ENV 797 - Time Series Analysis for Energy and Environment Applications | Spring 2026

Assignment 6 - Due date 02/27/26

Joy Wu

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_Sp26.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.

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

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

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

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

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

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

```r
library(cowplot)
```

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: For AR (2), the most important characteristics of the ACF plot are that it has a gradual decay over time. For PACF, the plot will show that there's a cut off after lag 2, with significant spikes at lag 1 and 2 since this is an autoregressive model of order 2.

- MA(1)

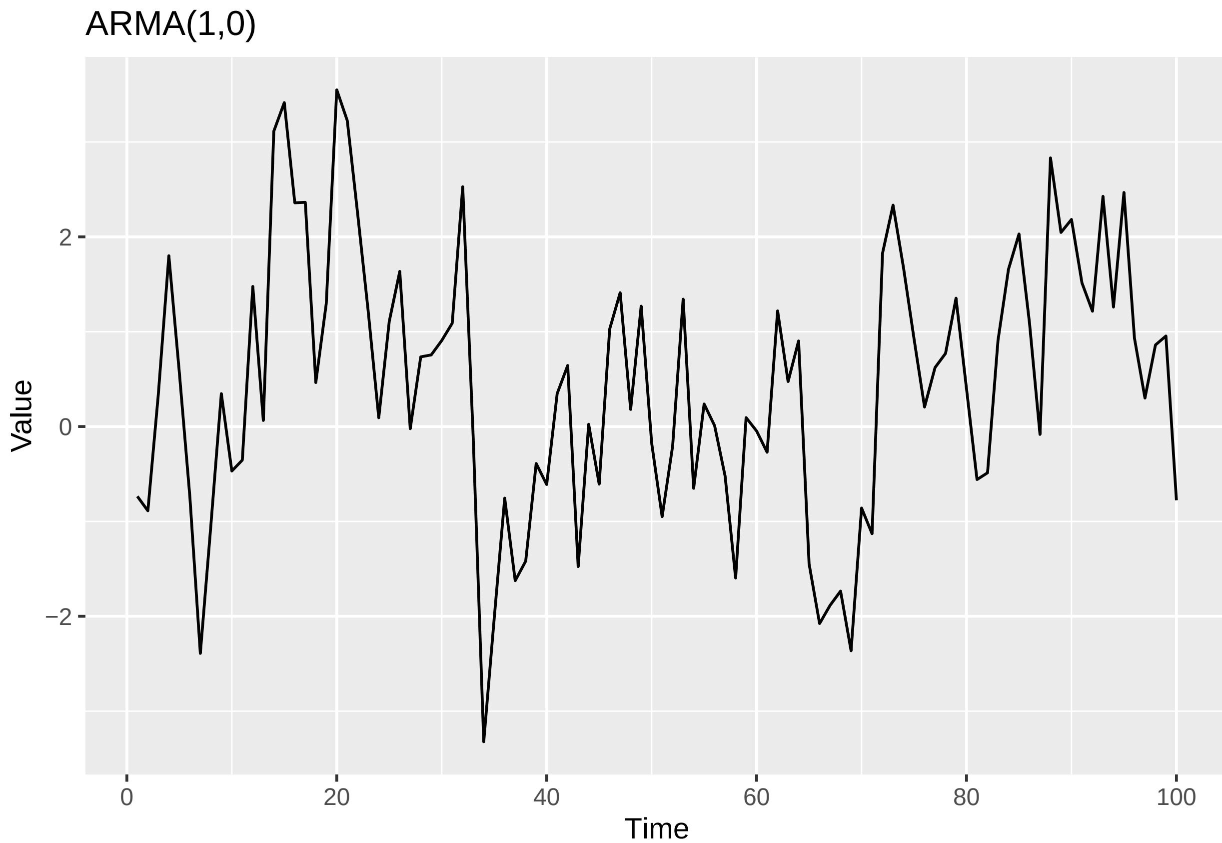
  Answer: For MA (1), the most important characteristics of the ACF plot are that it will show a spike at lag 1 and cuts off after that since MA has an order of 1 here. For PACF, the plot will show that a gradual decay over time.

Q2

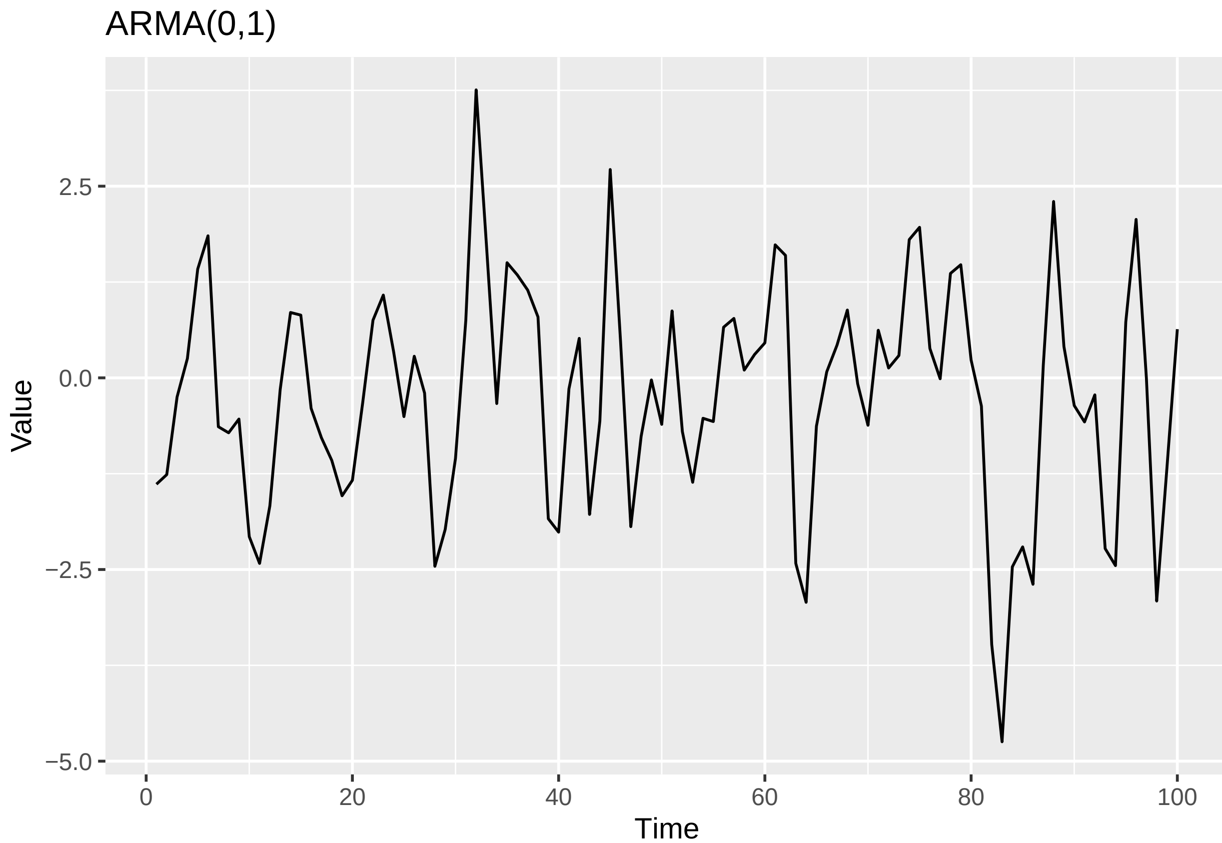
Recall that the non-seasonal ARIMA is described by three parameters ARIMA(p, d, q) where p is the order of the autoregressive component, d is the number of times the series need to be differenced to obtain stationarity and q is the order of the moving average component. If we don't need to difference 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.

