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

Assignment 6 - Due date 02/28/24

Shubhangi Gupta

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.

```
#Load/install required package here
library(ggplot2)
library(lubridate)
```

```
##
## Attaching package: 'lubridate'

## The following objects are masked from 'package:base':
##
##     date, intersect, setdiff, union
```

```
library(forecast)
```

```
## Registered S3 method overwritten by 'quantmod':
##   method            from
##   as.zoo.data.frame zoo
```

```
library(Kendall)
library(tseries)
#install.packages("outliers")
library(outliers)
library(tidyverse)
```

```
## -- Attaching core tidyverse packages ------------------------ tidyverse 2.0.0 --
## v dplyr   1.1.3     v stringr 1.5.0
## v forcats 1.0.0     v tibble  3.2.1
## v purrr   1.0.2     v tidyr   1.3.0
## v readr   2.1.4

## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
## i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors
```

```r
library(cowplot)
```

```
##
## Attaching package: 'cowplot'
##
## The following object is masked from 'package:lubridate':
##
##     stamp
```

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: AR(2) refers to an autoregressive model with lag 2 which means that y(t) linearly depends on y(t-2), y(t-4), etc. In this case the ACF would have two signficant lags and then begin to decay exponentially, whereas the PACF will have 2 significant bars and the remaining will be insignificant.

- MA(1)

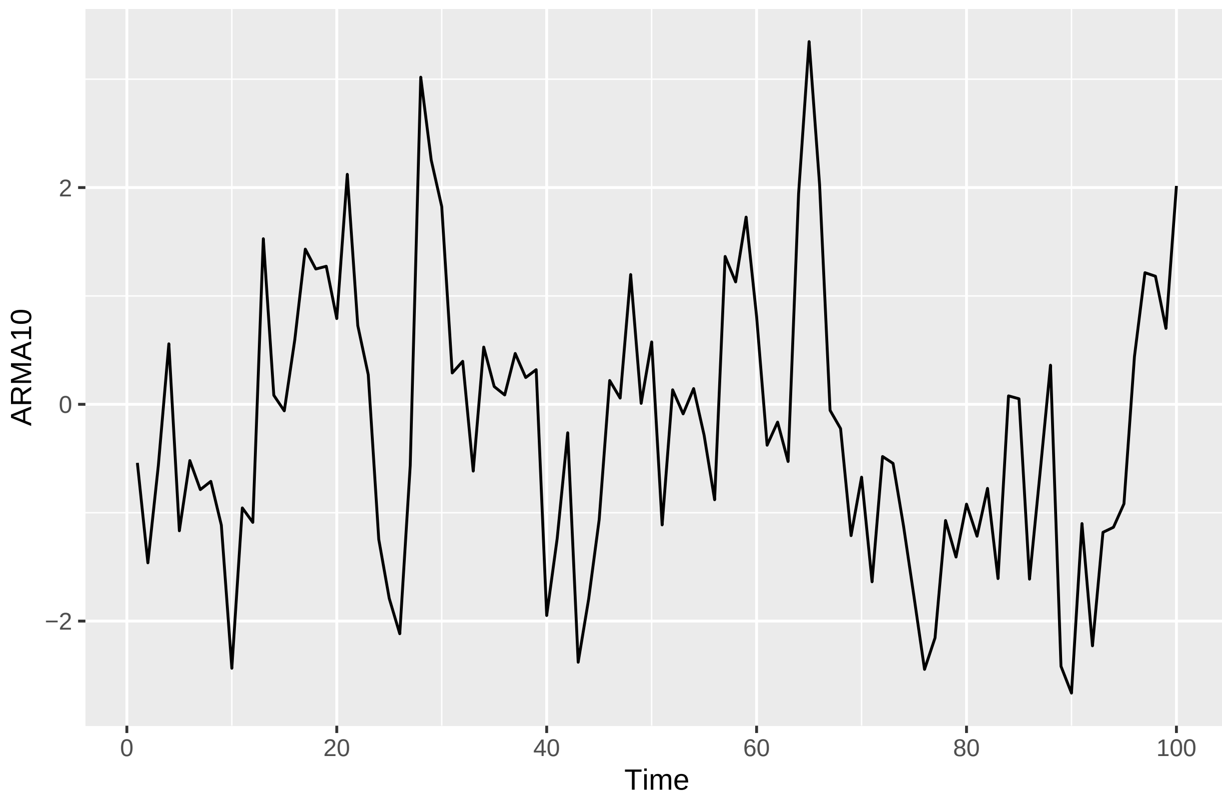
  Answer: MA(1) refers to a moving average model with lag 1, meaning that the deviation of y(t) from the mean depends on 1 previous deviation. In this case, ACF will have 1 significant lag and then cut off, whereas PACF will have 1 significant lag and then decay exponentially.

