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

Assignment 6 - Due date 02/27/26

Matthew Fischer

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.

```
#Load/install required package here
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
```

```
library(cowplot)
library(sarima)
```

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 tails off gradually and PACF cuts off after lag = 2.

- MA(1)

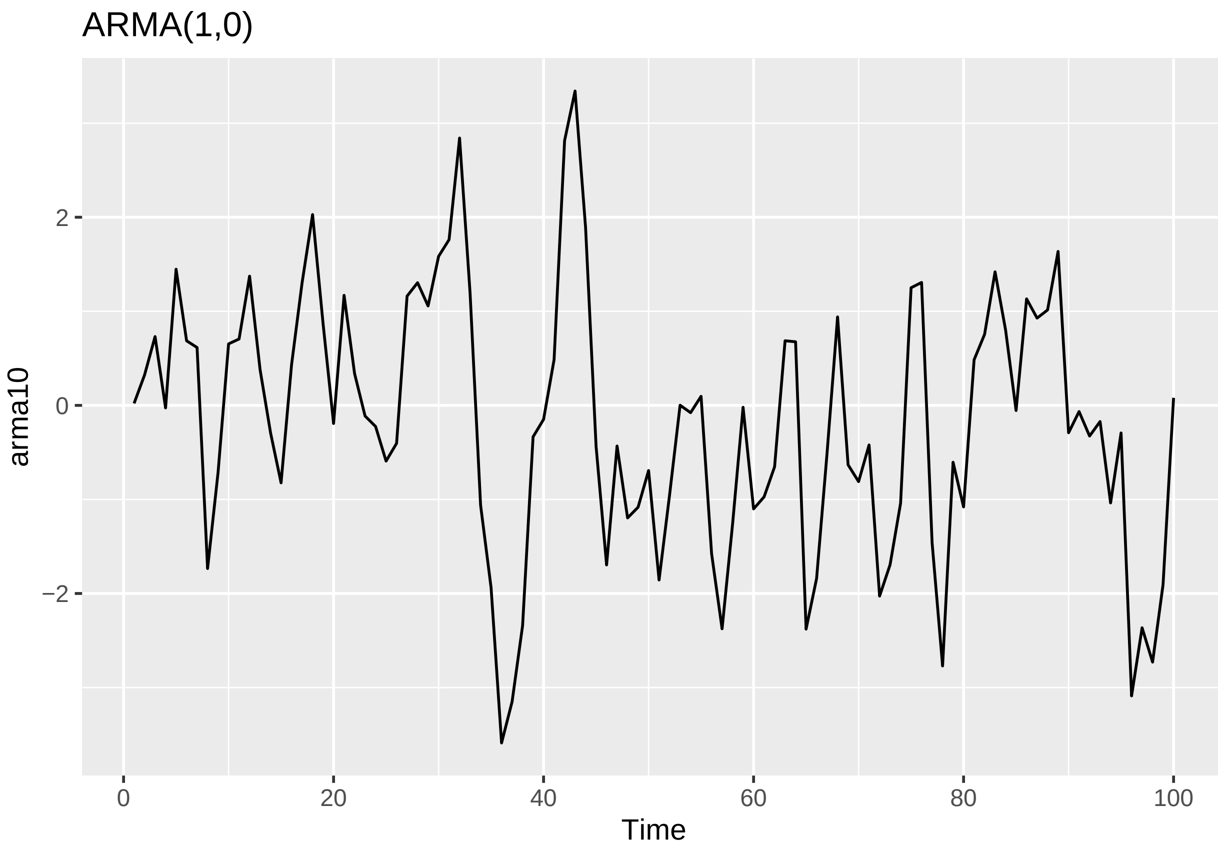
  Answer: ACF cuts off after lag = 1 and PACF tails off gradually.

