ENV 797 - Time Series Analysis for Energy and Environment Applications | Spring 2025

Assignment 6 - Due date 02/27/25

Lucy Wang

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_Sp25.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(forecast)
library(tseries)
library(ggplot2)
library(Kendall)
library(tidyverse)
library(dplyr)
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 for AR of order 2 decays gradually without a sharp cutoff point. The PACF shows significant spikes at lag 1 and lag 2, and it drops off to near zero showing a cutoff after lag 2.

- MA(1)

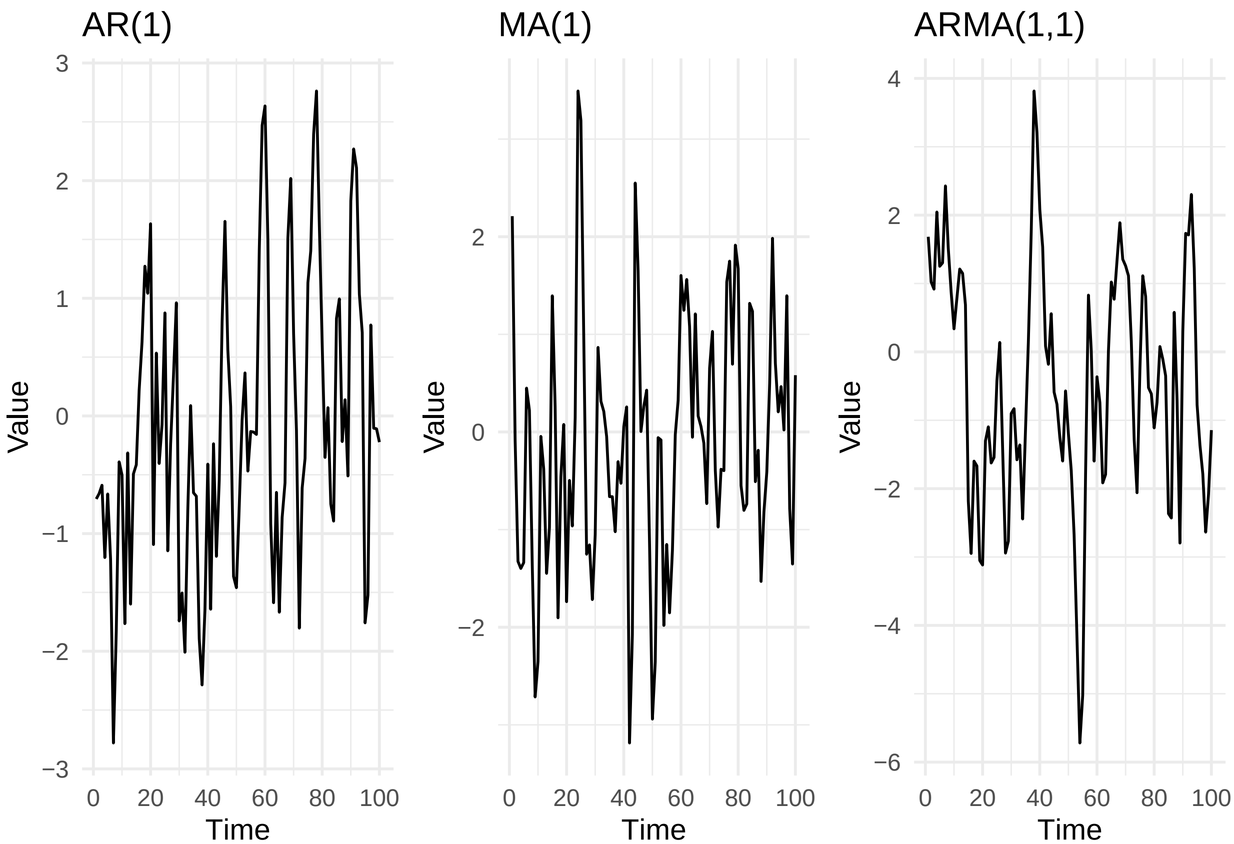
Answer: The ACF for MA of order 1 has a significiant spike at lag 1 a cutoff at lag 1. The PACF shows a gradual exponential decay after lag 1 without a sharp cutoff.

