Tutorial set #4

Question 1:

The following data represent the monthly sales (in thousand riyals) for a particular electrical appliance (read the data across from left to right).

53	43	66	48	52	42	44	56
44	58	41	54	51	56	38	56
49	52	32	59	34	57	39	60
41	52	43					

1- Plot the data, and comment on the stationarity of the data.

2- Based on the figure, can you say anything about the approximate value of the autocorrelation coefficient ρ_1?

3- Plot y_t against y_{t-1}, try to guess the value of ρ_1.

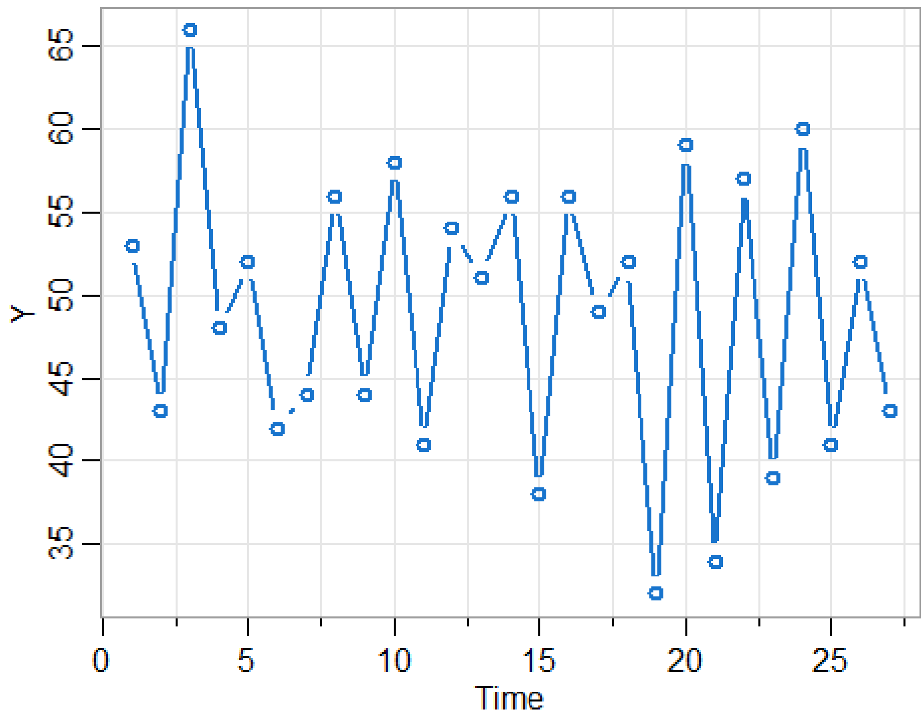
4- Find and plot the sample autocorrelation function r_k for $k = 0,1,2,3,4,5$. Comment on the shape of this function.

5- Find and plot the sample partial autocorrelation function r_{kk} for $k = 0,1,2,3,4,5$. Comment on the shape of this function.

↓	C1	C2	C3	C4	C5	C6
	y(t)	y(t)-y(t-1)	y(t-1)	AVER1	FITS1	RESI1
Entry Direction	3	*	*	*	*	*
2	43	-10	53	*	*	*
3	66	23	43	54.0000	*	*
4	48	-18	66	52.3333	54.0000	-6.0000
5	52	4	48	55.3333	52.3333	-0.3333
6	42	-10	52	47.3333	55.3333	-13.3333
7	44	2	42	46.0000	47.3333	-3.3333
8	56	12	44	47.3333	46.0000	10.0000
9	44	-12	56	48.0000	47.3333	-3.3333
10	58	14	44	52.6667	48.0000	10.0000
11	41	-17	58	47.6667	52.6667	-11.6667
12	54	13	41	51.0000	47.6667	6.3333
13	51	-3	54	48.6667	51.0000	0.0000
14	56	5	51	53.6667	48.6667	7.3333
15	38	-18	56	48.3333	53.6667	-15.6667
16	56	18	38	50.0000	48.3333	7.6667
17	49	-7	56	47.6667	50.0000	-1.0000
18	52	3	49	52.3333	47.6667	4.3333
19	32	-20	52	44.3333	52.3333	-20.3333
20	59	27	32	47.6667	44.3333	14.6667
21	34	-25	59	41.6667	47.6667	-13.6667
22	57	23	34	50.0000	41.6667	15.3333
23	39	-18	57	43.3333	50.0000	-11.0000
24	60	21	39	52.0000	43.3333	16.6667
25	41	-19	60	46.6667	52.0000	-11.0000
26	52	11	41	51.0000	46.6667	5.3333
27	43	-9	52	45.3333	51.0000	-8.0000

