Spring Semester 2023-2024

MATH3026/4022: Time Series Analysis/Time Series and Forecasting
Computer Practical 2 - SOLUTIONS

During this practical, we will use R by loading RStudio and running commands. When writing R commands, it is good practice to write comments so that you understand your code.

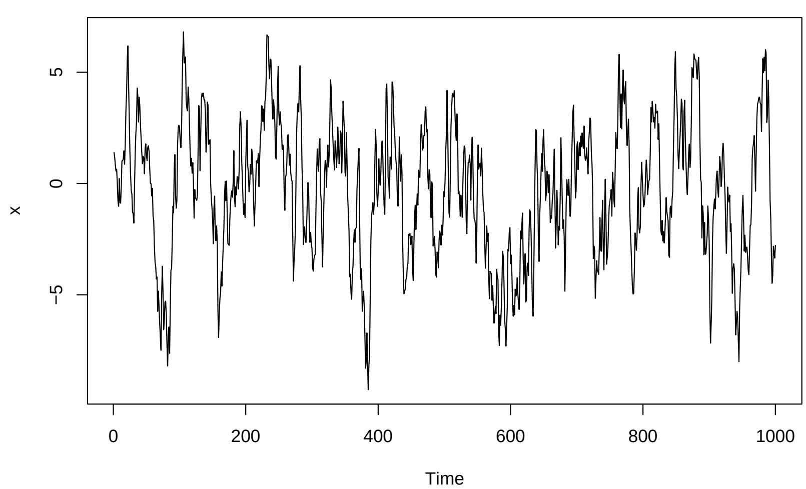
Before beginning the practical, it is recommended that you create a folder named `MATH3026` or `MATH4022` within your `Documents` folder on your University workspace. Then, within your `MATH3026` or `MATH4022` folder, create a new folder called `Practical2` in which you can save all of the material covered during this practical session.

You should set your working directory to be the `Practical2` folder that you've created. In RStudio, this can be done by selecting the `Session` menu and then `Set Working Directory` and `Choose Directory`, navigating to the correct folder. You can then easily save your R code and plots in this directory.

1. (b) Run the following code to produce a time plot of the simulated data from (a).

```
#Produce a time plot for the simulated data 'x'
ts.plot(x)
```

SOLUTION:



(c) Run the following code to produce a plot of the sample ACF and sample PACF for the simulated data from (a), for lags up to $h = 30$. Note the command `par(mfrow=c(1,2))` puts the plots in the same window.

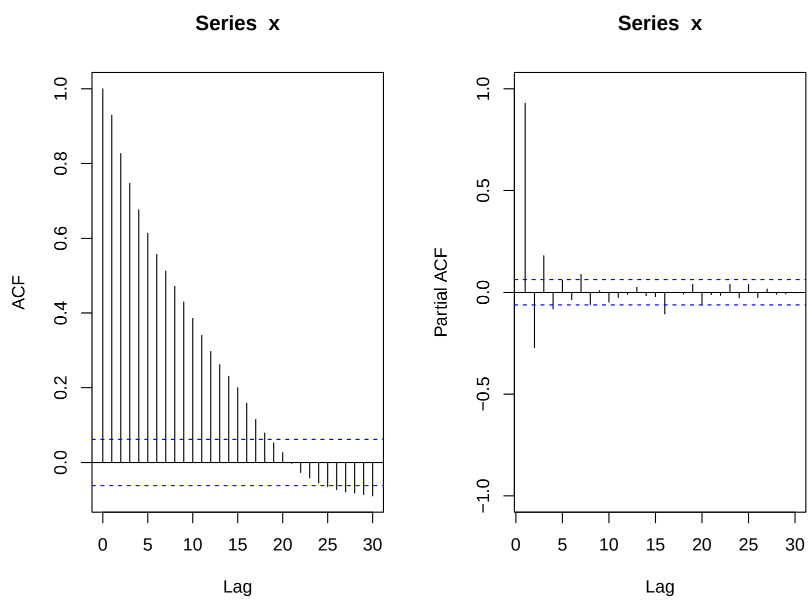
SOLUTION:

```
#Set the R graphics device to contain two plots (1 row, 2 columns)
par(mfrow=c(1,2))

#Plot the sample ACF
acf(x)

#Plot the sample PACF
#ylim=c(-1,1) ensures that the vertical axis runs from -1 to 1.

pacf(x,ylim=c(-1,1))
```



