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

Assignment 6 - Due date 02/27/25

Alex Lopez

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.

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 is an autoregressive model with an order of 2. For AR models, like this one, the ACF will decay exponentially with time. The PACF will have a cut off after lag 2, with 2 also identifying the order of this model.

- MA(1)

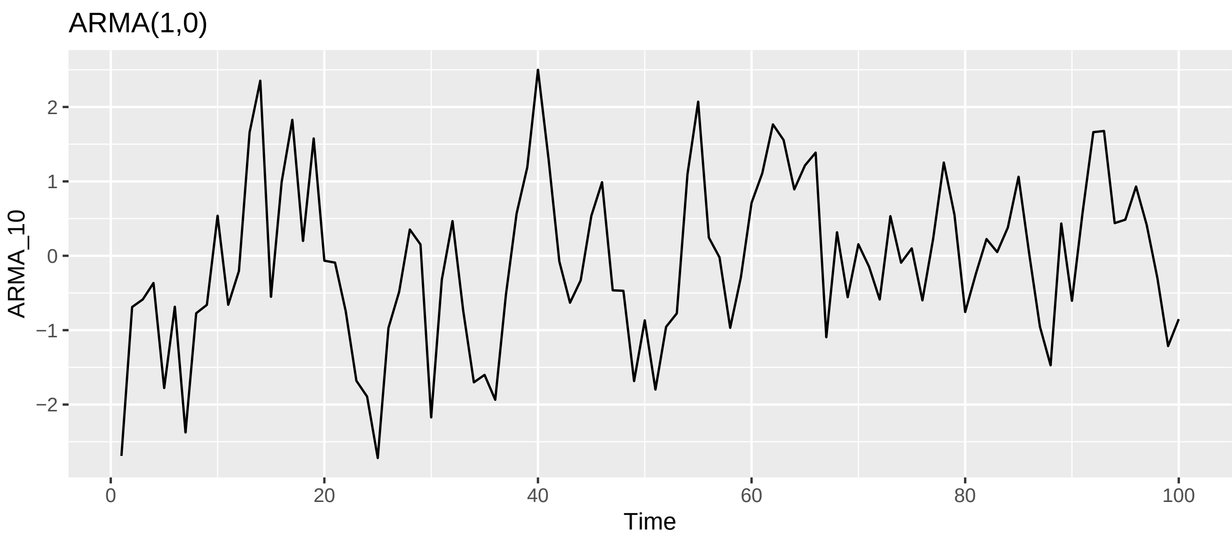
  Answer: This is a moving average model with an order of 1. For MA models, the ACF plot identifies the order of the model. Since this is a first order MA model, there is a cut off after lag 1 in the ACF plot. The PACF plot, on the other hand, will decay exponentially.

Q2

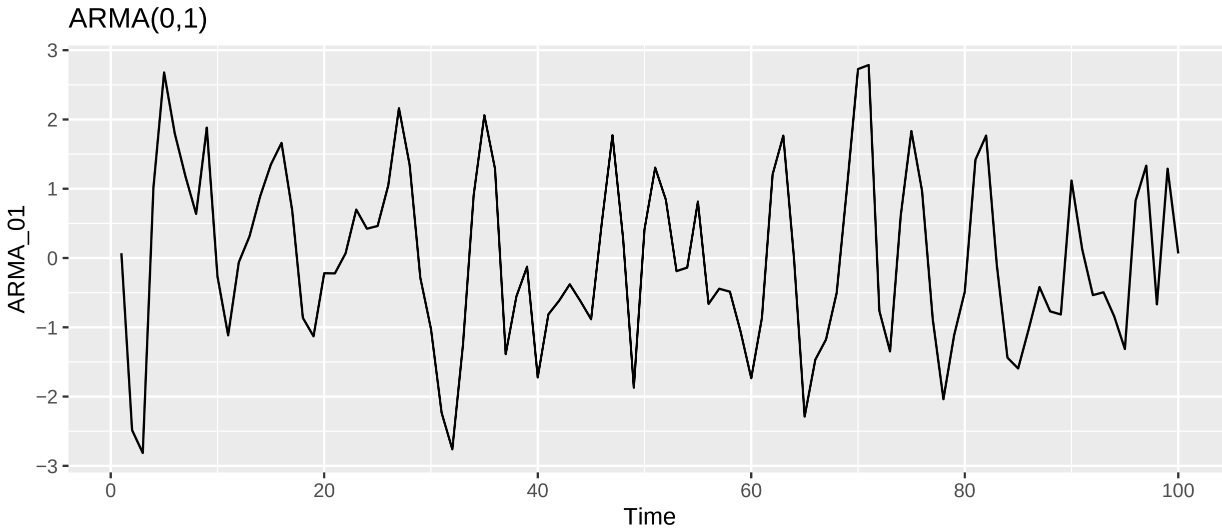
Recall that the non-seasonal ARIMA is described by three parameters ARIMA(p, d, q) where p is the order of the autoregressive component, d is the number of times the series need to be differenced to obtain stationarity and q is the order of the moving average component. If we don't need to difference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

