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

Assignment 6 - Due date 02/28/24

Faustin Kambale

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.

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

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: __ ACF plot typically shows a gradual decay, i.e. values decreasing slowly with the increase in lags. - PACF plot exhibits sharp cutoffs after a certain lag, i.e the the autocorrelation is explained by earlier lags.

- MA(1)

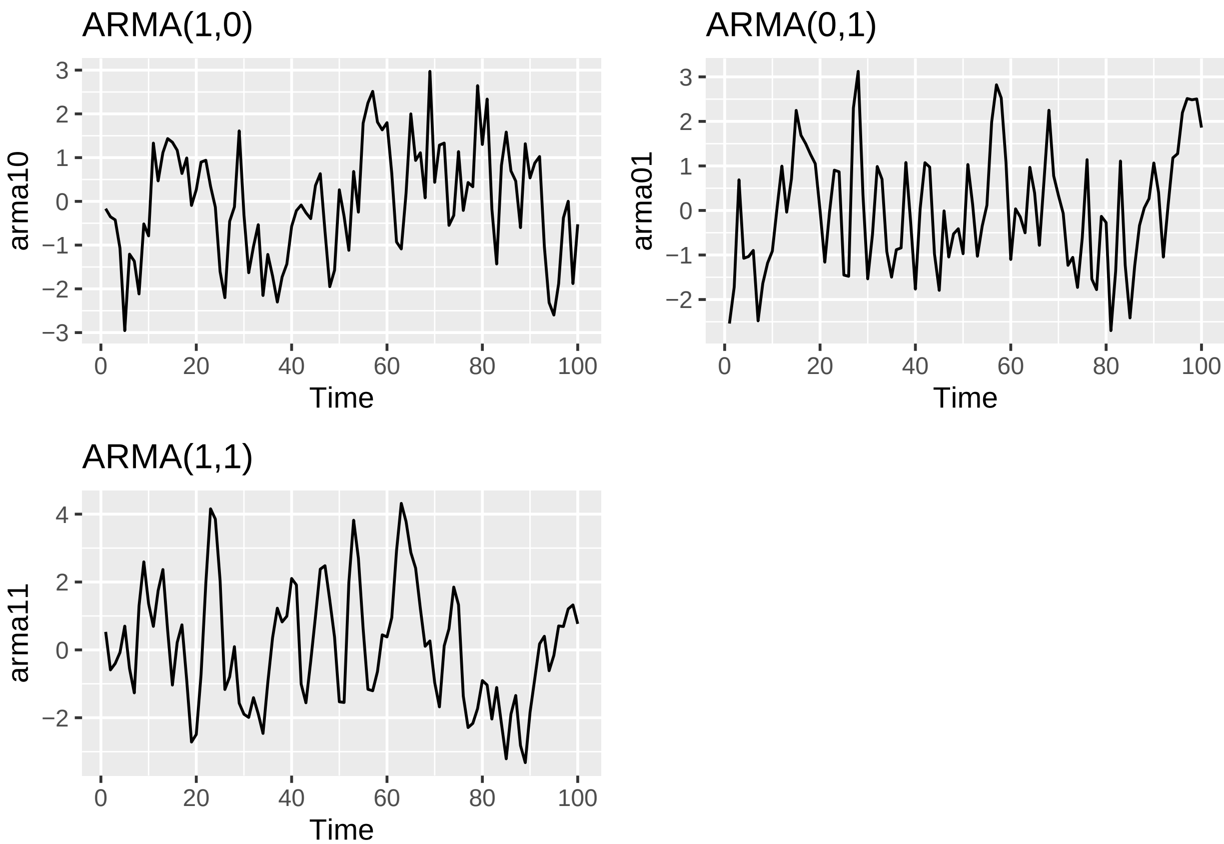
  Answer: For an MA process, the ACF has a sharp cutoff after a certain number of lags indicating the order of the MA process. The PACF decays gradually.

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.

