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

Assignment 6 - Due date 02/27/25

Rachael Stephan

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
#Load/install required package here
library(ggplot2); library(forecast); library(tseries)
library(sarima); 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)

  Autoregressive (AR) models determine future observations through the relationship between an observation and its previous observations. The ACF plot will decay exponentially with time for AR models. The PACF plots for AR models have a sharp decrease in correlation after the lag that determines the order. Therefore, for the AR(2) model, the ACF will decay exponentially and the PACF will have a sharp decrease to insignificant correlations after the second lag.

- MA(1)

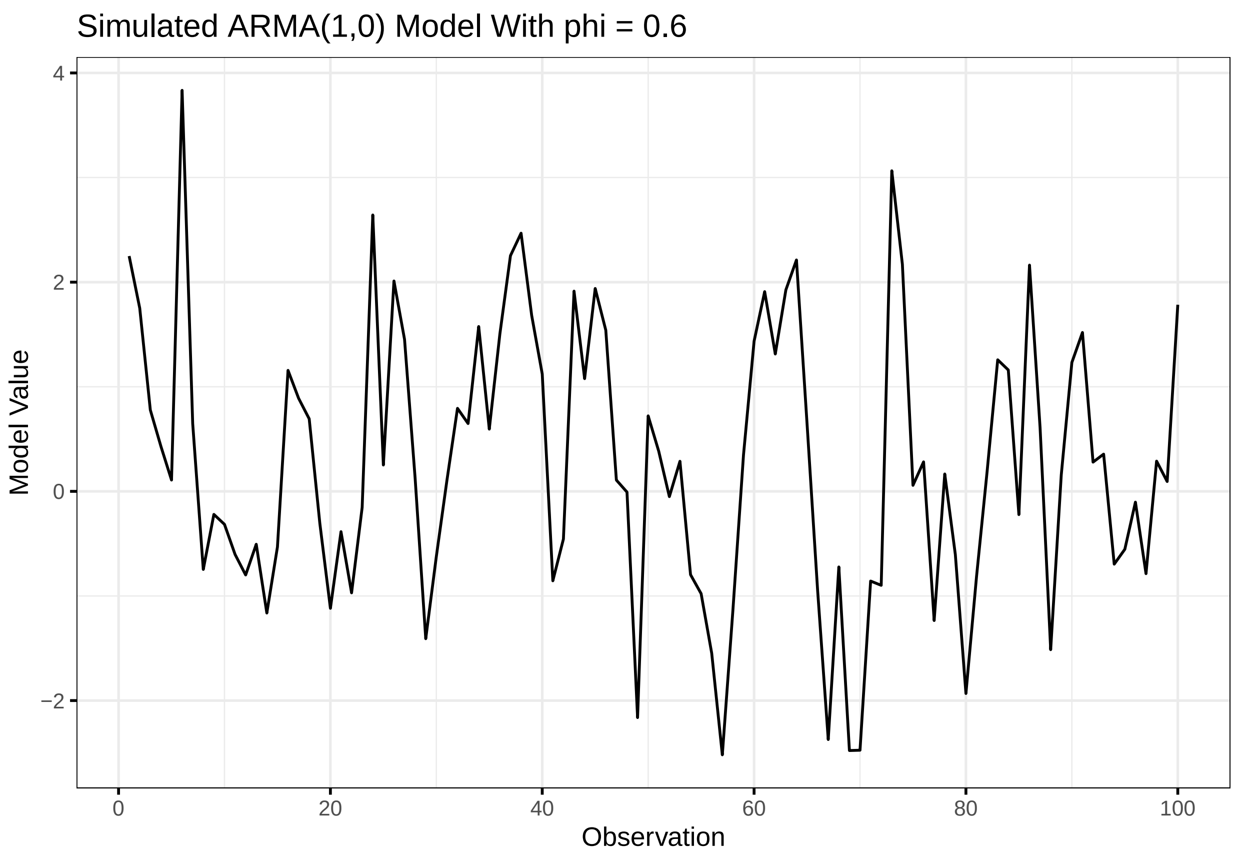
Moving average (MA) models determine future observations through the relationship between an observation residual and the residuals from a moving average. The PACF plot will decay exponentially with time for MA models. The ACF plots for MA models have a sharp decrease in correlation after the lag that determines the order. Therefore, for the MA(1) model, the PACF will decay exponentially and the ACF will have a sharp decrease to insignificant correlations after the first lag. MA models often have a negative correlation within the first few lags in both PACF and ACF plots. This may be present as well.