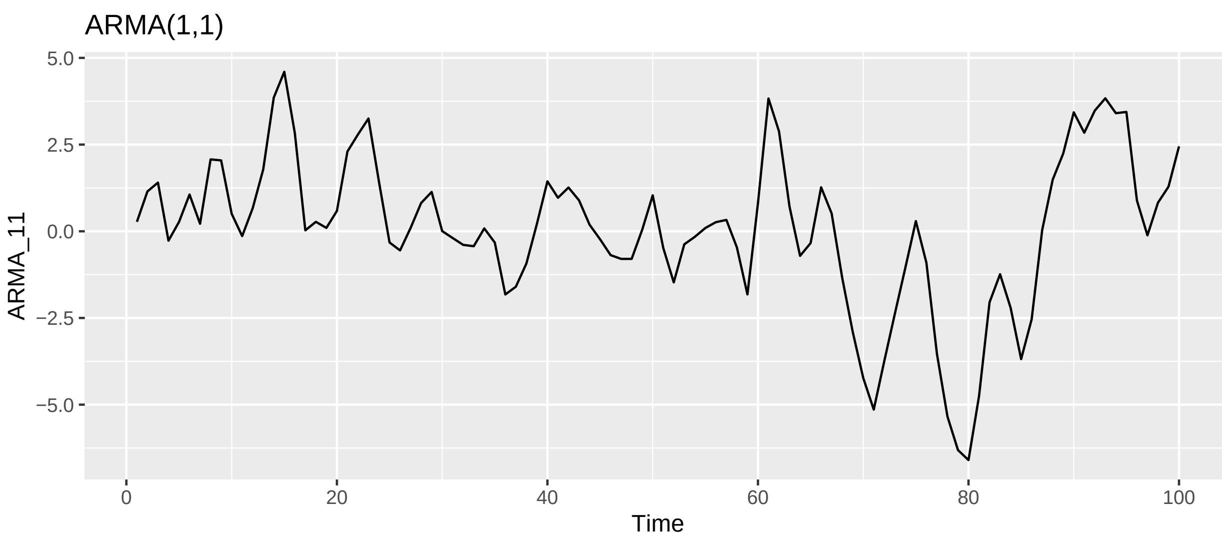
```
#ARMA(1,0)
ARMA_10 <- arima.sim(n=100, model = list(ar=0.6))

#ARMA(0,1)
ARMA_01 <- arima.sim(n=100, model = list(ma=0.9))

#ARMA(1,1)
ARMA_11 <- arima.sim(n=100, model = list(ar=0.6, ma=0.9))

autoplot(ARMA_10) + ggtitle("ARMA(1,0)")
```




```
autoplot(ARMA_01) + ggtitle("ARMA(0,1)")
```