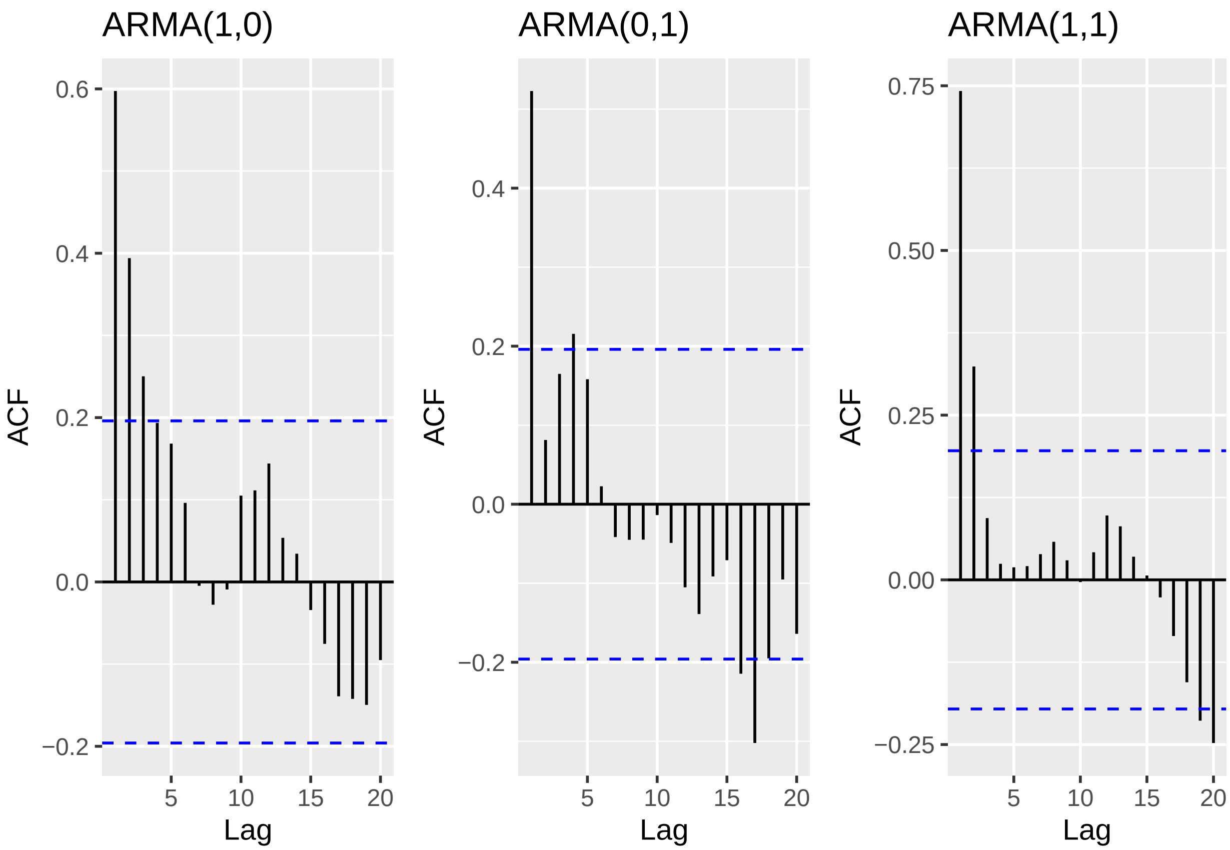
```
# Define parameters
phi <- 0.6
theta <- 0.9
n <- 100

# I create Data for ARMA series
arma10 <- arima.sim(model = list(ar = phi), n = n, innov = rnorm(n))
arma01 <- arima.sim(model = list(ma = theta), n = n, innov = rnorm(n))
arma11 <- arima.sim(model = list(ar = phi, ma = theta), n = n, innov = rnorm(n))

# Plot the series
plot_grid (
autoplot(arma10, main = "ARMA(1,0)"),
autoplot(arma01, main = "ARMA(0,1)"),
autoplot(arma11, main = "ARMA(1,1)")
)
```



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

```
# I Create ACF
acf10 <- acf(arma10, lag = 20, plot = FALSE)
acf01 <- acf(arma01, lag = 20, plot = FALSE)
acf11 <- acf(arma11, lag = 20, plot = FALSE)

#plot ACF
plot_grid(
  autoplot(acf10, main = "ARMA(1,0)"),
  autoplot(acf01, main = "ARMA(0,1)"),
  autoplot(acf11, main = "ARMA(1,1)"),
  nrow = 1
)
```