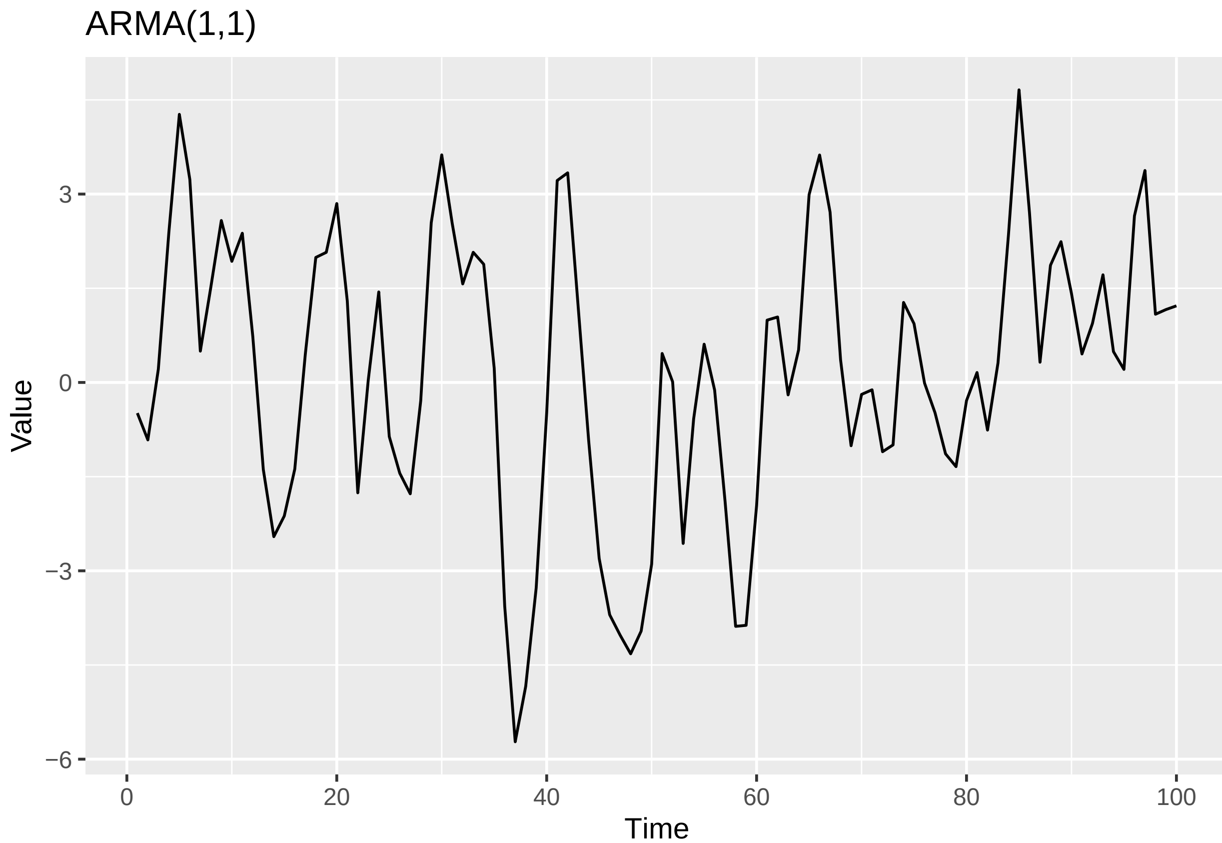
```r
arma.10 <- arima.sim(model = list(ar = 0.6), n = 100)
arma.01 <- arima.sim(model = list(ma = 0.9), n = 100)
arma.11 <- arima.sim(model = list(ar = 0.6, ma = 0.9), n = 100)
autoplot(arma.10) +
  ggtitle("ARMA(1,0)") +
  ylab("Value")
```




```r
autoplot(arma.01) +
  ggtitle("ARMA(0,1)") +
  ylab("Value")
```