Q2

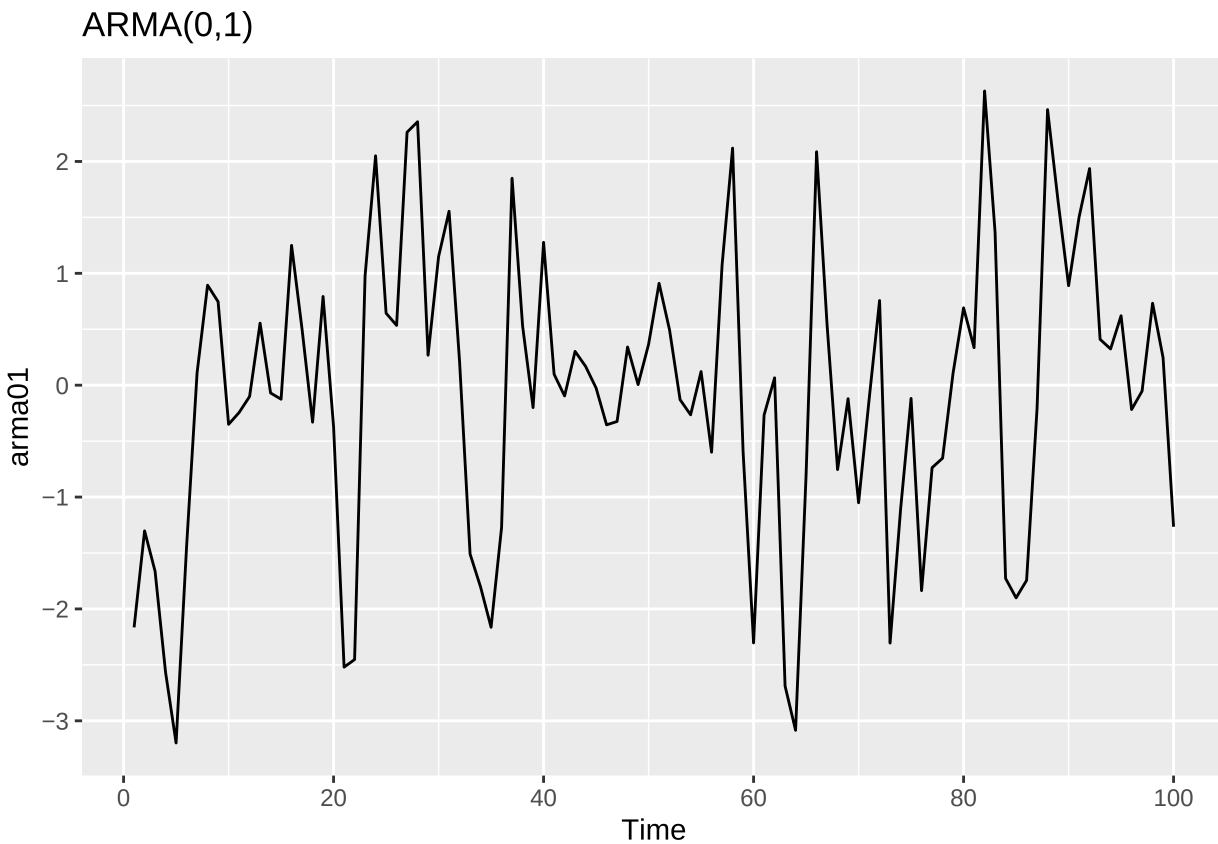
Recall that the non-seasonal ARIMA is described by three parameters ARIMA(p, d, q) where p is the order of the autoregressive component, d is the number of times the series need to be differenced to obtain stationarity and q is the order of the moving average component. If we don't need to difference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

