ENV 797 - Time Series Analysis for Energy and Environment Applications | Spring 2024

Assignment 6 - Due date 02/28/24

Ina Liao

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_Sp24.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(forecast)
library(tseries)
library(ggplot2)
library(Kendall)
library(lubridate)
library(tidyverse)
library(here)
library(knitr)
library(ggthemes)
library(cowplot)
library(dplyr)
#install.packages("sarima")
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: The ACF plot shows an exponential decay in autocorrelation. The PACF plot can be utilized to determine the order of the AR model.

- MA(1)

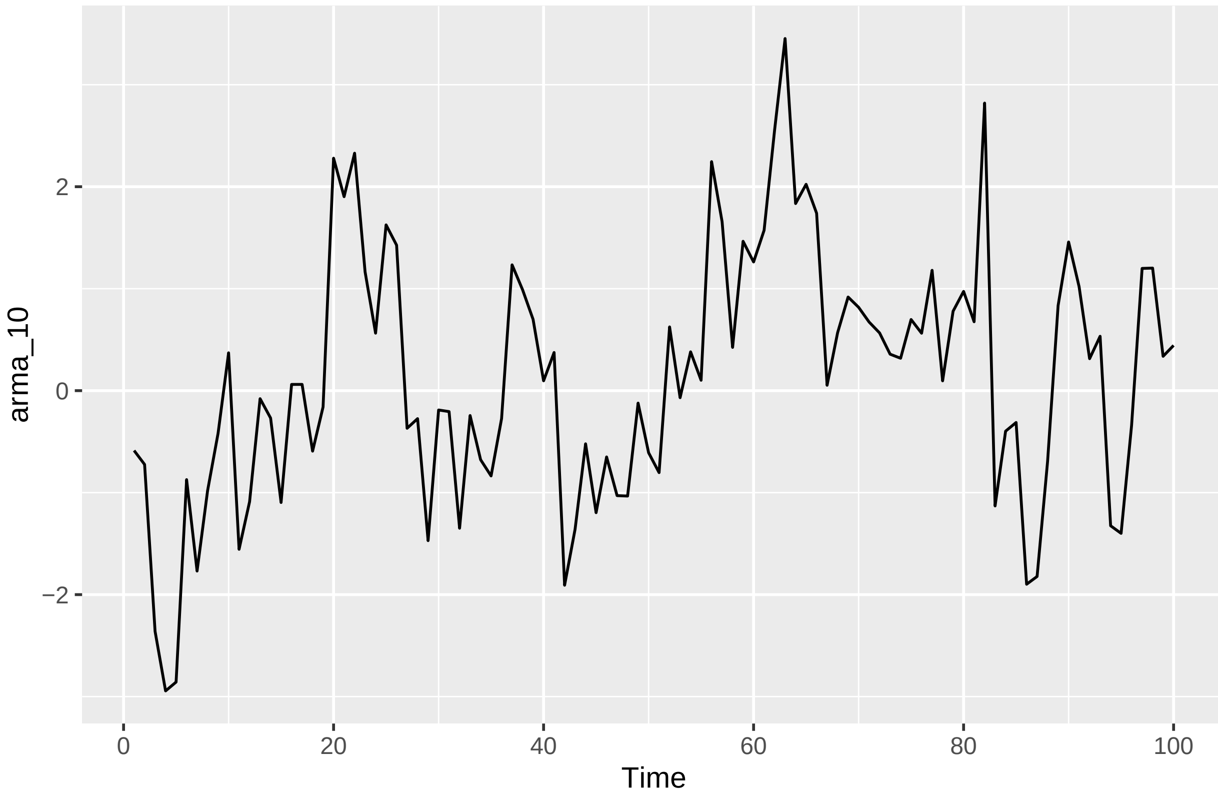
  Answer: The ACf plot can help us identify the order of the MA model. The PACF plot displays an exponential decay in partial autocorrelation.

Q2

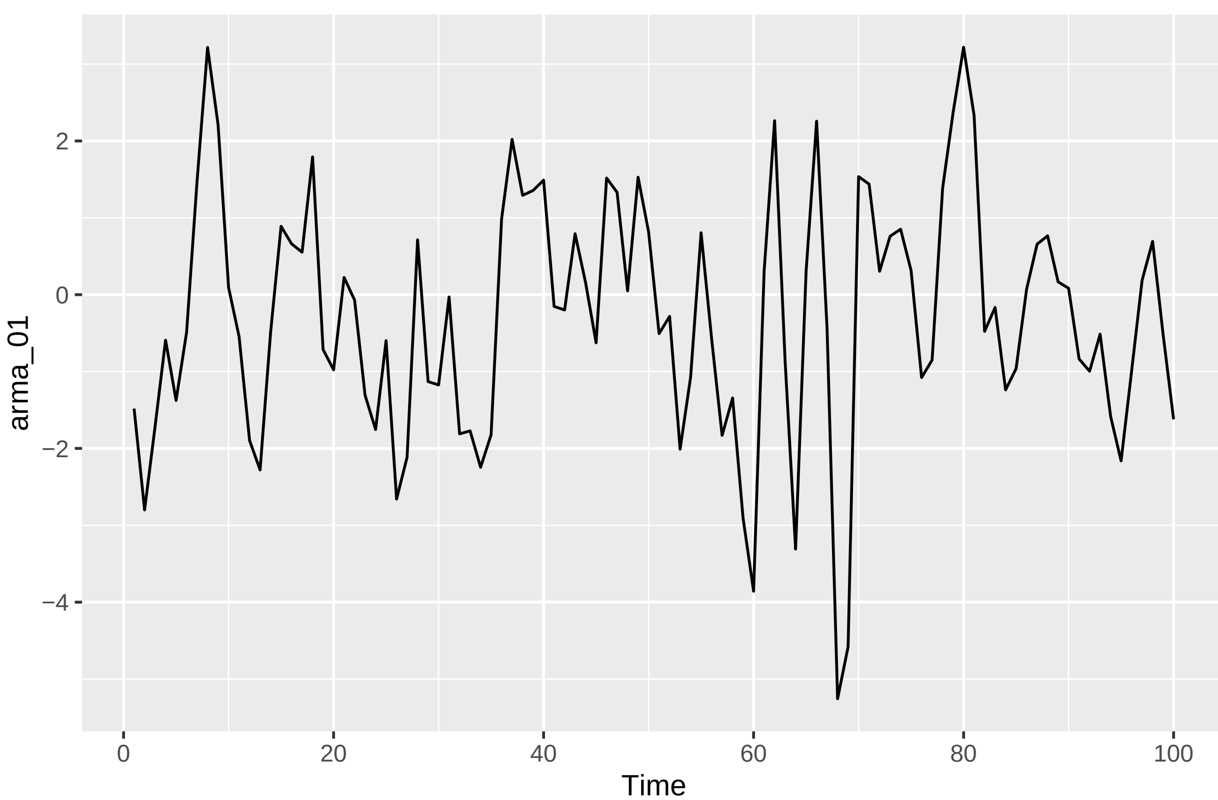
Recall that the non-seasonal ARIMA is described by three parameters ARIMA(p, d, q) where p is the order of the autoregressive component, d is the number of times the series need to be differenced to obtain stationarity and q is the order of the moving average component. If we don't need to difference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