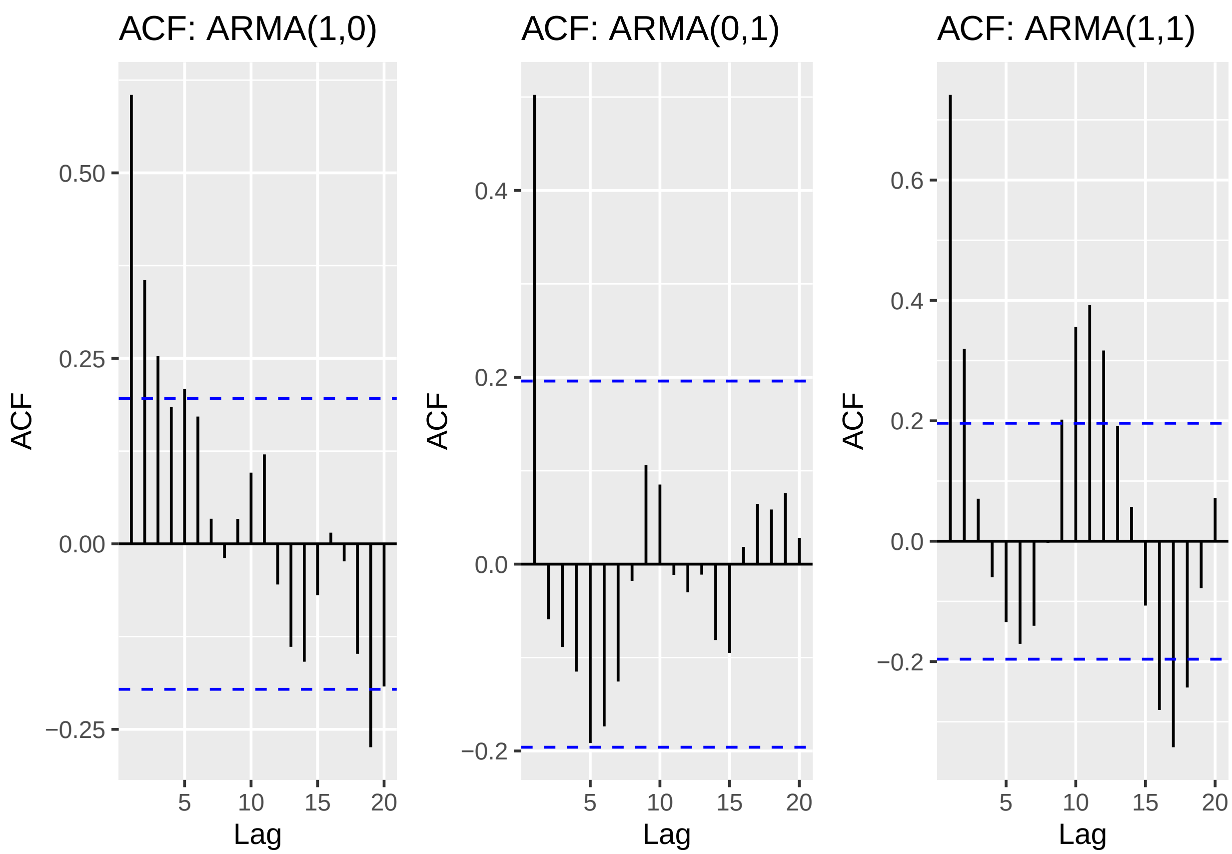



```r
autoplot(arma.11) +
  ggtitle("ARMA(1,1)") +
  ylab("Value")
```



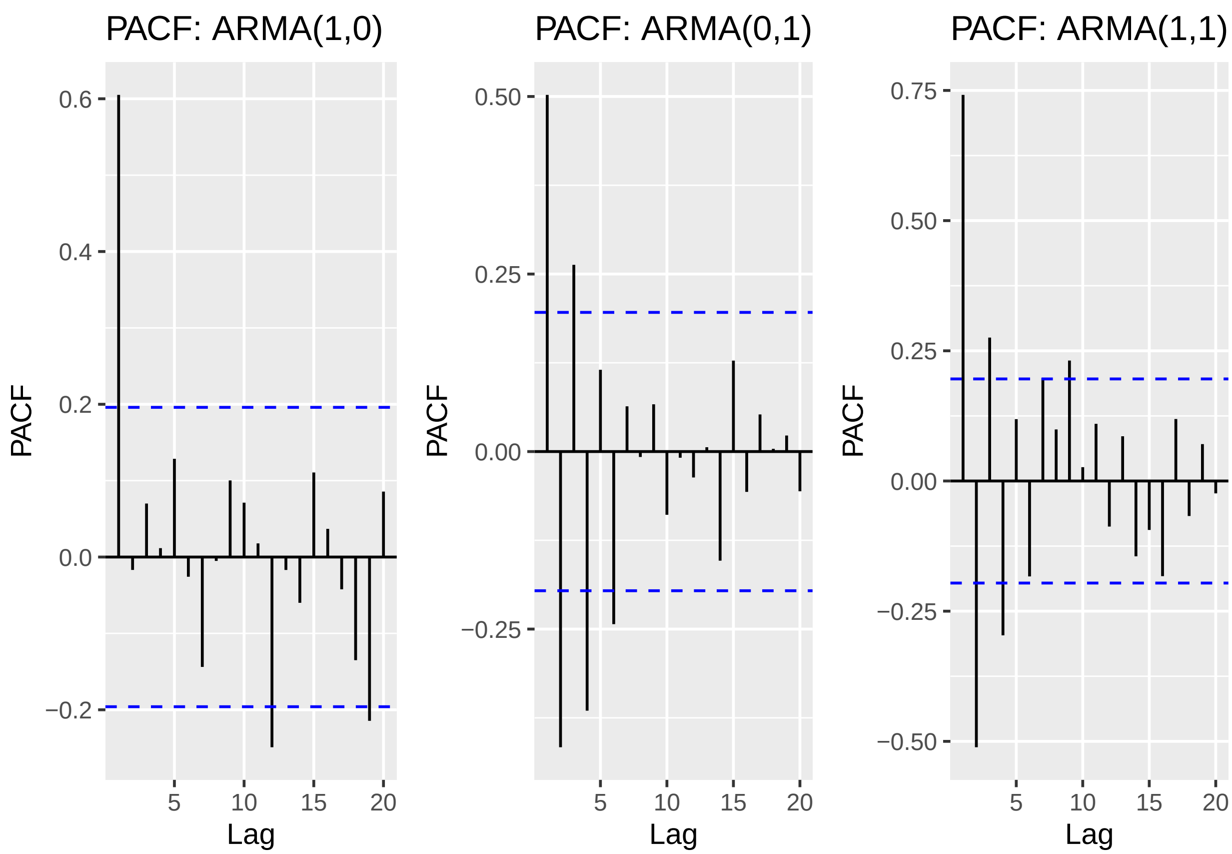
(b) Plot the sample ACF for each of these models in one window to facilitate comparison (Hint: use `cowplot::plot_grid()`).

```
acf.10 <- autoplot(acf(arma.10, plot = FALSE)) +
  ggtitle("ACF: ARMA(1,0)") +
  ylab("ACF") + xlab("Lag")
acf.01 <- autoplot(acf(arma.01, plot = FALSE)) +
  ggtitle("ACF: ARMA(0,1)") +
  ylab("ACF") + xlab("Lag")
acf.11 <- autoplot(acf(arma.11, plot = FALSE)) +
  ggtitle("ACF: ARMA(1,1)") +
  ylab("ACF") + xlab("Lag")
plot_grid(acf.10, acf.01, acf.11, ncol = 3)
```



