ENV 797 - Time Series Analysis for Energy and Environment Applications | Spring 2024

Assignment 6 - Due date 02/28/24

Vincient Whatley

Directions

You should open the .rmd file corresponding to this assignment on RStudio. The file is available on our class repository on Github.

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_Sp24.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: "ggplot2", "forecast", "tseries" and "sarima". 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.

```
#Load/install required package here
library("ggplot2")
library("forecast")
```

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

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

```
## Loading required package: stats4
```

```
##
## Attaching package: 'sarima'
```

```
## The following object is masked from 'package:stats':
##
##     spectrum
```

```
library("cowplot")
```

```
## Warning: package 'cowplot' was built under R version 4.3.2
```

This assignment has general questions about ARIMA Models.

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: This stands for the Auto regrissive model of the second order.This models most important charetristic is that it will take current value (2) (Which represents how many lags) and model it after our last two lags. This is done when we suspect that the last two values are correlated to the future value.

- MA(1)

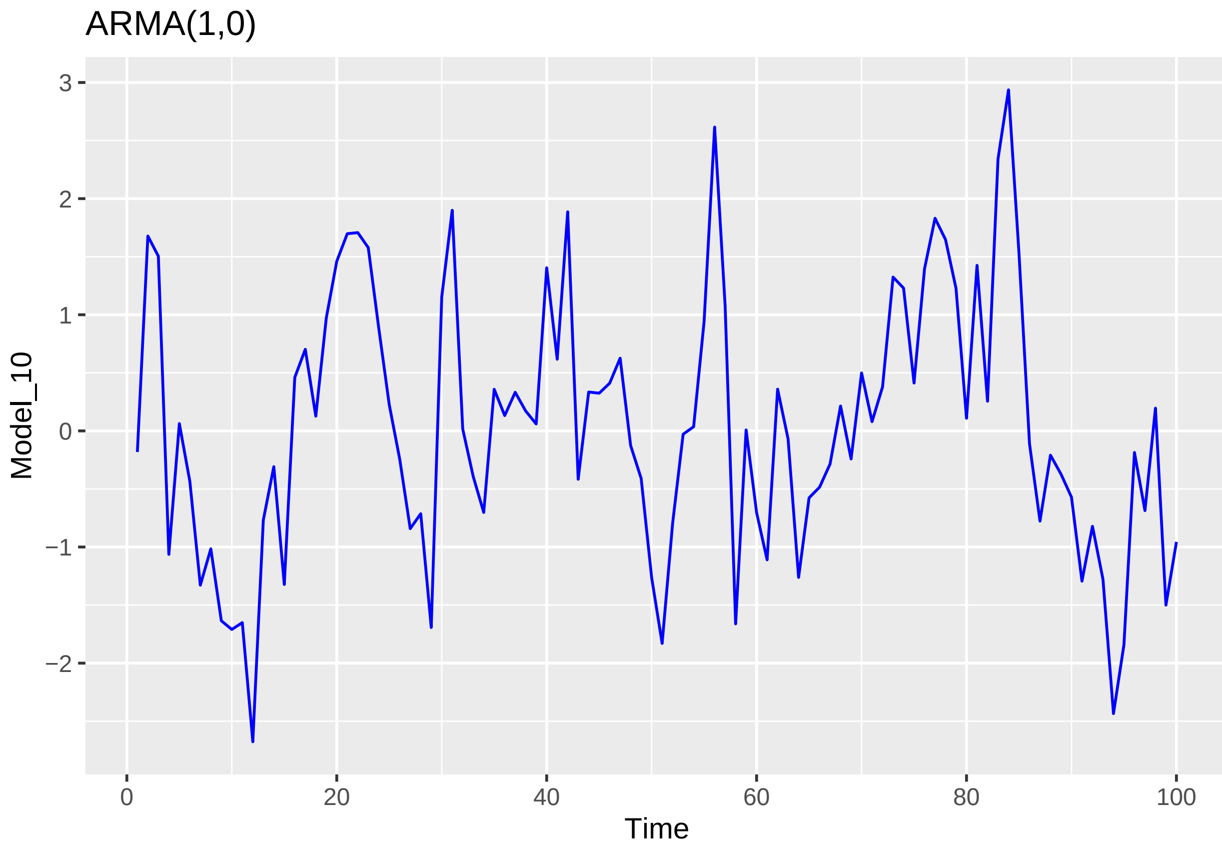
  Answer: This represents a moving average model with an order of 1. This model allows users to see the correlation between shortterm dependceies in the data. Within this model there is 1 lagged error term which is coupled with the current oberservation which is impacted by the previous error term.

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).

(a) 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 the `arima.sim()` function in R to generate $n = 100$ observations from each of these three models. Then, using `autoplot()` plot the generated series in three separate graphs.

