Financial Modelling

Avinash Kori | ED15B006
Engineering Design Department,
Indian Institute of Technology, Madras
koriavinash1@gmail.com

Contents

Introduction

In this case study various time series modelling methods are discussed along with R implementation for each of them. Main idea followed is to analyse the data, identify trend, seperate stationary part of the data, build model on that residual data. In this case study we model closing price for all the stocks.

Data

In this case study 5 different pharmaceutical companies (on NSE India) were considered for analysis. Data obtained from yahoo finance starting from 6th November 2016 to 7th November 2018. Companies selected were:

- GlaxoSmithkline Pharmaceuticals Limited (GLAXO.NS)
- Glenmark Pharmaceuticals Limited (GLENMARK.NS)
- Aurobindo Pharma Limited (AUROPHARMA.NS)
- Sun Pharmaceutical Industries Limited (SUNPHARMA.NS)
- Alembic Pharmaceuticals Limited (APLLTD.NS)

Data Analysis

In this section pre-processing steps involved in data analysis are explored, few of them involve:

- Data Standardization: let Y be the time-series data standardization involves converting the data to zero mean and unit standard deviation data:

$$Y = \frac{Y - \mu_Y}{\sigma_Y}$$

- Variance Stationarity (remove hetroskedasticity):

$$Y = \begin{cases} ln(Y) & if \quad \lambda = 0 \\ \frac{Y^\lambda - 1}{\lambda} & otherwise \end{cases}$$

  The conditions for predictability of any time series is illustrated using auto correlation and partial auto correlation functions.

- Auto correlation:

$$Y[l] = \frac{E((Y[k] - \mu)(Y[k-l] - \mu))}{\sigma_Y^2}$$

```
loadCSVData <- function(path){
  data = read.csv(path, header = TRUE)
  print(summary(data$Close))
  return(data$Close)
}

standardize <- function(data){
  data = (data - mean(data)) / sd(data)
  print(summary(data))
  return(data)
}
```

```r
boxcox <- function(y, lambda){
  if (lambda != 0){
    return((y^{lambda} - 1.0)/lambda)
  }
  else{
    return(log(y))
  }
}
```

Analysis of Alembic Pharmaceuticals Limited (APLLTD.NS)

```r
aplltd = NULL
aplltd$raw_daily = loadCSVData('../Data/APLLTD/APLLTD.NS_daily.csv')
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   418.1   515.7   546.0   553.7   597.5   661.9
```

```r
aplltd$raw_weekly = loadCSVData('../Data/APLLTD/APLLTD.NS_weekly.csv')
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   420.0   516.4   547.7   554.4   595.0   654.2
```

```r
aplltd$raw_monthly = loadCSVData('../Data/APLLTD/APLLTD.NS_monthly.csv')
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   427.5   519.3   552.8   553.0   590.5   629.0
```

```r
aplltd$sd_daily = standardize(aplltd$raw_daily)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## -2.7140 -0.7624 -0.1560  0.0000  0.8758  2.1650
```

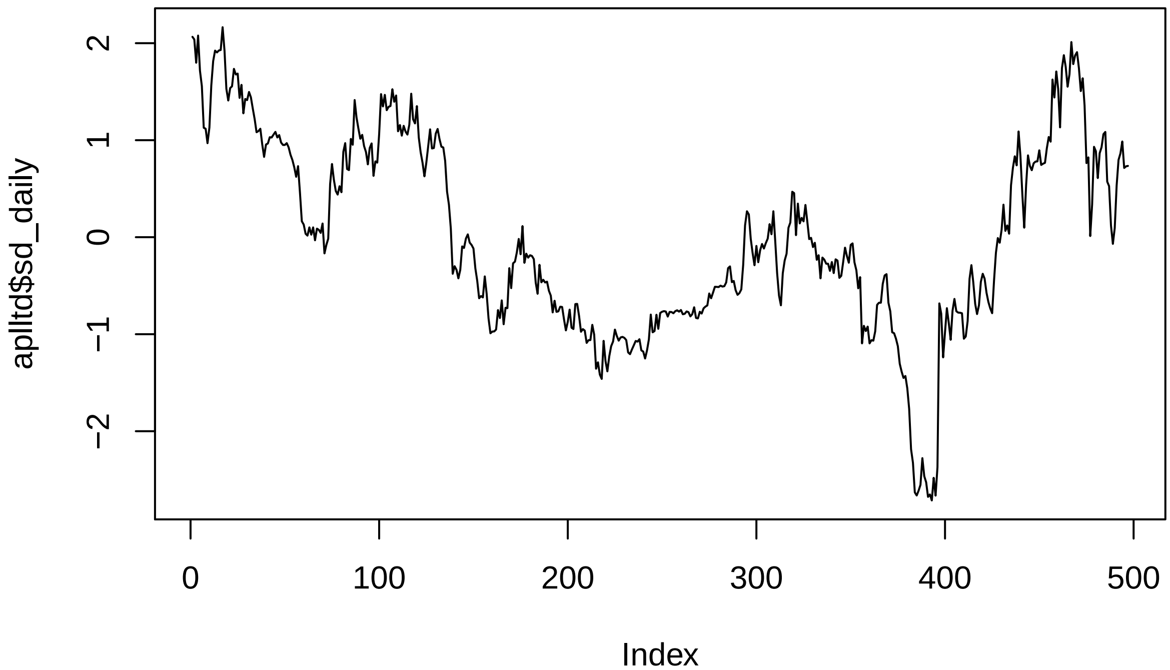
```r
aplltd$sd_weekly = standardize(aplltd$raw_weekly)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## -2.6640 -0.7536 -0.1326  0.0000  0.8057  1.9800
```

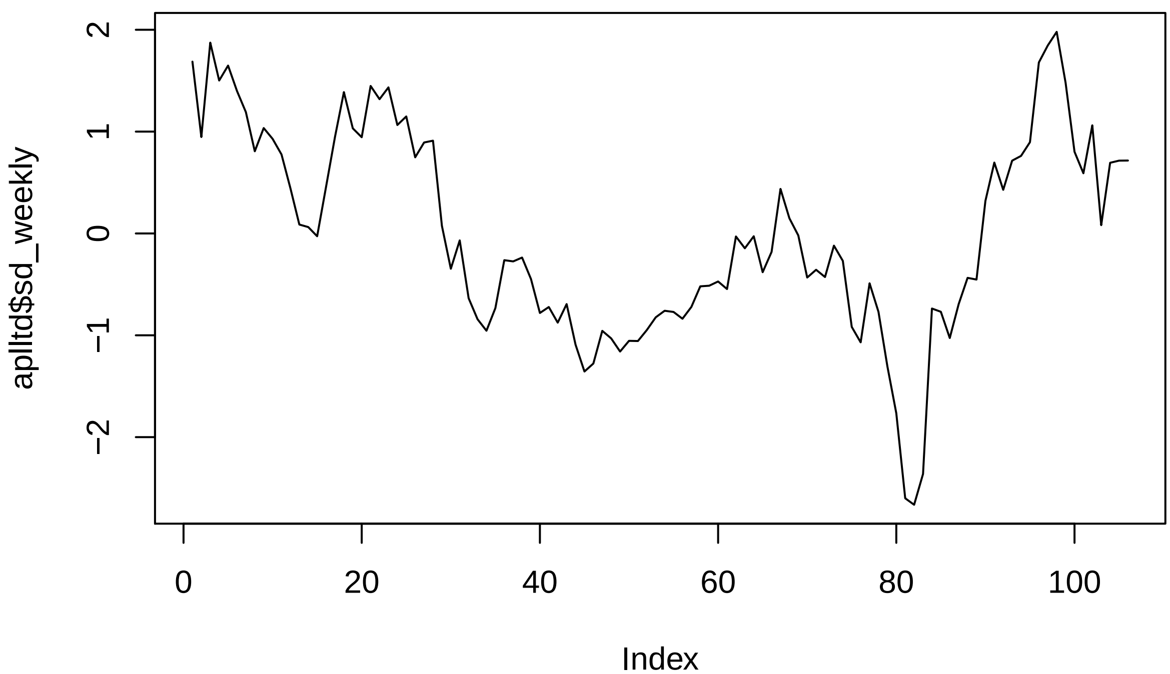
```r
aplltd$sd_monthly = standardize(aplltd$raw_monthly)
```

```
##      Min.    1st Qu.    Median      Mean    3rd Qu.      Max.
## -2.657000 -0.713800 -0.005926  0.000000  0.793000  1.608000
```

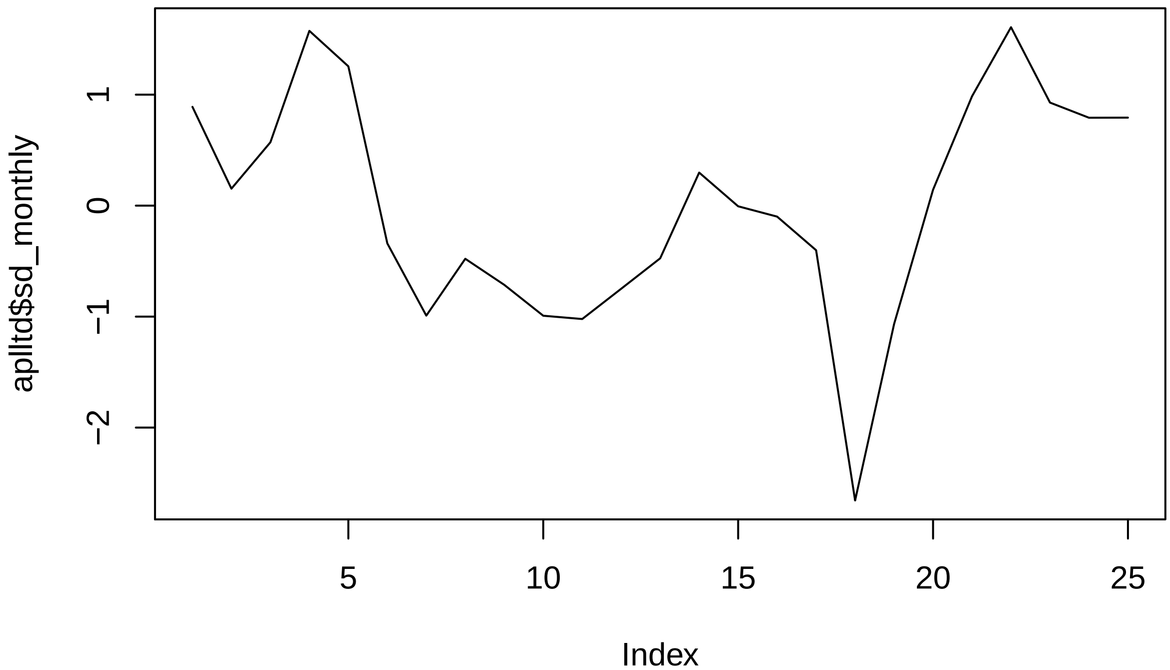
```r
# plot all the data
plot(aplltd$sd_daily, type='l')
```



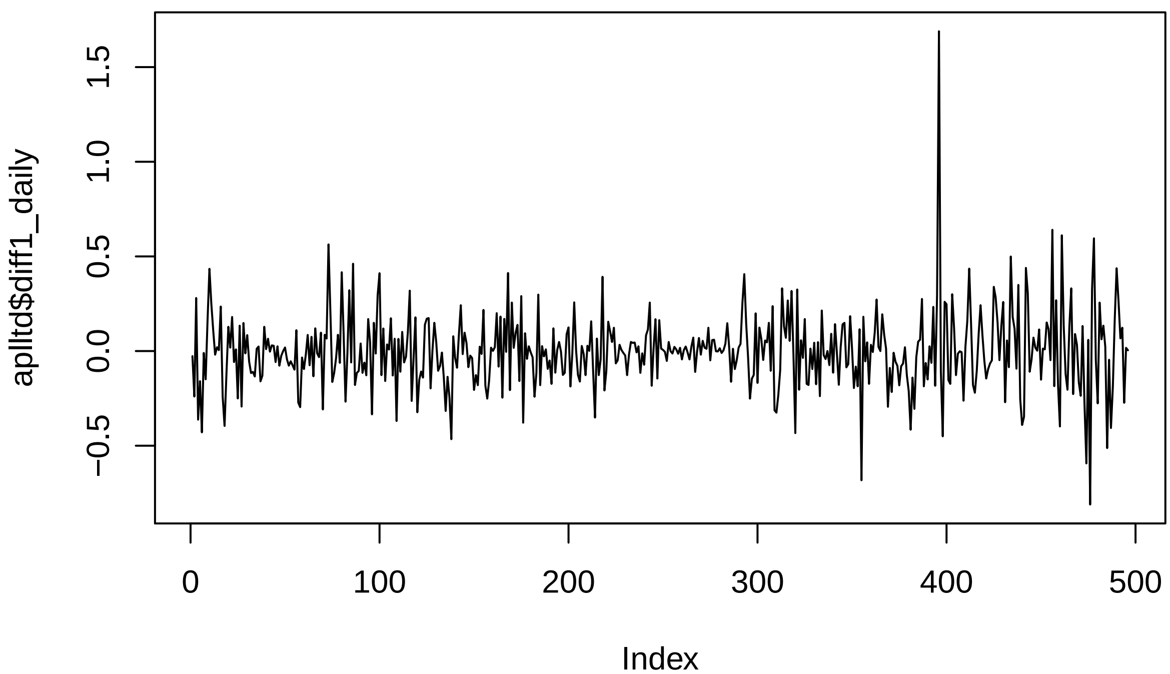

```
plot(aplltd$sd_weekly, type='l')
```



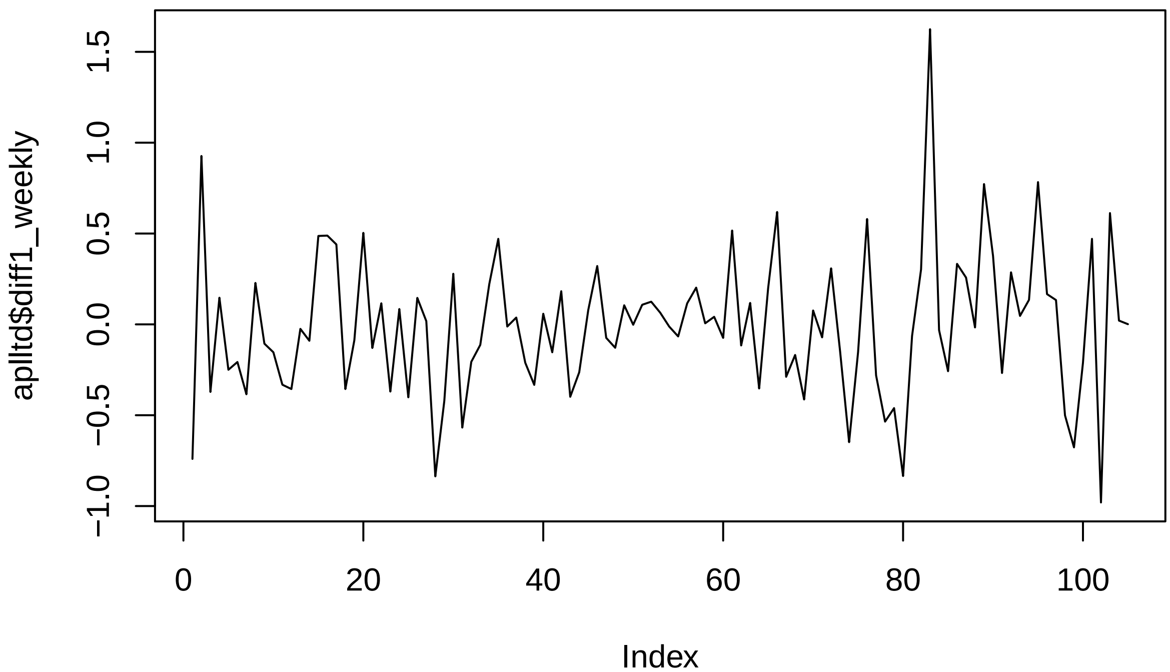

```
plot(aplltd$sd_monthly, type='l')
```



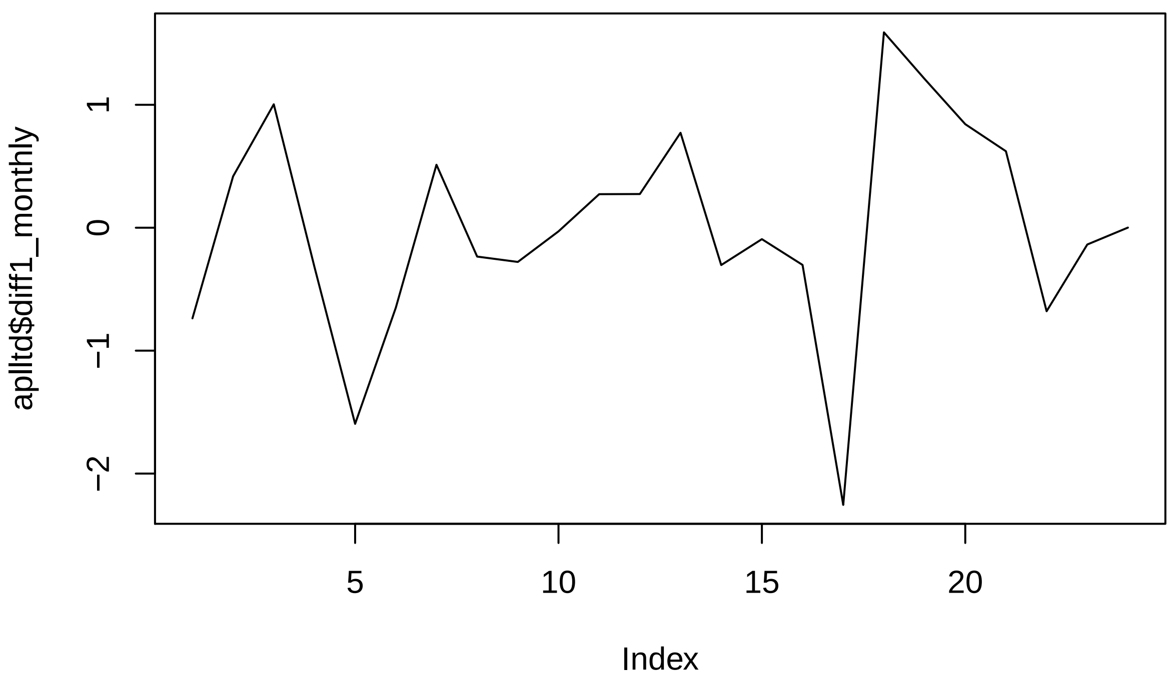

```r
# Based on plots we can clearly see that given data is
# non stationary (data has some integrating effect, which should estimated)
# which can be converted to stationary by taking difference of order = n

aplltd$diff1_daily = diff(aplltd$sd_daily)
aplltd$diff1_weekly = diff(aplltd$sd_weekly)
aplltd$diff1_monthly = diff(aplltd$sd_monthly)

# Plotting differenced data
plot(aplltd$diff1_daily, type='l')
```



```r
plot(aplltd$diff1_weekly, type='l')
```




```r
plot(aplltd$diff1_monthly, type='l')
```




```r
# Differenced plots seems to be stationary this can also be verified by summary of the data
print(summary(aplltd$diff1_daily))
```

```
##      Min.   1st Qu.    Median      Mean   3rd Qu.      Max.
## -0.810600 -0.115100 -0.003003 -0.002683  0.091820  1.689000
```

```r
print(summary(aplltd$diff1_weekly))
```

```
##       Min.    1st Qu.     Median       Mean    3rd Qu.       Max.
## -0.980200 -0.257700 -0.011890 -0.009251  0.182400  1.624000
```

```r
print(summary(aplltd$diff1_monthly))
```

```
##       Min.    1st Qu.     Median       Mean    3rd Qu.       Max.
## -2.255000 -0.307700 -0.061370 -0.004056  0.539700  1.589000
```