Q2

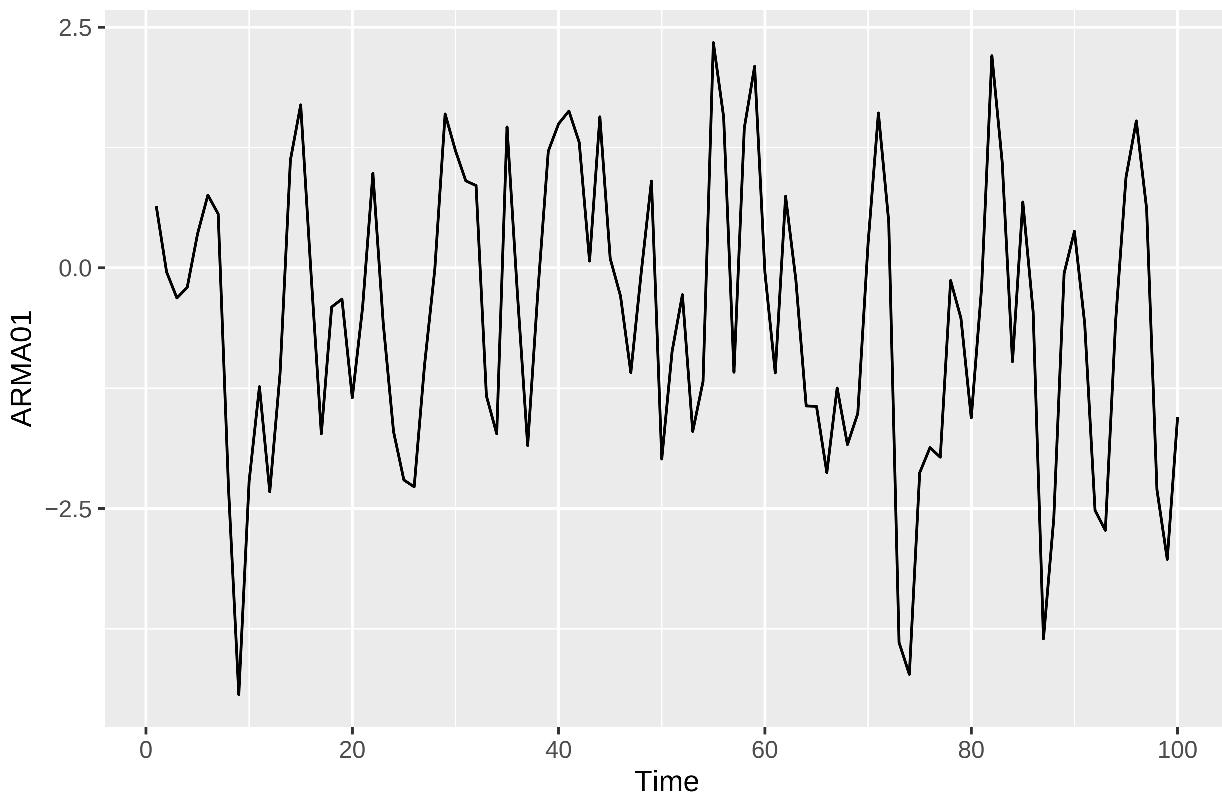
Recall that the non-seasonal ARIMA is described by three parameters ARIMA(p, d, q) where p is the order of the autoregressive component, d is the number of times the series need to be differenced to obtain stationarity and q is the order of the moving average component. If we don't need to difference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

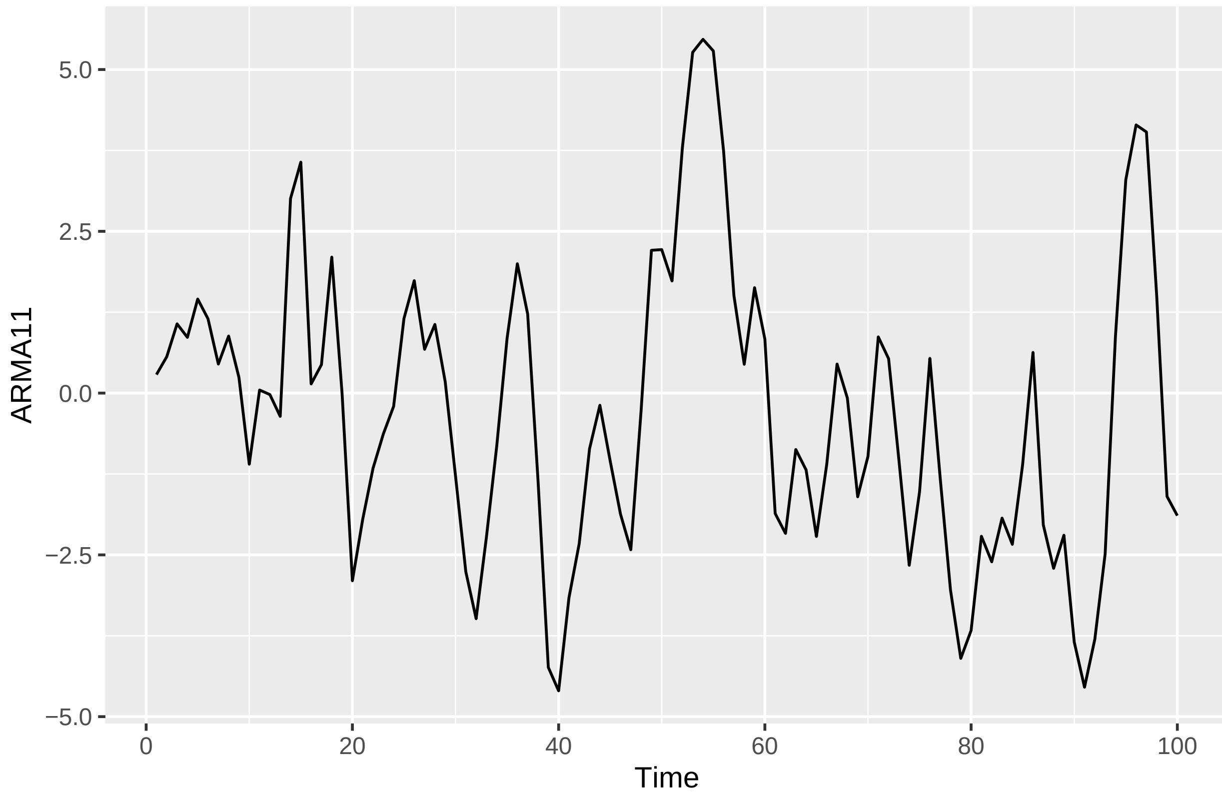
```r
ARMA10 <- arima.sim(model = list(order = c(1,0,0), ar=0.6), n=100)
ARMA01 <- arima.sim(model = list(order = c(0,0,1), ma=0.9), n=100)
ARMA11 <- arima.sim(model = list(order=c(1,0,1), ar=0.6, ma=0.9), n=100)
autoplot(ARMA10)
```




```r
autoplot(ARMA01)
```




```
autoplot(ARMA11)
```