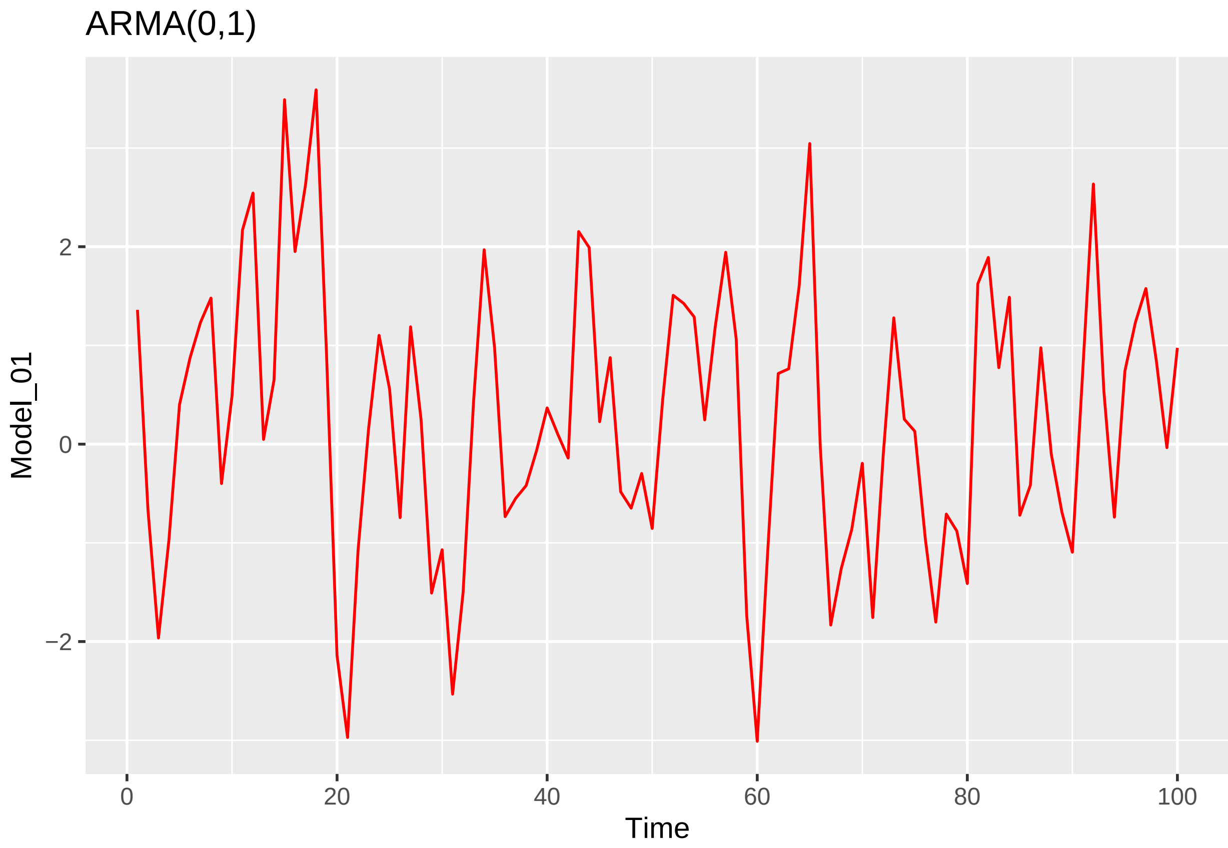
```
#Create parameters for phi, theata, and n
phi <-0.6
theta <-0.9
n <-100
n2 <-1000


#ARMA(1,0)
set.seed(123)  #same random values are used each time code is ran
Model_10 <- arima.sim(model=list(ar=phi, ma=0), n=n)

#Plot
autoplot(Model_10,colour="blue") +
  labs(title = "ARMA(1,0)")
```



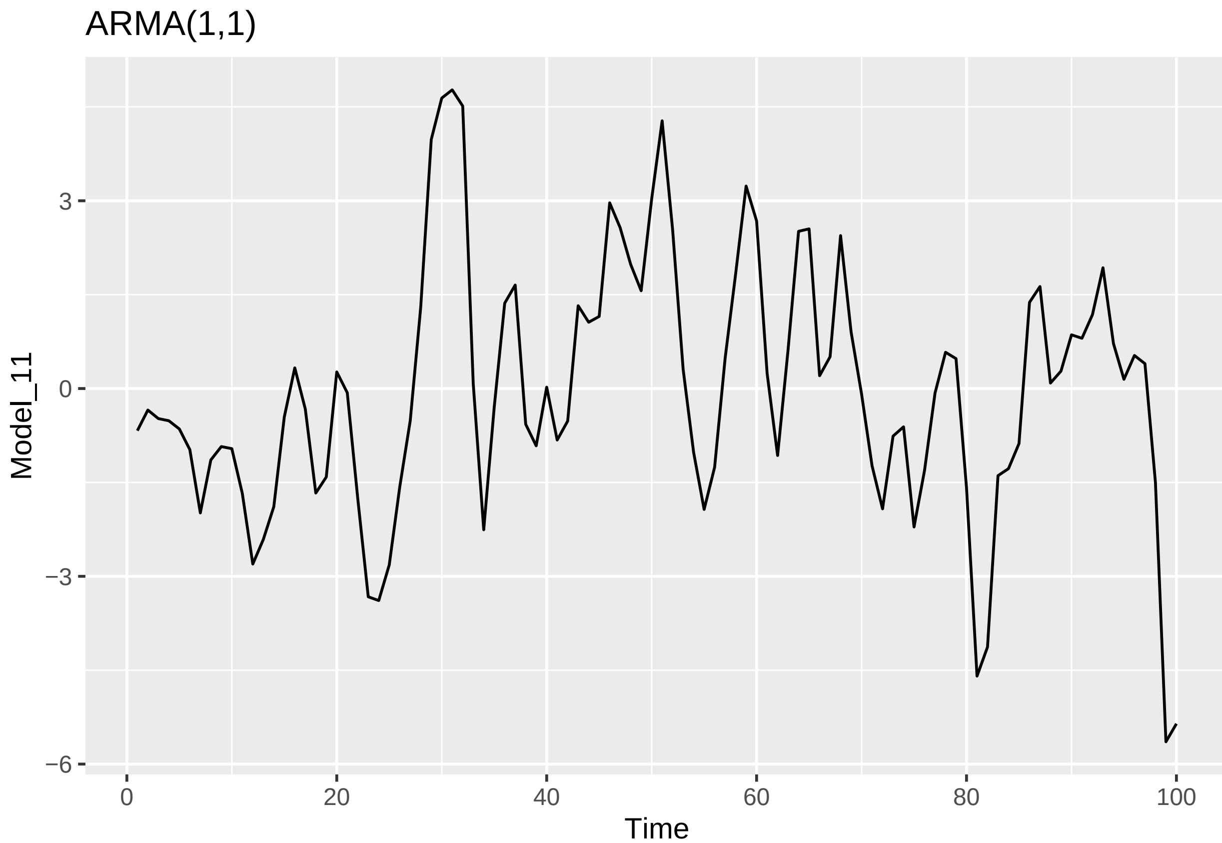

```
#ARMA(0,1)
set.seed(456)
Model_01 <- arima.sim(model=list(ar=0, ma=theta), n=n)
```

```
## Warning in min(Mod(polyroot(c(1, -model$ar)))): no non-missing arguments to
## min; returning Inf
```

```
autoplot(Model_01, colour="red") +
  labs(title = "ARMA(0,1)")
```




```
#ARMA(1,1)
set.seed(789)
Model_11 <- arima.sim(model=list(ar=phi,ma=theta), n=n)

autoplot(Model_11, colour="black") +
  labs(title = "ARMA(1,1) ")
```