R Code

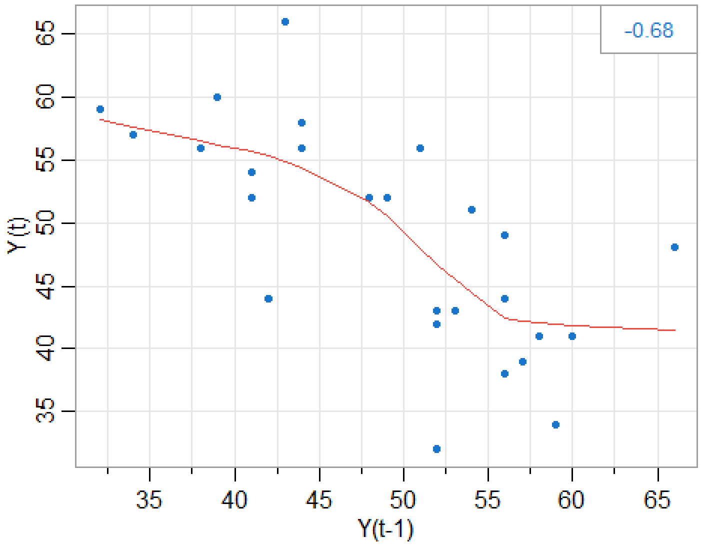
```
#Tutorial4_Q1
rm(list = ls()) #removes all objects from the current workspace (R memory)
data1 <- read.delim("C:/STAT 336-Time Series Analysis/data_tutorial4.txt",
header = TRUE)
#install.packages("astsa")
library(astsa)
Y <- ts(data1$Y) #this makes sure R knows that x is a time series.

summary(Y)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  32.00   42.50   51.00   48.89   56.00   66.00


#plotting time series of Y with points marked as "o" (part 1)
tsplot(Y, type="b",col=4,lwd=2)
```



1- From the figure, it appears that the data are stationary in the mean, as there do not seem to be a clear trend component in the data. Also, the variance seems to be constant over the time, hence, the series seem to be stationary.

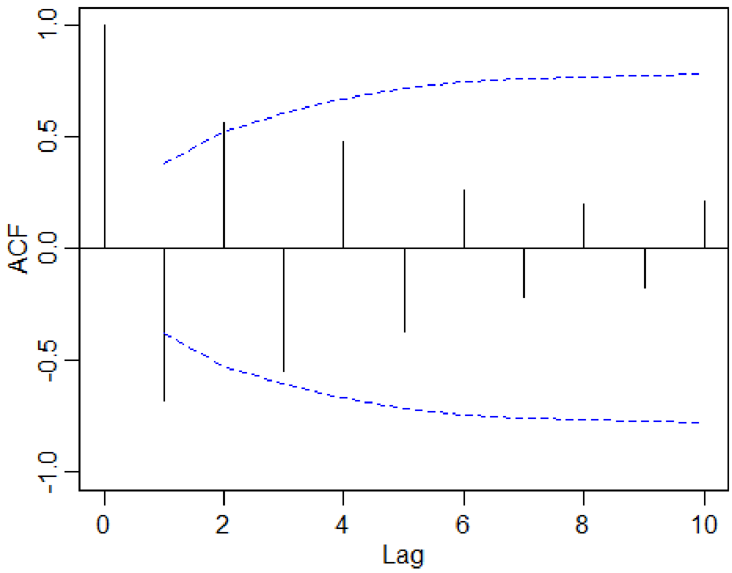
2- As we see from the plot of the series, most of the time there exist a value above the mean followed by a value beneath the mean and so on. Thus, we expect that the value of ρ_1 to be negative, however its exact value is difficult to guess from the figure, but we don't expect that it will be a high value (i.e. near to one) because the fluctuations are not the same across the series.

```
#Plot x versus lag 1 of x (part 2)
lag1.plot(Y,1,col=4,pch=20, cex=1)
```



