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

Assignment 6 - Due date 02/28/24

Samantha Pace

## 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.

```
#Load/install required package here
#install.packages("ggplot2")
#install.packages("forecast")
#install.packages("tseries")
#install.packages("sarima")
#install.packages("ggfortify")
#install.packages("cowplot")

library(ggplot2)
library(forecast)
```

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

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

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

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

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

```r
library(ggfortify)
```

```
## Registered S3 methods overwritten by 'ggfortify':
##   method                 from
##   autoplot.Arima         forecast
##   autoplot.acf           forecast
##   autoplot.ar            forecast
##   autoplot.bats          forecast
##   autoplot.decomposed.ts forecast
##   autoplot.ets           forecast
##   autoplot.forecast      forecast
##   autoplot.stl           forecast
##   autoplot.ts            forecast
##   fitted.ar              forecast
##   fortify.ts             forecast
##   residuals.ar           forecast
```

```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 for AR(2) models should have an important characteristic of exponential decay over time. The PACF for AR models identifies the order of the model. For an AR(2) model, the PACF will be significant at lag 1 and lag 2, since the order of the model is 2, and insignificant for the rest of the lags.

- MA(1)

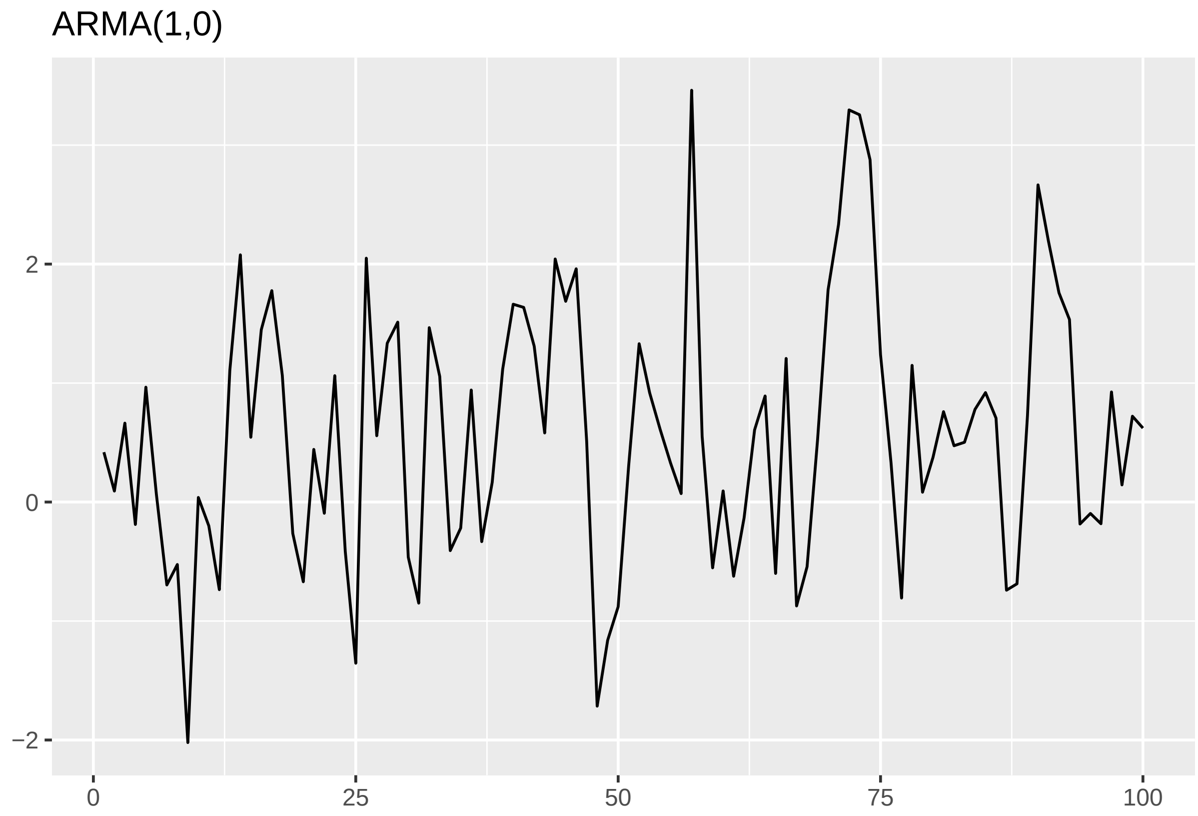
  Answer: The ACF for MA(1) models will indicate the order of the model. There will usually only be one or two significant lags, indicating it is a MA model. The PACF will show the exponential decay with many significant lags.

## Q2

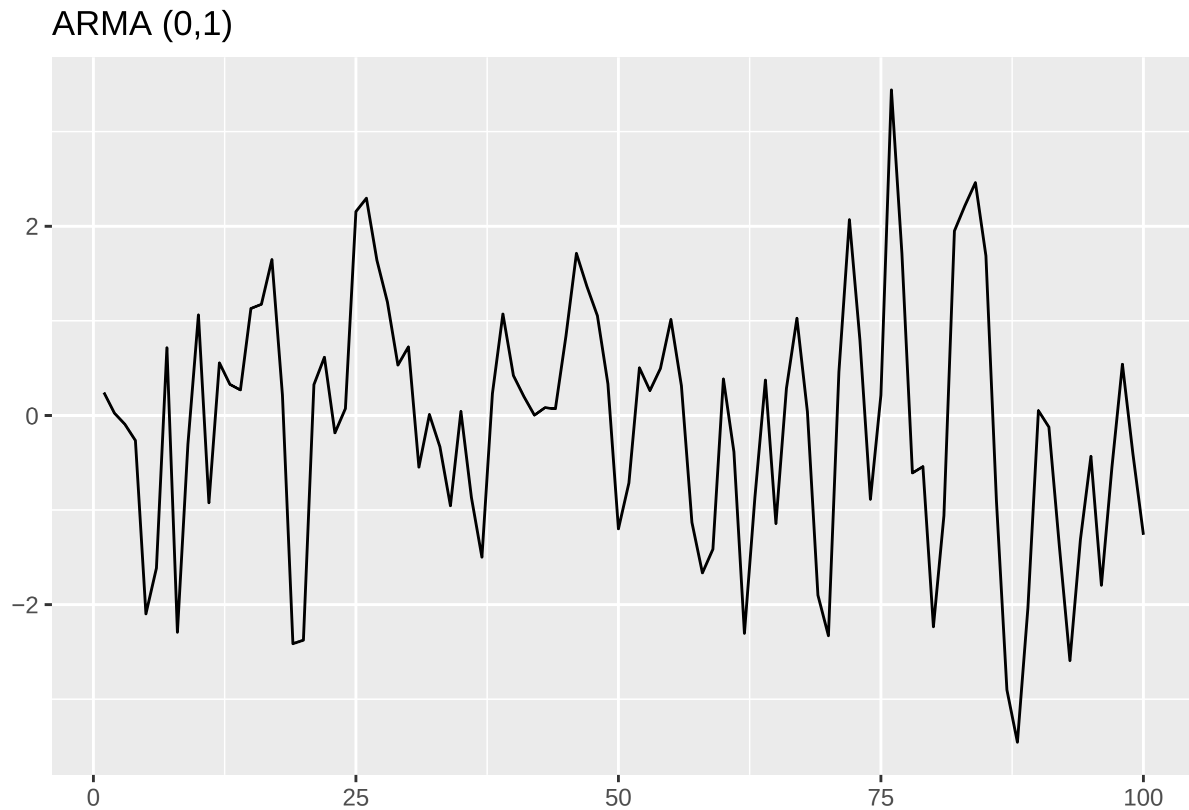
Recall that the non-seasonal ARIMA is described by three parameters ARIMA$(p, d, q)$ where $p$ is the order of the autoregressive component, $d$ is the number of times the series need to be differenced to obtain stationarity and $q$ is the order of the moving average component. If we don't need to difference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

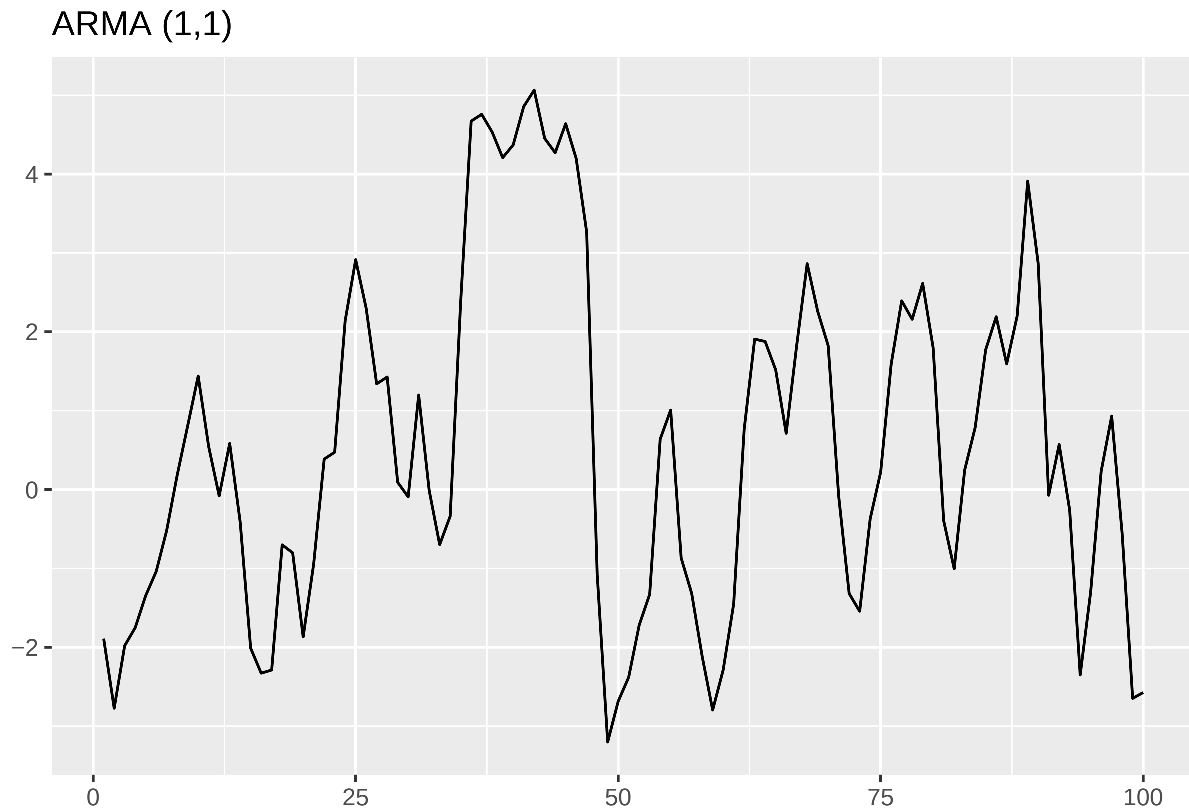
```r
# ARMA(1,0)
model10 <- arima.sim(model = list(order = c(1,0,0), ar = 0.6), n = 100)
autoplot(model10) +
  ggtitle("ARMA(1,0)")
```