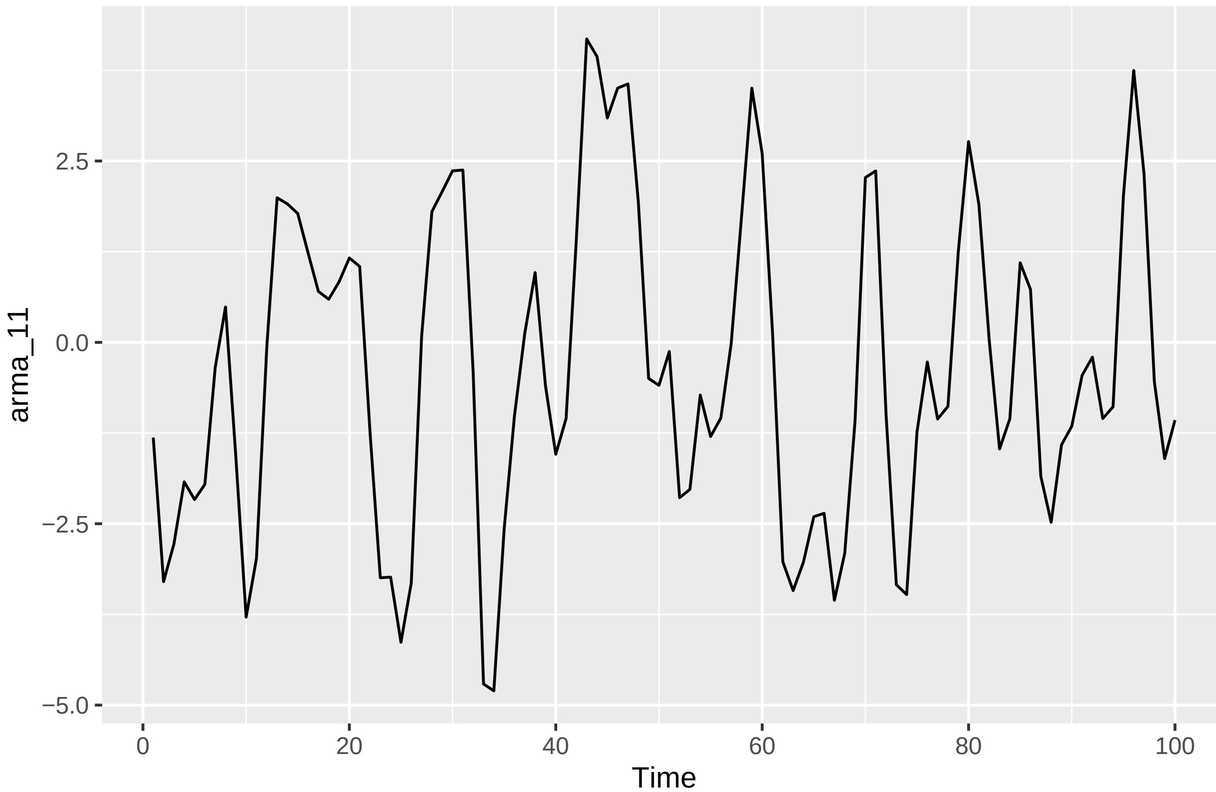
```
#set the parameters
phi<-0.6 #AR coefficient
theta<-0.9 #MA coefficient
n<-100

#generate data from ARMA(p,q)
#question: why in the arima.sim function, the argument order=c(p,d,q) is not needed
arma_10<-arima.sim(model=list(ar=phi),n=n)
arma_01<-arima.sim(model=list(ma=theta),n=n)
arma_11<-arima.sim(model=list(ar=phi,ma=theta),n=n)

#plot the series
plot_10<-autoplot(arma_10)
plot_10
```




```
plot_01<-autoplot(arma_01)
plot_01
```

