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

Assignment 6 - Due date 02/28/24

Zhenghao Lin

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(lubridate)
library(ggplot2)
library(forecast)
library(Kendall)
library(tseries)
library(outliers)
library(tidyverse)
library(cowplot)
library(sarima)
```

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:The ACF plot for an AR(2) model typically shows a gradual decay, while the PACF plot shows a significant spike at lag 2 and cuts off after lag 2.

- MA(1)

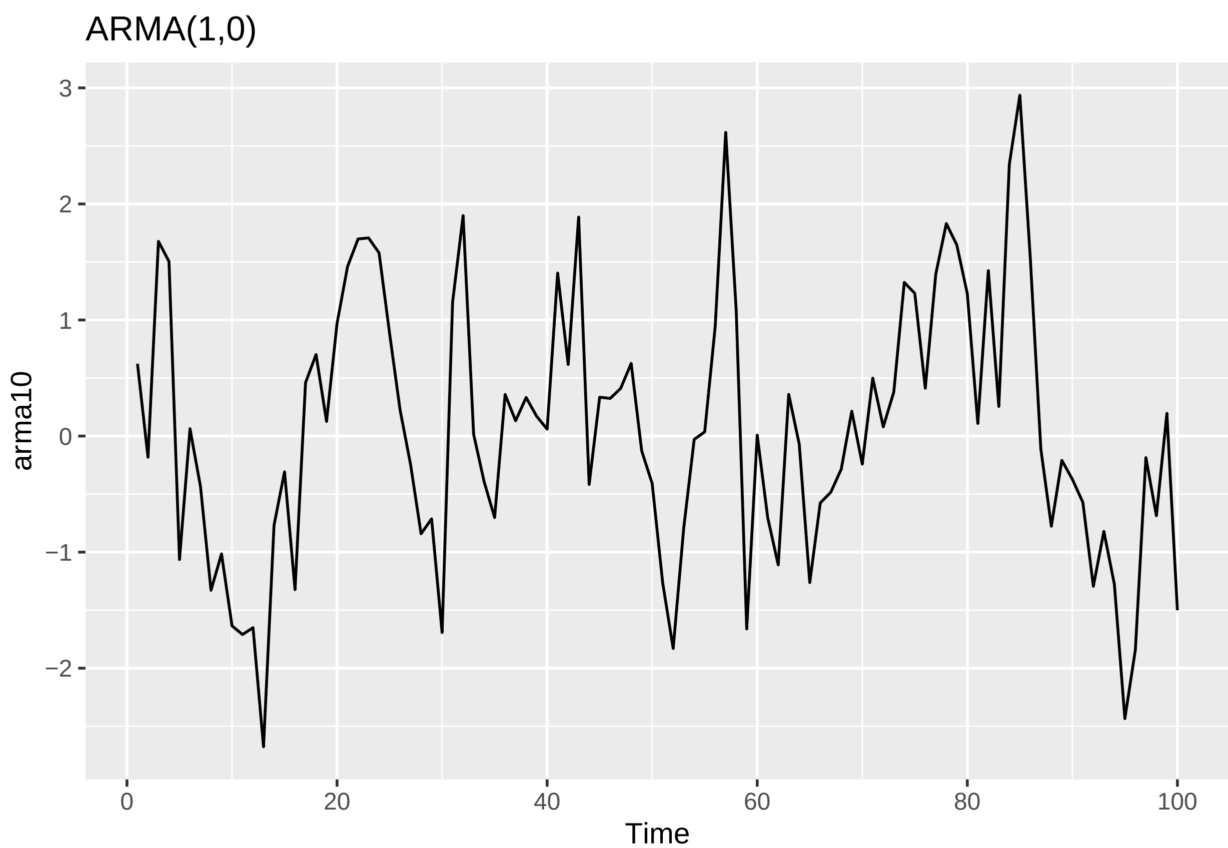
  Answer:The ACF plot for an MA(1) model shows a significant spike at lag 1 and cuts off after lag 1, while the PACF plot typically shows a gradual decay.