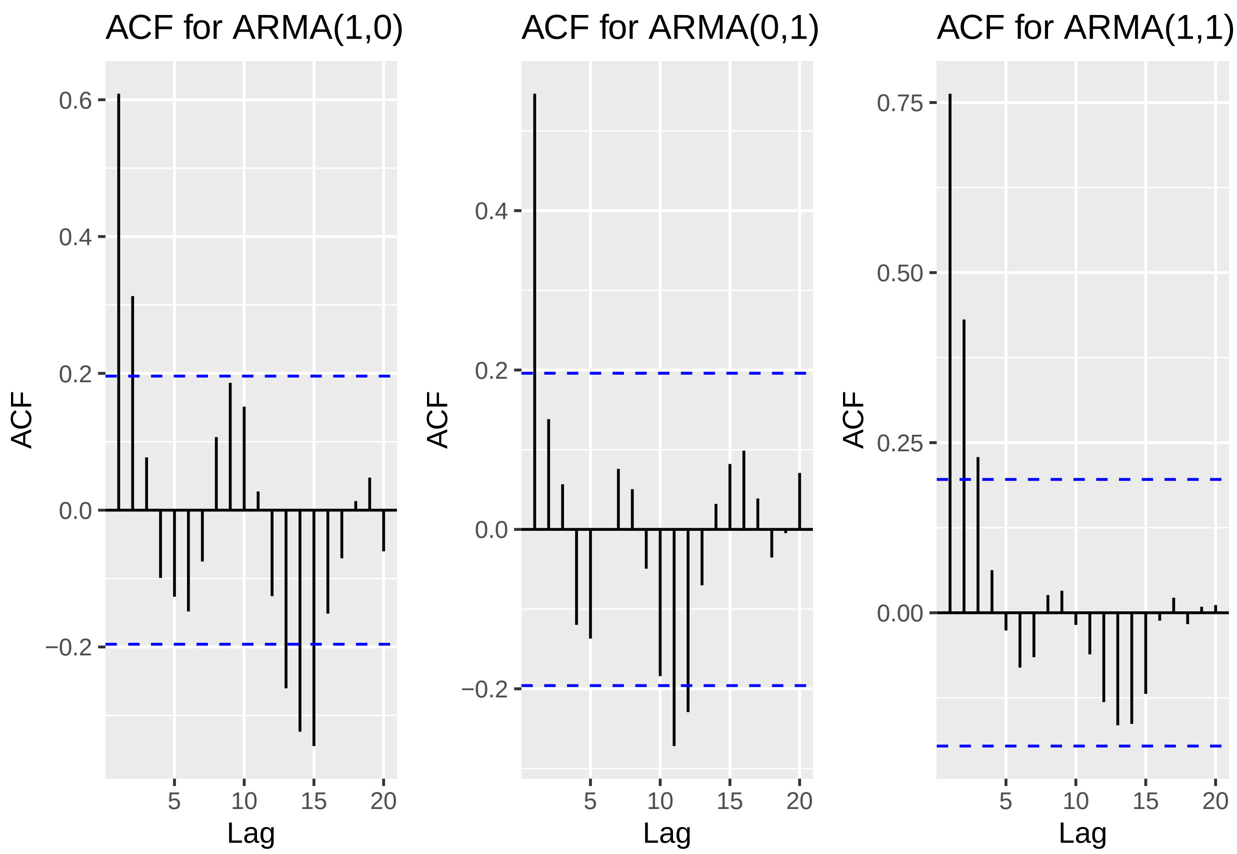
(b) Plot the sample ACF for each of these models in one window to facilitate comparison (Hint: use `cowplot::plot_grid()`).

```
#creating the data for each of the models
set.seed(456)
data_ARMA_10 <- arima.sim(model=list(ar=phi, ma=0), n=n)
data_ARMA_01 <- arima.sim(model=list(ar=0, ma=theta), n=n)
```

```
## Warning in min(Mod(polyroot(c(1, -model$ar)))): no non-missing arguments to
## min; returning Inf
```

```
data_ARMA_11 <- arima.sim(model=list(ar=phi, ma=theta), n=n)


#Plot the sample ACF
acf_ARMA_10 <- autoplot(acf(data_ARMA_10, plot=FALSE))+
  ggtitle("ACF for ARMA(1,0)")
acf_ARMA_01 <- autoplot(acf(data_ARMA_01, plot=FALSE))+
  ggtitle("ACF for ARMA(0,1)")
acf_ARMA_11 <- autoplot(acf(data_ARMA_11, plot=FALSE))+
  ggtitle("ACF for ARMA(1,1)")

#Combine ACF plots
combined_acf_plots <- plot_grid(acf_ARMA_10, acf_ARMA_01, acf_ARMA_11, nrow = 1)
```

```
#Display combined plots
plot_grid(combined_acf_plots, ncol = 1)
```



