ENV 797 - Time Series Analysis for Energy and Environment Applications | Spring 2025

Assignment 6 - Due date 02/27/25

Chloe Young

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_Sp25.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
library(lubridate)
library(ggplot2)
library(forecast)
library(Kendall)
library(tseries)
library(outliers)
library(tidyverse)
library(cowplot)
library(sarima)
library(openxlsx)
library(readxl)
library(dplyr)
```

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 will show an exponential decay over time.The PACF plot will show two significant spikes at lags 1 and 2, and after that the lags will drop off to be insignificant for the rest of the plot.

- MA(1)

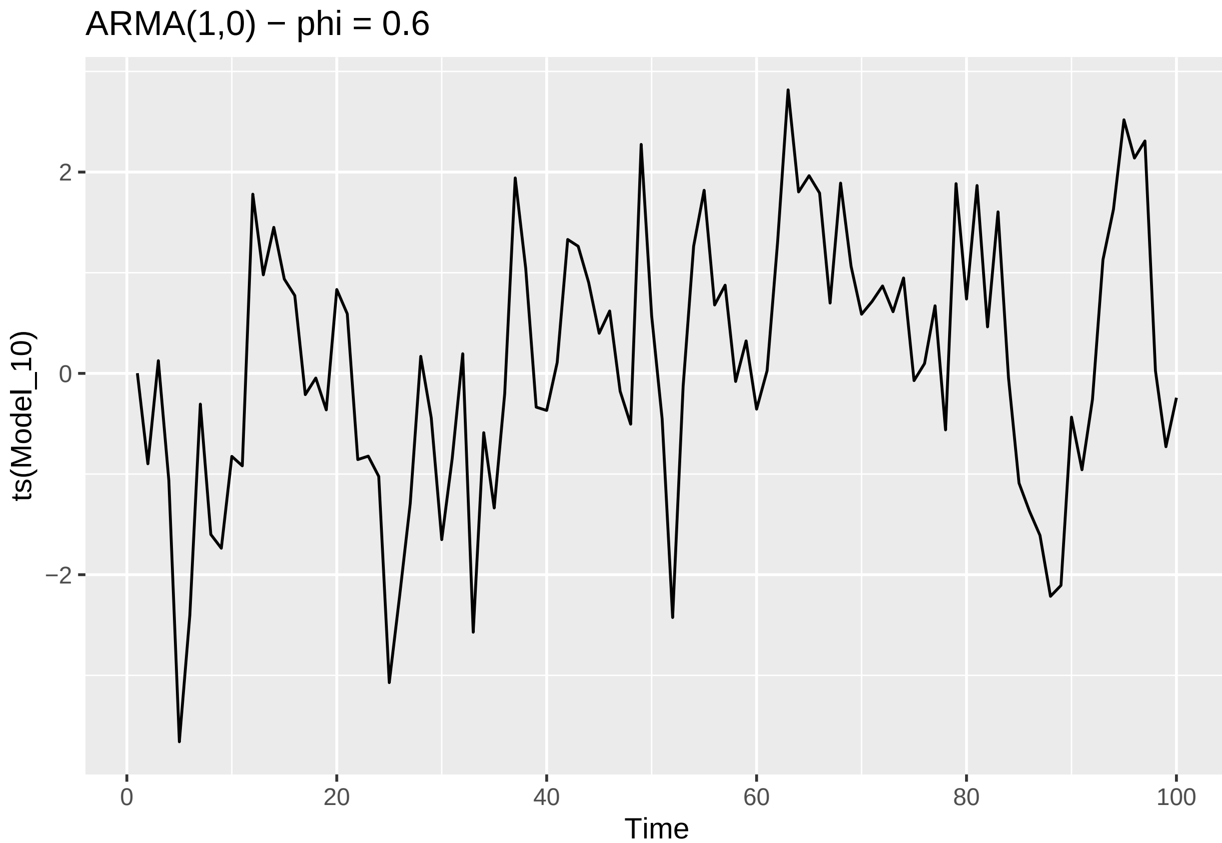