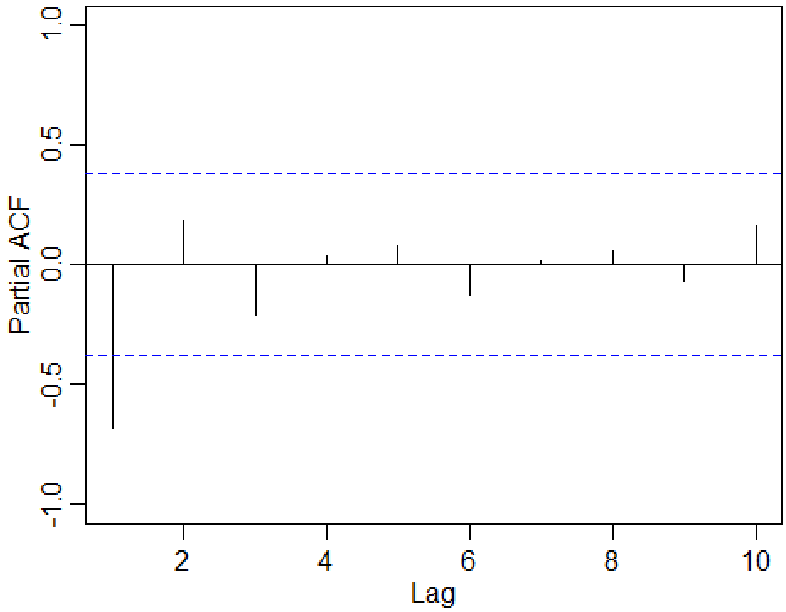
3- As we see from the plot, the behavior we anticipated for the relation between any two observations that are one time apart is clear. As we notice the negative correlation between the observations, as the regression line between y_t and y_{t-1} is decreasing. We can estimate visually the value of ρ_1 (which is the slope of the line) maybe between -0.6 and -0.7.

```
# ACF & PACF (part 4 & 5)
acf(Y,lag.max = 10, plot = TRUE,ylim=c(-1,1),ci.type = "ma")

#The confidence interval plotted in plot.acf is based on an uncorrelated s
eries and should be treated with appropriate caution. Using ci.type ="ma".
```



4- Where we got the SACF for the data for any time lags, we notice here that $\hat{\rho}_1 = r_1 = -0.68117$, we also notice that the autocorrelation decrease as time lag increase (this is a characteristic of the stationary processes). Notice also that all values of r_k after the first time lag lie within the 95% C.I , we thus can test the hypothesis that all autocorrelation coefficient after time lag 1 are not different from zero.

```r
pacf(Y,lag.max = 10,plot = TRUE,ylim=c(-1,1))
```



