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

Assignment 6 - Due date 02/28/24

Aditi Jackson

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.

```r
#Load/install required package here
#install.packages("sarima")

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: The ACF plot of an AR model decays exponentially with time, and the PACF plot will identify the order of the AR model. In this case, the order of the series is 2. The PACF for AR(2) will indicate that autocorrelation is only significant for two lags.

- MA(1)

  Answer: The ACF plot of an MA model will identify the order, which in this case is 1. Graphically we would see that only lag 1 is significant. The PACF will decay exponentially.

Q2

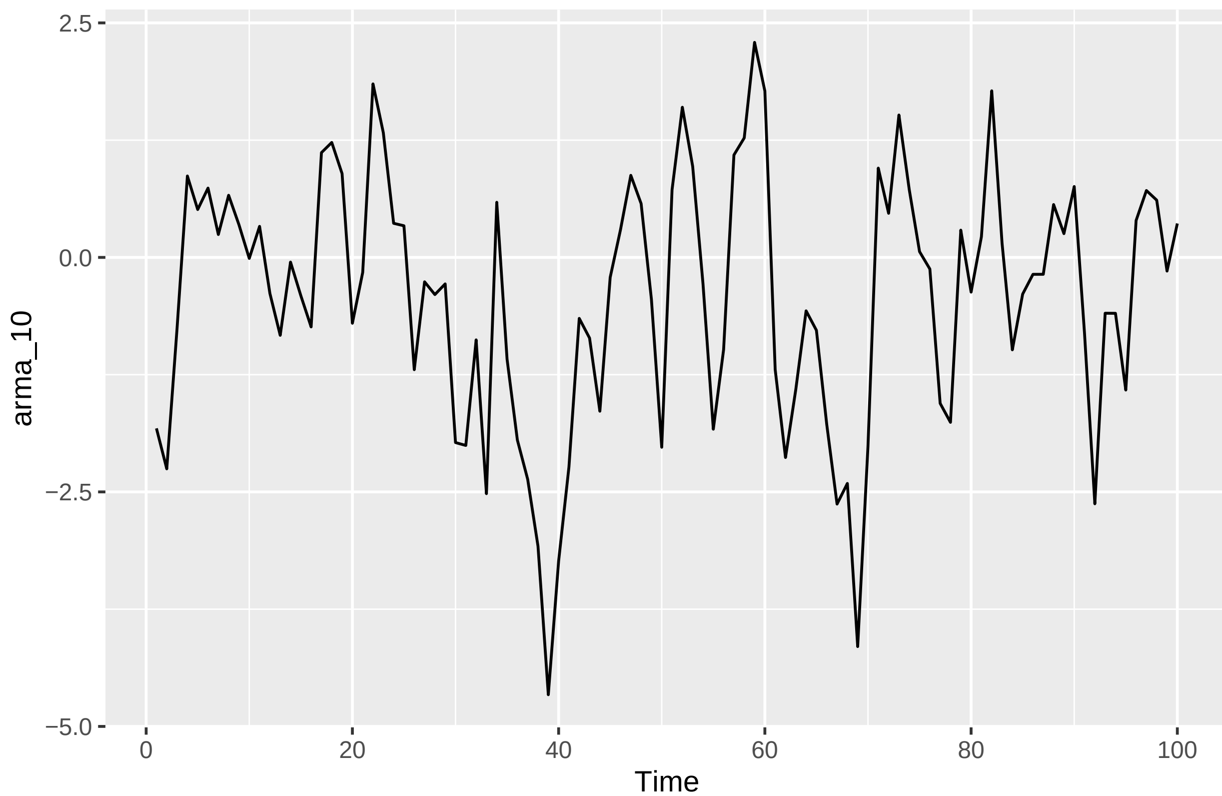
Recall that the non-seasonal ARIMA is described by three parameters ARIMA(p, d, q) where p is the order of the autoregressive component, d is the number of times the series need to be differenced to obtain stationarity and q is the order of the moving average component. If we don't need to difference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

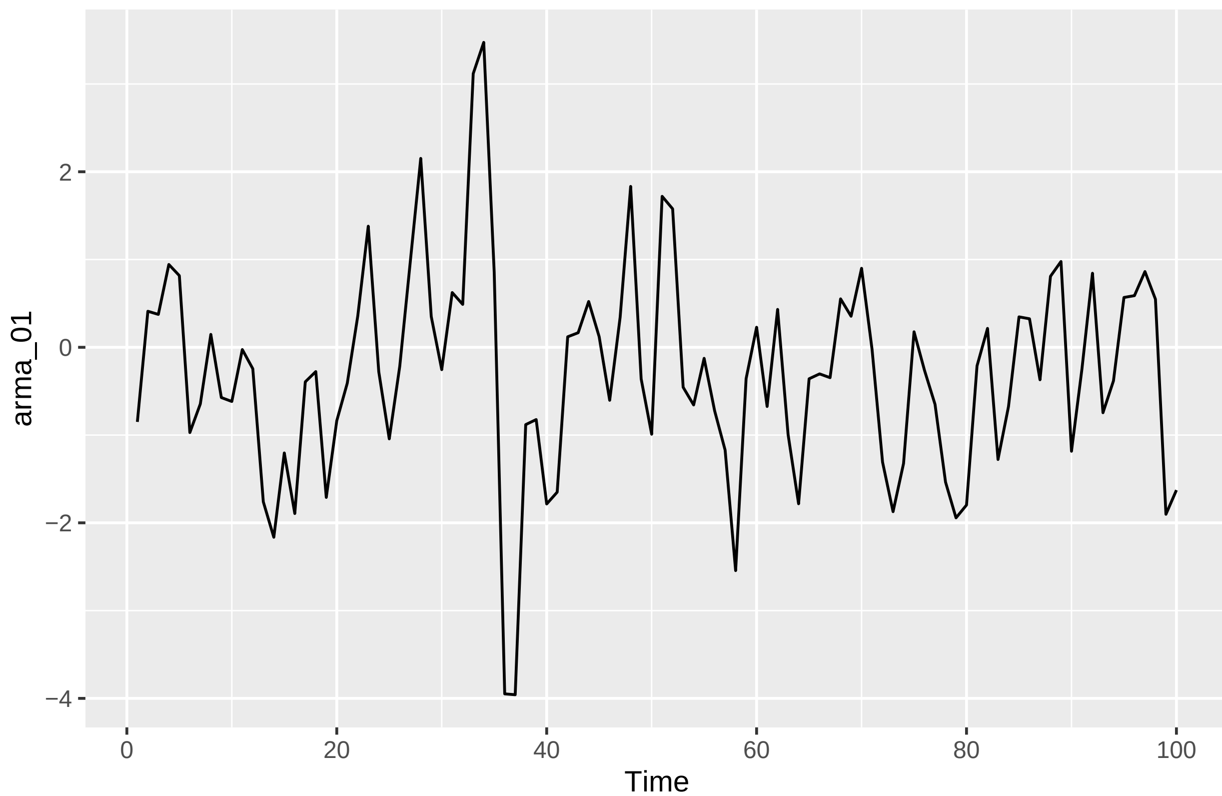
```
# Parameters
phi <- 0.6
theta <- 0.9