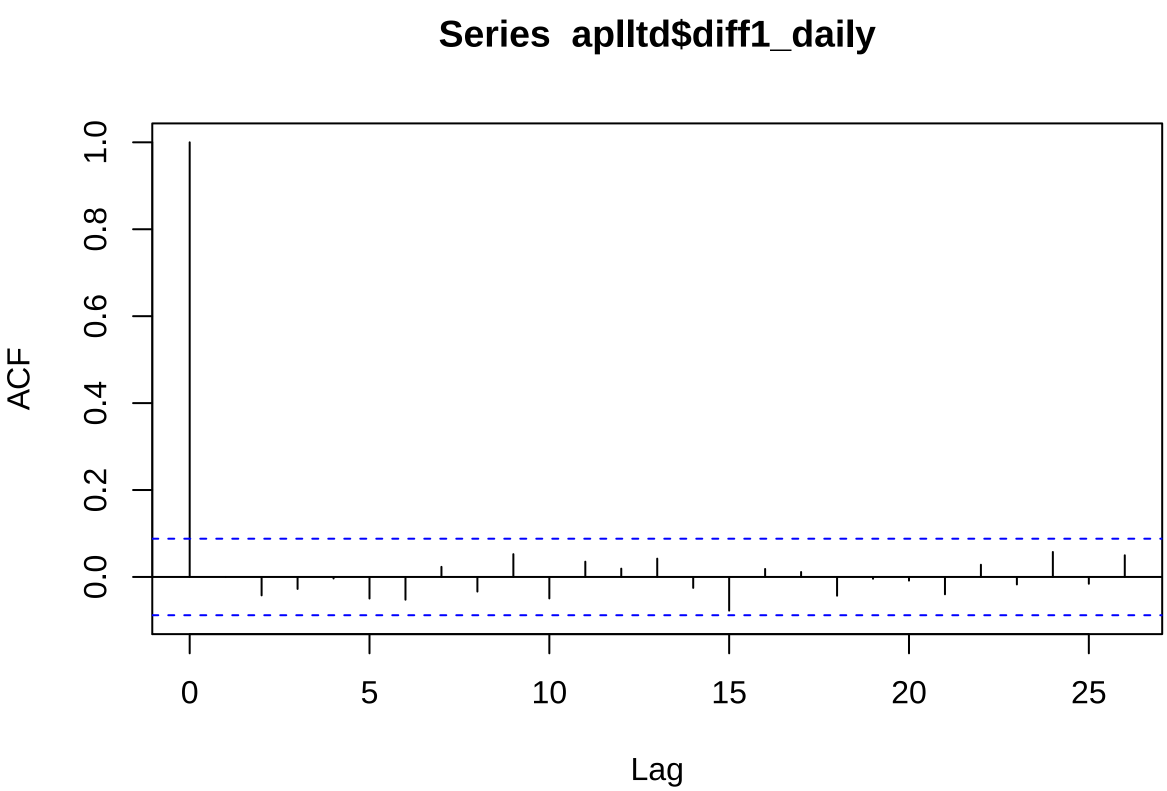
- Actual Raw data is non stationary
- By analysing the data it's clear that some integrating effect exists, which can be removed by taking first order difference of the signal
- Stationary component of the data is estimated using first order difference of the time series
- Resulting plot after removing integrating effect are stationary

Trend Estimation

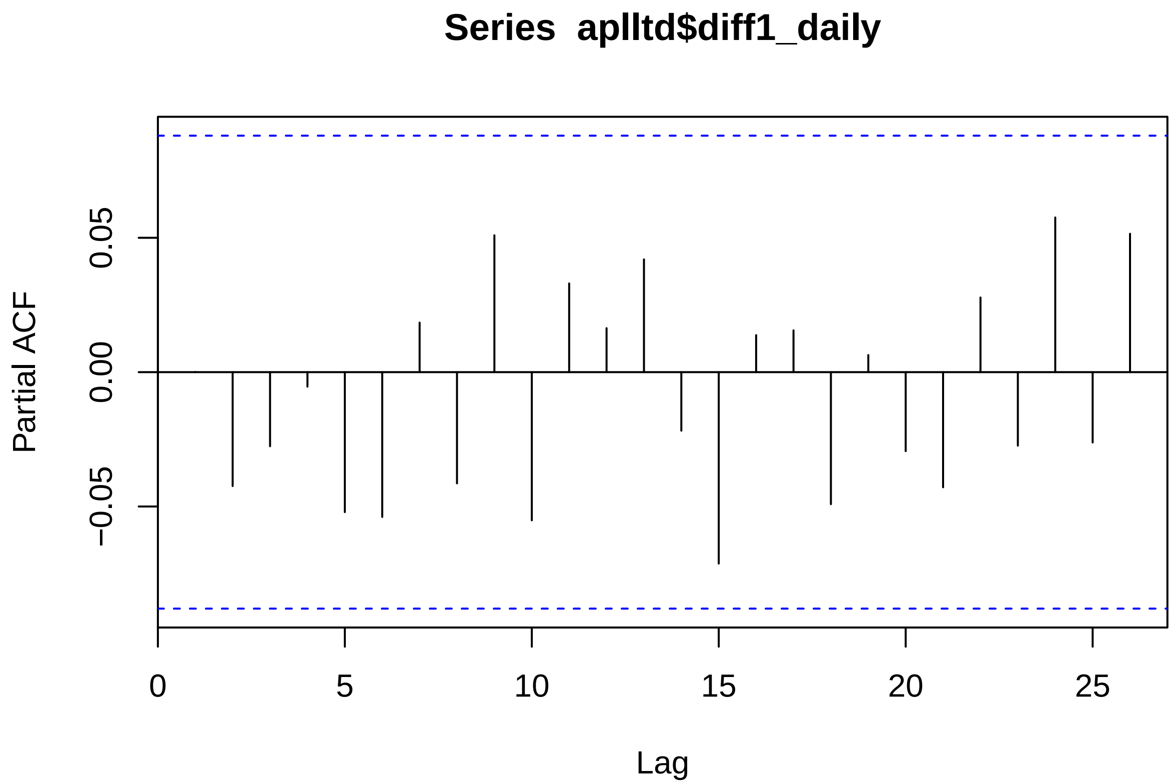
In this section we'll try to seperate predictable part of the series from random white noise. This is acheved using ACF PACF analysis.

Daily Series

```r
acf(aplltd$diff1_daily)
```




```r
pacf(aplltd$diff1_daily)
```

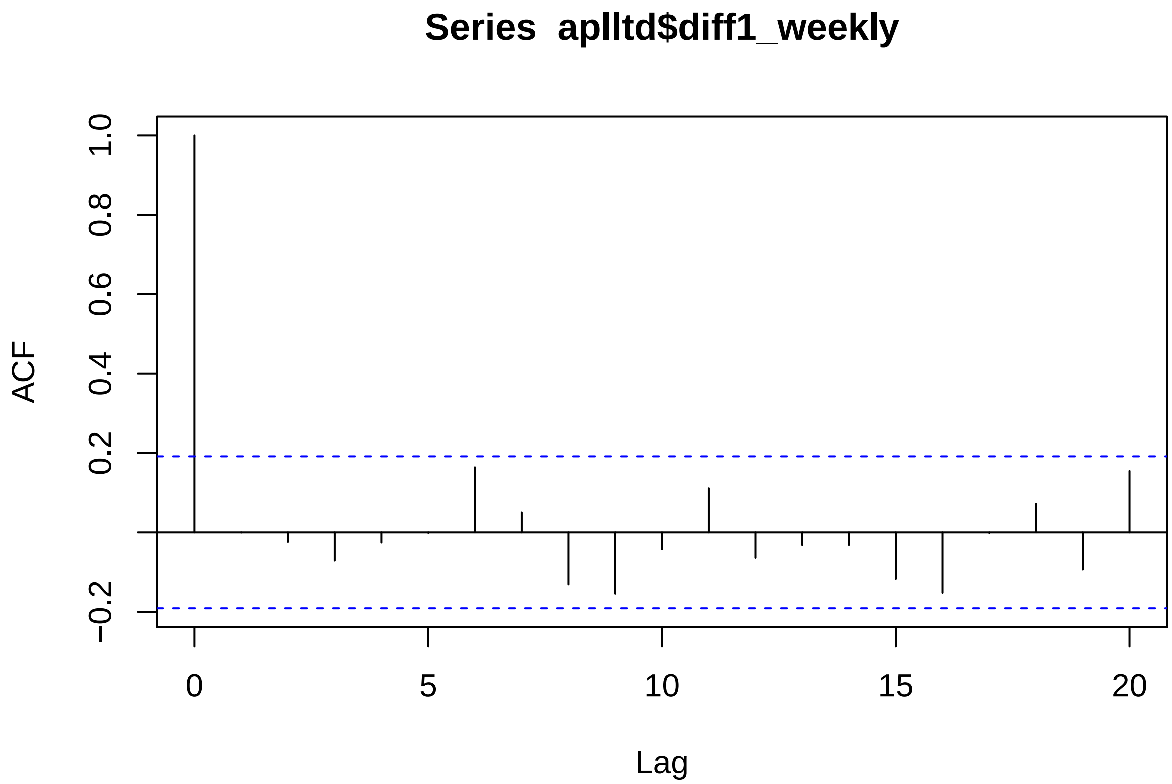



```
# As ACF and PACF are inbetween significant bands, this shows that signal
# is random (i.e there exists no trend to model) and stationary
```

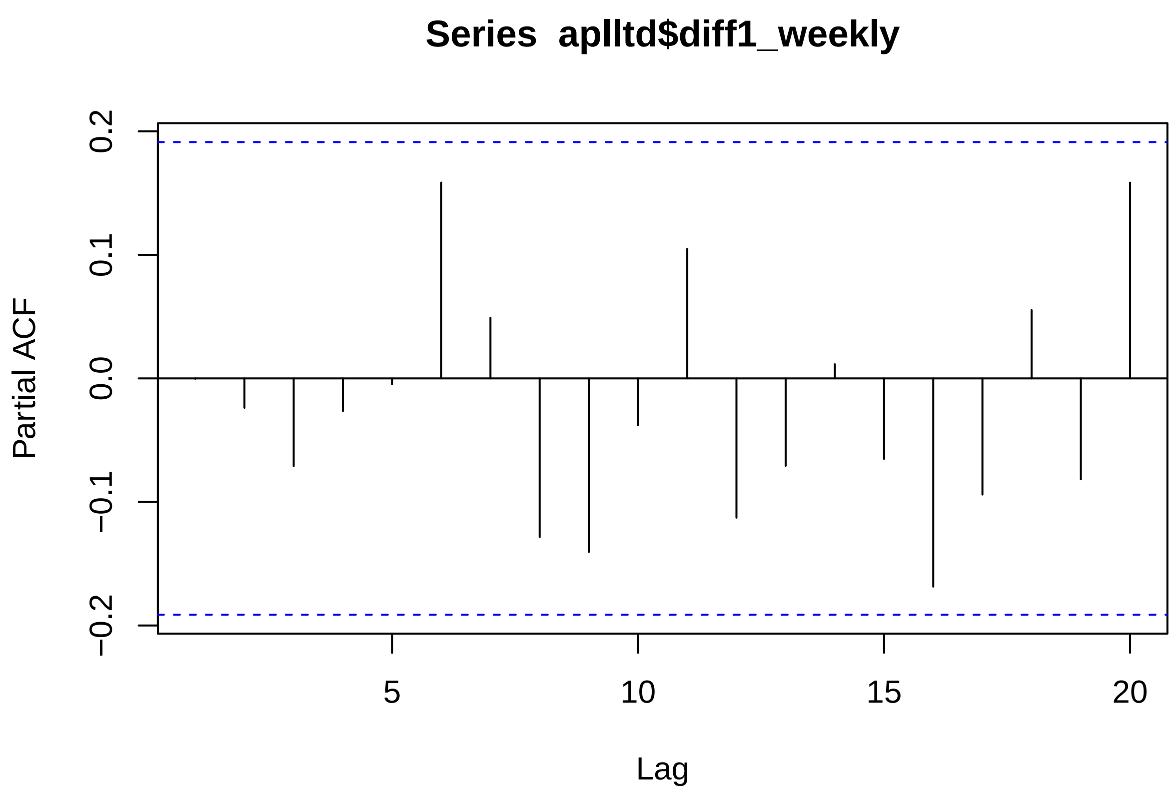
- As it can be seen that acf at any lag is bounded by significant limits, which means that the stationary series obtained after taking first order difference is white noise.

Weekly Series

```
acf(aplltd$diff1_weekly)
```




```r
pacf(aplltd$diff1_weekly)
```

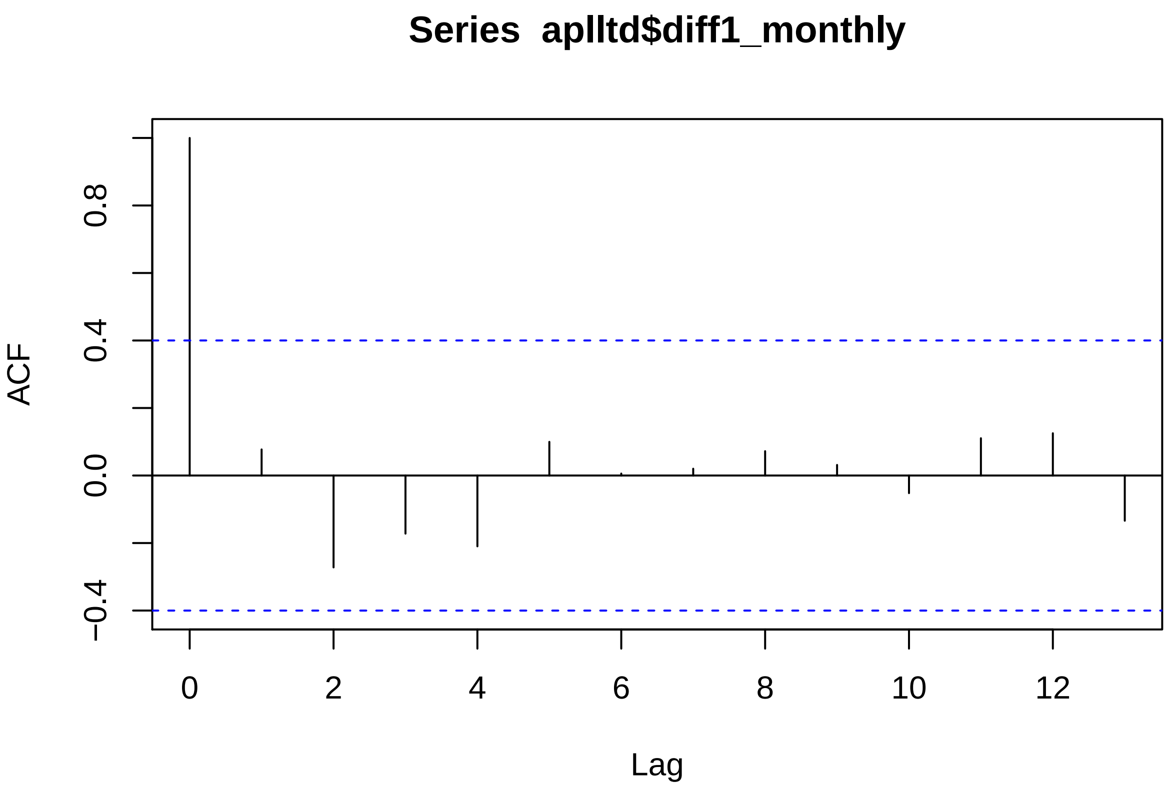



```r
# As ACF and PACF are inbetween significant bands, this shows that signal
# is random (i.e there exists no trend to model) and stationary
```

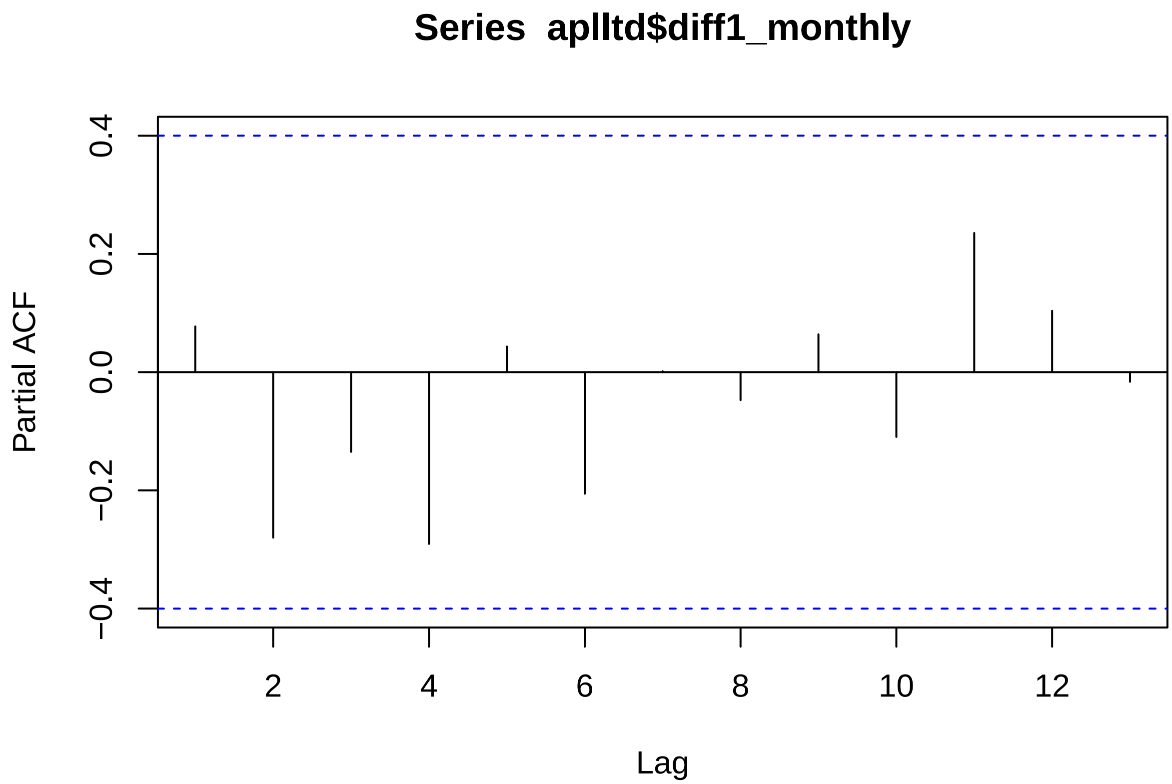
- As it can be seen that acf at any lag is bounded by significant limits, which means that the stationary series obtained after taking first order difference is white noise.

Monthly Series

```
acf(aplltd$diff1_monthly)
```




```
pacf(aplltd$diff1_monthly)
```




```
# As ACF and PACF are inbetween significant bands, this shows that signal
# is random (i.e there exists no trend to model) and stationary
```

- As it can be seen that acf at any lag is bounded by significant limits, which means that the stationary series obtained after taking first order difference is white noise.