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).

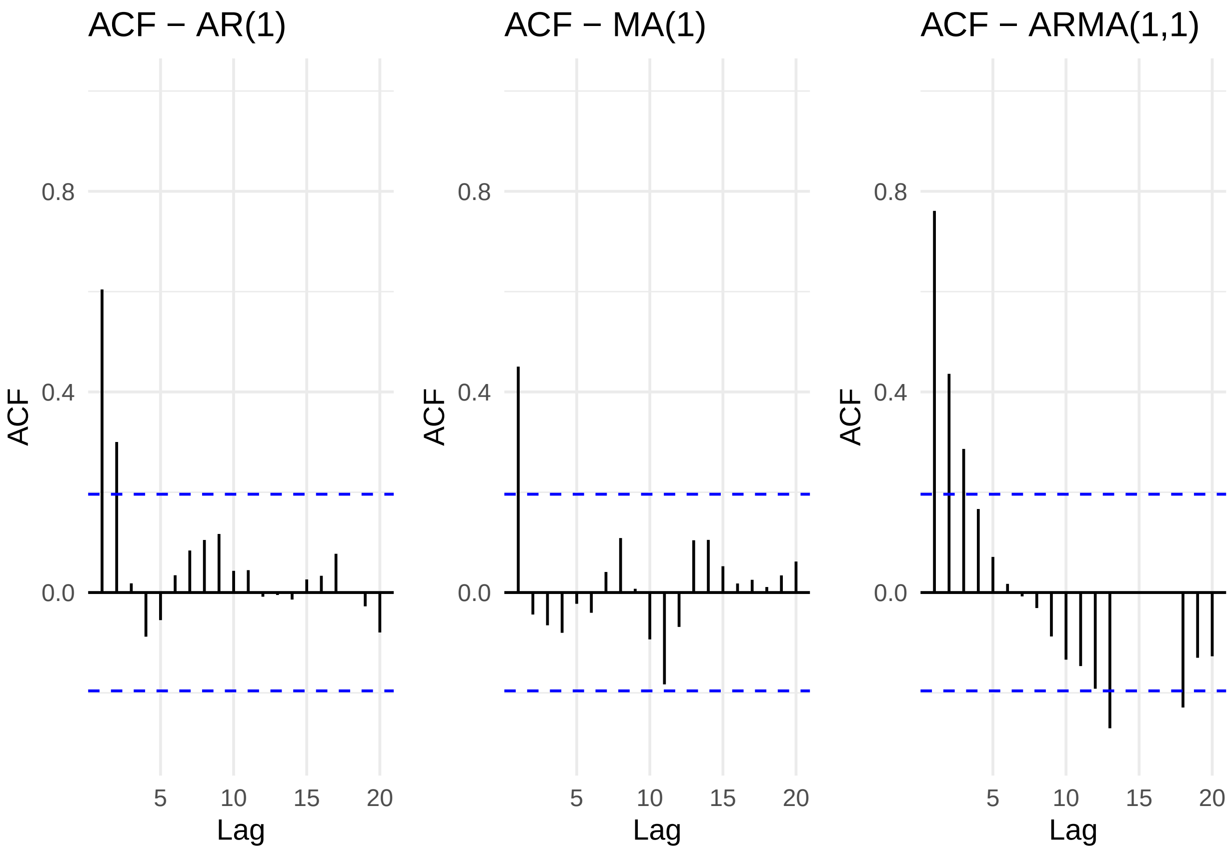
(a) Consider three models: ARMA(1,0), ARMA(0,1) and ARMA(1,1) with parameters $\phi = 0.6$ and $\theta = 0.9$. The ϕ refers to the AR coefficient and the θ refers to the MA coefficient. Use the `arima.sim()` function in R to generate $n = 100$ observations from each of these three models. Then, using `autoplot()` plot the generated series in three separate graphs.

```r
# Set seed
set.seed(101)

# ARMA(1,0) Model
arma_1_0 <- arima.sim(n = 100, list(ar = 0.6))
arma_1_0_plot <- autoplot(arma_1_0) +
  ggtitle("AR(1)") +
  ylab("Value") + xlab("Time") +
  theme_minimal()

# ARMA(0,1) Model
arma_0_1 <- arima.sim(n = 100, list(ma = 0.9))
arma_0_1_plot <- autoplot(arma_0_1) +
  ggtitle("MA(1)") +
  ylab("Value") + xlab("Time") +
  theme_minimal()

# ARMA(1,1) Model
arma_1_1 <- arima.sim(n = 100, list(ar = 0.6, ma = 0.9))
arma_1_1_plot <- autoplot(arma_1_1) +
  ggtitle("ARMA(1,1)") +
  ylab("Value") + xlab("Time") +
  theme_minimal()

plot_grid(arma_1_0_plot, arma_0_1_plot, arma_1_1_plot, ncol = 3)
```



