Methods of detrending of time series data of U.K. imports: goods and services (Pound millions) from 1960-1970.

**INTRODUCTION**

***Problem Description:*** Here in this problem we are interested in

1. Demonstarating the method of differencing, OLS estimation and moving average smoothing to eliminate the deterministic component from the data set and hence extract the stationary component from the data with respect to each of these methods.
2. And further we want to Validate our results with respect to necessary plots and tests.

***Objective:*** The main objective of this problem is to demonstrate the elimination of the deterministic component from the data set using the method of differencing, ordinary least square method of estimation and method of moving average smoothing. We also want to extract the stationary component from the data with respect to each of these methods and validate it using autocorrelation function plot and augmented dickey feller test.

*#Setting and getting the current working directory.*  
**setwd**("E:/M.Sc/SEM III/TIME\_SERIES\_ANALYSIS(MST371)/Practical Labs")  
**getwd**()

## [1] "E:/M.Sc/SEM III/TIME\_SERIES\_ANALYSIS(MST371)/Practical Labs"

***Data Description:***

The data set consists of Quarterly series of U.K. imports: goods and services (Pound millions) from 1960 – 1970. The data set has total 44 records of UK imports and and the year when the data is recorded. Also it is observed that in each year the data has been recorded in four quarters. There two variables in the the dataset i.e. time and imports: goods and services (Pound millions)

*#Loading the package required to load the dataset.*  
**library**(readxl)

*#Loading the UK import dataset.*  
data <- **read\_excel**("E:/M.Sc/SEM III/TIME\_SERIES\_ANALYSIS(MST371)/data.xlsx")

## New names:  
## \* `` -> ...2

*#Obtaining the first few records of the dataset.*  
**head**(data)

## # A tibble: 6 x 2  
## Quarter ...2  
## <chr> <dbl>  
## 1 1960 Q1 1382  
## 2 1960 Q2 1417  
## 3 1960 Q3 1432  
## 4 1960 Q4 1438  
## 5 1961 Q1 1457  
## 6 1961 Q2 1403

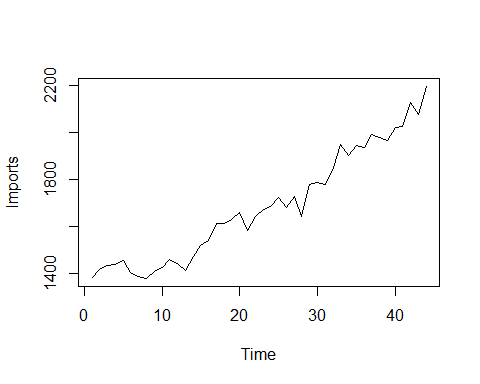
**ANALYSIS**

*#Extracing the data for the imports variable which we are interested in.*  
imports=data**$**...2  
  
*#Now converting the it into a time series data.*  
imports1=**ts**(imports)  
  
*#Here. we are checking if the dataset has been concverted into a timeseries plot.*  
**class**(imports1)

## [1] "ts"

Hence, now the dataset we are interested in is a timeseries data.

*#Now we will first obtain the time series plot of the data to understand the nature of the time series data.*  
**ts.plot**(imports1, gpars = **list**(xlab="Time",ylab="Imports",lty=**c**(1**:**20)))



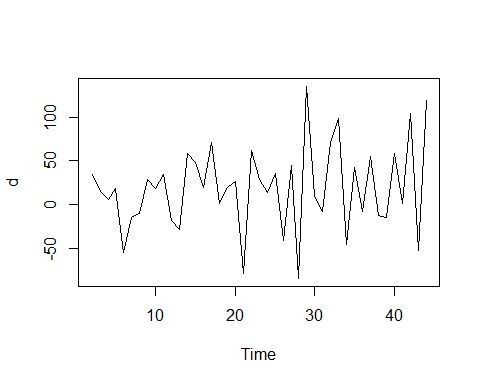


***Interpretation:*** From the above time series plot, figure 1 we observe that there exists a trend component in the dataset since there is observed a increase or decrese pattern for a longer period of time. Also we observe that there is some kind of irregularity in the dataset hence we can say that there also exists a error component in the dataset.