```r
# ARMA (0,1)
model01 <- arima.sim(model = list(order = c(0,0,1), ma = 0.9), n = 100)
autoplot(model01) +
  ggtitle("ARMA (0,1)")
```

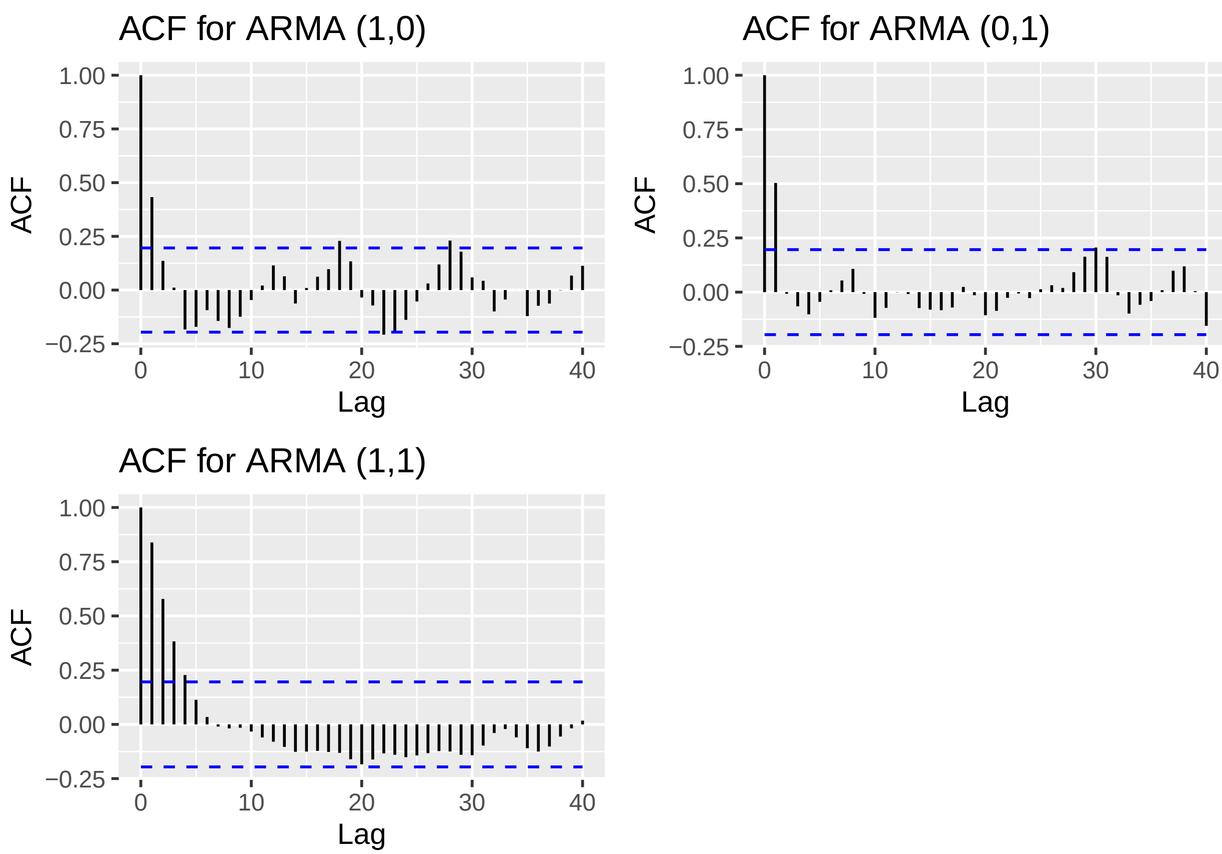