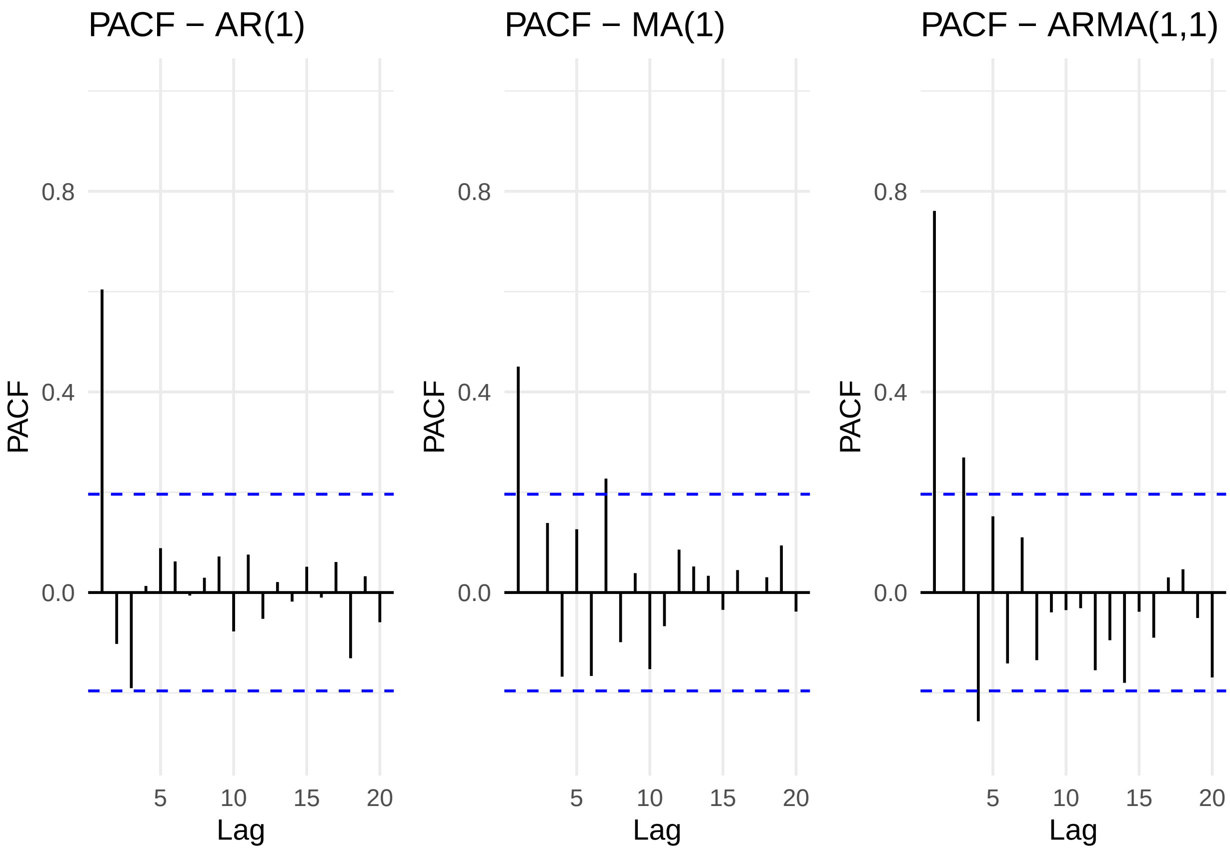
(b) Plot the sample ACF for each of these models in one window to facilitate comparison (Hint: use `cowplot::plot_grid()`).

```
plot_acf <- function(m1, m2, m3){
  # ACF for ARMA(1,0)
  acf_arma_1_0 <- ggAcf(m1, lag.max = 20) +
    ggtitle("ACF - AR(1)") +
    ylim(c(-0.3,1))+
    theme_minimal()

  #  ACF for ARMA(0,1)
  acf_arma_0_1 <- ggAcf(m2, lag.max = 20) +
    ggtitle("ACF - MA(1)") +
    ylim(c(-0.3,1))+
    theme_minimal()

  # ACF for ARMA(1,1)
  acf_arma_1_1 <- ggAcf(m3, lag.max =20) +
    ggtitle("ACF - ARMA(1,1)") +
    ylim(c(-0.3,1))+
    theme_minimal()

  plot_grid(acf_arma_1_0, acf_arma_0_1, acf_arma_1_1, ncol = 3)
}

plot_acf(arma_1_0, arma_0_1, arma_1_1)
```