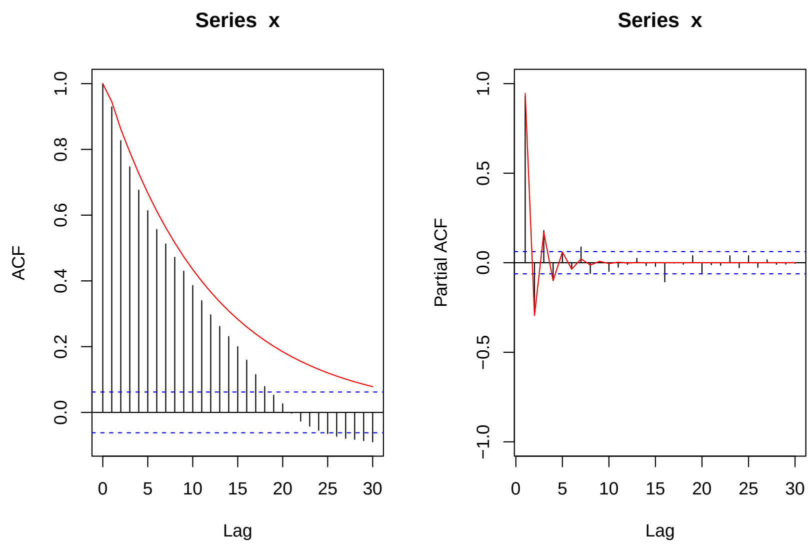
Do these plots appear consistent with those you would expect for an $ARMA(2, 1)$ process? Justify your answer.

SOLUTION:

The ACF does not cut off to zero after a particular lag which suggests that there is an AR component to the process. The sample PACF does not cut off after a particular lag, so this would not suggest that the process is not a pure AR process (i.e. that there is an MA component present).

(e) To compare the theoretical ACF and PACF to the sample ACF and PACF, run the following commands to produce plots of the sample ACF and PACF with the theoretical versions included on the plots as red lines.

SOLUTION:



Are the sample ACF and PACF close to the theoretical versions for all lags?

SOLUTION:

The sample PACFs are very close to the true values for many of the lags. The sample ACF is the correct shape but does not approximate the theoretical ACF exactly. If the sample size was larger, we'd expect the sample ACF to lie closer to the theoretical ACF for all lags.

2. (a) Using the code from 1. as a guide, simulate 10000 observations from the $ARMA(1,1)$ process with $\phi_1 = -0.8, \theta_1 = 0.4$ and $\sigma_z^2 = 0.64$. Remember to save and comment your code.

SOLUTION:

The code to simulate the data (using a seed of 215) would be

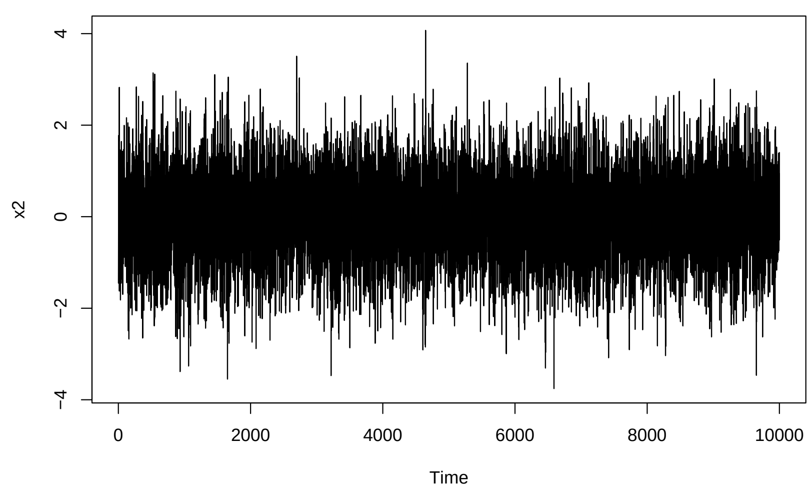
```
set.seed(215)
x2<-arima.sim(n=10000,model=list(ar=c(-0.8),ma=c(0.4)),sd=0.8)
```

