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

Assignment 6 - Due date 02/27/26

Aisha Shen

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(lubridate)
library(ggplot2)
```

```
## Warning: package 'ggplot2' was built under R version 4.5.2
```

```
library(forecast)
```

```
## Warning: package 'forecast' was built under R version 4.5.2
```

```
library(Kendall)
```

```
## Warning: package 'Kendall' was built under R version 4.5.2
```

```
library(tseries)
```

```
## Warning: package 'tseries' was built under R version 4.5.2
```

```r
library(outliers)
library(tidyverse)
```

```
## Warning: package 'tibble' was built under R version 4.5.2
```

```
## Warning: package 'tidyr' was built under R version 4.5.2
```

```
## Warning: package 'readr' was built under R version 4.5.2
```

```
## Warning: package 'purrr' was built under R version 4.5.2
```

```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: For an AR(2) model the ACF will show an exponential decay or a damped sinusoidal pattern (tailing off) while the PACF will have significant spikes at lag 1 and lag 2, and then "cut off" (become statistically insignificant) after lag 2.

- MA(1)

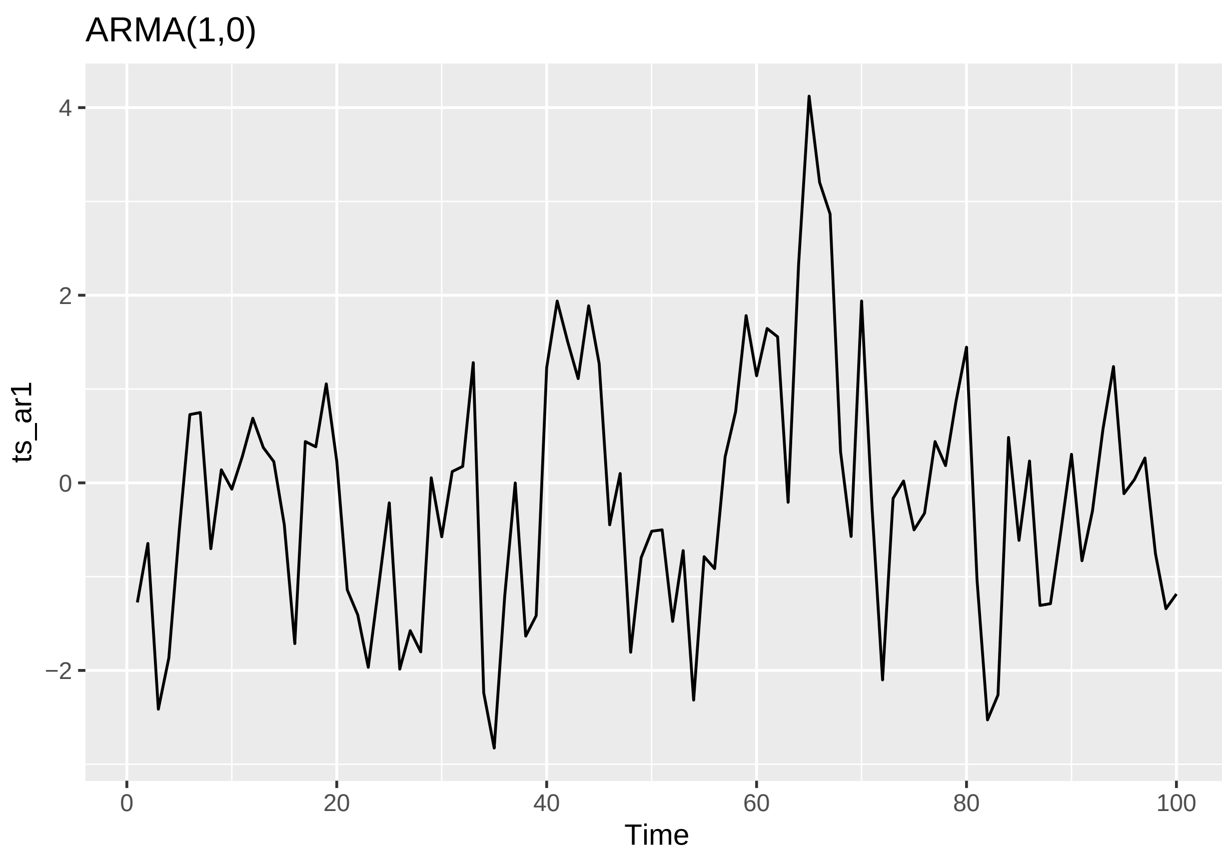
  Answer: For an MA(1) model the ACF will have a single significant spike at lag 1 and then "cut off" immediately after. The PACF will show an exponential decay or a damped sinusoidal pattern (tailing off) toward zero.