Q2

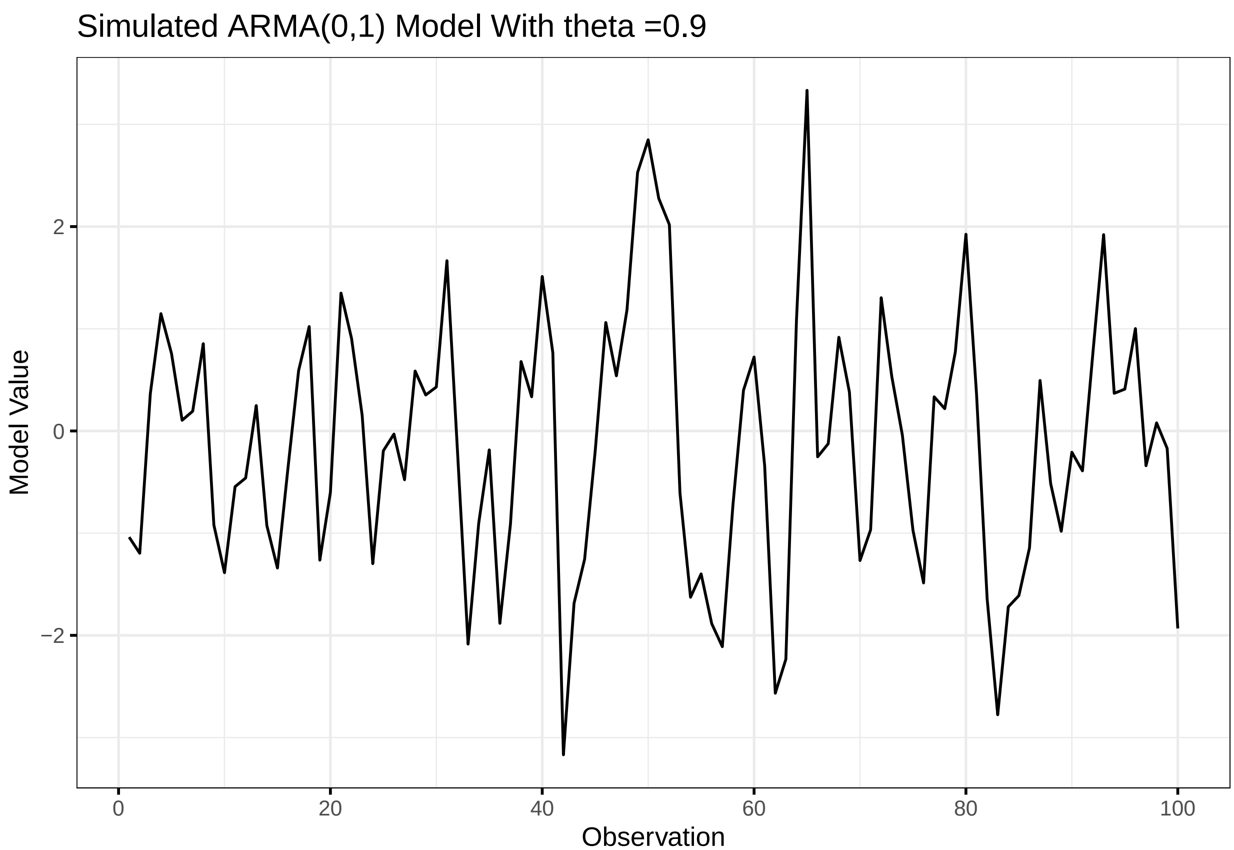
Recall that the non-seasonal ARIMA is described by three parameters ARIMA(p, d, q) where p is the order of the autoregressive component, d is the number of times the series need to be differenced to obtain stationarity and q is the order of the moving average component. If we don't need to difference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