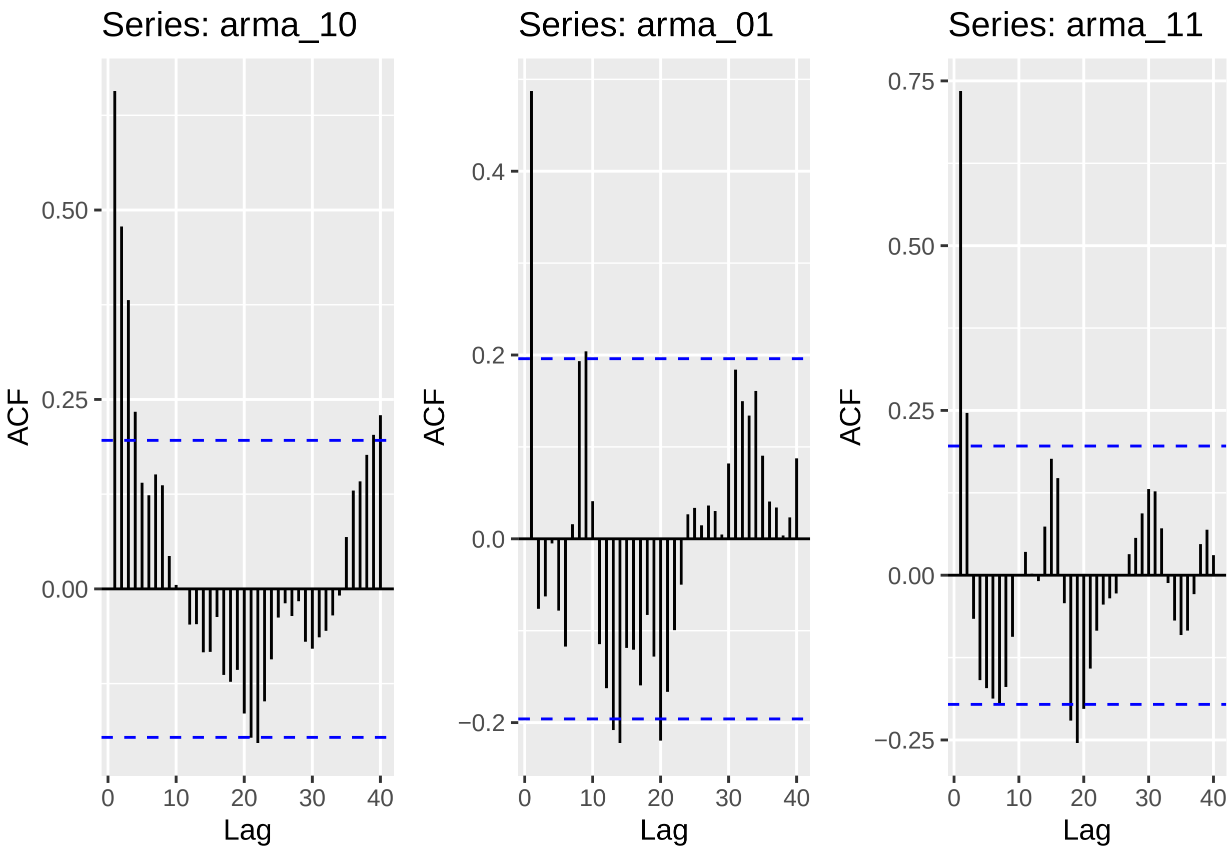


```r
plot_11<-autoplot(arma_11)
plot_11
```



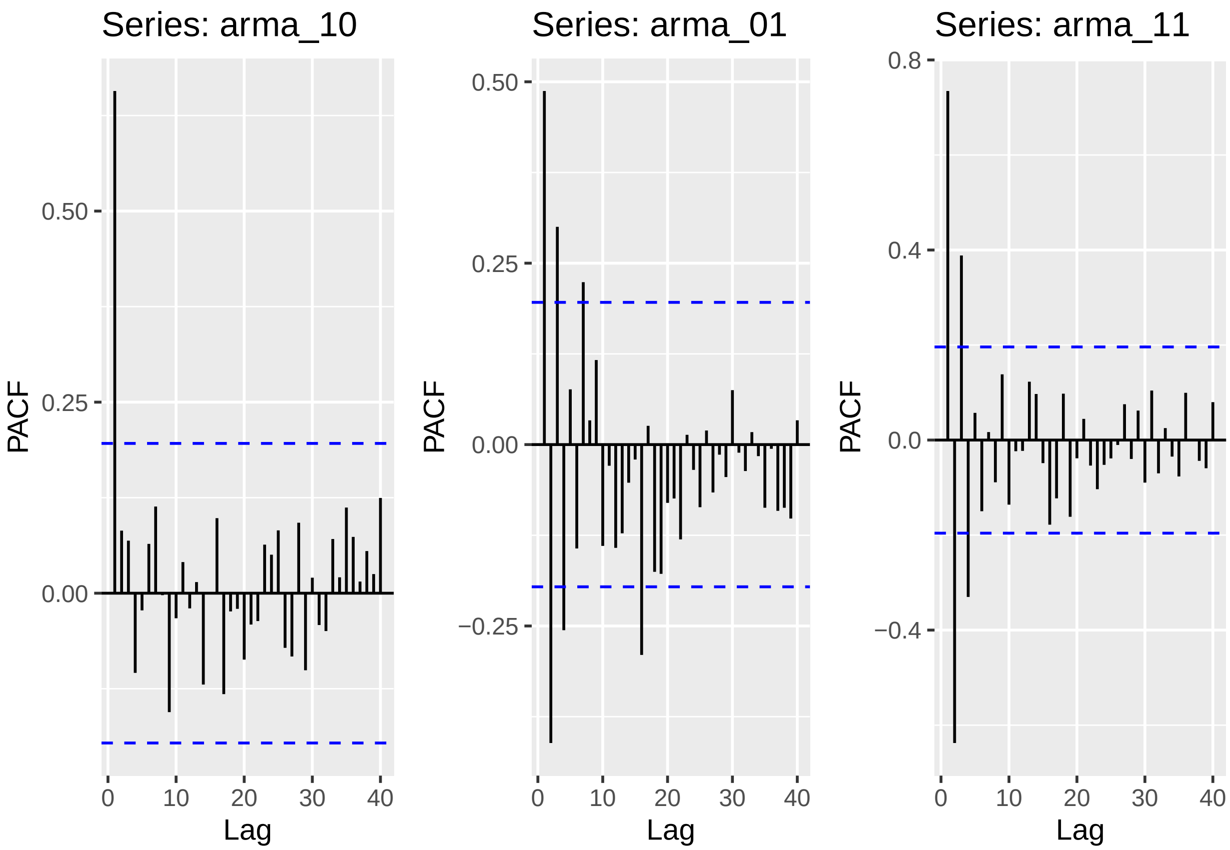
(b) Plot the sample ACF for each of these models in one window to facilitate comparison (Hint: use `cowplot::plot_grid()`).

```r
plot_combine_acf<-cowplot::plot_grid(
  autoplot(Acf(arma_10,lag.max=40,plot=FALSE)),
  autoplot(Acf(arma_01,lag.max=40,plot=FALSE)),
  autoplot(Acf(arma_11,lag.max=40,plot=FALSE)),
  ncol=3
)
plot_combine_acf
```