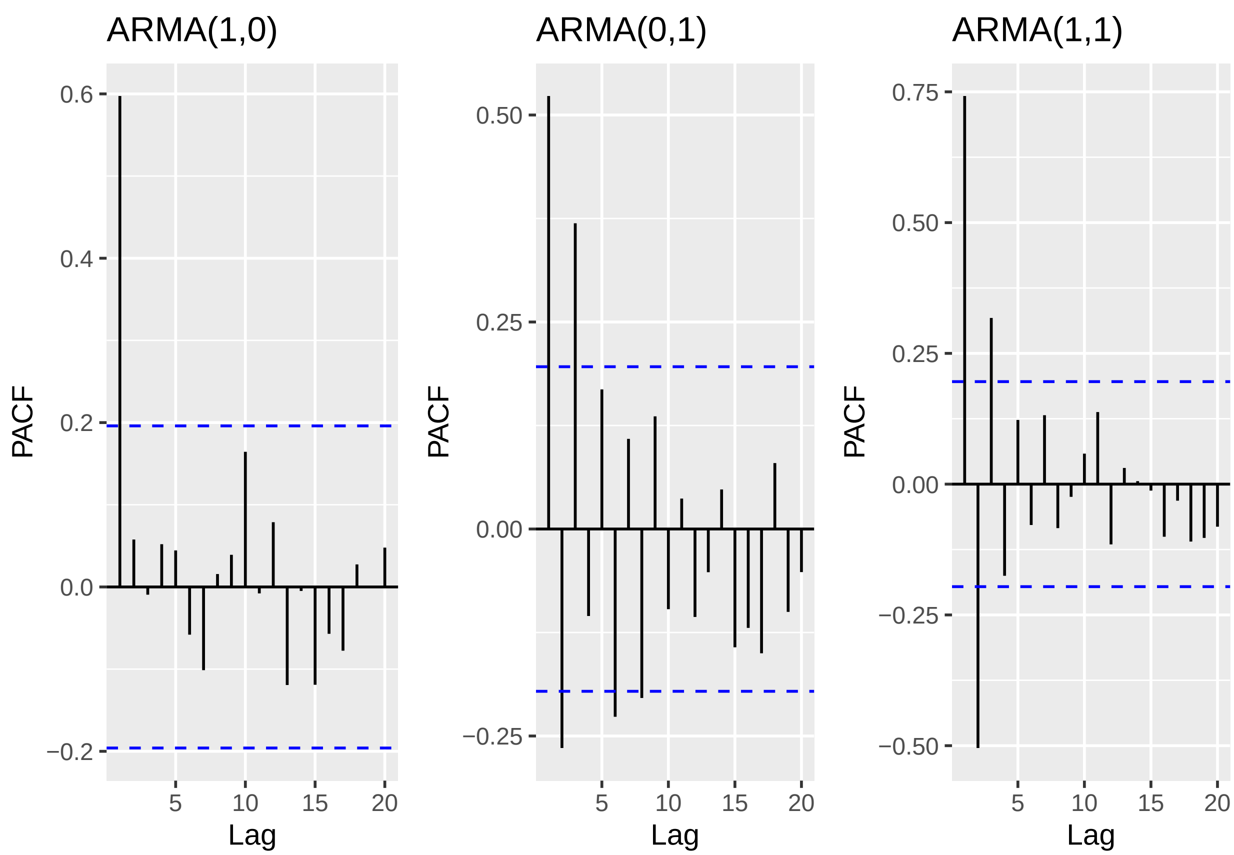
```
## Warning in ggplot2::geom_segment(lineend = "butt", ...): Ignoring unknown parameters: `main`
## Ignoring unknown parameters: `main`
## Ignoring unknown parameters: `main`
```