Q2

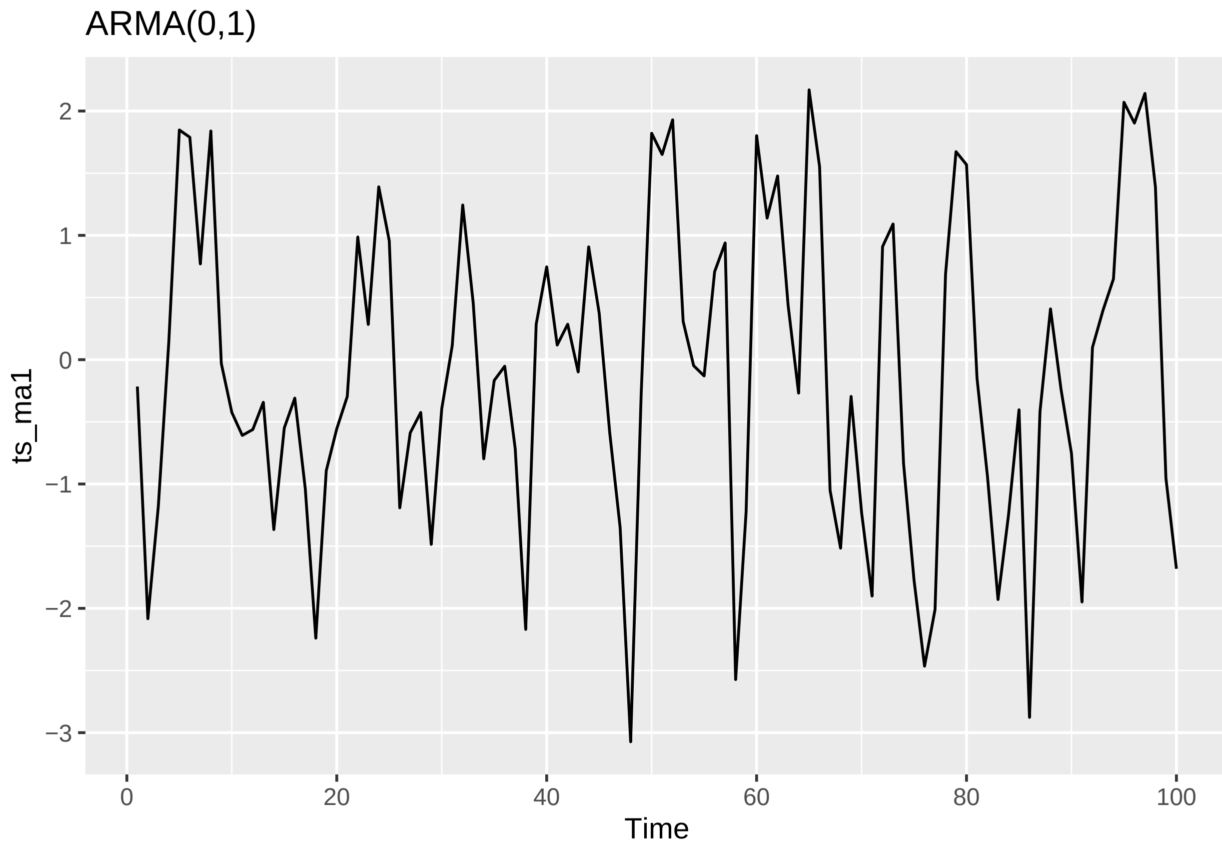
Recall that the non-seasonal ARIMA is described by three parameters ARIMA(p, d, q) where p is the order of the autoregressive component, d is the number of times the series need to be differenced to obtain stationarity and q is the order of the moving average component. If we don't need to difference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