  Answer: The ACF plot will have a spike at lag 1 and then following that the spikes will be insignificant, so following lag 1 there is a cutoff. There is only one significant spike since the order is 1. The PACF plot will show an exponential decay over time.

Q2

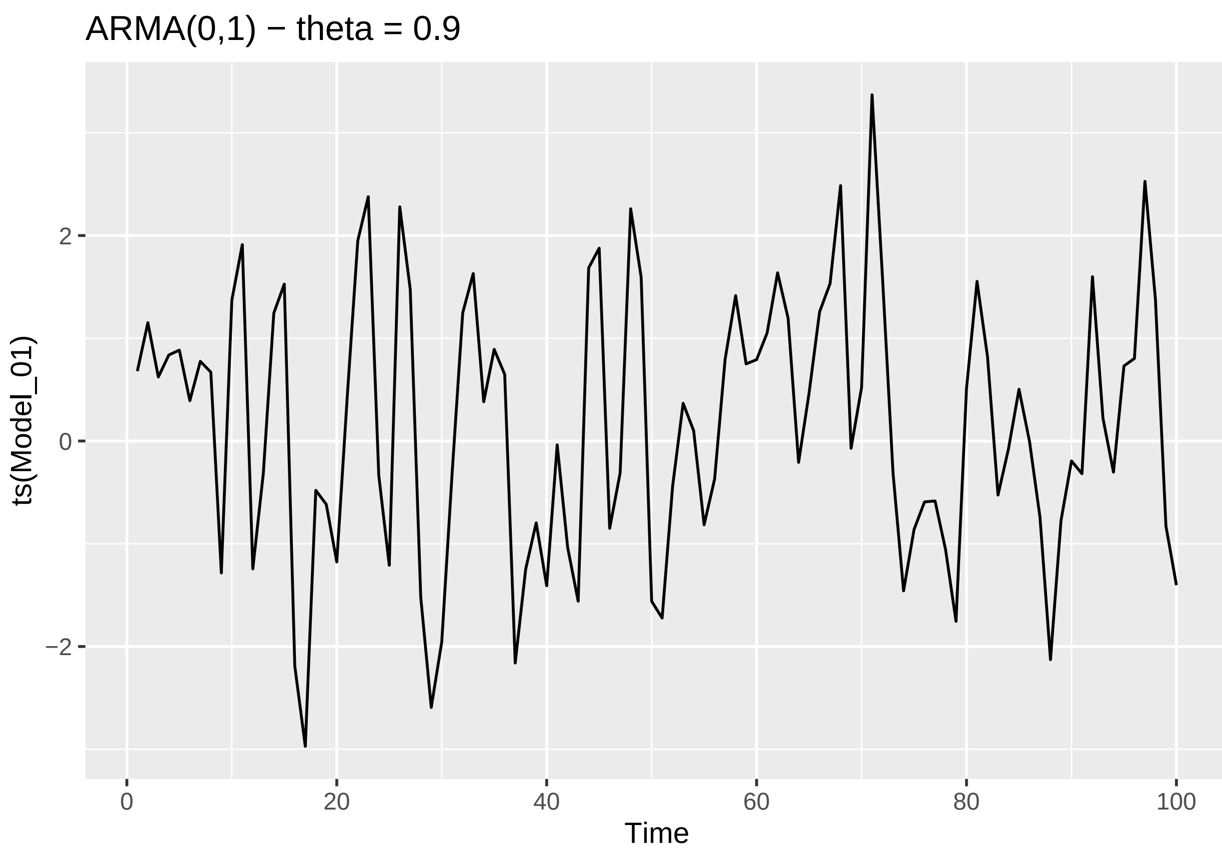
Recall that the non-seasonal ARIMA is described by three parameters ARIMA(p, d, q) where p is the order of the autoregressive component, d is the number of times the series need to be differenced to obtain stationarity and q is the order of the moving average component. If we don't need to difference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

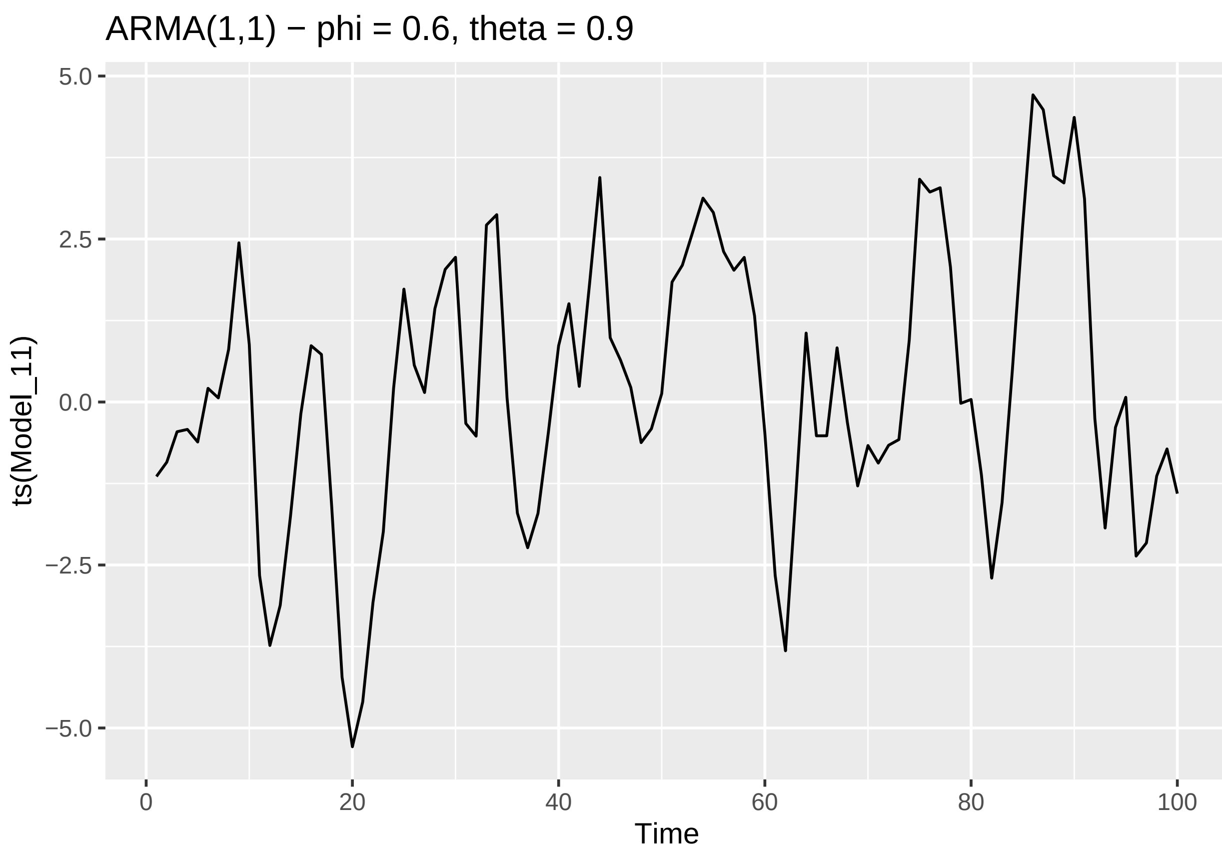
```
n <- 100