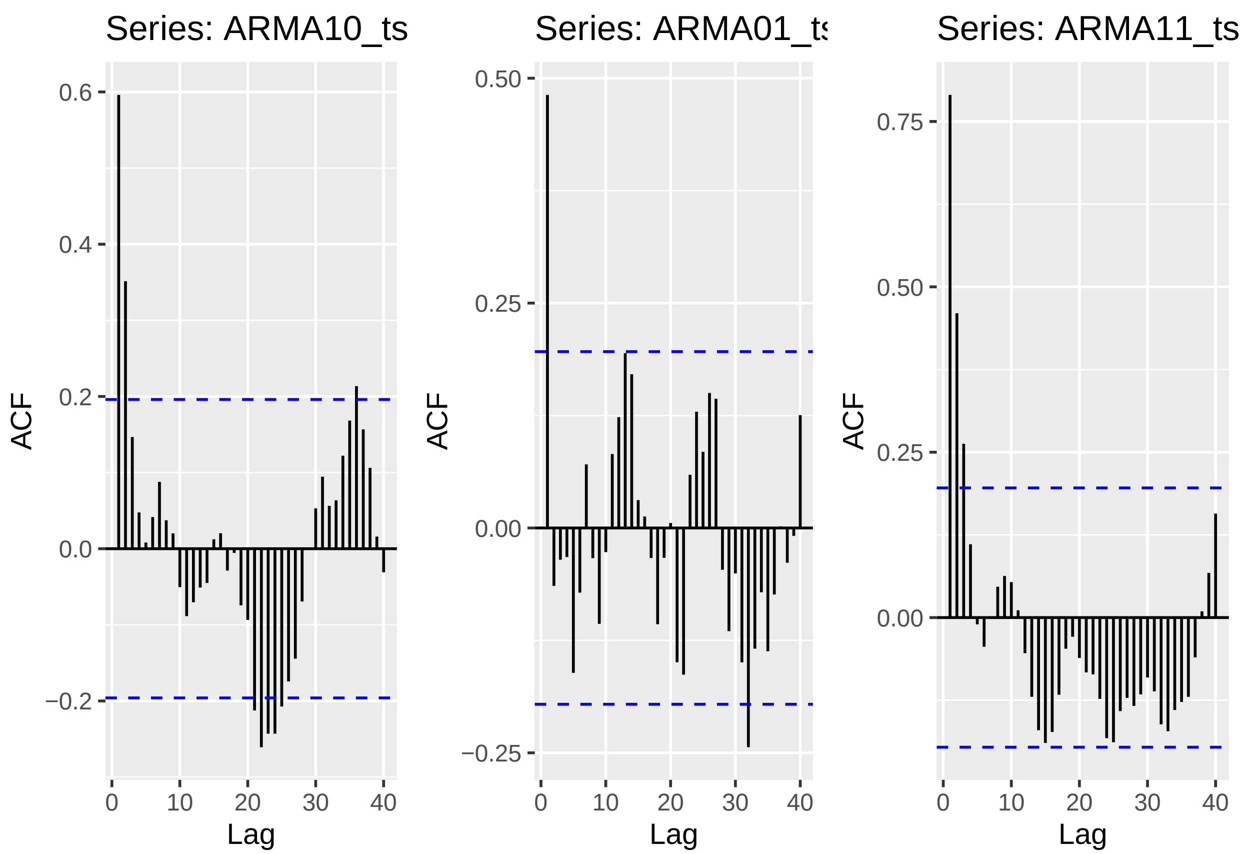


(b) Plot the sample ACF for each of these models in one window to facilitate comparison (Hint: use `cowplot::plot_grid()`).

```
#Converting the three series above to time series
ARMA10_ts <- ts(ARMA10)
ARMA01_ts <- ts(ARMA01)
ARMA11_ts <- ts(ARMA11)

#Calculating their ACF
ARMA10_ts_ACF <- Acf(ARMA10_ts, lag.max=40, plot=FALSE)
ARMA01_ts_ACF <- Acf(ARMA01_ts, lag.max=40, plot=FALSE)
ARMA11_ts_ACF <- Acf(ARMA11_ts, lag.max=40, plot=FALSE)

#plotting the three graphs
ARMA10_ts_ACF_plot <- autoplot(ARMA10_ts_ACF)
ARMA01_ts_ACF_plot <- autoplot(ARMA01_ts_ACF)
ARMA11_ts_ACF_plot <- autoplot(ARMA11_ts_ACF)
plot_grid(ARMA10_ts_ACF_plot, ARMA01_ts_ACF_plot, ARMA11_ts_ACF_plot, nrow=1)
```