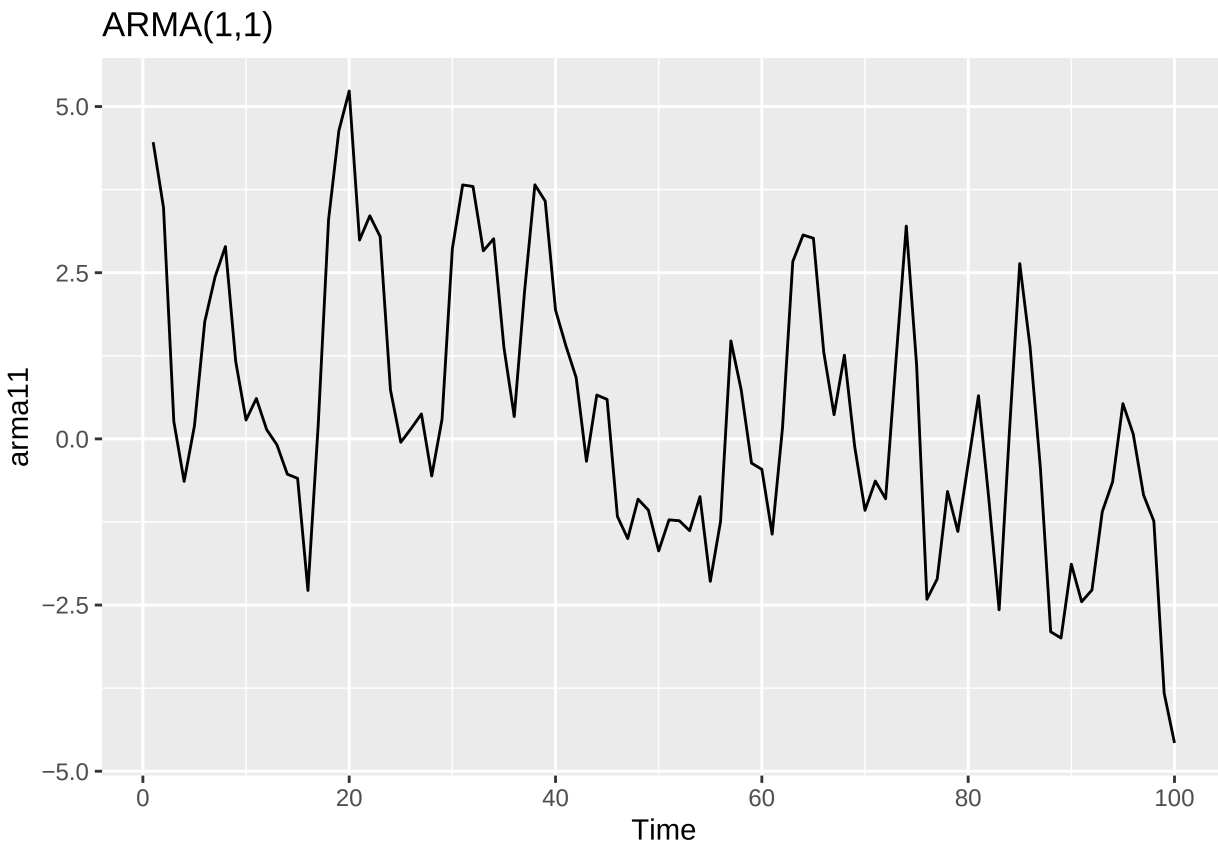
```r
arma10 <- arima.sim(
  n = 100,
  model = list(ar = 0.6)
)

autoplot(arma10) +
  ggtitle("ARMA(1,0)")
```




```r
arma01 <- arima.sim(
  n = 100,
  model = list(ma = 0.9)
)

autoplot(arma01) +
  ggtitle("ARMA(0,1)")
```

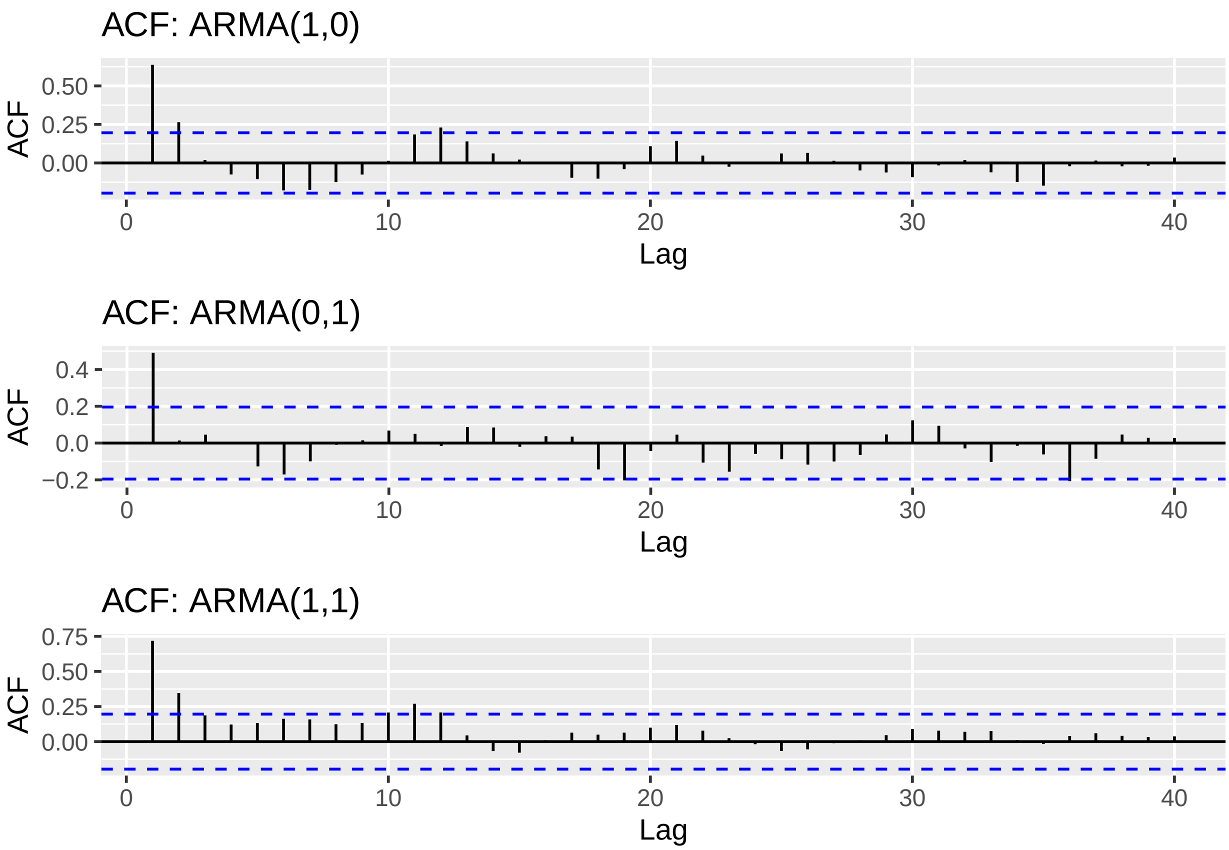



```r
arma11 <- arima.sim(
  n = 100,
  model = list(ar = 0.6, ma = 0.9)
)

autoplot(arma11) +
  ggtitle("ARMA(1,1)")
```