Model_10 <- arima.sim(n = n,
                      model = list(ar = 0.6))

Model_01 <- arima.sim(n = n,
                      model = list(ma = 0.9))

Model_11 <- arima.sim(n = n,
                      model = list(ar = 0.6, ma = 0.9))

autoplot(ts(Model_10)) + ggtitle("ARMA(1,0) - phi = 0.6")
```




```
autoplot(ts(Model_01)) + ggtitle("ARMA(0,1) - theta = 0.9")
```

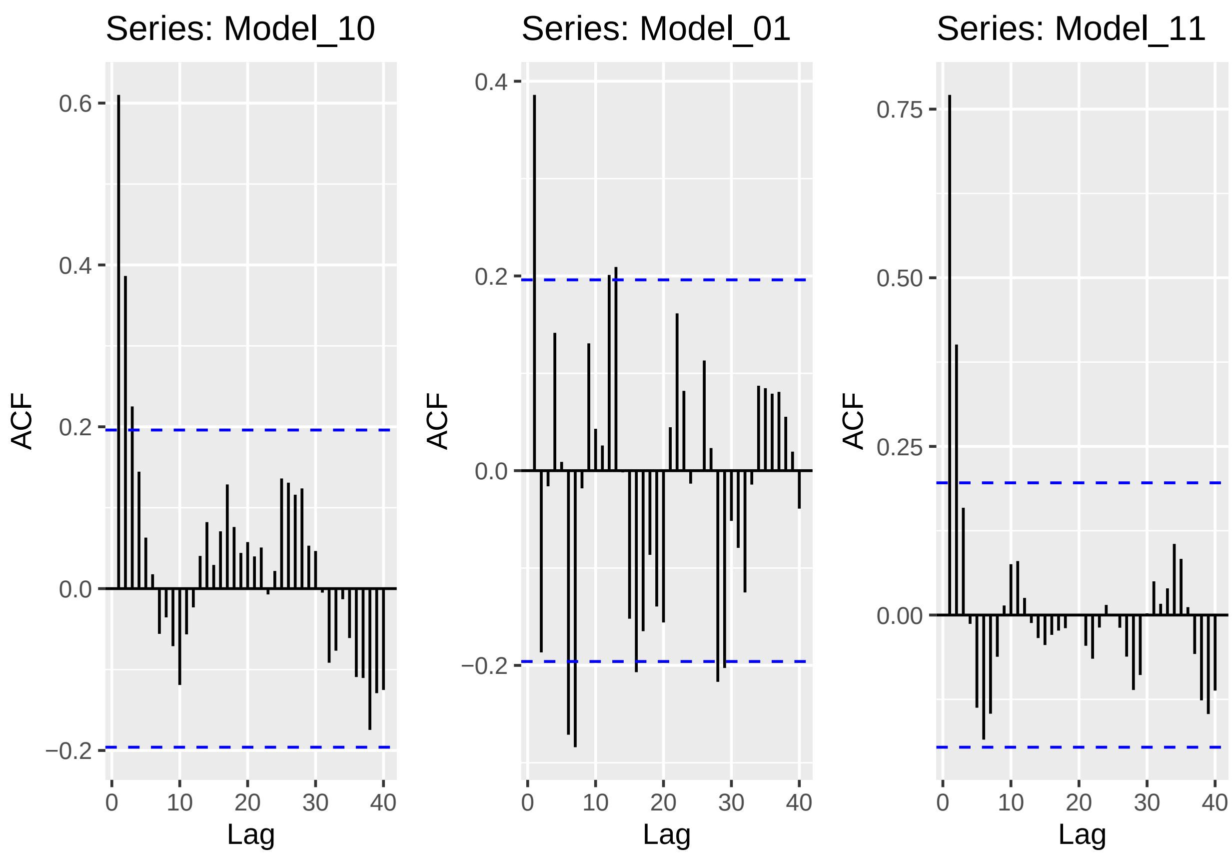