# Generating 100 observations for each ARMA model
## AMRA(1,0)
arma_10 <- arima.sim(model = list(ar = phi, ma = 0), n = 100)

## ARMA(0,1)
arma_01 <- arima.sim(model = list(ar = 0, ma = theta), n = 100)
```

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

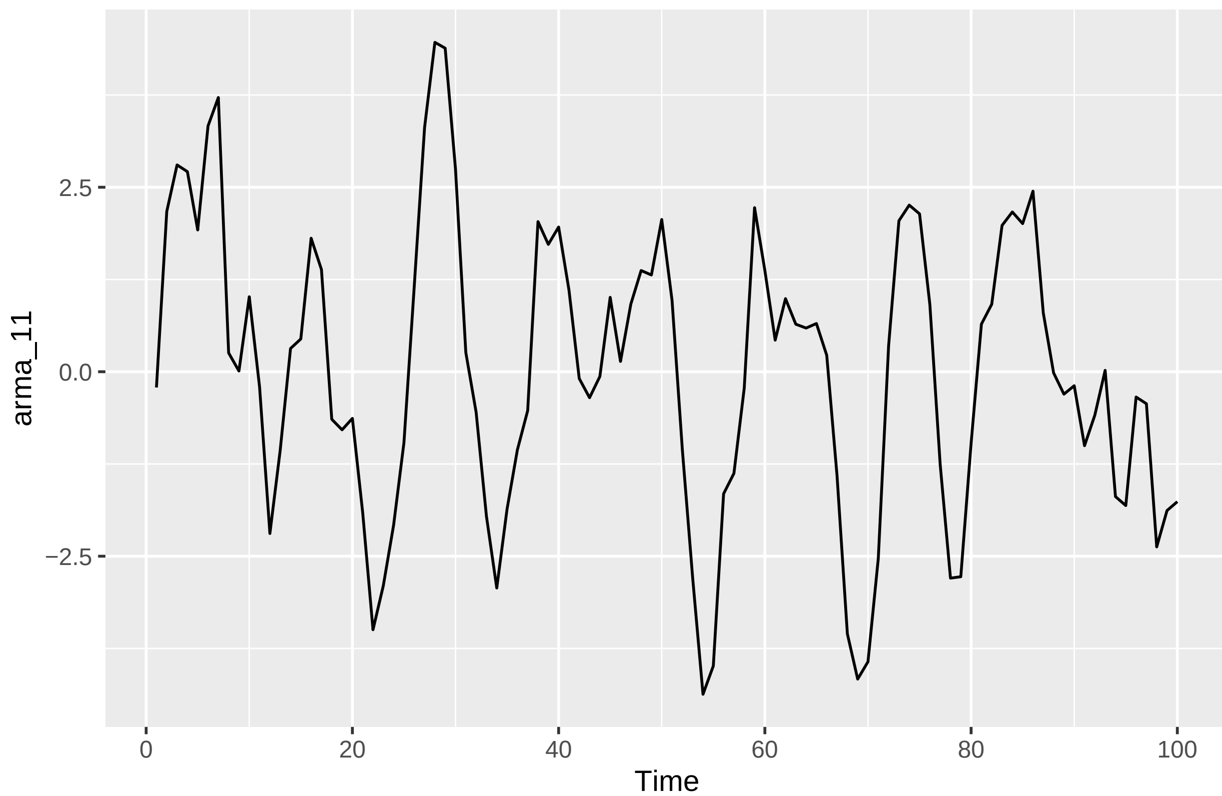
```
## ARMA(1,1)
arma_11 <- arima.sim(model = list(ar = phi, ma = theta), n = 100)