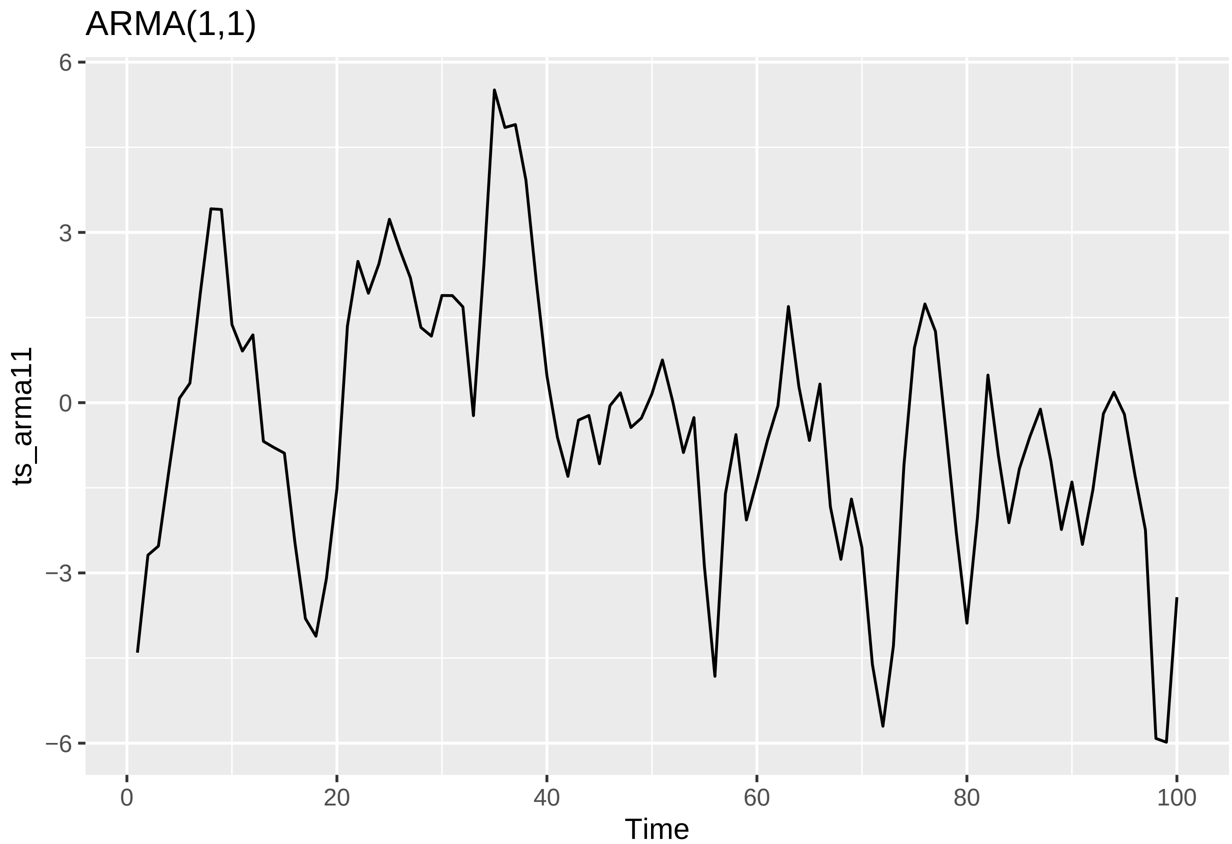
```r
n <- 100
phi <- 0.6
theta <- 0.9

ts_ar1 <- arima.sim(model = list(ar = phi), n = n)
ts_ma1 <- arima.sim(model = list(ma = theta), n = n)
ts_arma11 <- arima.sim(model = list(ar = phi, ma = theta), n = n)

p1 <- autoplot(ts_ar1) + ggtitle("ARMA(1,0)")
p2 <- autoplot(ts_ma1) + ggtitle("ARMA(0,1)")
p3 <- autoplot(ts_arma11) + ggtitle("ARMA(1,1)")

print(p1)
```




```r
print(p2)
```