Sun Pharmaceutical Industries Limited (SUNPHARMA.NS)

```
sunpharma = NULL
sunpharma$raw_daily = loadCSVData('../Data/SUNPHARMA/SUNPHARMA.NS_daily.csv')
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   443.8   523.8   567.7   577.4   637.3   721.5
```

```
sunpharma$raw_weekly = loadCSVData('../Data/SUNPHARMA/SUNPHARMA.NS_weekly.csv')
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   450.0   525.0   570.4   578.4   640.3   720.3
```

```
sunpharma$raw_monthly = loadCSVData('../Data/SUNPHARMA/SUNPHARMA.NS_monthly.csv')
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   480.4   531.7   568.5   571.1   623.2   688.2
```

```
sunpharma$sd_daily = standardize(sunpharma$raw_daily)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## -2.0070 -0.8044 -0.1459  0.0000  0.9009  2.1650
```

```r
sunpharma$sd_weekly = standardize(sunpharma$raw_weekly)
```

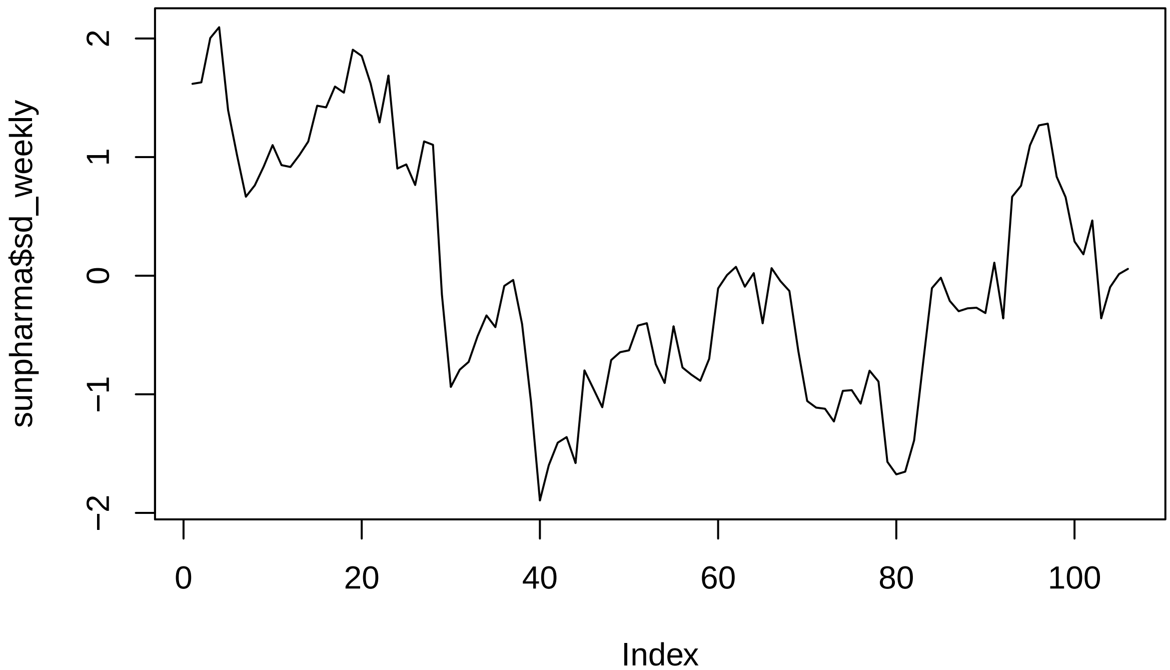
```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## -1.8950 -0.7879 -0.1180  0.0000  0.9132  2.0950
```

```r
sunpharma$sd_monthly = standardize(sunpharma$raw_monthly)
```

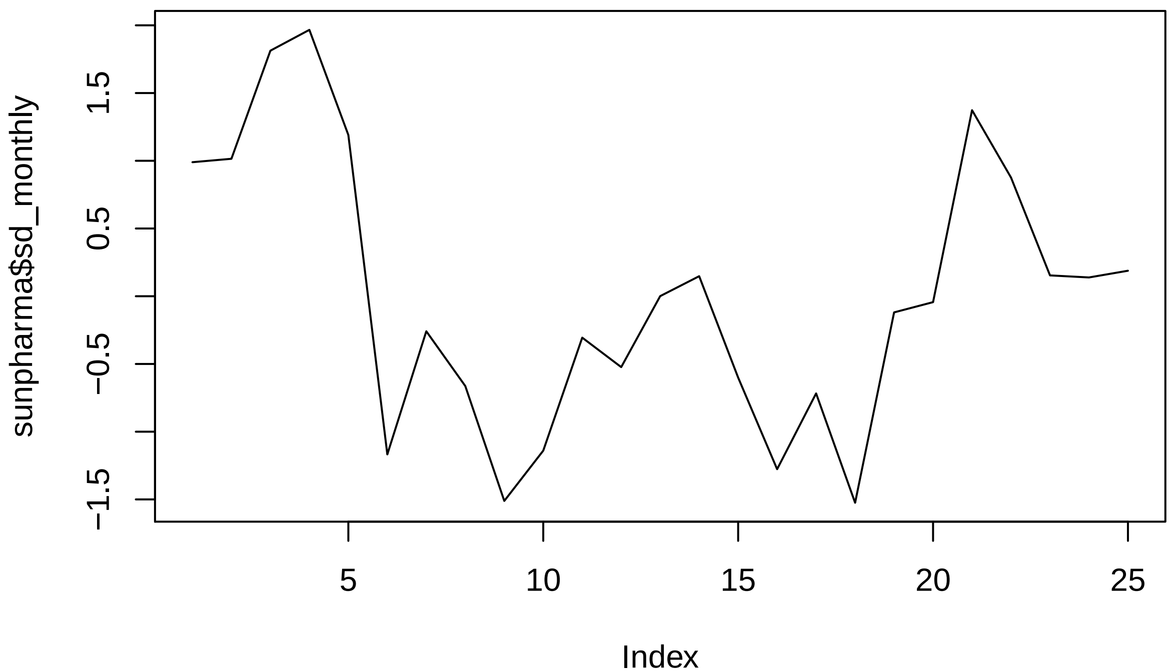
```
##     Min.  1st Qu.   Median     Mean  3rd Qu.     Max.
## -1.52500 -0.66280 -0.04372  0.00000  0.87610  1.96600
```

```r
# plot all the data
plot(sunpharma$sd_daily, type='l')
```


```r
plot(sunpharma$sd_weekly, type='l')
```

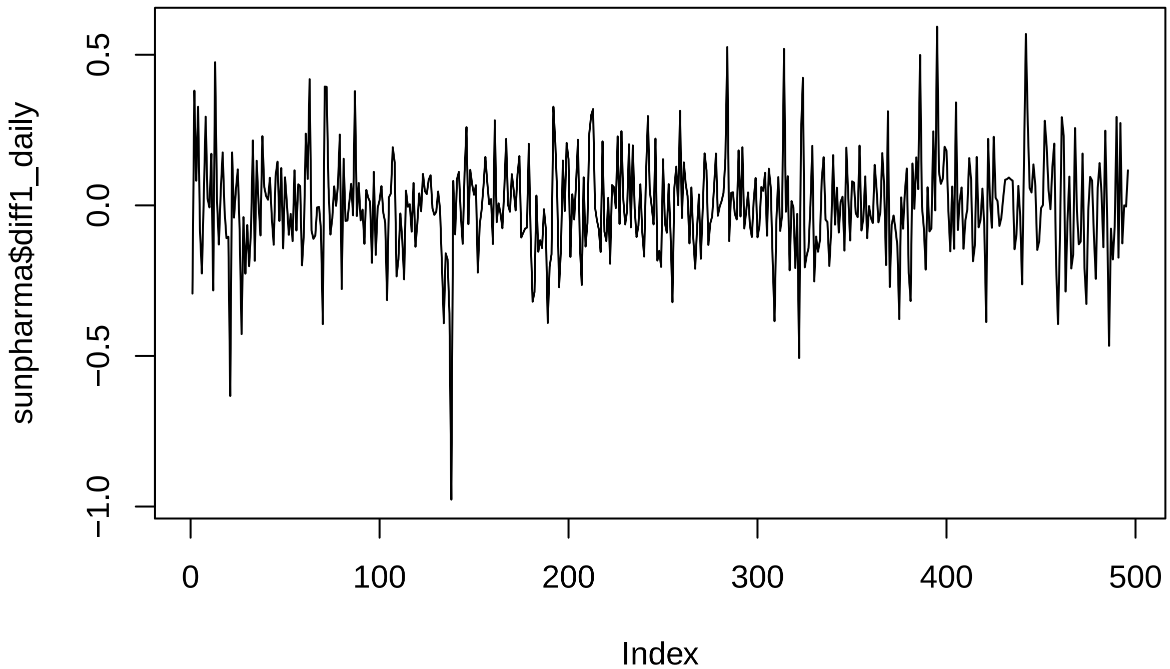



```
plot(sunpharma$sd_monthly, type='l')
```

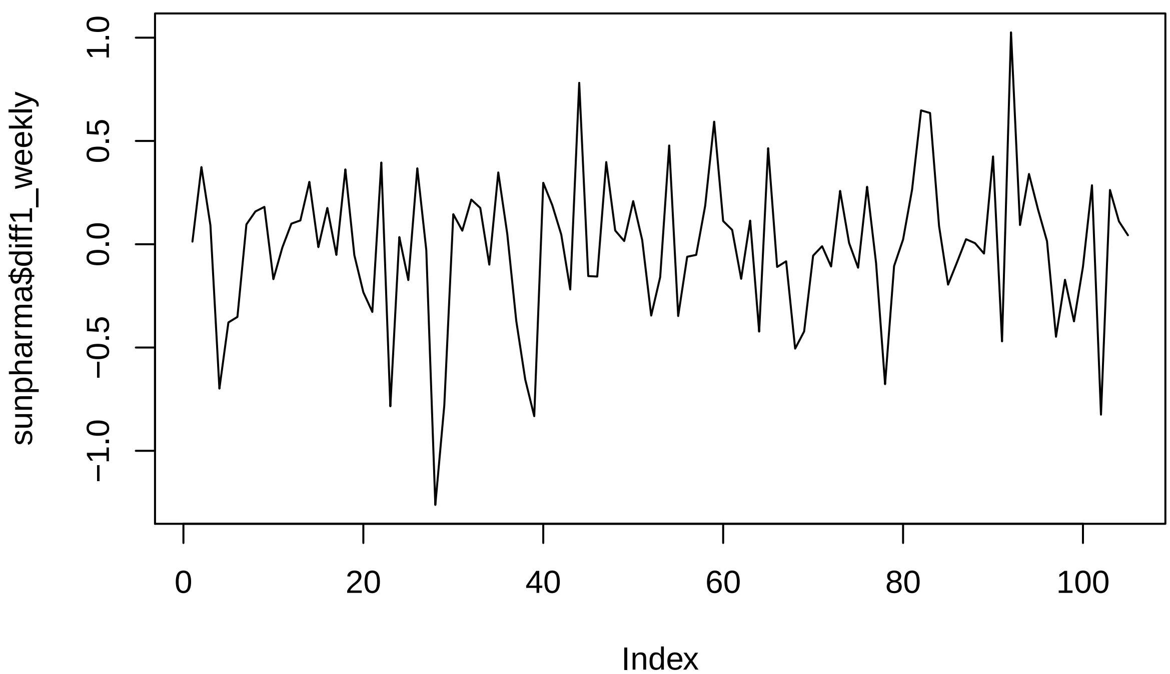



```
# Based on plots we can clearly see that given data is
# non stationary (data has some integrating effect, which should estimated)
# which can be converted to stationary by taking difference of order = n

sunpharma$diff1_daily = diff(sunpharma$sd_daily)
sunpharma$diff1_weekly = diff(sunpharma$sd_weekly)
sunpharma$diff1_monthly = diff(sunpharma$sd_monthly)

# Plotting differenced data
```

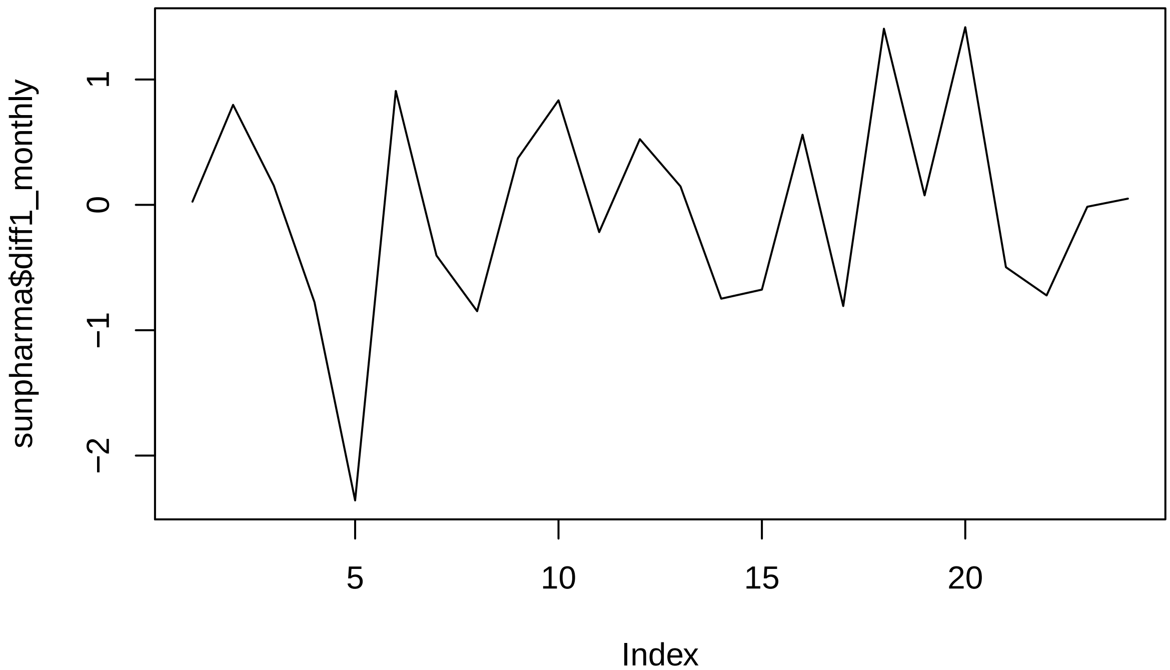
```
plot(sunpharma$diff1_daily, type='l')
```




```
plot(sunpharma$diff1_weekly, type='l')
```




```
plot(sunpharma$diff1_monthly, type='l')
```




```r
# Differenced plots seems to be stationary this can also be verified by summary of the data
print(summary(sunpharma$diff1_daily))
```

```
##      Min.   1st Qu.    Median      Mean   3rd Qu.      Max.
## -0.976900 -0.097060 -0.009011 -0.002288  0.087290  0.593200
```

```r
print(summary(sunpharma$diff1_weekly))
```

```
##     Min.  1st Qu.   Median     Mean  3rd Qu.     Max.
## -1.26200 -0.16750  0.01476 -0.01486  0.18080  1.02600
```

```r
print(summary(sunpharma$diff1_monthly))
```

```
##     Min.  1st Qu.   Median     Mean  3rd Qu.     Max.
## -2.35800 -0.68780  0.03738 -0.03339  0.53300  1.41700
```

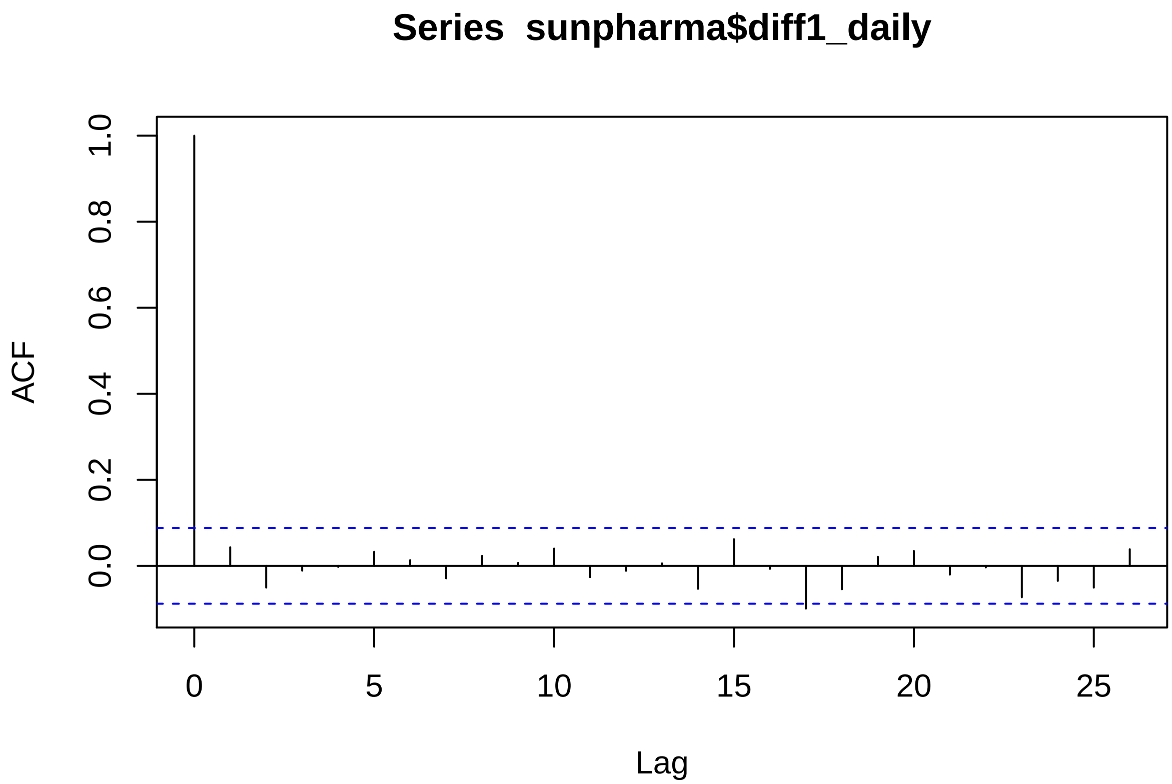
- Actual Raw data is non stationary
- By analysing the data it's clear that some integrating effect exists, which can be removed by taking first order difference of the signal
- Stationary component of the data is estimated using first order difference of the time series
- Resulting plot after removing integrating effect are stationary

Trend Estimation

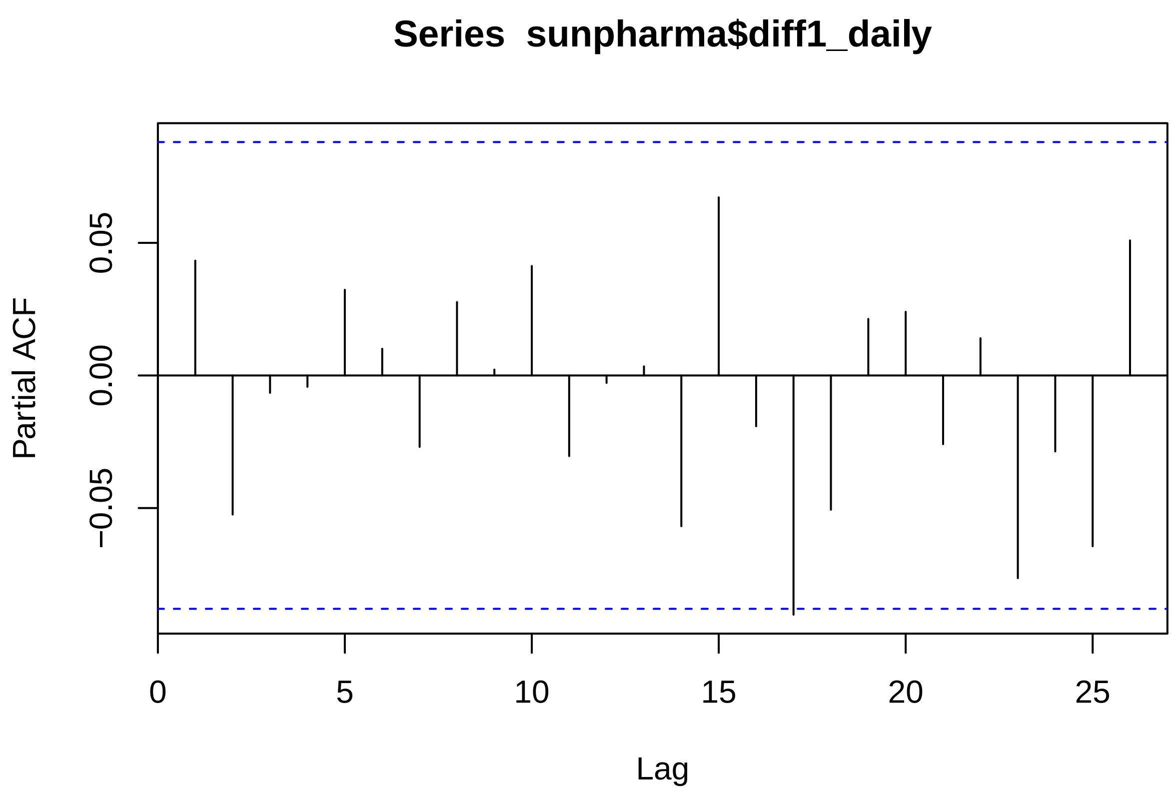
In this section we'll try to seperate predictable part of the series from random white noise. This is acheved using ACF PACF analysis.