Q2

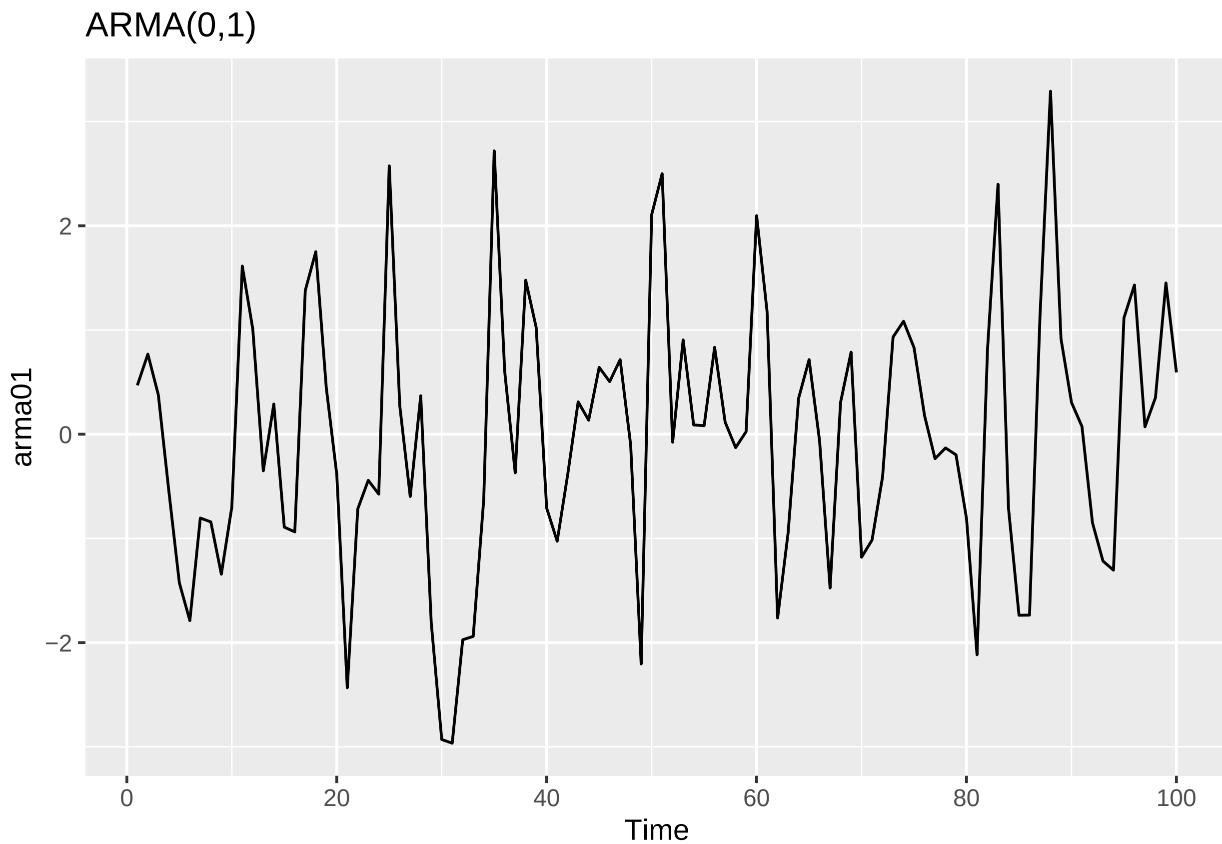
Recall that the non-seasonal ARIMA is described by three parameters ARIMA(p, d, q) where p is the order of the autoregressive component, d is the number of times the series need to be differenced to obtain stationarity and q is the order of the moving average component. If we don't need to difference the series, we don't need to specify the "I" part and we can use the short version, i.e., the ARMA(p, q).

(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.