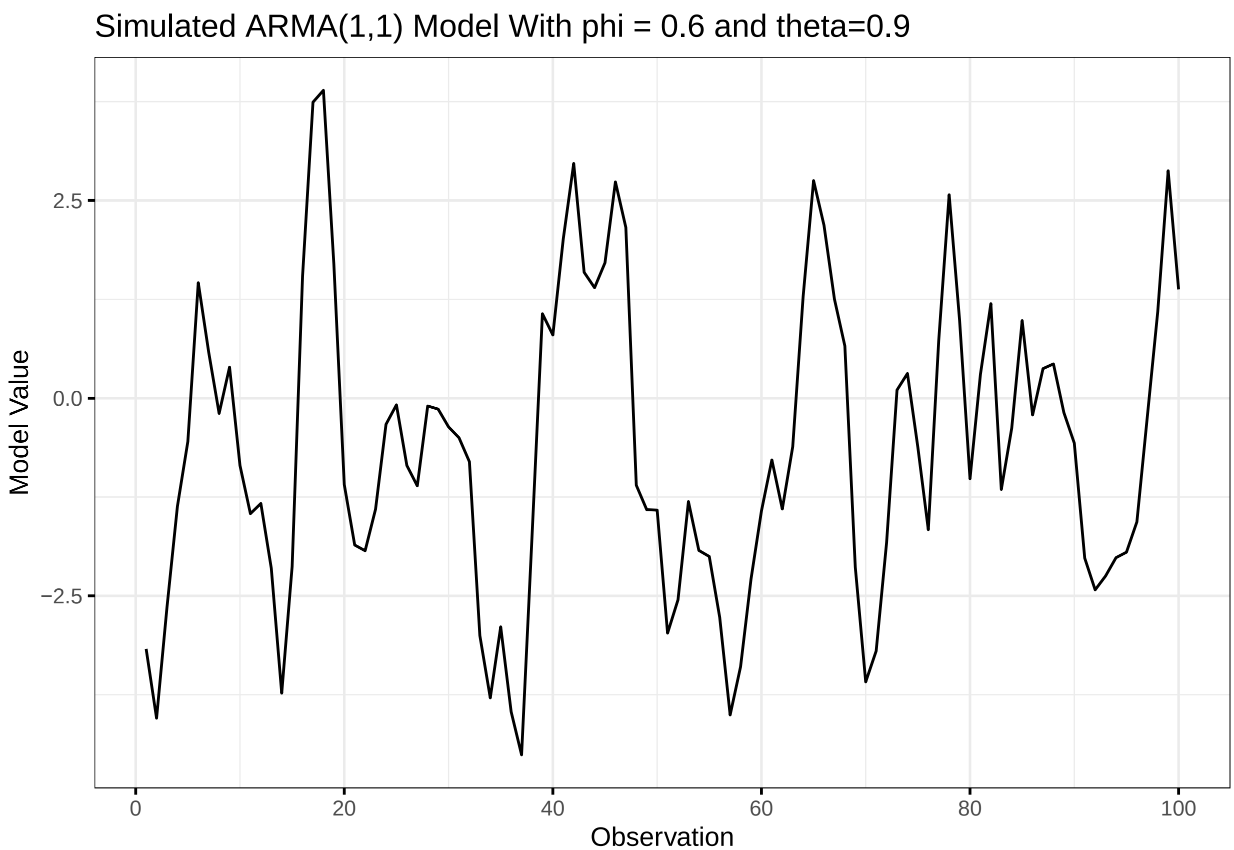
```r
#set seed for reproduceability
set.seed(45)

#create models
arma_10 <- arima.sim(model = list(order = c(1,0,0), ar = 0.6),
                     n = 100)

arma_01 <- arima.sim(model = list(order = c(0,0,1), ma = 0.9),
                     n = 100)

arma_11 <- arima.sim(model = list(order = c(1,0,1), ar = 0.6, ma = 0.9),
                     n = 100)

#plot models
autoplot(arma_10)+
  labs(x = "Observation",
       y = "Model Value",
       title = paste0("Simulated ARMA(1,0) Model With phi = 0.6"))
```




```r
autoplot(arma_01)+
  labs(x = "Observation",
       y = "Model Value",
       title = "Simulated ARMA(0,1) Model With theta =0.9")
```

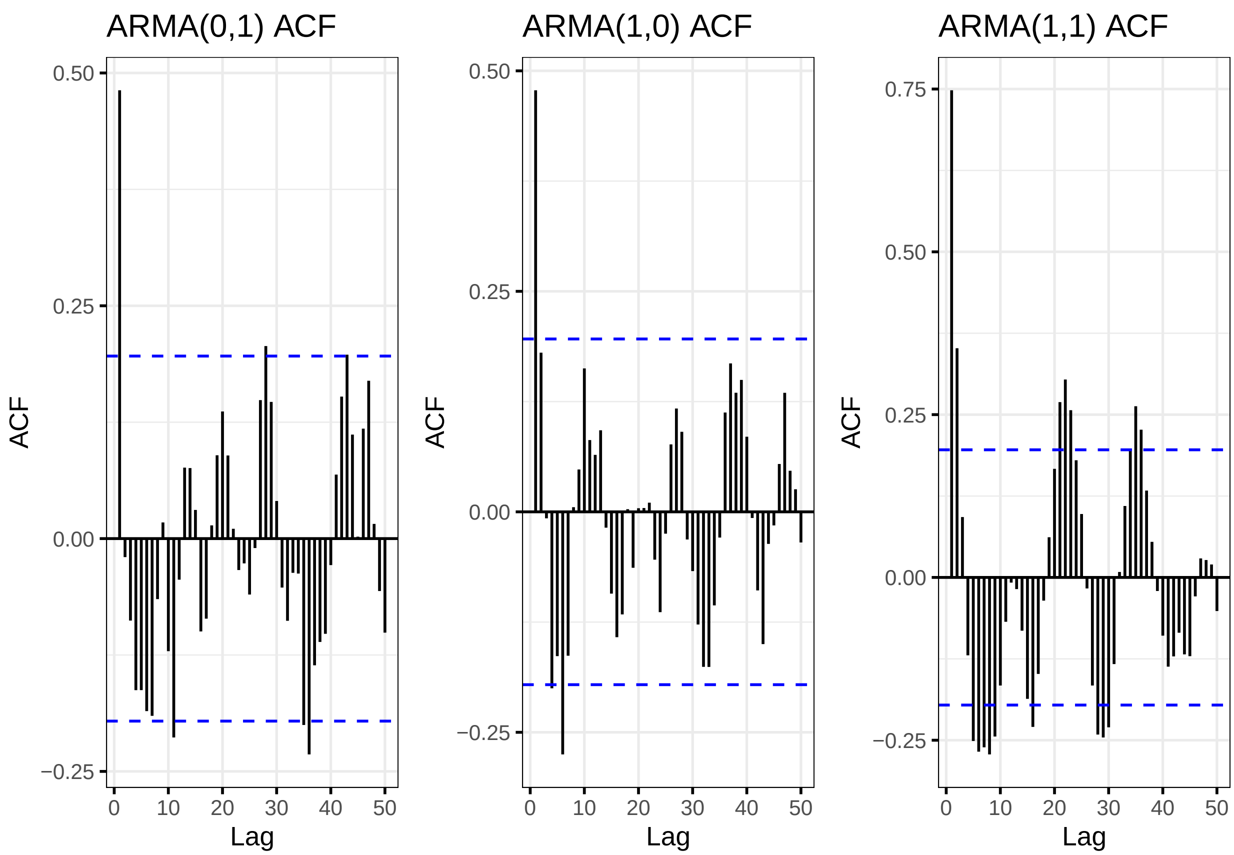



```r
autoplot(arma_11)+
  labs(x = "Observation",
       y = "Model Value",
       title = "Simulated ARMA(1,1) Model With phi = 0.6 and theta=0.9")
```