# Plotting series
## AMRA(1,0)
arma_10_plot <- autoplot(arma_10)
arma_10_plot
```




```
## AMRA(0,1)
arma_01_plot <- autoplot(arma_01)
arma_01_plot
```

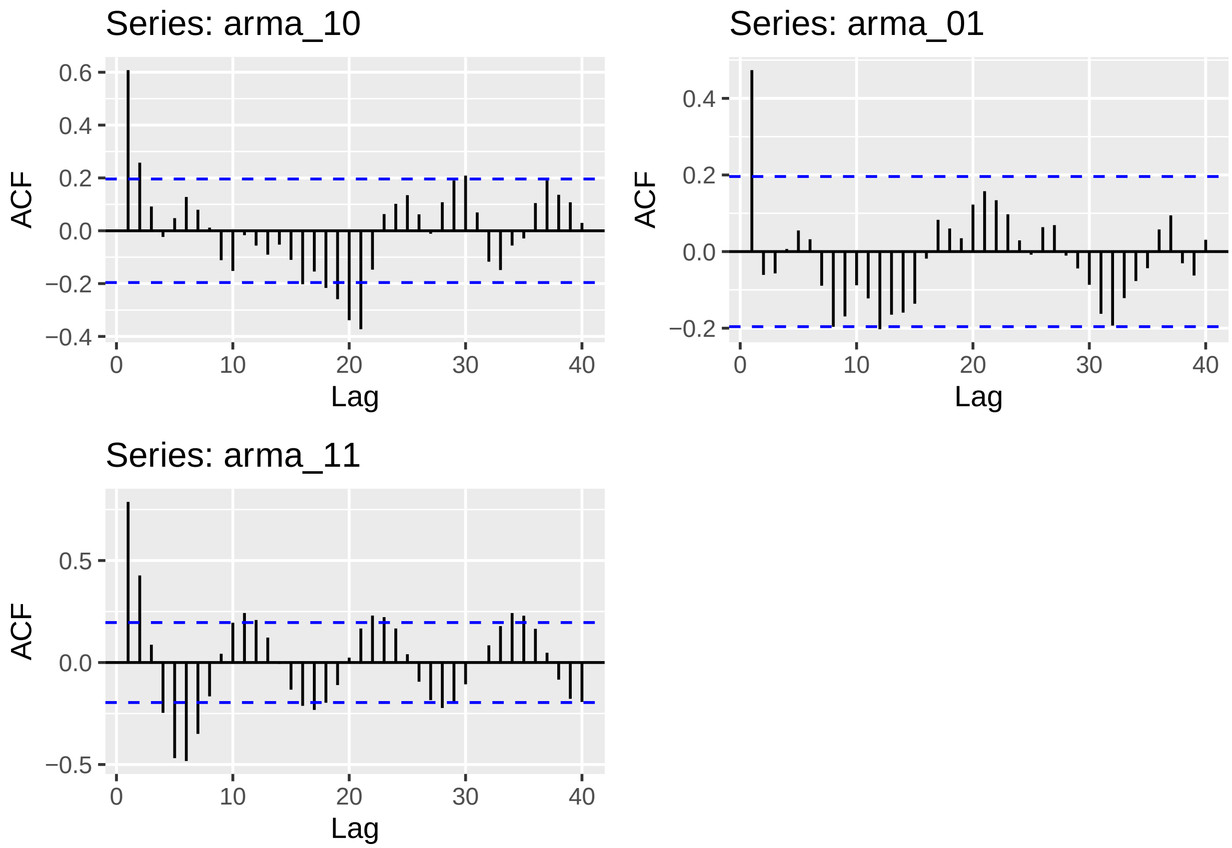



```r
## AMRA(1,1)
arma_11_plot <- autoplot(arma_11)
arma_11_plot
```



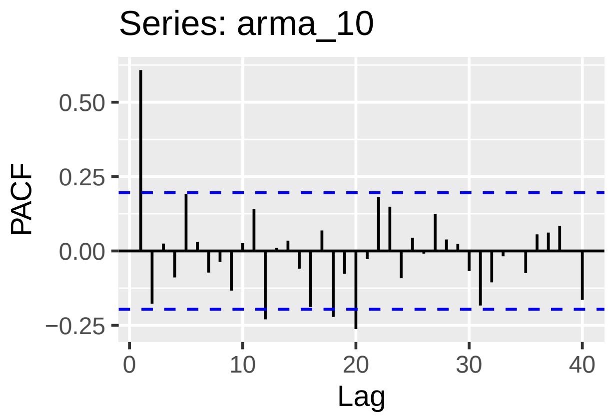
(b) Plot the sample ACF for each of these models in one window to facilitate comparison (Hint: use `cowplot::plot_grid()`).

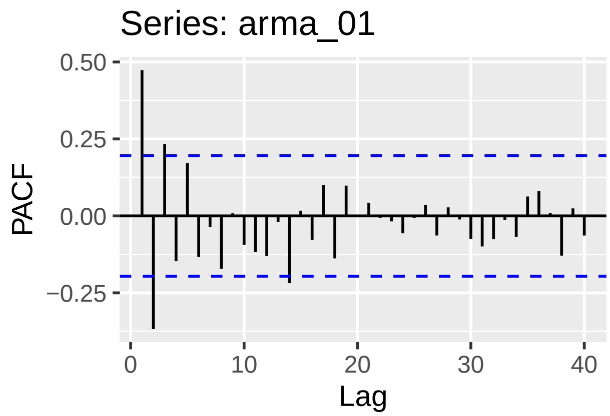
```
# creating ACFs
Acf_arma_10 <- Acf(arma_10,lag.max=40,plot=FALSE)
Acf_arma_01 <- Acf(arma_01,lag.max=40,plot=FALSE)
Acf_arma_11 <- Acf(arma_11,lag.max=40,plot=FALSE)

# plotting ACFs
plot_grid(
  autoplot(Acf_arma_10),
  autoplot(Acf_arma_01),
  autoplot(Acf_arma_11))
```

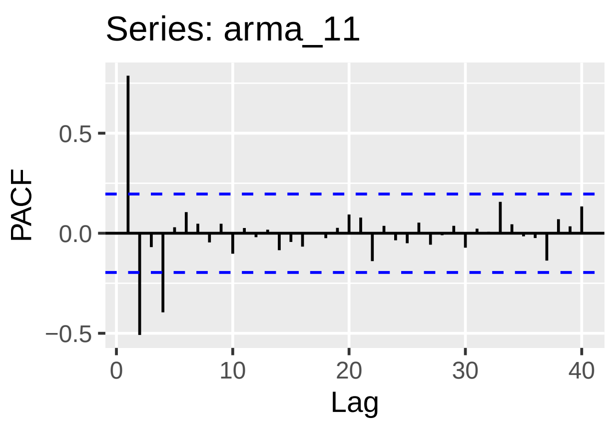


(c) Plot the sample PACF for each of these models in one window to facilitate comparison.