Daily Series

```r
acf(sunpharma$diff1_daily)
```



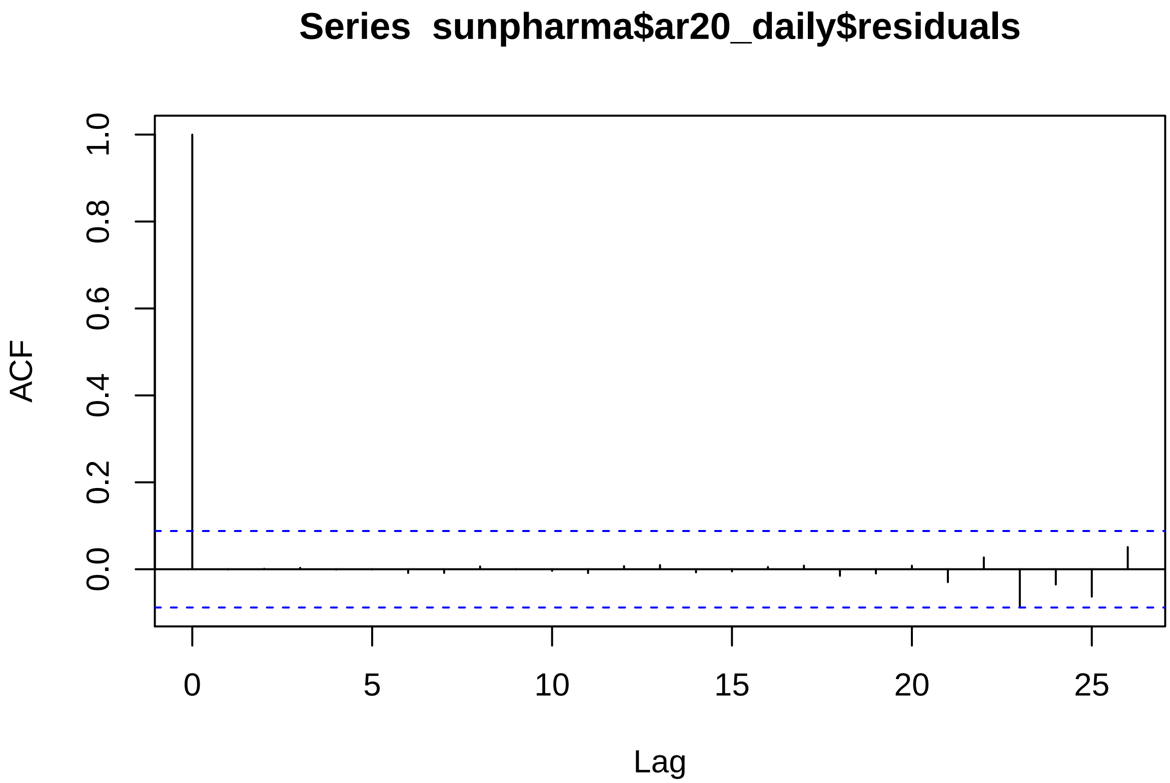

```
pacf(sunpharma$diff1_daily)
```



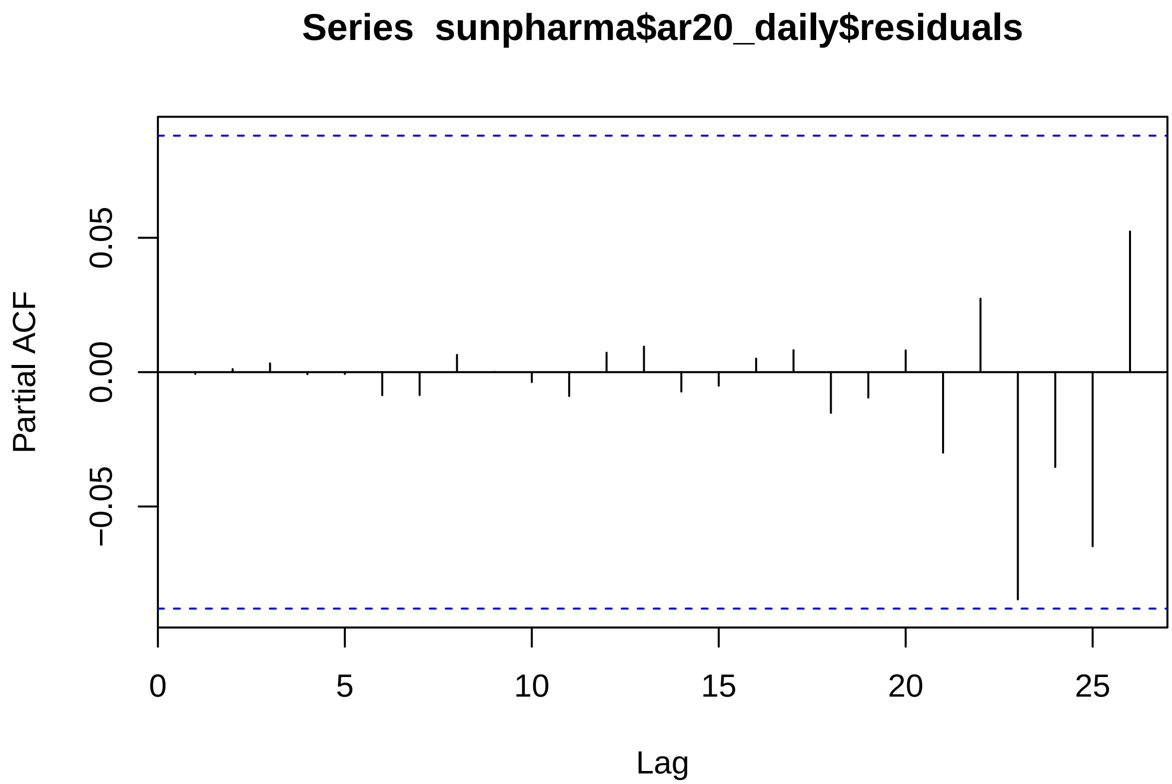

```
# As ACF and PACF are inbetween significant bands, this shows that signal
# is random (i.e there exists no trend to model) and stationary
```

- As it can be seen that pacf at lag = 20 exceeds significant bounds, which means Autoregressive model AR(20) can fit the data

```
sunpharma$ar20_daily = arima(sunpharma$diff1_daily, order=c(20,0,0))

# ACF of residuals should be white if model captures entire information
acf(sunpharma$ar20_daily$residuals)
```



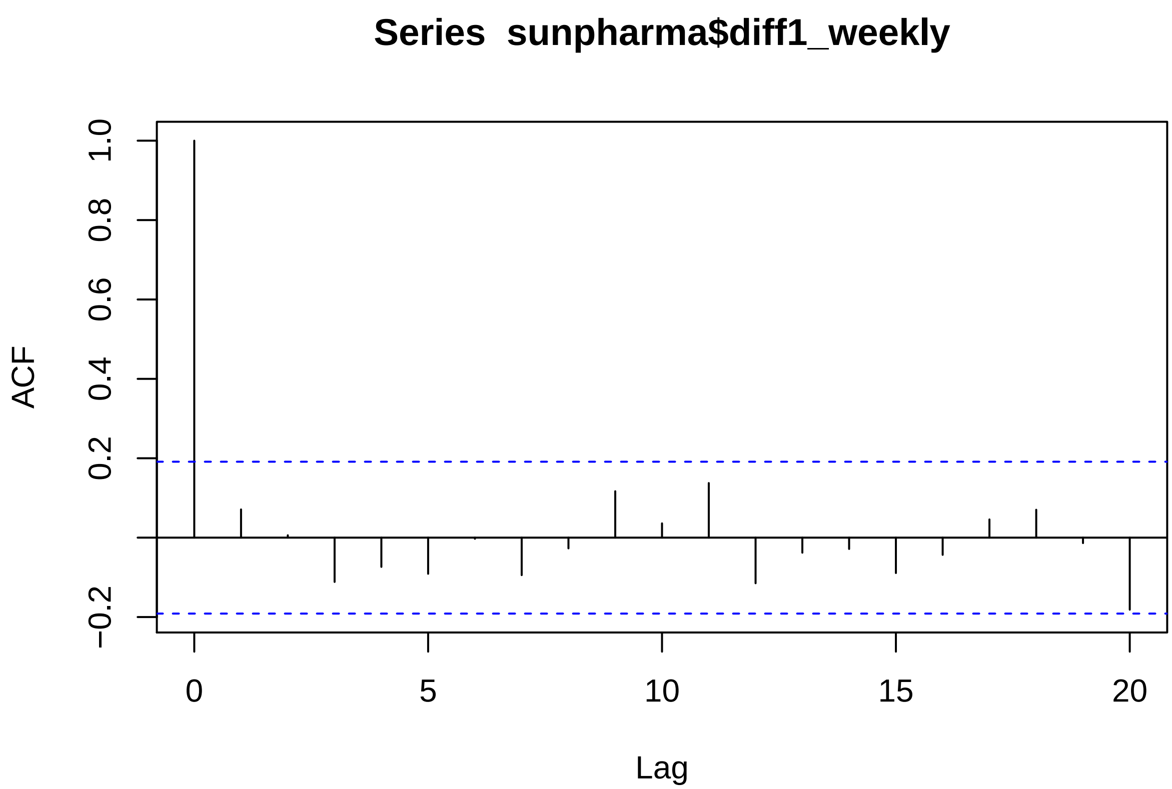

```
pacf(sunpharma$ar20_daily$residuals)
```



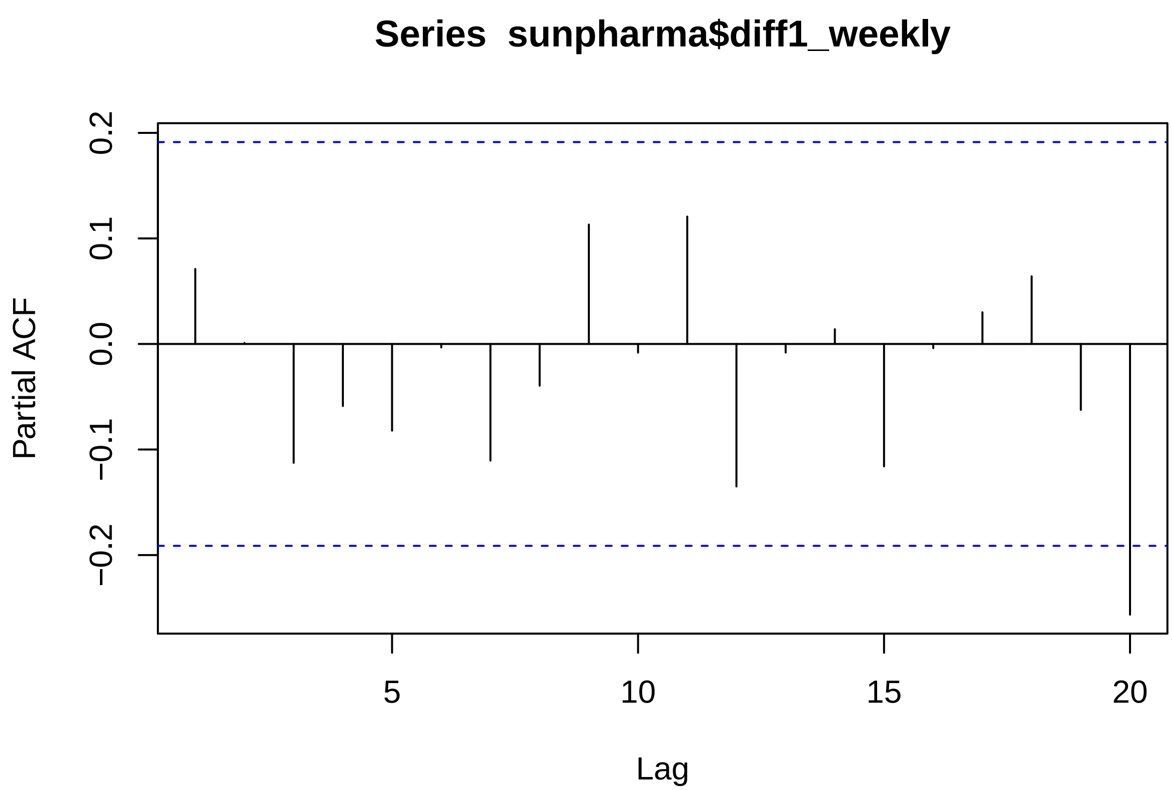
- By fitting AR(20) model predictable component of the data is exploited, residuals obtained forms white noise
- Auto regressive trend is followed in case of Daily data, Data can be made stationary after considering first order difference

Weekly Series

```
acf(sunpharma$diff1_weekly)
```




```r
pacf(sunpharma$diff1_weekly)
```

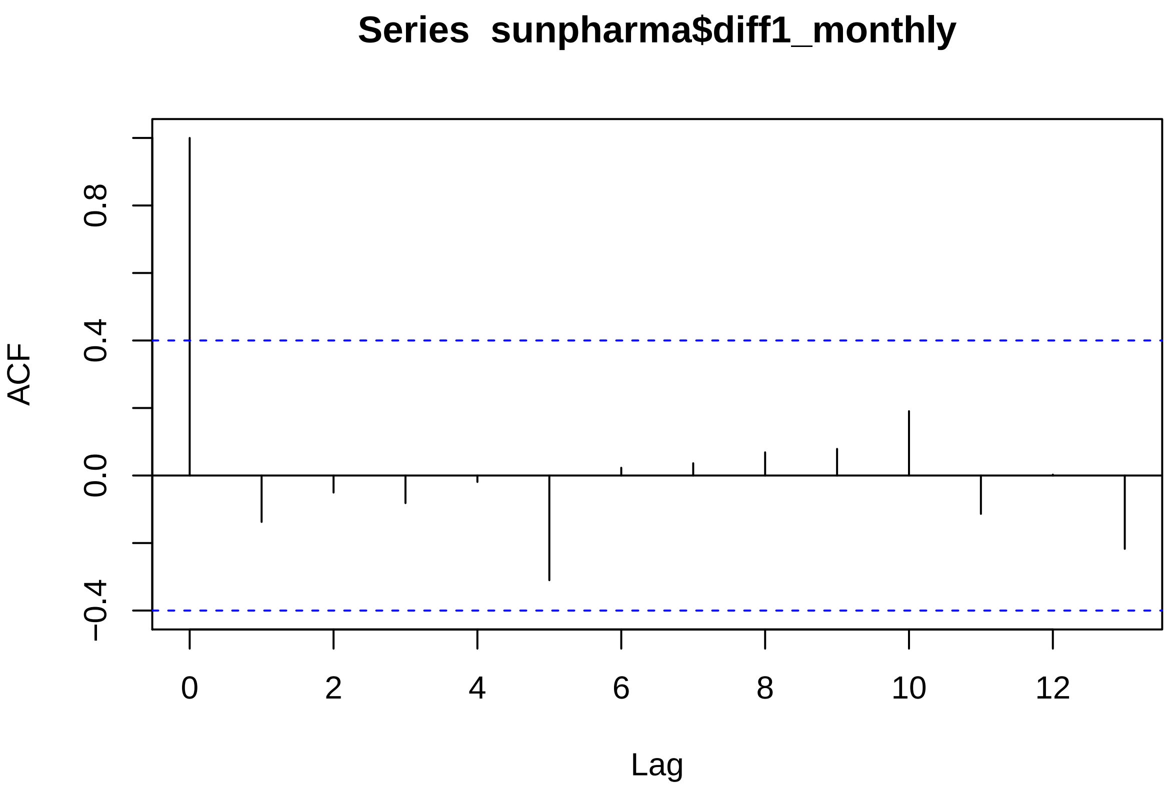



```r
# As ACF and PACF are inbetween significant bands, this shows that signal
# is random (i.e there exists no trend to model) and stationary
```

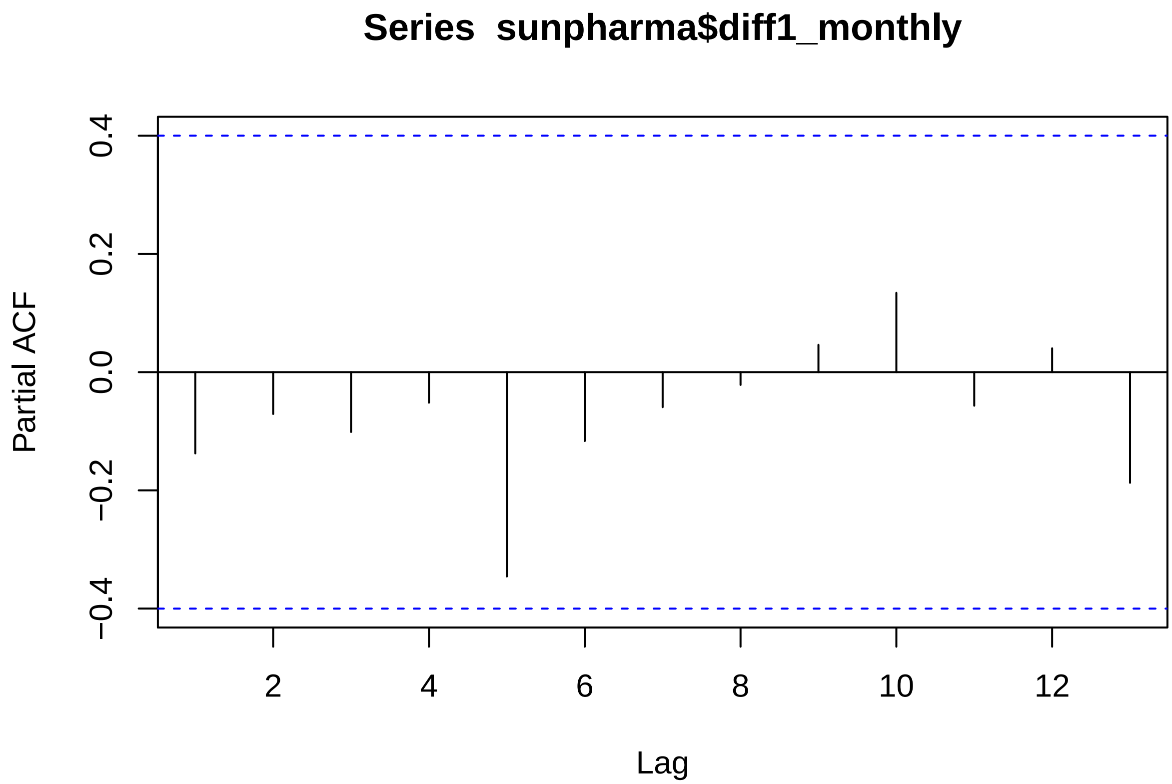
- As it can be seen that acf at any lag is bounded by significant limits, which means that the stationary series obtained after taking first order difference is white noise.

Monthly Series

```r
acf(sunpharma$diff1_monthly)
```




```r
pacf(sunpharma$diff1_monthly)
```




```
# As ACF and PACF are inbetween significant bands, this shows that signal
# is random (i.e there exists no trend to model) and stationary
```

- As it can be seen that acf at any lag is bounded by significant limits, which means that the stationary series obtained after taking first order difference is white noise.