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

```
#Calculating their PACF
ARMA10_ts_PACF <- Pacf(ARMA10_ts, lag.max=40, plot=FALSE)
ARMA01_ts_PACF <- Pacf(ARMA01_ts, lag.max=40, plot=FALSE)
ARMA11_ts_PACF <- Pacf(ARMA11_ts, lag.max=40, plot=FALSE)

ARMA10_ts_PACF
```

```
##
## Partial autocorrelations of series 'ARMA10_ts', by lag
##
##      1      2      3      4      5      6      7      8      9     10     11
##  0.596 -0.006 -0.094 -0.003  0.010  0.071  0.060 -0.093  0.012 -0.075 -0.033
##     12     13     14     15     16     17     18     19     20     21     22
##  0.038 -0.018 -0.032  0.079 -0.022 -0.066  0.071 -0.127 -0.013 -0.199 -0.108
##     23     24     25     26     27     28     29     30     31     32     33
##  0.020 -0.115 -0.051  0.000 -0.071  0.115  0.054  0.012  0.076 -0.097  0.046
##     34     35     36     37     38     39     40
##  0.161  0.003  0.108 -0.037 -0.075  0.039 -0.055
```

```
ARMA01_ts_PACF
```

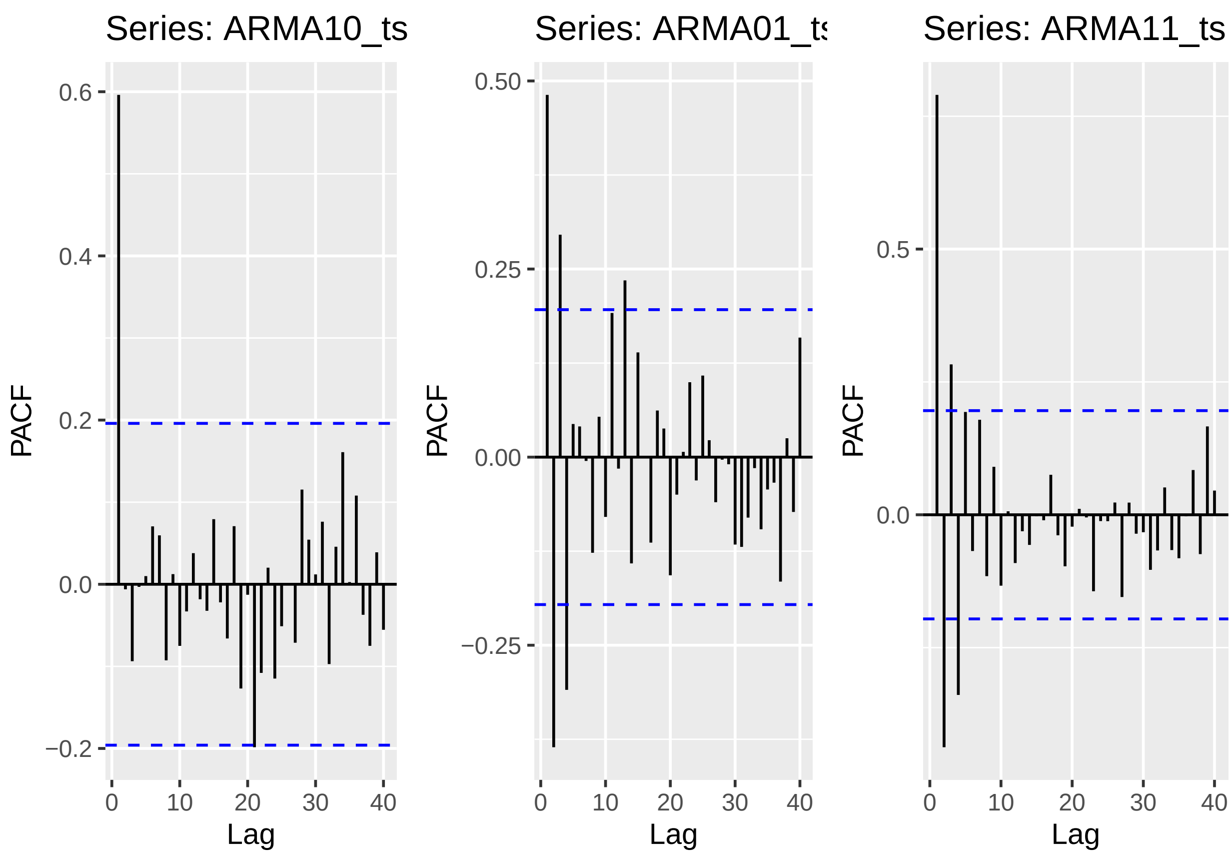
```
##
```

```
## Partial autocorrelations of series 'ARMA01_ts', by lag
##
##      1      2      3      4      5      6      7      8      9     10     11
##  0.481 -0.386  0.296 -0.309  0.044  0.041 -0.005 -0.127  0.054 -0.079  0.192
##     12     13     14     15     16     17     18     19     20     21     22
## -0.015  0.235 -0.141  0.139 -0.001 -0.114  0.062  0.038 -0.157 -0.050  0.007
##     23     24     25     26     27     28     29     30     31     32     33
##  0.100 -0.031  0.108  0.022 -0.060 -0.003 -0.009 -0.116 -0.119 -0.080 -0.015
##     34     35     36     37     38     39     40
## -0.096 -0.043 -0.034 -0.165  0.025 -0.073  0.159
```

```
ARMA11_ts_PACF
```

```
##
## Partial autocorrelations of series 'ARMA11_ts', by lag
##
##      1      2      3      4      5      6      7      8      9     10     11
##  0.790 -0.437  0.283 -0.339  0.193 -0.068  0.179 -0.115  0.090 -0.133  0.007
##     12     13     14     15     16     17     18     19     20     21     22
## -0.091 -0.031 -0.057 -0.003 -0.010  0.075 -0.039 -0.097 -0.022  0.011 -0.005
##     23     24     25     26     27     28     29     30     31     32     33
## -0.144 -0.012 -0.012  0.023 -0.155  0.023 -0.036 -0.033 -0.104 -0.067  0.051
##     34     35     36     37     38     39     40
## -0.066 -0.082 -0.003  0.084 -0.074  0.166  0.046
```

```
#plotting the three graphs
ARMA10_ts_PACF_plot <- autoplot(ARMA10_ts_PACF)
ARMA01_ts_PACF_plot <- autoplot(ARMA01_ts_PACF)
ARMA11_ts_PACF_plot <- autoplot(ARMA11_ts_PACF)
plot_grid(ARMA10_ts_PACF_plot, ARMA01_ts_PACF_plot, ARMA11_ts_PACF_plot, 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) has an exponential decay in AR and lag 1 in PACF indicating that it is an autoregressive model with lag 1. ARMA(0,1) cuts off after 1 lag in the ACF and PACF, indicating a moving average of lag 1. In the third model, there is a rapid decay of the ACF and the PACF cuts off after lag 1 so while I would veer towards guessing that it is an AR model, it is hard to tell.

(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: No they don't match, and they shouldn't because ARMA (1,0) does not consist of an MA component whereas ARMA (1,1) does, and so the PACF of both will be different.

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