```
# creating ACFs
Pacf_arma_10 <- Pacf(arma_10,lag.max=40,plot=FALSE)
Pacf_arma_01 <- Pacf(arma_01,lag.max=40,plot=FALSE)
Pacf_arma_11 <- Pacf(arma_11,lag.max=40,plot=FALSE)

# plotting ACFs
plot_grid(
  autoplot(Pacf_arma_10),
  autoplot(Pacf_arma_01),
  autoplot(Pacf_arma_11))
```







(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): I think I would be able to identify this series correctly as AR by looking at the ACF, which appears to decay exponentially, and the PACF, which shows one sigificant lag (indicating an order of 1). ARMA(0,1): I'm not as certain I'd be able to identify this as MA. By looking at the ACF I can tell that the order is 1 since only lag 1 is significant. However, it's hard to see an exponential decay in the PACF, which is characteristic of MA models. ARMA(1,1): I do not think I would not be able to identify this as ARMA. The ACF decays exponentially and looks similar to the ACF of the AR while the PACF has 2 significant lags and appears more similar to the PACF of the MA.

(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?

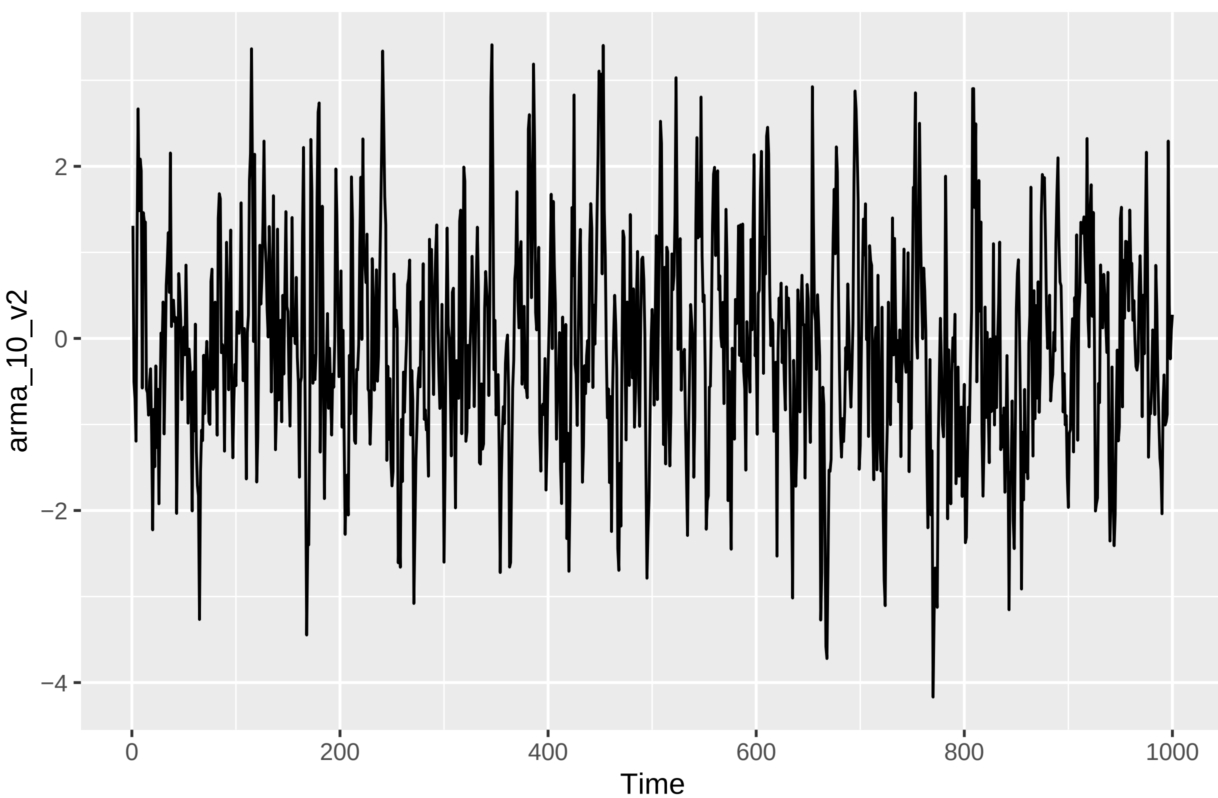
Answer: ARMA(1,0): The calculated correlation coefficient at lag 1 has a magnitude of 0.6, which matches the given value of Phi. This seems to an ideal situation where the population parameter (Phi) is equal to the sample estimate given by the magnitude of lag 1. ARMA(0,1): The calculated correlation coefficient at lag 1 has a magnitude of ~0.4. It does not match the given value of Theta (0.9). However, this may be becasue we have only run the model for n=100 observations. Perhaps running it with more observations will give us a sample estimate closer to theta. ARMA(1,1): Lag 1 of the PACF gives us a value of ~0.77, which is almost halfway between phi=0.6 and theta=0.9. This seems reasonable as it reflects some sort of midpoint between phi and theta.

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

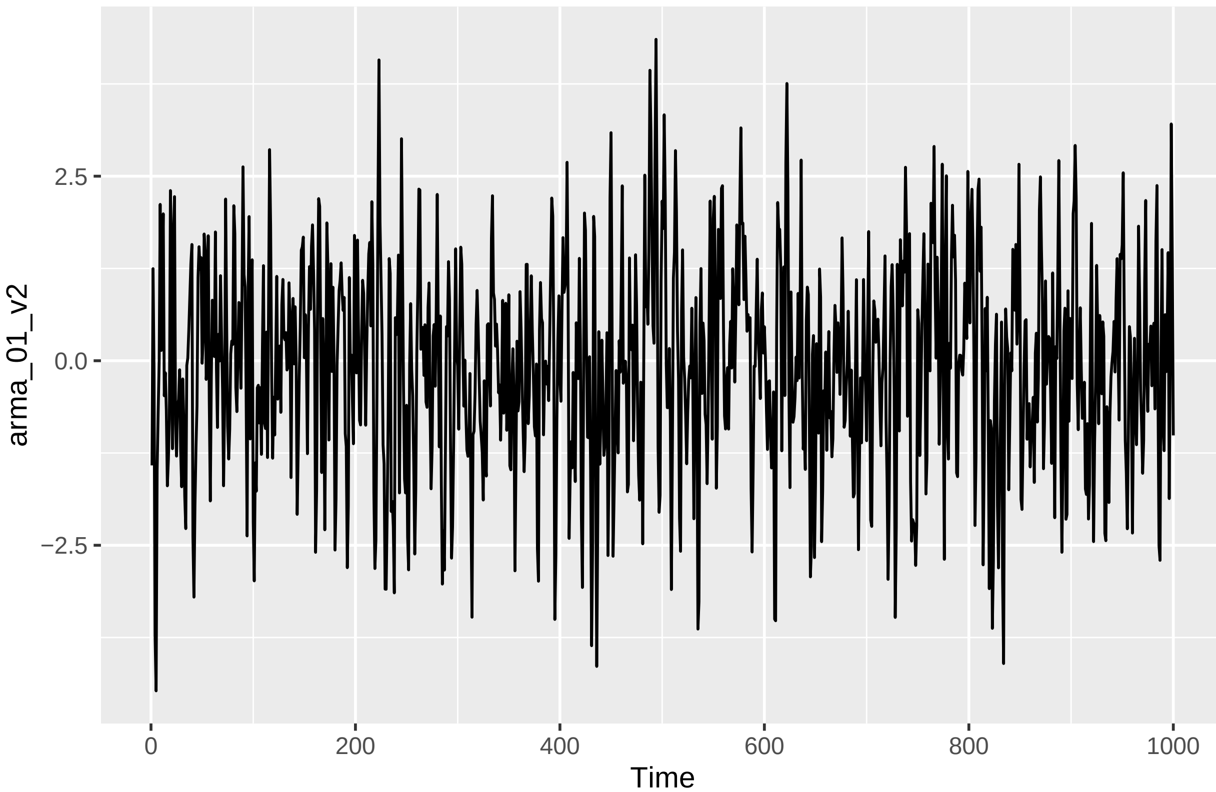
```r
# a
# Generating 100 observations for each ARMA model
## AMRA(1,0)
arma_10_v2 <- arima.sim(model = list(ar = phi, ma = 0), n = 1000)
## ARMA(0,1)
arma_01_v2 <- arima.sim(model = list(ar = 0, ma = theta), n = 1000)
```

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

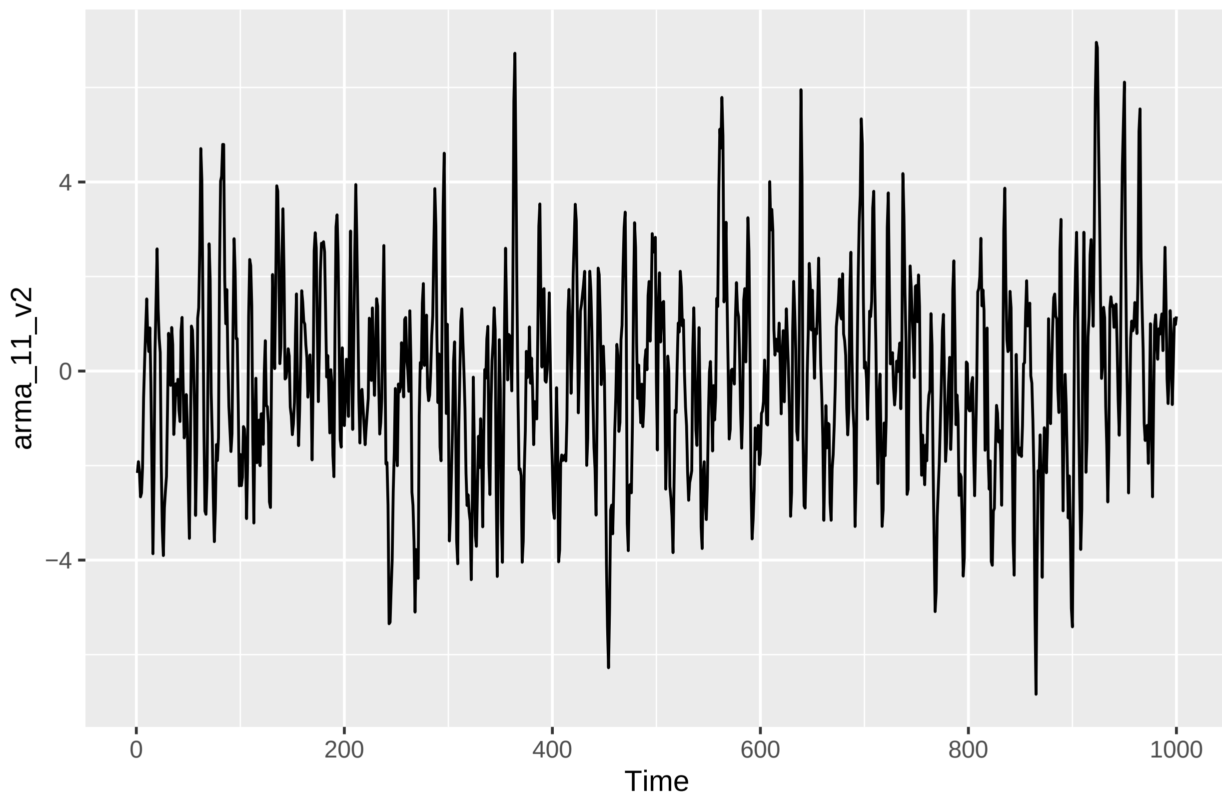
```r
## ARMA(1,1)
arma_11_v2 <- arima.sim(model = list(ar = phi, ma = theta), n = 1000)