```r
# ARMA (1,1)
model11 <- arima.sim(model = list(order = c(1,0,1),
                                  ar = 0.6,
                                  ma = 0.9), n = 100)
autoplot(model11) +
  ggtitle("ARMA (1,1)")
```



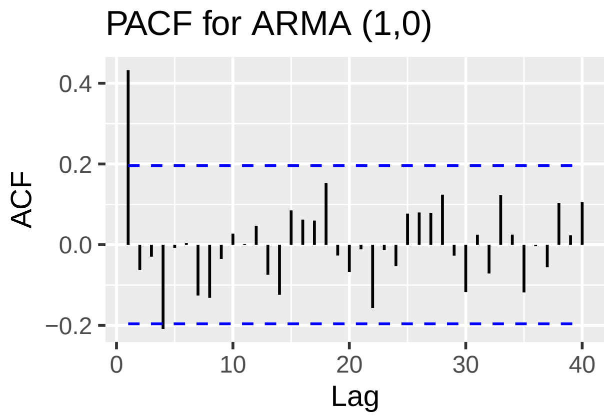

(b) Plot the sample ACF for each of these models in one window to facilitate comparison (Hint: use `cowplot::plot_grid()`).

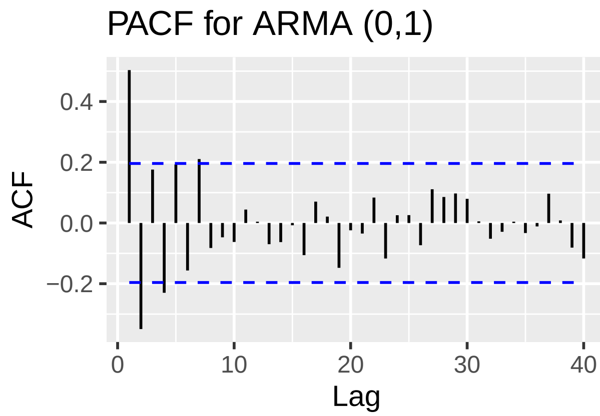
```
# plotting ACFs for all three models
plot_grid(
  autoplot(Acf(model10, lag = 40, plot=FALSE),
                 main = "ACF for ARMA (1,0)"),
  autoplot(Acf(model01, lag = 40, plot=FALSE),
                  main = "ACF for ARMA (0,1)"),
  autoplot(Acf(model11, lag = 40, plot=FALSE),
                  main = "ACF for ARMA (1,1)")
)
```

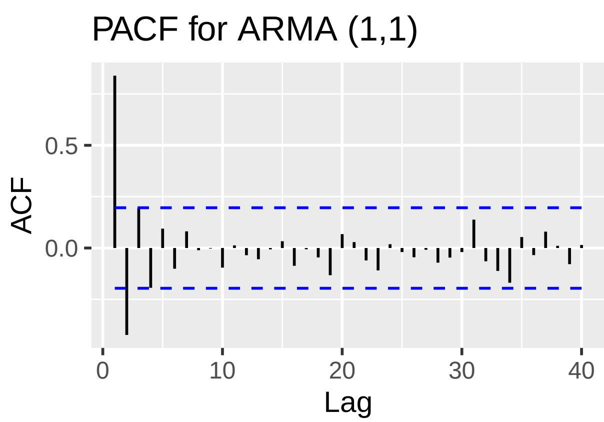



(c) Plot the sample PACF for each of these models in one window to facilitate comparison.

```
# plotting PACFs for all three models
plot_grid(
  autoplot(Pacf(model10, lag = 40, plot=FALSE),
                  main = "PACF for ARMA (1,0)"),
  autoplot(Pacf(model01, lag = 40, plot=FALSE),
                    main = "PACF for ARMA (0,1)"),
  autoplot(Pacf(model11, lag = 40, plot=FALSE),
                    main = "PACF for 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: I would not be able to identify them and their order correctly. The plots for the ACFs and PACFs are not consistent with the expectations of AR or MA, so if I were to identify the model I would probably guess ARMA. Additionally,I would guess it to be a low number for the order because the number of significant lags is generally between 1-3.