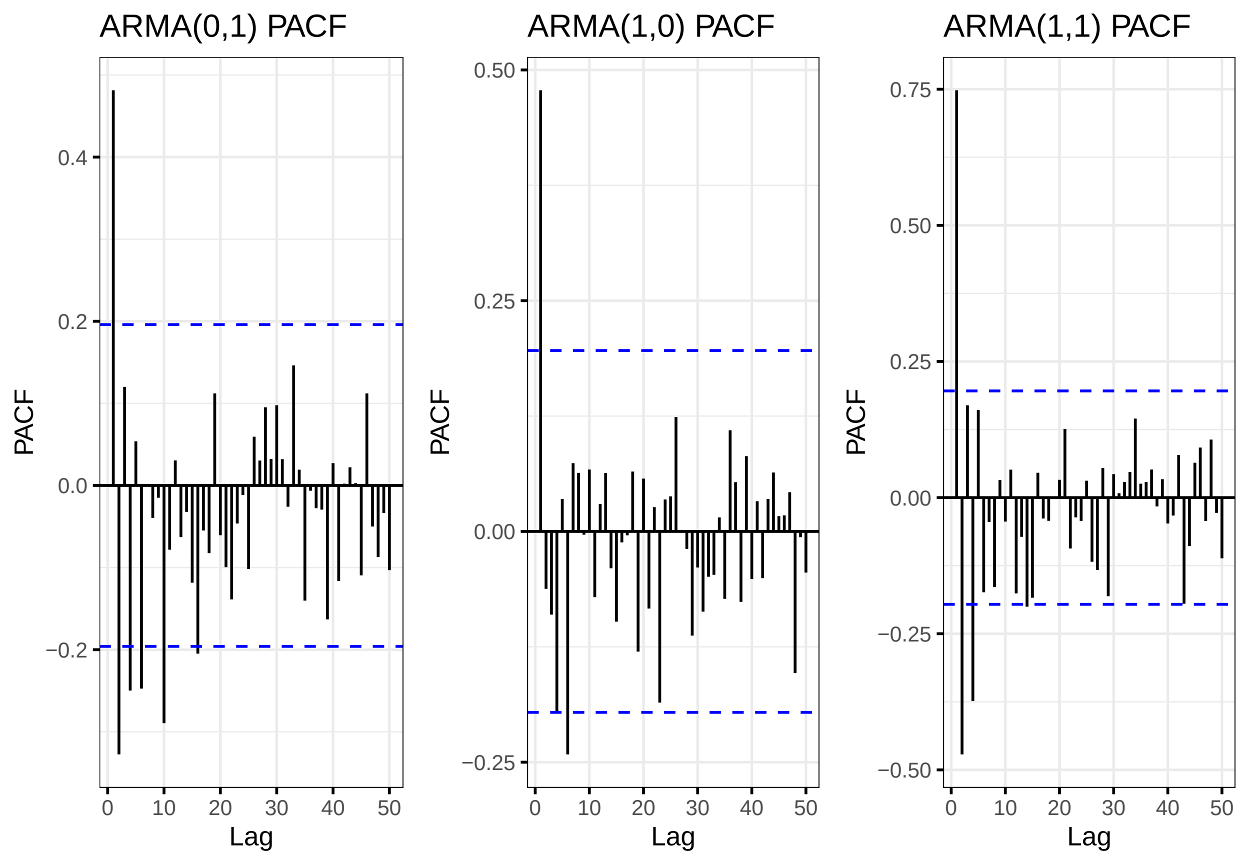
(b) Plot the sample ACF for each of these models in one window to facilitate comparison (Hint: use `cowplot::plot_grid()`).

```
#create ACF plots
acf_arma_10 <- Acf(arma_10,
                   lag = 50,
                   plot=FALSE)

acf_plot_arma_10 <- autoplot(acf_arma_10, main = "ARMA(1,0) ACF")

acf_arma_01 <- Acf(arma_01,
                   lag = 50,
                   plot=FALSE)

acf_plot_arma_01 <- autoplot(acf_arma_01, main = "ARMA(0,1) ACF")

acf_arma_11 <- Acf(arma_11,
                   lag = 50,
                   plot=FALSE)

acf_plot_arma_11 <- autoplot(acf_arma_11, main = "ARMA(1,1) ACF")

#plot the ACF onto a grid
plot_grid(acf_plot_arma_01, acf_plot_arma_10, acf_plot_arma_11, nrow = 1)
```