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

```
plot_combine_pacf<-cowplot::plot_grid(
  autoplot(Pacf(arma_10,lag.max=40,plot=FALSE)),
  autoplot(Pacf(arma_01,lag.max=40,plot=FALSE)),
  autoplot(Pacf(arma_11,lag.max=40,plot=FALSE)),
  ncol=3
)
plot_combine_pacf
```



(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: We can see an exponential decay in the ACF plot, and a sharp cutoff at lag-1 in the PACF plot, which suggests the AR model. The ACF plot shows a sharp cutoff at lag-1 and the PACF plot show an exponential decay, which suggests the MA model. We can observe an exponential decay in ACF and PACF plots, which suggests the ARMA model. Relying solely on PACF and ACF plots to determine the model may be insufficient, as the patterns in these plots can often overlap and be difficult to interpret without additional information.

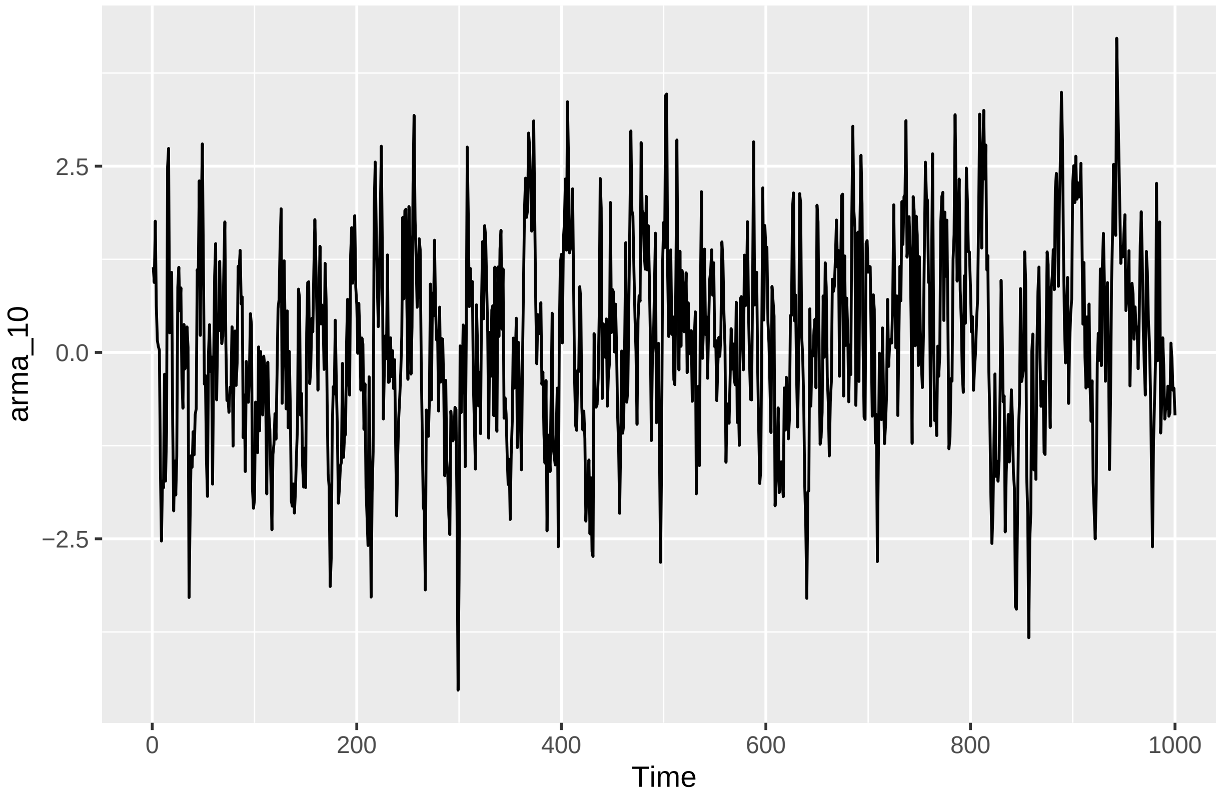
(e) Compare the PACF values R computed with the values you provided for the lag 1 correlation coefficient, i.e., does $\phi = 0.6$ match what you see on PACF for ARMA(1,0), and ARMA(1,1)? Should they match?

Answer: In ARMA(1,0) and ARMA(1,1), the PACF values at lag-1 match the AR coefficient (phi=0.6). If there is no randomness in the stimulated data, the partial autocoefficient should match the specified coefficient.