Aurobindo Pharma Limited (AUROPHARMA.NS)

```
auropharma = NULL
auropharma$raw_daily = loadCSVData('../Data/AUROPHARMA/AUROPHARMA.NS_daily.csv')
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   512.4   613.8   672.4   669.5   719.8   800.8
```

```
auropharma$raw_weekly = loadCSVData('../Data/AUROPHARMA/AUROPHARMA.NS_weekly.csv')
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   532.5   614.8   673.7   672.8   723.9   800.8
```

```
auropharma$raw_monthly = loadCSVData('../Data/AUROPHARMA/AUROPHARMA.NS_monthly.csv')
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   556.3   613.9   681.8   674.9   718.6   794.2
```

```
auropharma$sd_daily = standardize(auropharma$raw_daily)
```

```
##     Min.  1st Qu.   Median     Mean  3rd Qu.     Max.
## -2.44700 -0.86660  0.04531  0.00000  0.78350  2.04600
```

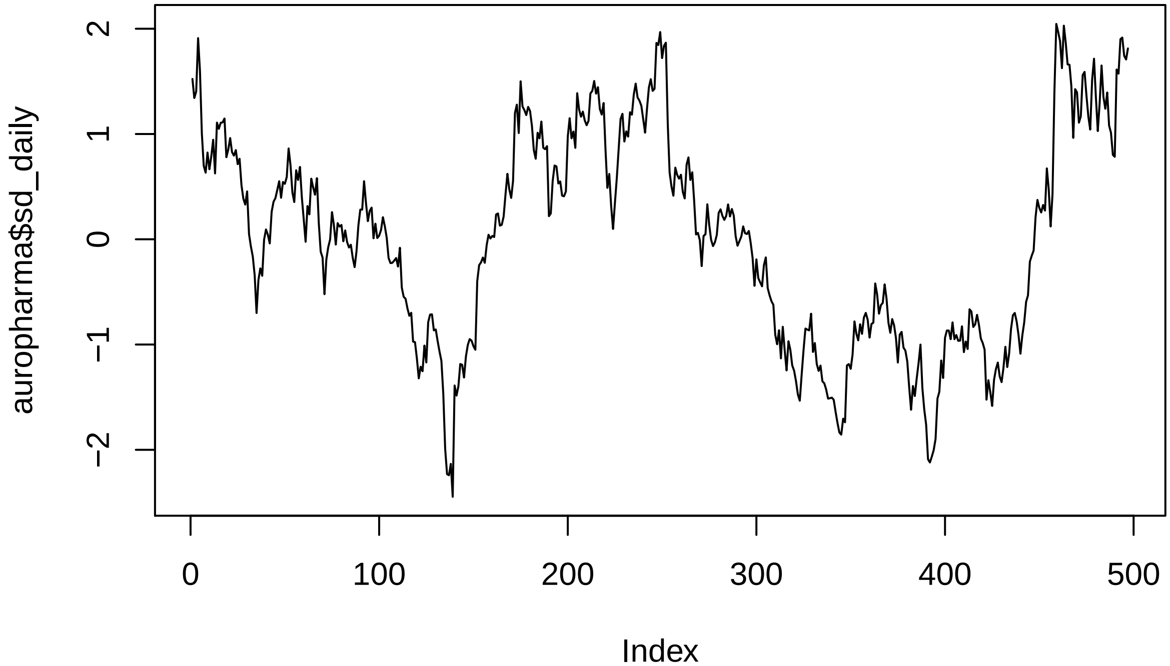
```r
auropharma$sd_weekly = standardize(auropharma$raw_weekly)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## -2.1430 -0.8861  0.0139  0.0000  0.7808  1.9560
```

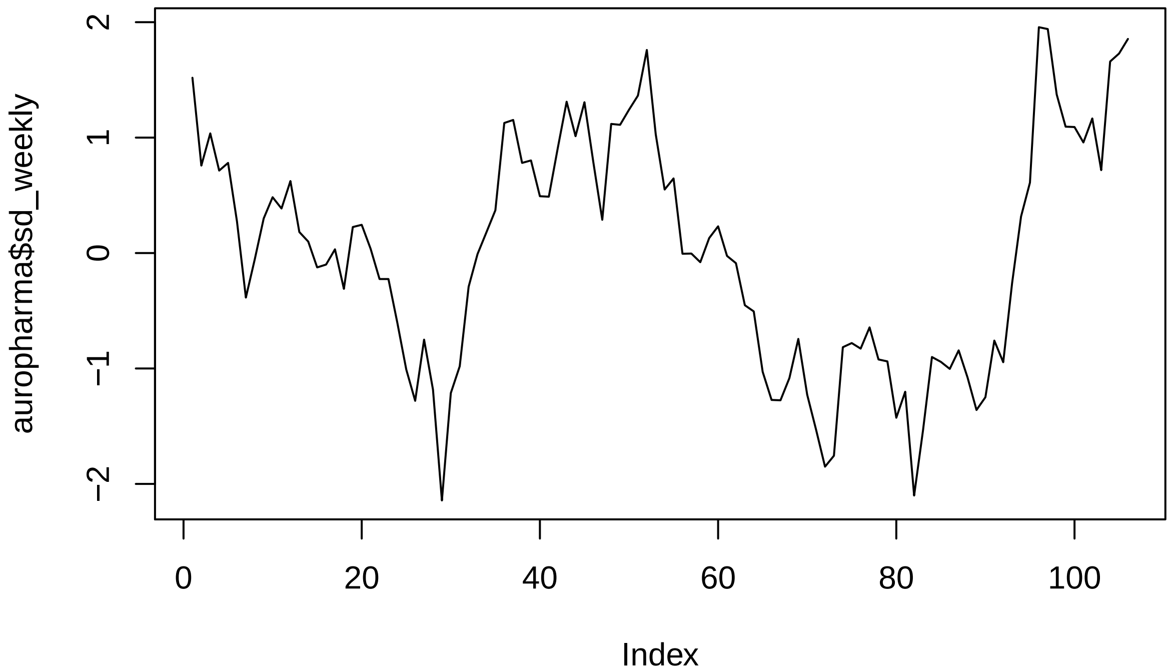
```r
auropharma$sd_monthly = standardize(auropharma$raw_monthly)
```

```
##     Min.  1st Qu.   Median     Mean  3rd Qu.     Max.
## -1.67700 -0.86250  0.09762  0.00000  0.61800  1.68700
```

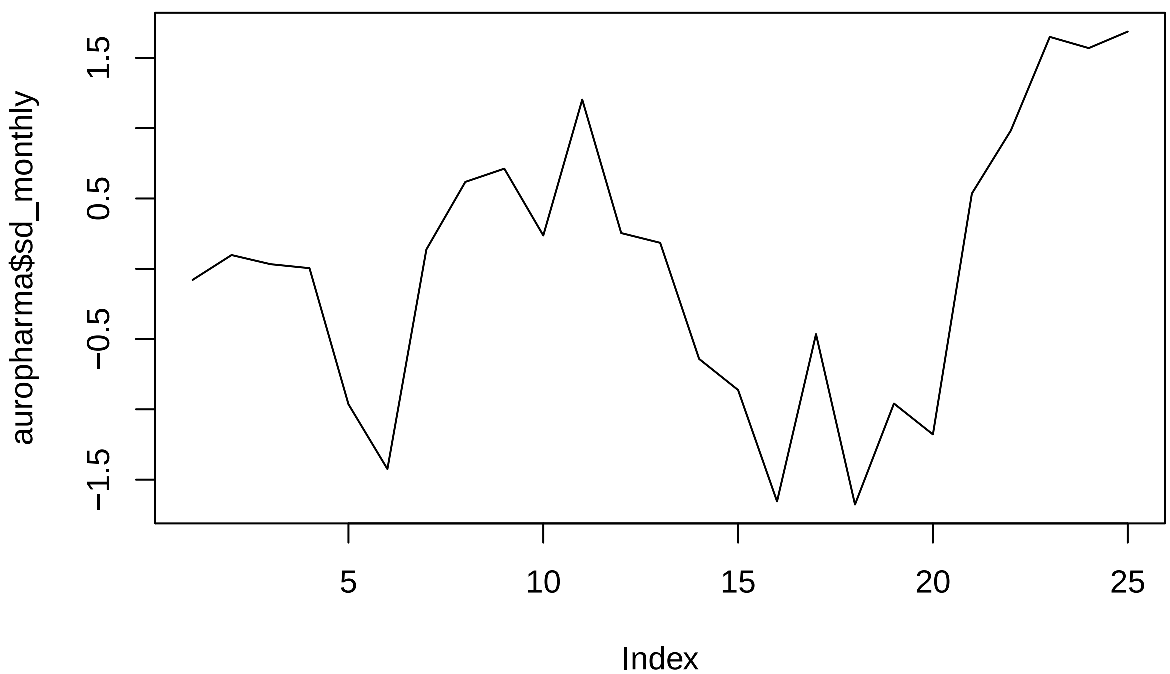
```r
# plot all the data
plot(auropharma$sd_daily, type='l')
```




```r
plot(auropharma$sd_weekly, type='l')
```

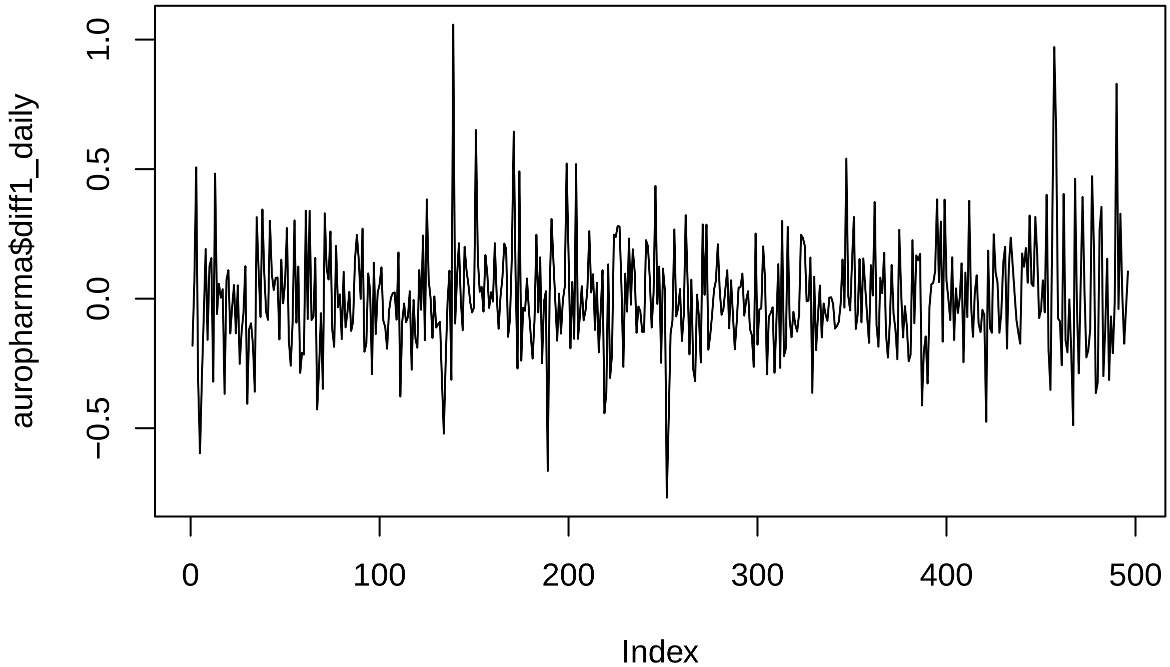



```
plot(auropharma$sd_monthly, type='l')
```

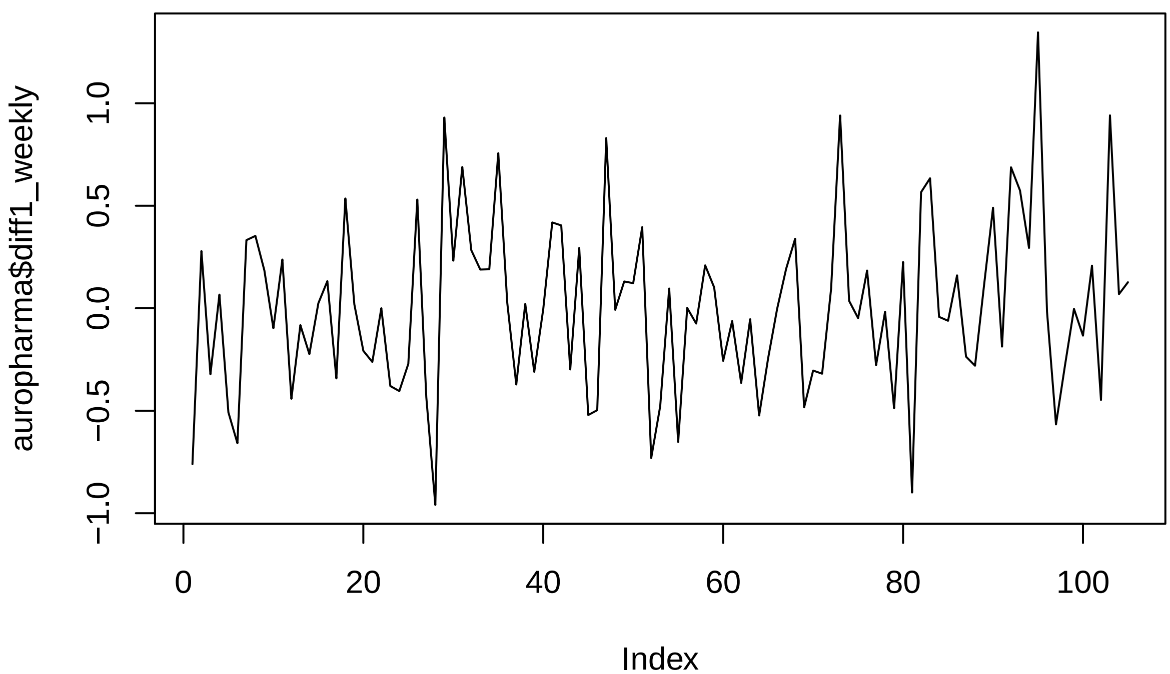



```
# Based on plots we can clearly see that given data is
# non stationary (data has some integrating effect, which should estimated)
# which can be converted to stationary by taking difference of order = n

auropharma$diff1_daily = diff(auropharma$sd_daily)
auropharma$diff1_weekly = diff(auropharma$sd_weekly)
auropharma$diff1_monthly = diff(auropharma$sd_monthly)

# Plotting differenced data
```

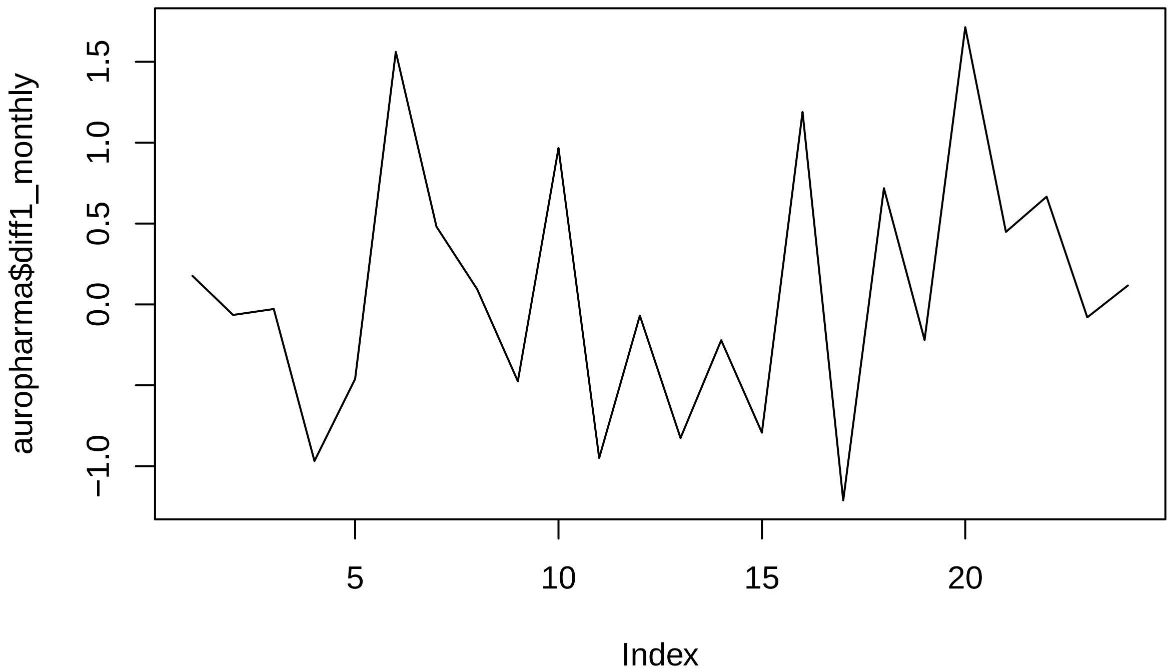
```r
plot(auropharma$diff1_daily, type='l')
```




```r
plot(auropharma$diff1_weekly, type='l')
```




```r
plot(auropharma$diff1_monthly, type='l')
```




```r
# Differenced plots seems to be stationary this can also be verified by summary of the data
print(summary(auropharma$diff1_daily))
```

```
##      Min.   1st Qu.    Median      Mean   3rd Qu.      Max.
## -0.767800 -0.127100 -0.018690  0.000584  0.110000  1.057000
```

```r
print(summary(auropharma$diff1_weekly))
```

```
##      Min.   1st Qu.    Median      Mean   3rd Qu.      Max.
## -0.959400 -0.298700 -0.003056  0.003201  0.225300  1.346000
```

```r
print(summary(auropharma$diff1_monthly))
```

```
##     Min.  1st Qu.   Median     Mean  3rd Qu.     Max.
## -1.21200 -0.46450 -0.04666  0.07359  0.52710  1.71400
```

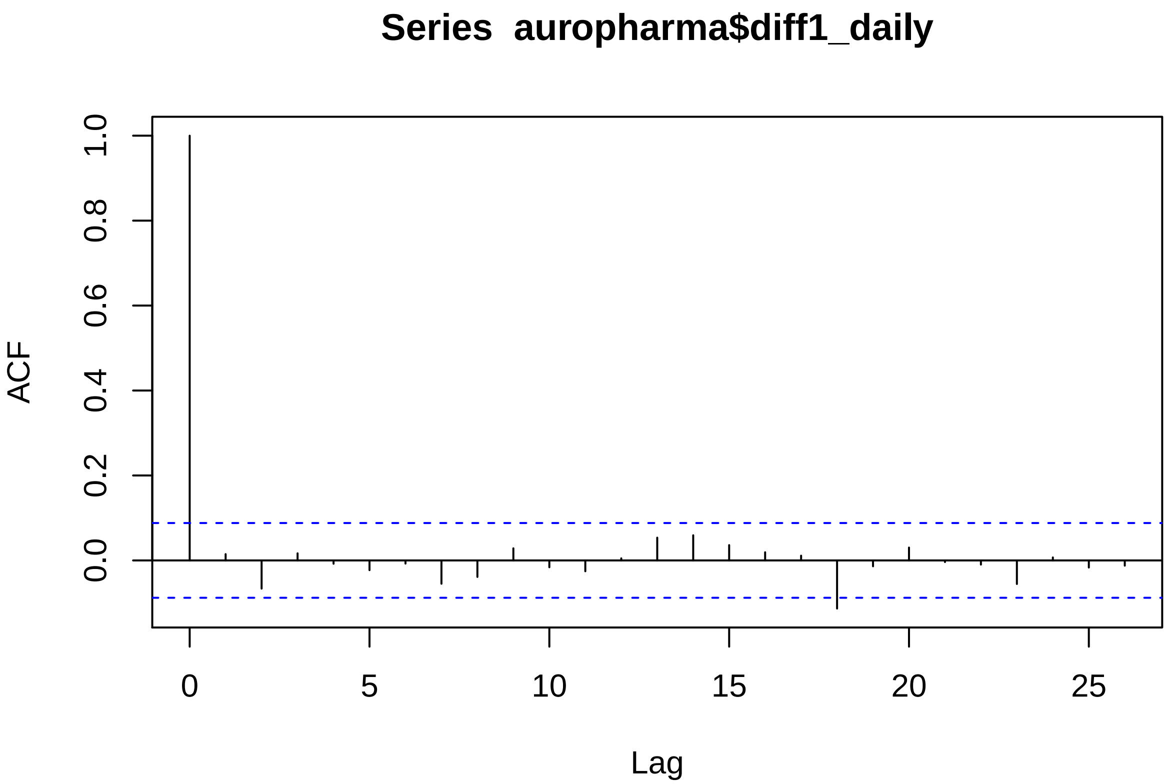
- Actual Raw data is non stationary
- By analysing the data it's clear that some integrating effect exists, which can be removed by taking first order difference of the signal
- Stationary component of the data is estimated using first order difference of the time series
- Resulting plot after removing integrating effect are stationary