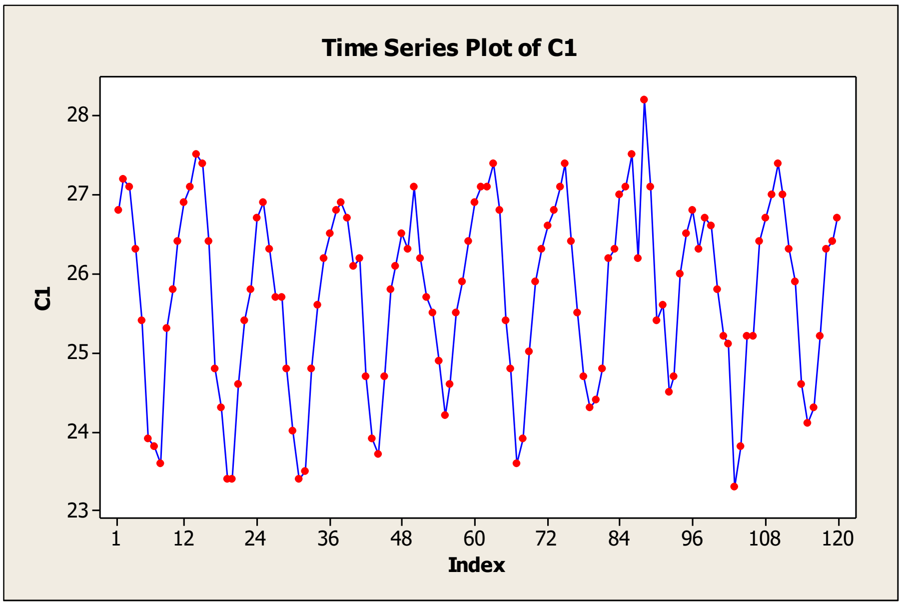
5- We notice from the figure that only one value of r_{kk} is outside the 95% C.I.(cutoff after lag one), whereas all the rest of the values are within the 95% C.I. which means that they are not significantly different from zero. Notice also that $r_1 = r_{11} = -0.68117$ which is always true.

```r
acf2(Y,max.lag = 10, plot=FALSE) # Value of ACF & PACF

             ACF        PACF
 [1,] -0.6811684 -0.68116842
 [2,]  0.5617521  0.18238796
 [3,] -0.5468817 -0.21187138
 [4,]  0.4815203  0.03553379
 [5,] -0.3726019  0.08128773
 [6,]  0.2628556 -0.12723856
 [7,] -0.2164267  0.01697477
 [8,]  0.2001462  0.05640012
 [9,] -0.1784259 -0.07046835
[10,]  0.2122418  0.16324654
```

Question 2:

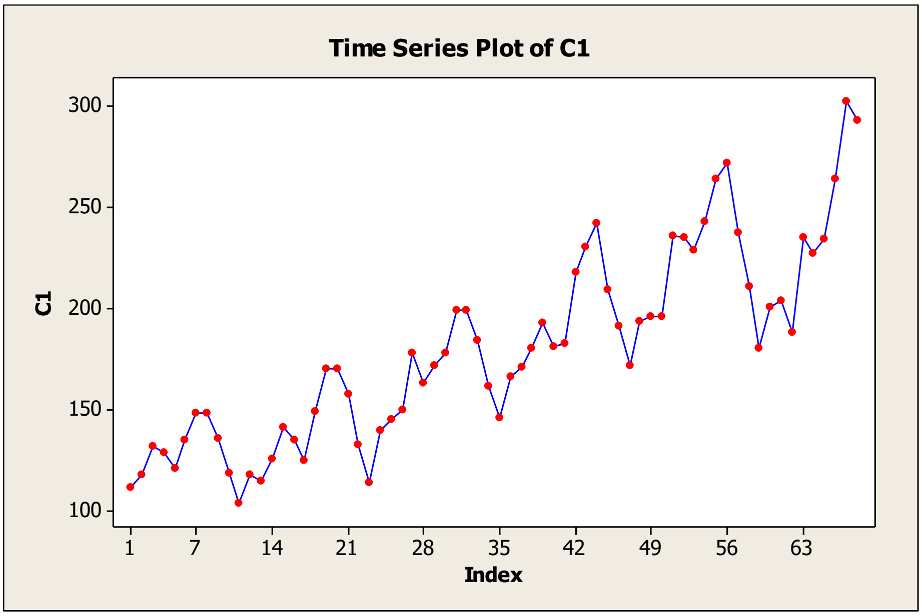
In the following cases, comment on the stationarity of the time series, and in case of non-stationarity, briefly explain how you will deal with the problem:

1- The following series represent average monthly temperatures for a period of 10 years:



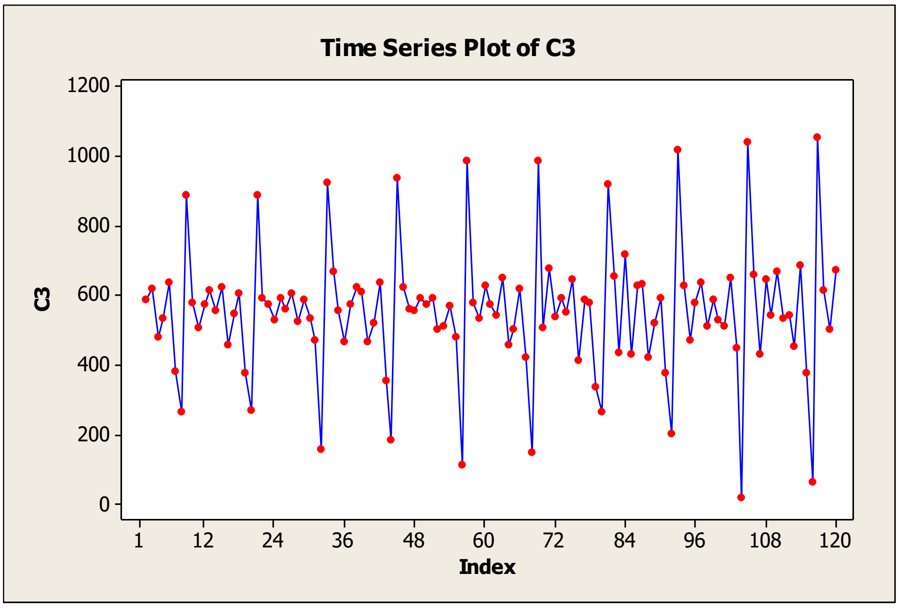
The mean of the series looks constant over time, the same could be said about the variance. Which indicate that the series is **stationary**. Also, notice that the series exhibit a **seasonal pattern**, where the average temperatures decrease for months 6,7 and 8 every year. Whereas the temperatures increase gradually for the rest of the year. So, the model we use for the data should incorporate a seasonal component and must estimate its coefficients and test their significance.

2- The following series represent monthly numbers (in thousands) of international travelers for a period of 10 years:



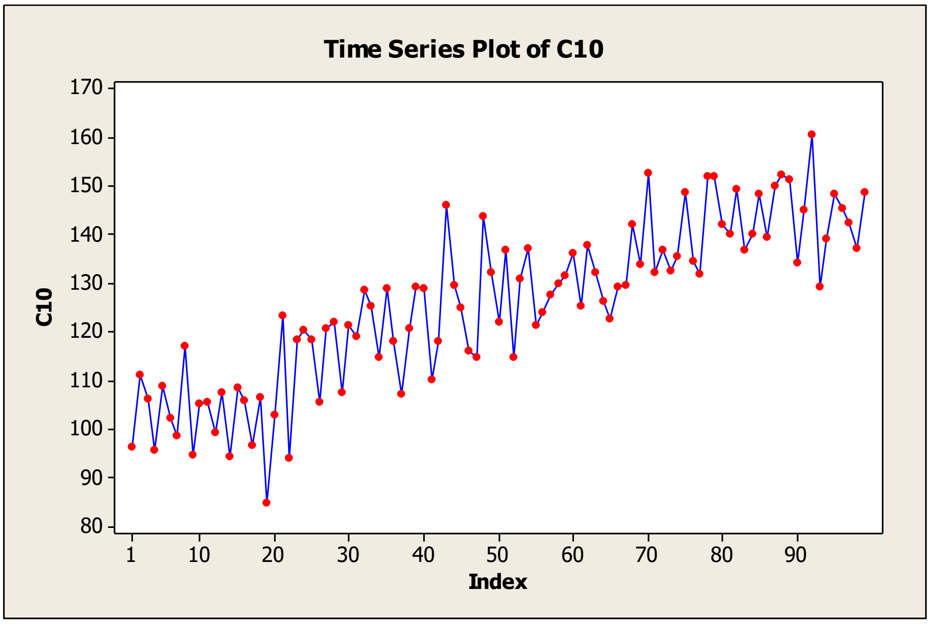
It is clear that the series exhibit an increasing **trend component** and thus the series is **not stationary.** We can deal with the problem of ***non-stationarity in the mean*** by applying the ***difference operator ∇,*** so we take the first order difference and inspect the resulting series to see if it succeeded in turning it to a stationary series or not, otherwise we can take the second difference. We also notice that the variance of the series increases with time, hence it is ***not stationary in variance*** as well. We can use for example the ***logarithmic*** transformation or any other transformation in the ***Box-Cox family*** of transformations. But be aware that if there is a need to apply both transformations for the data, then ***logarithmic transformation must be applied before the differences.***