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

```
# I Create PACF
pacf10 <- Pacf(arma10, lag = 20, plot = FALSE)
pacf01 <- Pacf(arma01, lag = 20, plot = FALSE)
pacf11 <- Pacf(arma11, lag = 20, plot = FALSE)

#plot PACF
plot_grid(
  autoplot(pacf10, main = "ARMA(1,0)"),
  autoplot(pacf01, main = "ARMA(0,1)"),
  autoplot(pacf11, 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: ARMA(1,0): ACF shows a gradual decay and PACF shows a sharp cut-off after the first lag (AR process). ARMA(0,1): ACF and PACF plots show a sharp cut-off after the first lag: MA process. ARMA(1,1): ACF and PACF autocorrelations are significant at the first lag and then gradually decrease : indicating the mixture of AR and MA processes.

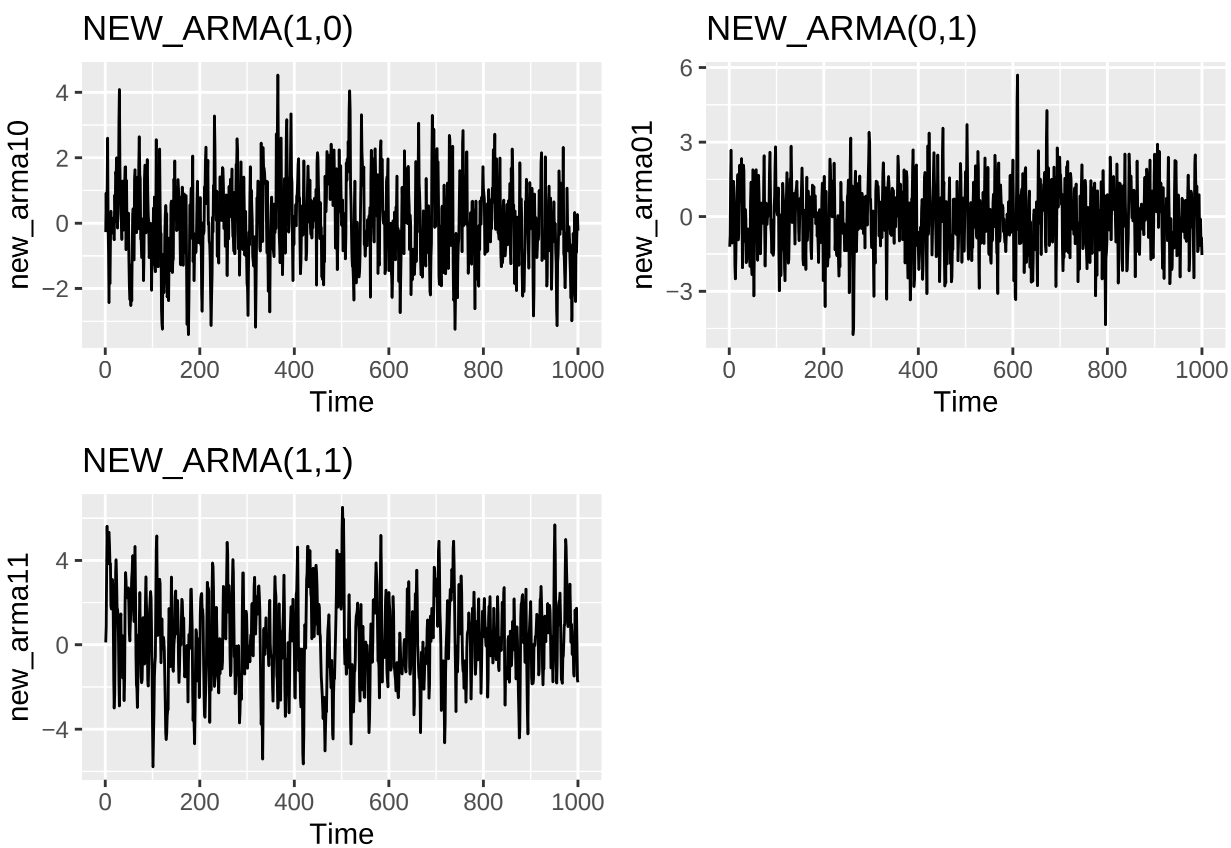
(e) Compare the PACF values R computed with the values you provided for the lag 1 correlation coefficient, i.e., does $\phi = 0.6$ match what you see on PACF for ARMA(1,0), and ARMA(1,1)? Should they match?

Answer: In perfect estimation, the theorethical value $\phi = 0.6$ should match the lag 1 PACF for ARMA(1.0) and ARMA(1,1).

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

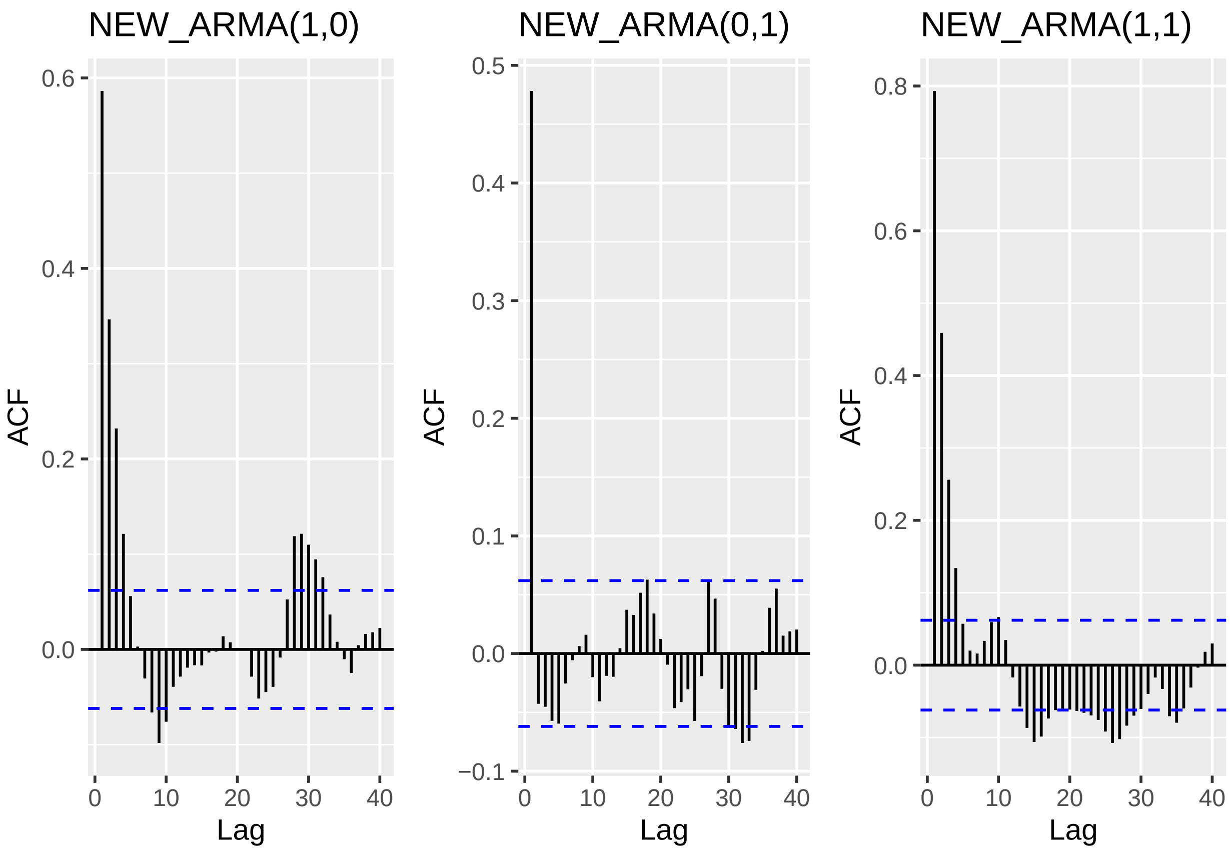
```
# b) Define parameters
phi <- 0.6
theta <- 0.9
new_n <- 1000