# Plotting series
## AMRA(1,0)
arma_10_plot_v2 <- autoplot(arma_10_v2)
arma_10_plot_v2
```



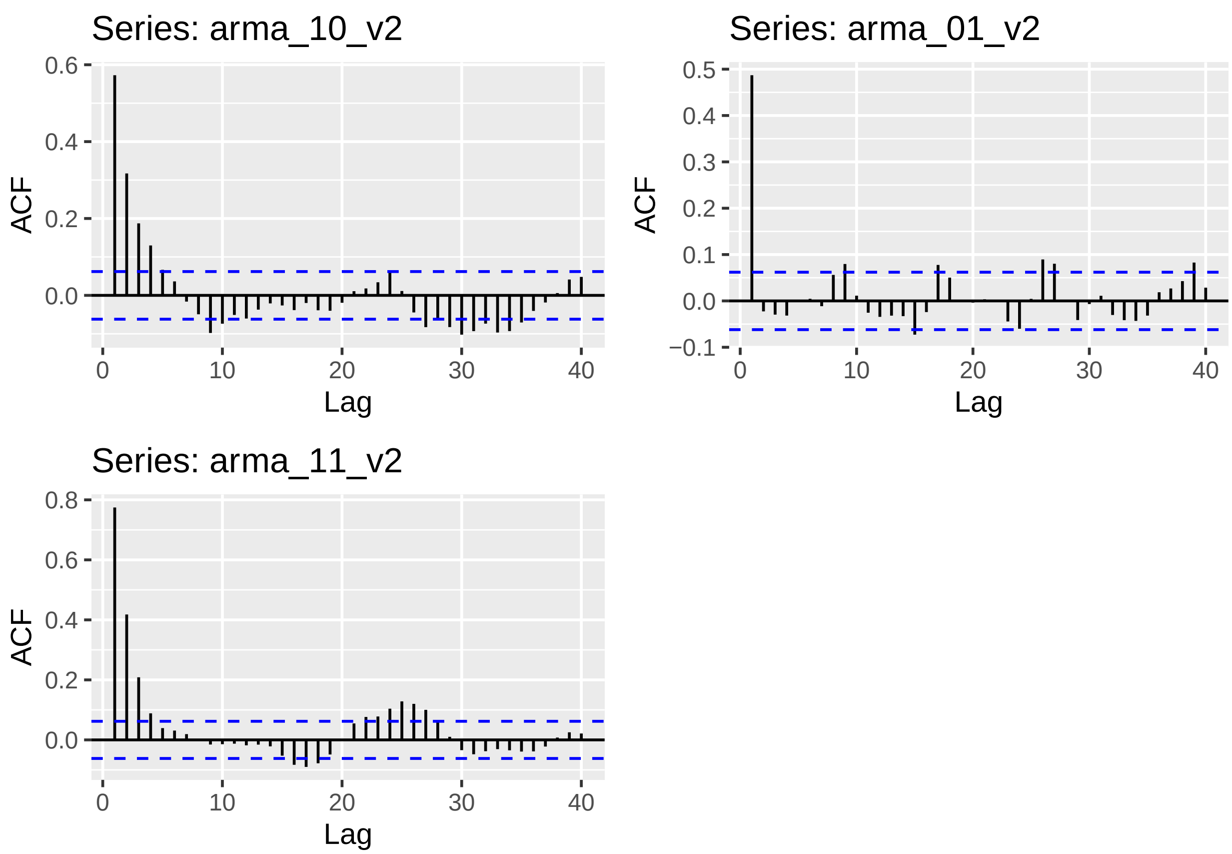