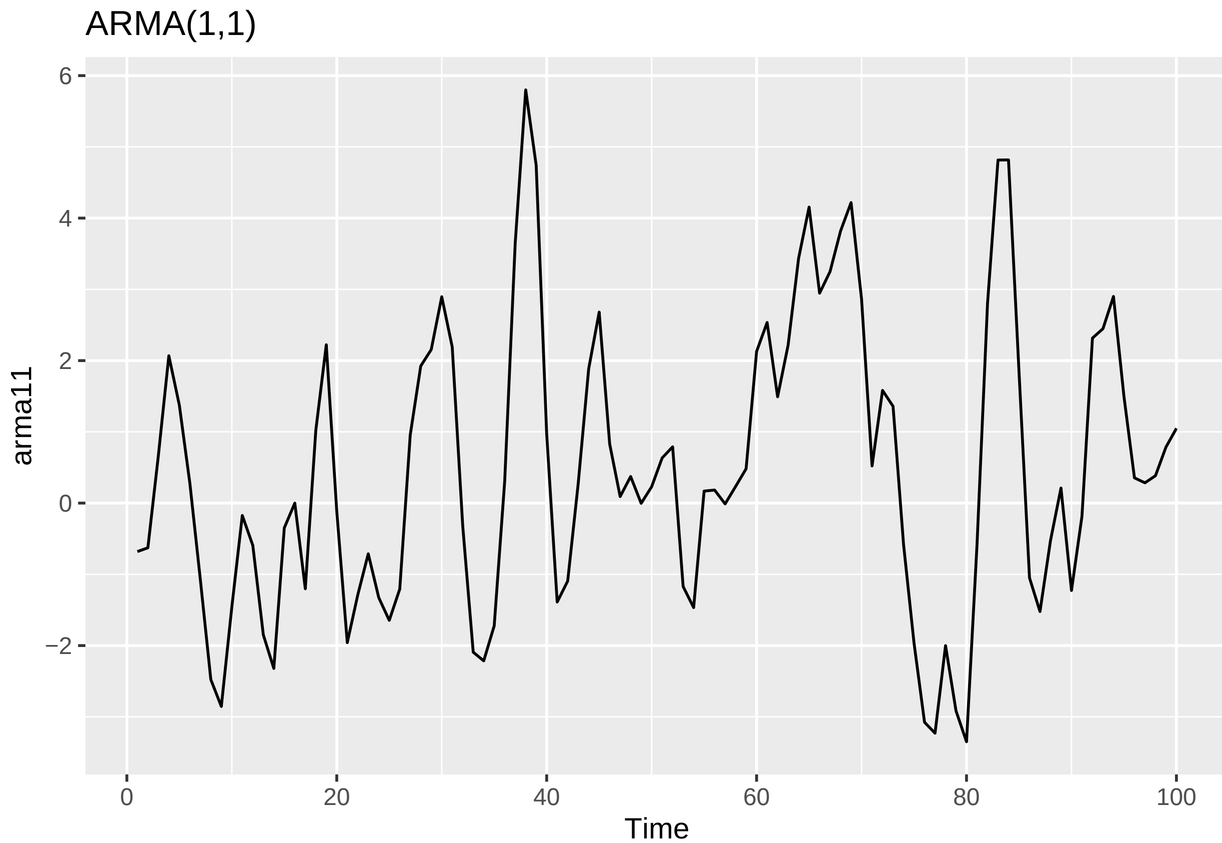
```
# ARMA(1,0)
set.seed(123)
arma10 <- arima.sim(n = 100, list(ar = 0.6))
autoplot(arma10) + ggtitle("ARMA(1,0)")
```



```r
# ARMA(0,1)
arma01 <- arima.sim(n = 100, list(ma = 0.9))
autoplot(arma01) + ggtitle("ARMA(0,1)")
```

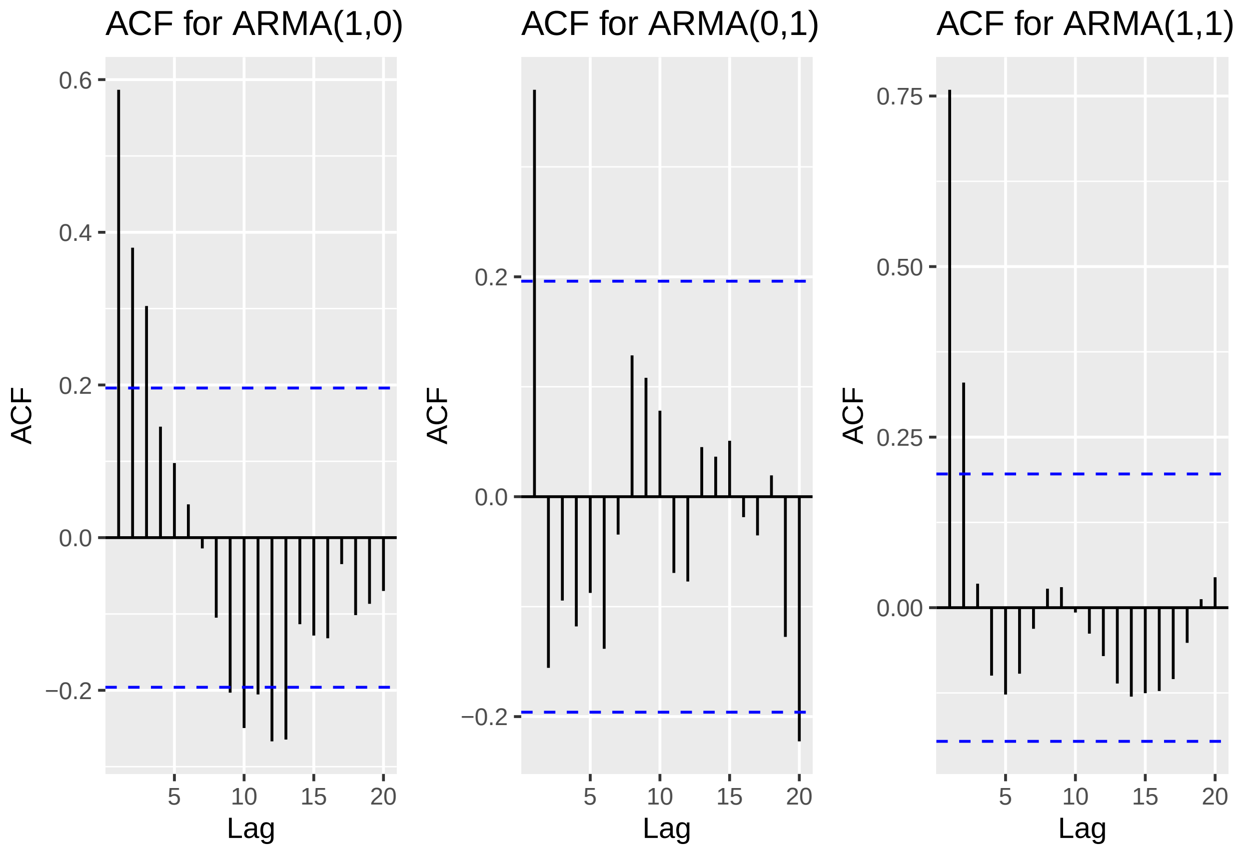