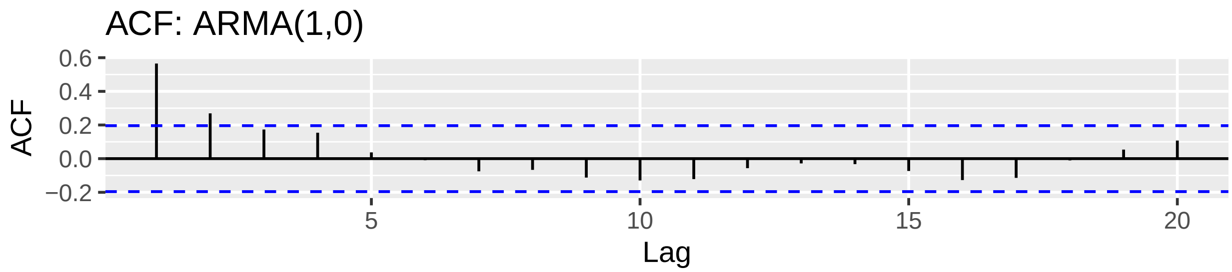


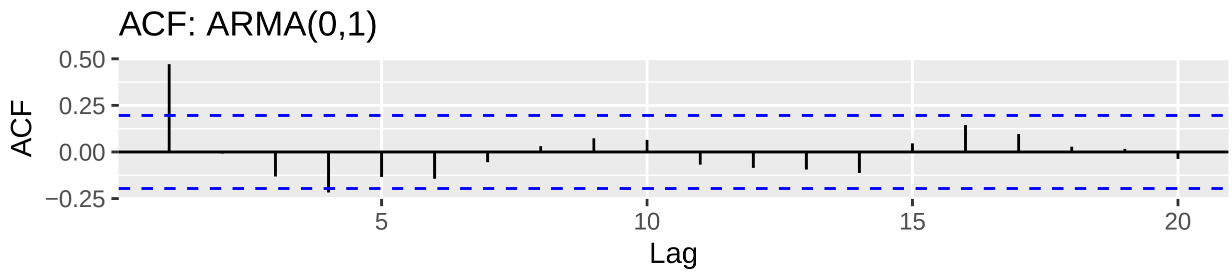

```r
print(p3)
```

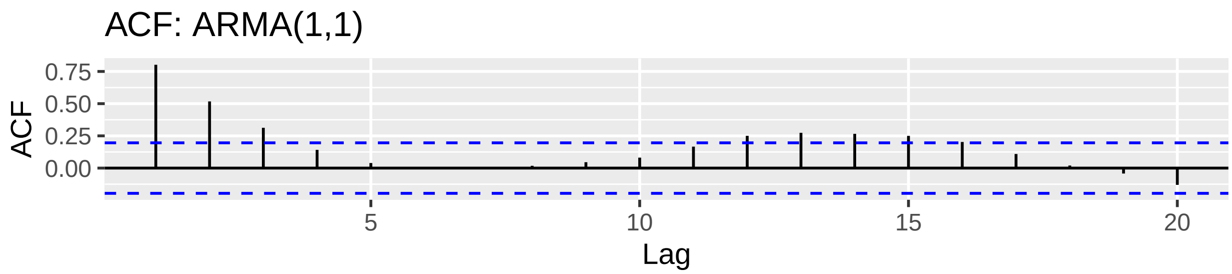


(b) Plot the sample ACF for each of these models in one window to facilitate comparison (Hint: use `cowplot::plot_grid()`).

```
a1 <- ggAcf(ts_ar1) + ggtitle("ACF: ARMA(1,0)")
a2 <- ggAcf(ts_ma1) + ggtitle("ACF: ARMA(0,1)")
a3 <- ggAcf(ts_arma11) + ggtitle("ACF: ARMA(1,1)")

plot_grid(a1, a2, a3, ncol = 1)
```