Trend Estimation

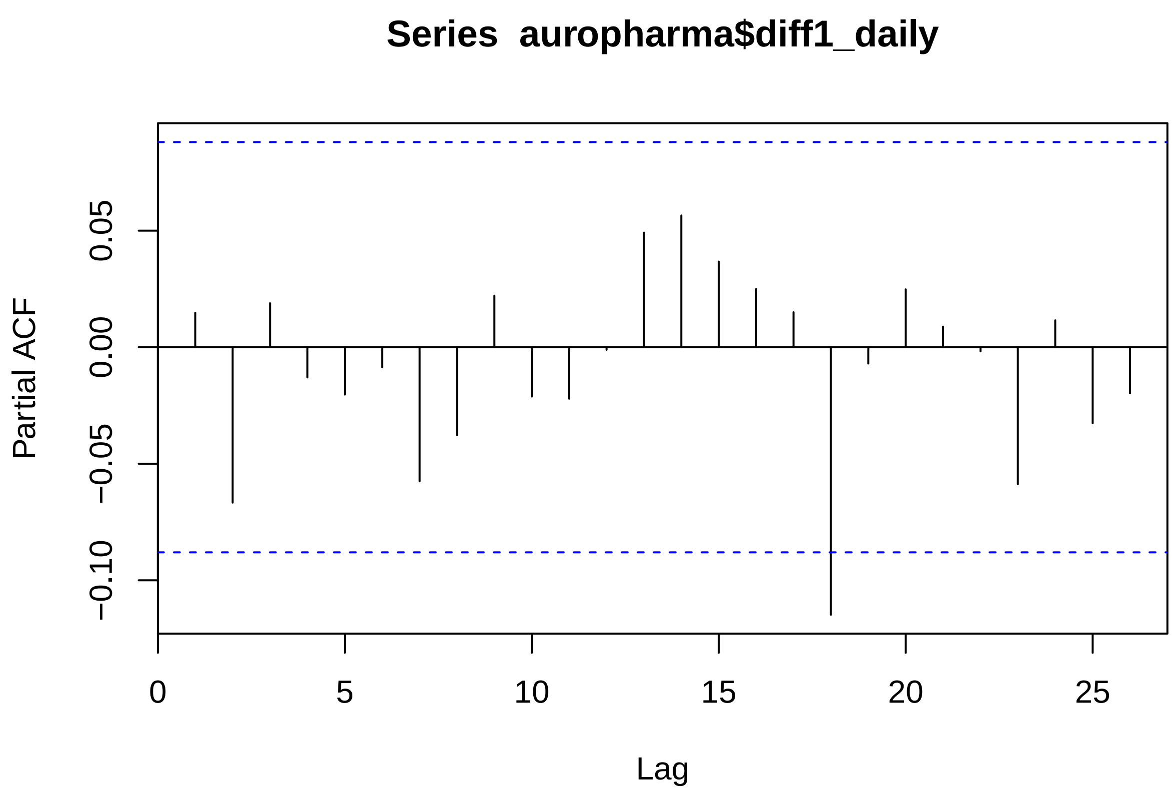
In this section we'll try to seperate predictable part of the series from random white noise. This is acheved using ACF PACF analysis.

Daily Series

```r
acf(auropharma$diff1_daily)
```




```r
pacf(auropharma$diff1_daily)
```



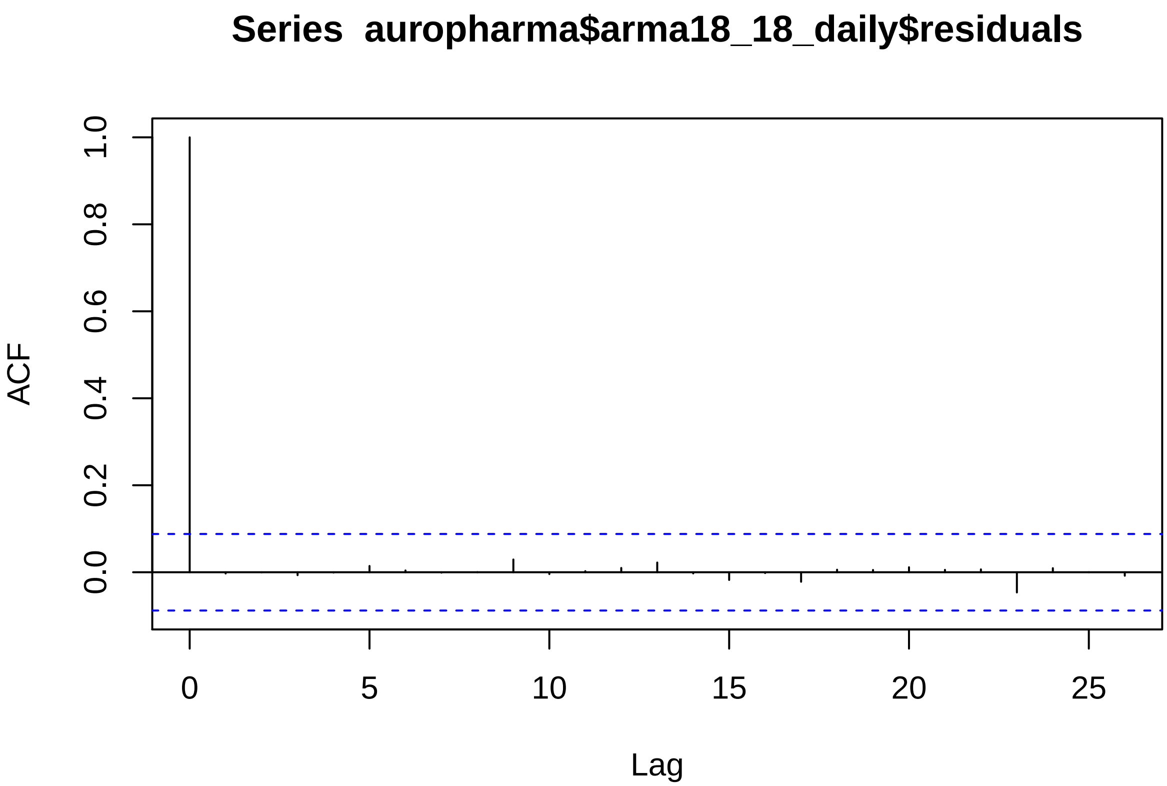

```r
# As ACF and PACF are inbetween significant bands, this shows that signal
# is random (i.e there exists no trend to model) and stationary
```

- As it can be seen that acf and pacf at lag = 18 exceeds significant bounds, which means Autoregressive model ARMA(18,18) can fit the data

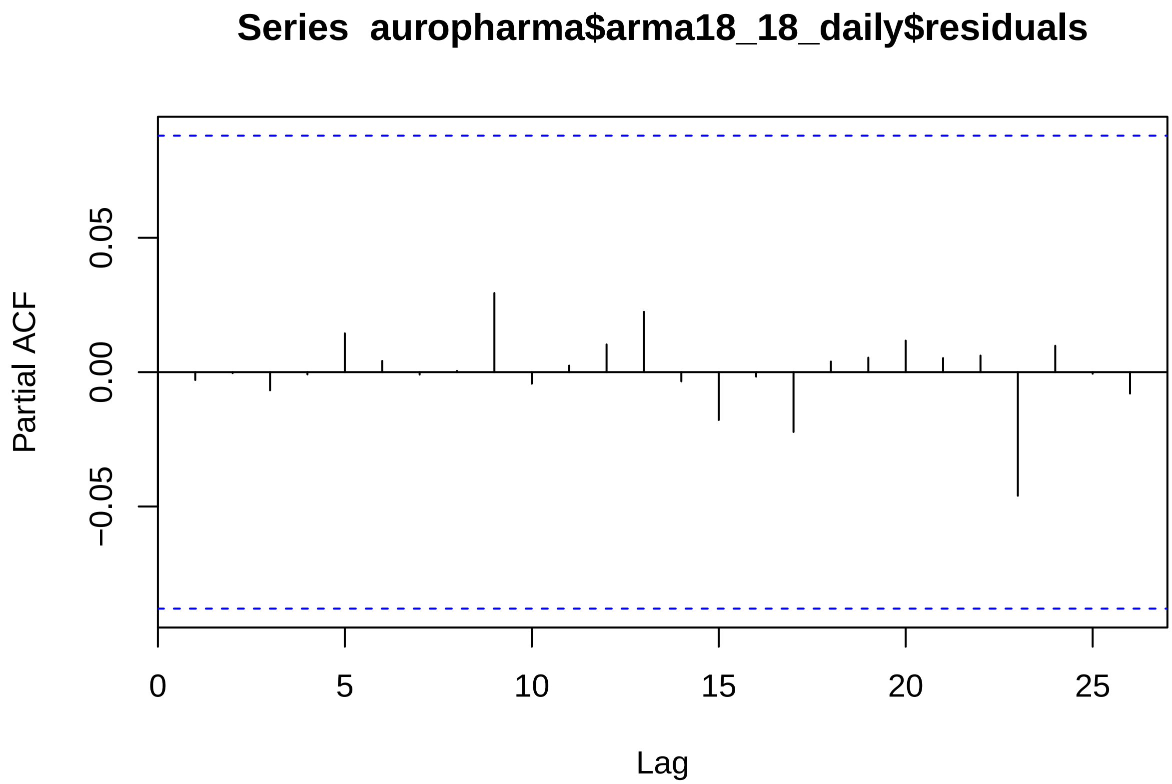
```
auropharma$arma18_18_daily = arima(auropharma$diff1_daily, order=c(18,0,18))
```

```
## Warning in arima(auropharma$diff1_daily, order = c(18, 0, 18)): possible
## convergence problem: optim gave code = 1
```

```
# ACF of residuals should be white if model captures entire information
acf(auropharma$arma18_18_daily$residuals)
```



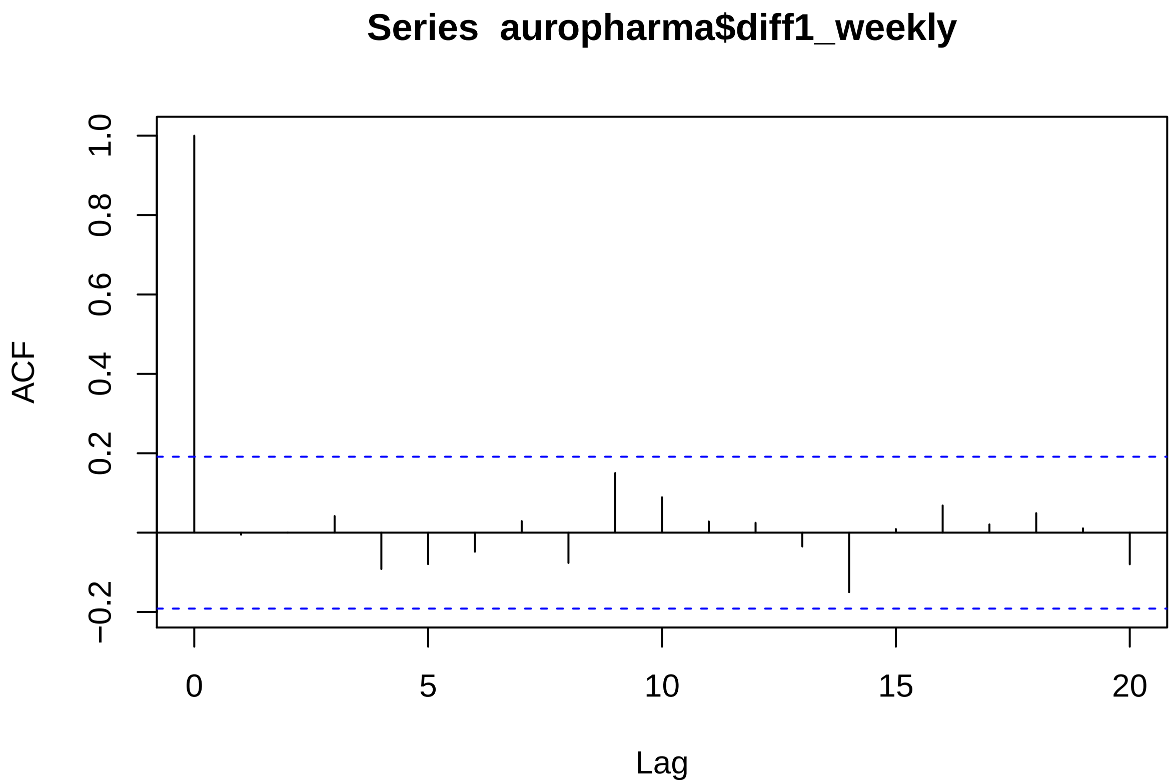

```
pacf(auropharma$arma18_18_daily$residuals)
```



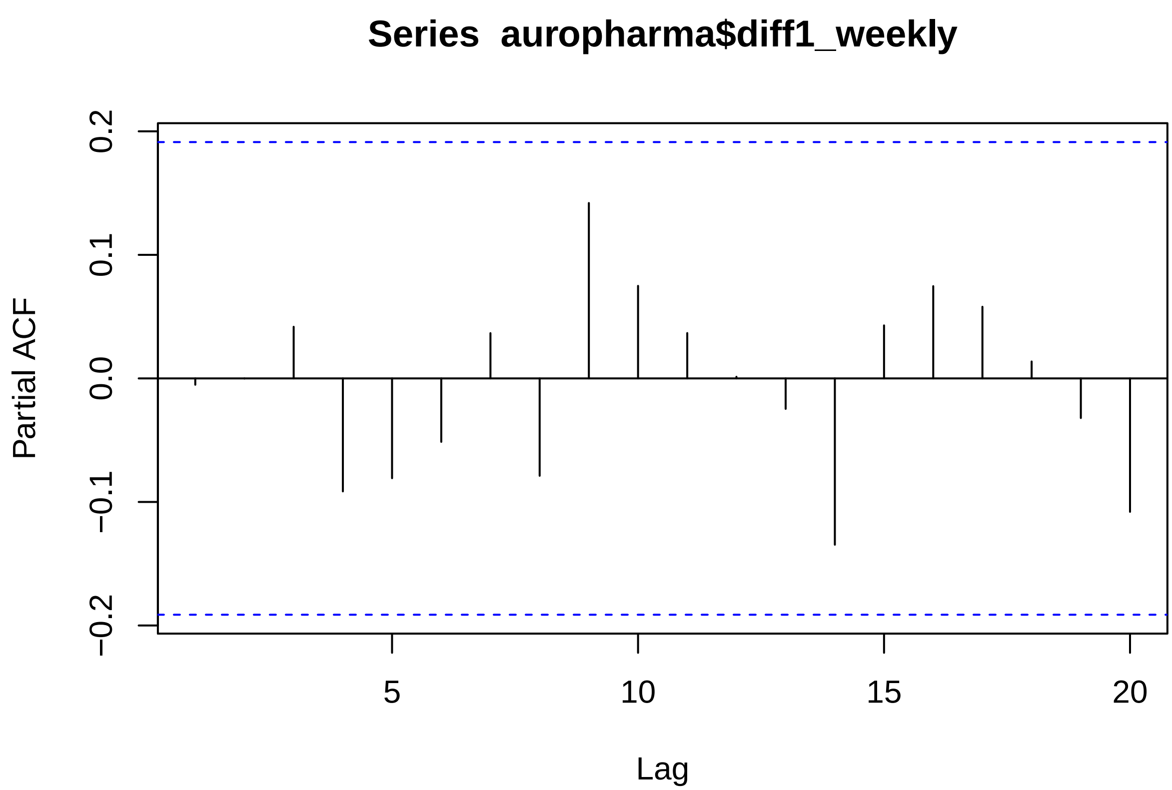
- By fitting ARMA(18, 18) model predictable component of the data is exploited, residuals obtained forms white noise
- AMA trend is followed in case of Daily data, Data can be made stationary after considering first order difference

Weekly Series

```
acf(auropharma$diff1_weekly)
```




```
pacf(auropharma$diff1_weekly)
```

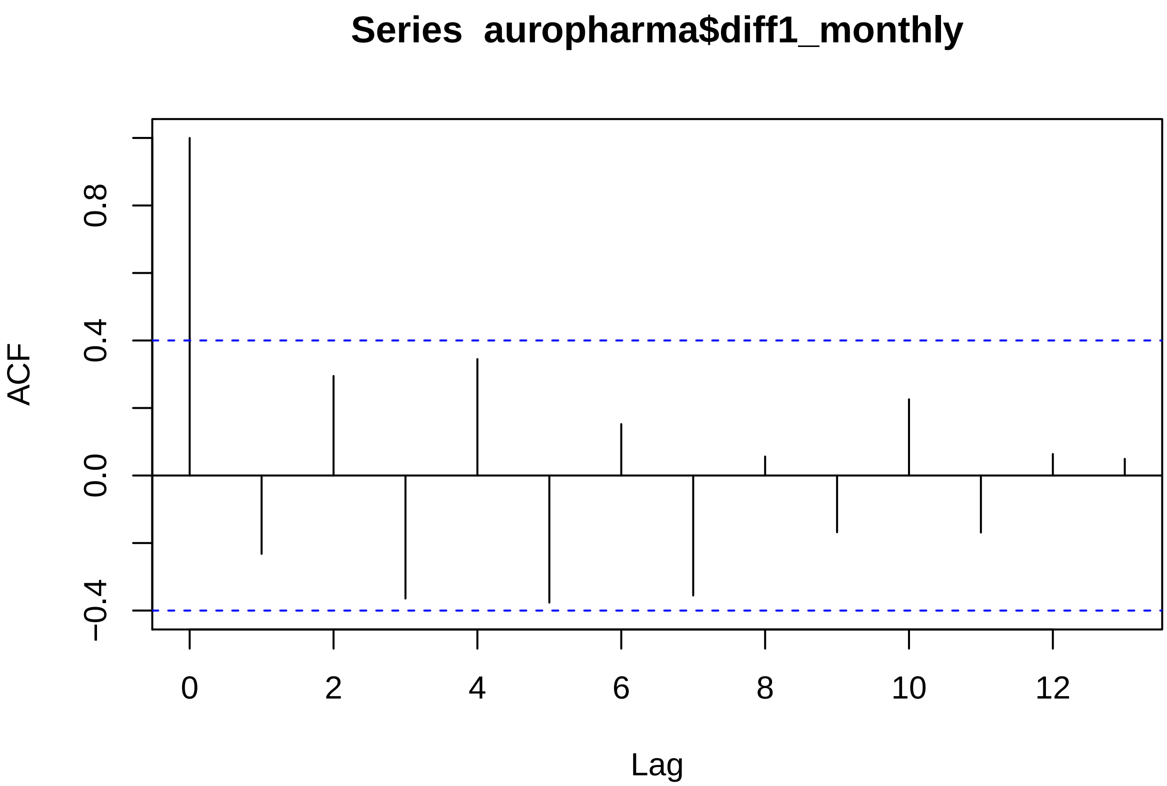



```
# As ACF and PACF are inbetween significant bands, this shows that signal
# is random (i.e there exists no trend to model) and stationary
```