(b) For the data simulated in (a), produce:

(i) A time plot.

SOLUTION:

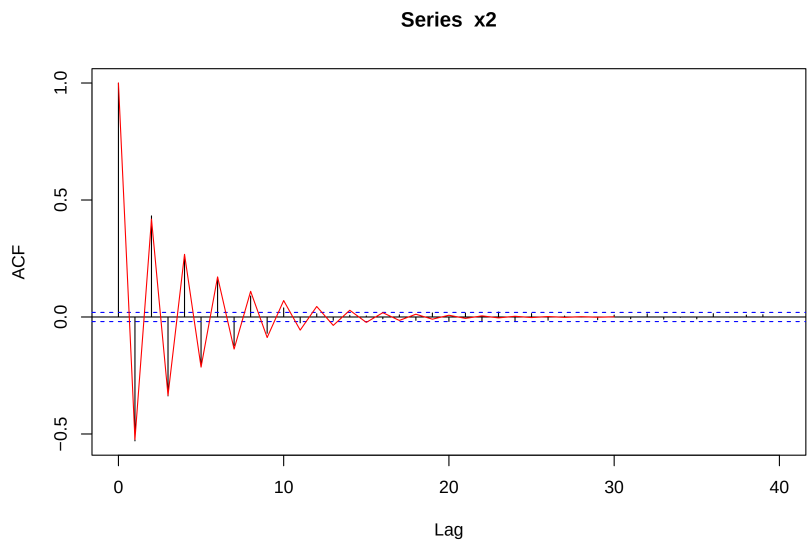
```
ts.plot(x2)
```



(ii) A plot of the sample ACF with the theoretical ACF overlayed on the plot, for lags up to 30.

SOLUTION:

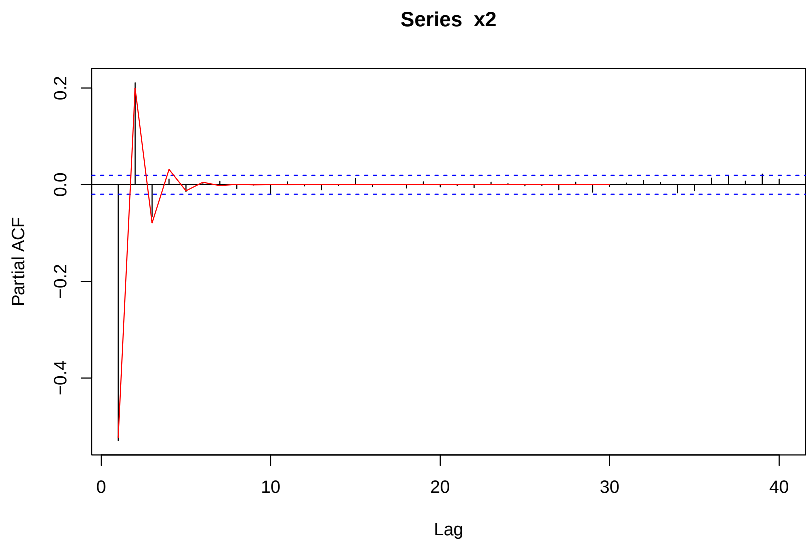
```
#Calculate the theoretical ACF
t_acf_x2<-ARMAacf(ar=c(-0.8),ma=c(0.4),lag.max=30)
```



(iii) A plot of the sample PACF with the theoretical PACF overlayed on the plot, for lags up to 30.