# Simulate new ARMA series
new_arma10 <- arima.sim(model = list(ar = phi), n = new_n, innov = rnorm(new_n))
new_arma01 <- arima.sim(model = list(ma = theta), n = new_n, innov = rnorm(new_n))
```

```
new_arma11 <- arima.sim(model = list(ar = phi, ma = theta), n = new_n, innov = rnorm(new_n))

# Plot the new series
plot_grid(
autoplot(new_arma10, main = "NEW_ARMA(1,0)"),
autoplot(new_arma01, main = "NEW_ARMA(0,1)"),
autoplot(new_arma11, main = "NEW_ARMA(1,1)")
)
```



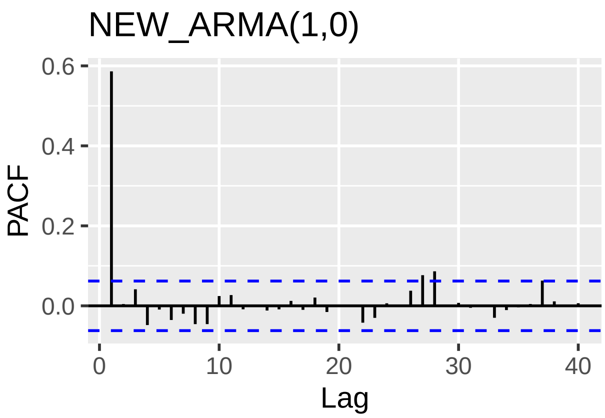

```
#I Create new ACF
new_acf10 <- acf(new_arma10, lag = 40, plot = FALSE)
new_acf01 <- acf(new_arma01, lag = 40, plot = FALSE)
new_acf11 <- acf(new_arma11, lag = 40, plot = FALSE)

#plot NEW ACF
plot_grid(
  autoplot(new_acf10, main = "NEW_ARMA(1,0)"),
  autoplot(new_acf01, main = "NEW_ARMA(0,1)"),
  autoplot(new_acf11, main = "NEW_ARMA(1,1)"),
  nrow = 1
)
```

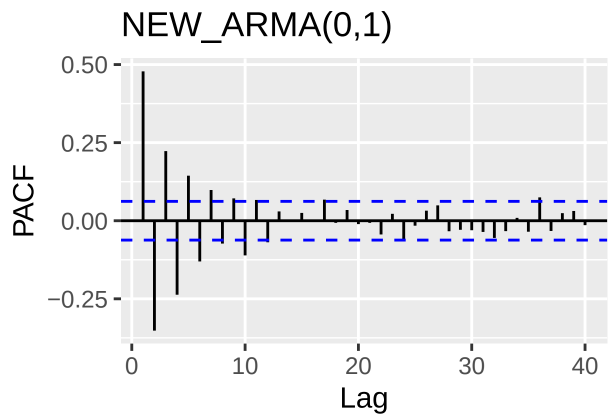
```
## Warning in ggplot2::geom_segment(lineend = "butt", ...): Ignoring unknown parameters: `main`
## Ignoring unknown parameters: `main`
## Ignoring unknown parameters: `main`
```

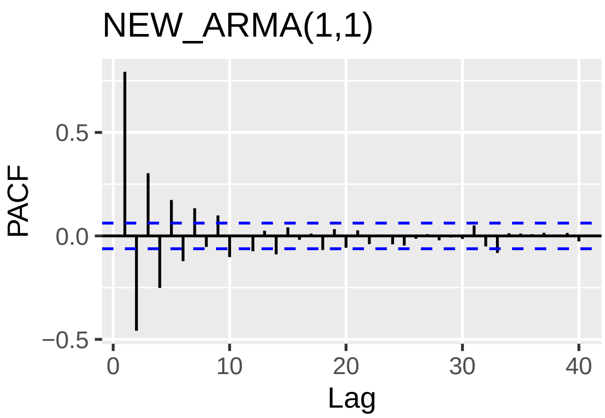



```
# c) I Create New PACF
new_pacf10 <- Pacf(new_arma10, lag = 40, plot = FALSE)
new_pacf01 <- Pacf(new_arma01, lag = 40, plot = FALSE)
new_pacf11 <- Pacf(new_arma11, lag = 40, plot = FALSE)

#plot PACF
plot_grid(
  autoplot(new_pacf10, main = "NEW_ARMA(1,0)"),
  autoplot(new_pacf01, main = "NEW_ARMA(0,1)"),
  autoplot(new_pacf11, main = "NEW_ARMA(1,1)")
)
```

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








