ENV 790.30 - Time Series Analysis for Energy Data | Spring 2021

Assignment 5 - Due date 03/12/21

Rajat Khandelwal

Directions

You should open the .rmd file corresponding to this assignment on RStudio. The file is available on our class repository on Github. And to do so you will need to fork our repository and link it to your RStudio.

Once you have the project open the first thing you will do is change "Student Name" on line 3 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. Rename the pdf file such that it includes your first and last name (e.g., "LuanaLima_TSA_A05_Sp21.Rmd"). Submit this pdf using Sakai.

Questions

This assignment has general questions about ARIMA Models.

Packages needed for this assignment: "forecast","tseries". Do not forget to load them before running your script, since they are NOT default packages.\

```
#Load/install required package here
library(forecast)
library(tseries)
library(stats)
library(sarima)
```

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:
> **ACF:** Exponential decay with increasing lag.
> **PACF:** Cut-off observed at lag = 2. Helps to identify the order of the AR model.

MA(1)

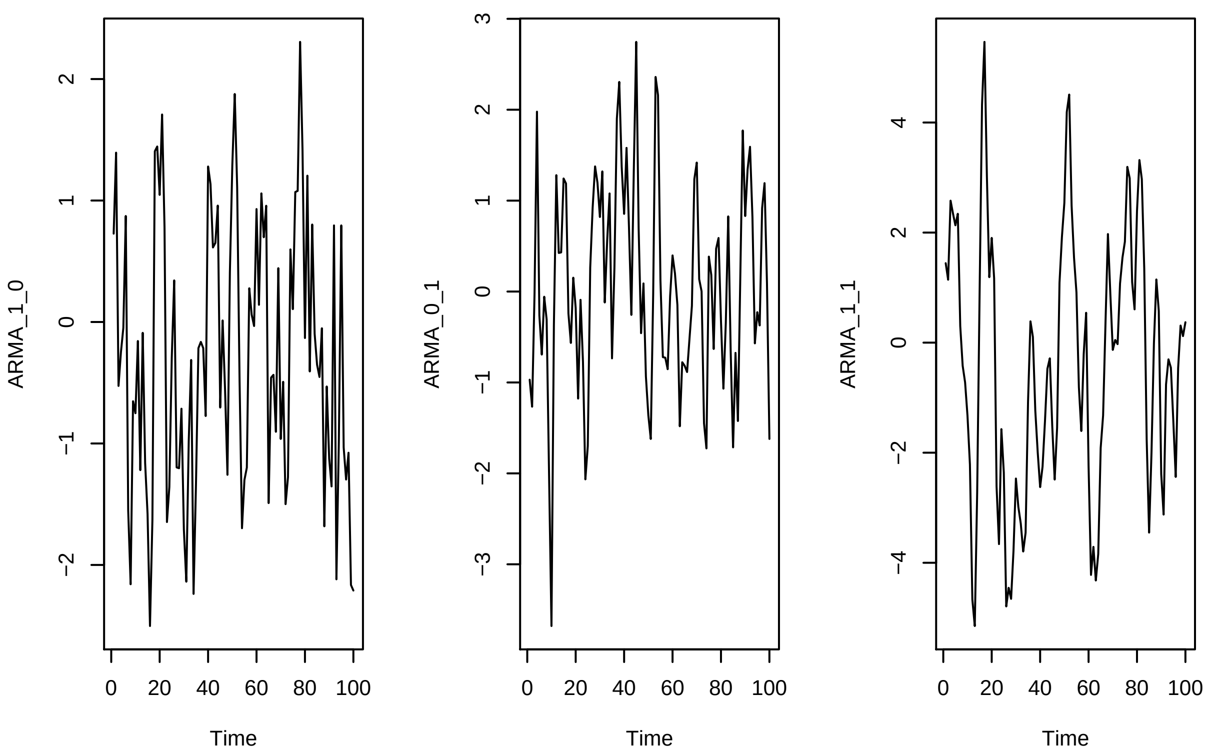
> Answer:
> **ACF:** Cut-off observed at lag = 1. Helps to identify the order of the MA model.
> **PACF:** Exponential delay with increasing lag.

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). 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 R to generate $n = 100$ observations from each of these three models