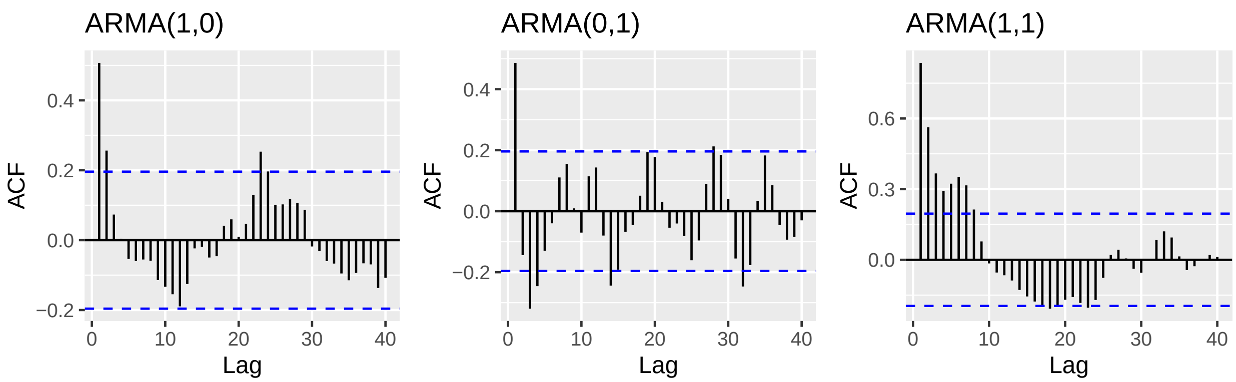

```
autoplot(ARMA_11) + ggtitle("ARMA(1,1)")
```



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

```
#ACFs
plot_grid(
  autoplot(Acf(ARMA_10, lag = 40, plot=FALSE), main = "ARMA(1,0)"),
  autoplot(Acf(ARMA_01, lag = 40, plot=FALSE), main = "ARMA(0,1)"),
  autoplot(Acf(ARMA_11, lag = 40, plot=FALSE), main = "ARMA(1,1)"),
  nrow=1
)
```

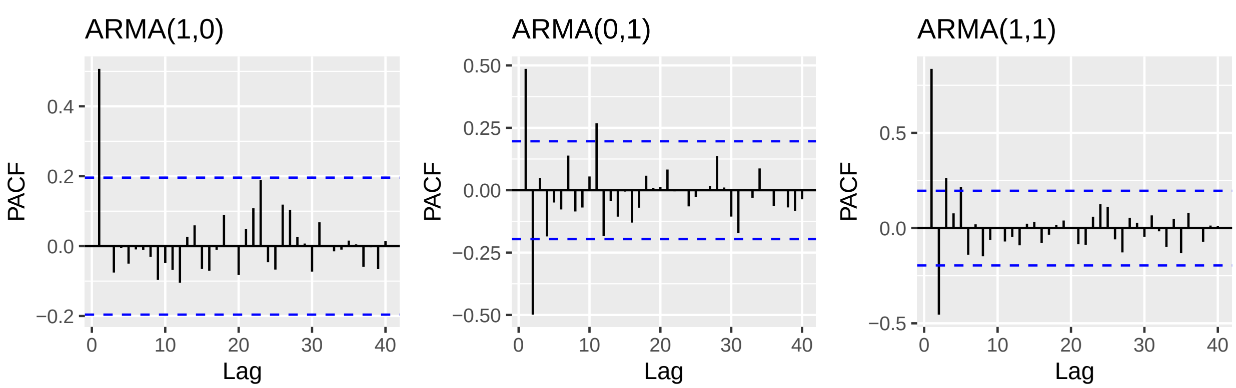
```
## Warning in ggplot2::geom_segment(lineend = "butt", ...): Ignoring unknown parameters: ‘main‘
## Ignoring unknown parameters: ‘main‘
## Ignoring unknown parameters: ‘main‘
```