```r
# ARMA(1,1)
arma11 <- arima.sim(n = 100, list(ar = 0.6, ma = 0.9))
autoplot(arma11) + ggtitle("ARMA(1,1)")
```



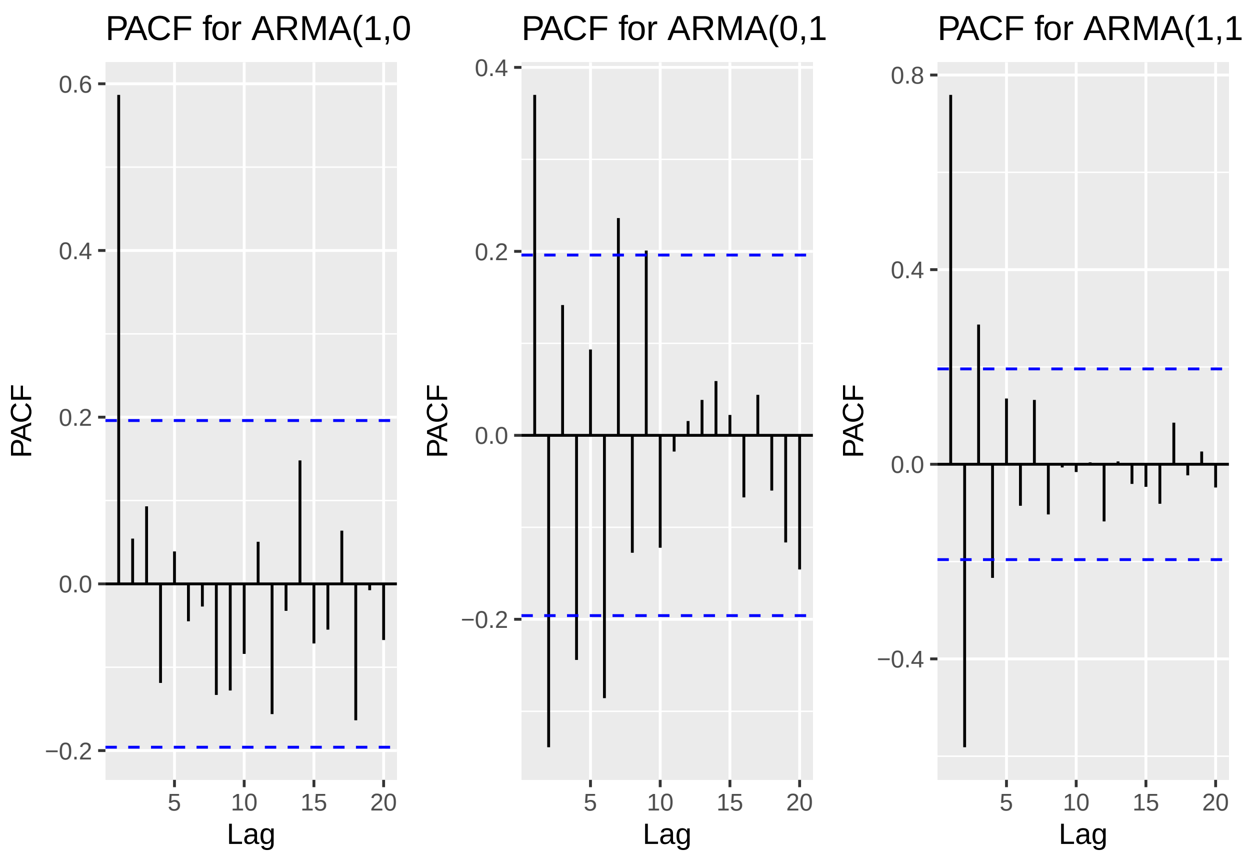
(b) Plot the sample ACF for each of these models in one window to facilitate comparison (Hint: use `cowplot::plot_grid()`).