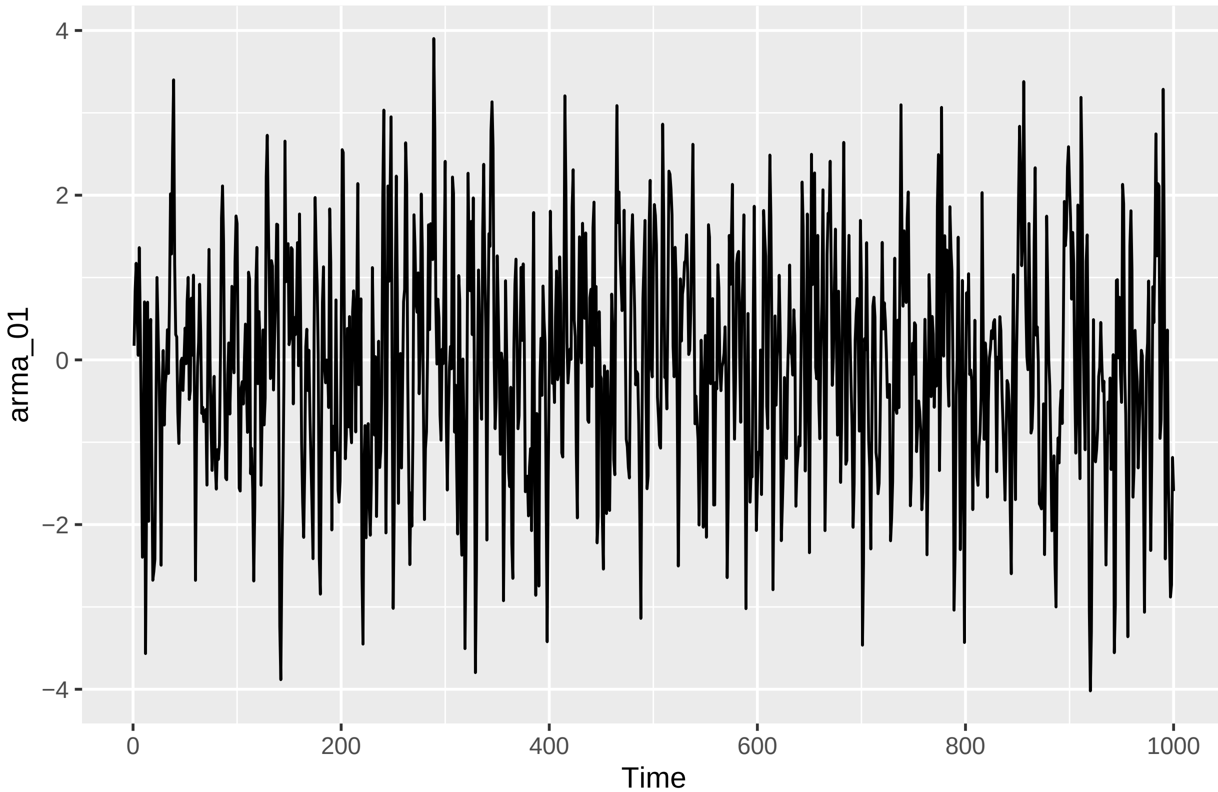
(f) Increase number of observations to $n = 1000$ and repeat parts (b)-(e).

```
#set the parameters
phi<-0.6 #AR coefficient
theta<-0.9 #MA coefficient
n<-1000

#generate data from ARMA(p,q)
arma_10<-arima.sim(model=list(ar=phi),n=n)
arma_01<-arima.sim(model=list(ma=theta),n=n)
arma_11<-arima.sim(model=list(ar=phi,ma=theta),n=n)
```

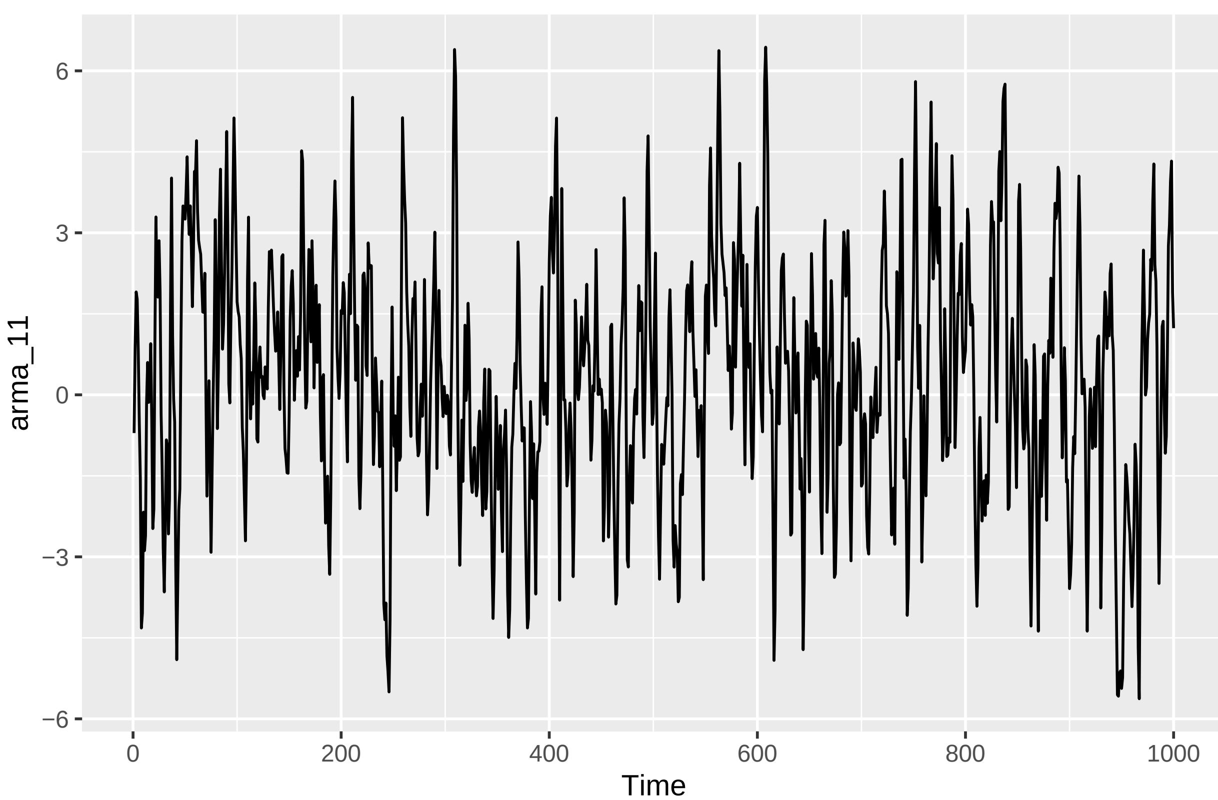
```r
#plot the series
plot_10<-autoplot(arma_10)
plot_10
```