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

```
#PACFs
plot_grid(
  autoplot(Pacf(ARMA_10, lag = 40, plot=FALSE), main = "ARMA(1,0)"),
  autoplot(Pacf(ARMA_01, lag = 40, plot=FALSE), main = "ARMA(0,1)"),
  autoplot(Pacf(ARMA_11, lag = 40, plot=FALSE), main = "ARMA(1,1)"),
  nrow=1
)
```

```
## Warning in ggplot2::geom_segment(lineend = "butt", ...): Ignoring unknown parameters: ‘main‘
## Ignoring unknown parameters: ‘main‘
## Ignoring unknown parameters: ‘main‘
```



(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: Yes, by identifying whether there is a gradual decay or cut offs in their respective ACF and PACF plots. Specifically, if the ACF plot cuts off after lag 1, it is likely an MA model. If the PACF plot cuts off after lag 1, it is likely an AR model. If both ACF and PACF plots show gradual decay, it is likey an ARMA model.

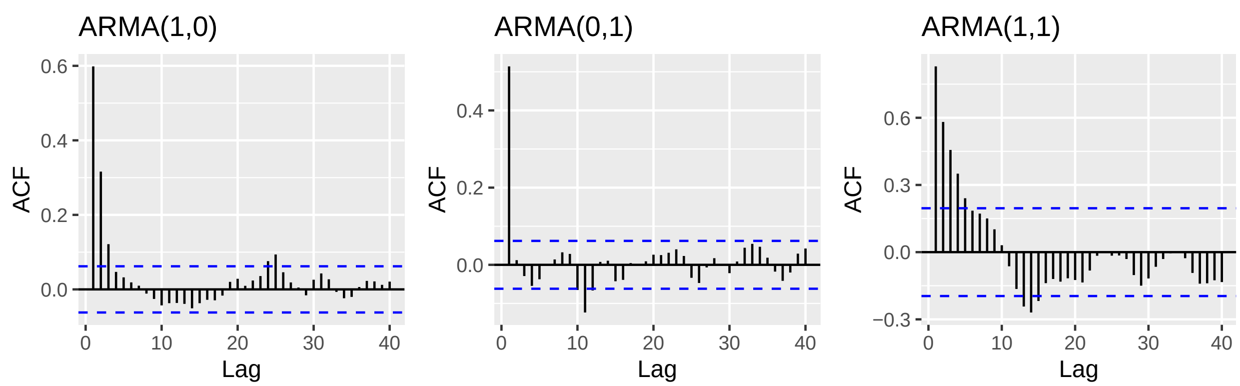
(e) Compare the PACF values R computed with the values you provided for the lag 1 correlation coefficient, i.e., does $\phi = 0.6$ match what you see on PACF for ARMA(1,0), and ARMA(1,1)? Should they match?

Answer: The PACF at lag 1 should be approximately 0.6, the given AR coefficient, since it is the PACF plot that identifies the order of the AR model. The PACF at lag 1 for the ARMA model might not match completely due to the addition of the MA term.