```
plot_grid(
  autoplot(Acf(arma10, plot = FALSE)) + ggtitle("ACF for ARMA(1,0)"),
  autoplot(Acf(arma01, plot = FALSE)) + ggtitle("ACF for ARMA(0,1)"),
  autoplot(Acf(arma11, plot = FALSE)) + ggtitle("ACF for ARMA(1,1)"),
  nrow = 1
)
```



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

```
plot_grid(
  autoplot(Pacf(arma10, plot = FALSE)) + ggtitle("PACF for ARMA(1,0)"),
  autoplot(Pacf(arma01, plot = FALSE)) + ggtitle("PACF for ARMA(0,1)"),
  autoplot(Pacf(arma11, plot = FALSE)) + ggtitle("PACF for ARMA(1,1)"),
  nrow = 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: ARMA(1,0): PACF cuts off after lag 1, indicating an AR(1) process. ARMA(0,1): ACF cuts off after lag 1, indicating an MA(1) process. ARMA(1,1): Both ACF and PACF gradually decay, indicating a mixed ARMA(1,1) process.

(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?

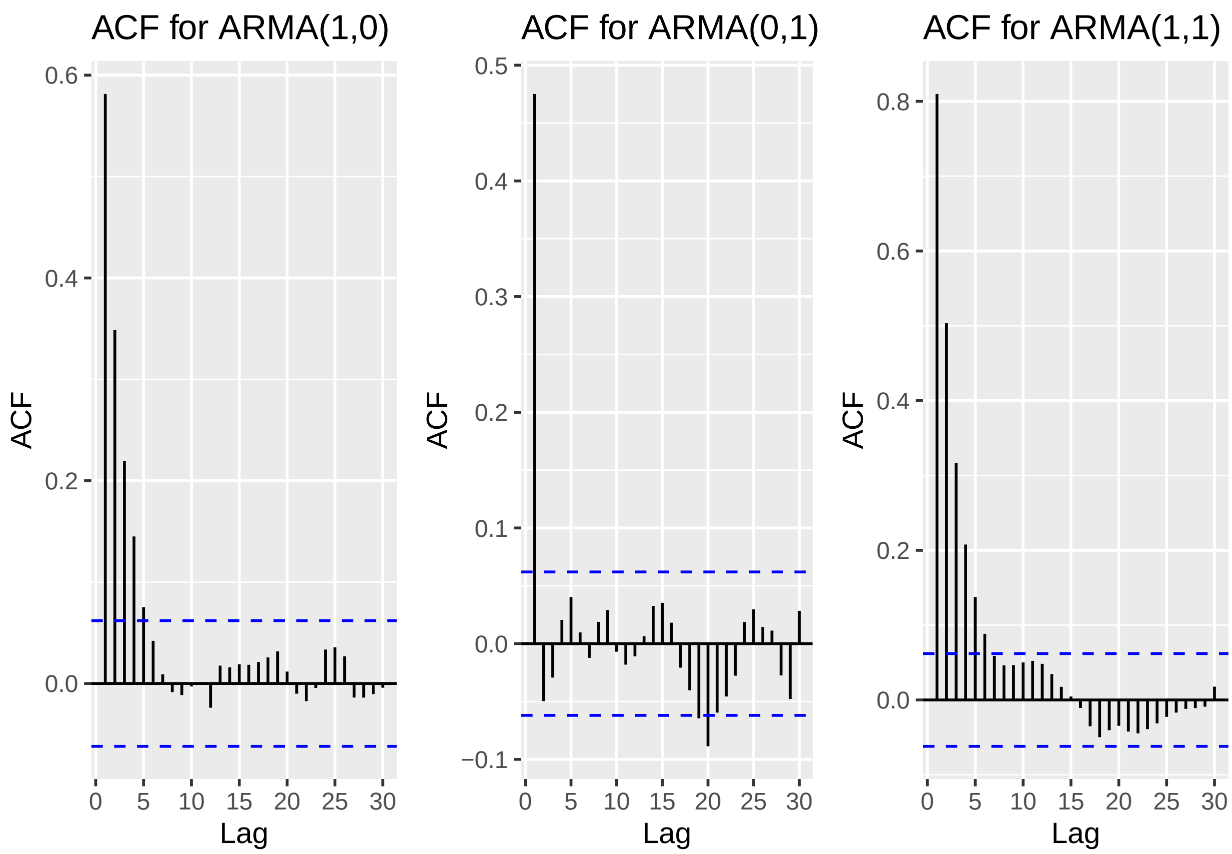
Answer:For ARMA(1,0), the PACF at lag 1 should be close to the AR coefficient $\phi = 0.6$. For ARMA(1,1), the PACF at lag 1 may not match exactly due to the presence of the MA component, but it should still provide some indication of the AR component.