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

```
#create PACF plots
pacf_arma_10 <- Pacf(arma_10,
                     lag = 50,
                     plot=FALSE)

pacf_plot_arma_10 <- autoplot(pacf_arma_10, main = "ARMA(1,0) PACF")

pacf_arma_01 <- Pacf(arma_01,
                     lag = 50,
                     plot=FALSE)

pacf_plot_arma_01 <- autoplot(pacf_arma_01, main = "ARMA(0,1) PACF")

pacf_arma_11 <- Pacf(arma_11,
                     lag = 50,
                     plot=FALSE)

pacf_plot_arma_11 <- autoplot(pacf_arma_11, main = "ARMA(1,1) PACF")

#plot the PACF onto a grid
plot_grid(pacf_plot_arma_01, pacf_plot_arma_10, pacf_plot_arma_11, 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.

- *ARMA(0,1) model:* I would likely be able to identify this model correctly as a MA model. The ACF has a relatively clear drop-off after the first lag, and the PACF decays gradually, with half of the first ten lags being significant.
- *ARMA(1,0) model:* I may or may not be able to correctly identify this model as an AR model. The ACF model is somewhat ambiguous as to whether it has a cut-off or a decay. However, lag 1 is positive, which is more typical of AR models. The PACF model does look more like a cut-off, but this is again somewhat ambiguous. The ambiguity might throw off my answers.
- *ARMA(1,1) model:* I would possibly be able to identify this model as an ARMA model. There appears to be some element of decay in both the ACF and PACF plots. However, the cut-off in the ACF plot is less pronounced than in the PACF plot.

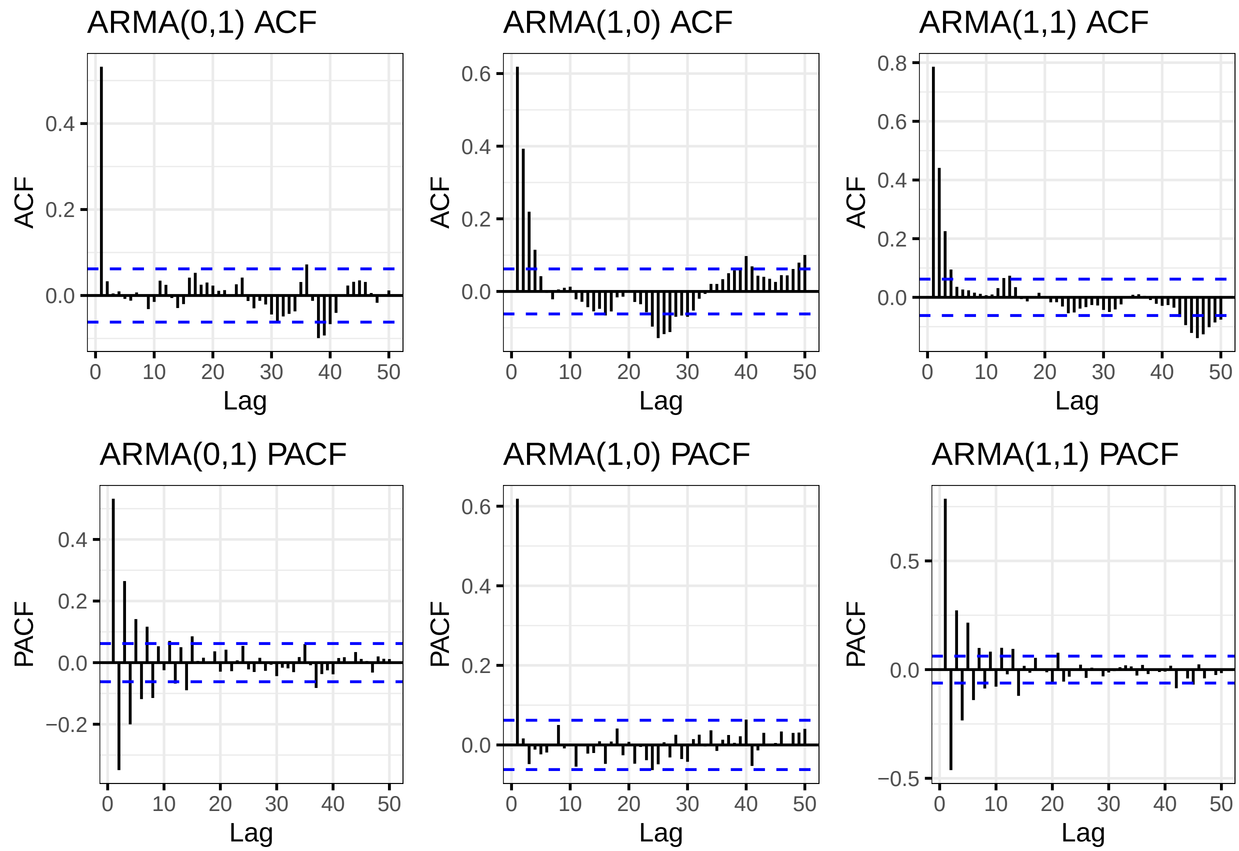
(e) Compare the PACF values R computed with the values you provided for the lag 1 correlation coefficient, i.e., does $\phi = 0.6$ match what you see on PACF for ARMA(1,0), and ARMA(1,1)? Should they match?

The lag 1 correlation coefficient on the PACF for the ARMA(1,0) model is 0.4779534. This is not close to $\phi = 0.6$. They should match but they do not, potentially because of the number of observations. The lag 1 correlation coefficient on the PACF for the ARMA(1,1) model is 0.7480868. This does not match $\phi = 0.6$. However, this is expected because the θ coefficient is also playing a part in the ARMA(1,1) model.

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