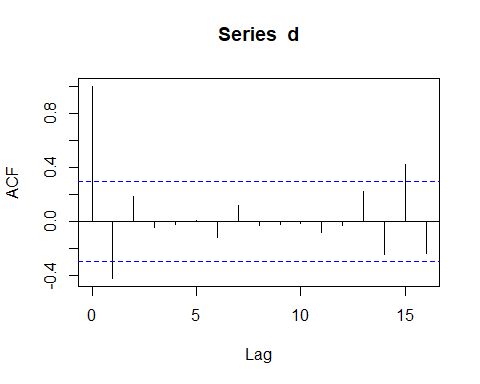
1. **Method of differencing**

Since, from the above time series plot we observed that our time series data is non stationary and there exist a trend component in the time series data, we try to eliminate the trend component to make it stationary by using the detrending method of differencing.

*#Now, we proceed to perform the first differencing and check for the stationarity.*  
d=**diff**(imports1)  
  
*#obtaining the time series plot for the new dataset.*  
**ts.plot**(d)



*#Obtaining the acf plot for the above time series after first differencing to check for stationarity.*  
**acf**(d)



*#loading the package 'tseries'*  
**library**(tseries)

## Warning: package 'tseries' was built under R version 4.0.5

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

*#Now we want to validate for the stationarity of the new time series dataset using Augmented Dickey Fuller(ADF test) test.*  
**adf.test**(d)

##   
## Augmented Dickey-Fuller Test  
##   
## data: d  
## Dickey-Fuller = -3.8136, Lag order = 3, p-value = 0.02772  
## alternative hypothesis: stationary

***Interpretation:*** From the acf plot (Figure 3) the series seems to be non stationary because lag 1 and lag 15 are beyond the threshold line but since from the above statistical test Augmented Dickey-Fuller Test it is observed tha pvalue = 0.02772 < 0.05, thus we reject the null hypothesis and conclude that the timeseries have been converted into stationary series and since the statiscal tests are more accurate we go with the tests results.

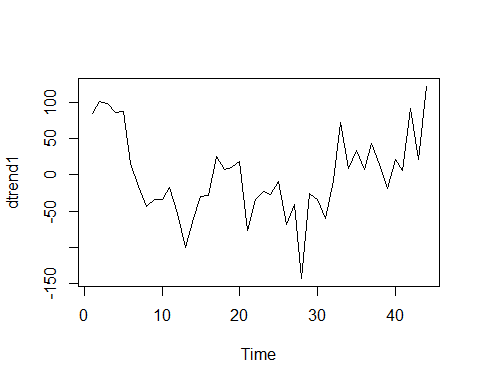
1. **Method of ordinary least square**

Now, here we try to estimate the trend component using OLS method and then we detrend the series by eliminating the estimated trend component and extract the stationary component from the data and validate the results.

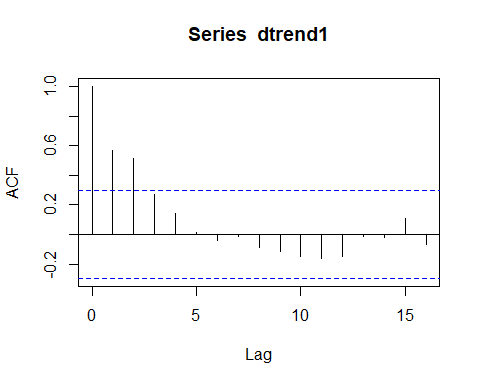
Here, time is our independent variable and U.K. imports is the dependent variable.

*#Creating the independent variable time.*  
time = **seq**(1**:**44)

*#Obtaining a simple linear regression model for imports on time to get the estimated trend required to detrend the series.*  
reg1=**lm**(imports1**~**time)  
  
*#Here, we are trying to obtain the estimated trend values.*  
estim\_trend1=**fitted.values**(reg1)  
  
*#Now, we are eliminating the estimated trend from the time series by subtracting the estimated trend values from the corresponding time series values.*  
dtrend1=imports1**-**estim\_trend1  
  
*#Obtaining the time series plot of detrended data.*  
**ts.plot**(dtrend1)



*#Obtaining the acf plot for the detrended series*  
**acf**(dtrend1)