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

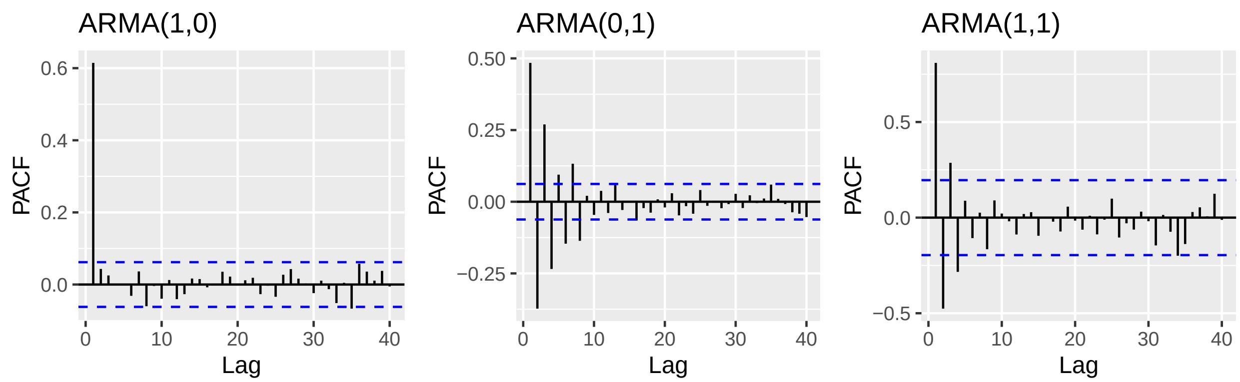
```
#ACFs with 1000 observations
plot_grid(
  autoplot(Acf(arima.sim(n=1000, model = list(ar=0.6)), lag = 40, plot=FALSE),
              main = "ARMA(1,0)"),
  autoplot(Acf(arima.sim(n=1000, model = list(ma=0.9)), lag = 40, plot=FALSE),
                main = "ARMA(0,1)"),
  autoplot(Acf(arima.sim(n=100, model = list(ar=0.6, ma=0.9)), lag = 40, plot=FALSE),
                main = "ARMA(1,1)"),
  nrow=1
)
```

```
## Warning in ggplot2::geom_segment(lineend = "butt", ...): Ignoring unknown parameters: ‘main‘
## Ignoring unknown parameters: ‘main‘
## Ignoring unknown parameters: ‘main‘
```




```
#PACFs with 1000 observations
plot_grid(
  autoplot(Pacf(arima.sim(n=1000, model = list(ar=0.6)), lag = 40, plot=FALSE),
              main = "ARMA(1,0)"),
  autoplot(Pacf(arima.sim(n=1000, model = list(ma=0.9)), lag = 40, plot=FALSE),
                main = "ARMA(0,1)"),
  autoplot(Pacf(arima.sim(n=100, model = list(ar=0.6, ma=0.9)), lag = 40, plot=FALSE),
                main = "ARMA(1,1)"),
  nrow=1
)
```

```
## Warning in ggplot2::geom_segment(lineend = "butt", ...): Ignoring unknown parameters: ‘main‘
## Ignoring unknown parameters: ‘main‘
## Ignoring unknown parameters: ‘main‘
```



>Answer: With an increased number of observations, it becomes much easier to identify whether there is a gradual decay or cut offs in the models' respective ACF and PACF plots. If the ACF plot cuts off after lag 1, it is likely an MA model. If the PACF plot cuts off after lag 1, it is likely an AR model. If both ACF and PACF plots show gradual decay, it is likey an ARMA model. The PACF at lag 1 should be approximately 0.6, the given AR coefficient, since it is the PACF plot that identifies the order of the AR model. The PACF at lag 1 for the ARMA model might not match completely due to the addition of the MA term; however, with the increased number of observations, it appears that the PACF at lag 1 for the ARMA model is closer to the value of the given AR coefficient, 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: ARIMA(1,0,1)(1,0,0)_12

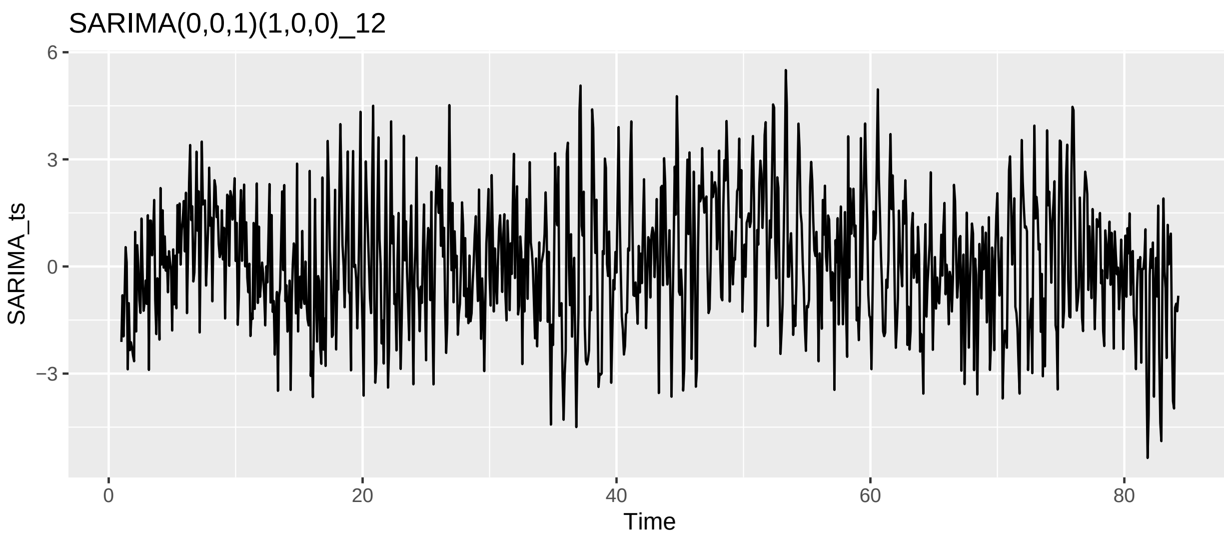
(b) Also from the equation what are the values of the parameters, i.e., model coefficients.