```r
#Creating three new models with n=1000
ARMA10_2 <- arima.sim(model = list(order = c(1,0,0), ar=0.6), n=1000)
ARMA01_2 <- arima.sim(model = list(order = c(0,0,1), ma=0.9), n=1000)
ARMA11_2 <- arima.sim(model = list(order=c(1,0,1), ar=0.6, ma=0.9), n=1000)

#Converting the three series above to time series
```

```r
ARMA10_2_ts <- ts(ARMA10_2)
ARMA01_2_ts <- ts(ARMA01_2)
ARMA11_2_ts <- ts(ARMA11_2)

#Calculating their ACF
ARMA10_2_ts_ACF <- Acf(ARMA10_2_ts, lag.max=40, plot=FALSE)
ARMA01_2_ts_ACF <- Acf(ARMA01_2_ts, lag.max=40, plot=FALSE)
ARMA11_2_ts_ACF <- Acf(ARMA11_2_ts, lag.max=40, plot=FALSE)

#Calculating their PACF
ARMA10_2_ts_PACF <- Pacf(ARMA10_2_ts, lag.max=40, plot=FALSE)
ARMA01_2_ts_PACF <- Pacf(ARMA01_2_ts, lag.max=40, plot=FALSE)
ARMA11_2_ts_PACF <- Pacf(ARMA11_2_ts, lag.max=40, plot=FALSE)

ARMA10_2_ts_PACF
```

```
##
## Partial autocorrelations of series 'ARMA10_2_ts', by lag
##
##      1      2      3      4      5      6      7      8      9     10     11
##  0.604  0.051  0.033  0.006 -0.020  0.015 -0.044 -0.059  0.021 -0.035  0.004
##     12     13     14     15     16     17     18     19     20     21     22
##  0.049 -0.016  0.000  0.009  0.028  0.014  0.037  0.057  0.005 -0.034 -0.044
##     23     24     25     26     27     28     29     30     31     32     33
## -0.010  0.022 -0.005  0.024  0.008 -0.027  0.025 -0.004  0.001  0.027  0.000
##     34     35     36     37     38     39     40
##  0.008 -0.060 -0.007  0.024  0.014 -0.026 -0.011
```