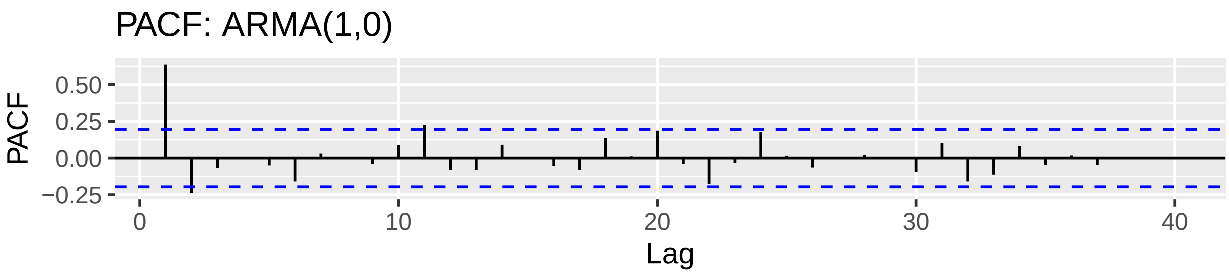
(b) Plot the sample ACF for each of these models in one window to facilitate comparison (Hint: use `cowplot::plot_grid()`).

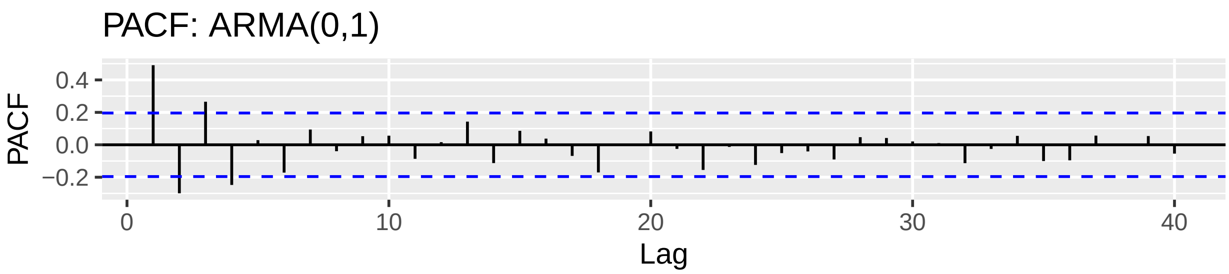
```
acf_10 <- ggAcf(arma10, lag.max = 40) +
  ggtitle("ACF: ARMA(1,0)")

acf_01 <- ggAcf(arma01, lag.max = 40) +
  ggtitle("ACF: ARMA(0,1)")

acf_11 <- ggAcf(arma11, lag.max = 40) +
  ggtitle("ACF: ARMA(1,1)")

plot_grid(acf_10, acf_01, acf_11, ncol = 1)
```