  Answer: AR term = 0.7 ; Seasonal AR term = -0.25; MA term = -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_model <- sim_sarima(n=1000, model = list(ar = c(rep(0, 11), 0.8), ma = 0.5))

#convert to ts object
SARIMA_ts <- ts(SARIMA_model, frequency = 12)

autoplot(SARIMA_ts) + ggtitle("SARIMA(0,0,1)(1,0,0)_12")
```



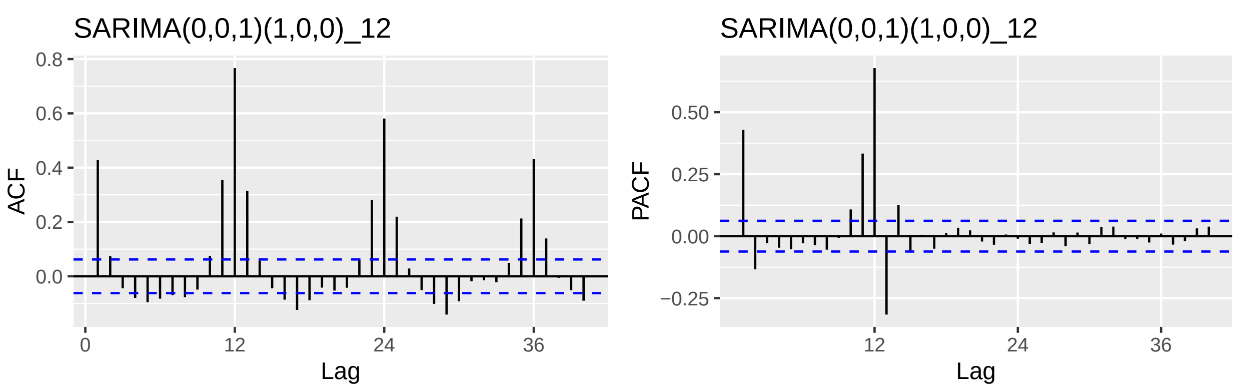
>Answer: The plot appears seasonal.

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_grid(
  autoplot(Acf(SARIMA_ts, lag = 40, plot=FALSE), main = "SARIMA(0,0,1)(1,0,0)_12"),
  autoplot(Pacf(SARIMA_ts, lag = 40, plot=FALSE), main = "SARIMA(0,0,1)(1,0,0)_12"),
  nrow=1
)
```

```
## Warning in ggplot2::geom_segment(lineend = "butt", ...): Ignoring unknown parameters: ‘main‘
## Ignoring unknown parameters: ‘main‘
```



>Answer: Yes, the plots should be able to help us determine the order of both non-seasonal and seasonal components. Due to q = 1, the order of the non-seasonal component, there should be a sharp cut off after lag 1 in the ACF plot, which can be observed in the plot above. Due to P = 1, the order of the seasonal component, there should be a spike at lag 12 in the PACF plot, after which there is a cut off, as seen in the plot above.