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

Assignment 6 - Due date 02/27/25

Aye Nyein Thu

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.

```
#Load/install required package here
library(forecast)
library(cowplot)
library(tseries)
library(ggplot2)
library(Kendall)
library(lubridate)
library(tidyverse)
library(openxlsx)
library(dplyr)
library(sarima) #install.packages("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: The Autoregressive (AR) Model exhibits long memory, meaning it retains information from previous values. In an AR model, the current value has a autocorrelation with previous values. The Autocorrelation Function (ACF) of an AR process typically decays exponentially over time. The Partial Autocorrelation Function (PACF) is used to determine the order (p) of the AR model. If an AR(2) model is appropriate, the PACF will cut off at lag 2 while the ACF will exhibit an exponential decay in general. The AR model equation would be $y_t = c + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \ a_t$ where c and ϕ are constants and a_t is independent and identically distributed (i.i.d).

- MA(1)

  Answer: The Moving Average (MA) Model has a short memory and it depends on past error terms rather than past values of the series itself. The Partial Autocorrelation Function (PACF) of an MA model typically decays gradually over time rather than cutting off. The Autocorrelation Function (ACF) is used to determine the order (q) of the MA model. If an MA(1) model is appropriate, the ACF will cut off at lag 1, while the PACF will overall exhibit an exponential decay. The MA model will look like $y_t = \mu + a_t - \theta a_{t-1}$ where μ is the process mean.

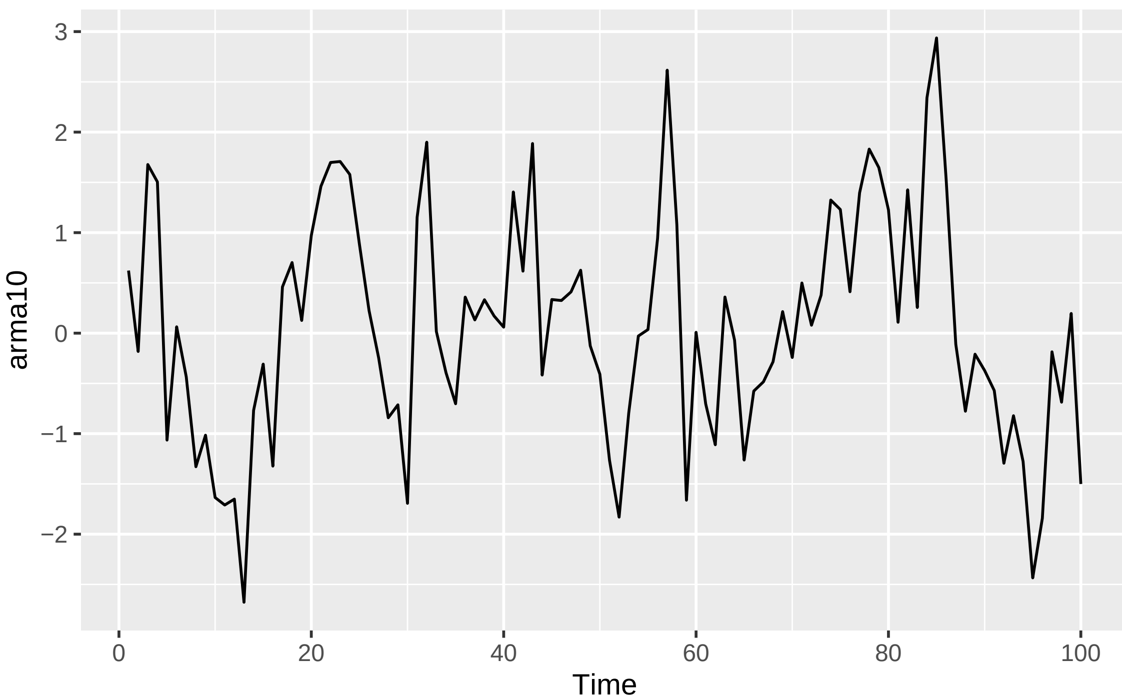
Q2

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.

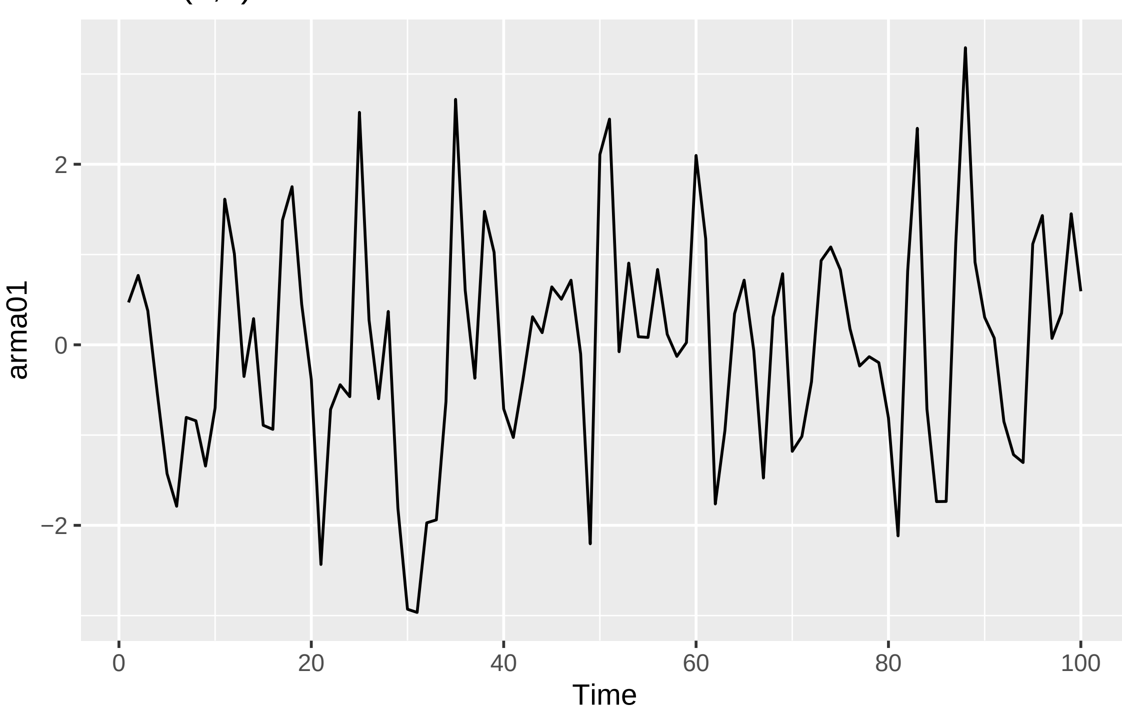
```
# Set seed for reproducibility
set.seed(123)