3- A time series representing the monthly demand of a particular item:



The series seems to be ***stationary in the mean***, as it does not change over time. There is a slight indication of ***non-stationarity in the variance***, we can confirm this by applying the logarithmic transformation to the data and study the resulting series.

4- A time series representing the weekly sales of a large company:



It is clear that the series exhibit an increasing ***trend component*** and thus the series is ***not stationary***. We can deal with the problem of ***non-stationarity in the mean*** by applying the difference operator ∇, so we take the first order difference and inspect the resulting series to see if it succeeded in turning it to a stationary series or not, otherwise we can take the second difference. We do not notice any problem of the variance as it seems constant as time increase.

Question 3:

In the general linear process (GLP) , $Y_t = \mu_Y + \sum_{j=0}^{\infty} \psi_j \varepsilon_{t-j}$, we used the following ψ_j weights:

1- $\psi_j = \phi^j$ for j=1,2,… , where $|\phi| < 1$. What is the form of the resulting process, and derive its autocorrelation function?

$$
\begin{aligned}
Y_t &= \mu_Y + \textstyle\sum_{j=0}^{\infty} \psi_j \varepsilon_{t-j} \\
&= \mu_Y + \textstyle\sum_{j=0}^{\infty} \phi^j \varepsilon_{t-j} \\
Y_t - \mu_Y &= \varepsilon_t + \phi\, \varepsilon_{t-1} + \phi^2 \varepsilon_{t-2} + \phi^3 \varepsilon_{t-3} + \cdots \\
&= \varepsilon_t + \phi\, [\varepsilon_{t-1} + \phi\, \varepsilon_{t-2} + \phi^2 \varepsilon_{t-3} + \cdots] \\
&= \varepsilon_t + \phi\, [Y_{t-1} - \mu_Y]
\end{aligned}
$$

Note: This process is called ***Autoregressive*** process of order one, since it is a regression of process at time t on its value at time $t-1$.

The ACF:

Taking the variance of both sides of the general linear process:

$$\mathrm{var}(Y_t) = \gamma_0 = \mathrm{var}\left(\sum_{j=0}^{\infty} \psi_j \varepsilon_{t-j}\right) = \sigma^2 \sum_{j=0}^{\infty} \psi_j^2$$

Also, finding the autocovariance for the process:

$$
\begin{aligned}
\mathrm{cov}(Y_t, Y_{t-k}) = \gamma_k &= \mathrm{cov}\left(\sum_{j=0}^{\infty} \psi_j \varepsilon_{t-j} , \sum_{j=0}^{\infty} \psi_j\ \varepsilon_{t-k-j}\right) \\
&= \mathrm{cov}\left(\sum_{i=0}^{k-1} \psi_i \varepsilon_{t-i} + \sum_{i=k}^{\infty} \psi_i \varepsilon_{t-i} , \sum_{j=0}^{\infty} \psi_j\ \varepsilon_{t-k-j}\right)
\end{aligned}
$$