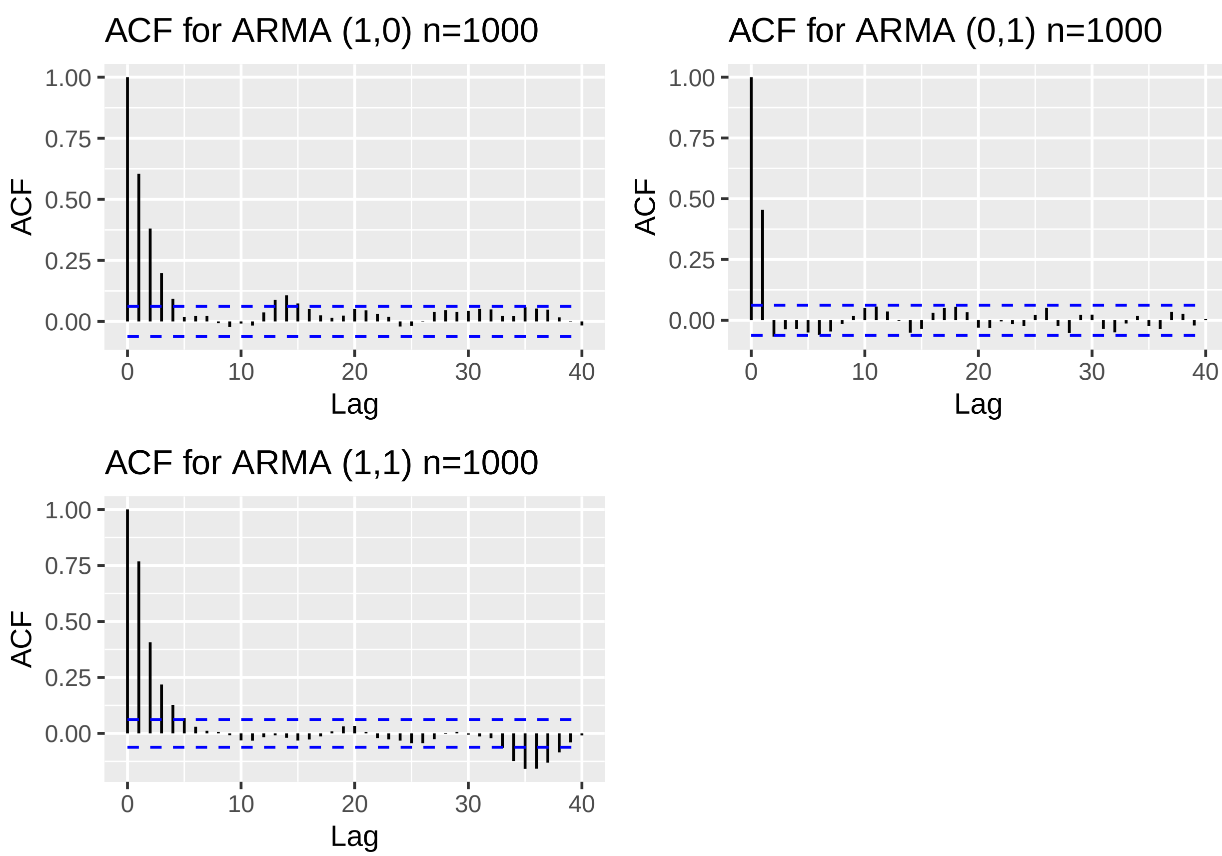
(e) Compare the PACF values R computed with the values you provided for the lag 1 correlation coefficient, i.e., does $\phi = 0.6$ match what you see on PACF for ARMA(1,0), and ARMA(1,1)? Should they match?

Answer: They do not match. Lag 1 coefficient for ARMA(1,0) is about 0.04. Lag 1 coefficient for ARMA(1,1) is about -0.5. They should not match because ARMA(1,0) has only the first order of AR, while ARMA(1,1) has first order of AR and first order of MA, which affect the ultimate PACF values produced.

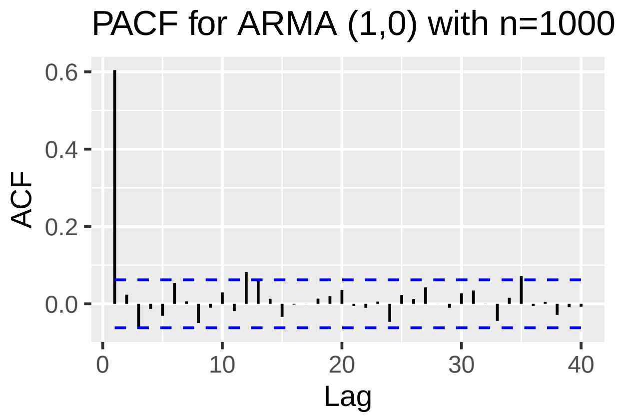
(f) Increase number of observations to $n = 1000$ and repeat parts (b)-(e).