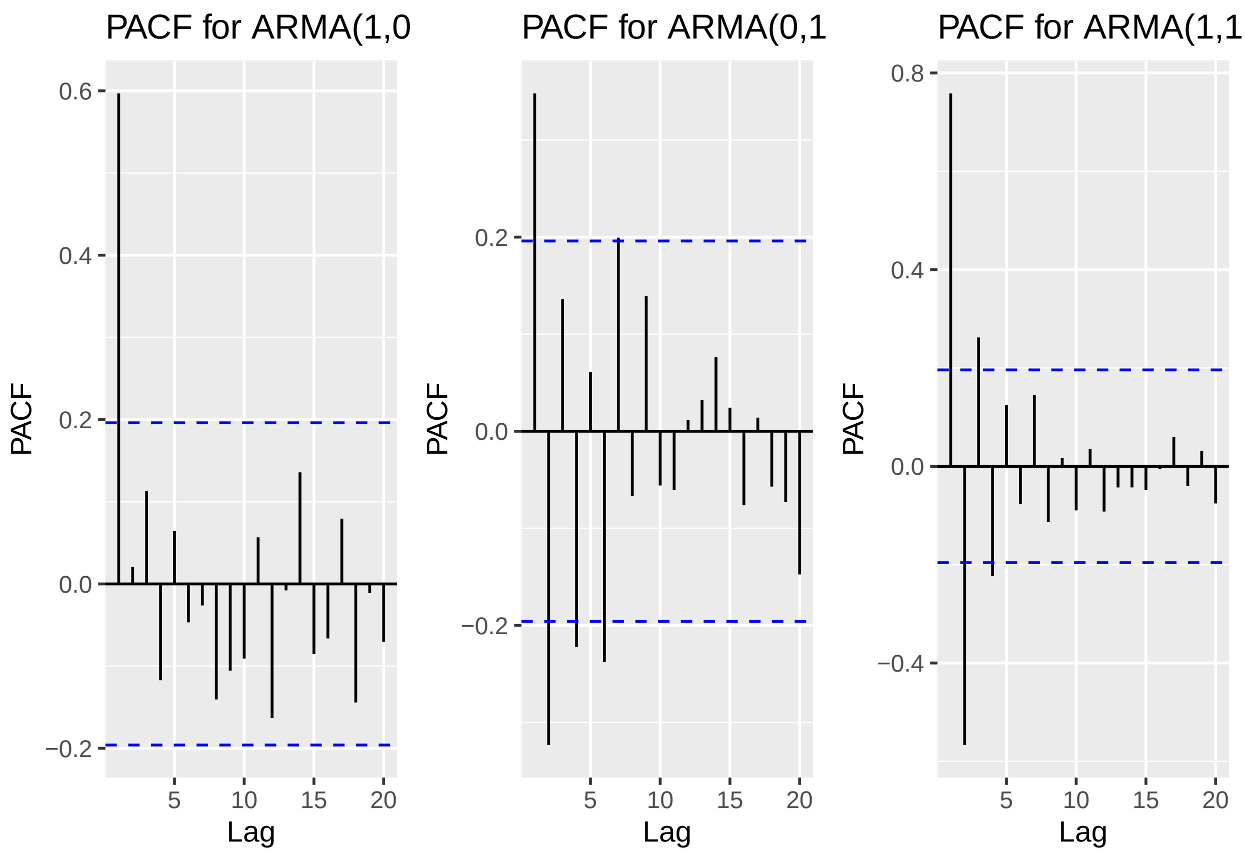
(c) Plot the sample PACF for each of these models in one window to facilitate comparison.

```
#Generate data for each model
set.seed(123)
data_ARMA_10 <- arima.sim(model=list(ar=phi, ma=0), n=n)
data_ARMA_01 <- arima.sim(model=list(ar=0, ma=theta), n=n)
```

```
## Warning in min(Mod(polyroot(c(1, -model$ar)))): no non-missing arguments to
## min; returning Inf
```

```
data_ARMA_11 <- arima.sim(model=list(ar=phi, ma=theta), n=n)


#Plot the PACF for each model
pacf_ARMA_10 <- autoplot(pacf(data_ARMA_10, plot = FALSE))+
 ggtitle("PACF for ARMA(1,0)")
pacf_ARMA_01 <- autoplot(pacf(data_ARMA_01, plot = FALSE))+
 ggtitle("PACF for ARMA(0,1)")
pacf_ARMA_11 <- autoplot(pacf(data_ARMA_11, plot = FALSE))+
 ggtitle("PACF for ARMA(1,1)")
```

```r
#Combine PACF plots
combined_pacf_plots <-
  plot_grid(pacf_ARMA_10, pacf_ARMA_01, pacf_ARMA_11, nrow = 1)

#Display combined plots
plot_grid(combined_pacf_plots, ncol=1)
```



(d) 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 able identify them correctly? Explain your answer.