```r
ARMA01_2_ts_PACF
```

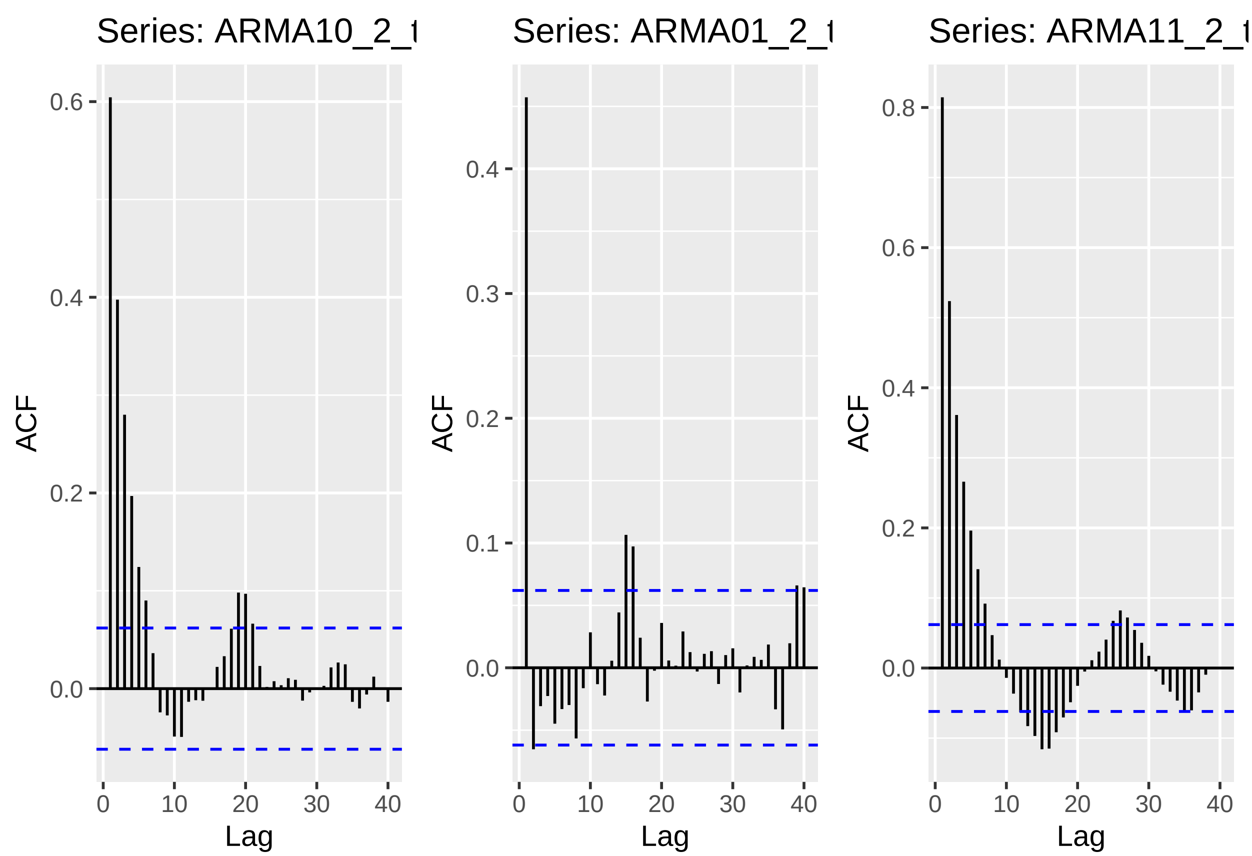
```
##
## Partial autocorrelations of series 'ARMA01_2_ts', by lag
##
##      1      2      3      4      5      6      7      8      9     10     11
##  0.457 -0.347  0.242 -0.220  0.128 -0.127  0.063 -0.121  0.120 -0.087  0.031
##     12     13     14     15     16     17     18     19     20     21     22
## -0.031  0.026  0.044  0.091 -0.011  0.017 -0.040  0.064 -0.009  0.004  0.030
##     23     24     25     26     27     28     29     30     31     32     33
##  0.018 -0.011  0.011  0.013  0.005 -0.019  0.049 -0.067  0.033 -0.009  0.000
##     34     35     36     37     38     39     40
##  0.025 -0.012 -0.072  0.037  0.010  0.046  0.026
```

```r
ARMA11_2_ts_PACF
```

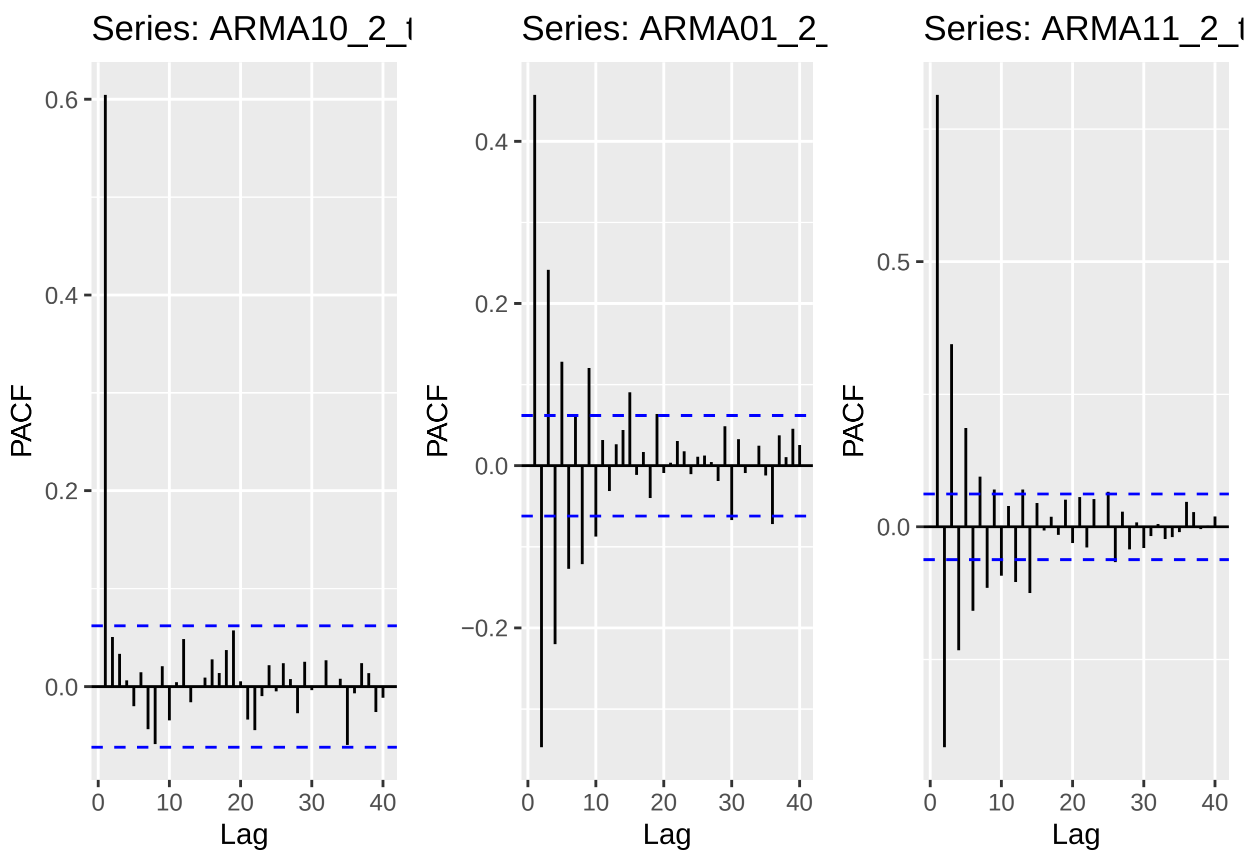
```
##
## Partial autocorrelations of series 'ARMA11_2_ts', by lag
##
##      1      2      3      4      5      6      7      8      9     10     11
##  0.815 -0.415  0.344 -0.233  0.187 -0.158  0.095 -0.115  0.070 -0.092  0.040
##     12     13     14     15     16     17     18     19     20     21     22
## -0.104  0.070 -0.124  0.045 -0.007  0.019 -0.015  0.051 -0.030  0.056 -0.039
##     23     24     25     26     27     28     29     30     31     32     33
```