# Create ARMA models
arma10 <- arima.sim(model = list(ar = 0.6), n = 100)
arma01 <- arima.sim(model = list(ma = 0.9), n = 100)
arma11 <- arima.sim(model = list(ar = 0.6, ma = 0.9), n = 100)

# Plot ARMA models separately
autoplot(arma10) + ggtitle("ARMA(1,0) Model")
```

ARMA(1,0) Model

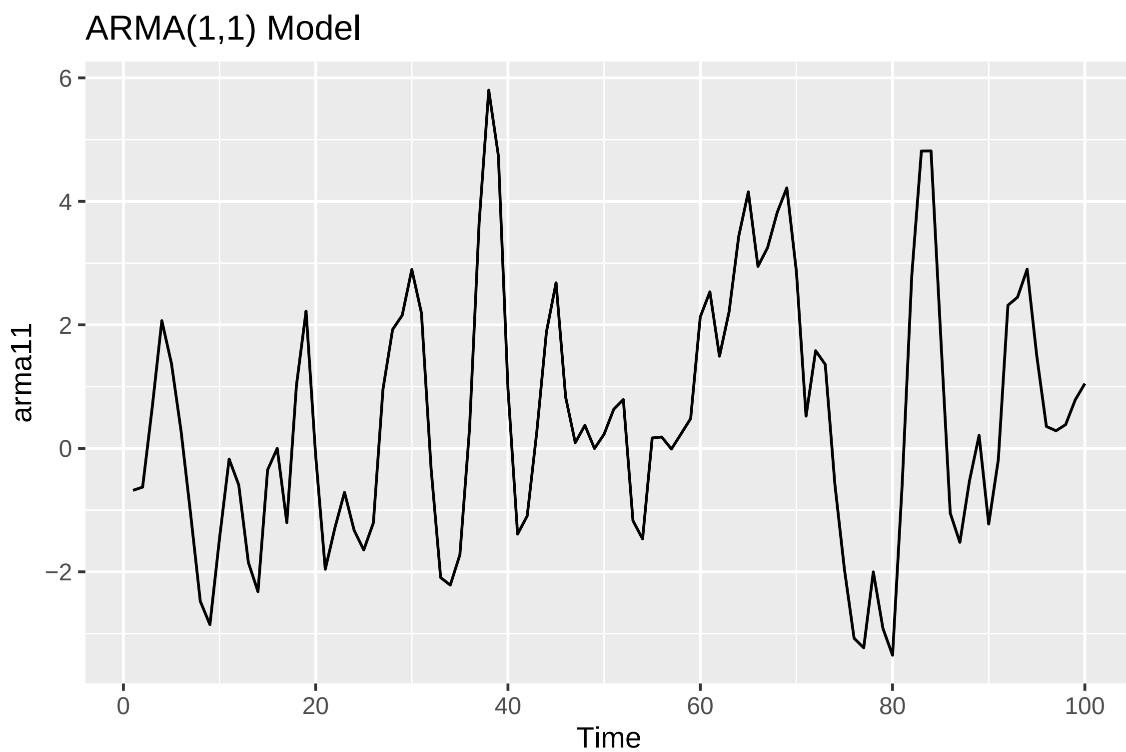