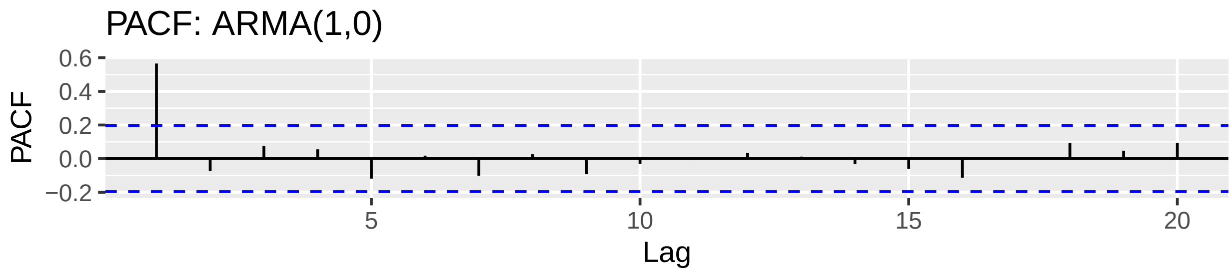


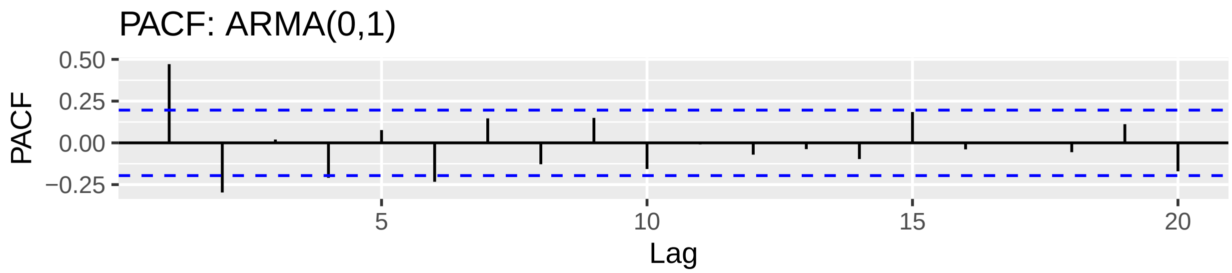


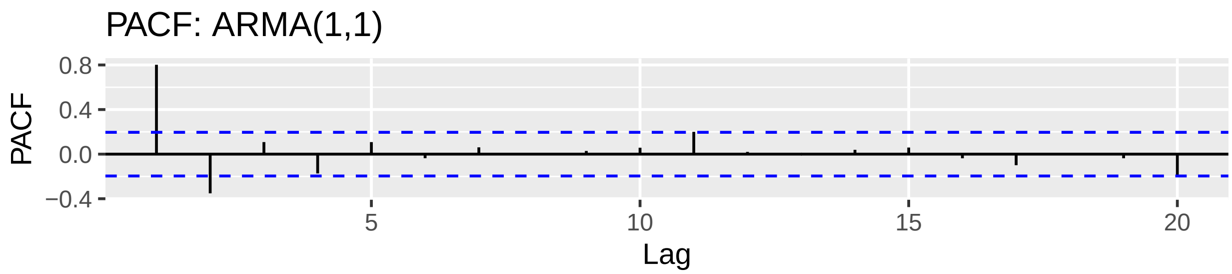


(c) Plot the sample PACF for each of these models in one window to facilitate comparison.

```
pa1 <- ggPacf(ts_ar1) + ggtitle("PACF: ARMA(1,0)")
pa2 <- ggPacf(ts_ma1) + ggtitle("PACF: ARMA(0,1)")
pa3 <- ggPacf(ts_arma11) + ggtitle("PACF: ARMA(1,1)")

plot_grid(pa1, pa2, pa3, 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: With $n = 100$, it is difficult to identify the model due to sampling variation. For AR(1), we expect a cut-off in PACF at lag 1 which we observe above, and for MA(1) we expect a a cut-off in ACF at lag 1 which we also observe. However, the ARMA(1,1) is harder to identify because both plots "tail off," making it easy to confuse these with a higher-order AR or MA model unless the sample size is large enough to see the patterns clearly.

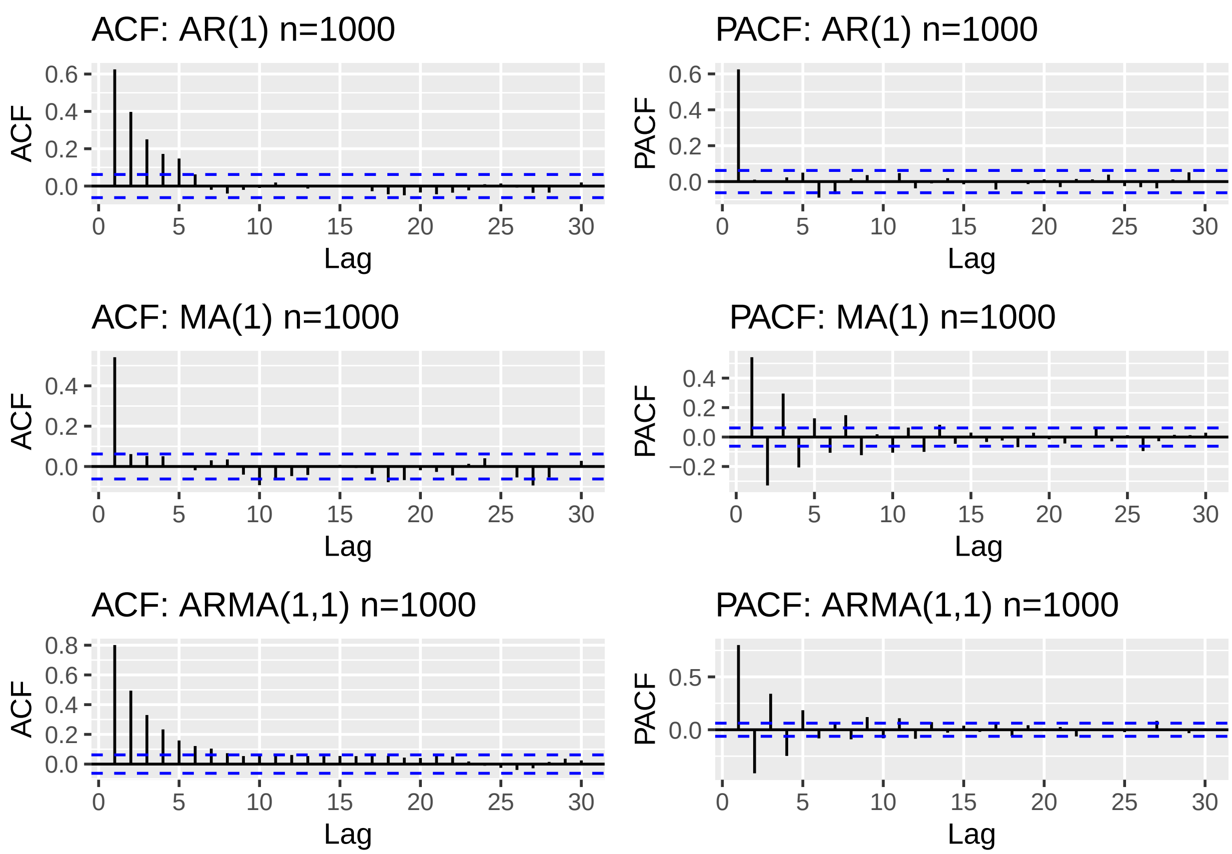