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

```
pacf.10 <- autoplot(pacf(arma.10, plot = FALSE)) +
  ggtitle("PACF: ARMA(1,0)") +
  ylab("PACF") + xlab("Lag")
pacf.01 <- autoplot(pacf(arma.01, plot = FALSE)) +
  ggtitle("PACF: ARMA(0,1)") +
  ylab("PACF") + xlab("Lag")
pacf.11 <- autoplot(pacf(arma.11, plot = FALSE)) +
  ggtitle("PACF: ARMA(1,1)") +
  ylab("PACF") + xlab("Lag")
plot_grid(pacf.10, pacf.01, pacf.11, ncol = 3)
```



(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.

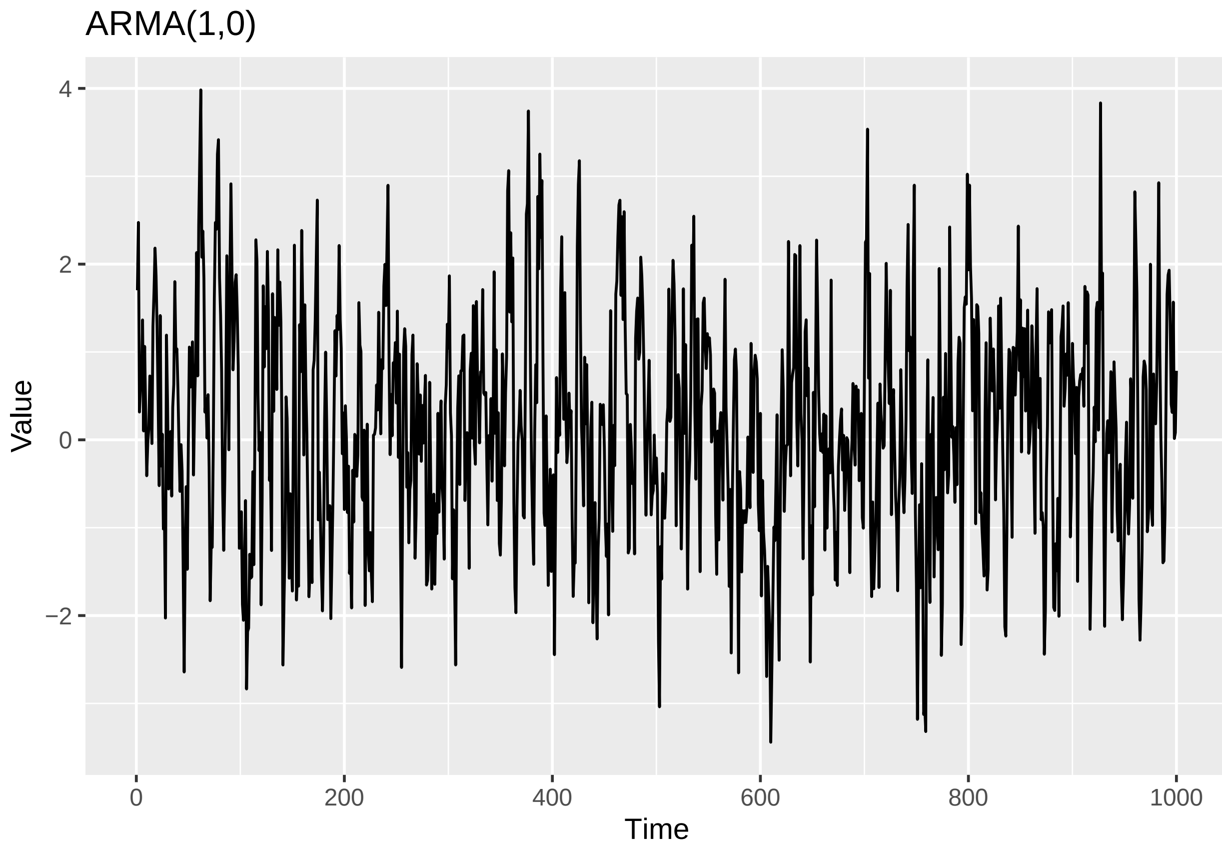
Answer: For ARMA (1,0), we do see a gradual decay in the ACF plot and a sudden cut off after lag 1 in the PACF plot, indicating that it's an AR (1). For ARMA (0,1), we see a gradual decay in the PACF plot and a sudden cut off after lag 1 in the ACF plot, indicating that it's a MA (1). For ARMA (1,1), since both AR and MA components are present, we see that the ACF and PACF plots show similar patterns where they gradually decay over lags.

(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?