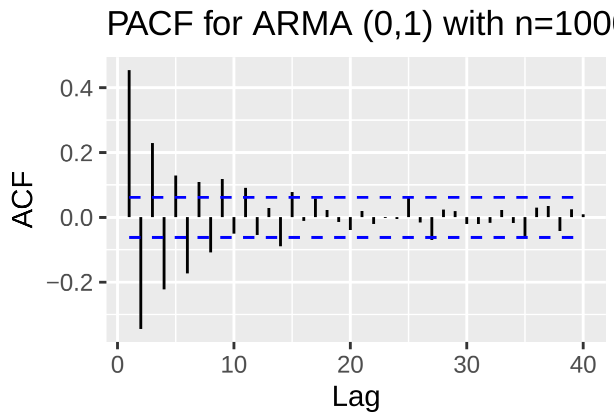
```
# ARMA(1,0)
model10_1000 <- arima.sim(model = list(order = c(1,0,0), ar = 0.6), n = 1000)

# ARMA (0,1)
model01_1000 <- arima.sim(model = list(order = c(0,0,1), ma = 0.9), n = 1000)
```

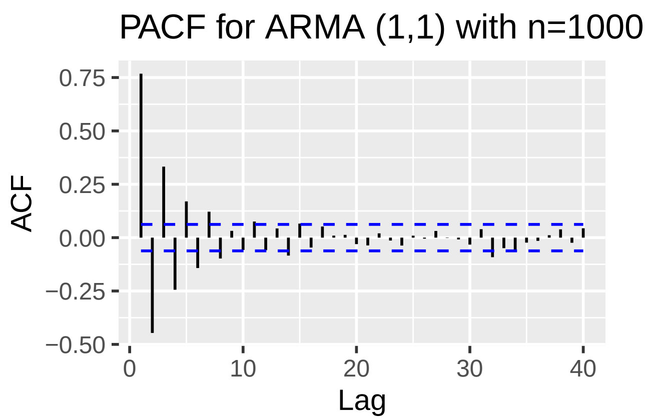
```r
# ARMA (1,1)
model11_1000 <- arima.sim(model = list(order = c(1,0,1),
                                       ar = 0.6,
                                       ma = 0.9), n = 1000)

# plotting ACFs for all three models with n=1000
plot_grid(
  autoplot(Acf(model10_1000, lag = 40, plot=FALSE),
               main = "ACF for ARMA (1,0) n=1000"),
  autoplot(Acf(model01_1000, lag = 40, plot=FALSE),
                 main = "ACF for ARMA (0,1) n=1000"),
  autoplot(Acf(model11_1000, lag = 40, plot=FALSE),
                 main = "ACF for ARMA (1,1) n=1000")
)
```




```r
# plotting PACFs for all three models with n = 1000
plot_grid(
  autoplot(Pacf(model10_1000, lag = 40, plot=FALSE),
               main = "PACF for ARMA (1,0) with n=1000"),
  autoplot(Pacf(model01_1000, lag = 40, plot=FALSE),
                 main = "PACF for ARMA (0,1) with n=1000"),
  autoplot(Pacf(model11_1000, lag = 40, plot=FALSE),
                 main = "PACF for ARMA (1,1) with n=1000")
)
```




Answer: Based on the ARMA models with n=1000, the expected characteristics of the ACFs and PACFs come to light more than the models with n=100 so I may be able to tell what type of model it is and what the order is. The ACF for ARMA(1,0) shows exponential decay, which appears to be an AR model, and there is only one significant lag on the PACF showing that is probably of the first order. The ACF for ARMA(0,1) shows significance at lag 0 and 1, meaning that it is probably a MA model with an order of 1, and the PACF is showing something that may be considered exponential decay. As for the ARMA(1,1) the ACF and PACF do not behave according to the expectations for either AR or MA models, so I would expect this to be an ARMA model, but unsure of the order number.

The PACF values for ARMA(1,0) and ARMA(1,1) at lag 1 do appear to be about the same.