SOLUTION:

```
#Calculate the theoretical PACF
t_acf_x2<-ARMAacf(ar=c(-0.8),ma=c(0.4),lag.max=30,pacf=TRUE)
```



3. The dataset `practical2.csv` can be found within the Week 24 section of the course moodle page. This dataset contains 2000 observations of a time series in one column named `y`.

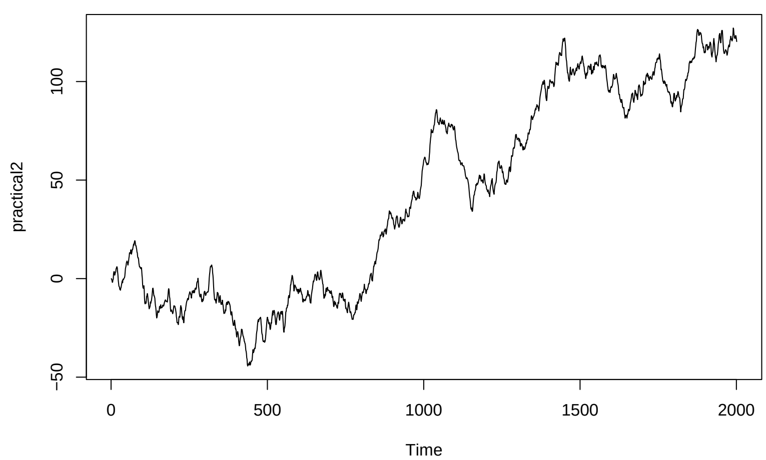
   Download this dataset from moodle and save it in your `Practical2` folder. Ensuring that the working directory in RStudio is set to your `Practical2` folder (see instructions at the beginning of the practical sheet for guidance on how to do this), run the following command to read the dataset into RStudio and declare the data to be time series data:

```
#Read in the dataset "practical2"
practical2<-read.csv("practical2.csv",header=TRUE)

#Declare the dataset to be time series data
practical2<-ts(practical2$y,start=1,end=2000)
```

   (a) Produce a time plot of the data. Does this time series appear to be (weakly) stationary? Explain your answer.

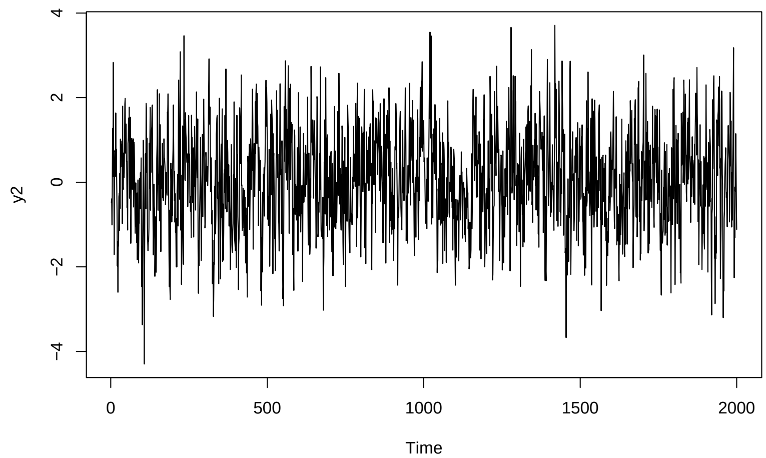
   SOLUTION:



   The time plot indicates that the time series is not stationary.

(c) Produce a time plot for the differenced series `y2`. Does the differenced time series appear to be stationary?