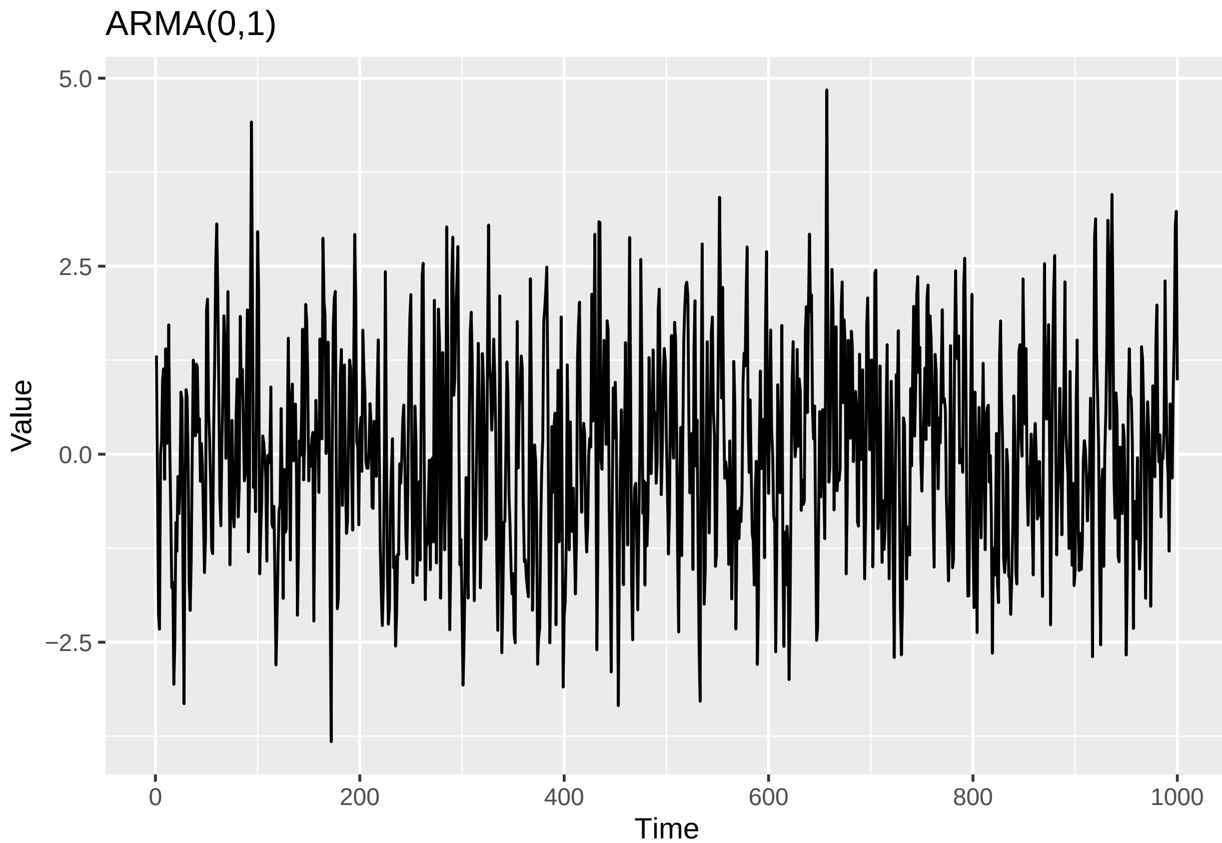
Answer: For ARMA(1,0), yes, they should match, and we can see that from the PACF plot too. The PACF at lag 1 should theoretically be equal to ϕ, which is 0.6 here. For ARMA(1,1), not quite. This is because it has both AR and MA component, so the PACF at lag 1 doesn't quite equal to ϕ and would be influenced by θ, thus not showing exactly 0.6 here.

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

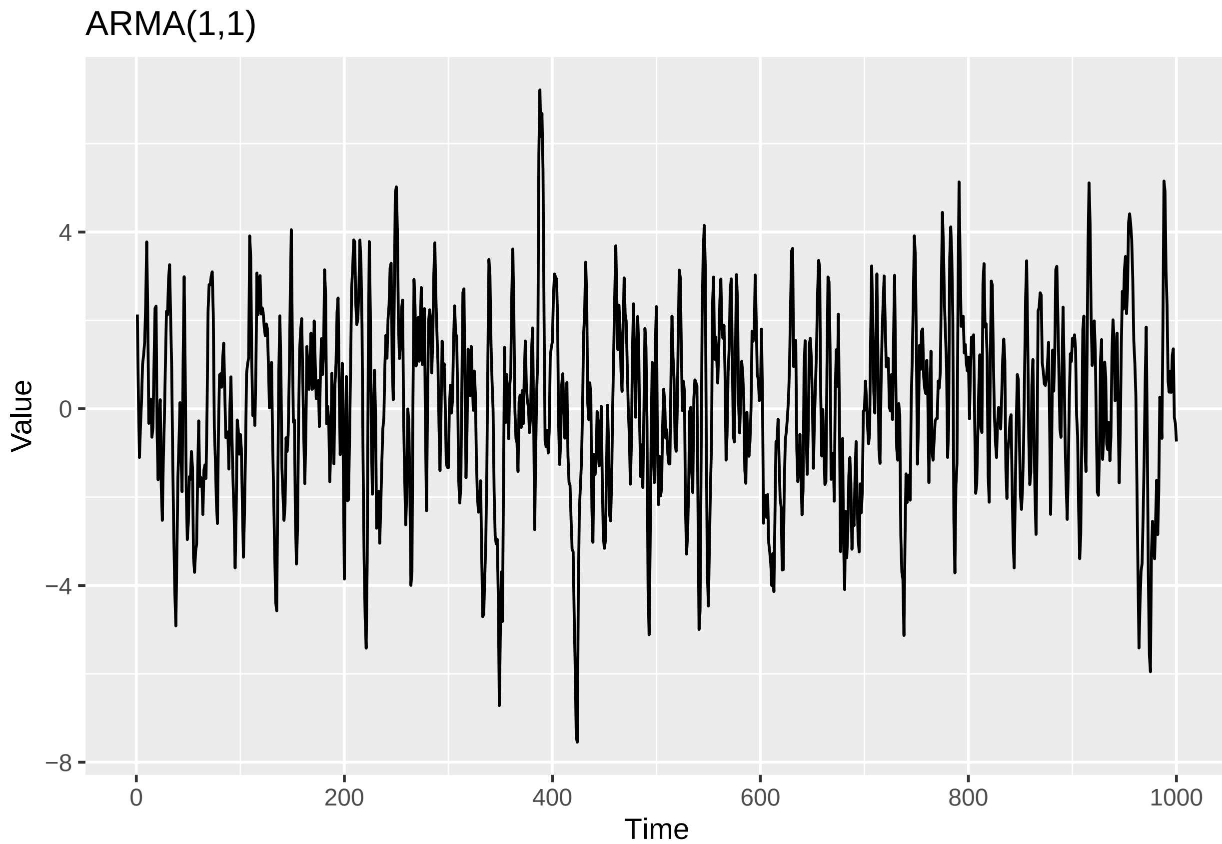
```
arma.10.1000 <- arima.sim(model = list(ar = 0.6), n = 1000)
arma.01.1000 <- arima.sim(model = list(ma = 0.9), n = 1000)
arma.11.1000 <- arima.sim(model = list(ar = 0.6, ma = 0.9), n = 1000)
autoplot(arma.10.1000) +
  ggtitle("ARMA(1,0)") +
  ylab("Value")
```