```r
autoplot(arma01) + ggtitle("ARMA(0,1) Model")
```

ARMA(0,1) Model

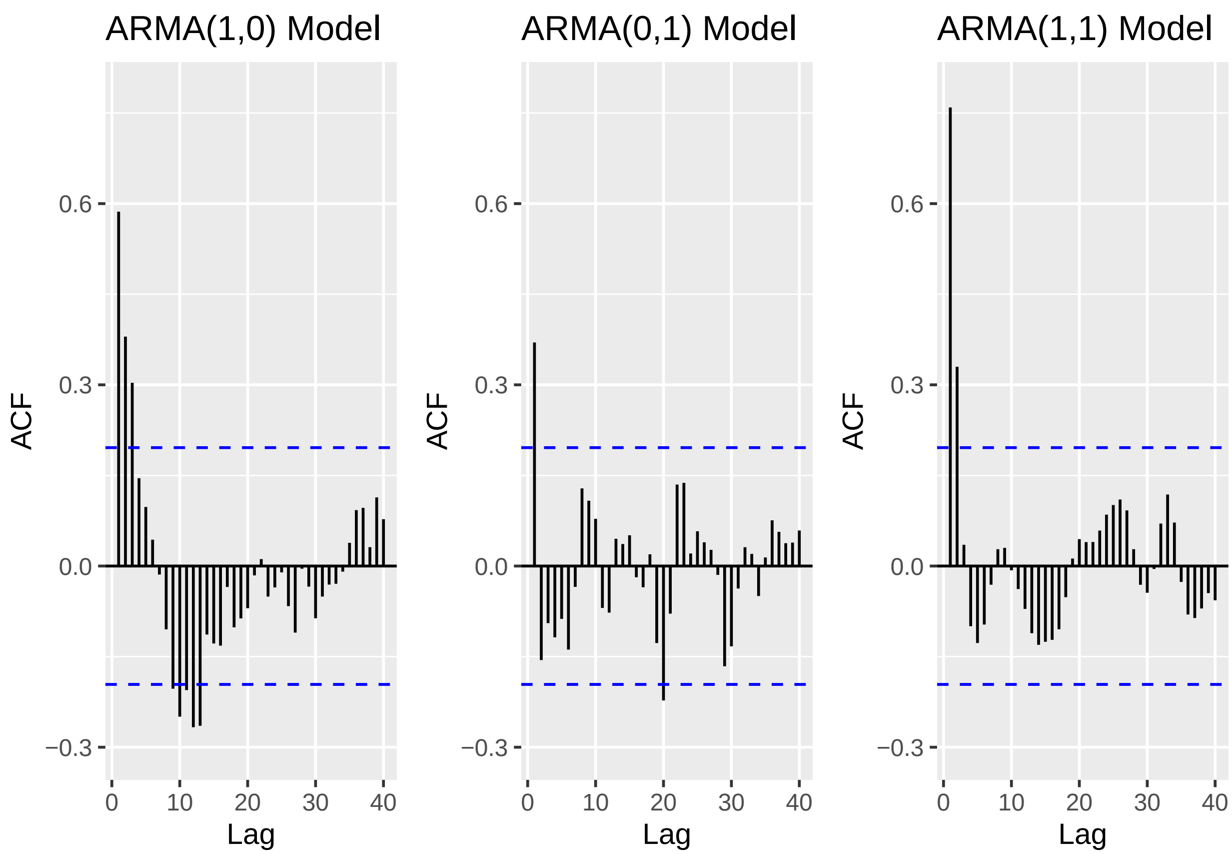


```r
autoplot(arma11) + ggtitle("ARMA(1,1) Model")
```



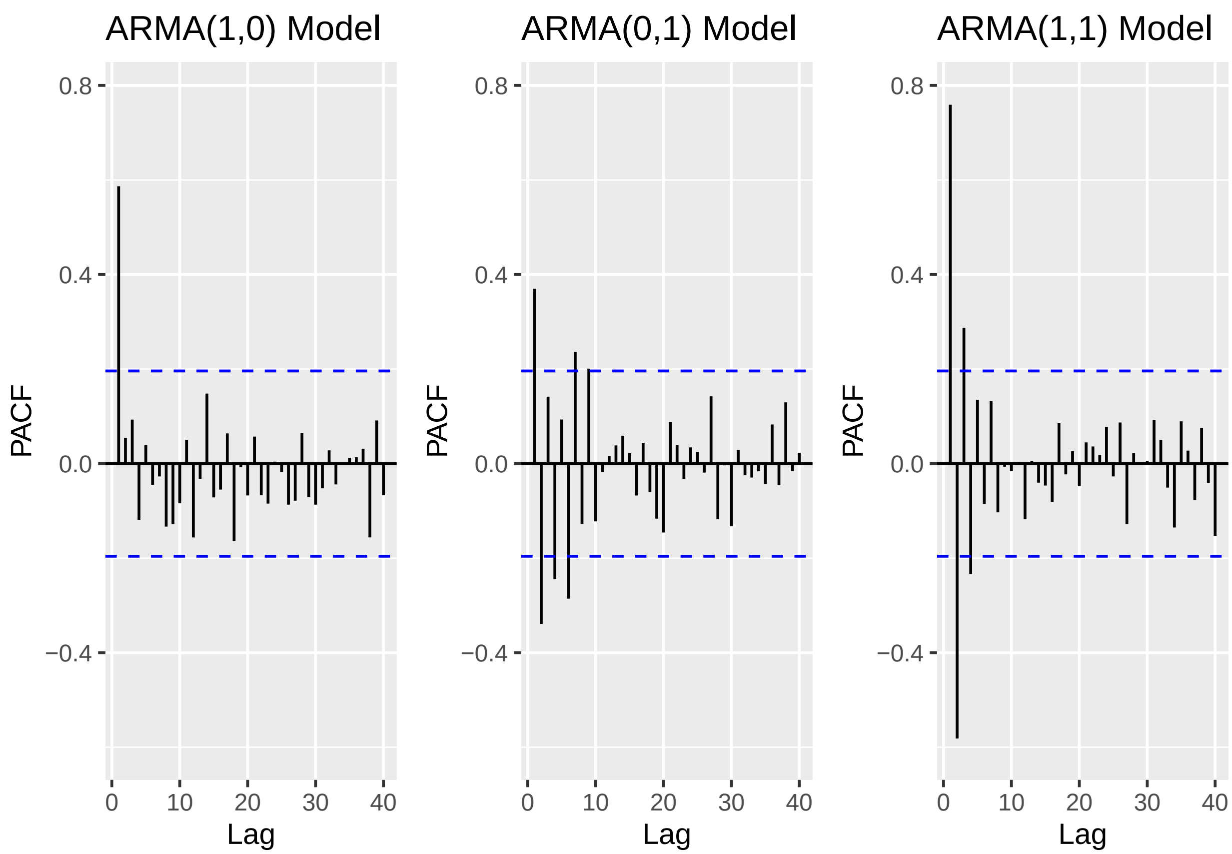
(b) Plot the sample ACF for each of these models in one window to facilitate comparison (Hint: use `cowplot::plot_grid()`).