```r
## AMRA(0,1)
arma_01_plot_v2 <- autoplot(arma_01_v2)
arma_01_plot_v2
```



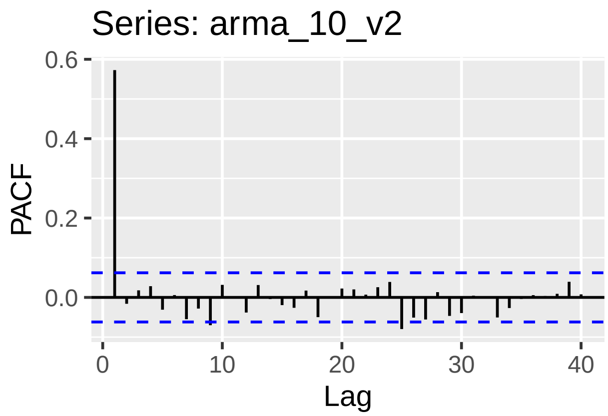

```r
## AMRA(1,1)
arma_11_plot_v2 <- autoplot(arma_11_v2)
arma_11_plot_v2
```

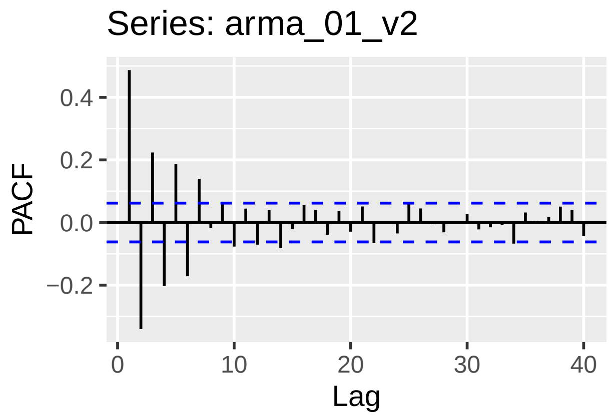


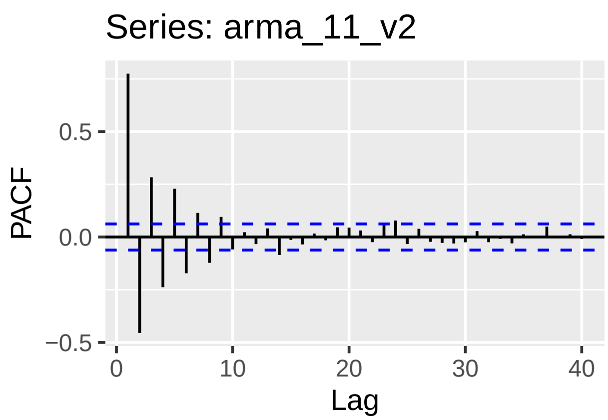

```r
# b
# creating ACFs
Acf_arma_10_v2 <- Acf(arma_10_v2,lag.max=40,plot=FALSE)
Acf_arma_01_v2 <- Acf(arma_01_v2,lag.max=40,plot=FALSE)
Acf_arma_11_v2 <- Acf(arma_11_v2,lag.max=40,plot=FALSE)

# plotting ACFs
plot_grid(
  autoplot(Acf_arma_10_v2),
  autoplot(Acf_arma_01_v2),
  autoplot(Acf_arma_11_v2))
```




```r
# c
# creating PACFs
Pacf_arma_10_v2 <- Pacf(arma_10_v2,lag.max=40,plot=FALSE)
Pacf_arma_01_v2 <- Pacf(arma_01_v2,lag.max=40,plot=FALSE)
Pacf_arma_11_v2 <- Pacf(arma_11_v2,lag.max=40,plot=FALSE)