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

```
plot_pacf <- function(m1, m2, m3){

  # PACF for ARMA(1,0)
  pacf_arma_1_0 <- ggPacf(m1, lag.max = 20) +
    ggtitle("PACF - AR(1)") +
    ylim(c(-0.3,1))+
    theme_minimal()

  # PACF for ARMA(0,1)
  pacf_arma_0_1 <- ggPacf(m2, lag.max = 20) +
    ggtitle("PACF - MA(1)") +
    ylim(c(-0.3,1))+
    theme_minimal()

  # PACF for ARMA(1,1)
  pacf_arma_1_1 <- ggPacf(m3, lag.max = 20) +
    ggtitle("PACF - ARMA(1,1)") +
    ylim(c(-0.3,1))+
    theme_minimal()

  plot_grid(pacf_arma_1_0, pacf_arma_0_1, pacf_arma_1_1, ncol = 3)
}

plot_pacf(arma_1_0, arma_0_1, arma_1_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:

AR(1) model: The PACF cuts off after lag 1, which indicates possibility of an AR model of order 1. However, there is seemingly a cutoff after lag 2 in the ACF plot without an obvious observation of a gradual decay. Therfore, it is a bit hard to tell if it is an AR model of order 1.

MA(1) model: The ACF plot shows a crisp cutoff after lag 1. The PACF shows a decay, though the decay is not as gradual as expected. We can basically confirm this is an MA model of order 1.