```
autoplot(ts(Model_11)) + ggtitle("ARMA(1,1) - phi = 0.6, theta = 0.9")
```



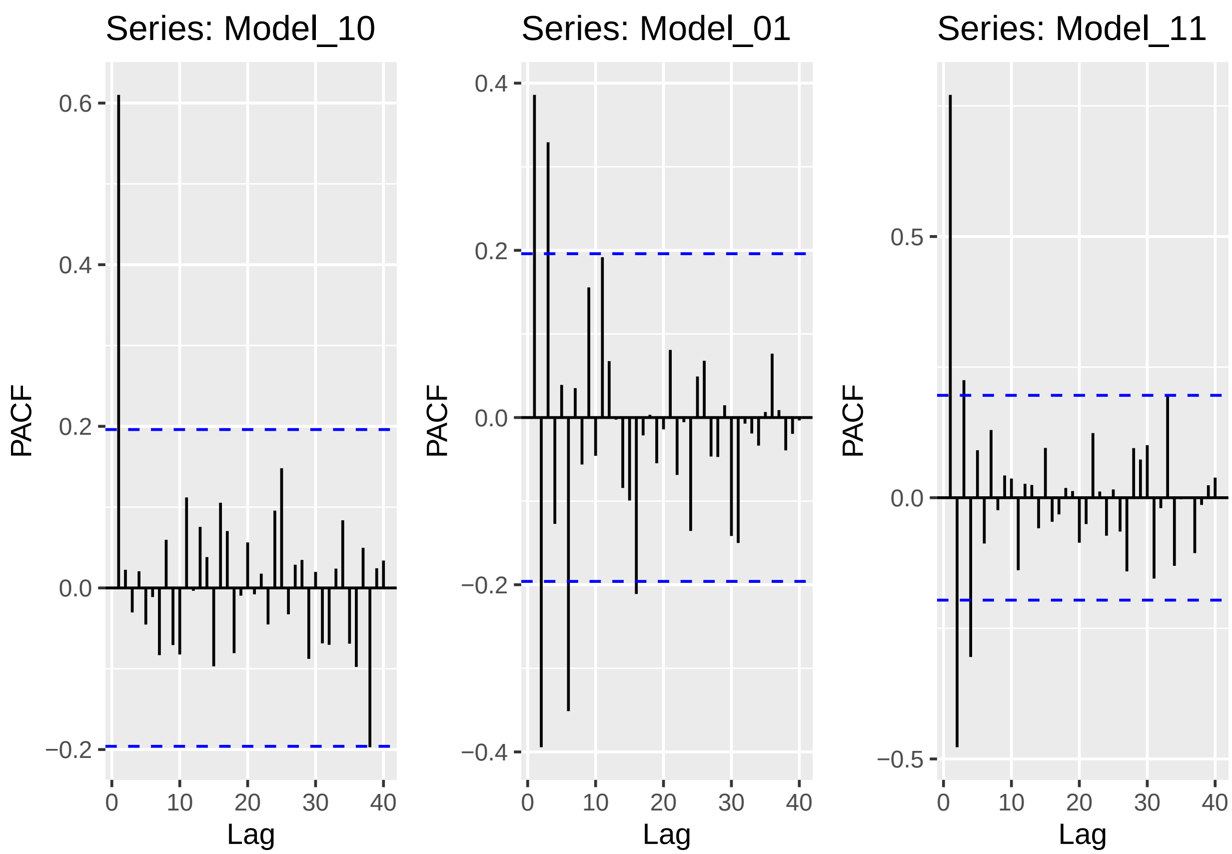
(b) Plot the sample ACF for each of these models in one window to facilitate comparison (Hint: use `cowplot::plot_grid()`).