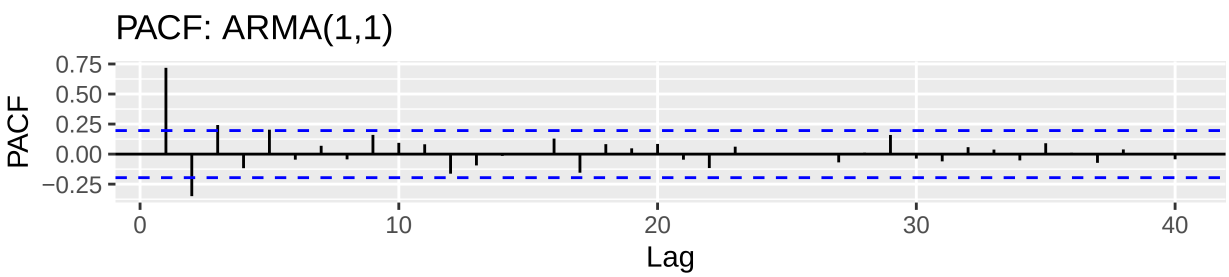


(c) Plot the sample PACF for each of these models in one window to facilitate comparison.

```
pacf_10 <- ggPacf(arma10, lag.max = 40) +
  ggtitle("PACF: ARMA(1,0)")

pacf_01 <- ggPacf(arma01, lag.max = 40) +
  ggtitle("PACF: ARMA(0,1)")

pacf_11 <- ggPacf(arma11, lag.max = 40) +
  ggtitle("PACF: ARMA(1,1)")

plot_grid(pacf_10, pacf_01, pacf_11, ncol = 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: ARMA(1,0) would not be very hard to identify as AR(1) as the ACF decreases gradually and the PACF cuts off after lag = 1. ARMA(0,1) would also not be very hard to identify as MA(1) as the ACF cuts off after lag = 1 and the PACF gradually decreases. ARMA(1,1) would be harder to identify since the ACF and PACF both tail off, making it hard to tell if it's ARMA or a higher-order AR or MA process.

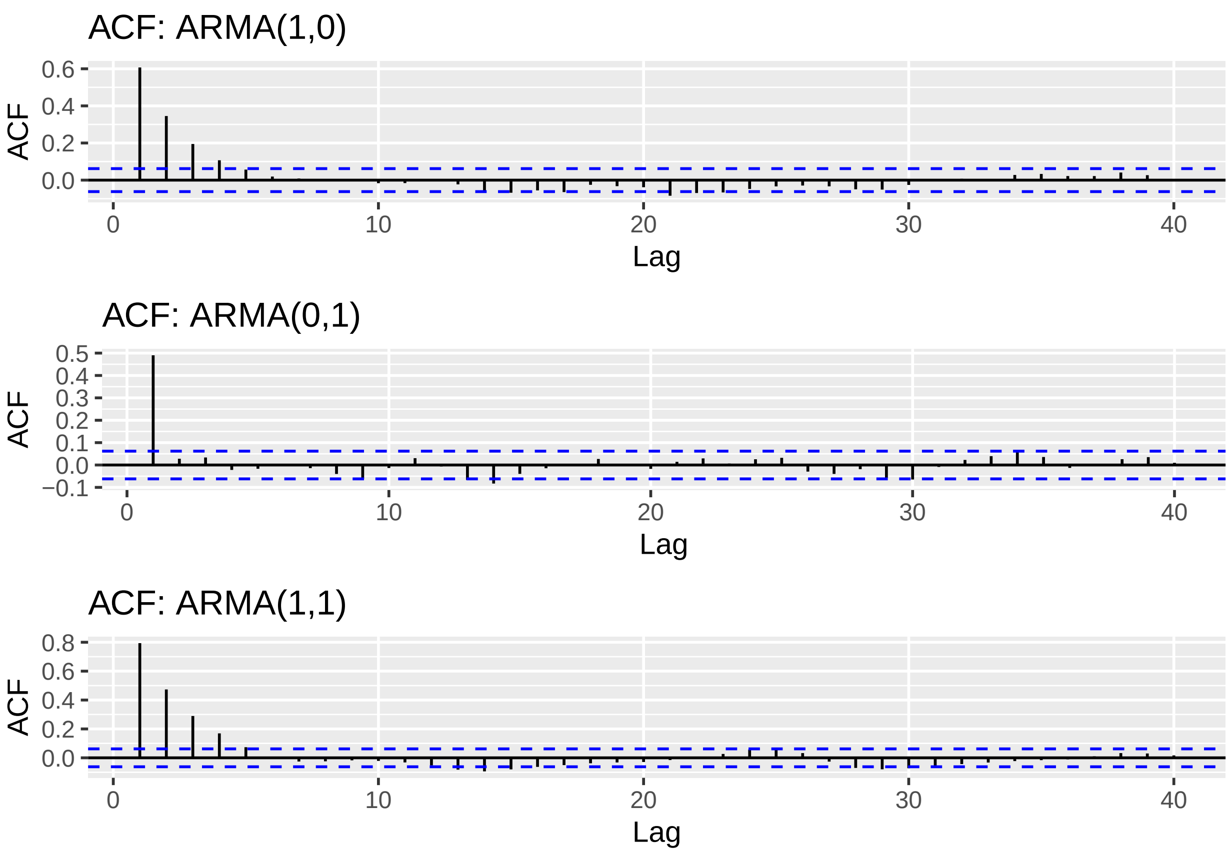
(e) Compare the PACF values R computed with the values you provided for the lag 1 correlation coefficient, i.e., does $\phi = 0.6$ match what you see on PACF for ARMA(1,0), and ARMA(1,1)? Should they match?

Answer: For ARMA(1,0), the PACF value at lag 1 is very close to 0.6 but not exactly. For ARMA(1,1), the PACF value at lag 1 is farther away, around 0.75. ARMA(1,0) should match, ARMA(1,1) likely would not due to the MA component.

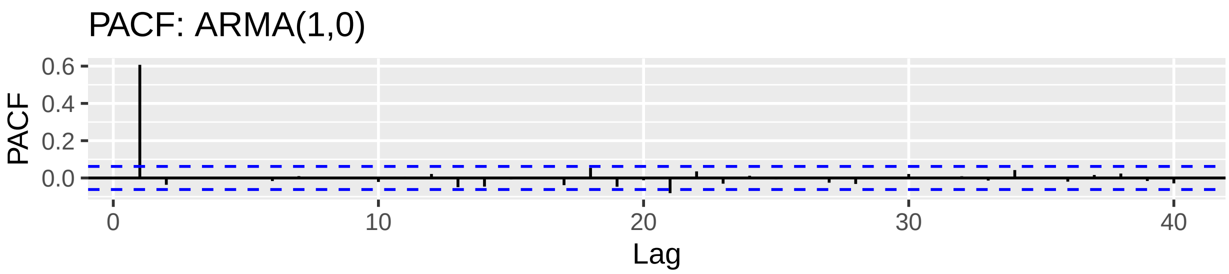
(f) Increase number of observations to $n = 1000$ and repeat parts (b)-(e).

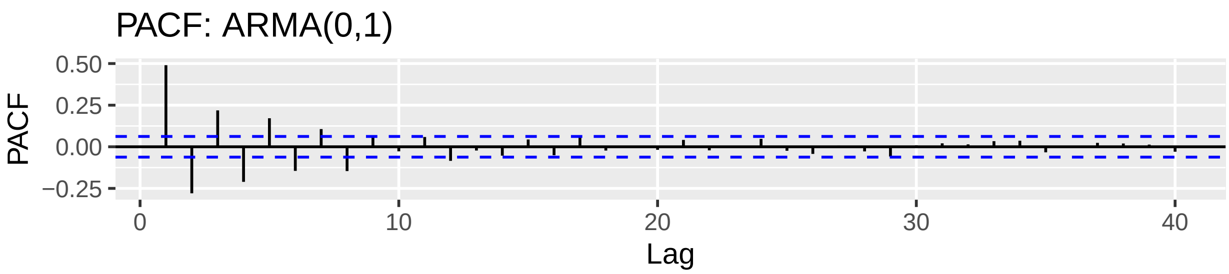
```
arma10_1000 <- arima.sim(
  n = 1000,
  model = list(ar = 0.6)
)

