Spring Semester 2023–2024

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

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. (a) Use the following code to simulate 1000 observations from an $ARMA(2,1)$ process, with parameters $\phi_1 = 0.7, \phi_2 = 0.2, \theta_1 = 0.6$ and $\sigma_z^2 = 1$.

```
#Set the seed (1 can be replaced with a number of your choice)
set.seed(1)

#Simulate 1000 observations from the ARMA(2,1) process with
#phi1=0.7, phi2=0.2, theta1=0.6 and sigma_z=1
x<-arima.sim(n=1000,model=list(ar=c(0.7,0.2),ma=c(0.6)),sd=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)
```

(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.

```
#Set the R graphics device to contain two plots (1 row, 2 columns)
x11()
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.

(d) Now we will compute the theoretical ACF and PACF for the $ARMA(2,1)$ process simulated in (a) and compare these to the sample plots obtained in (c). Run the following commands to compute the theoretical ACF and PACF for the $ARMA(2,1)$ process up to lag 30.

```
#Calculate the theoretical ACF for the ARMA(2,1) process from (a)
#call this t_acf

t_acf<-ARMAacf(ar=c(0.7,0.2),ma=c(0.6),lag.max=30)

#Calculate the theoretical PACF for the ARMA(2,1) process from (a)
#call this t_pacf

t_pacf<-ARMAacf(ar=c(0.7,0.2),ma=c(0.6),lag.max=30,pacf=TRUE)
```

(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.

```
#Set the R graphics device to contain two plots (1 row, 2 columns)
x11()
par(mfrow=c(1,2))

#Plot the sample ACF
acf(x)

#Add the theoretical ACF to this plot as a red line
lines(c(0:30),t_acf,col="red")

#Plot the sample PACF
pacf(x,ylim=c(-1,1))

#Add the theoretical PACF to this plot as a red line
lines(c(1:30),t_pacf,col="red")
```

Are the sample ACF and PACF close to the theoretical versions 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.

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

      (i) A time plot.

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

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

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.

   (b) Using the following command, take the first difference of the `practical2` dataset and call the resulting dataset `y2`.

```
#Take the first difference of the dataset "practical2"
y2<-diff(practical2)
```

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