Answer:For the ACF model you can tell that this would be a ARMA model due to the exponential decay that you see across all three of them while the while the PACF model tells us that it is ARMA because of the significant spike at lag 1 and then exponetial decay.

(e) Compare the PACF values R computed with the values you provided for the lag 1 correlation coefficient, i.e., does $\phi = 0.6$ match what you see on PACF for ARMA(1,0), and ARMA(1,1)? Should they match?

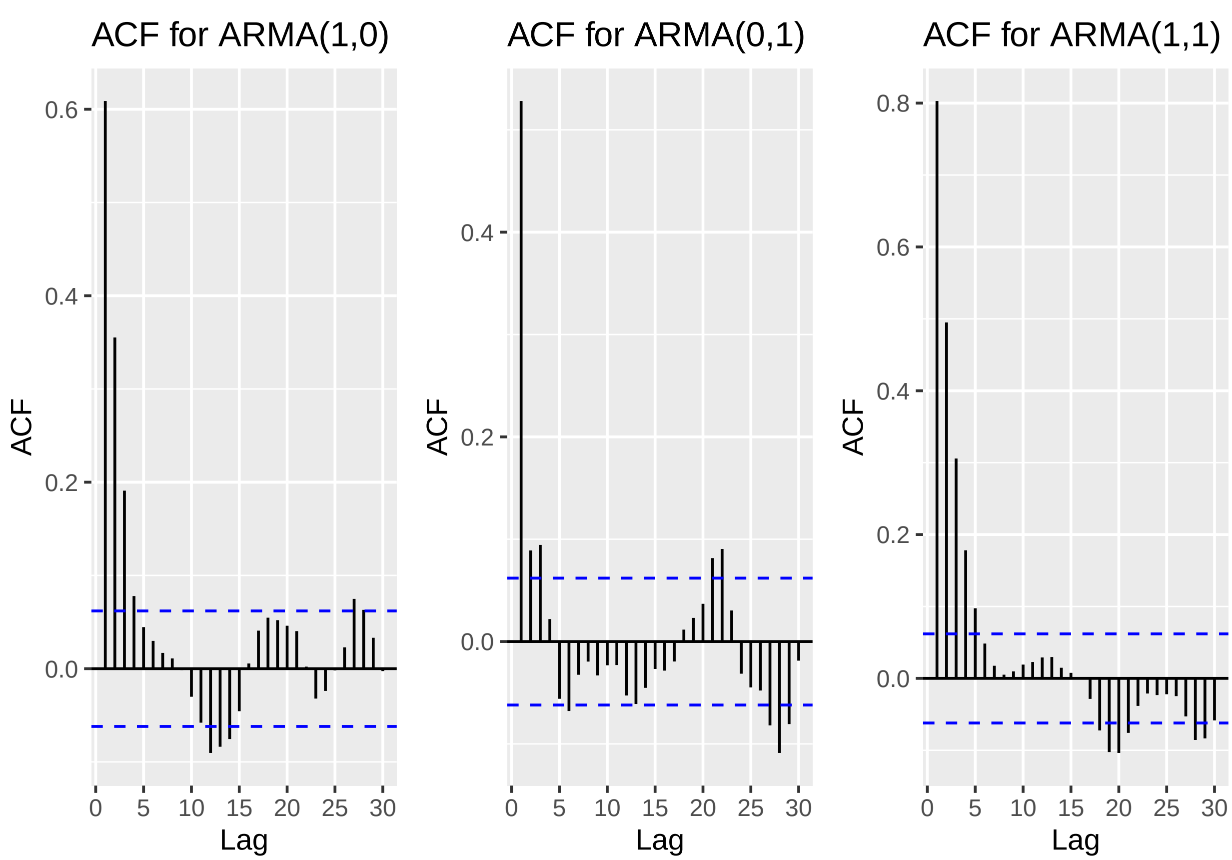
Answer:We only see that this is true in the case of ARMA (1,0) while the other two graphs have diffrening points. They should match or at the very least hover around 0.6.