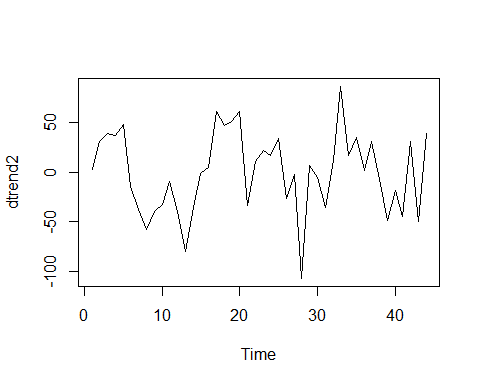


*#Testing for the stationarity of detrended series using augmented dickey feller test.*  
**adf.test**(dtrend1)

##   
## Augmented Dickey-Fuller Test  
##   
## data: dtrend1  
## Dickey-Fuller = -2.5343, Lag order = 3, p-value = 0.3623  
## alternative hypothesis: stationary

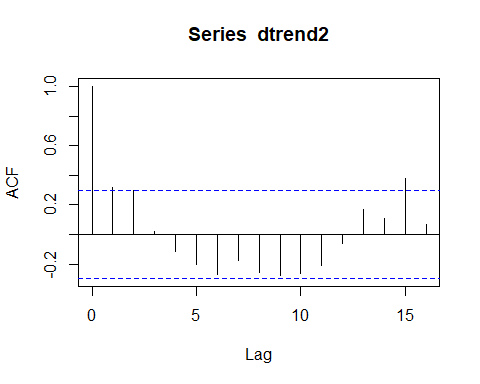
***Interpretation:*** From the acf plot (Figure 5) the series seems to be non stationary because lag 1 and lag 2 are beyond the threshold line and also from the above statistical test Augmented Dickey-Fuller Test it is observed tha pvalue = 0.3623 > 0.05, thus we fail to reject the null hypothesis and conclude that the detrended timeseries is a non stationary series.

*#Since the time series data is still non stationary we proceed to Obtain a polynomial regression model of degree 2 for imports on time to get the estimated trend required to detrend the series.*  
reg2=**lm**(imports1**~**time **+** **I**(time**^**2))  
  
*#Here, we are trying to obtain the estimated trend values.*  
estim\_trend2=**fitted.values**(reg2)  
  
*#Now, we are eliminating the estimated trend from the time series by subtracting the estimated trend values from the corresponding time series values.*  
dtrend2=imports1**-**estim\_trend2  
  
*#Obtaining the time series plot of detrended data.*  
**ts.plot**(dtrend2)





*#Obtaining the acf plot for the detrended series*  
**acf**(dtrend2)



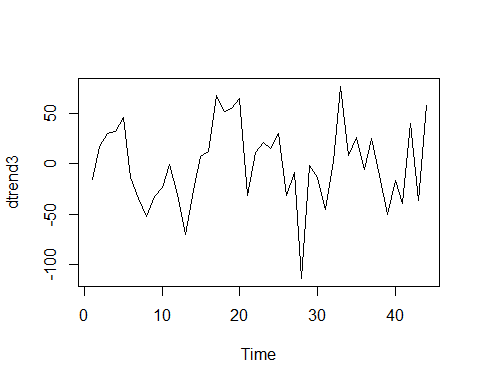


*#Testing for the stationarity of detrended series using augmented dickey feller test.*  
**adf.test**(dtrend2)

##   
## Augmented Dickey-Fuller Test  
##   
## data: dtrend2  
## Dickey-Fuller = -3.2564, Lag order = 3, p-value = 0.091  
## alternative hypothesis: stationary

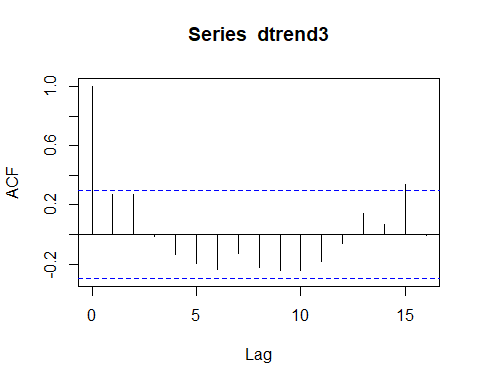
***Interpretation:*** From the acf plot (Figure 7) the series seems to be non stationary because lag 1, lag 2, and lag 15 are beyond the threshold line and also from the above statistical test Augmented Dickey-Fuller Test it is observed tha pvalue = 0.091 > 0.05, thus we fail to reject the null hypothesis and conclude that the detrended timeseries is a non stationary series.