```
##  0.052  0.001  0.066 -0.067  0.029 -0.043  0.008 -0.040 -0.017  0.006 -0.023
##     34     35     36     37     38     39     40
## -0.019 -0.010  0.047  0.028 -0.004  0.002  0.020
```

```r
#plotting the three graphs
ARMA10_2_ts_ACF_plot <- autoplot(ARMA10_2_ts_ACF)
ARMA01_2_ts_ACF_plot <- autoplot(ARMA01_2_ts_ACF)
ARMA11_2_ts_ACF_plot <- autoplot(ARMA11_2_ts_ACF)
plot_grid(ARMA10_2_ts_ACF_plot, ARMA01_2_ts_ACF_plot, ARMA11_2_ts_ACF_plot, nrow=1)
```




```r
#plotting the three graphs
ARMA10_2_ts_PACF_plot <- autoplot(ARMA10_2_ts_PACF)
ARMA01_2_ts_PACF_plot <- autoplot(ARMA01_2_ts_PACF)
ARMA11_2_ts_PACF_plot <- autoplot(ARMA11_2_ts_PACF)
plot_grid(ARMA10_2_ts_PACF_plot, ARMA01_2_ts_PACF_plot, ARMA11_2_ts_PACF_plot, nrow=1)
```



Analysis : The ACFs and PACFs are a lot more clear when n=1000. ARMA(1,0) has a decaying ACF and PACF cuts off at 1, indicating an AR model at lag=1. ARMA(0,1) has an ACF that cuts off at lag 1 and PACF that decays slowly indicating a moving average model of lag 1. Meanwhile, in ARMA(1,1), both ACF and PACF have a slow decay indicating both, an AR and MA component that is a lot more clean than when n=100. The PACF of ARMA(0,1) and (1,1) still don't match - as they shouldn't, given that the latter has an AR component that the former doesn't.

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.

Ans: This model is is a SARIMA(1,0,1)(0,0,1) ie p=1, d=0, q=1, P=0, D=0, Q=1.

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

Ans: $\phi_1 = 0.7$, $\theta_1 = 0.1$, $\theta_{12} = 0.25$

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?

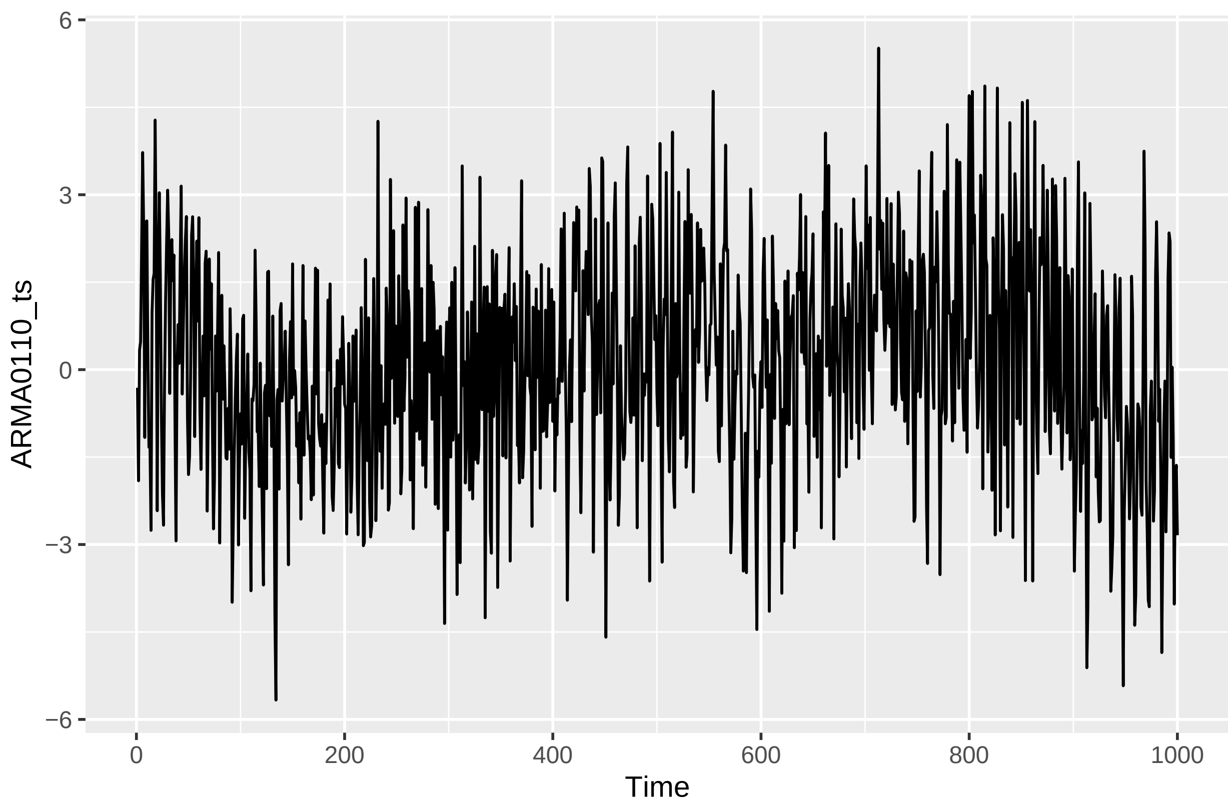
```r
#installing and loading the package
#install.packages("sarima")
library(sarima)
```