arma01_1000 <- arima.sim(
  n = 1000,
  model = list(ma = 0.9)
)
```

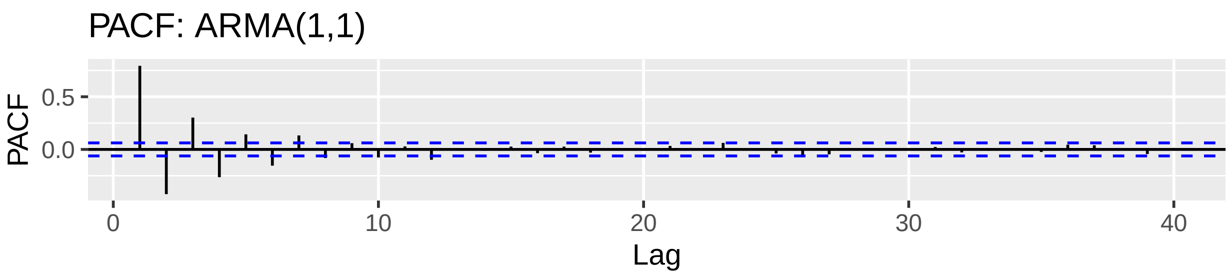
```
arma11_1000 <- arima.sim(
  n = 1000,
  model = list(ar = 0.6, ma = 0.9)
)

acf_10_1000 <- ggAcf(arma10_1000, lag.max = 40) +
  ggtitle("ACF: ARMA(1,0)")

acf_01_1000 <- ggAcf(arma01_1000, lag.max = 40) +
  ggtitle("ACF: ARMA(0,1)")

acf_11_1000 <- ggAcf(arma11_1000, lag.max = 40) +
  ggtitle("ACF: ARMA(1,1)")

plot_grid(acf_10_1000, acf_01_1000, acf_11_1000, ncol = 1)
```




```
pacf_10_1000 <- ggPacf(arma10_1000, lag.max = 40) +
  ggtitle("PACF: ARMA(1,0)")