*#Since the time series data is still non stationary we proceed to Obtain a polynomial regression model of degree 3 for imports on time to get the estimated trend required to detrend the series.*  
reg3=**lm**(imports1**~**time **+** **I**(time**^**2) **+** **I**(time**^**3))  
  
*#Here, we are trying to obtain the estimated trend values.*  
estim\_trend3=**fitted.values**(reg3)  
  
*#Now, we are eliminating the estimated trend from the time series by subtracting the estimated trend values from the corresponding time series values.*  
dtrend3=imports1**-**estim\_trend3  
  
*#Obtaining the time series plot of detrended data.*  
**ts.plot**(dtrend3)





*#Obtaining the acf plot for the detrended series*  
**acf**(dtrend3)



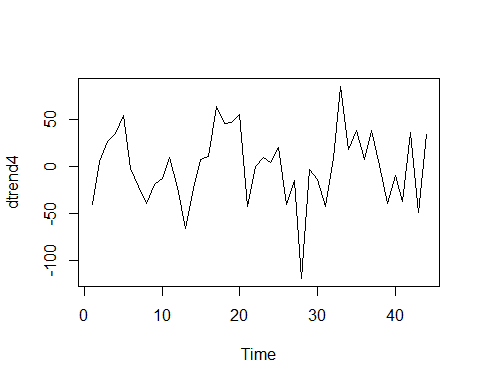


*#Testing for the stationarity of detrended series using augmented dickey feller test.*  
**adf.test**(dtrend3)

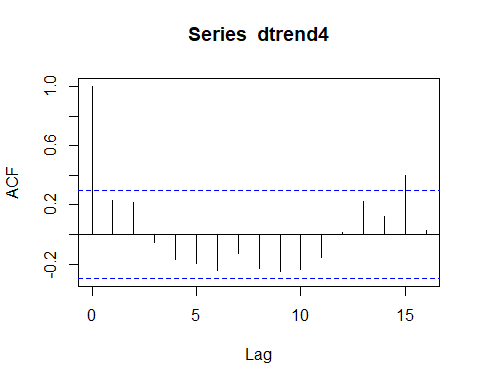
##   
## Augmented Dickey-Fuller Test  
##   
## data: dtrend3  
## Dickey-Fuller = -3.1852, Lag order = 3, p-value = 0.1046  
## alternative hypothesis: stationary

***Interpretation:*** From the acf plot (Figure 9) the series seems to be non stationary because lag 15 is beyond the threshold line and also from the above statistical test Augmented Dickey-Fuller Test it is observed tha pvalue = 0.1046 > 0.05, thus we fail to reject the null hypothesis and conclude that the detrended timeseries is a non stationary series.

*#Since the time series data is still non stationary we proceed to Obtain a polynomial regression model of degree 4 for imports on time to get the estimated trend required to detrend the series.*  
reg4=**lm**(imports1**~**time **+** **I**(time**^**2) **+** **I**(time**^**3) **+** **I**(time**^**4))  
  
*#Here, we are trying to obtain the estimated trend values.*  
estim\_trend4=**fitted.values**(reg4)  
  
*#Now, we are eliminating the estimated trend from the time series by subtracting the estimated trend values from the corresponding time series values.*  
dtrend4=imports1**-**estim\_trend4  
  
*#Obtaining the time series plot of detrended data.*  
**ts.plot**(dtrend4)



*#Obtaining the acf plot for the detrended series*  
**acf**(dtrend4)



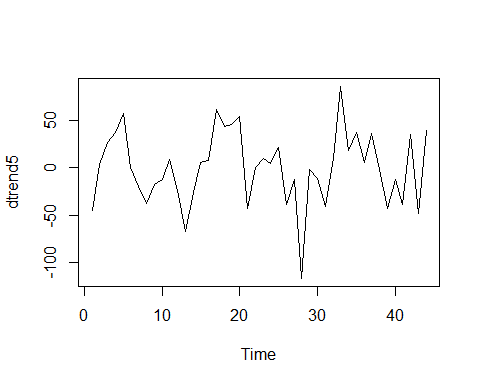


*#Testing for the stationarity of detrended series using augmented dickey feller test.*  
**adf.test**(dtrend4)

##   
## Augmented Dickey-Fuller Test  
##   
## data: dtrend4  
## Dickey-Fuller = -3.3396, Lag order = 3, p-value = 0.07845  
## alternative hypothesis: stationary