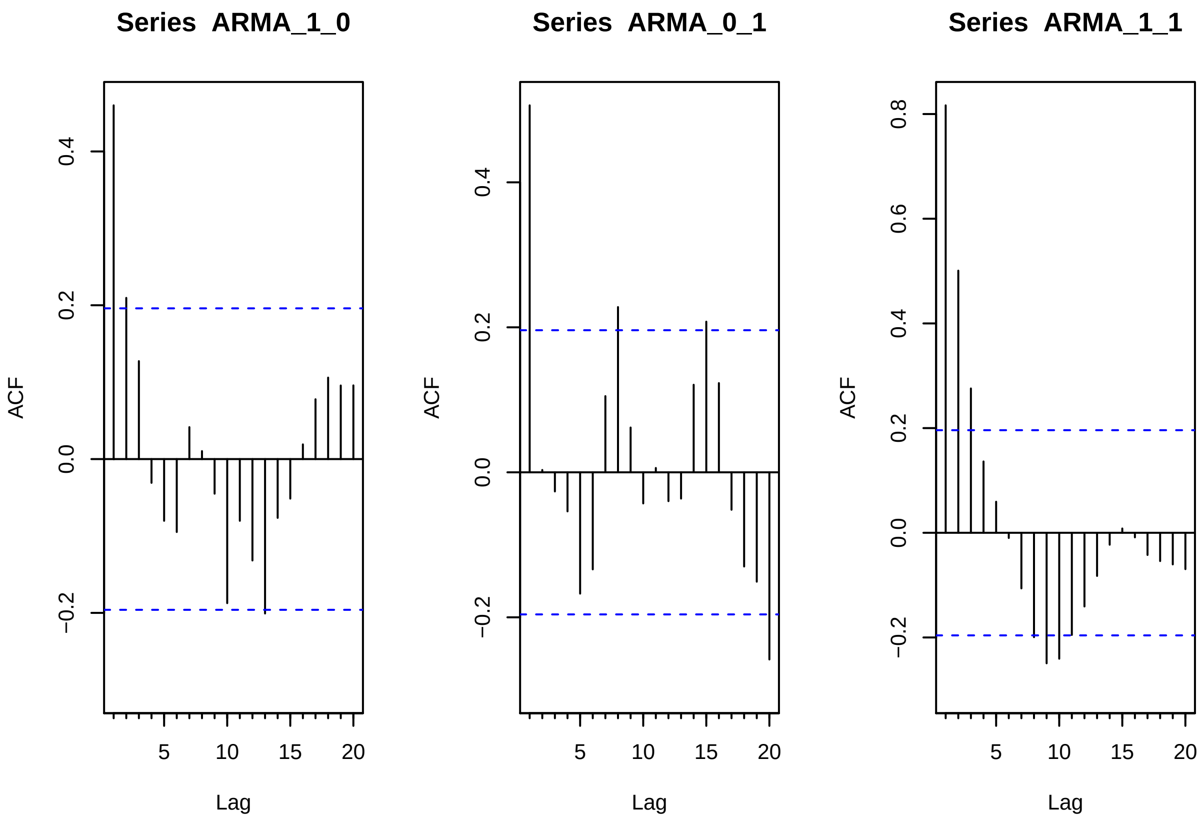
```
set.seed(999)
ARMA_1_0 <- arima.sim(n = 100,  list(order = c(1,0,0), ar = c(0.6)))
ARMA_0_1 <- arima.sim(n = 100,  list(order = c(0,0,1), ma = c(0.9)))
ARMA_1_1 <- arima.sim(n = 100,  list(order = c(1,0,1), ar = c(0.6), ma = c(0.9)))
```

```
par(mfrow = c(1,3))
ts.plot(ARMA_1_0)
ts.plot(ARMA_0_1)
ts.plot(ARMA_1_1)
```



Plot the sample ACF for each of these models in one window to facilitate comparison (Hint: use command $par(mfrow = c(1, 3))$ that divides the plotting window in three columns).

```
par(mfrow = c(1,3))
Acf(ARMA_1_0)
Acf(ARMA_0_1)
Acf(ARMA_1_1)
```