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

```r
#creating the data for each of the models
set.seed(456)
data_ARMA_10 <- arima.sim(model=list(ar=phi, ma=0), n=n2)
data_ARMA_01 <- arima.sim(model=list(ar=0, ma=theta), n=n2)
```

```
## Warning in min(Mod(polyroot(c(1, -model$ar)))): no non-missing arguments to
## min; returning Inf
```

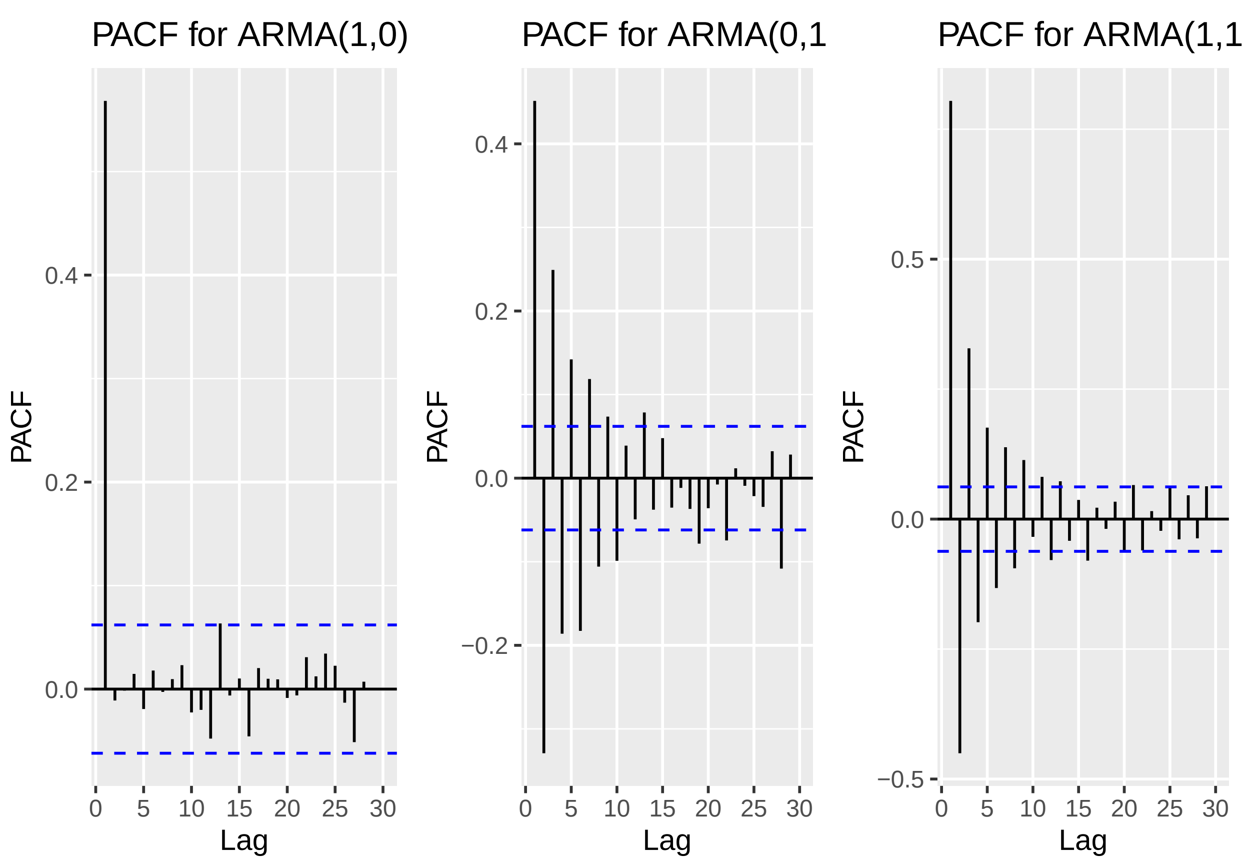
```r
data_ARMA_11 <- arima.sim(model=list(ar=phi, ma=theta), n=n2)


#Plot the ACF
acf_ARMA_10 <- autoplot(acf(data_ARMA_10, plot=FALSE))+
  ggtitle("ACF for ARMA(1,0)")
acf_ARMA_01 <- autoplot(acf(data_ARMA_01, plot=FALSE))+
  ggtitle("ACF for ARMA(0,1)")
acf_ARMA_11 <- autoplot(acf(data_ARMA_11, plot=FALSE))+
  ggtitle("ACF for ARMA(1,1)")

#Combine ACF plots
combined_acf_plots <- plot_grid(acf_ARMA_10, acf_ARMA_01, acf_ARMA_11, nrow = 1)

#Display combined plots
plot_grid(combined_acf_plots, ncol = 1)
```



```r
#Generate data for each model
set.seed(123)  # for reproducibility
data_ARMA_10 <- arima.sim(model=list(ar=phi, ma=0), n=n2)
data_ARMA_01 <- arima.sim(model=list(ar=0, ma=theta), n=n2)
```

```
## Warning in min(Mod(polyroot(c(1, -model$ar)))): no non-missing arguments to
## min; returning Inf
```

```r
data_ARMA_11 <- arima.sim(model=list(ar=phi, ma=theta), n=n2)


#Plot the PACF
pacf_ARMA_10 <- autoplot(pacf(data_ARMA_10, plot = FALSE))+
 ggtitle("PACF for ARMA(1,0)")
pacf_ARMA_01 <- autoplot(pacf(data_ARMA_01, plot = FALSE))+
 ggtitle("PACF for ARMA(0,1)")
pacf_ARMA_11 <- autoplot(pacf(data_ARMA_11, plot = FALSE))+
 ggtitle("PACF for ARMA(1,1)")

#Combine PACF plots
combined_pacf_plots <-
  plot_grid(pacf_ARMA_10, pacf_ARMA_01, pacf_ARMA_11, nrow = 1)

# Display combined plots
plot_grid(combined_pacf_plots, ncol=1)
```



> Answer part d again: With this we see a similar case as before with a huge spike a lag 1 and then exponential decay.

Answer part e again:These graphs match a lot more what it should be, all lines at lag 1 are at or hover around 0.6.