```r
# Create ACF plots for all ARMA models
plot_grid(
  autoplot(Acf(arma10,lag.max=40, plot = FALSE),
           main = "ARMA(1,0) Model") + ylim(-0.3,0.78),
  autoplot(Acf(arma01,lag.max=40, plot = FALSE),
           main = "ARMA(0,1) Model") + ylim(-0.3,0.78),
  autoplot(Acf(arma11,lag.max=40, plot = FALSE),
           main = "ARMA(1,1) Model") + ylim(-0.3,0.78),
  nrow=1
)
```



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

```
# Create PACF plots for all ARMA models
plot_grid(
  autoplot(Pacf(arma10,lag.max=40, plot = FALSE),
           main = "ARMA(1,0) Model") + ylim(-0.6,0.78),
  autoplot(Pacf(arma01,lag.max=40, plot = FALSE),
           main = "ARMA(0,1) Model") + ylim(-0.6,0.78),
  autoplot(Pacf(arma11,lag.max=40, plot = FALSE),
           main = "ARMA(1,1) Model") + ylim(-0.6,0.78),
  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.

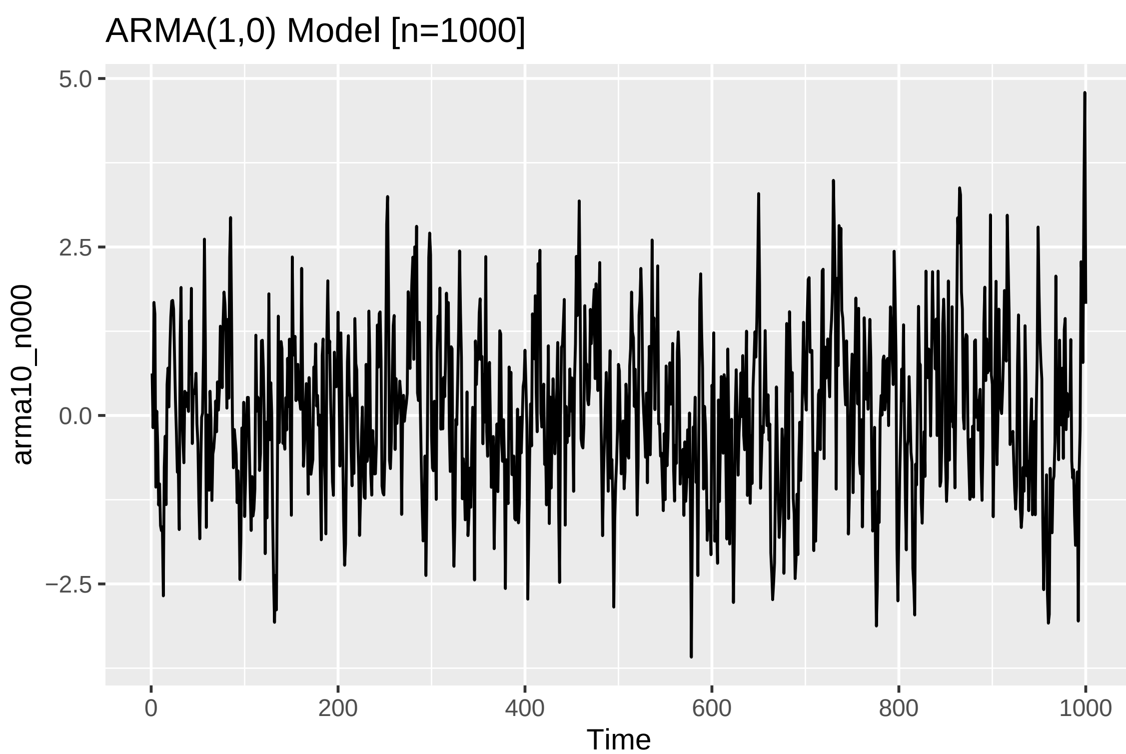
Answer:
ARMA (1,0) Model: We could identify the AR model and its order. The ACF plot of the ARMA (1,0) Model shows an apparent decaying trend over time and the PACF plot displays a clear cut-off at lag 1 (p=1).
ARMA(0,1) Model: In this model, the ACF plot shows a cut-off after lag 1, hence, we could identify q=1. Despite a few ups and downs in the initial lags, the PACF plot exhibits a slow decaying pattern over time. Therefore, we could also identify the MA pattern.
ARMA(1,1) Model: For the ARMA model, the ACF plot shows significant spikes at both lag 1 and 2. We might want to add lag 2 or q = 2 in this model. Further, a gradual decaying pattern after lag 2 is not as sharply pronounced as AR model. The PACF plot, on the other hand, shows a cut-off after lag 1 (p=1) but a few initial significant ups and downs lags before its gradual decay. Compared to AR and MA model, ARMA model is a bit more difficult to identify.

(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?