```
#d)
#interpretation : The NEW_ARMA(1,0) show an AR process, the NEW_ARMA(0,1),
                  #an MA process, and the NEW_ARMA(1,1) shows a mixture of both AR
#e) interpretation : For the NEW_ARMA(1,0), the PACF at lag 1 is significant
                     #For the NEW_ARMA(1,1) plot, since it includes AR and MA, PACF at
```

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.

#For the non-seasonal part: There is one autoregressive term at lag 1, so p=1 There is no differencing term indicated directly,d=0 There is one moving average term at lag 1, so q=1.

#For the seasonal part:

There is one seasonal autoregressive term at lag 12, so P=1. There is no direct indication of seasonal differencing,D=0. There is no seasonal moving average term, so Q=0. The seasonal period, s being the the lag of the seasonal term, which is 12.

#The model can be written ARIMA(1,0,1)(1,0,0)_12.

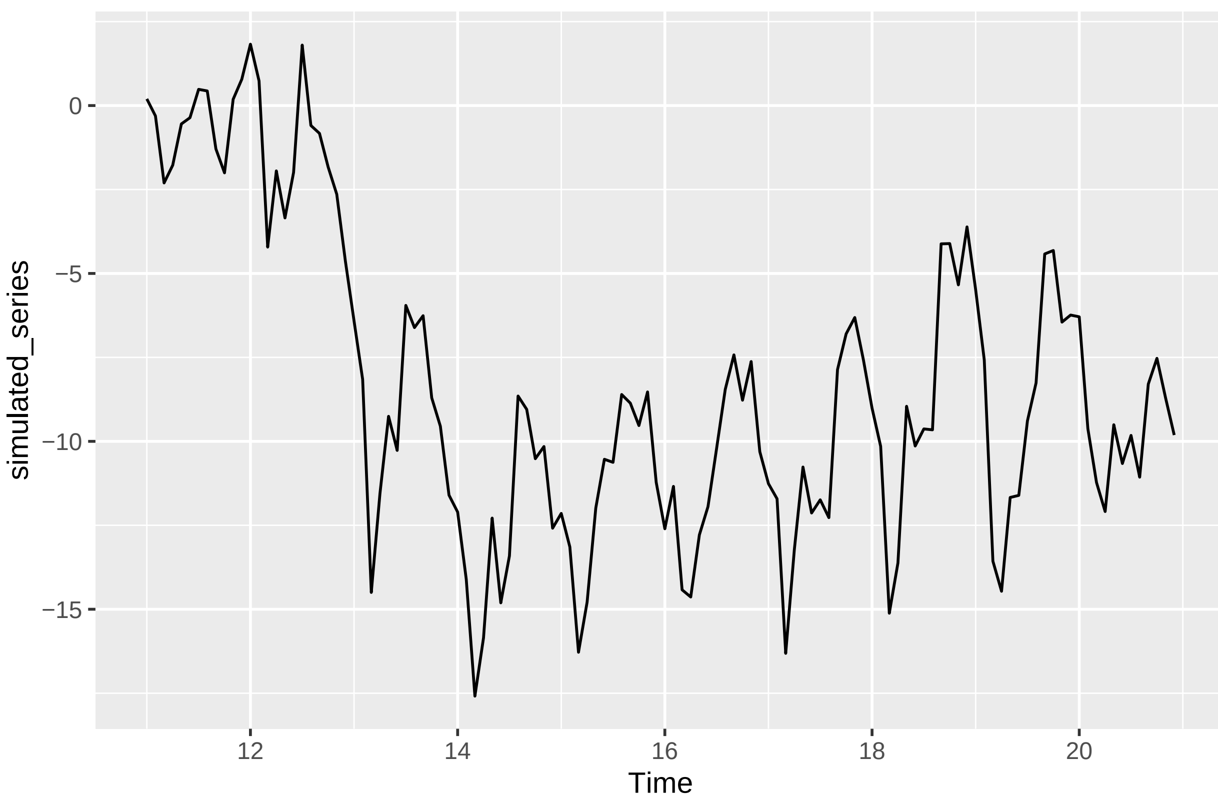
(b) Also from the equation what are the values of the parameters, i.e., model coefficients.

#Here are the coefficients Non-seasonal AR coefficient at lag 1: 1 = 0.7 Seasonal AR coefficient at lag 12:Φ1= −0.25 Non-seasonal MA coefficient at lag 1: 1=−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?