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

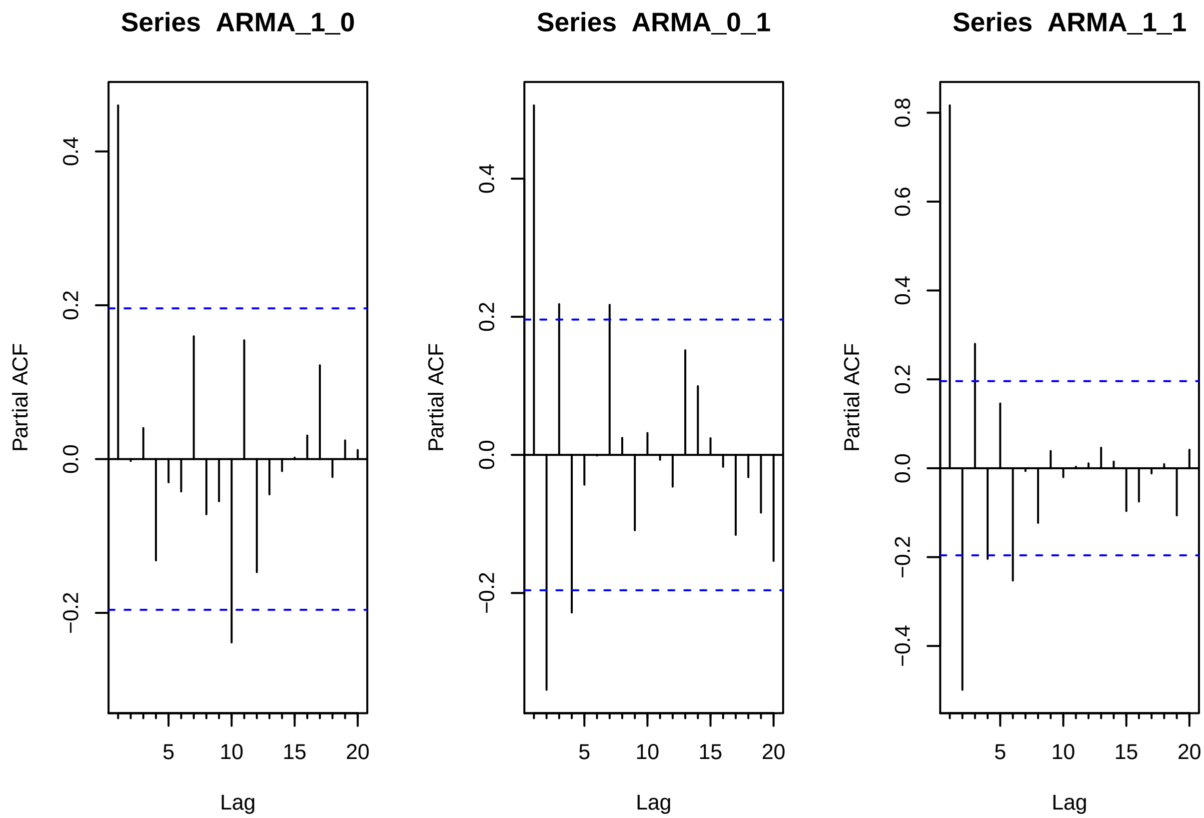
```r
par(mfrow = c(1,3))
lag_ARMA_1_0 <- Pacf(ARMA_1_0)
print(lag_ARMA_1_0[1])
```

```
##
## Partial autocorrelations of series 'ARMA_1_0', by lag
##
##    1
## 0.46
```

```r
lag_ARMA_0_1 <- Pacf(ARMA_0_1)
print(lag_ARMA_0_1[1])
```

```
##
## Partial autocorrelations of series 'ARMA_0_1', by lag
##
##     1
## 0.506
```

```r
lag_ARMA_1_1 <- Pacf(ARMA_1_1)
```




```
print(lag_ARMA_1_1[1])
```

```
##
## Partial autocorrelations of series 'ARMA_1_1', by lag
##
##     1
## 0.817
```

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 identify them correctly? Explain your answer.

Answer:

Simply based on the ACFs and PACFs, I note that:

Series_ARMA_1_0: Since the ACF is exponentially decreasing and the value at lag 1 is positive, it appears to be an Auto-Regressive (AR) process. As can be seen from the time-series plot, we observe long memory tails which indicate an AR process. Further, since the PACF cut-offs at lag = 1, this time-series looks like an AR process with order = 1, i.e. AR(1).

Series_ARMA_0_1: Since the PACF is exponentially decreasing, it appears to be a Moving Average (MA) process. As can be seen from the time-series plot, we observe short memory tails which indicate a MA process. Further, since the ACF cut-offs at lag = 1, this time-series looks like a MA process with order = 1, i.e. MA(1).

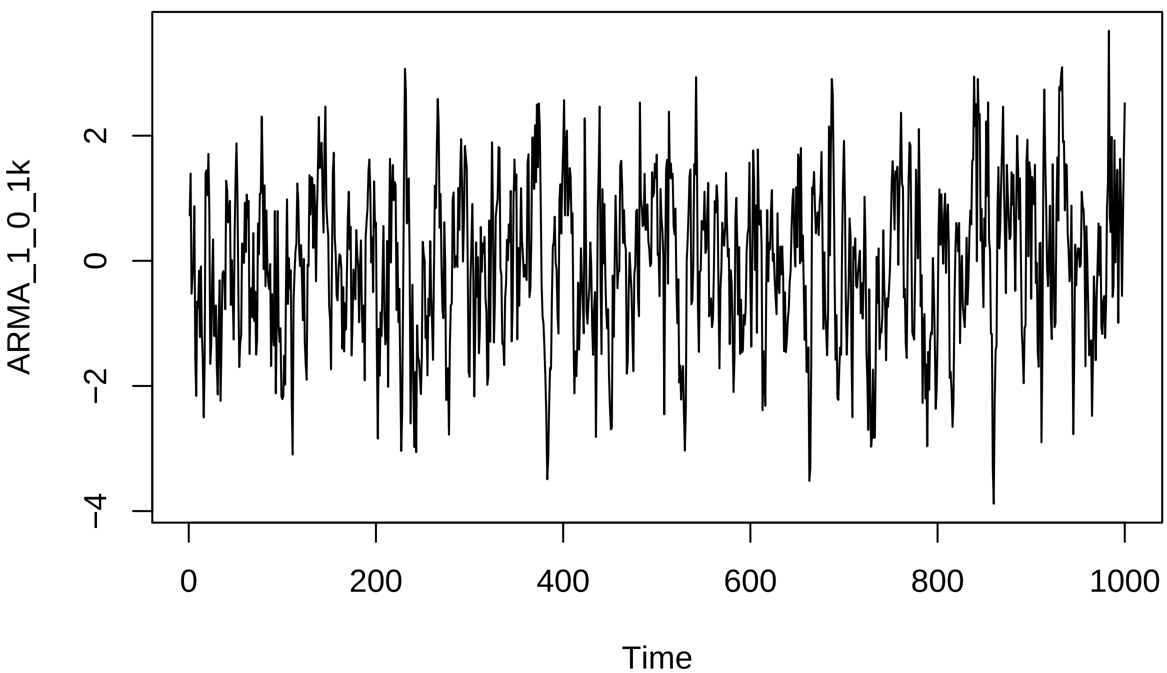
Series_ARMA_1_1: Since the ACF and PACF are exponentially decreasing, it appears to be an Auto-Regressive Moving Average (ARMA) process.Further, since the ACF cut-offs at lag = 3 and PACF cuts off at lag = 4, this time-series looks like a ARMA process with AR order = 4 and MA order = 3, i.e. ARMA(4,3). This observation is in contradiction with how we generated the series to be an ARMA(1,1) process. This illustrates the difficulty in finding the process orders graphically for ARMA processes.