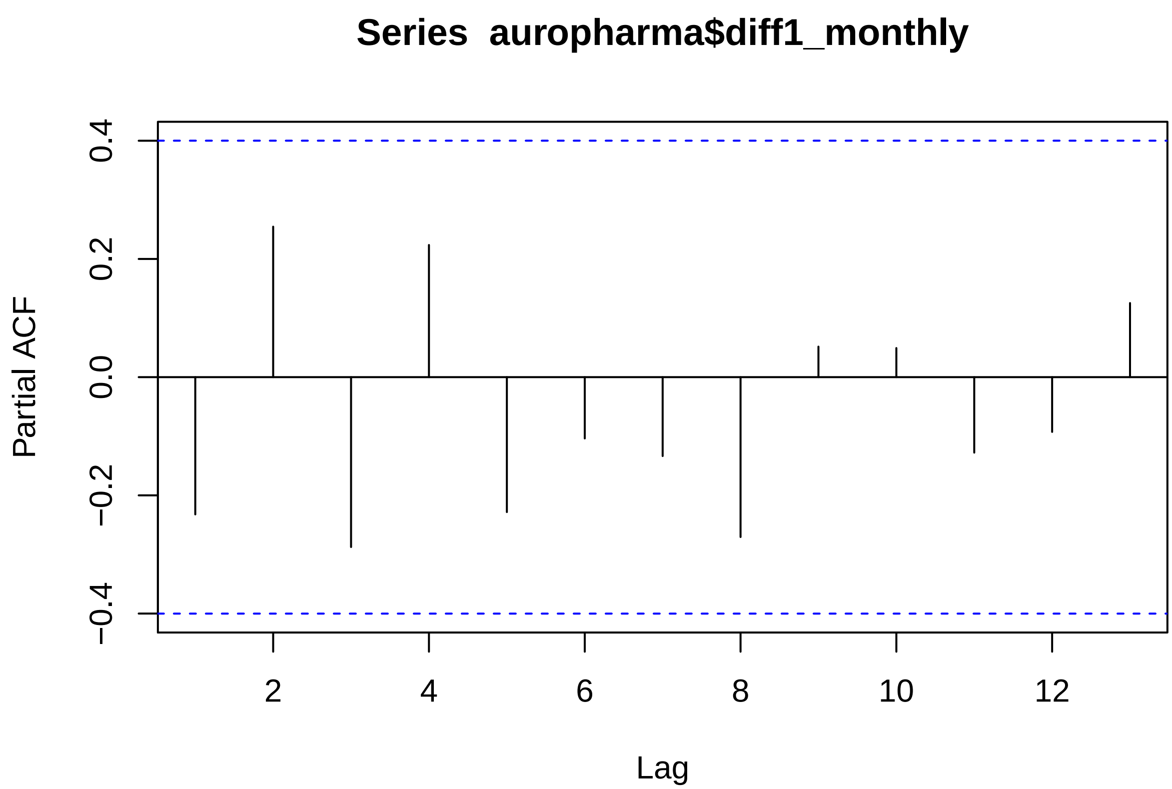
- As it can be seen that acf at any lag is bounded by significant limits, which means that the stationary series obtained after taking first order difference is white noise.

Monthly Series

```r
acf(auropharma$diff1_monthly)
```




```r
pacf(auropharma$diff1_monthly)
```




```r
# As ACF and PACF are inbetween significant bands, this shows that signal
# is random (i.e there exists no trend to model) and stationary
```

- As it can be seen that acf at any lag is bounded by significant limits, which means that the stationary series obtained after taking first order difference is white noise.

GlaxoSmithkline Pharmaceuticals Limited (GLAXO.NS)

```r
glaxo = NULL
glaxo$raw_daily = loadCSVData('../Data/GLAXO/GLAXO.NS_daily.csv')
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##    1044    1216    1272    1308    1366    1795
```

```r
glaxo$raw_weekly = loadCSVData('../Data/GLAXO/GLAXO.NS_weekly.csv')
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##    1044    1220    1271    1311    1369    1749
```

```r
glaxo$raw_monthly = loadCSVData('../Data/GLAXO/GLAXO.NS_monthly.csv')
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##    1044    1230    1273    1313    1372    1714
```

```r
glaxo$sd_daily = standardize(glaxo$raw_daily)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## -1.9620 -0.6849 -0.2666  0.0000  0.4267  3.6130
```

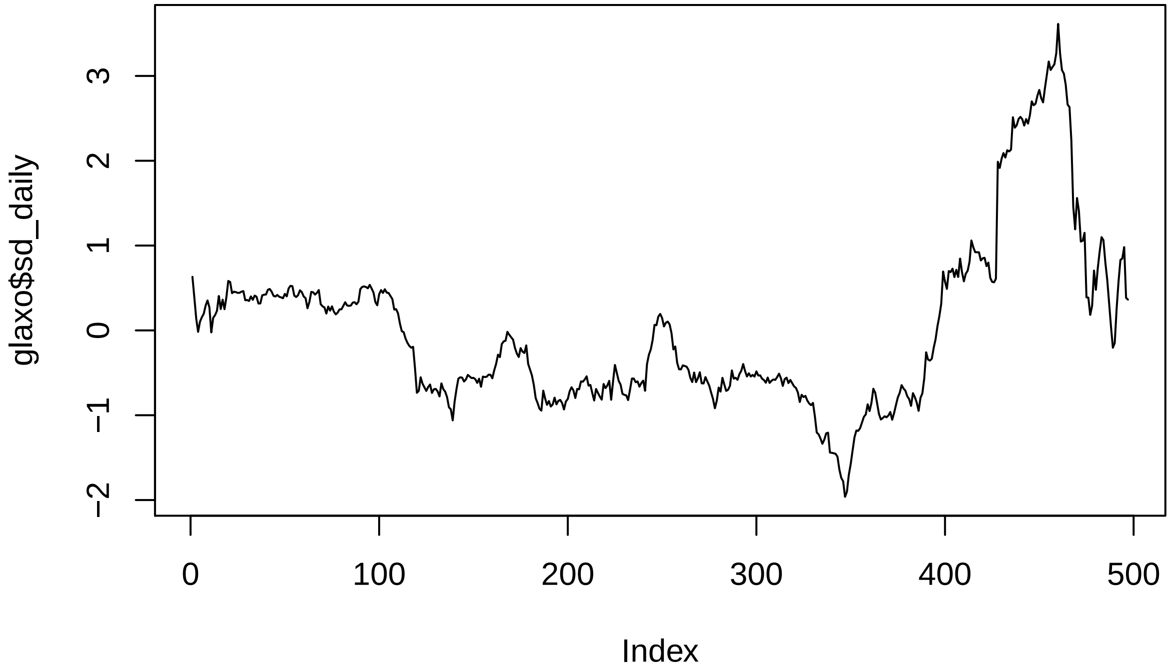
```r
glaxo$sd_weekly = standardize(glaxo$raw_weekly)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## -1.9470 -0.6658 -0.2936  0.0000  0.4184  3.1850
```

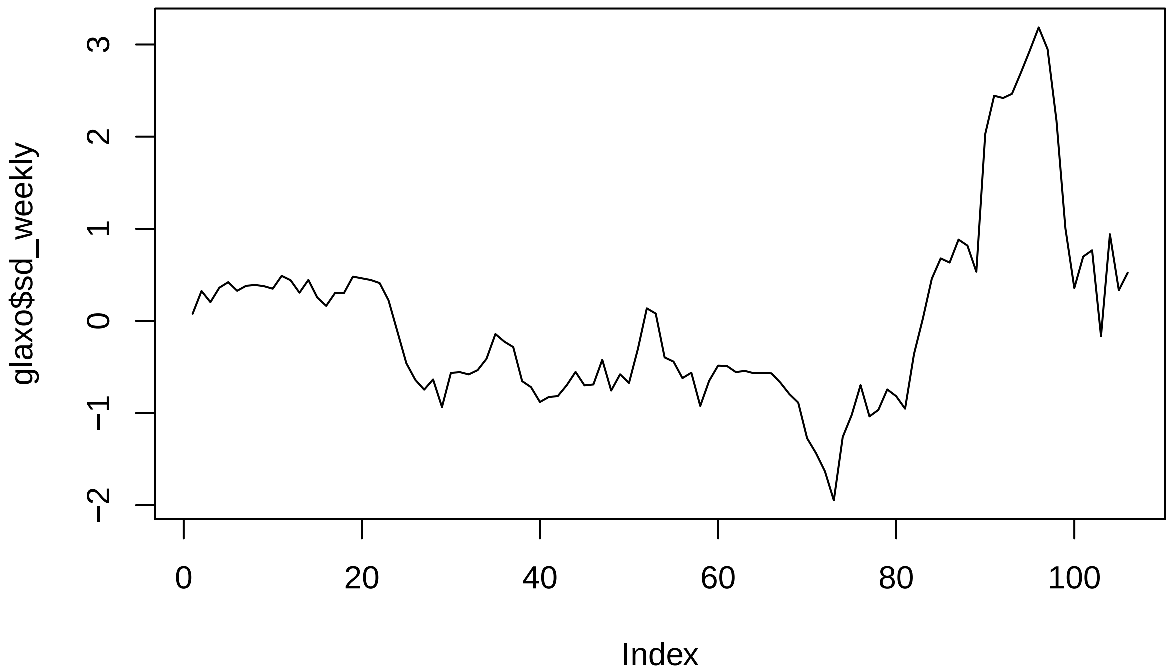
```r
glaxo$sd_monthly = standardize(glaxo$raw_monthly)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## -1.9310 -0.5927 -0.2828  0.0000  0.4269  2.8820
```

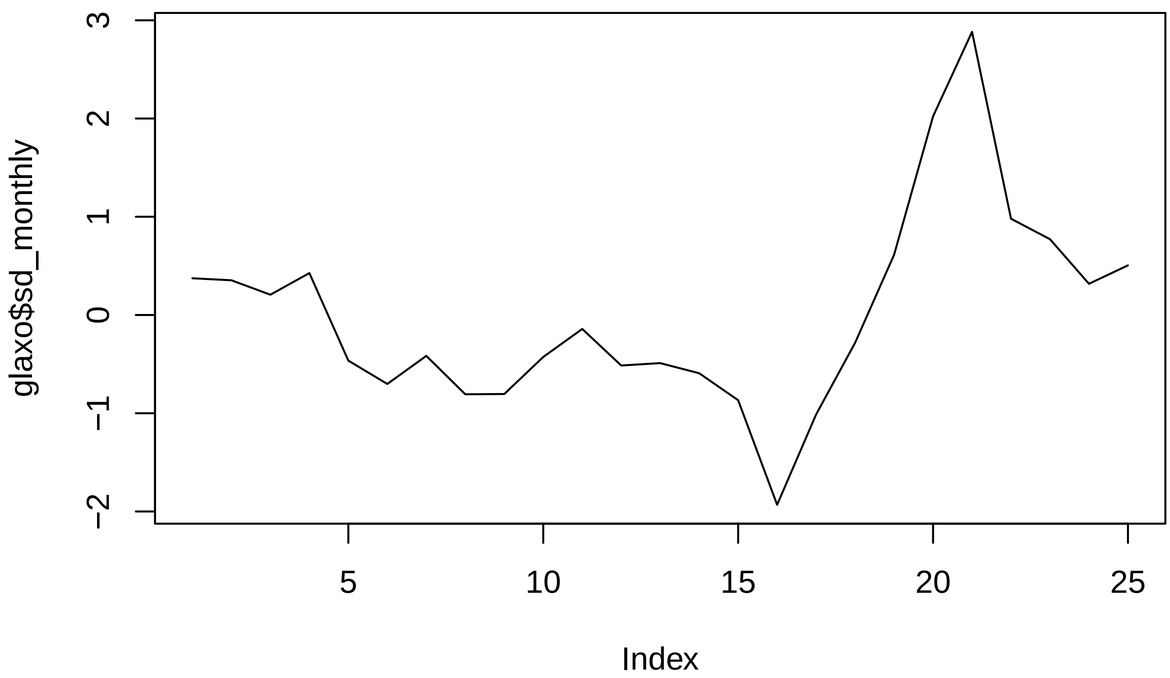
```r
# plot all the data
plot(glaxo$sd_daily, type='l')
```




```r
plot(glaxo$sd_weekly, type='l')
```

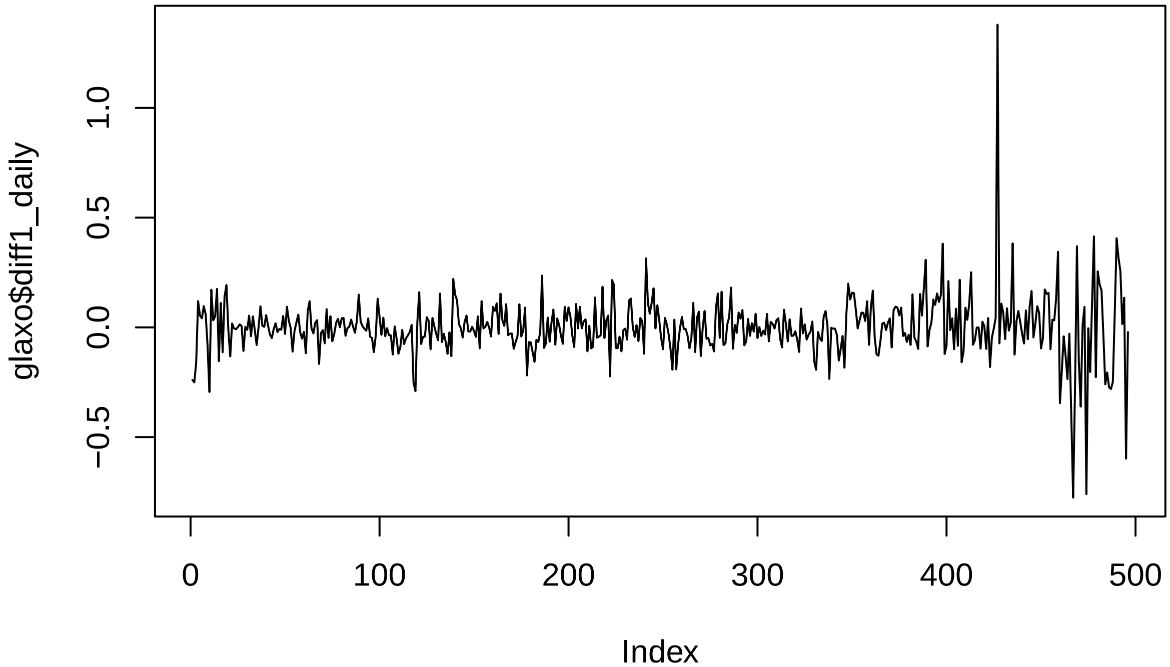



```
plot(glaxo$sd_monthly, type='l')
```

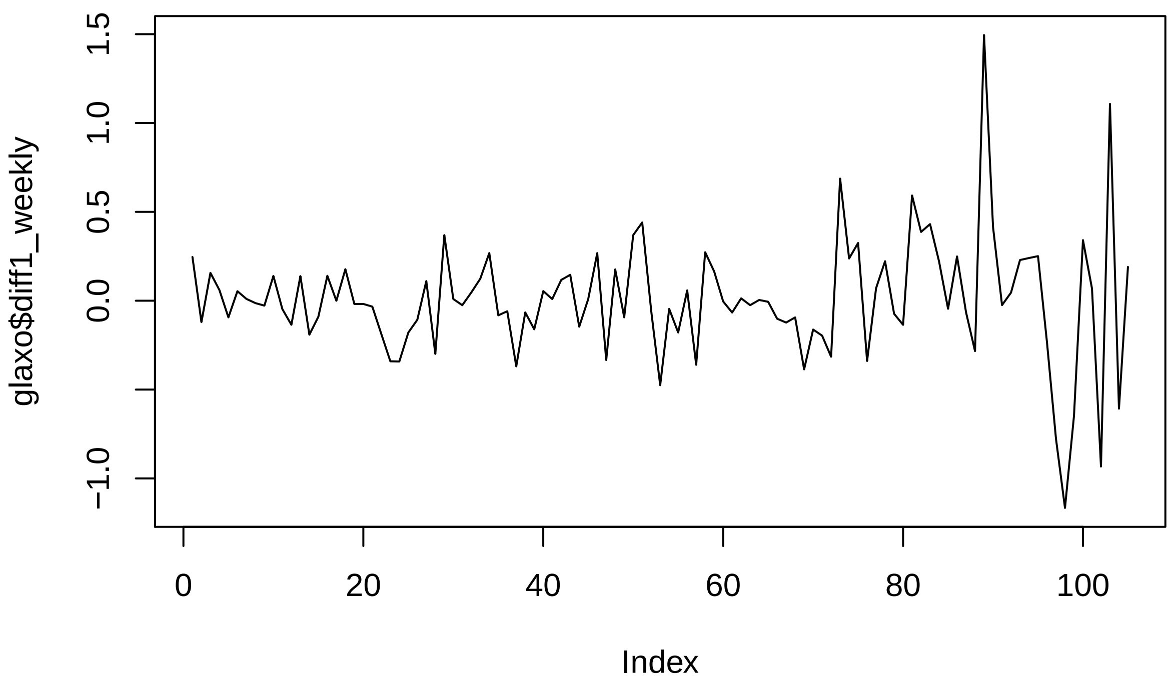



```
# Based on plots we can clearly see that given data is
# non stationary (data has some integrating effect, which should estimated)
# which can be converted to stationary by taking difference of order = n

glaxo$diff1_daily = diff(glaxo$sd_daily)
glaxo$diff1_weekly = diff(glaxo$sd_weekly)
glaxo$diff1_monthly = diff(glaxo$sd_monthly)

# Plotting differenced data
```

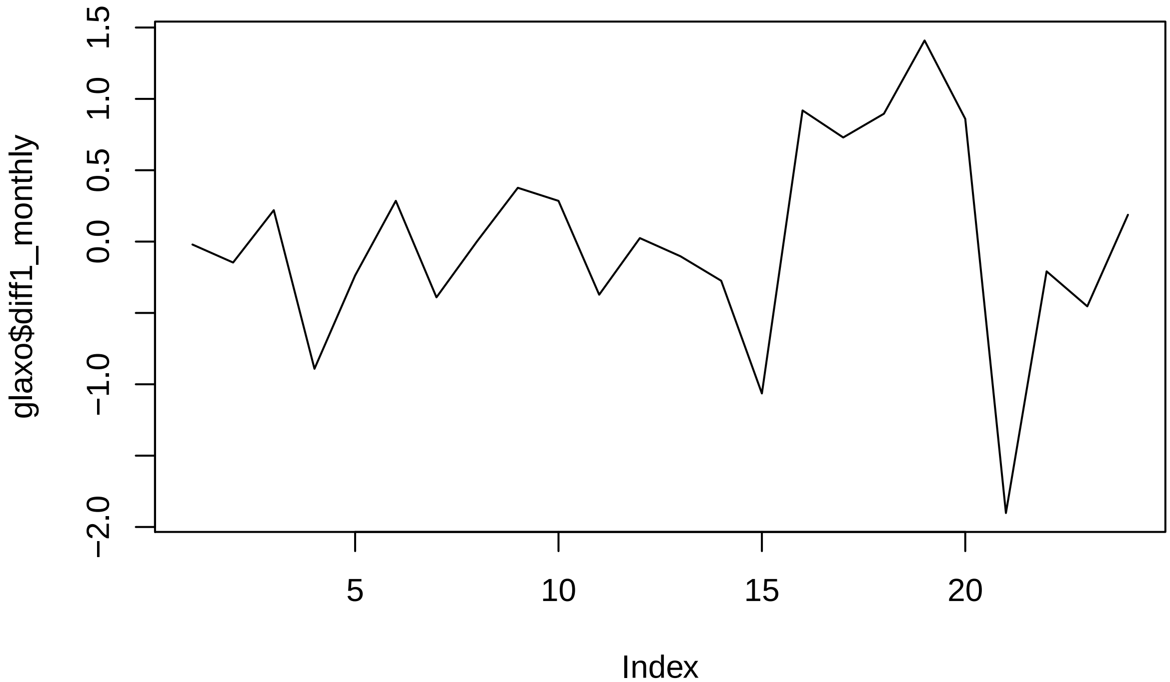
```r
plot(glaxo$diff1_daily, type='l')
```




```r
plot(glaxo$diff1_weekly, type='l')
```




```r
plot(glaxo$diff1_monthly, type='l')
```




```r
# Differenced plots seems to be stationary this can also be verified by summary of the data
print(summary(glaxo$diff1_daily))
```

```
##       Min.    1st Qu.     Median       Mean    3rd Qu.       Max.
## -0.7756000 -0.0572600 -0.0045650 -0.0005432  0.0537000  1.3790000
```

```r
print(summary(glaxo$diff1_weekly))
```

```
##      Min.   1st Qu.    Median      Mean   3rd Qu.      Max.
## -1.166000 -0.135500 -0.013330  0.004248  0.164700  1.495000
```

```r
print(summary(glaxo$diff1_monthly))
```

```
##      Min.   1st Qu.    Median      Mean   3rd Qu.      Max.
## -1.902000 -0.298900 -0.008869  0.005476  0.308500  1.409000
```