```
#Simulate data using the correct SARIMA model
set.seed(123) # For reproducibility
simulated_data <- Arima(ts(rnorm(120), frequency = 12),
                        order=c(0,1,0), seasonal=c(1,0,0),
                        fixed=c(phi=0.8), include.constant=FALSE)

# Simulate 120 observations from the SARIMA model
simulated_series <- simulate(simulated_data, nsim=120)

# I Plot the simulated series
autoplot(simulated_series)
```

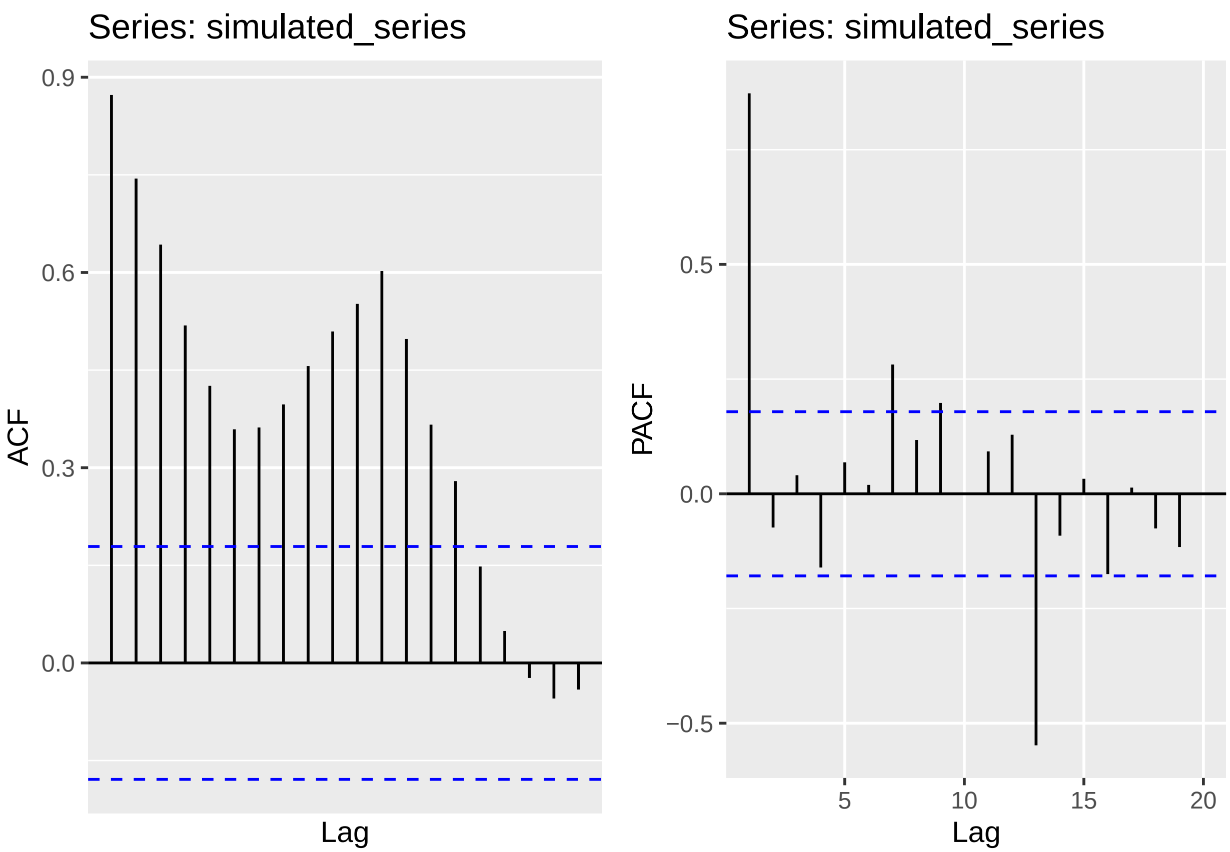


We see a seasonal pattern in the data.

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.

```
simul_acf <- acf(simulated_series, lag = 20, plot = FALSE)
simul_pacf <- Pacf(simulated_series, lag = 20, plot = FALSE)

plot_grid (
  autoplot(simul_acf),
  autoplot(simul_pacf)
)
```



ACF Plot shows a gradual decline in autocorrelations. PACF Plot: exhibits a significant spike at lag 1 and then cuts off, indicating a non-seasonal AR(1) process.