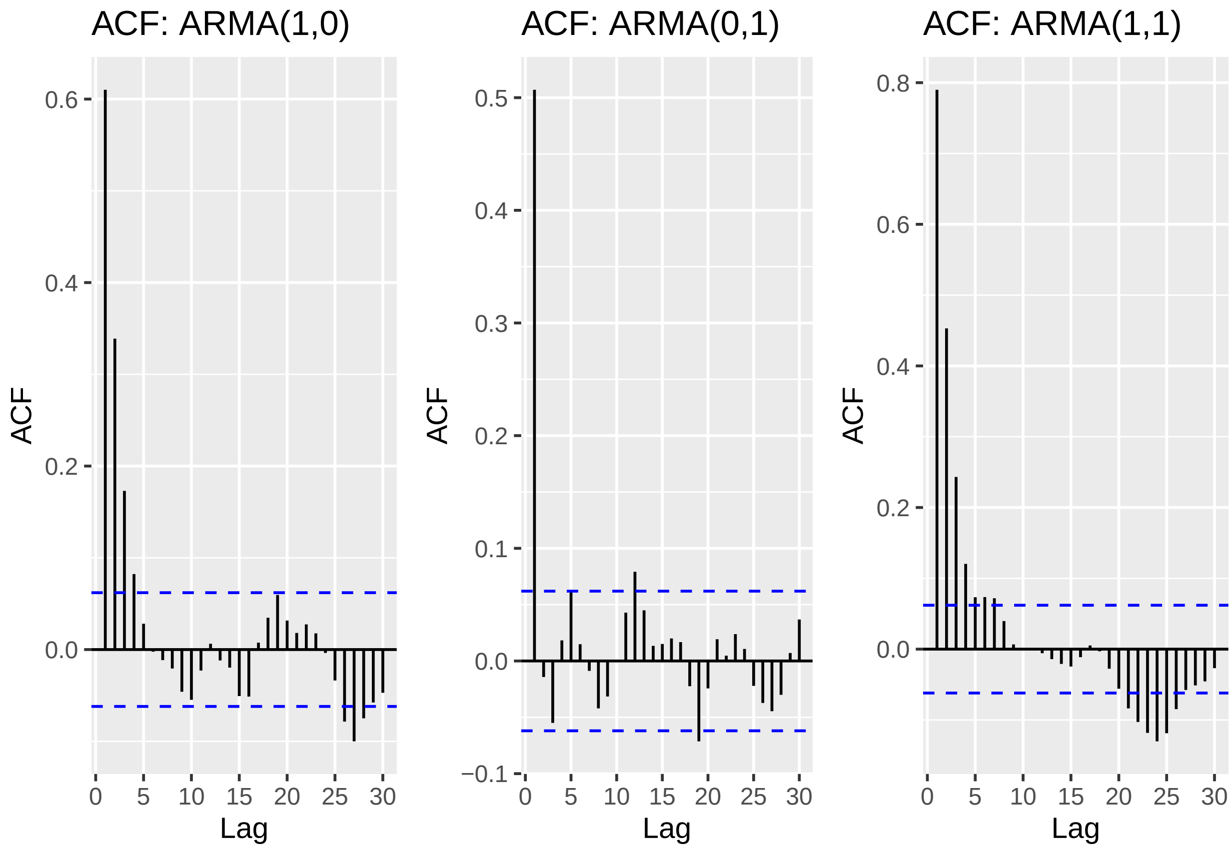

```r
autoplot(arma.01.1000) +
  ggtitle("ARMA(0,1)") +
  ylab("Value")
```



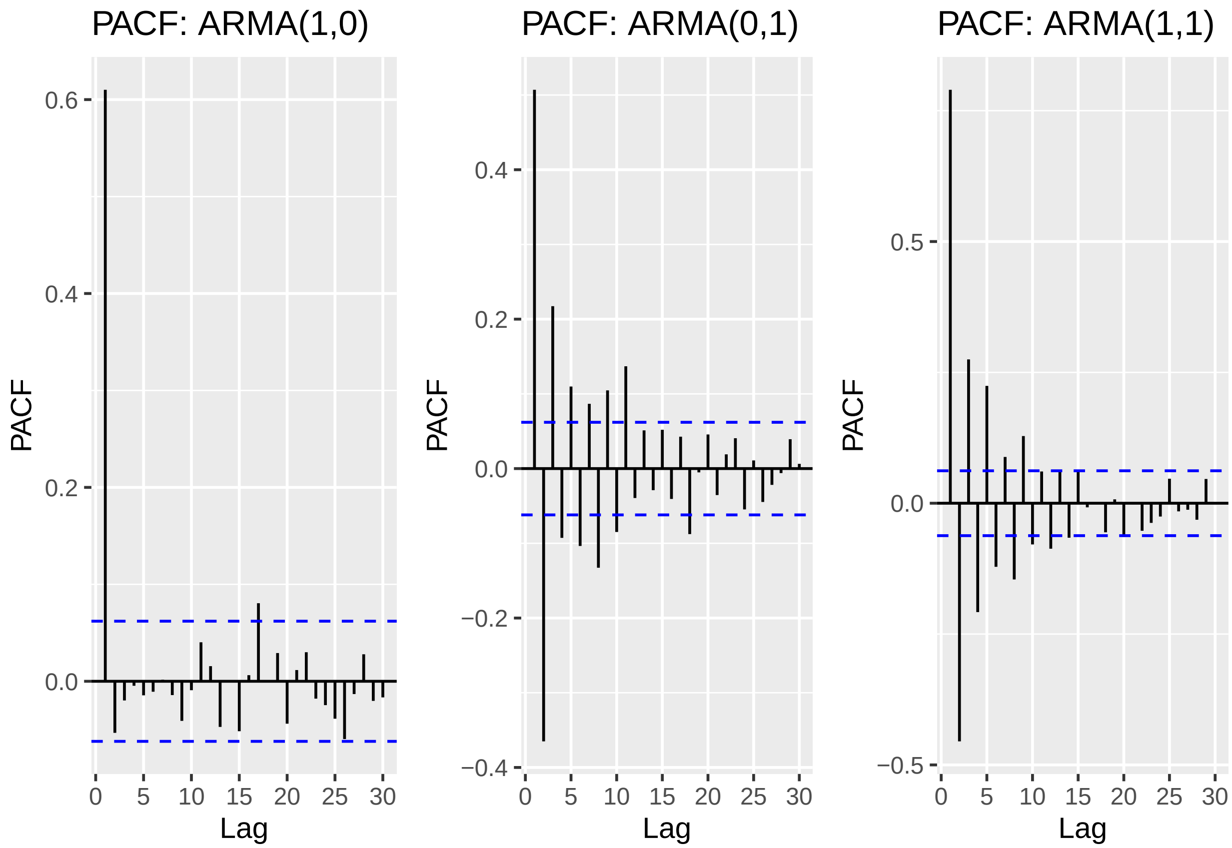

```
autoplot(arma.11.1000) +
  ggtitle("ARMA(1,1)") +
  ylab("Value")
```




```r
acf.10.1000 <- autoplot(acf(arma.10.1000, plot = FALSE)) +
  ggtitle("ACF: ARMA(1,0)") +
  ylab("ACF") + xlab("Lag")

acf.01.1000 <- autoplot(acf(arma.01.1000, plot = FALSE)) +
  ggtitle("ACF: ARMA(0,1)") +
  ylab("ACF") + xlab("Lag")

acf.11.1000 <- autoplot(acf(arma.11.1000, plot = FALSE)) +
  ggtitle("ACF: ARMA(1,1)") +
  ylab("ACF") + xlab("Lag")

plot_grid(acf.10.1000, acf.01.1000, acf.11.1000, ncol = 3)
```




```
pacf.10.1000 <- autoplot(pacf(arma.10.1000, plot = FALSE)) +
  ggtitle("PACF: ARMA(1,0)") +
  ylab("PACF") + xlab("Lag")
pacf.01.1000 <- autoplot(pacf(arma.01.1000, plot = FALSE)) +
  ggtitle("PACF: ARMA(0,1)") +
  ylab("PACF") + xlab("Lag")
pacf.11.1000 <- autoplot(pacf(arma.11.1000, plot = FALSE)) +
  ggtitle("PACF: ARMA(1,1)") +
  ylab("PACF") + xlab("Lag")
plot_grid(pacf.10.1000, pacf.01.1000, pacf.11.1000, ncol = 3)
```