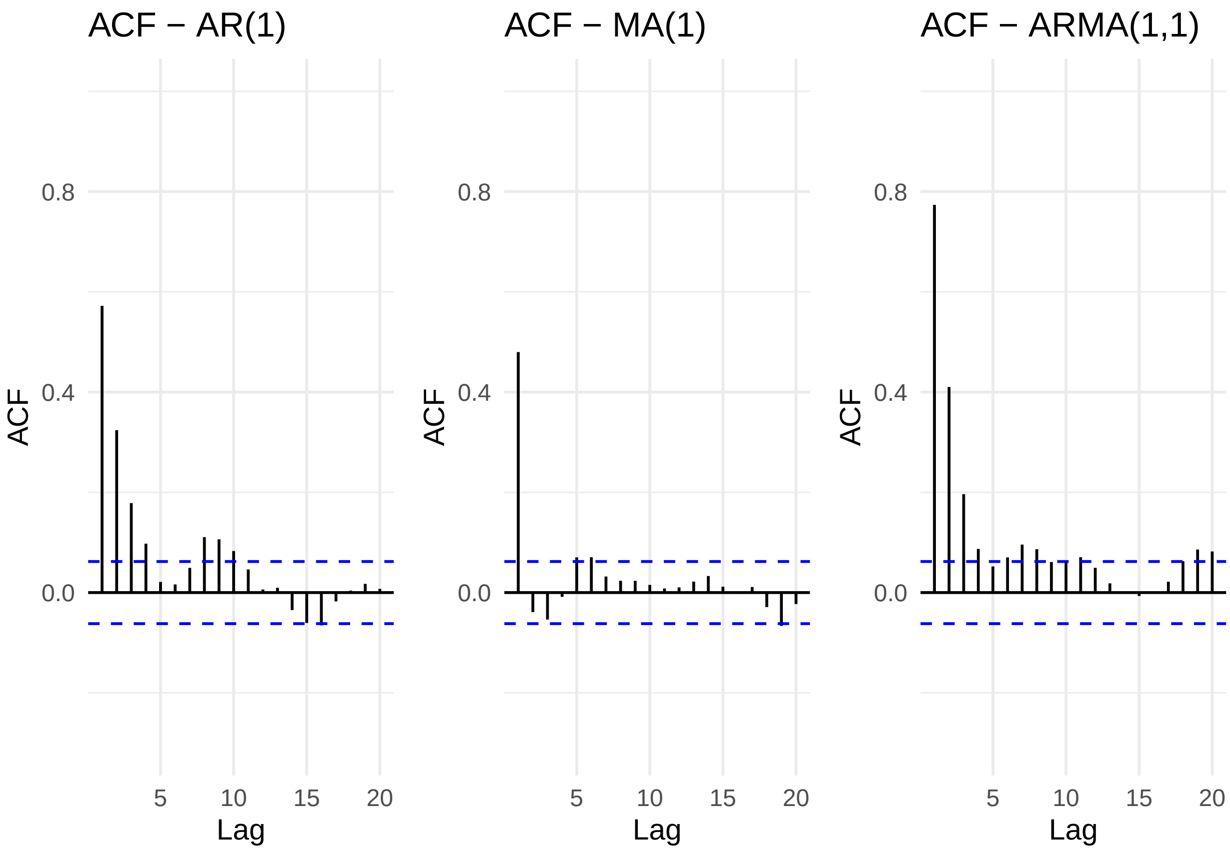
ARMA model: The ACF and PACF plots shows a mix of bahaviors that could be in ACF or PACF plots. Both ACF and PACF plots show a gradual decay, which indicates that the model may be influenced by both AR and MA components. Therefore, this is an ARMA model.

(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?

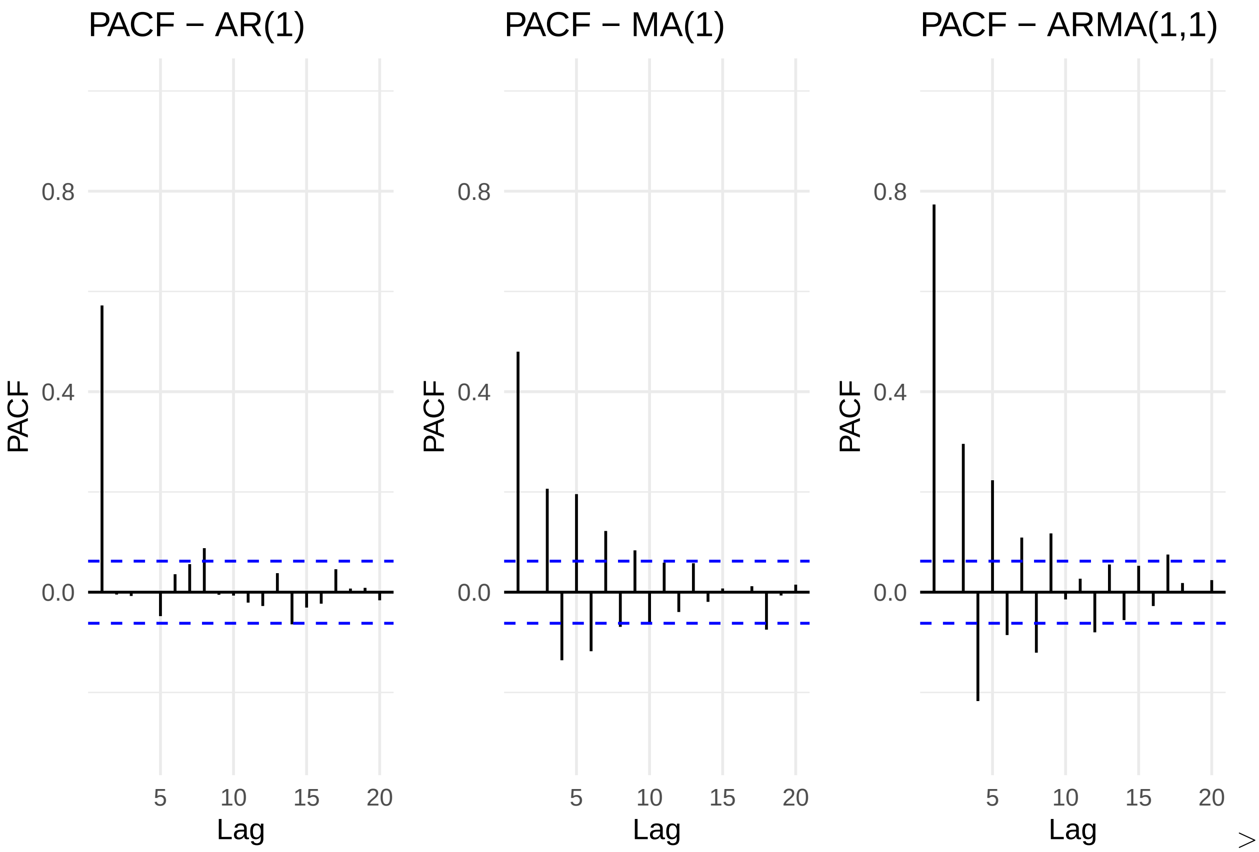
Answer: $\phi = 0.6$ does match with what's shown on the PACF plot for ARMA(1,0) but not necessarily for ARMA (1,1). For the ARMA (1,0) model, the spike on the PACF plot at lag 1 is very close to 0.6. They should match becasue it represents the partial correlation with its lagged value after removing the effect of intermediate lags. However, the PACF value is close to 0.8 rather than 0.6 for the ARMA (1,1) model. Since it is affected by both AR and MA components, the correlation at lag 1 is influenced by both ϕ and θ. Therefore, the actual value falls in between. The PACF for ARMA(1,1) should not match exactly with $\phi = 0.6$.

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

