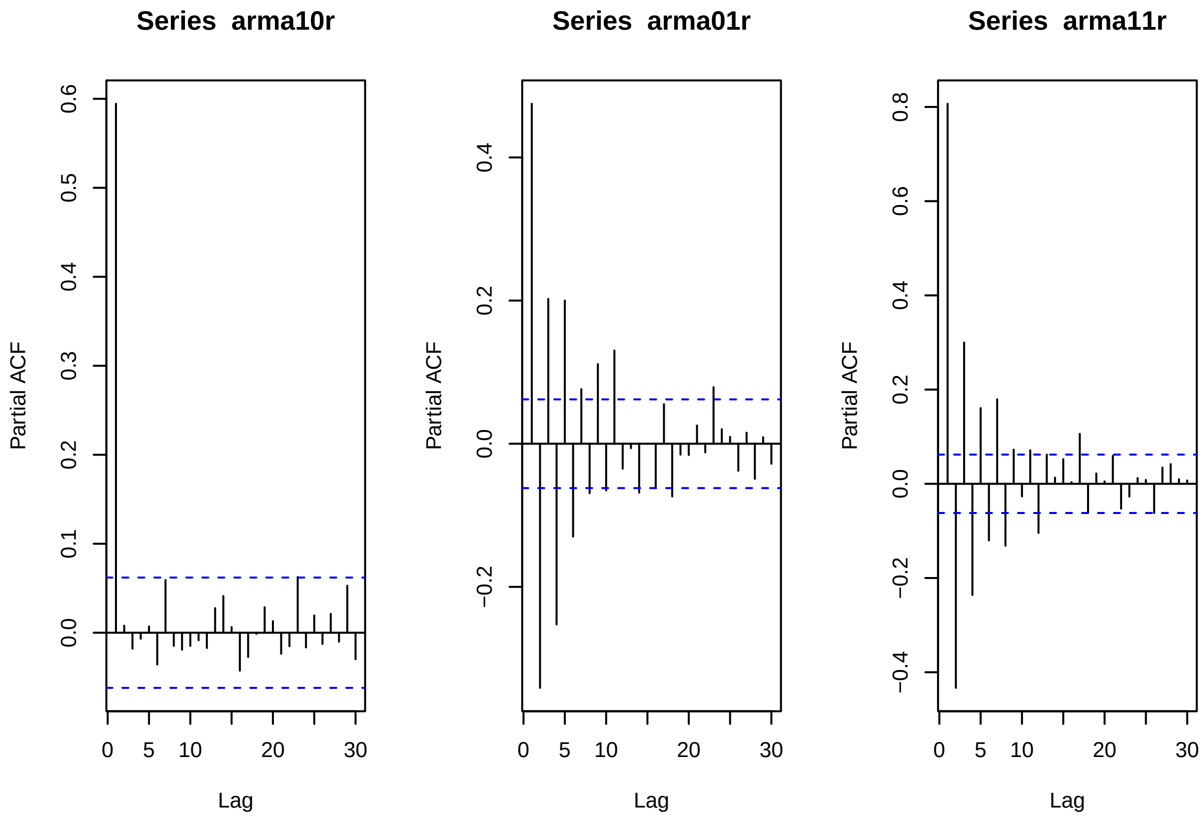
Increase number of observations to $n = 1000$ and repeat parts (a)-(d).

I commented out the running of the n = 1000 runs of the arma models since they would take up a lot of space.

```
arma10r <- arima.sim(model = list(order = c(1,0,0), ar = 0.6), n = 1000)  ## ARMA(1,0)
arma01r <- arima.sim(model = list(order = c(0,0,1), ma = 0.9), n = 1000)  ## ARMA(0,1)
arma11r <- arima.sim(model = list(order = c(1,0,1), ar = 0.6, ma = 0.9), n = 1000)  ## ARMA(1,1)

#arma10r
#arma01r
#arma11r



par(mfrow=c(1,3))
acf(arma10r)
acf(arma01r)
acf(arma11r)
```




```
par(mfrow=c(1,3))
pacf(arma10r)
pacf(arma01r)
pacf(arma11r)
```



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 would have been able to identify the AR and MA plots. For ARMA(1,0), the ACF plot has an exponential decay while the PACF has a single high bar cutoff (that is not lag 0) and does not have a strong exponential decay trend. For ARMA(0,1), the ACF plot does not have a strong exponential decay trend and has 1 large bar that is above the threshold and not lag 0. On the other hand, the PACF has some semblance of exponential decay. It is very difficult to identify ARMA models, but I suppose the indicator here would be that we could see that there is no strong exponential decay in either ACF or PACF, but there still remain bars above the the thresholds, indicating cutoff bars and that the data is not white noise. However, I would not have been able to identify the order of the ARMA model since I see 3-4 bars for ACF and 3-4 bars for PACF making me think it could be a larger order ARMA model in terms of both AR and MA, but this is not the case.