And to let the summation in the term $\sum_{i=k}^{\infty} \psi_i \varepsilon_{t-i}$ to start from zero, we use the index transformation:

$\quad$ let $j = i - k \ \Rightarrow i = j + k$

Then,

$$\mathrm{cov}(Y_t, Y_{t-k}) = \mathrm{cov}\left(\sum_{i=0}^{k-1} \psi_i \varepsilon_{t-i} + \sum_{j=0}^{\infty} \psi_{j+k}\varepsilon_{t-(j+k)}\ ,\sum_{j=0}^{\infty} \psi_j\ \varepsilon_{t-k-j}\right)$$
$$= \mathrm{cov}\left(\sum_{j=0}^{\infty} \psi_{j+k}\ \varepsilon_{t-k-j}\ ,\sum_{j=0}^{\infty} \psi_j\ \varepsilon_{t-k-j}\right)$$
$$= \sigma^2 \sum_{j=0}^{\infty} \psi_j\ \psi_{j+k}$$

And hence the ACF for the G.L.P. has the form:

$$\boxed{\rho_k = \frac{\gamma_k}{\gamma_0} = \frac{\sum_{j=0}^{\infty} \psi_j\ \psi_{j+k}}{\sum_{j=0}^{\infty} \psi_j^2}}$$

Now, substituting for $\psi_j = \phi^j$, we get:

$$\rho_k = \frac{\sum_{j=0}^{\infty} \phi^j \phi^{j+k}}{\sum_{j=0}^{\infty} \phi^{2j}} = \frac{\phi^k \sum_{j=0}^{\infty} \phi^{2j}}{\sum_{j=0}^{\infty} \phi^{2j}} = \phi^k\ , k = 0,1,2, ..$$

Thus, we deduce that the resulting process is an AR(1) process (a special case of the GLP), and it has the ability to model data that has the property of autocorrelation that decline in an exponential fashion (for $0 < \phi < 1$), or in a declining sine wave fashion if $-1 < \phi < 0$. Try using different values for ϕ.

2- $\psi_0 = 1, \psi_1 = -\theta\ , \psi_j = 0$, for $j = 2,3, \ldots$, where, $|\theta| < 1$. What is the form of the resulting process, and derive its autocorrelation function?

$$Y_t - \mu_Y = \sum_{j=0}^{\infty} \psi_j \varepsilon_{t-j}$$
$$= (\varepsilon_t - \theta\varepsilon_{t-1} + 0\ \times\ \varepsilon_{t-2} + 0\ \times\ \varepsilon_{t-3} + \cdots)$$

$$\therefore Y_t - \mu_Y = \varepsilon_t - \theta\varepsilon_{t-1}$$

Note: This process is called a moving average of order 1, and it relates the process at time t with the errors (or shocks) at time t and time t-1.

The ACF:

Substituting for the value $\psi_0 = 1,\ \psi_1 = -\theta$ and the rest of the weights $\psi_j = 0, j > 1$ in the general form of the autocorrelation function of the GLP, we get:

$$\rho_k = \frac{\sum_{j=0}^{\infty} \psi_j\ \psi_{j+k}}{\sum_{j=0}^{\infty} \psi_j^2} = \frac{\psi_0\ \psi_k + \psi_1\ \psi_{1+k} + \psi_2\ \psi_{2+k} + \cdots}{\psi_0^2 + \psi_1^2 + \psi_2^2 + \cdots} = \frac{\psi_k + -\theta\psi_{1+k}}{1 + \theta^2}$$

For **k=1**:

$$\therefore \rho_1 = \frac{-\theta}{1 + \theta^2}$$

Note that using **k=2**, then all the terms in the numerator equals **zero**, this also true for any value **$k \geq 2$**. So, we note that the MA(1) process is a special case of the GLP, and it has the ability of modeling data that are correlated at one time lag only, and are uncorrelated for data that are further apart.