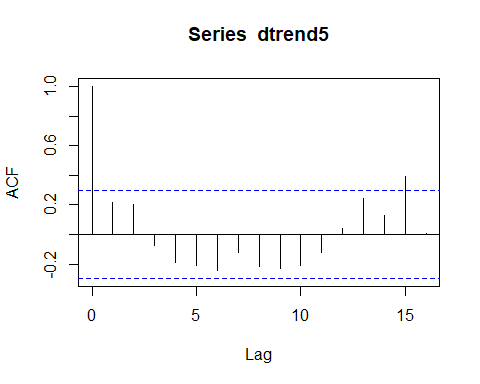
***Interpretation:*** From the acf plot (Figure 11) the series seems to be non stationary because lag 15 is beyond the threshold line and also from the above statistical test Augmented Dickey-Fuller Test it is observed tha pvalue = 0.07845 > 0.05, thus we fail to reject the null hypothesis and conclude that the detrended timeseries is a non stationary series.

*#Since the time series data is still non stationary we proceed to Obtain a polynomial regression model of degree 5 for imports on time to get the estimated trend required to detrend the series.*  
reg5=**lm**(imports1**~**time **+** **I**(time**^**2) **+** **I**(time**^**3) **+** **I**(time**^**4) **+** **I**(time**^**5))  
  
*#Here, we are trying to obtain the estimated trend values.*  
estim\_trend5=**fitted.values**(reg5)  
  
*#Now, we are eliminating the estimated trend from the time series by subtracting the estimated trend values from the corresponding time series values.*  
dtrend5=imports1**-**estim\_trend5  
  
*#Obtaining the time series plot of detrended data.*  
**ts.plot**(dtrend5)





*#Obtaining the acf plot for the detrended series*  
**acf**(dtrend5)



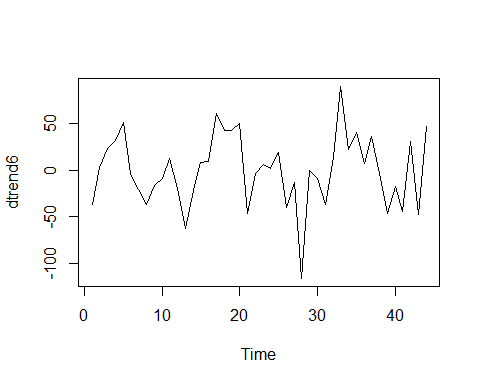


*#Testing for the stationarity of detrended series using augmented dickey feller test.*  
**adf.test**(dtrend5)

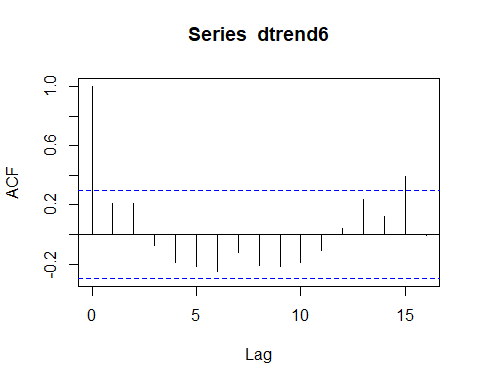
##   
## Augmented Dickey-Fuller Test  
##   
## data: dtrend5  
## Dickey-Fuller = -3.4345, Lag order = 3, p-value = 0.06412  
## alternative hypothesis: stationary

***Interpretation:*** From the acf plot (Figure 13) the series seems to be non stationary because lag 15 is beyond the threshold line and also from the above statistical test Augmented Dickey-Fuller Test it is observed tha pvalue = 0.06412 > 0.05, thus we fail to reject the null hypothesis and conclude that the detrended timeseries is a non stationary series.

*#Since the time series data is still non stationary we proceed to Obtain a polynomial regression model of degree 6 for imports on time to get the estimated trend required to detrend the series.*  
reg6=**lm**(imports1**~**time **+** **I**(time**^**2) **+** **I**(time**^**3) **+** **I**(time**^**4) **+** **I**(time**^**5) **+** **I**(time**^**6))  
  
*#Here, we are trying to obtain the estimated trend values.*  
estim\_trend6=**fitted.values**(reg6)  
  
*#Now, we are eliminating the estimated trend from the time series by subtracting the estimated trend values from the corresponding time series values.*  
dtrend6=imports1**-**estim\_trend6  
  
*#Obtaining the time series plot of detrended data.*  
**ts.plot**(dtrend6)