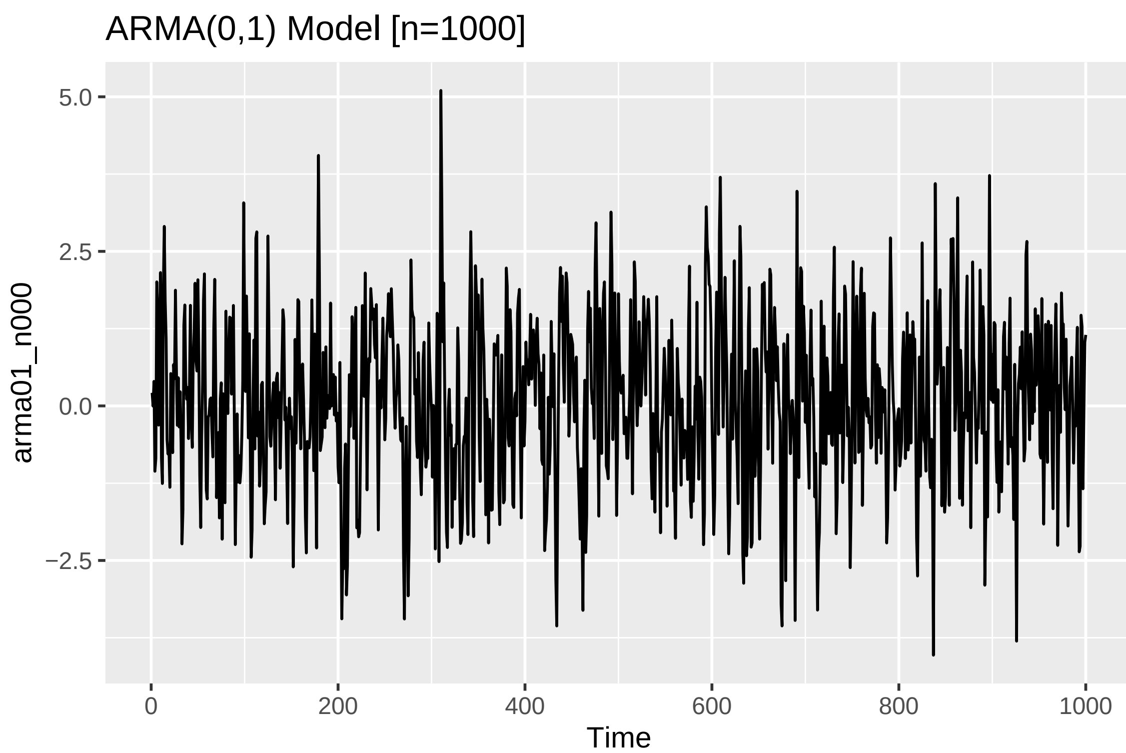
Answer: In the ARMA(1,0) model, the lag 1 correlation coefficient in the PACF plot is approximately close to 0.6 and we could say that R-coomputed coefficient match with $\phi = 0.6$. On the other hand, the lag 1 correlation coefficient of the ARMA model is approximately 0.8 and it does not match with $\phi = 0.6$.

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

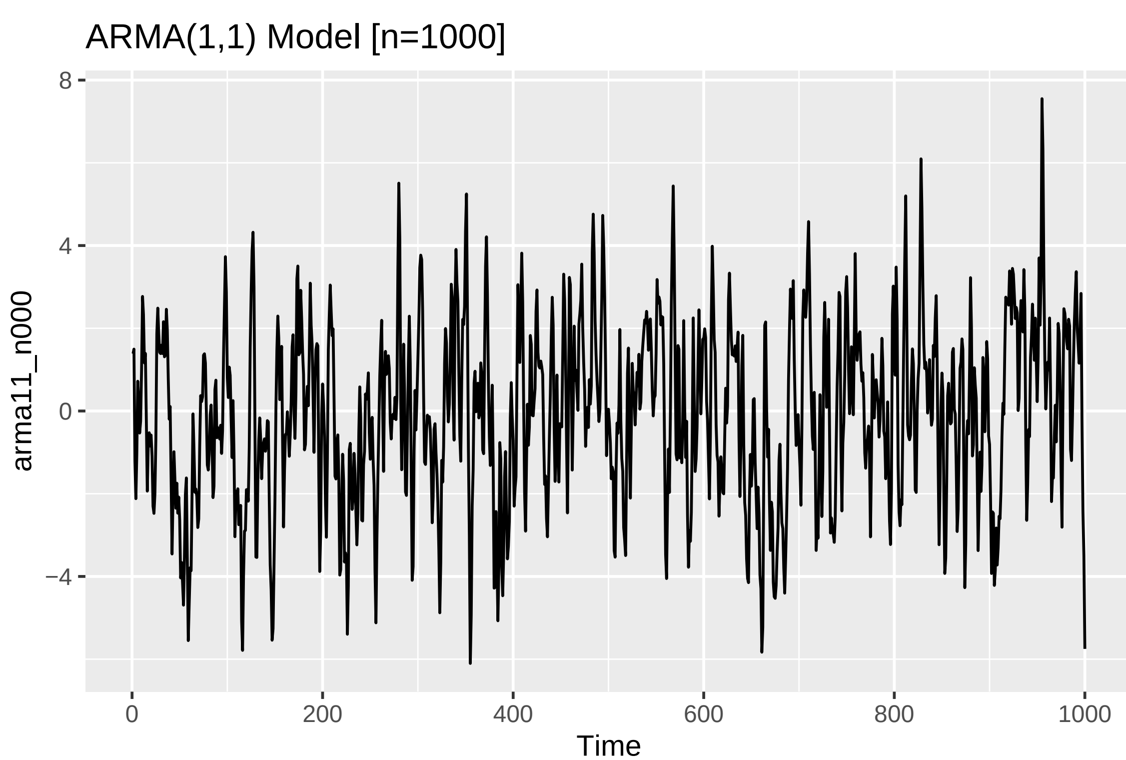
```r
# Set seed for reproducibility
set.seed(123)

# Create ARMA models
arma10_n000 <- arima.sim(model = list(ar = 0.6), n = 1000)
arma01_n000 <- arima.sim(model = list(ma = 0.9), n = 1000)
arma11_n000 <- arima.sim(model = list(ar = 0.6, ma = 0.9), n = 1000)

# Plot ARMA models separately
autoplot(arma10_n000) + ggtitle("ARMA(1,0) Model [n=1000]")
```