# plotting PACFs
plot_grid(
  autoplot(Pacf_arma_10_v2),
  autoplot(Pacf_arma_01_v2),
  autoplot(Pacf_arma_11_v2))
```







(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): I am confident that I would be able to identify this series correctly as AR by looking at the ACF, which decays exponentially, and the PACF, which shows one sigificant lag (indicating an order of 1). ARMA(0,1): With n=1000 observations I am more confident that I'd be able to identify this as an MA proess. The exponential decay pattern is more pronounced in the PACF with increasing observations. The ACF still shows an order of 1. ARMA(1,1): Again, this one is hard to tell. It seems to exhibit both characteristics of the AR and MA processes in its ACF and PACF. It's ACF decays exponentially per an AR process. It's PACF appears to also decay exponentially and looks similar to the MA PACF.

(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: ARMA(1,0): The calculated correlation coefficient is a little above 0.6, which is very close to the given value of phi=0.6. ARMA(0,1): The calculated correlation coefficient is ~0.5, which is still below the below the population parameter of theta=0.9. However, it's notable that the calculated value increased by ~0.1 when running the model with more obsevations. ARMA(1,1): The first lag of the PACF has a calculated correlation coefficient a little below 0.6. This is close to Phi=0.6, and seems to be getting closer as we increase the number of observations.

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. > We know that p=1 because there is only one AR term in the equation (y_{t-1}). Similarly, we know that q=1 because there is only one MA term (a_{t-1}). Further, we know that P=1 because there is only one SAR term (t_{t-12}). We then know Q=0 because P+Q must be < 1. We cannot explicitly identify d or D (the number of times the series was differenced) from the provided equation, so I am taking it to be 0. We know that s=12 because the subscript of the SAR term is minus 12.

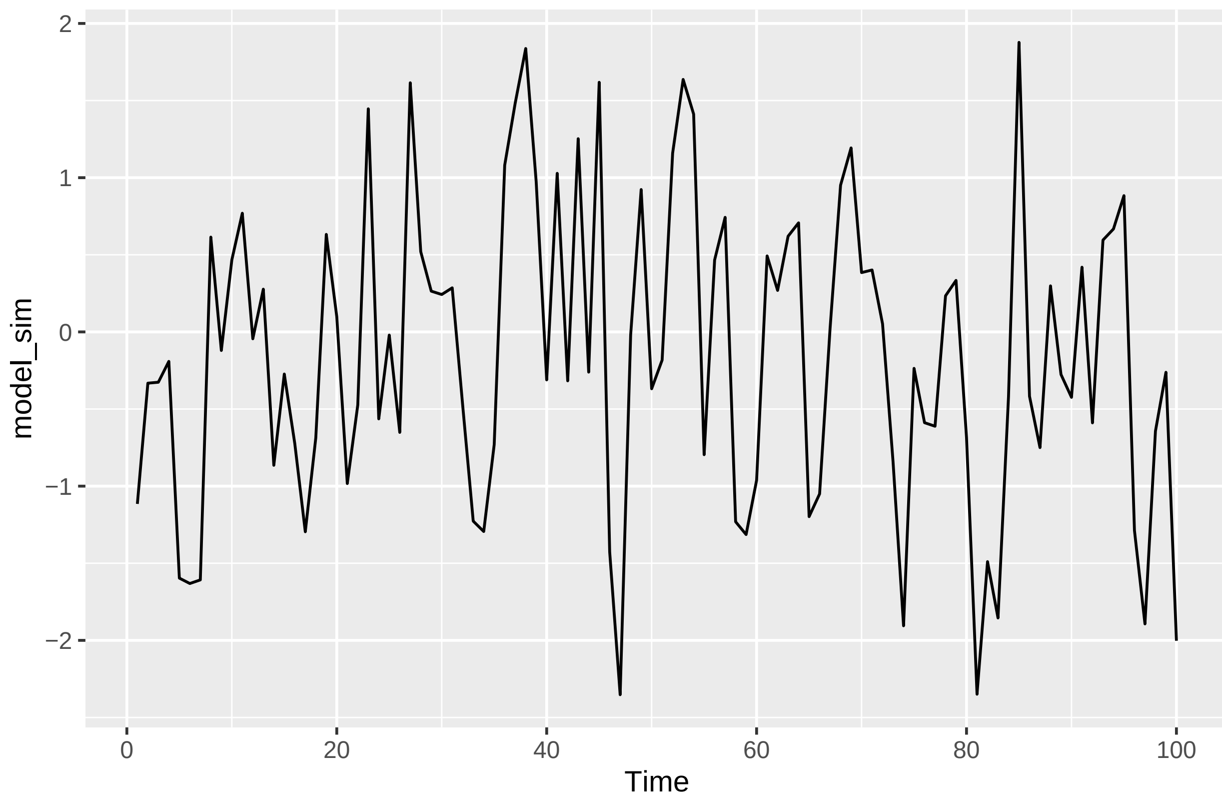
SARIMA$(1,0,1)(1,0,0)_{12}$.

(b) Also from the equation what are the values of the parameters, i.e., model coefficients. > The values of the model coefficients are: Autoregressive coefficient (phi) = 0.7 Seasonal autoregressive coefficient = -0.25 Moving average 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?

```
# Parameters
phi <- 0.8
theta <- 0.5

# defining SARIMA model
sarima_model_001100 <- list(order = c(0,0,1), seasonal = list(order = c(1,0,0)), period = 12,
                    ma = theta, sar = phi)