pacf_01_1000 <- ggPacf(arma01_1000, lag.max = 40) +
  ggtitle("PACF: ARMA(0,1)")

pacf_11_1000 <- ggPacf(arma11_1000, lag.max = 40) +
  ggtitle("PACF: ARMA(1,1)")

plot_grid(pacf_10_1000, pacf_01_1000, pacf_11_1000, ncol = 1)
```







Answer: It is now very easy to identify the patterns in all models, making ARMA(1,0) and ARMA(0,1) easy to identify while ARMA(1,1) wouls still be somewhat ambiguous. The PACF value at lag 1 is nearly equal to 0.6 for ARMA(1,0) and 0.75 for ARMA(1,1).

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: p = 1, d = 0, q = 1, P = 1, D = 0, Q = 0. So the model is ARIMA(1,0,1)(1,0,0)

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

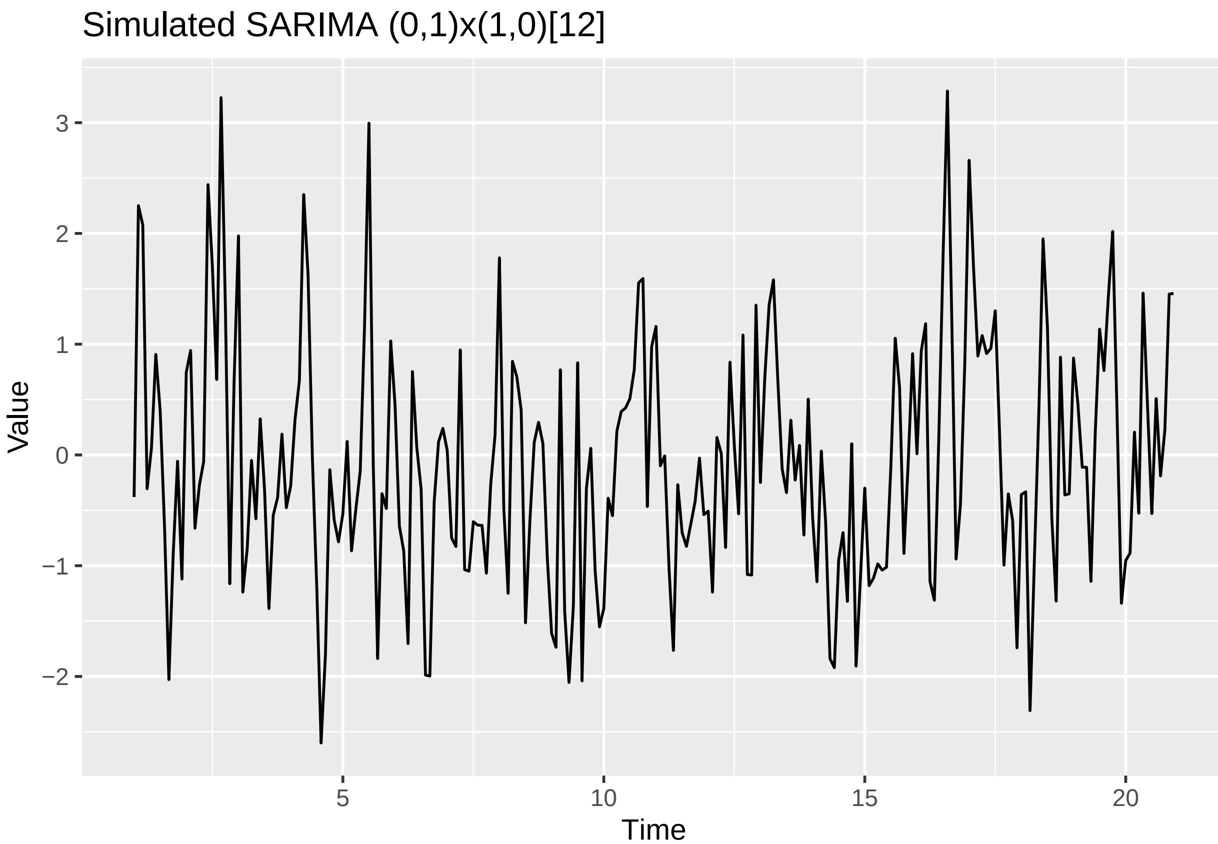
Answer: The non-seasonal AR_1 coefficient (phi_1) is 0.7, the non-seasonal MA_1 coefficient (theta_1) is -0.1, and the seasonal AR_1 coefficient (capital phi_1) is -0.25.

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?