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 R computed value for the AR model is clearer and matches the theoretical phi value since we can see the cutoff bar in the PACF plot at lag 1 is equal to 0.6. However, for the MA model, this is still not as close since the cutoff bar in the ACF plot at lag 1 is 0.5, not close to the theoretical theta of 0.9. The reason for it being off is because for the MA model, you cannot determine the error term of the previous term from the ACF plot. For the ARMA model, they do not match at all since the MA bar (ACF) is around 0.8 (not like theoretical theta of 0.9) and the AR bar (PACF) is around 0.78, not like theoretical phi of 0.6. However, they are close.

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.

ARIMA(1,0,1) X (1,0,0)12

I can't tell if D = 1 or D = 0 because in class we learned D = 1, then it's yt - yt-s, but then if D = 0, the seasonal yt-s term still exists as long as P = 1. If I had to pick one, it would be D = 0.

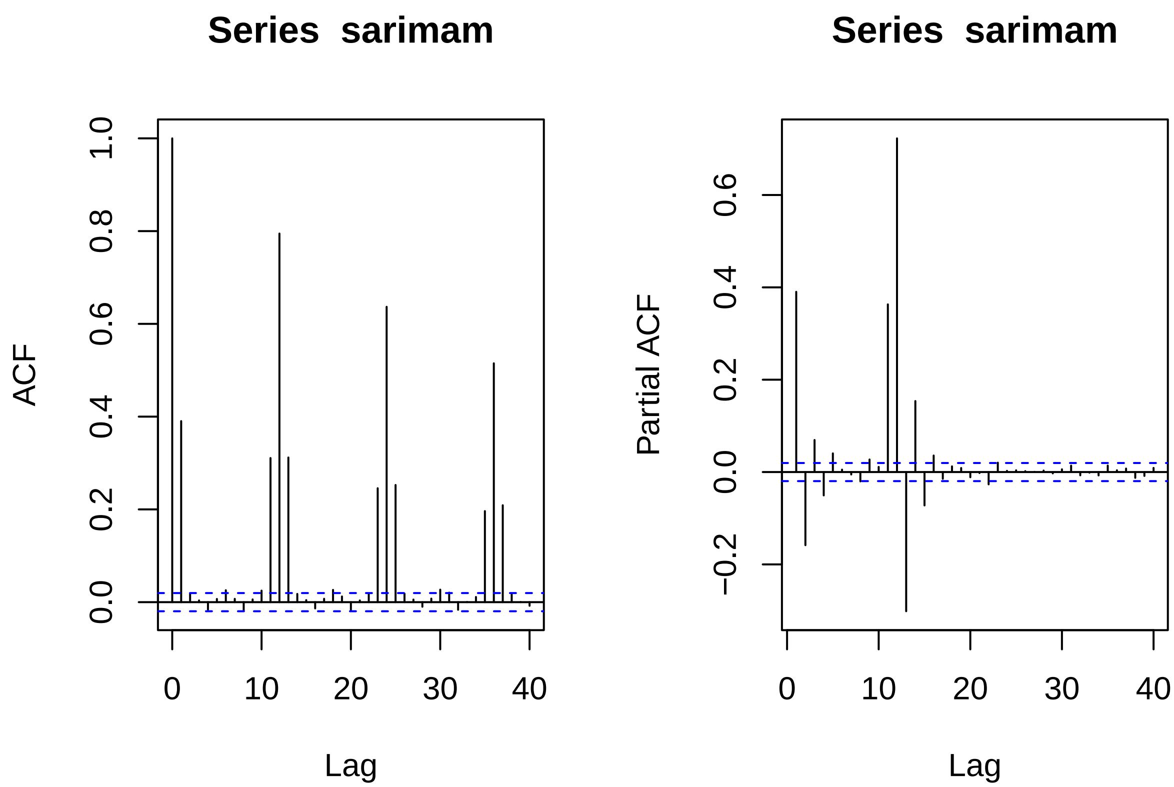
b) Also from the equation what are the values of the parameters, i.e., model coefficients.

The 0.7 is phi1 and 0.1 is theta1. 0.25 is phi12.

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.

```
sarimam <- sim_sarima(model = list(ma = 0.5, sar = 0.8, nseasons = 12), n = 10000)

par(mfrow=c(1,2))
sarima_acf <- acf(sarimam)
sarima_pacf <- pacf(sarimam)
```



For analyzing the seasonal component, the ACF plot represents the seasonality of the model as the large spikes occur in lag multiples of 12. Furthermore, the PACF has one large spike, combining with the previous statement to indicate that we have a seasonal AR model. Thus, the plots represent the model accurately in terms of the seasonal component.

For the non-seasonal component, we can see an exponential decay in the PACF plot and a significant cut off at lag 1 in the ACF plot, indicating that this is an MA model with order 1 ($q = 1$) and no AR component. Thus, the plots represent the model accurately in terms of the non-seasonal component as well.