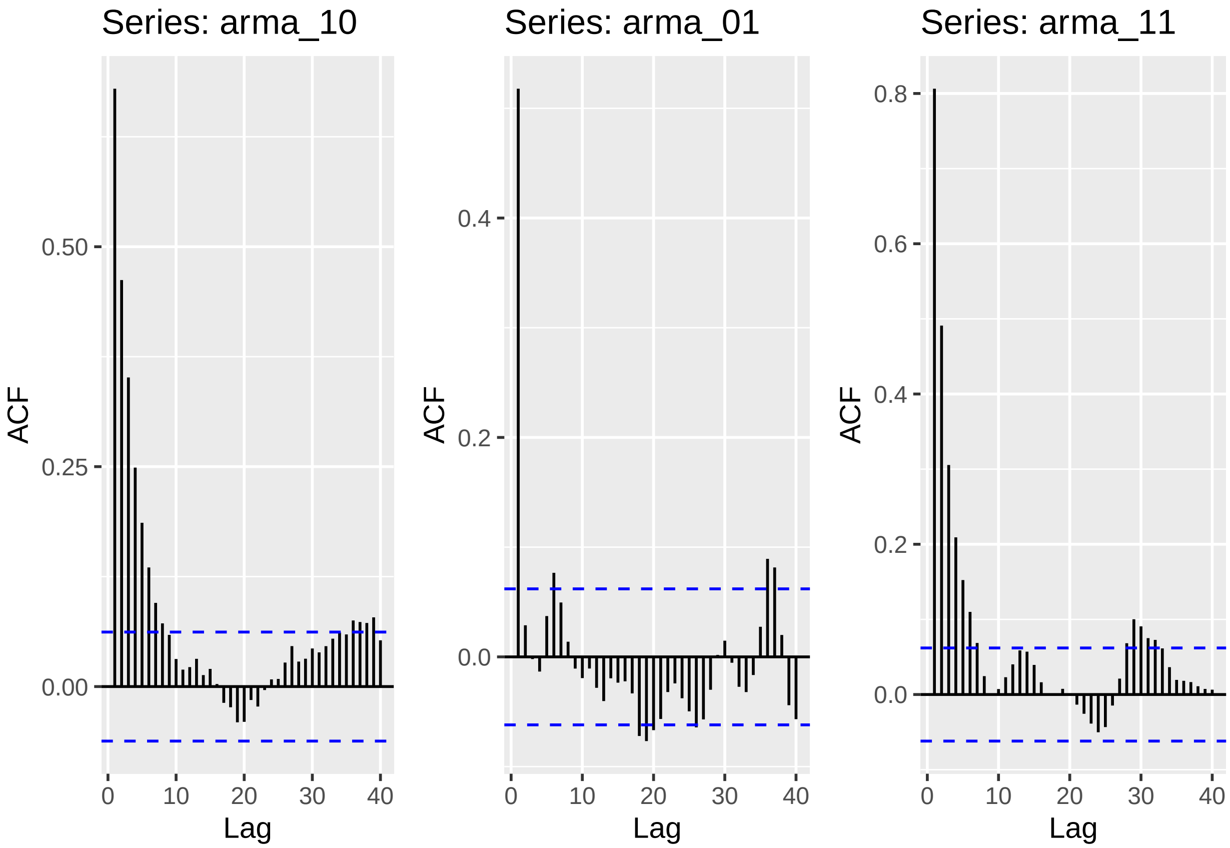

```r
plot_01<-autoplot(arma_01)
plot_01
```