## 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$(1, 0, 1)(1, 0, 0)_s$

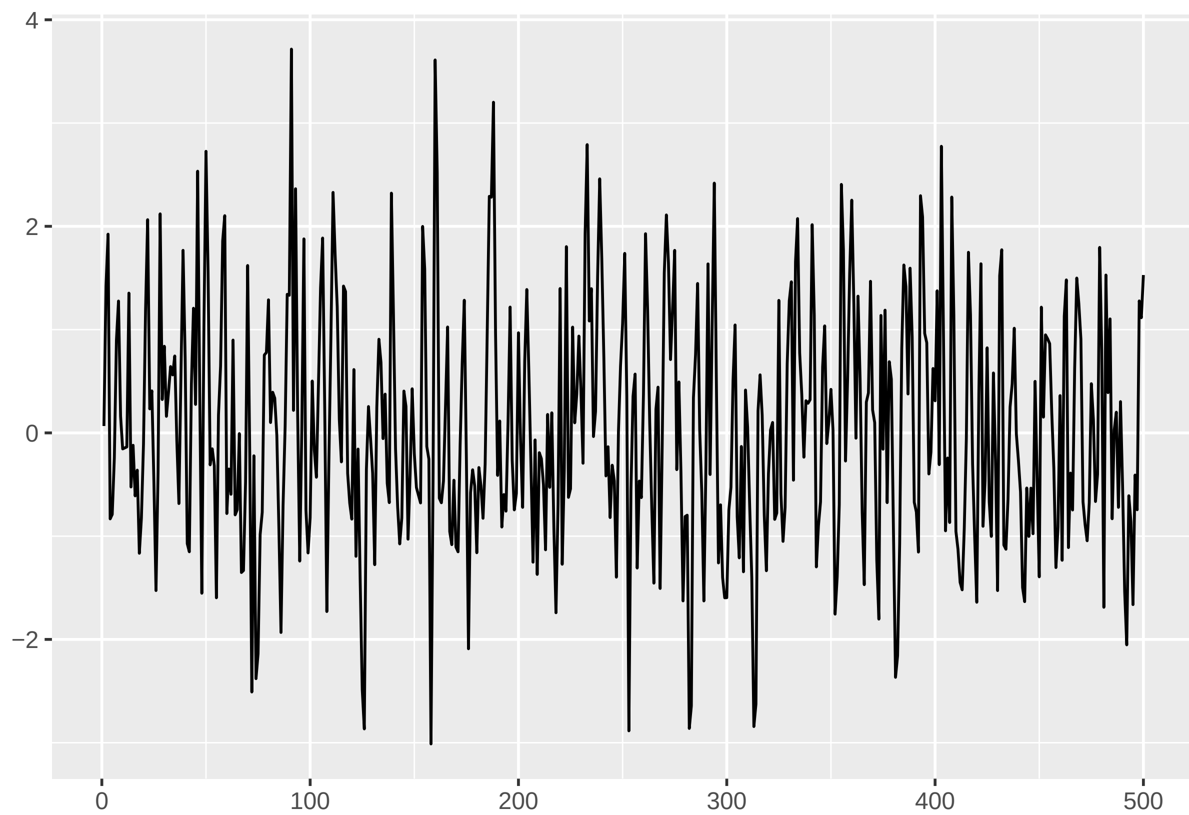
(b) Also from the equation what are the values of the parameters, i.e., model coefficients.