```
plot_grid(
  autoplot(Acf(Model_10,lag.max=40, plot = FALSE)),
  autoplot(Acf(Model_01,lag.max=40, plot = FALSE)),
  autoplot(Acf(Model_11,lag.max=40, plot = FALSE)),
  nrow=1
)
```



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

```
plot_grid(
  autoplot(Pacf(Model_10,lag.max=40,plot=FALSE)),
  autoplot(Pacf(Model_01,lag.max=40,plot=FALSE)),
  autoplot(Pacf(Model_11,lag.max=40,plot=FALSE)),
  nrow=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:

- Model (1,0): you could potentially identify correctly that it's AR because there is somewhat of a gradual decay, but it isn't super clear. This is similar to the PACF since there is a second significant lag at around 12, so it isn't necessarily a sharp cutoff. Nonetheless, you could still see the order 1 since there isn't any significant lag next to lag 1. Therefore, for 1,0, it would be possible but perhaps difficult and could be confused for an ARMA model.
- Model (0,1): this model is much easier to identify. The ACF plot shows a cutoff after lag 1 with no more significant lags, and the PACF displays a gradual decay, and both of these point to a MA model. The order would be 1 since there is only 1 significant lag in the ACF plot.
- Model (1,1): This model displays a gradual decay in both the ACF plot and the PACF plot, and this is indicative of an ARMA model since it shows both AR and MA characteristics. Furthermore, since both the ACF and PACF plots show 3 significant spikes before they cutoff to become insignificant, the order can be identified as (3,3).

(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?

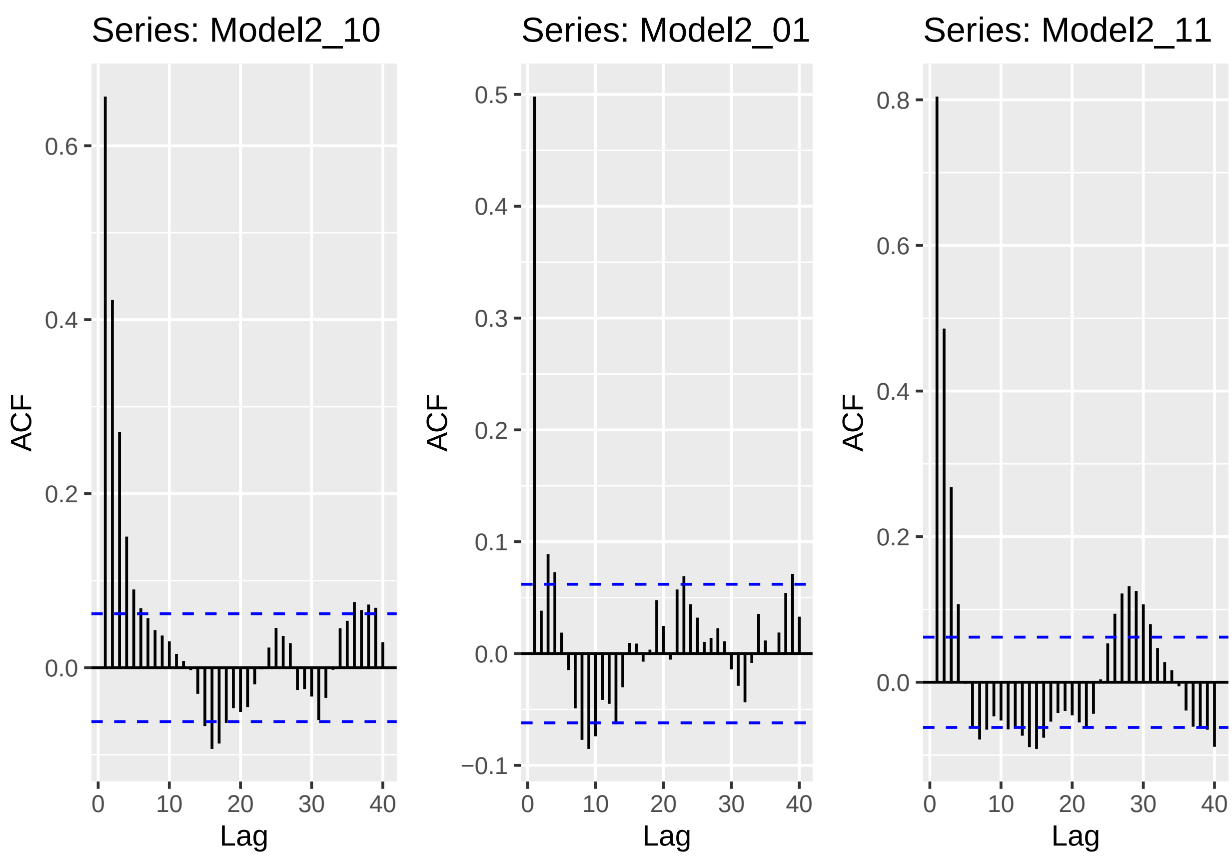
Answer:

For the ARMA (1,0) model, $\phi = 0.6$ doesn't match what is seen on the PACF, as the spike at lag 1 looks like it is just above 0.5. In this case, the the value of lag 1 should match and be 0.6, even though it doesn't.