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

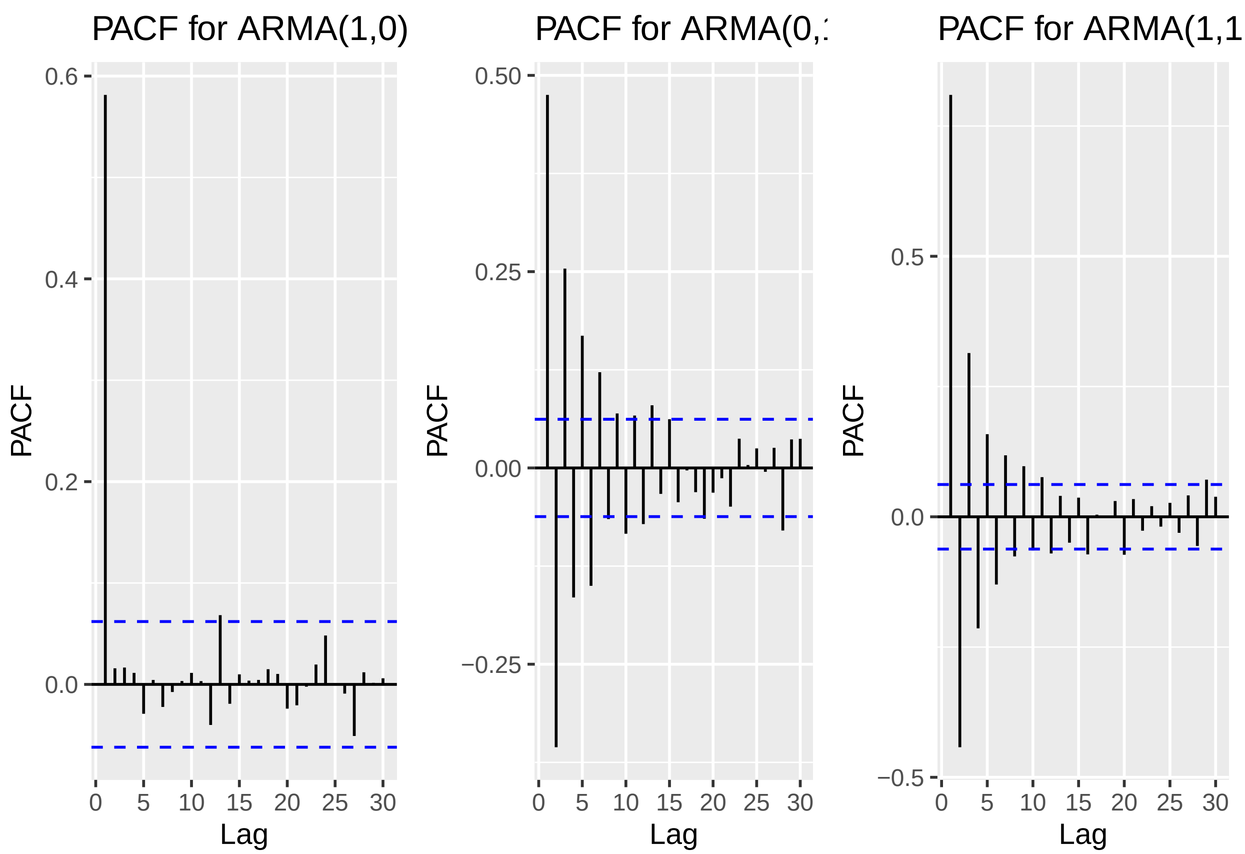
```r
# Increase n to 1000
arma10 <- arima.sim(n = 1000, list(ar = 0.6))
arma01 <- arima.sim(n = 1000, list(ma = 0.9))
arma11 <- arima.sim(n = 1000, list(ar = 0.6, ma = 0.9))

# ACF plots
plot_grid(
```

```r
  autoplot(Acf(arma10, plot = FALSE)) + ggtitle("ACF for ARMA(1,0)"),
  autoplot(Acf(arma01, plot = FALSE)) + ggtitle("ACF for ARMA(0,1)"),
  autoplot(Acf(arma11, plot = FALSE)) + ggtitle("ACF for ARMA(1,1)"),