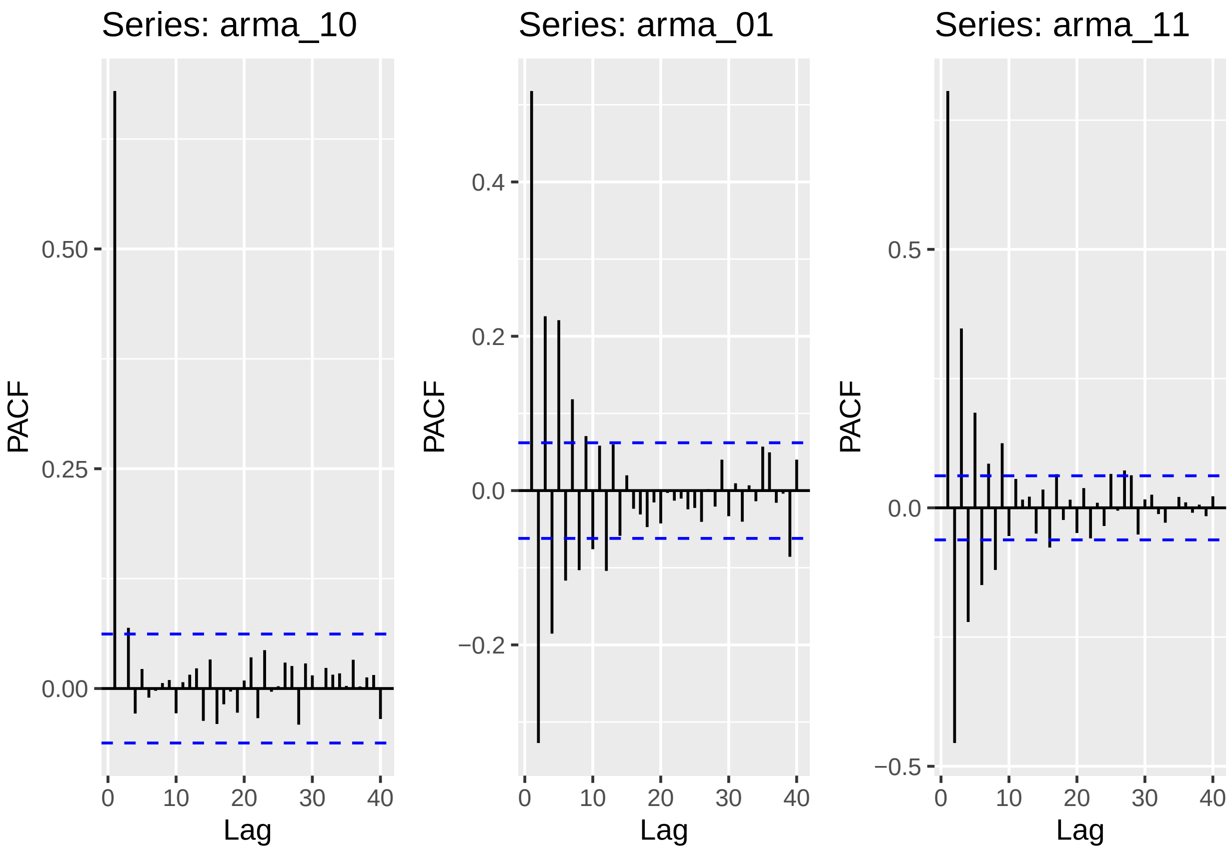

```
plot_11<-autoplot(arma_11)
plot_11
```



```r
#ACF plot
plot_combine_acf<-cowplot::plot_grid(
  autoplot(Acf(arma_10,lag.max=40,plot=FALSE)),
  autoplot(Acf(arma_01,lag.max=40,plot=FALSE)),
  autoplot(Acf(arma_11,lag.max=40,plot=FALSE)),
  ncol=3
)
plot_combine_acf
```




```r
#PACF plot
plot_combine_pacf<-cowplot::plot_grid(
  autoplot(Pacf(arma_10,lag.max=40,plot=FALSE)),
  autoplot(Pacf(arma_01,lag.max=40,plot=FALSE)),
  autoplot(Pacf(arma_11,lag.max=40,plot=FALSE)),
  ncol=3
)
plot_combine_pacf
```



Answer: We can see an exponential decay in the ACF plot, and a sharp cutoff at lag-1 in the PACF plot, which suggests the AR model. The ACF plot shows a sharp cutoff at lag-1 and the PACF plot show an exponential decay, which suggests the MA model. We can observe an exponential decay in ACF and PACF plots, which suggests the ARMA model. Relying solely on PACF and ACF plots to determine the model may be insufficient, as the patterns in these plots can often overlap and be difficult to interpret without additional analysis.

Answer: In ARMA(1,0) and ARMA(1,1), the PACF values at lag-1 do not match the AR coefficient (phi=0.6), which might due to the randomness in the stimulated data.

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 $\text{ARIMA}(p, d, q)(P, D, Q)_s$, i.e., identify the integers p, d, q, P, D, Q, s (if possible) from the equation.

1. p is the number of autoregressive term. p=1 because of y_{t-1}.
2. d is the number of differences. d=0 because there is no differencing term applied to y_t.
3. q is the number of moving average term. q=1 because of $0.1 * a_{t-1}$.
4. P is the seasonal order of autoregressive term. P=1 because of y_{t-1}.
5. D is the seasonal order of differences. Q=0 because there is no MA terms.
6. s is the length of the seasonal cycle. s=12 because y_{t-12}

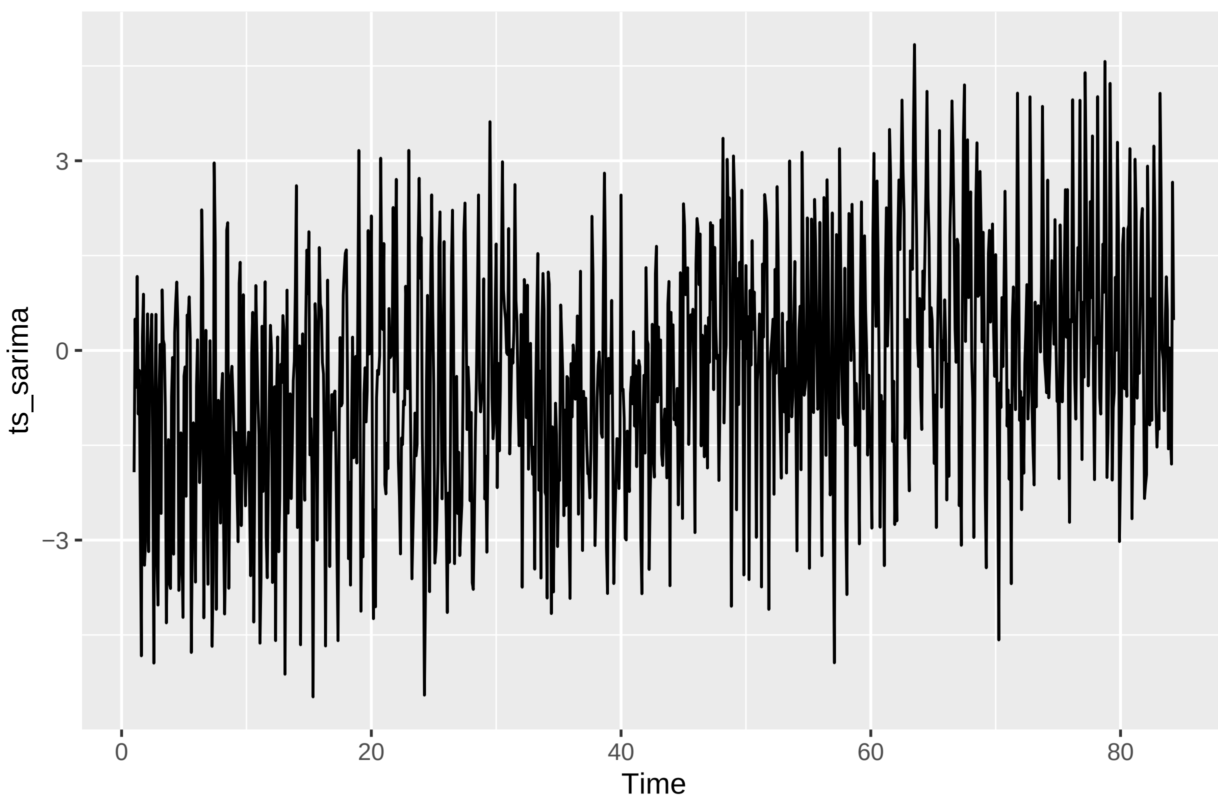
(b) Also from the equation what are the values of the parameters, i.e., model coefficients.