```
sim_data <- sim_sarima(
  n = 240,
  model = list(ma = 0.5, sar = 0.8),
  S = 12
```

```
)

sim_ts <- ts(sim_data, frequency = 12)

autoplot(sim_ts) +
  ggtitle("Simulated SARIMA (0,1)x(1,0)[12]") +
  xlab("Time") +
  ylab("Value")
```



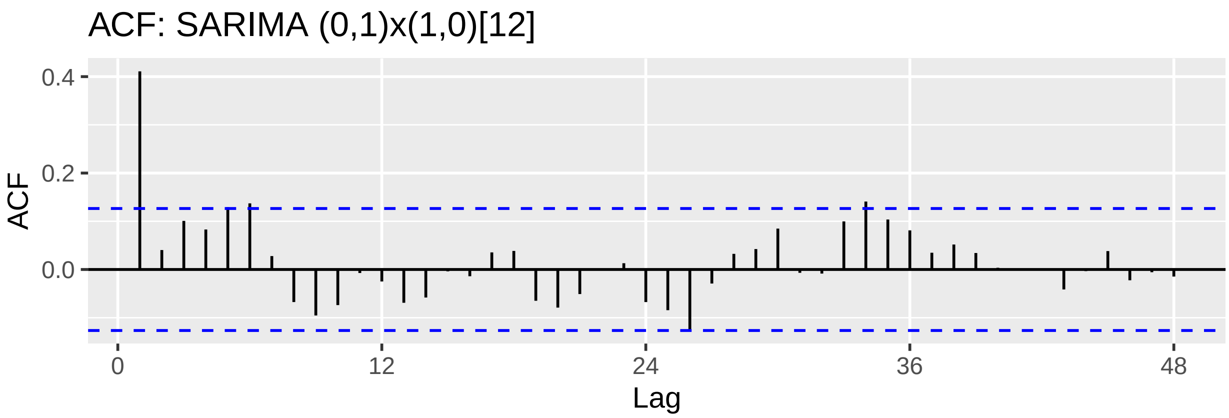
>

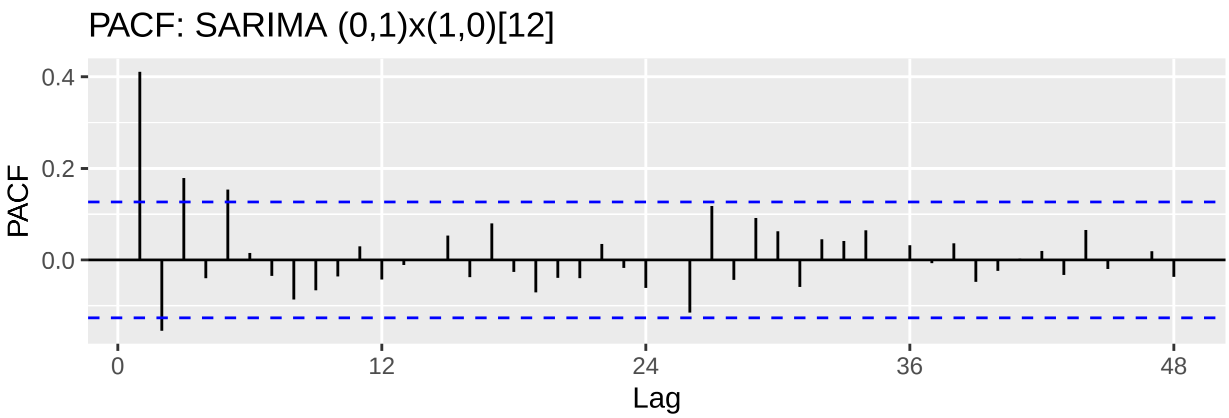
Answer: The data do look seasonal although there is a lot of noise as well that makes them look messy.

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.

```
acf_01_x_10 <- ggAcf(sim_ts, lag.max = 48) +
  ggtitle("ACF: SARIMA (0,1)x(1,0)[12]")

pacf_01_x_10 <- ggPacf(sim_ts, lag.max = 48) +
  ggtitle("PACF: SARIMA (0,1)x(1,0)[12]")

plot_grid(acf_01_x_10, pacf_01_x_10, ncol = 1)
```





>

Answer: There is a large spike at lag = 1 in the ACF which corresponds to MA(1) and the ACF cuts off after that with only a couple other significant points. There do not seem to be very noticeable values around lags = multiples of 12 (seasonal AR) which is not something I would expect. For the PACF, there is no sharp cutoff at lag = 1 (MA(1)) and it tails off after that, there do not seem to be spikes at lag = 12 (seaosnal AR(1)) which is not something I would expect. It does not look like the plots are entirely representative of the model I simulated.