(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 ARMA(1,0), the PACF cutoff after lag 1 is very sharp now and the ACF plot shows a more apparent gradual decay in the ACF plot. For ARMA(0,1), the ACF cutoff after lag 1 largely and PACF plot shows a more apparent gradual decay. For ARMA(1,1), both ACF and PACF still show similar patterns due to having both components present. They both show gradual decay with no obvious spikes and cutoffs.

(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), yes, they should match, and we can see that from the PACF plot too (with slight deviations). The PACF at lag 1 should theoretically be equal to ϕ, which is 0.6 here. For ARMA(1,1), not quite. This is because it has both AR and MA component like previously with 100 observations, so the PACF at lag 1 doesn't quite equal to ϕ and would be influenced by θ, thus not showing exactly 0.6 here.

Q3

Consider the ARIMA model $y_t = 0.7 * y_{t-1} - 0.25 * y_{t-12} + a_t - 0.1 * a_{t-1}$

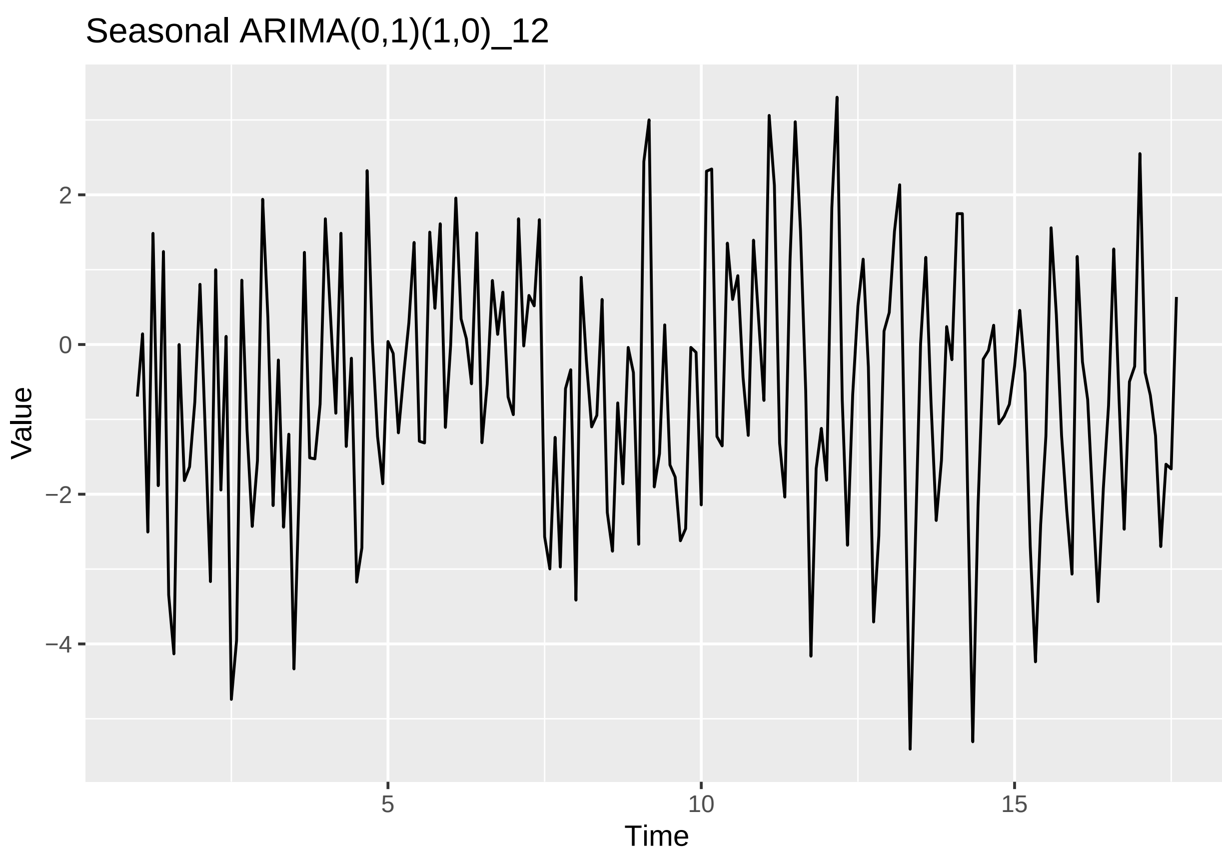
(a) Identify the model using the notation ARIMA$(p, d, q)(P, D, Q)_s$, i.e., identify the integers p, d, q, P, D, Q, s (if possible) from the equation. > Answer: non-seasonal: p = 1, d = 0, q = 1; seasonal: P = 1, s = 12, D = 0, Q = 0