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: p=1,d=0,q=0,P=1,D=0,Q=0,S=12. This can be displayed as ARIMA(1,0,1)(1,0,0)s=12

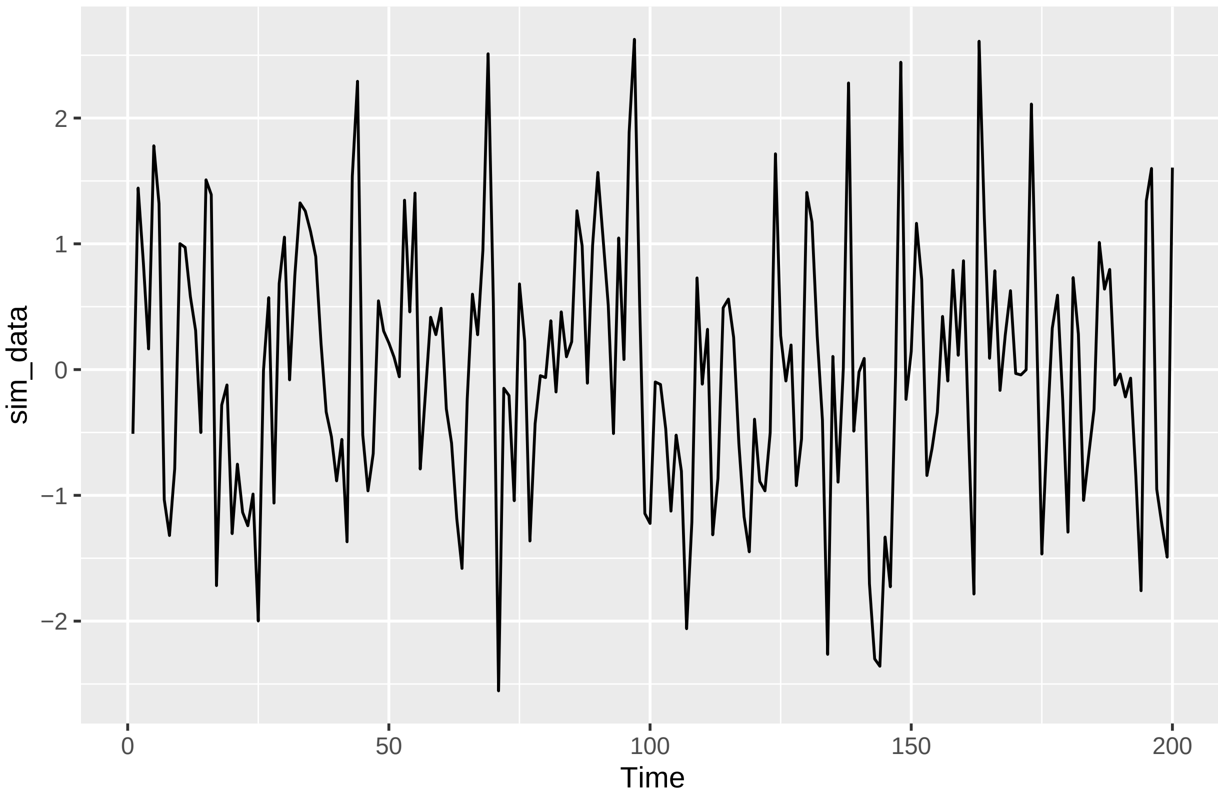
(b) Also from the equation what are the values of the parameters, i.e., model coefficients.

Answer:Theta is 0.7, Phi is -0.25, and the error term is -0.1

Q4

Simulate a seasonal ARIMA$(0, 1) \times (1, 0)_{12}$ model with $\phi = 0.8$ and $\theta = 0.5$ using the `sim_sarima()` function from package `sarima`. 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 the generated series using `autoplot()`. Does it look seasonal?