(e) Compare the PACF values R computed with the values you provided for the lag 1 correlation coefficient, i.e., does $\phi = 0.6$ match what you see on PACF for ARMA(1,0), and ARMA(1,1)? Should they match?

Answer: $\phi = 0.6$ matches what we see for PACF for ARMA(1,0) but does not match for ARMA(1,1). This should be expected as For the ARMA(1,1) model, the PACF at lag 1 is a function of both ϕ and θ, so it will not be exactly 0.6.

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

```
n_large <- 1000

ts_ar1_l <- arima.sim(model = list(ar = phi), n = n_large)
ts_ma1_l <- arima.sim(model = list(ma = theta), n = n_large)
ts_arma11_l <- arima.sim(model = list(ar = phi, ma = theta), n = n_large)
```

```r
# ACF Plots
l_a1 <- ggAcf(ts_ar1_l) + ggtitle("ACF: AR(1) n=1000")
l_a2 <- ggAcf(ts_ma1_l) + ggtitle("ACF: MA(1) n=1000")
l_a3 <- ggAcf(ts_arma11_l) + ggtitle("ACF: ARMA(1,1) n=1000")

# PACF Plots
l_pa1 <- ggPacf(ts_ar1_l) + ggtitle("PACF: AR(1) n=1000")
l_pa2 <- ggPacf(ts_ma1_l) + ggtitle("PACF: MA(1) n=1000")
l_pa3 <- ggPacf(ts_arma11_l) + ggtitle("PACF: ARMA(1,1) n=1000")