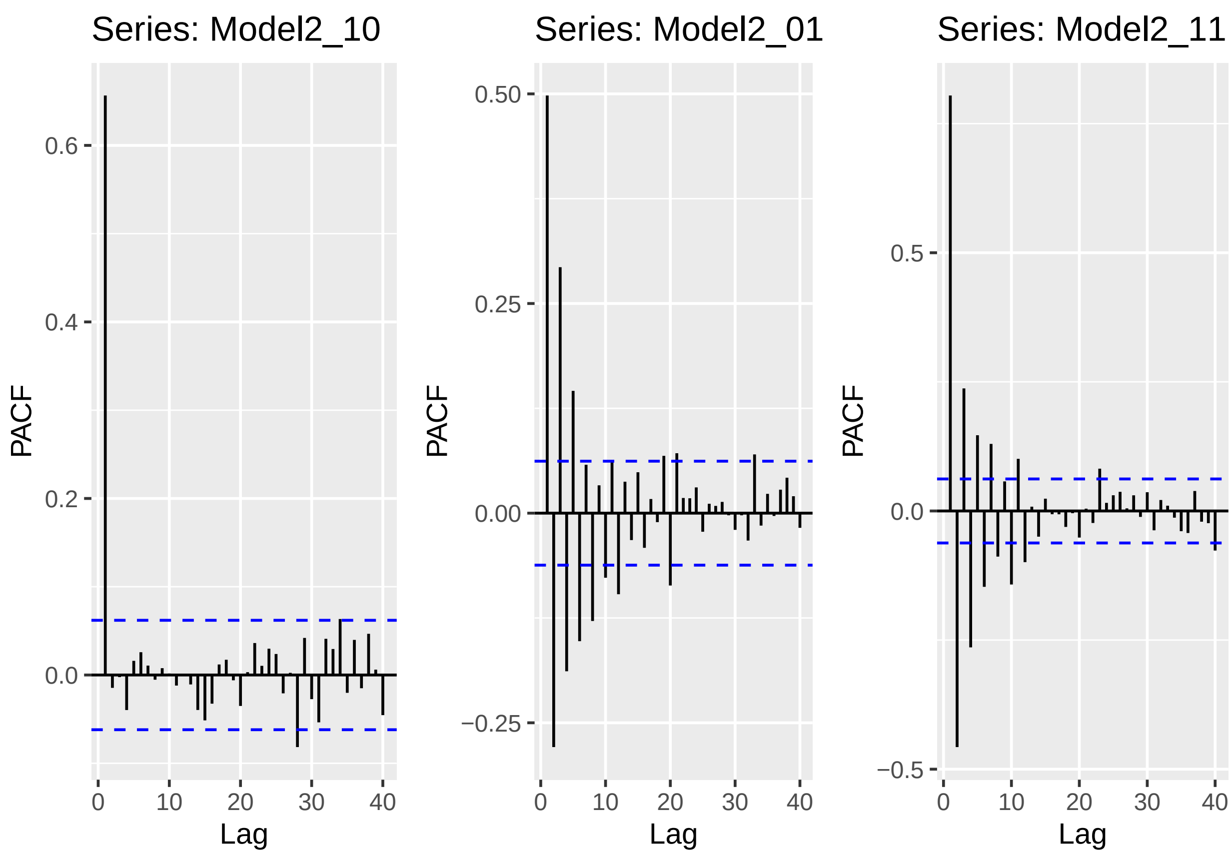
For the ARMA (1,1) model,$\phi = 0.6$ doesn't match what is seen on the PACF, but it shouldn't because of the MA component of this model. The MA component affects the correlations so we wouldn't expect the value at lag 1 on the PACF to equal 0.6.

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

```
n <- 1000

Model2_10 <- arima.sim(n = n,
                       model = list(ar = 0.6))
Model2_01 <- arima.sim(n = n,
                       model = list(ma = 0.9))
Model2_11 <- arima.sim(n = n,
                       model = list(ar = 0.6, ma = 0.9))

plot_grid(
  autoplot(Acf(Model2_10,lag.max=40, plot = FALSE)),
  autoplot(Acf(Model2_01,lag.max=40, plot = FALSE)),
  autoplot(Acf(Model2_11,lag.max=40, plot = FALSE)),
  nrow=1
)
```



```
plot_grid(
  autoplot(Pacf(Model2_10,lag.max=40,plot=FALSE)),
  autoplot(Pacf(Model2_01,lag.max=40,plot=FALSE)),
  autoplot(Pacf(Model2_11,lag.max=40,plot=FALSE)),
  nrow=1
)
```



Repeat d:

- Model (1,0): With n=1000, it is significant easier to identift that this model is AR since the ACF shows a clear gradual decay and the PACF shows a sharp cutoff after lag 1. This indicates that this is an AR model and that the order is 1.
- Model (0,1): The ACF plot shows a cutoff after lag 1 with no more significant lags, and the PACF displays a gradual decay, and both of these point to a MA model. The order would be 1 since there is only 1 significant lag in the ACF plot.
- Model (1,1): This model displays a gradual decay in both the ACF plot and the PACF plot, and this is indicative of an ARMA model since it shows both AR and MA characteristics. However, identifying the order is much more difficult.