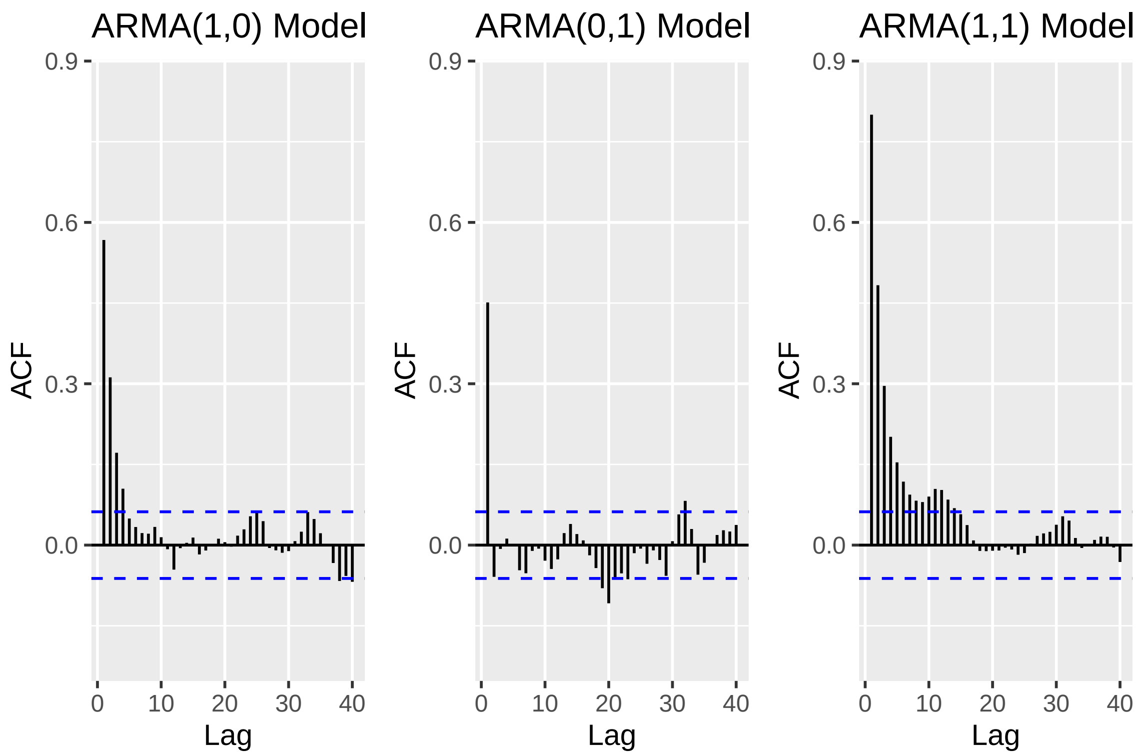

```r
autoplot(arma01_n000) + ggtitle("ARMA(0,1) Model [n=1000]")
```



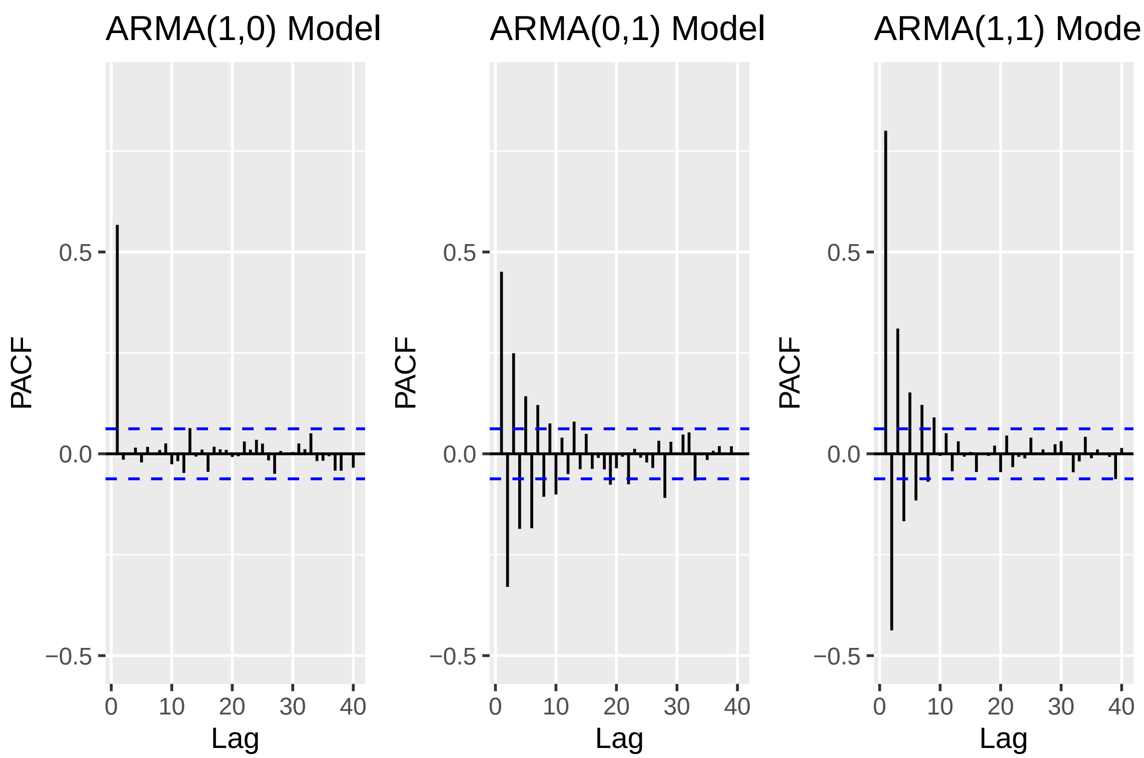

```
autoplot(arma11_n000) + ggtitle("ARMA(1,1) Model [n=1000]")
```



```r
# Create ACF plots for all ARMA models
plot_grid(
  autoplot(Acf(arma10_n000,lag.max=40, plot = FALSE),
           main = "ARMA(1,0) Model") + ylim(-0.2,0.85),
  autoplot(Acf(arma01_n000,lag.max=40, plot = FALSE),
           main = "ARMA(0,1) Model") + ylim(-0.2,0.85),
  autoplot(Acf(arma11_n000,lag.max=40, plot = FALSE),
           main = "ARMA(1,1) Model") + ylim(-0.2,0.85),
  nrow=1
)
```




```r
# Create PACF plots for all ARMA models
plot_grid(
  autoplot(Pacf(arma10_n000,lag.max=40, plot = FALSE),
           main = "ARMA(1,0) Model") + ylim(-0.5,0.9),
  autoplot(Pacf(arma01_n000,lag.max=40, plot = FALSE),
           main = "ARMA(0,1) Model") + ylim(-0.5,0.9),
  autoplot(Pacf(arma11_n000,lag.max=40, plot = FALSE),
           main = "ARMA(1,1) Model") + ylim(-0.5,0.9),
  nrow=1
)
```