plot_grid(l_a1,  l_pa1, l_a2,  l_pa2, l_a3, l_pa3, ncol = 2)
```



Inference: By increasing the number of observations we can much clearly see the model structure.

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.

This is a ARIMA$(1, 0, 1) \times (1, 0, 0)_{12}$ model where: $p = 1, d = 0, q = 1,\ P = 1, D = 0, Q = 0$, $s = 12$. The model is autoregressive because of y_{t-1} with a ϕ_1 of 0.7; it is a MA(1) model with coefficient θ of 0.1 as it has the a_{t-1} term, and it is seasonal as it has the y_{t-12} term showing annual seasonality.

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

AR coefficient (ϕ_1): 0.7 Seasonal AR coefficient (Φ_1): -0.25 Non-seasonal MA coefficient (θ_1): 0.1

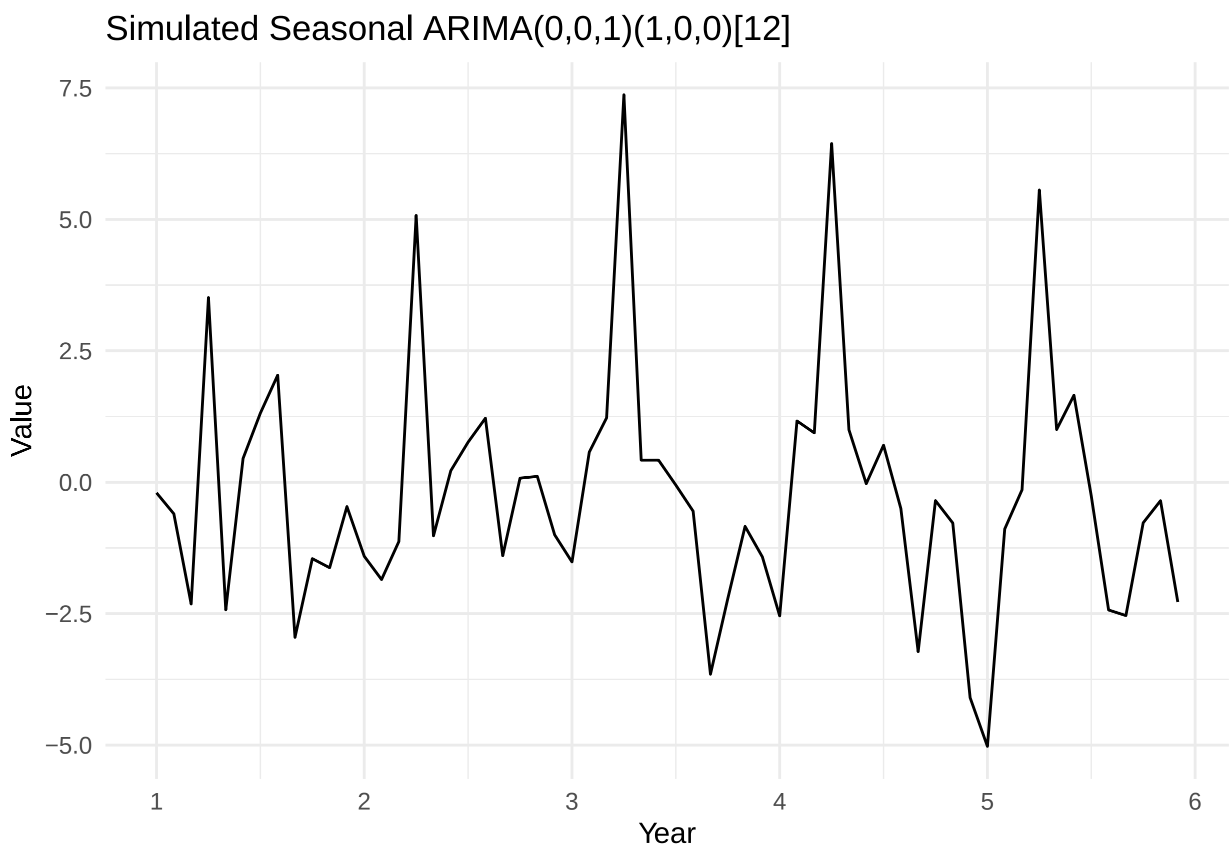
Q4

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?