Repeat e:

For the ARMA (1,0) model with n=1000, $\phi = 0.6$ doesn't matches seen on the PACF, as the spike at lag 1 looks is right around 0.6. This makes sense since the PACF value at lag 1 is meant to match the given value of phi.

For the ARMA (1,1) model,$\phi = 0.6$ almost matches what is seen on the PACF, but we wouldn't expect it to because of the MA component of this model. The MA component affects the correlations so we wouldn't expect the value at lag 1 on the PACF to equal 0.6, although in this case, it does. This could indicate that the AR component of the model is stronger which is shown with n=1000 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.

p= 1 d= 0 q= 1 P= 1 D= 0 Q= 0 s= 12

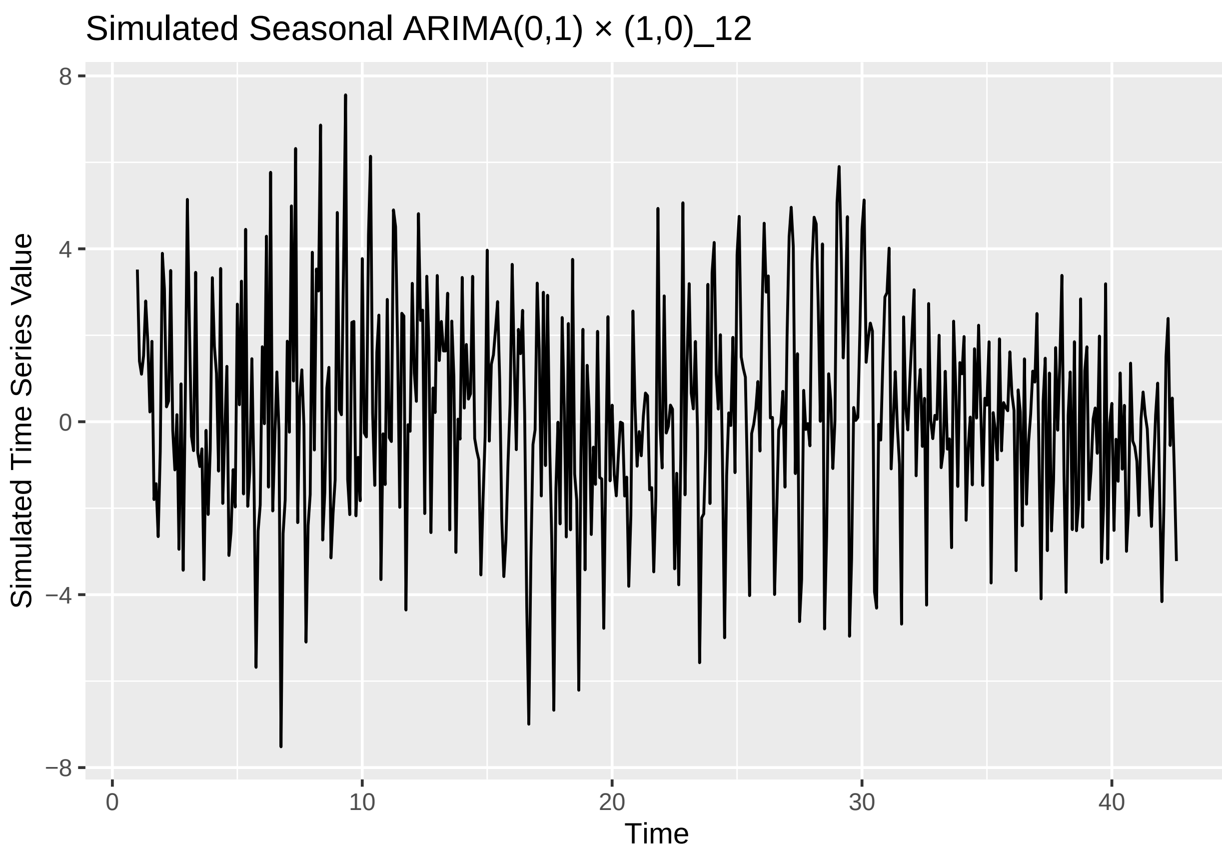
(b) Also from the equation what are the values of the parameters, i.e., model coefficients.