```r
arma_1_0_f <- arima.sim(n = 1000, list(ar = 0.6))
arma_0_1_f <- arima.sim(n = 1000, list(ma = 0.9))
arma_1_1_f <- arima.sim(n = 1000, list(ar = 0.6, ma = 0.9))

plot_acf(arma_1_0_f, arma_0_1_f, arma_1_1_f)
```




```r
plot_pacf(arma_1_0_f, arma_0_1_f, arma_1_1_f)
```



Answer: The plots show much clearer patterns to identify the AR/MA/ARMA models with 1000 observations. The decays are more gradual, and the cutoffs are more sharp after n=1000.

AR(1) model: The ACF shows a gradual decay rather than a confusing cutoff after lag 3. The PACF cuts off sharply after lag 1. So we could easily conclude that this is an AR model.

MA(1) model: The ACF plot shows a crisp cutoff after lag 1, while the PACF shows a decay. We can confirm this is an MA model.

ARMA model: Similarly to previous observation, the ACF and PACF plots shows a mix of bahaviors that could be in ACF or PACF plots. Both ACF and PACF plots show a gradual decay, which indicates that the model may be influenced by both AR and MA components. This is an ARMA model.

Answer: Similar to my previous answer in (e), $\phi = 0.6$ matches with what's shown on the PACF plot for ARMA(1,0) but not necessarily for ARMA (1,1). For the ARMA (1,0) model, the spike on the PACF plot at lag 1 is close to 0.6 as they should. The PACF value becomes closer to 0.8 for the ARMA (1,1) model as the PACF for ARMA(1,1) should not match exactly with $\phi = 0.6$. Since it is affected by both AR and MA components, the correlation at lag 1 is influenced by both ϕ and θ. Therefore, the actual value falls in between (0.6 and 0.9).

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$(1, 0, 1)(1, 0, 0)_1 2$

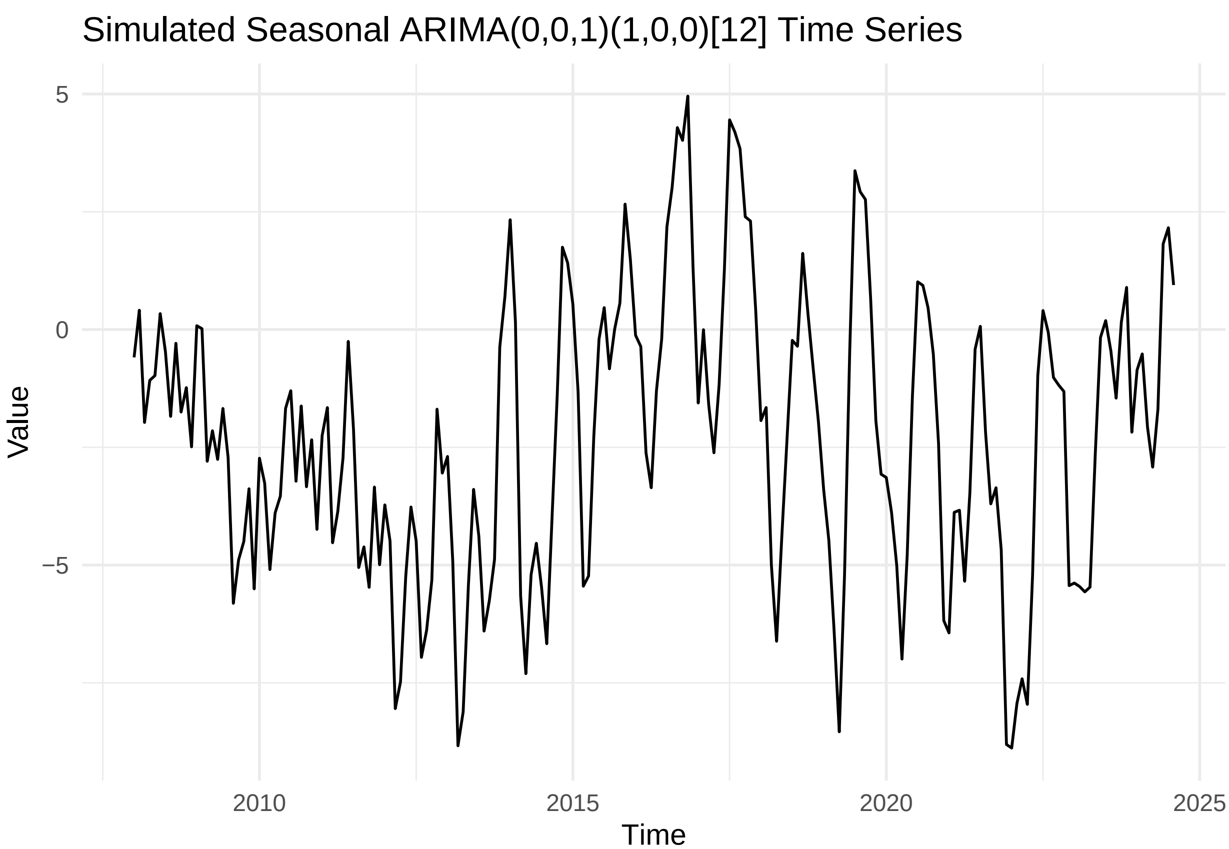
(b) Also from the equation what are the values of the parameters, i.e., model coefficients.

The non-seasonal AR coefficient is 0.7. The seasonal AR coefficient is -0.25. The non-seasonal MA coefficient 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?

```r
# Simulate model
set.seed(101)  # For reproducibility
simulated_data <- sim_sarima(n = 200, model=list(ar = 0.8, ma = 0.5, sar = 0.8, sma = 0, nseasons = 12))