# simulating model with 100 observations
model_sim <- arima.sim(n = 100, model = sarima_model_001100)

# plotting series
autoplot(model_sim)
```

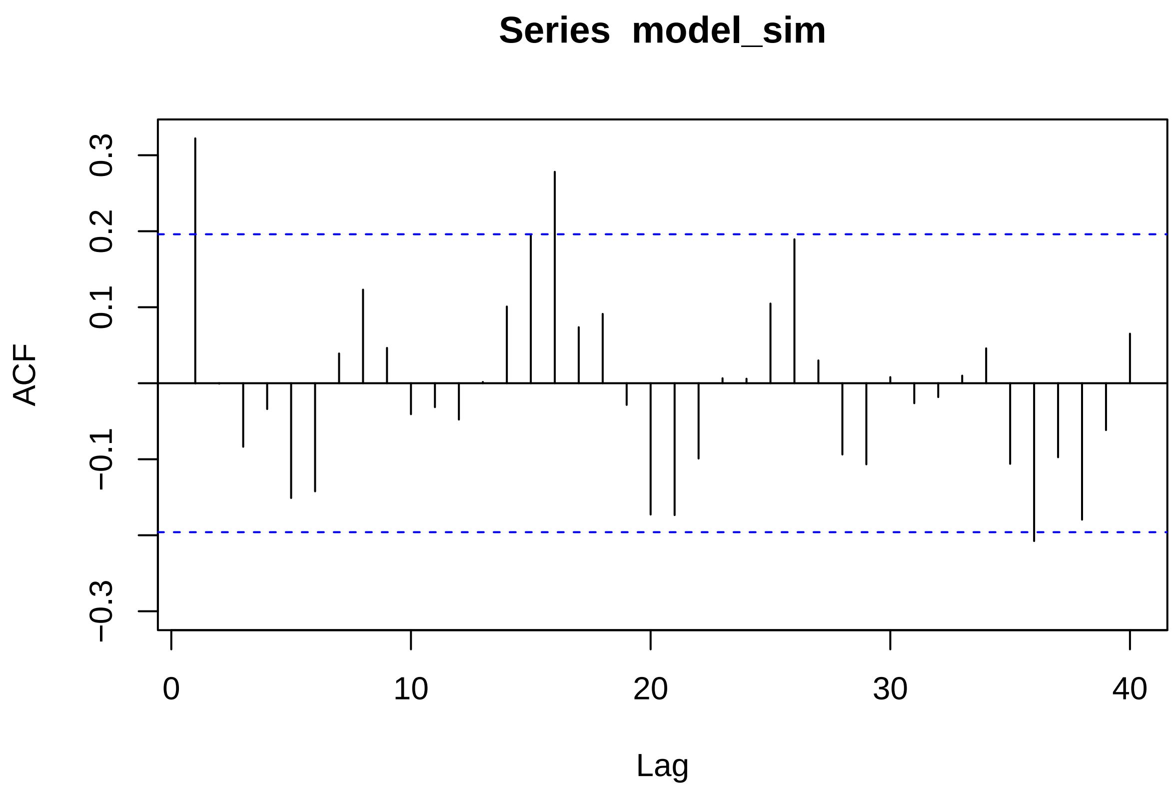


Answer: There is some sort of up and down pattern repeating approximately every 12 months, which suggests there may be some seasonality. Although I'm unsure.

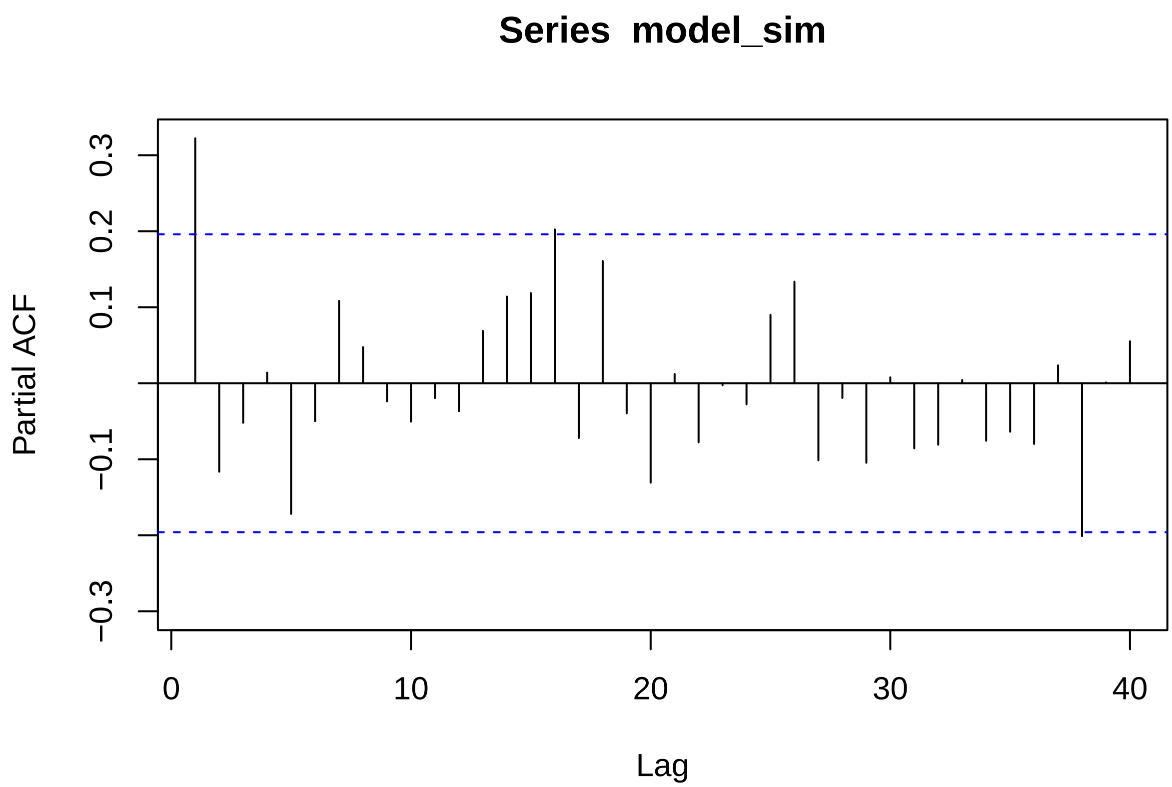
Q5

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.

```
# plotting ACF and PACF
acf_sarima_model_001100 <- Acf(model_sim, lag.max=40)
```




```
pacf_sarima_model_001100 <- Pacf(model_sim, lag.max=40)
```



> Answer: For the ACF plot, I would expect to see an exponential decay for non-seasonal lags and significant spikes at the seasonal lags for the seasonal component (multiples of 12). I do not observe either of these attributes in the ACF, which suggests to me that the model may not be representing the process well. For the PACF, I observe that the lags tail off, which suggests an MA process with an order of 1 (based on significnat spike at ACF). Further, there appear to be 2 highly significant lags in the PACF, which would suggest that the order of the seasonal model is 2. However, this is not reflective of the model from Q4 where Q=1.