Compare the ACF and PACF values R computed with the theoretical values you provided for the coefficients. Do they match? Explain your answer.

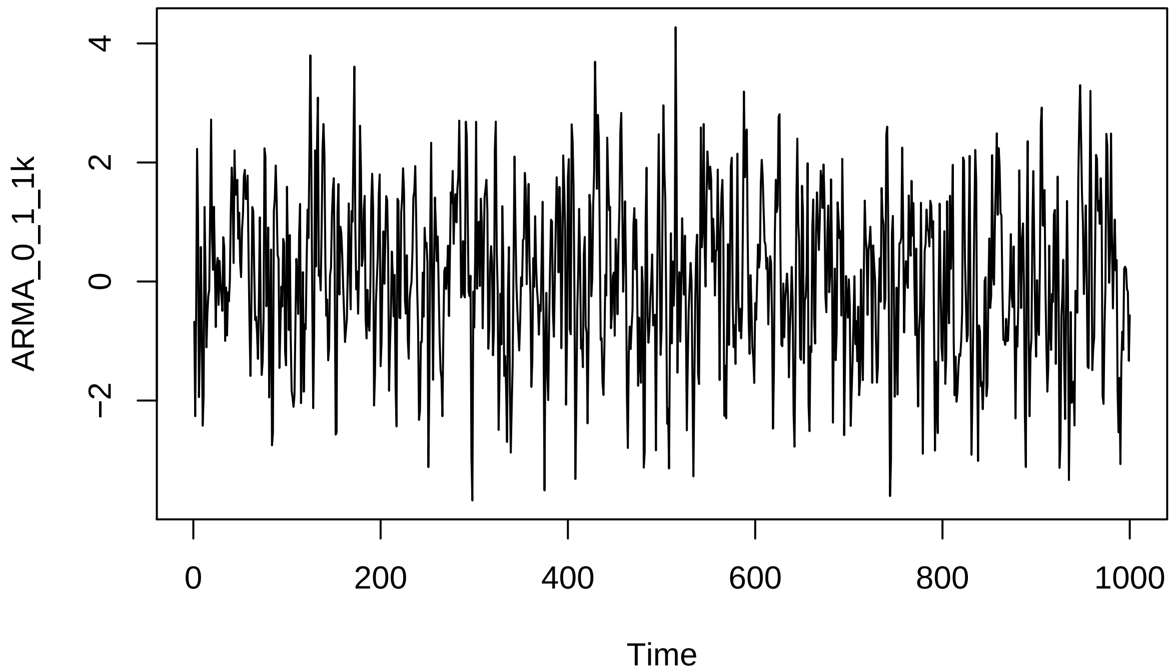
> Answer:
> We observe that the PACF lag 1 coefficient for the ARMA_1_0 process is 0.46 which is not equal to the value of $\phi = 0.6$ set by us.
>
> For the ARMA_1_1 process, the PACF lag 1 coefficient is 0.743 which is not equal to the value of $\phi = 0.6$ set by us.
>
> We can't compare the values of ACF coefficients with the value of $\theta = 0.8$ since θ is the coefficient of the dependence of an observation (say y_t) on a previous **deviation from the mean,** a_{t-1} and not the actual previous observation y_{t-1}.

Increase number of observations to $n = 1000$ and repeat parts (a)-(d).