```r
#set seed for reproduceability
set.seed(45)

#create models
arma2_10 <- arima.sim(model = list(order = c(1,0,0), ar = 0.6),
                      n = 1000)

arma2_01 <- arima.sim(model = list(order = c(0,0,1), ma = 0.9),
                      n = 1000)

arma2_11 <- arima.sim(model = list(order = c(1,0,1), ar = 0.6, ma = 0.9),
                      n = 1000)

#create ACF plots
acf_arma2_10 <- Acf(arma2_10,
                    lag = 50,
                    plot=FALSE)

acf_plot_arma2_10 <- autoplot(acf_arma2_10, main = "ARMA(1,0) ACF")

acf_arma2_01 <- Acf(arma2_01,
                    lag = 50,
                    plot=FALSE)

acf_plot_arma2_01 <- autoplot(acf_arma2_01, main = "ARMA(0,1) ACF")

acf_arma2_11 <- Acf(arma2_11,
                    lag = 50,
                    plot=FALSE)

acf_plot_arma2_11 <- autoplot(acf_arma2_11, main = "ARMA(1,1) ACF")

#create PACF plots
pacf_arma2_10 <- Pacf(arma2_10,
                    lag = 50,
                    plot=FALSE)

pacf_plot_arma2_10 <- autoplot(pacf_arma2_10, main = "ARMA(1,0) PACF")

pacf_arma2_01 <- Pacf(arma2_01,
                    lag = 50,
                    plot=FALSE)

pacf_plot_arma2_01 <- autoplot(pacf_arma2_01, main = "ARMA(0,1) PACF")

pacf_arma2_11 <- Pacf(arma2_11,
                    lag = 50,
                    plot=FALSE)

pacf_plot_arma2_11 <- autoplot(pacf_arma2_11, main = "ARMA(1,1) PACF")

#plot the ACF/PACF onto a grid
plot_grid(acf_plot_arma2_01, acf_plot_arma2_10, acf_plot_arma2_11,
```