*#Obtaining the acf plot for the detrended series*  
**acf**(dtrend6)



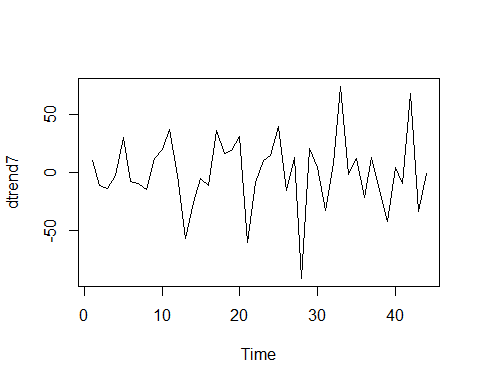


*#Testing for the stationarity of detrended series using augmented dickey feller test.*  
**adf.test**(dtrend6)

##   
## Augmented Dickey-Fuller Test  
##   
## data: dtrend6  
## Dickey-Fuller = -3.4359, Lag order = 3, p-value = 0.0639  
## alternative hypothesis: stationary

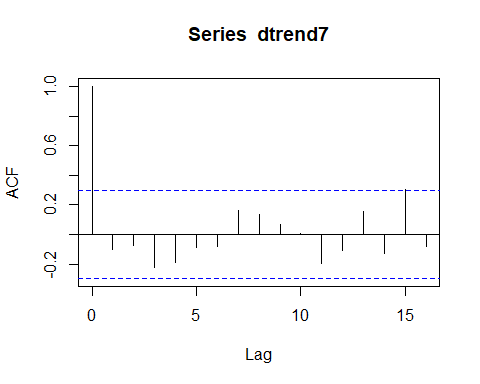
***Interpretation:*** From the acf plot (Figure 15) the series seems to be non stationary because lag 15 is beyond the threshold line and also from the above statistical test Augmented Dickey-Fuller Test it is observed tha pvalue = 0.0639 > 0.05, thus we fail to reject the null hypothesis and conclude that the detrended timeseries is a non stationary series.

*#Since the time series data is still non stationary we proceed to Obtain a polynomial regression model of degree 6 for imports on time to get the estimated trend required to detrend the series.*  
reg7=**lm**(imports1**~**time **+** **I**(time**^**2) **+** **I**(time**^**3) **+** **I**(time**^**4) **+** **I**(time**^**5) **+** **I**(time**^**6) **+** **I**(time**^**7))  
  
*#Here, we are trying to obtain the estimated trend values.*  
estim\_trend7=**fitted.values**(reg7)  
  
*#Now, we are eliminating the estimated trend from the time series by subtracting the estimated trend values from the corresponding time series values.*  
dtrend7=imports1**-**estim\_trend7  
  
*#Obtaining the time series plot of detrended data.*  
**ts.plot**(dtrend7)





*#Obtaining the acf plot for the detrended series*  
**acf**(dtrend7)





*#Testing for the stationarity of detrended series using augmented dickey feller test.*  
**adf.test**(dtrend7)*#p>0.05 we accept ho n conclude non stationary*

## Warning in adf.test(dtrend7): p-value smaller than printed p-value

##   
## Augmented Dickey-Fuller Test  
##   
## data: dtrend7  
## Dickey-Fuller = -4.8795, Lag order = 3, p-value = 0.01  
## alternative hypothesis: stationary

***Interpretation:*** From the acf plot (Figure 17) the series seems to be non stationary because lag 15 is beyond the threshold line by a little but since from the above statistical test Augmented Dickey-Fuller Test it is observed tha pvalue is less than 0.01 < 0.05, thus we reject the null hypothesis and conclude that the timeseries have been converted into stationary series and since the statiscal tests are more accurate we go with the tests results.

1. **Method of moving average**

Here in this method we first try to estimate the trend values using the method of moving average smoothing and further we try to eliminate the trend component to make it a stationary series and then validate the stationarity of the detrended series.