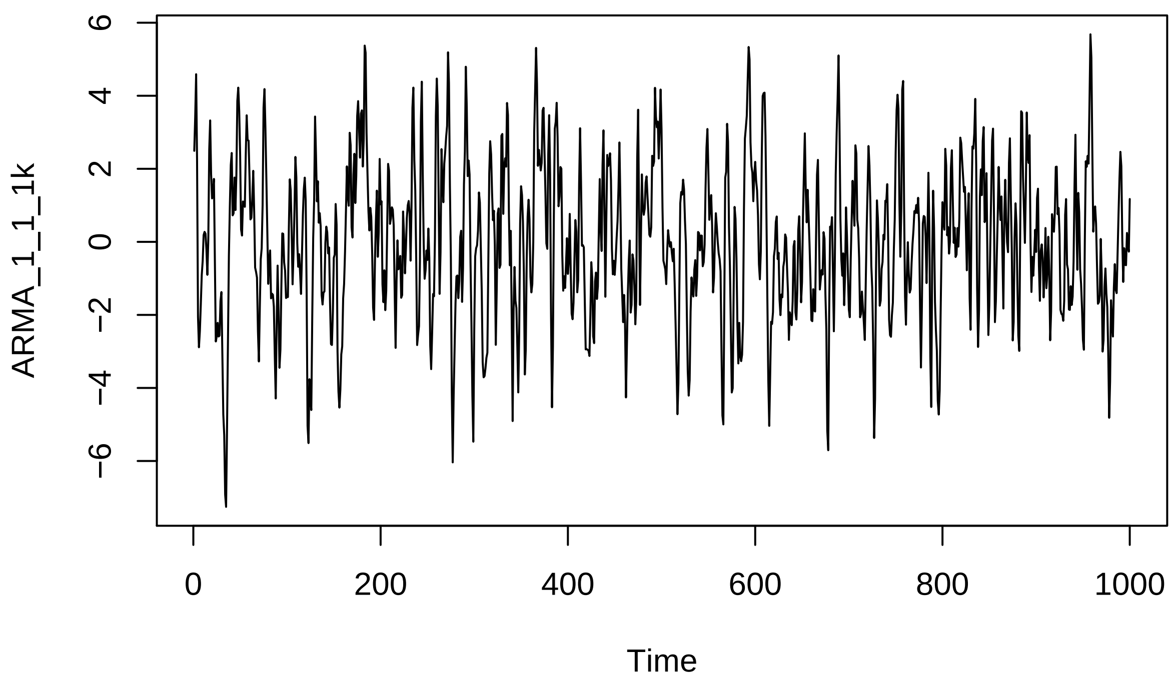
```
set.seed(999)
ARMA_1_0_1k <- arima.sim(n = 1000,  list(order = c(1,0,0), ar = c(0.6)))
ARMA_0_1_1k <- arima.sim(n = 1000,  list(order = c(0,0,1), ma = c(0.9)))
ARMA_1_1_1k <- arima.sim(n = 1000,  list(order = c(1,0,1), ar = c(0.6), ma = c(0.9)))
ts.plot(ARMA_1_0_1k)
```



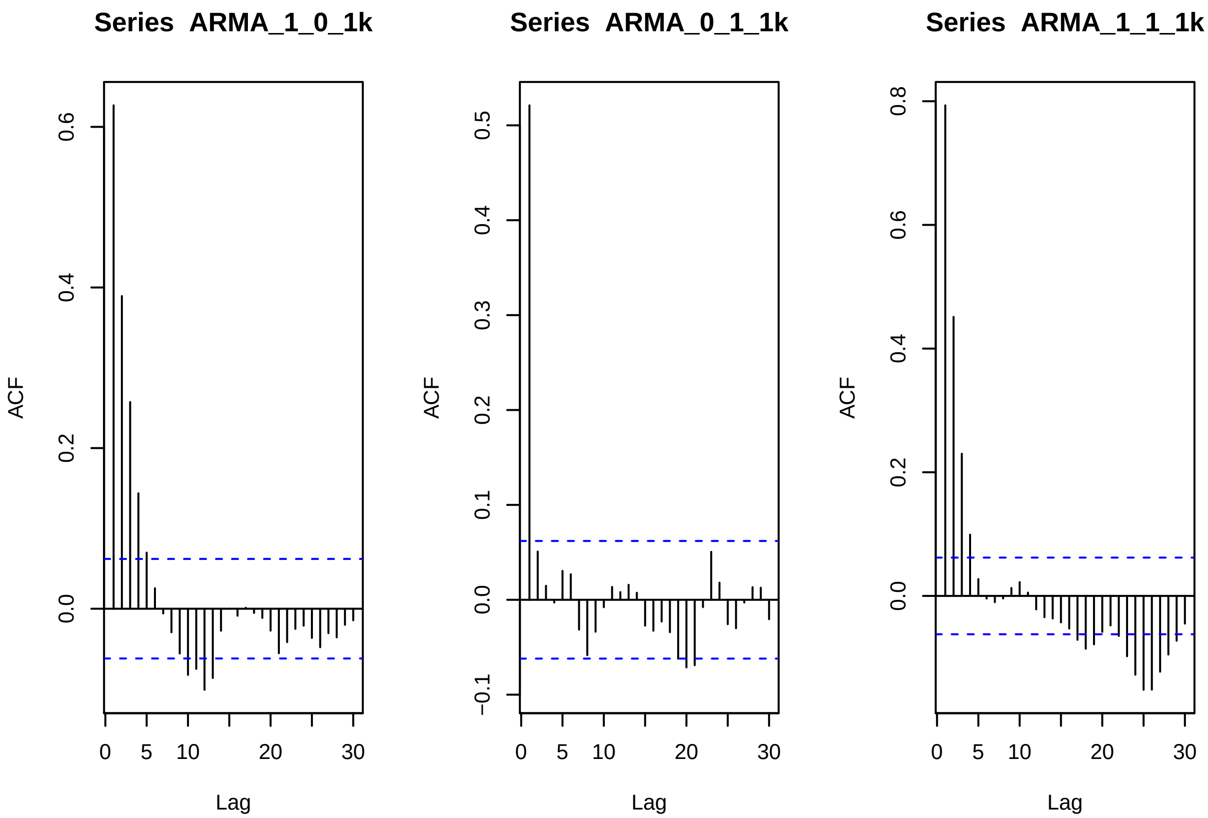

```
ts.plot(ARMA_0_1_1k)
```




```
ts.plot(ARMA_1_1_1k)
```




```
par(mfrow = c(1,3))
Acf(ARMA_1_0_1k)
Acf(ARMA_0_1_1k)
Acf(ARMA_1_1_1k)
```