Answer: When the total number of observations is increased from 100 to 1,000, the ACF plot of ARMA (1,0) model shows a slow decaying over time and the PACF of the model shows a clear cut-off at lag 1 (p=1). In the ARMA (0,1) model, the ACF shows a clear cutoff at lag 1 (q=1) and the PACF is also slowly decaying in general but some oscillate behaviors. In the ARMA(1,1) model, both ACF and PACF plots, this time, shows cut-off at lag 1 respectively (p=1 and q=1) with general gradual decay after a few initial spikes. In general, increasing the number of observations helps to identify the model and the order of the components better. The PACF values R computed for ARMA(1,0) and ARMA(1,1) models are approximately at 0.55 and 0.63 and they are close to $\phi = 0.6$ compared to the model with less 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.

Answer: p (order of AR, non-seasonal) = 1
d (differencing order, non-seasonal) = 0 q (order of MA, non-seasonal) = 1
P (order of AR, seasonal) = 1
D (differencing order, seasonal) = 0
Q (order of MA, seasonal) = 0
s (seasonal period) = 12
Therefore, the model would be ARIMA$(1, 0, 1)(1, 0, 0)_{12}$.

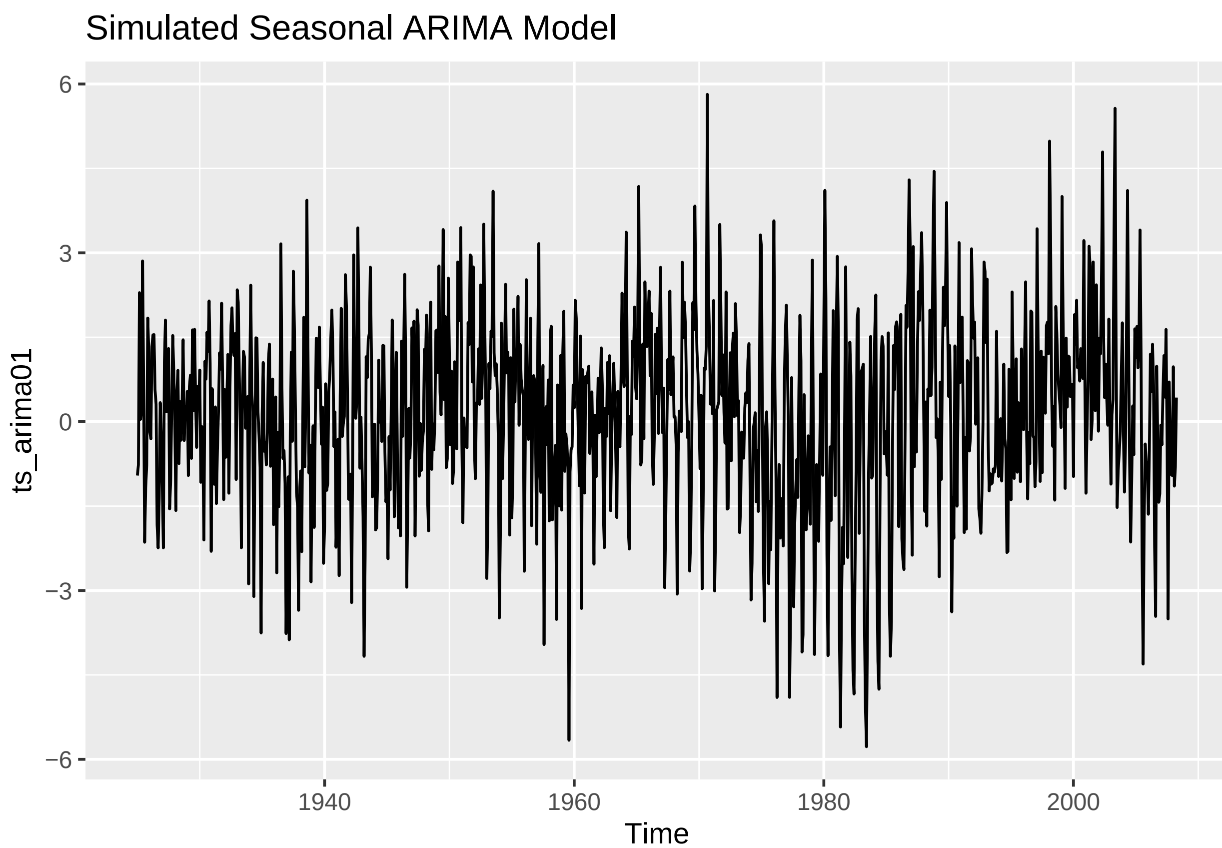
(b) Also from the equation what are the values of the parameters, i.e., model coefficients.