1. The autoregressive coefficient $\phi = 0.7$.
2. The moving average coefficient $\theta = -0.1$.
3. The seasonal autoregressive coefficient $\Phi = -0.25$. The seasonal moving average coefficient $\Theta = 0$.

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 the parameters
phi<-0.8 #AR coefficient
theta<-0.5 #MA coefficient
#integrated part was omitted so the series do not need differencing

#define ARIMA model
sarima<-sim_sarima(model=list(ma=0.5,sar=0.8, nseasons=12), n=1000)

#convert to time series object
#arima.sim will generate time series object
#sarima will generate numeric results
#create time series object
ts_sarima<-ts(sarima,frequency=12)

#plot the series
plot_sarima<-autoplot(ts_sarima)
plot_sarima
```

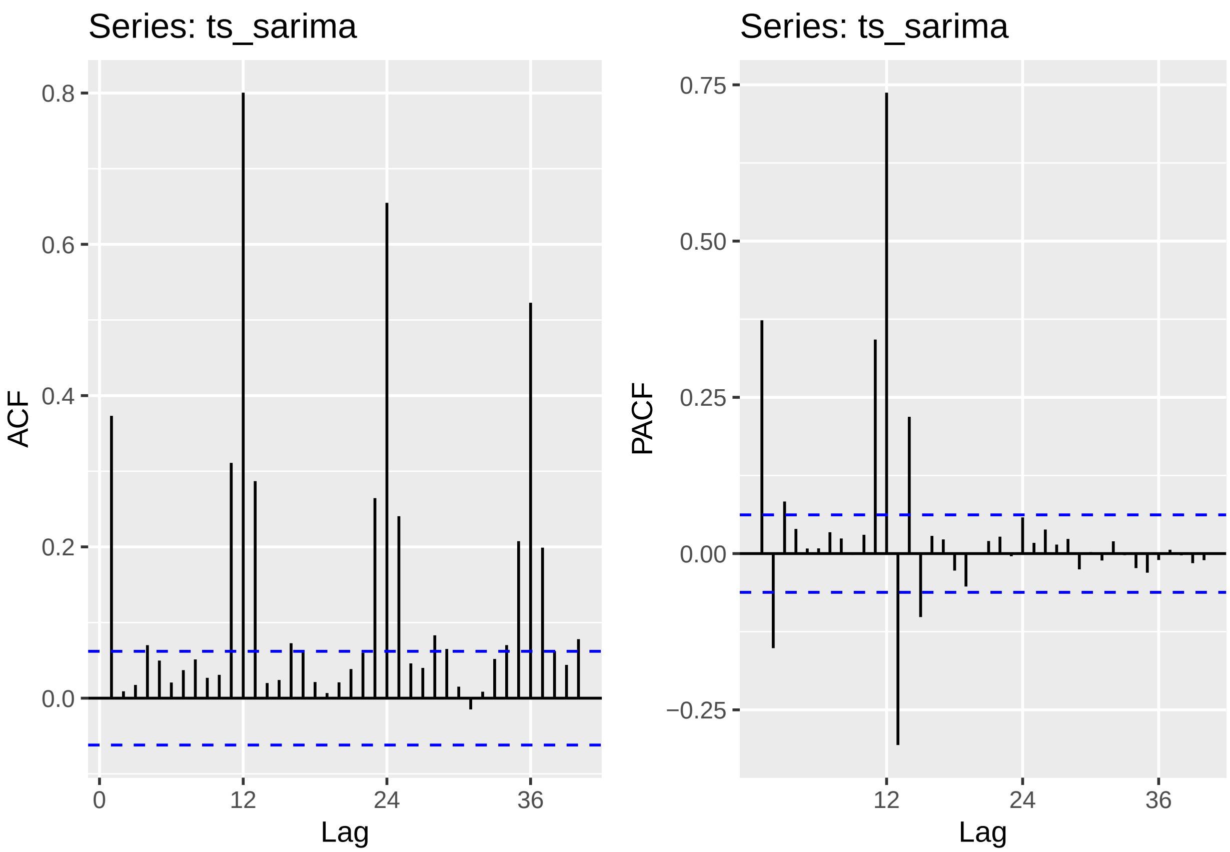


The plot show a wave-like pattern, which suggests possible presence of a seasonal component in the series.

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.

```
plot_sarima_acf_pacf<-cowplot::plot_grid(
  autoplot(Acf(ts_sarima,lag.max=40,plot=FALSE)),
  autoplot(Pacf(ts_sarima,lag.max=40,plot=FALSE)),
  ncol=2
)
plot_sarima_acf_pacf
```



There are spikes at lags 12, 24, and 36 on the ACF plot, and a spike at lag 12 on the PACF plot. These spikes suggest seasonal components might exist in the time series data. Further statistical testing is required to confirm the existence of a seasonal component.