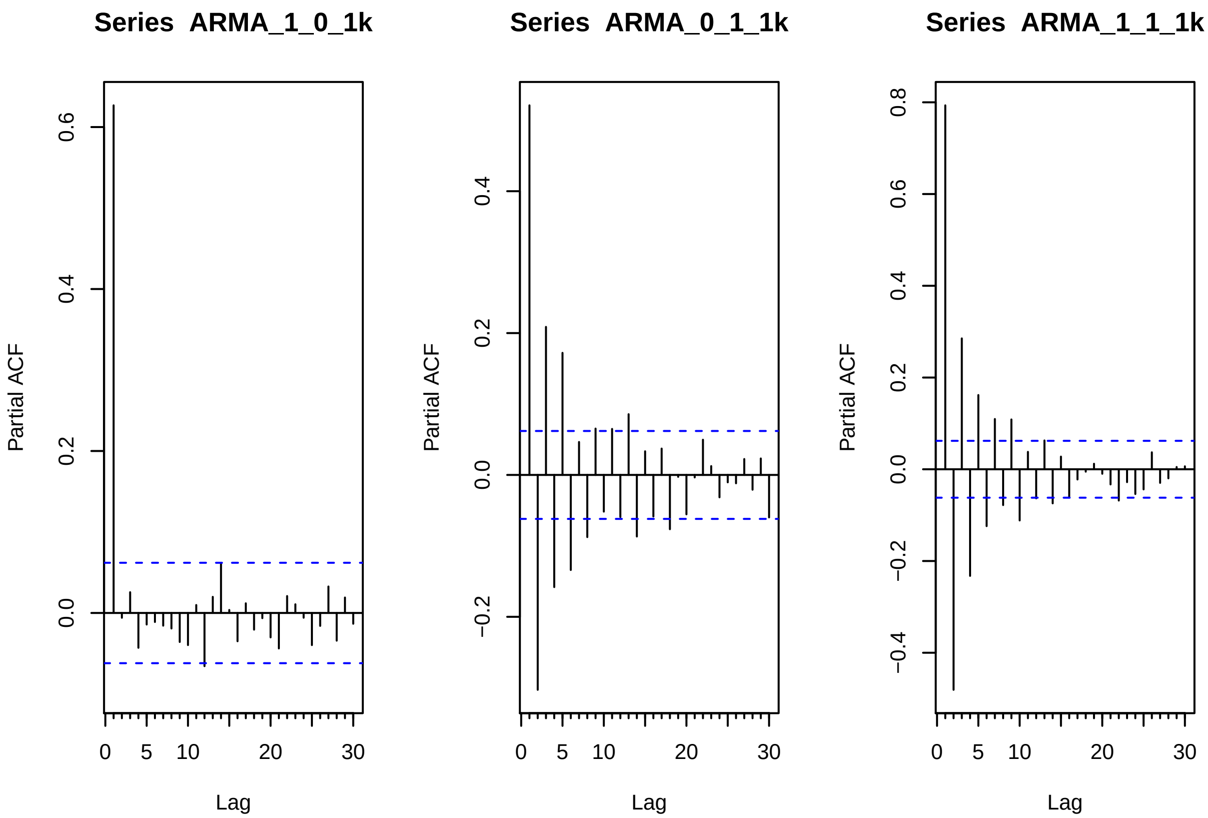

```
par(mfrow = c(1,3))
lag_ARMA_1_0_1k <- Pacf(ARMA_1_0_1k)
print(lag_ARMA_1_0_1k[1])
```

```
##
## Partial autocorrelations of series 'ARMA_1_0_1k', by lag
##
##     1
## 0.627
```

```
lag_ARMA_0_1_1k <- Pacf(ARMA_0_1_1k)
print(lag_ARMA_0_1_1k[1])
```

```
##
## Partial autocorrelations of series 'ARMA_0_1_1k', by lag
##
##     1
## 0.521
```

```
lag_ARMA_1_1_1k <- Pacf(ARMA_1_1_1k)
```




```
print(lag_ARMA_1_1_1k[1])
```

```
##
## Partial autocorrelations of series 'ARMA_1_1_1k', by lag
##
##     1
## 0.793
```

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 identify them correctly? Explain your answer.