Answer: AR parameter (non-seasonal): $\phi_1 = 0.7$
AR parameter (seasonal): $\Phi_1 = -0.25$
MA parameter (non-seasonal): $\theta_1 = -0.1$

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
# Set seed for reproducibility
set.seed(123)

# Simulate a seasonal ARIMA model
arima01 <- sim_sarima(n = 1000,
                      model = list(sar = 0.8, ma = 0.5, nseasons = 12))

# Change the seasonal ARIMA model to time series object
ts_arima01 <- ts(arima01, frequency=12, c(1925,1))

# Plot the seasonal ARMA model
autoplot(ts_arima01) + ggtitle("Simulated Seasonal ARIMA Model")
```

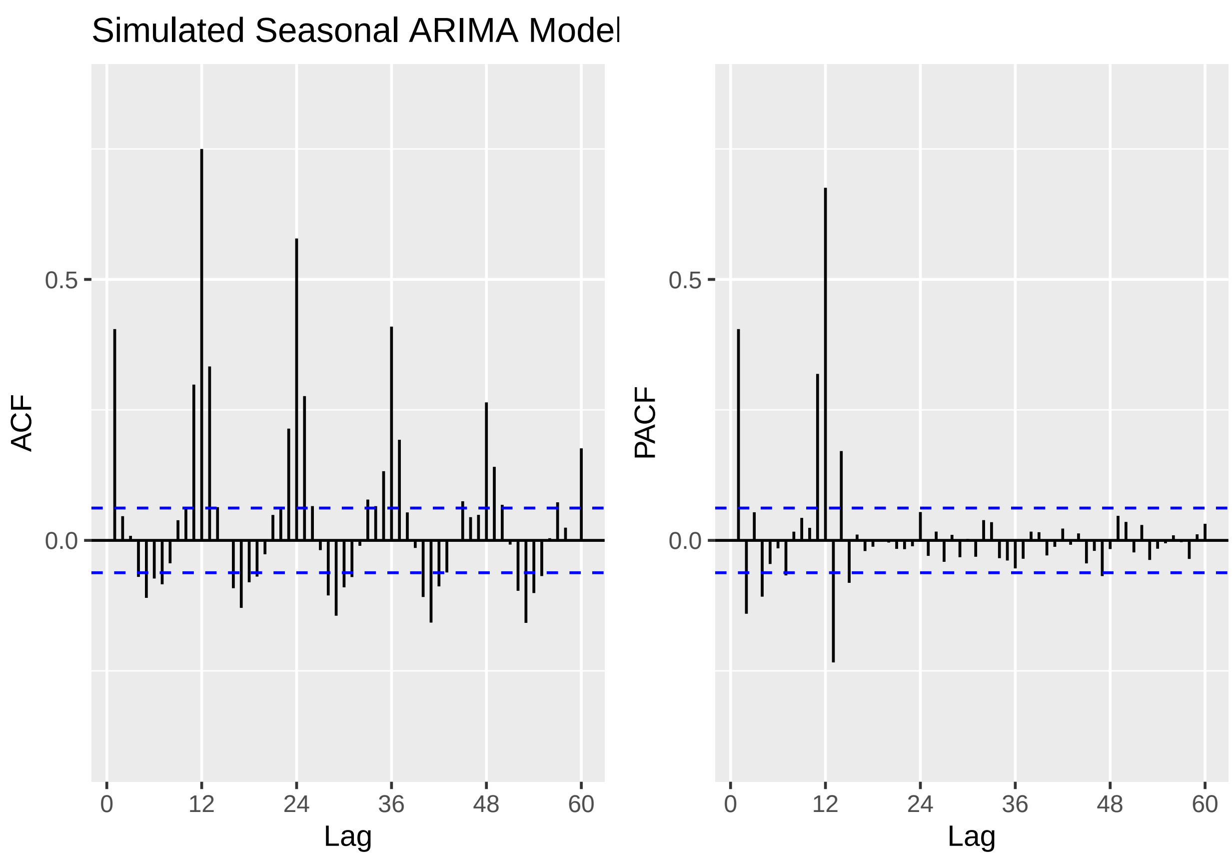


Answer: The plot of the seasonal ARIMA$(0,1) \times (1,0)_{12}$ model with 1,000 observations indicates some seasonal patterns. In the plot, there are some patterns of peaks and troughs repeating at a regular intervals.

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 and PACF plots of ARIMA Model
plot_grid(
  autoplot(Acf(ts_arima01,lag.max=60, plot = FALSE),
           main = "Simulated Seasonal ARIMA Model") + ylim(-0.4,0.85),
  autoplot(Pacf(ts_arima01,lag.max=60, plot = FALSE),
           main = "") + ylim(-0.4,0.85),
  nrow=1
)
```



Answer: Through the plots, we could clearly see the seasonal components. The spikes at lag 12, 24 and 36 in ACF plots indiate the strong AR seasonal pattern (Q=0). The spike at lag 12 in PACF and cut off after that period indicates that P = 1. For the non-seasonal parts, the spike at lag 1 and cut-off after that in the ACF plot indicates that q = 1 and that at lag 1 of PACF exhibits that p = 1. Therefore, we could conclude that the plots could fairly represent the seasonal ARIMA$(0,1) \times (1,0)_{12}$ model.