Answer: The AR coefficient is 0.7. The seasonal AR coefficient is -0.25. The 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
# making model
model0110_s <- arima.sim(model = list(order = c(0,0,1),
                                      seasonal = c(1,0,0),
                                      period = 12,
                         ma = 0.5,
                         sar = 0.8),
                         n = 500)
#plotting model
autoplot(model0110_s)
```

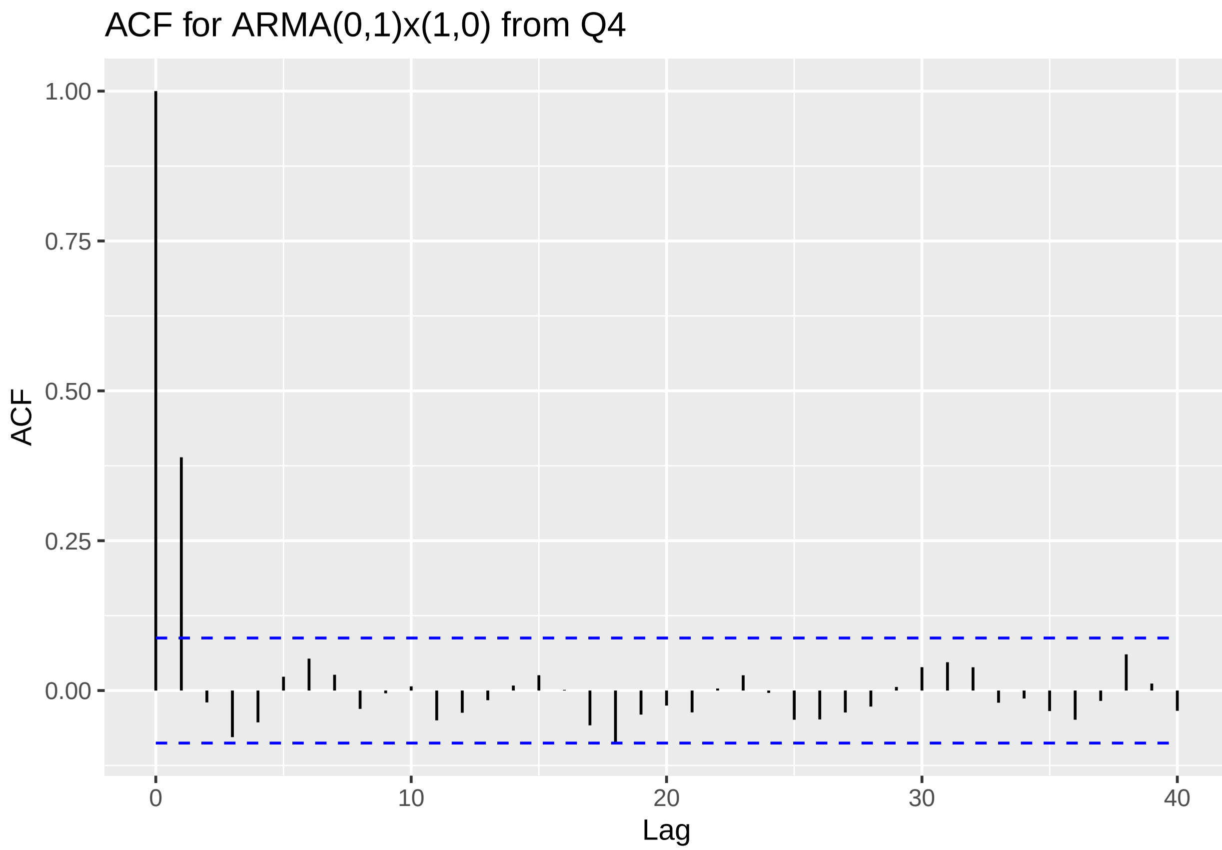



Answer: It is difficult to discern seasonality.

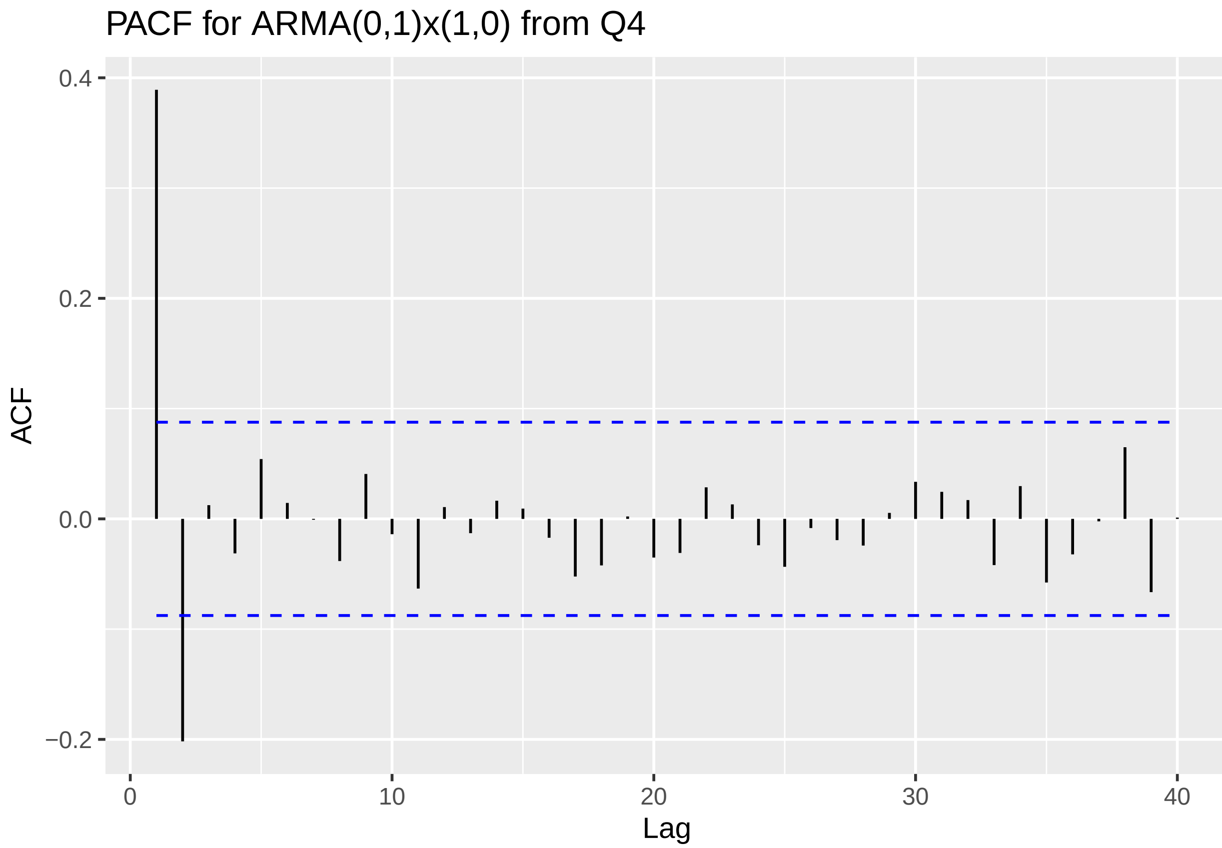
## Q5

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.

```r
# Plotting ACF for Q4 data
autoplot(Acf(model0110_s, lag = 40, plot=FALSE),
         main = "ACF for ARMA(0,1)x(1,0) from Q4")
```




```r
# Plotting PACF for Q4 data
autoplot(Pacf(model0110_s, lag = 40, plot=FALSE),
         main = "PACF for ARMA(0,1)x(1,0) from Q4")
```




Answer: The ACF shows a wavelike pattern, which indicates to me there is a seasonal component. Lag 1 in the ACF is significant, and there looks to be slight significance at lag 30, but I am thinking that is the seasonal component, but considering that lag 1 is the primary source of significance, that reflects the MA characteristics and that the order is 1. It seems difficult to determine the order of the seasonality based on the PACF and ACF. Overall, I think these plots do an okay job of representing the model – it tells some information, but not all.