Simply based on the ACFs and PACFs, I would have commented that:

Series_ARMA_1_0_1k: Since the ACF is exponentially decreasing and the value at lag 1 is positive, it appears to be an Auto-Regressive (AR) process. As can be seen from the time-series plot, we observe long memory tails which indicate an AR process. Further, since the PACF cut-offs at lag = 1, this time-series looks like an AR process with order = 1, i.e. AR(1).

Series_ARMA_0_1_1k: Since the PACF is exponentially decreasing, it appears to be a Moving Average (MA) process. As can be seen from the time-series plot, we observe short memory tails which indicate a MA process. Further, since the ACF cut-offs at lag = 1, this time-series looks like a MA process with order = 1, i.e. MA(1).

Series_ARMA_1_1_1k: Since the ACF and PACF are exponentially decreasing, it appears to be an Auto-Regressive Moving Average (ARMA) process.Further, since the ACF cut-offs at lag = 4 and PACF cuts off at lag = 10, this time-series looks like a ARMA process with AR order = 10 and MA order = 4, i.e. ARMA(10,4). Such a series is unlikely. This observation is in contradiction with how we generated the series to be an ARMA(1,1) process. This illustrates the difficulty in finding the process orders graphically for ARMA processes.

Compare the ACF and PACF values R computed with the theoretical values you provided for the coefficients. Do they match? Explain your answer.

> Answer:
> We observe that the PACF lag 1 coefficient for the ARMA_1_0 process is 0.627 which is closer equal to the value of $\phi = 0.6$ set by us as compared to case when n = 100. This shows that on increasing the number of observations from n = 100 to n = 1000, we are able to obtain a time-series which is more in accordance with the desired input parameters. Simply put, having more data points in the series gives us a better desired fit to the inputs.
>
> For the ARMA_1_1 process, the PACF lag 1 coefficient is 0.811 which is not equal to the value of $\phi = 0.6$ set by us.
>
> We still can't compare the values of ACF coefficients with the value of $\theta = 0.8$ since θ is the coefficient of the dependence of an observation (say y_t) on a previous **deviation from the mean,** a_{t-1} and not the actual previous observation y_{t-1}.

Q3

Consider the ARIMA model $y_t = 0.7 * y_{t-1} - 0.25 * y_{t-12} + a_t - 0.1 * a_{t-1}$

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 : Not possible
q = 1
P = 1
D : Not possible
Q = 0
s = 12

Note: It is not possible to find out the values of d and D since we the given relationship is for y_t and not a differenced series such as $y_t - y_{t-1}$ or $y_t - y_{t-12}$.

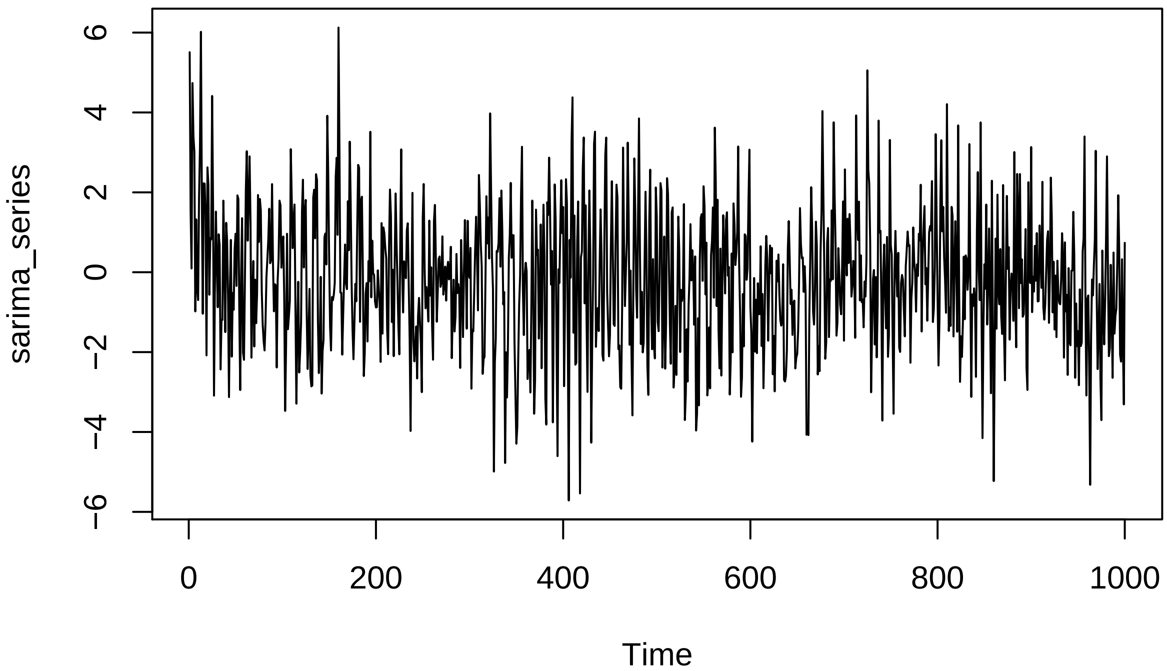
Also from the equation what are the values of the parameters, i.e., model coefficients.