```r
    pacf_plot_arma2_01, pacf_plot_arma2_10, pacf_plot_arma2_11,
    nrow = 2)
```



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.

- *ARMA(0,1) model:* I would be able to identify this model correctly as a MA model. The ACF has a very clear drop-off after the first lag,and the first lag is negative (which is typical of MA models). The PACF decays gradually and neatly without ambiguity.
- *ARMA(1,0) model:* I would be able to identify this model correctly as an AR model. The PACF has a very clear drop-off after the first lag. The ACF also decays gradually without much ambiguity.
- *ARMA(1,1) model:* I would be able to identify this model as an ARMA model. Both the ACF and PACF show an element of drop-off (0.7860555 to 0.4411995 in the ACF and 0.7860555 to -0.4623812 in the PACF), and both plots show a clear element of decay as well.

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?

> The lag 1 correlation coefficient on the PACF for the ARMA(1,0) model is 0.6187547. This is relatively close to $\phi = 0.6$. They should match, and this correlation coefficient is closer to 0.6 than the previous series. The lag 1 correlation coefficient on the PACF for the ARMA(1,1) model is 0.7860555. This does not match $\phi = 0.6$. However, this is expected because θ is also playing a part in the ARMA(1,1) model.

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.

$ARIMA(1, d, 1)(1, D, 0)_{12}$. There has been some differencing but it is unclear where.

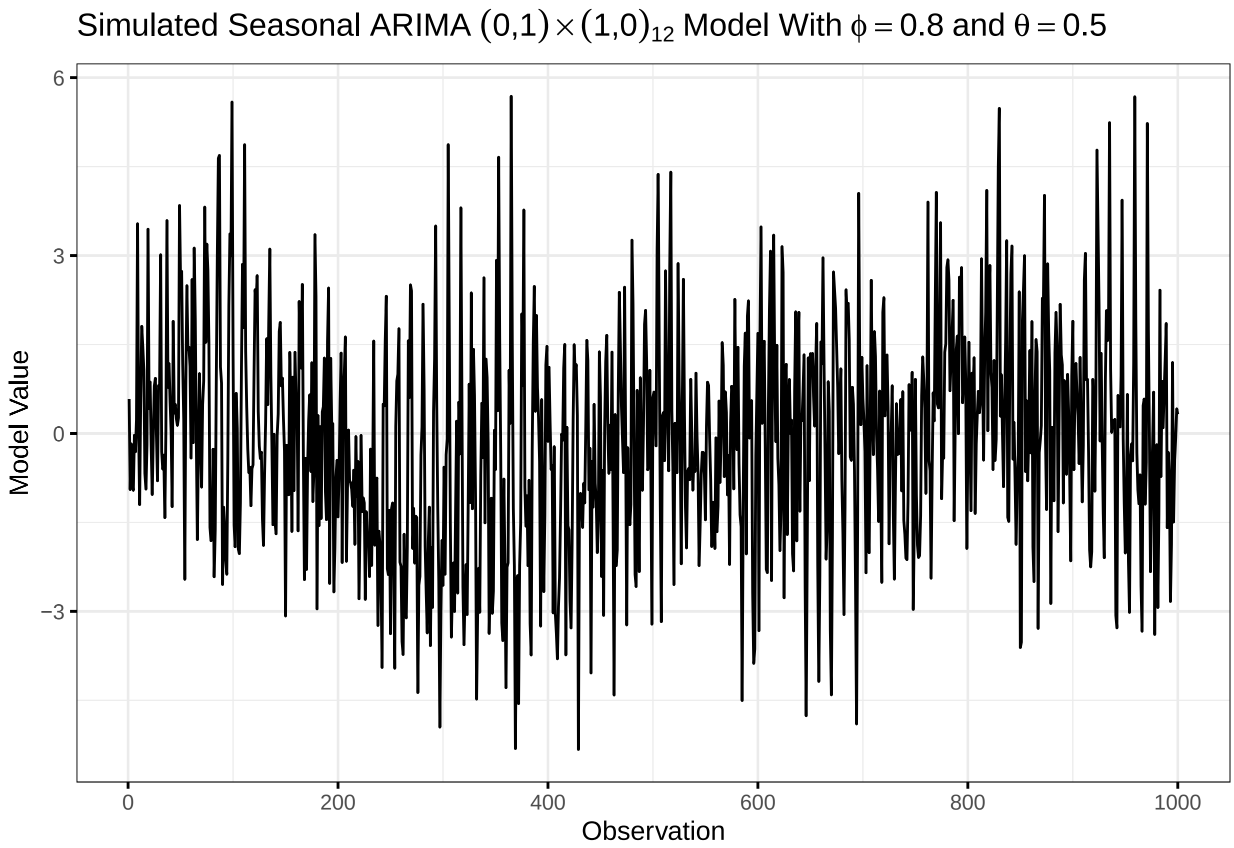
(b) Also from the equation what are the values of the parameters, i.e., model coefficients.

- ϕ_1 = 0.7
- θ_1 = 0.1
- ϕ_{12} = -0.25
- s = 12

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?