# Convert to time series object
ts_data <- ts(simulated_data, frequency = 12, start = c(2008, 1))

# Plot the time series
autoplot(ts_data) +
  ggtitle("Simulated Seasonal ARIMA(0,0,1)(1,0,0)[12] Time Series") +
  xlab("Time") +
  ylab("Value") +
  theme_minimal()
```

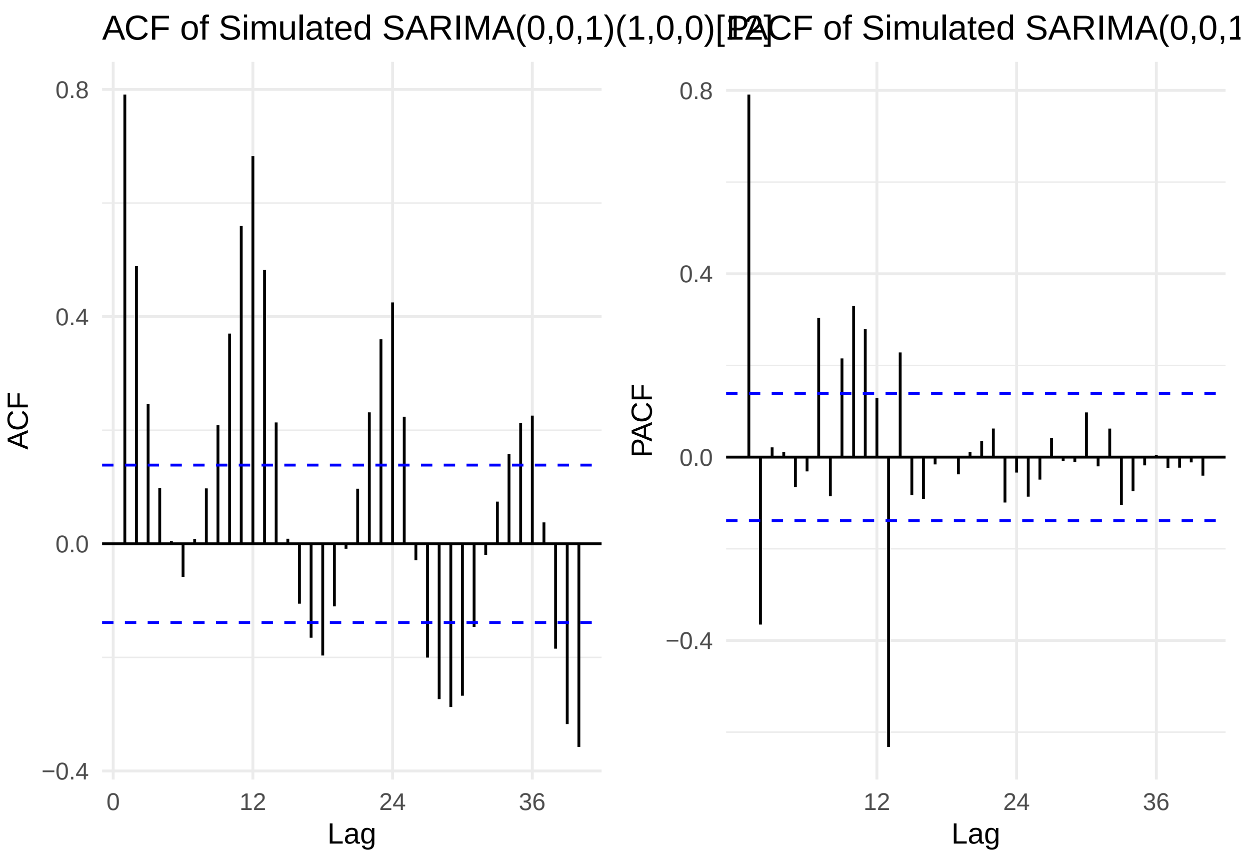


Answer: The plot does show periodic fluctuations, which might indicate seasonality. Though there is a differenced series (D=1), the seasonal pattern still exists.

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.

```
# Plot ACF
acf_plot_Q5 <- ggAcf(ts_data, lag.max = 40) +
  ggtitle("ACF of Simulated SARIMA(0,0,1)(1,0,0)[12]") +
  theme_minimal()

# Plot PACF
pacf_plot_Q5 <- ggPacf(ts_data, lag.max = 40) +
  ggtitle("PACF of Simulated SARIMA(0,0,1)(1,0,0)[12]") +
  theme_minimal()

plot_grid(acf_plot_Q5, pacf_plot_Q5, ncol = 2)
```



Answer: Yes, the plots present the models well. Both non-seasonal and seasonal components cna be identified. The ACF spikes every 12 lags, suggesting strong seasonality. The non-seasonal correlation in the ACF plot also decays gradually. The PACF cuts off after lag 1, indicating the non-seasonal MA of order 1 component. The strong spike at 12 also reflects the seasonal autoregressive component, consistent with the seasonal AR(1) at lag 12.