$\phi_1 = 0.7$
$\phi_{12} = -0.25$
$\theta_1 = -0.1$

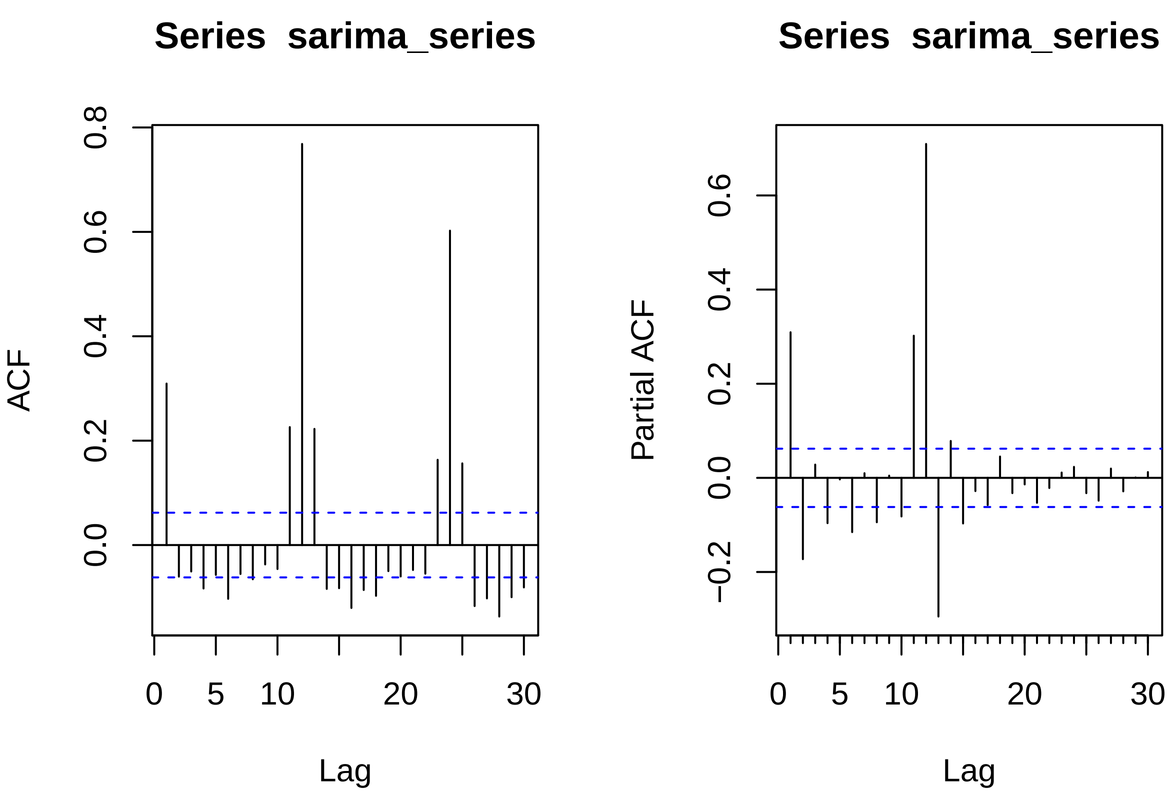
Q4

Plot the ACF and PACF of a seasonal ARIMA$(0, 1) \times (1, 0)_{12}$ model with $\phi = 0.8$ and $\theta = 0.5$ using R. 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 ACF and PACF for the simulated data. 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.

```r
sarima_series <- sim_sarima(n=1000, model = list(ma=0.5, sar = 0.8, nseasons = 12))
ts.plot(sarima_series)
```




```r
par(mfrow = c(1,2))
Acf(sarima_series)
Pacf(sarima_series)
```



On observing the ACF and PACF for the given model, we note that: There seems to be a sesasonality with frequency = 12. Therefore, s = 12.

PACF is decreasing exponentially in the initial few lags, which is indicative of a MA process. ACF cuts-off at a lag of 1, indicating that the MA process order = 1 Therefore, q = 1.

Further, we observe positive spikes in the ACF at lags = 12, 24, 36 which are accompanied with a positive spike at lag = 12, which is a strong indicator of a Sesasonal AR process or order = 1. Therefore, P = 1.

The plots do not give any indication of an AR or seasonal MA process, hence p = Q = 0.

The series look well-differenced as the ACF does not show postive values upto a high number of lags and hence we can conclude it does not need more differencing and therefore, d = D = 0.

Therefore, based on the ACF and PACF alone, I would comment that the series is:

$SARIMA(0,0,1)(1,0,0)_{[12]}$

Which matches with it's actual nature as given the question prompt. Hence, the plot seems to be well-matching the simulated model on using n = 1000.