  nrow = 1
)
```




```r
# PACF plots
plot_grid(
  autoplot(Pacf(arma10, plot = FALSE)) + ggtitle("PACF for ARMA(1,0)"),
  autoplot(Pacf(arma01, plot = FALSE)) + ggtitle("PACF for ARMA(0,1)"),
  autoplot(Pacf(arma11, plot = FALSE)) + ggtitle("PACF for ARMA(1,1)"),
  nrow = 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, s=12

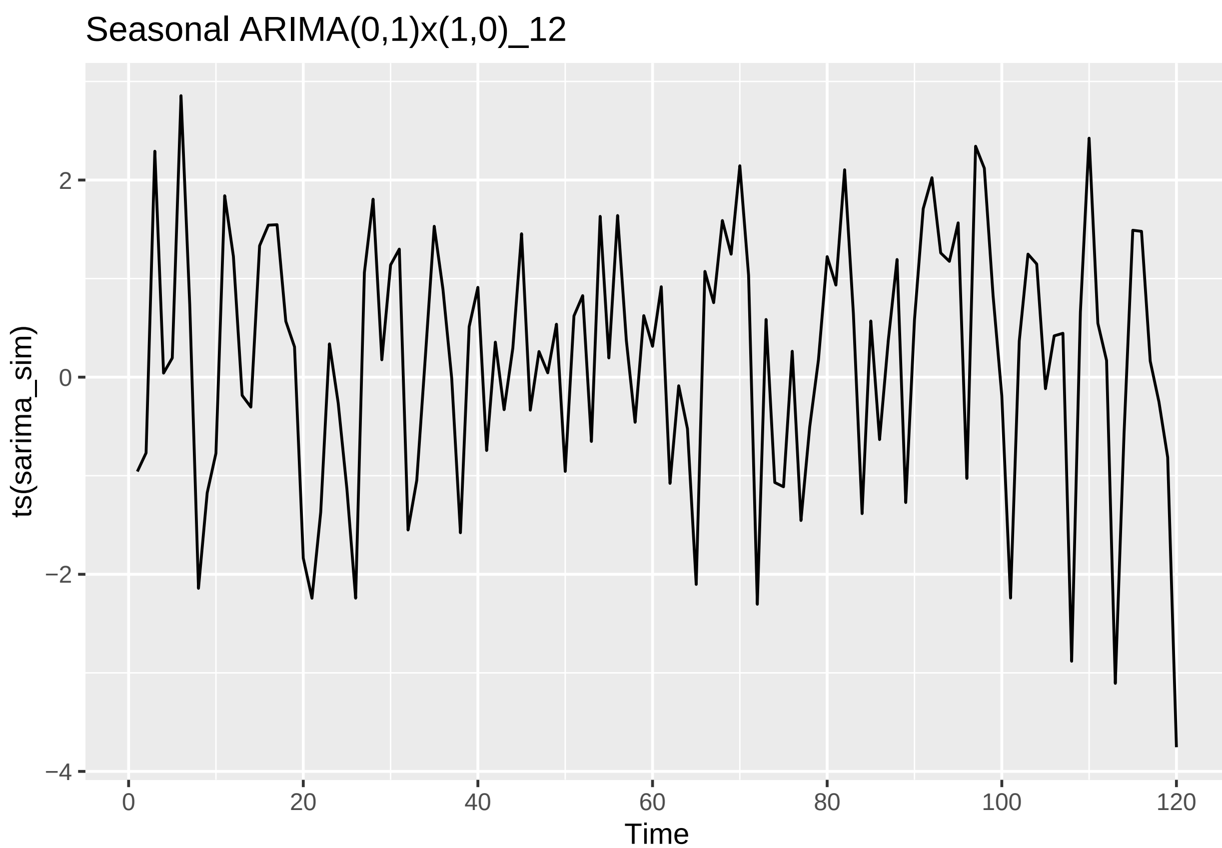
(b) Also from the equation what are the values of the parameters, i.e., model coefficients.