```
## Loading required package: stats4
```

```
##
## Attaching package: 'sarima'
```

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

```r
#Fitting the model
ARMA0110 <- sim_sarima(model = list(sar=0.8, ma=0.5, nseasons=12), n=1000)
ARMA0110_ts <- ts(ARMA0110)
autoplot(ARMA0110_ts)
```

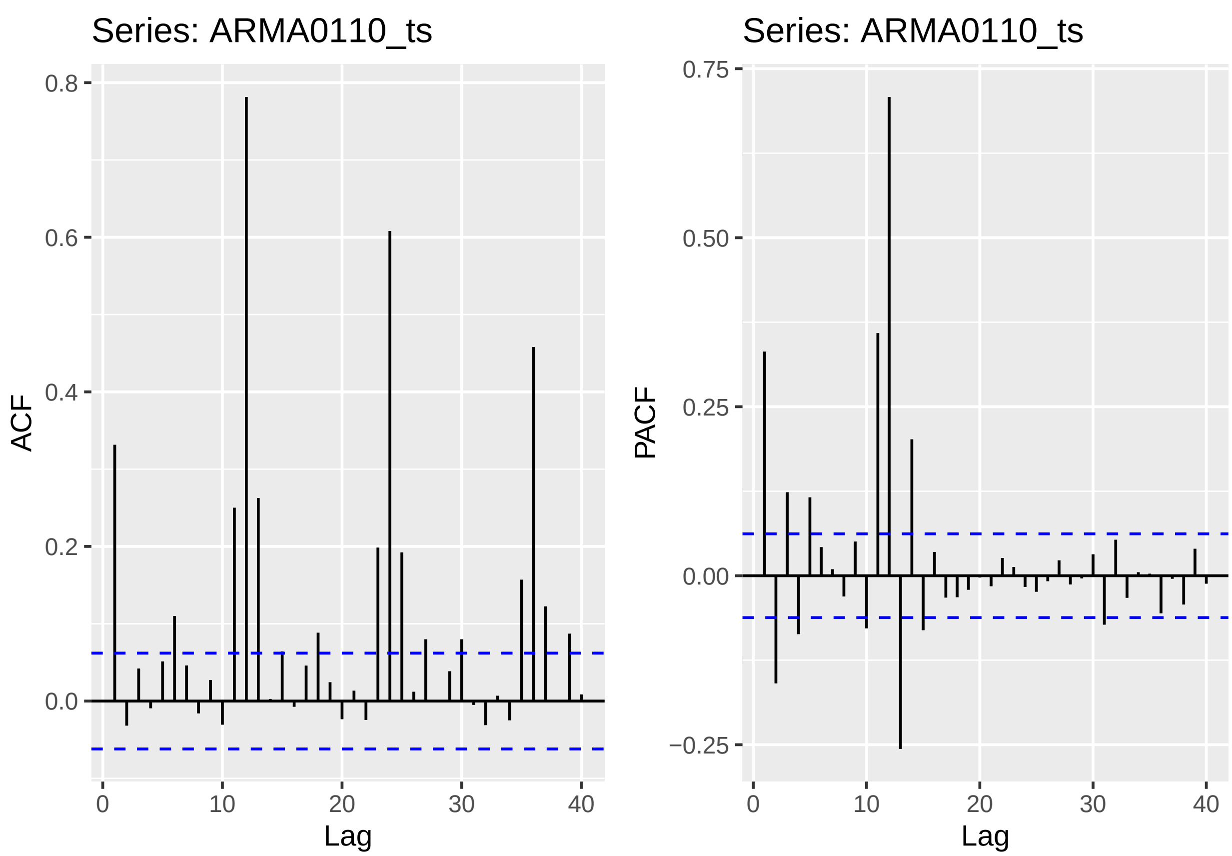


Answer: No, it does not seem to be seasonal.

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.

```
ARMA0110_ts_ACF <- Acf(ARMA0110_ts, lag.max=40, plot=FALSE)
ARMA0110_ts_PACF <- Pacf(ARMA0110_ts, lag.max=40, plot=FALSE)
ARMA0110_ts_ACF_plot <- autoplot(ARMA0110_ts_ACF)
ARMA0110_ts_PACF_plot <- autoplot(ARMA0110_ts_PACF)
plot_grid(ARMA0110_ts_ACF_plot, ARMA0110_ts_PACF_plot)
```



Analysis: The seasonal component of the model P=1, Q=0 is evident in the plots as the ACF shows significant spikes at all lags of 12 in a decaying fashion, whereas the PACF is signicant once at lag 12 and then cuts off. This indicates a strong seasonal autoregressive component (1,0). On the other hand, the stationary component of p=0 and q=1 is hard to tell because while the ACF cuts off after lag 1, the PACF also cuts off after lag 1 and does not properly showcase a decaying trend.