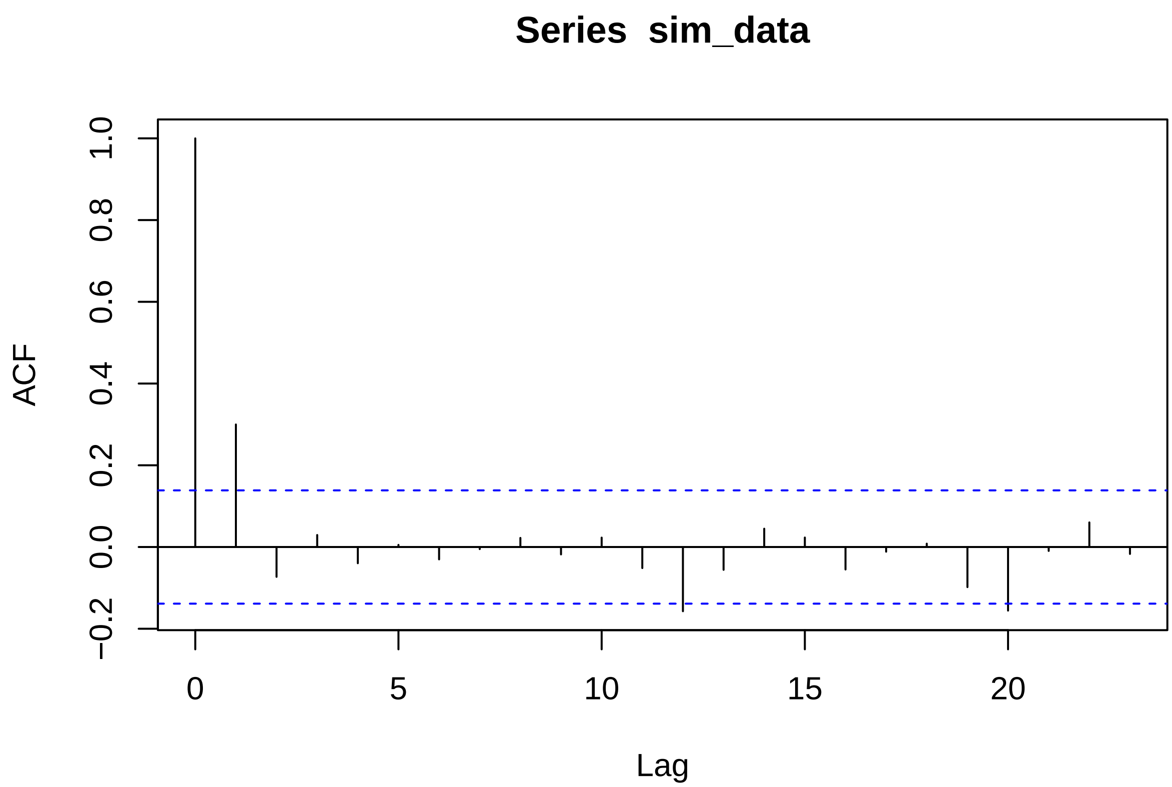
```
# Set parameters
phi <- 0.8
theta <- 0.5
s <- 12  #Seasonal lag
n <- 200  #Number of observations

model_list <- list(order = c(0, 0, 1),seasonal = list(order = c(1, 0, 0), period = s),sar = phi, ma = th

#using simulation fuction
set.seed(123)
sim_data <- arima.sim(n = n, model = model_list)

autoplot(sim_data)
```



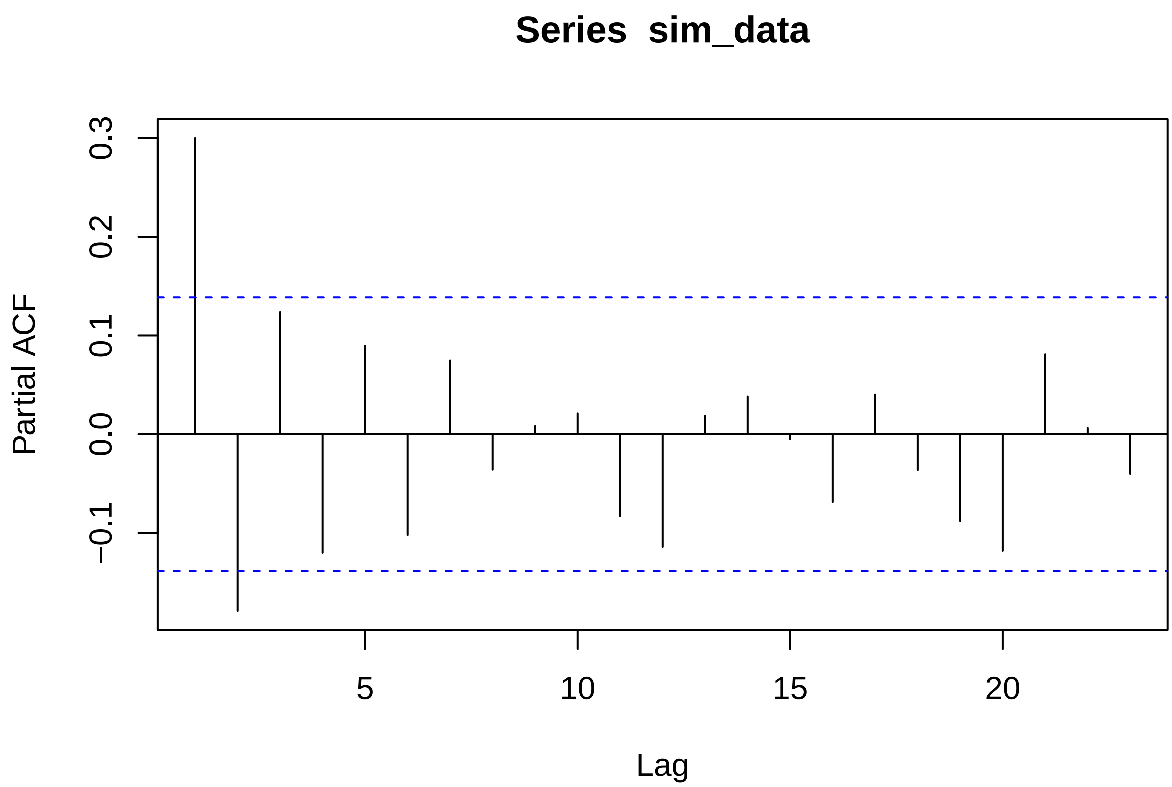
Q5

Plot ACF and PACF of the simulated series in Q4. 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
#ACF plot
acf(sim_data)
```




```r
#PACF plot
pacf(sim_data)
```



> Answer:This IS possoible by looking at the spikes from the lag. For acf note that there are multiple spikes in this model strategically placed. Indicaitong that their is some type of seasonality to this model. As for Pacf,this seasonality becomes even more prominent than the last series. Having mulitple speaks at clear intervals along with a gradual decay.