non-seasonal AR coefficient: 0.7 non-seasonal MA coefficient: -0.1 seasonal AR coefficient: -0.25

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?

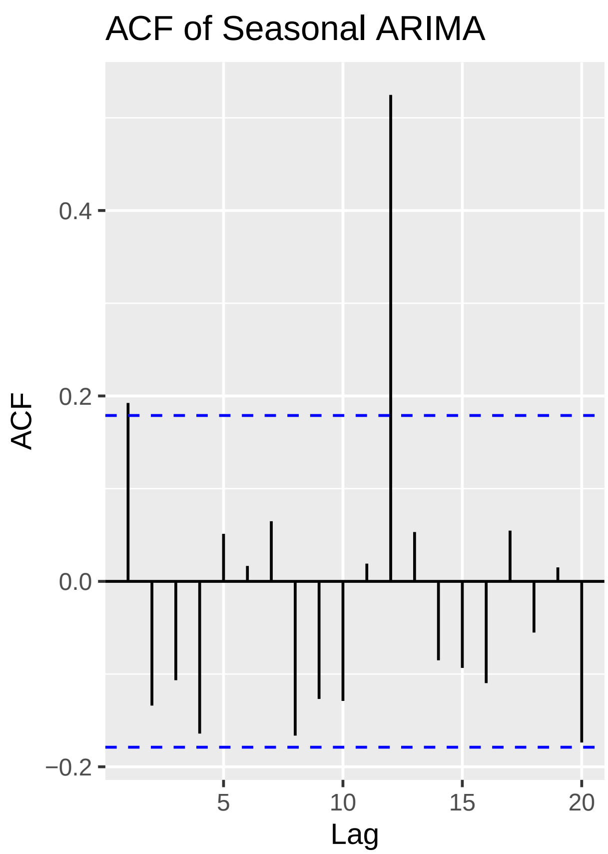
```r
set.seed(123)
sarima_sim <- sim_sarima(n = 120, model = list(sar=0.8, ma=0.5, iorder=0, siorder=0, nseasons=12))
autoplot(ts(sarima_sim)) + ggtitle("Seasonal ARIMA(0,1)x(1,0)_12")
```

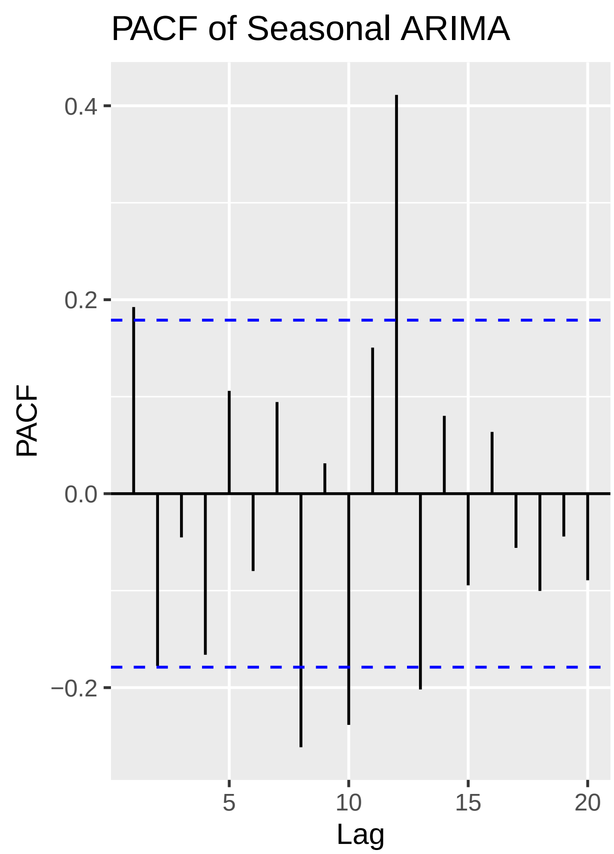


Q5

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 and PACF plots
plot_grid(
  autoplot(Acf(sarima_sim, plot = FALSE)) + ggtitle("ACF of Seasonal ARIMA"),
  autoplot(Pacf(sarima_sim, plot = FALSE)) + ggtitle("PACF of Seasonal ARIMA"),
  nrow = 1
)
```


ACF of Seasonal ARIMA
ACF
0.4
0.2
0.0
−0.2
5
10
15
20
Lag


PACF of Seasonal ARIMA
PACF
0.4
0.2
0.0
−0.2
5
10
15
20
Lag