SOLUTION:

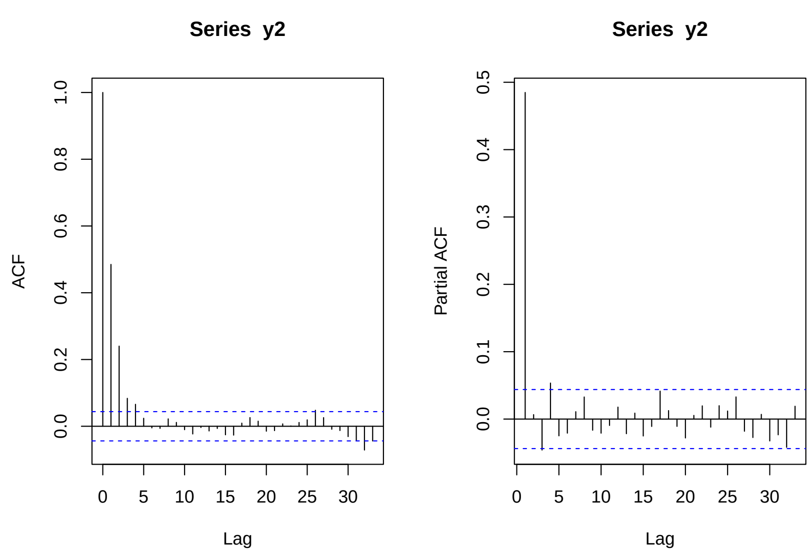


Yes, the differenced process appears to be stationary.

(d) Using plots of the sample ACF and sample PACF, suggest an $ARMA$ model that could be fitted to the data `y2`. Justify your answer. Approximately, what might be the parameter value(s) for your model?

SOLUTION:

A plot of the ACF and PACF is shown below.



The plot of the PACF suggests that the process only has an AR component because the PACF appears to decrease to zero after lag 1. This property suggests that the order of the AR component is equal to 1. With this in mind, and knowing that the ACF for an AR(1) process at lag h is $\phi^{|h|}$, the peak at 0.5 on the sample ACF at lag 1, might suggest that $\phi = 0.5$.

we know that, for an AR(1) process, the variance of the process is

$$\frac{\sigma_z^2}{1-\phi^2} = \frac{4\sigma_z^2}{3} \quad (\text{if } \phi = 0.5)$$

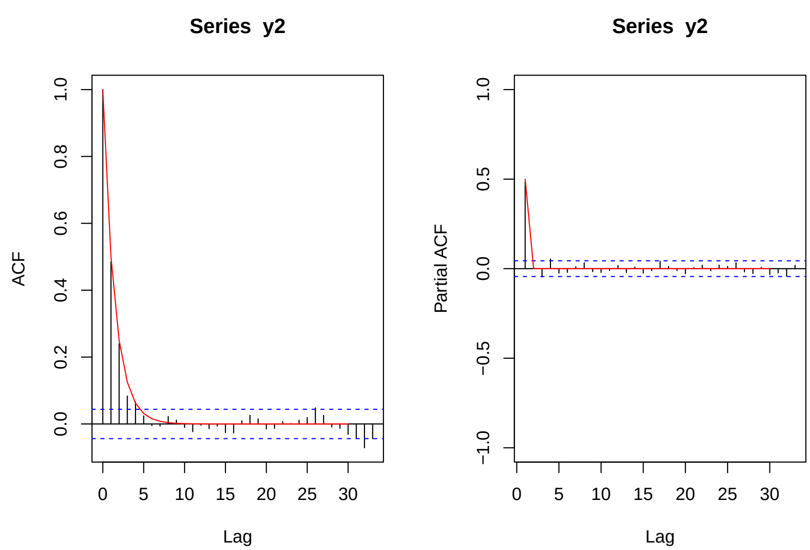
To estimate σ_z^2 (approximately) we calculate the sample variance of `y2`, s_{y2}^2, and solve

$$\frac{4\sigma_z^2}{3} = s_{y2}^2$$

The sample variance is $s_{y2}^2 = 1.365268$ and, hence

$$\sigma_z^2 \approx \frac{3 \times 1.365268}{4} = 1.024.$$

To check this, we produce another plot of the sample ACF and PACF, with the theoretical equivalents for the AR(1) process with $\phi = 0.5$ added as red lines (shown below).



We see that an AR(1) would appear to fit the differenced data well.