```
#set seed for reproduceability
set.seed(45)

#simulate seasonal arima
sarima_model <- sim_sarima(n = 1000,
                           model = list(ma = 0.5,
                                        sar = 0.8,
                                        nseasons = 12)) %>%
  ts()

#plot seasonal model
autoplot(sarima_model)+
  labs(x = "Observation",
       y = "Model Value",
       title = expression(
      "Simulated Seasonal ARIMA" ~ (0 * "," * 1)
      %*% (1 * "," * 0)[12]
      ~ "Model With" ~ phi == 0.8 ~ "and" ~ theta == 0.5))
```

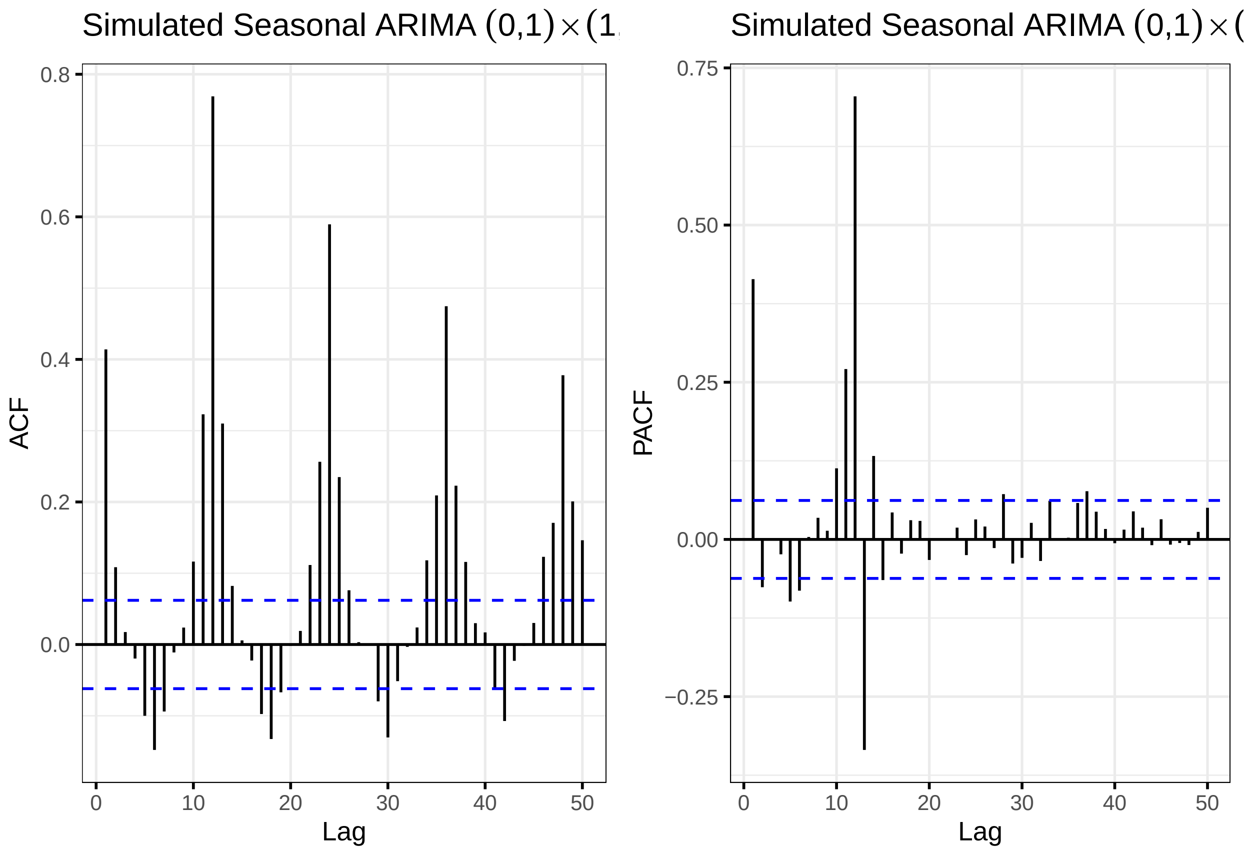


There does seem to be a seasonal component to this ARIMA model. There are semi-regular peaks and valleys in this time series.

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.

```
#create ACF plot
acf_sarima <- Acf(sarima_model,
                  lag = 50,
                  plot=FALSE)

acf_plot_sarima <- autoplot(acf_sarima, main = "Simulated Seasonal ARIMA" ~ (0 * "," * 1) %*% (1 * "," *

#create PACF plot
pacf_sarima <- Pacf(sarima_model,
                   lag = 50,
                   plot=FALSE)

pacf_plot_sarima <- autoplot(pacf_sarima, main = "Simulated Seasonal ARIMA" ~ (0 * "," * 1) %*% (1 * ",

#plot the ACF and PACF onto a grid
plot_grid(acf_plot_sarima, pacf_plot_sarima, nrow = 1)
```



The seasonal component is found in the AR component of the model. There are multiple spikes in the ACF plot of the model and a single large spike in the PACF plot. This is consistent with SAR processes. The spiked lag in the PACF is lag 12. This is consistent with s = 12 in the ARIMA notation. The non-seasonal component is not easily seen in the plots. There may be a drop off between lag 1 and lag 2 in the ACF plot, but this is not as stark compared to the seasonal trends. There is also no easily visible decay in the PACF plot.