(b) Also from the equation what are the values of the parameters, i.e., model coefficients. >Answer: non-seasonal AR coefficient ($\phi = 0.7$), seasonal AR coefficient ($\phi = -0.25$), non-seasonal MA coefficient ($\theta = 0.1$)

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?

```
sarima.sim <- sim_sarima(n = 200,
                      model = list(ma = 0.5,
                                   sar = 0.8,
                                   nseasons = 12))
sarima.sim.ts <- ts(sarima.sim, frequency = 12, start = c(1,1))

autoplot(sarima.sim.ts) +
  ggtitle("Seasonal ARIMA(0,1)(1,0)_12") +
  ylab("Value")
```

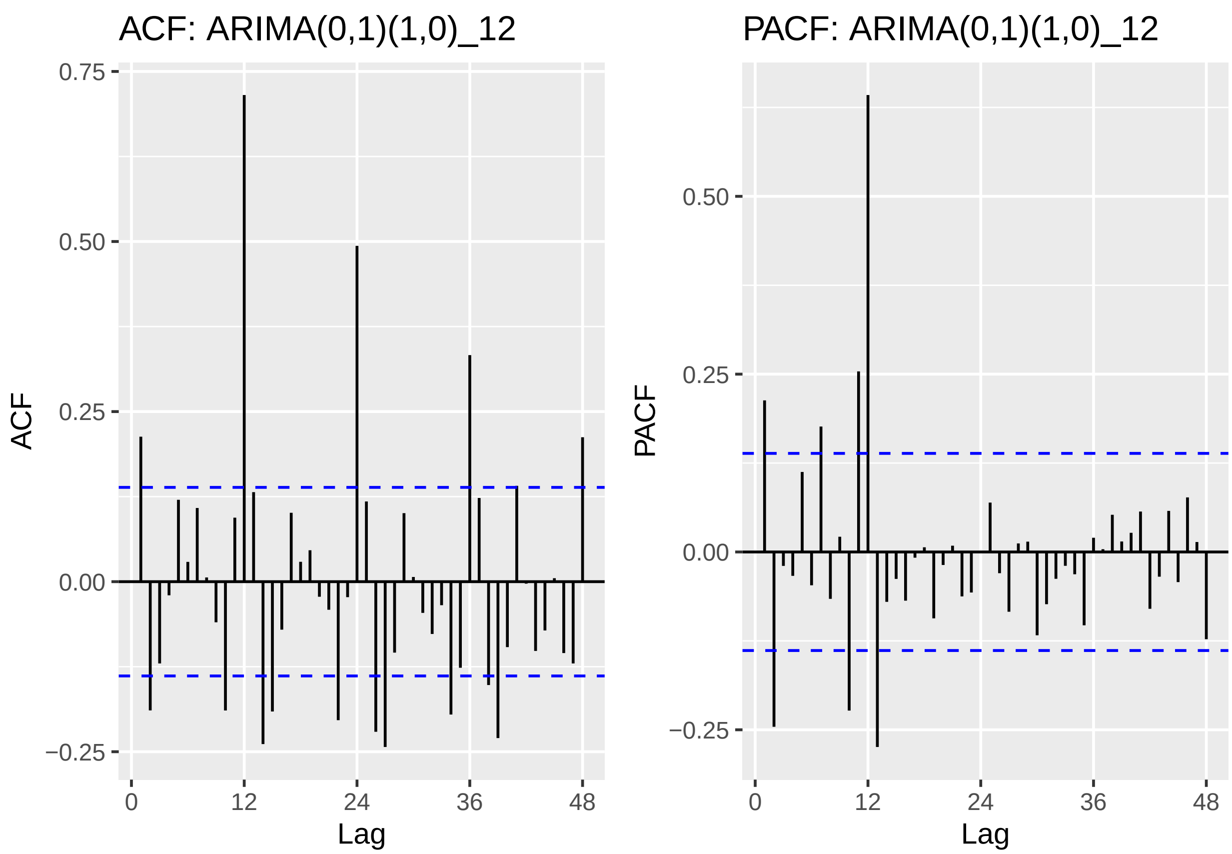


> Answer: Yes, the plot looks seasonal. After turning the series into a time series object, 200 observations turn into roughly 16 ish cycles that show on the x axis of the plot, and we can see that there is a regular pattern (ups and downs) throughout these cycles, indicating seasonal patterns.

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.

```
sarima.acf <- ggAcf(sarima.sim.ts, lag.max = 48) +
  ggtitle("ACF: ARIMA(0,1)(1,0)_12") +
  ylab("ACF") + xlab("Lag")

sarima.pacf <- ggPacf(sarima.sim.ts, lag.max = 48) +
  ggtitle("PACF: ARIMA(0,1)(1,0)_12") +
  ylab("PACF") + xlab("Lag")

plot_grid(sarima.acf, sarima.pacf, ncol = 2)
```



> Answer: The seasonal AR component (SAR) is easy to identify, with ACF plot showing spikes at 12, 24, 36, 48… and decreasing each time gradually over time. There's also a spike at lag 12 from the PACF plot and cuts off afterwards (at order of 1). These both show the seasonal AR component of the model. The non-seasonal MA component is harder to identify from both plots. The ACF plot does have a spike at lag 1 and then cuts off afterwards, but it's not very easy to spot since the spike and cutoff doesn't show the biggest differences. In addition, even though the PACF plot does show a consistent gradual decay pattern, it's not very clean and has some variations in between. But overall broadly speaking, the plots show the non-seasonal MA(1) patterns.