The formula to obtain the estimated trend values of order q by the method of moving average smoothing is,

Wt = , t=q+1, q+2,…,n-q

Where,

**Wt** is the estimated trend values.

**Zt** is the time series values.

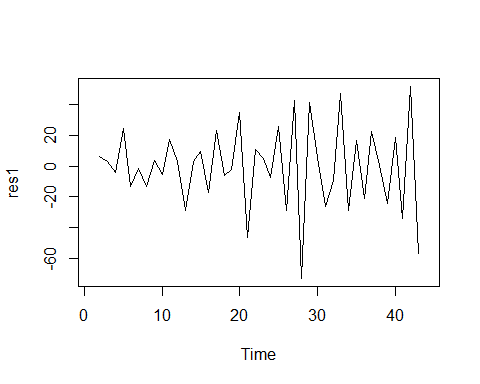
*#Loading the package 'forecast'*  
**library**(forecast)

## Warning: package 'forecast' was built under R version 4.0.5

*#Here we are estimating the trend values of order 3 using method of moving average smoothing.*  
ma1=**ma**(imports1,order=3)  
ma1

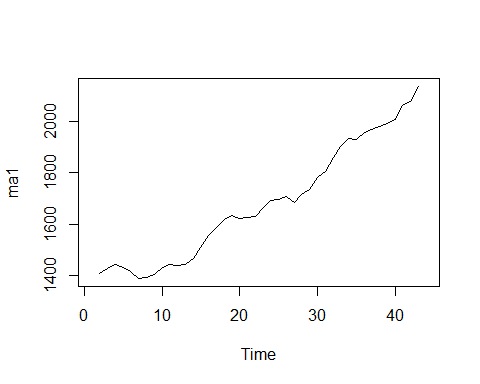
## Time Series:  
## Start = 1   
## End = 44   
## Frequency = 1   
## [1] NA 1410.333 1429.000 1442.333 1432.667 1416.333 1390.333 1392.000  
## [9] 1404.333 1431.333 1442.667 1438.667 1442.667 1468.667 1510.667 1557.000  
## [17] 1587.667 1618.333 1634.333 1624.000 1627.667 1632.000 1667.000 1693.333  
## [25] 1696.333 1709.667 1683.000 1715.000 1735.333 1781.000 1805.333 1859.000  
## [33] 1900.333 1932.000 1928.333 1958.000 1969.667 1979.333 1990.000 2005.333  
## [41] 2060.000 2078.000 2135.000 NA

*#Now, we are eliminating the estimated trend from the time series by subtracting the estimated trend values from the corresponding time series values.*  
res1=imports1**-**ma1  
  
*#Obtaing the time series plot for the detrended series by the method of moving average smoothing.*  
**ts.plot**(res1)





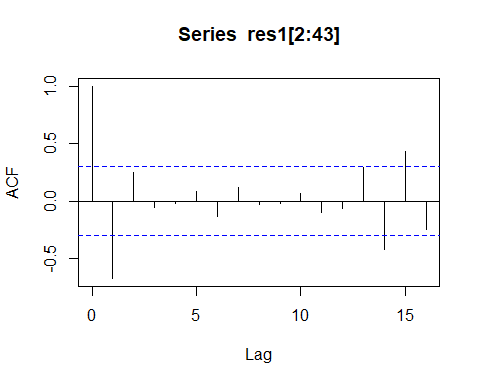
*#Obtaining the plot for estimated trends using moving average smoothing.*  
**ts.plot**(ma1)





Thus, from the above plot (Figure 19) it is observed that the plot is much smoother.

*#Now, we obtain the acf plot for the detrended dataset.*  
**acf**(res1[2**:**43])





*#Now, we check for the stationarity of the detrended dataset using Augmented Dickey-Fuller Test.*  
**adf.test**(res1[2**:**43])

## Warning in adf.test(res1[2:43]): p-value smaller than printed p-value

##   
## Augmented Dickey-Fuller Test  
##   
## data: res1[2:43]  
## Dickey-Fuller = -5.3463, Lag order = 3, p-value = 0.01  
## alternative hypothesis: stationary

***Interpretation:*** From the acf plot (Figure 20) the series seems to be non stationary because lag 1, lag 14, and lag 15 are beyond the threshold line but since from the above statistical test Augmented Dickey-Fuller Test it is observed tha pvalue is less than 0.01 < 0.05, thus we reject the null hypothesis and conclude that the timeseries have been converted into stationary series and since the statiscal tests are more accurate we go with the tests results.

**CONCLUSION**

Thus, on performing all the three methods of detrending the time series and extracting the stationary component we observe that for OLS estimation we need to go till fitting polynomial regression model of degree 7 which is time consuming hence, we can go for either moving average smoothing or method of differencing in this case, since it is easier to extract stationarity using both the methods and also we are not loosing much information about the dataset in this case.