$\phi = 0.7$ $\phi_1 = -0.25$ $\theta_1 = -0.1$

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
n <- 500

simulated_data <- sim_sarima(n = n,
                             model = list(sar = 0.8,
                                          sma = 0.5,
                                          nseasons = 12))

ts_simulatedseasonal <- ts(simulated_data, frequency = 12)

autoplot(ts_simulatedseasonal) +
  ggtitle("Simulated Seasonal ARIMA(0,1) × (1,0)_12") +
  xlab("Time") +
  ylab("Simulated Time Series Value")
```

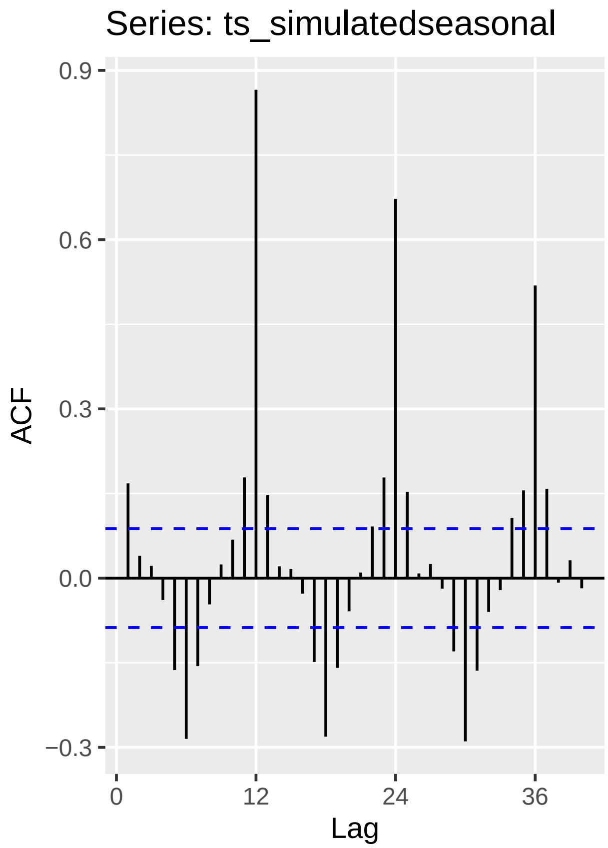


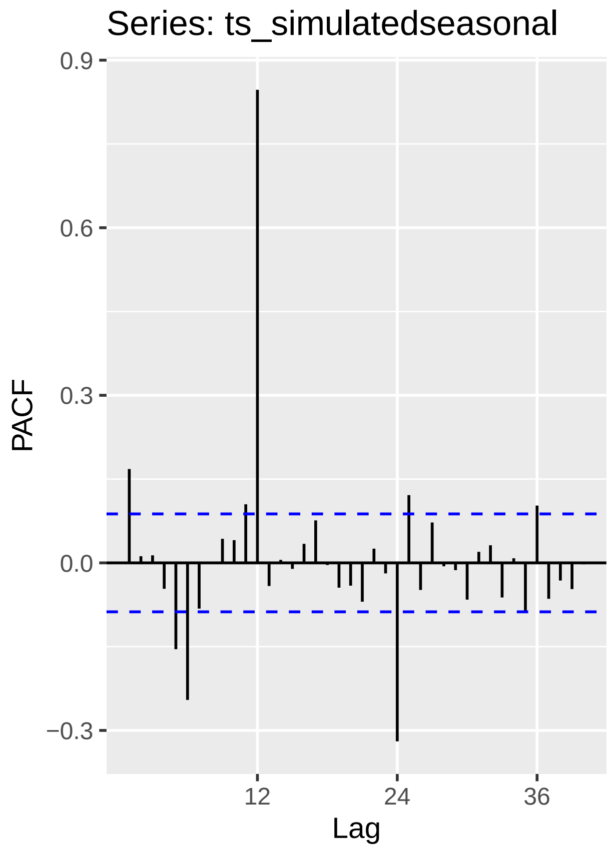
Visually, the plot doesn't look seasonal as there is no clear repeating 12 month pattern. There is some repetitive up down pattern, but the seasonality is unclear just from the model.

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.

```
plot_grid(
  autoplot(Acf(ts_simulatedseasonal,lag.max=40, plot = FALSE)),
  autoplot(Pacf(ts_simulatedseasonal,lag.max=40,plot=FALSE))
)
```





The plot represents the seasonal component of the model well. The positive spikes in the ACF at lags 12, 24, and 36 and the single positive spike in the PACF at lag 12 are indicative of P=1. (P+Q) should never be more than 1, so there is no MA component in this model since Q=0 in the model and P=1. This shows that the seasonal AR(1) aspect of the model can be identified from the ACF and PACF. For the nonseasonal component, the model is (0,1), so p=0 and q=1. This is harder to identify using the ACF and PACF. For the MA component, the ACF should identify the order and the PACF should show a gradual decay, which is slightly unclear in these models.Therefore, the plots don't represent the non-seasonal component well.