```r
library(sarima)
```

```
## Warning: package 'sarima' was built under R version 4.5.2

## Loading required package: stats4

##
## Attaching package: 'sarima'

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

```r
model_list <- list(sar = 0.8, ma = 0.5, nseasons = 12)

simulated_data <- sim_sarima(n = 60, model = model_list)

sim_ts <- ts(simulated_data, frequency = 12)

autoplot(sim_ts) +
  ggtitle("Simulated Seasonal ARIMA(0,0,1)(1,0,0)[12]") +
  xlab("Year") +
  ylab("Value") +
  theme_minimal()
```

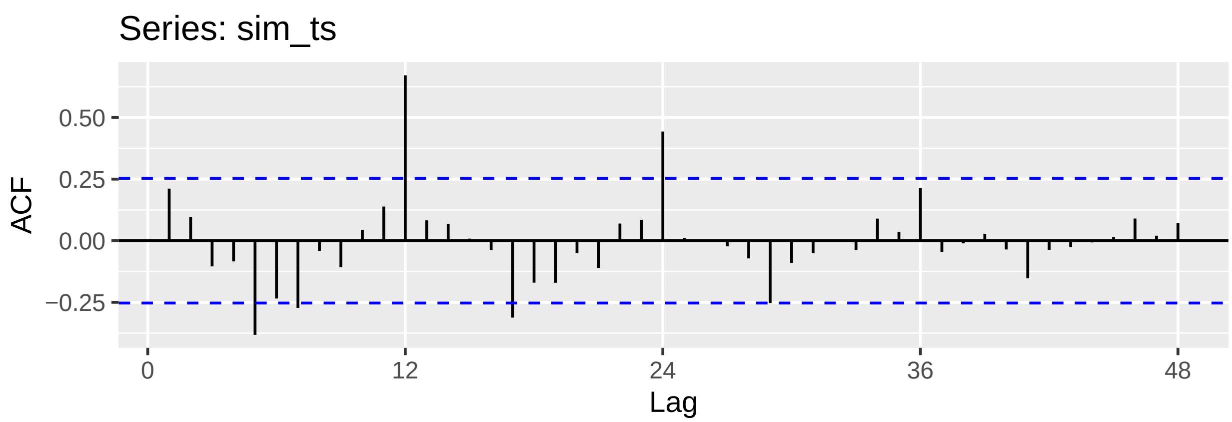


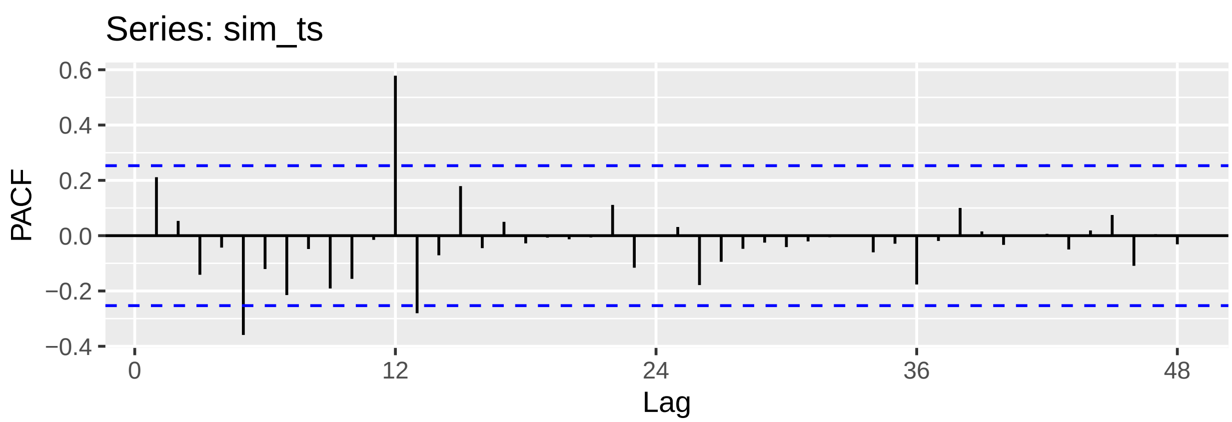
Based on just the visual plot the data shows a clear seasonal trend. The x-axis is a year and each rise a fall is a single season (12 month period)

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.

```
# Using the series from Q4
p_acf <- ggAcf(sim_ts, lag.max = 48)
p_pacf <- ggPacf(sim_ts, lag.max = 48)
plot_grid(p_acf, p_pacf, ncol = 1)
```





Answer: From the plot the seasonal component is well identified as we see the spikes at multiples of 12 in the ACF plot. The MA(1) component is seen in the spike at lag 1 with an observed parameter value of 0.4 which is close to the model. The plots well represent the simulated data as both trends can be ascertained by looking at them.