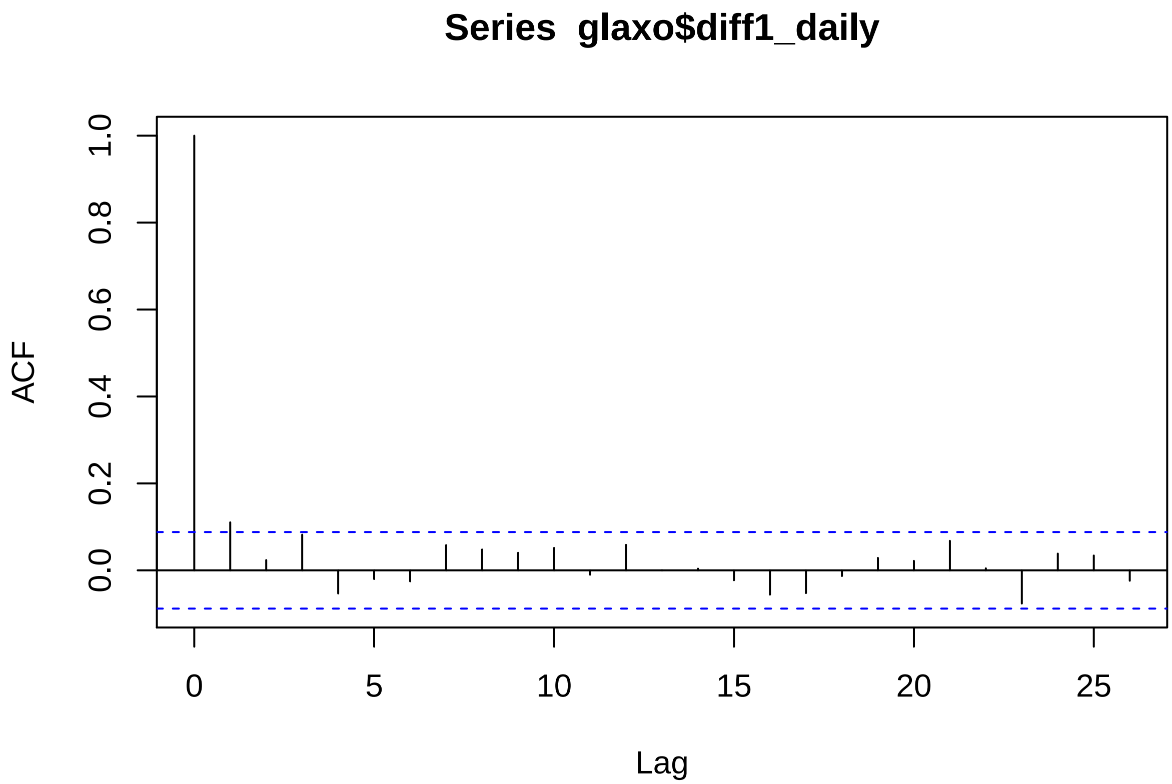
- Actual Raw data is non stationary
- By analysing the data it's clear that some integrating effect exists, which can be removed by taking first order difference of the signal
- Stationary component of the data is estimated using first order difference of the time series
- Resulting plot after removing integrating effect are stationary

Trend Estimation

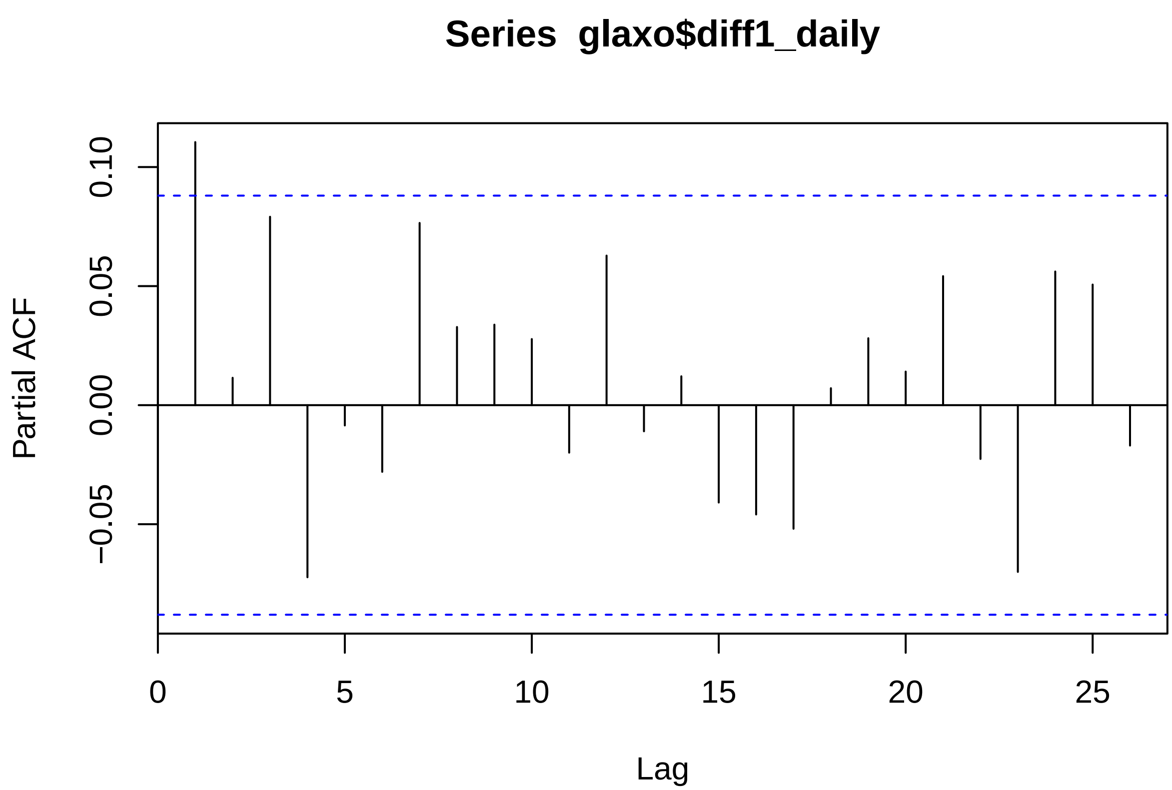
In this section we'll try to seperate predictable part of the series from random white noise. This is acheved using ACF PACF analysis.

Daily Series

```r
acf(glaxo$diff1_daily)
```



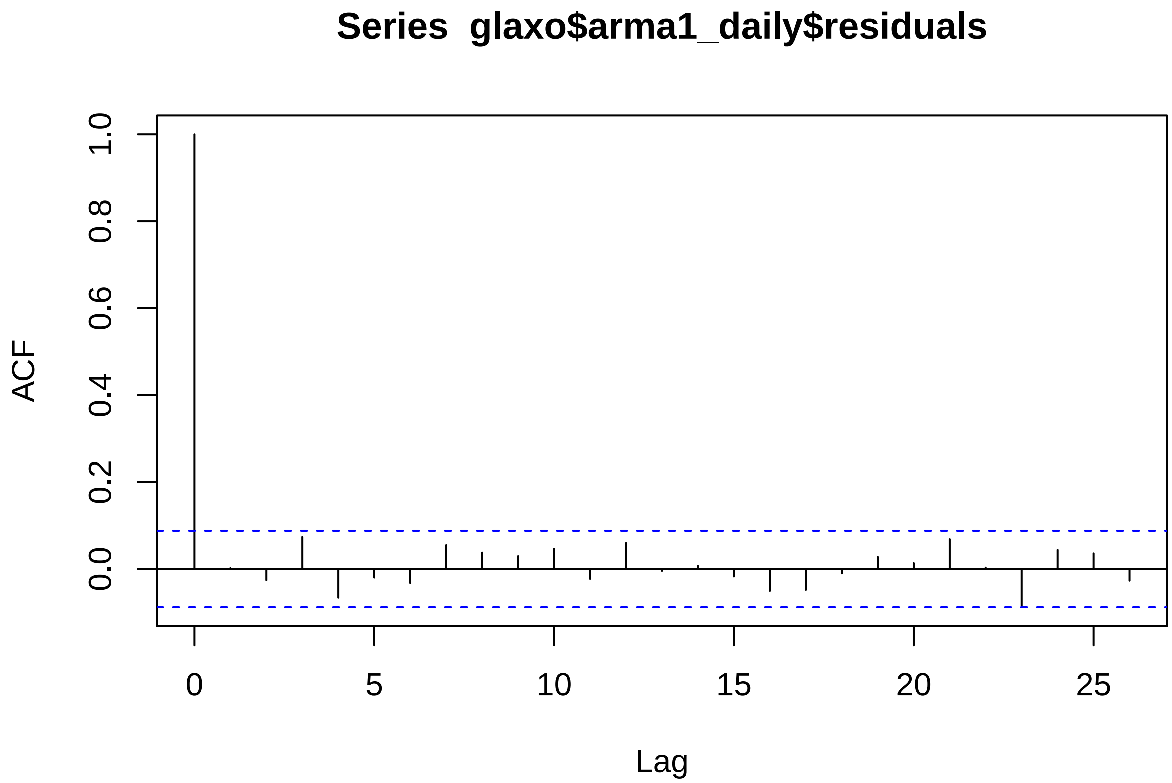

```r
pacf(glaxo$diff1_daily)
```



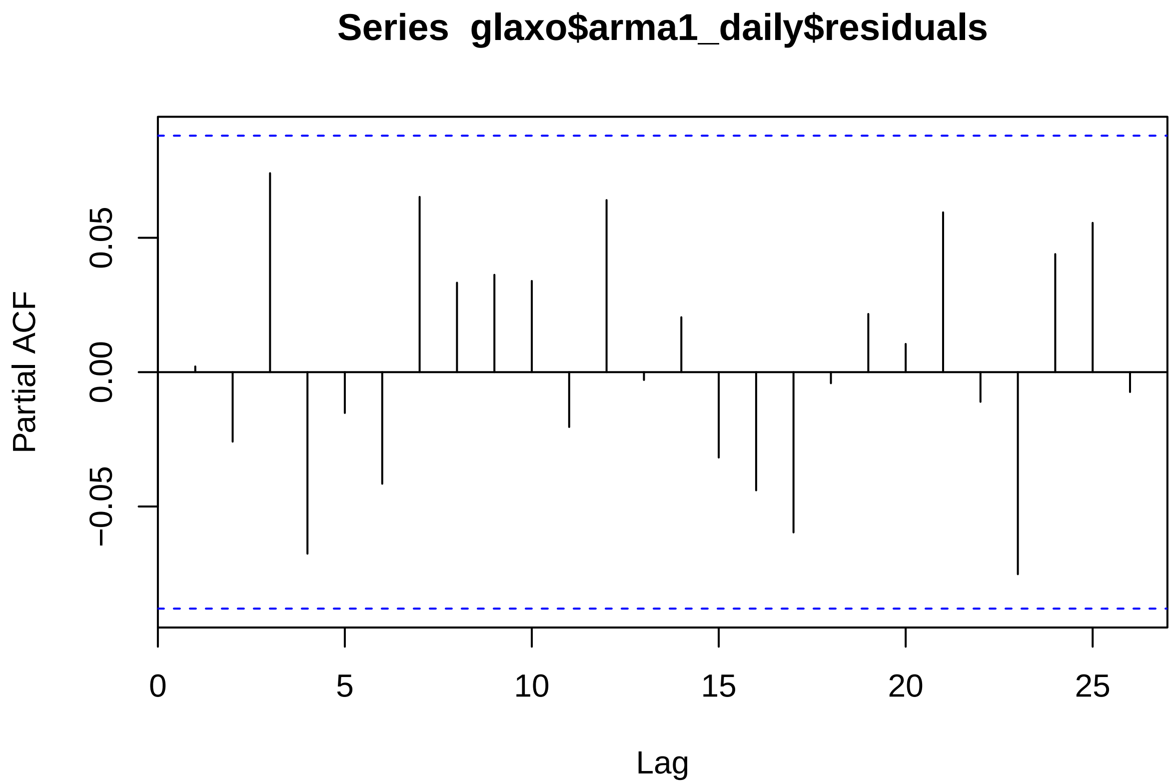

```r
# As ACF and PACF are inbetween significant bands, this shows that signal
# is random (i.e there exists no trend to model) and stationary
```

- As it can be seen that acf and pacf at lag = 1 exceeds significant bounds, which means Autoregressive model ARMA(1, 1) can fit the data

```
glaxo$arma1_daily = arima(glaxo$diff1_daily, order=c(1,0,1))

# ACF of residuals should be white if model captures entire information
acf(glaxo$arma1_daily$residuals)
```



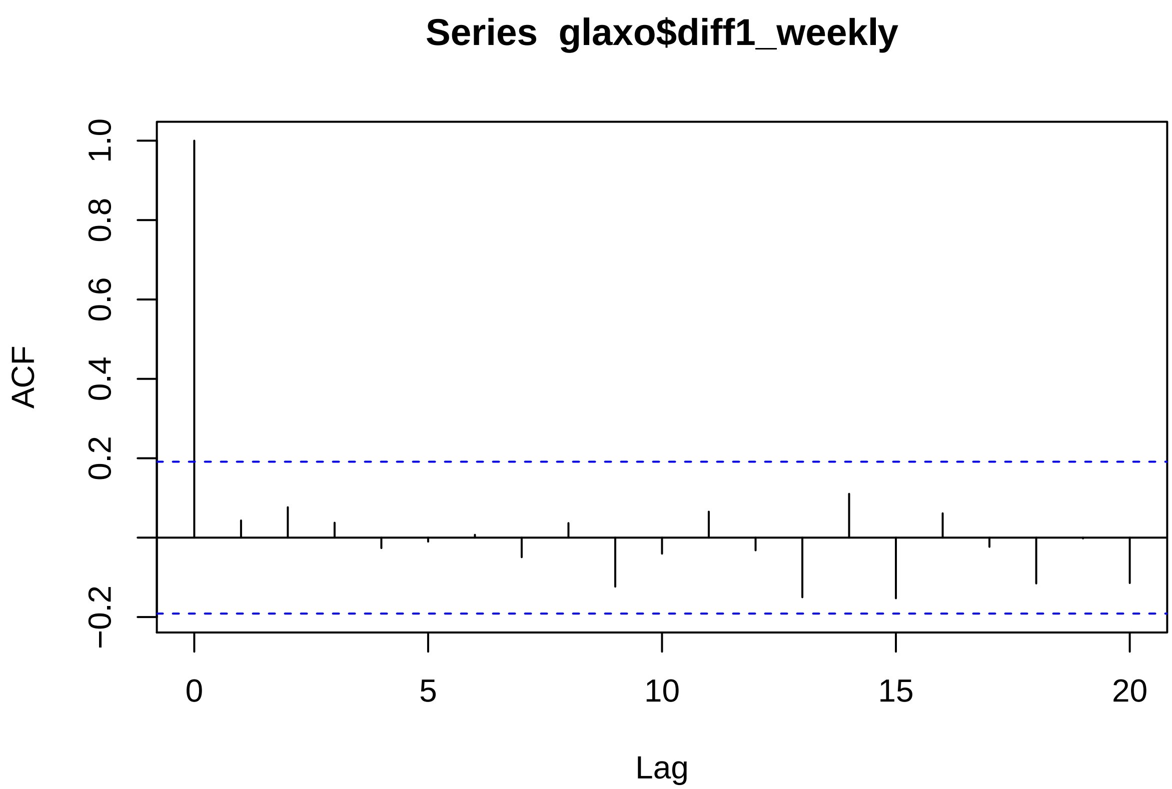

```
pacf(glaxo$arma1_daily$residuals)
```



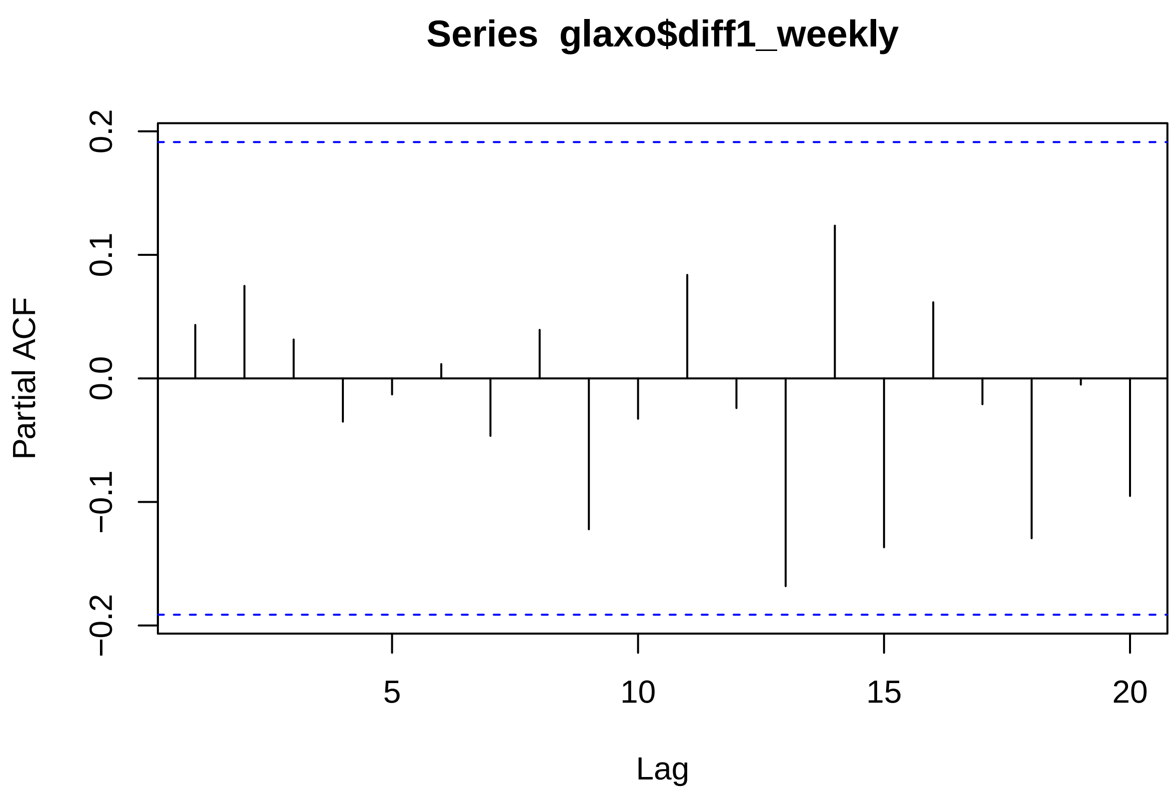
- By fitting ARMA(1,1) model predictable component of the data is exploited, residuals obtained forms white noise
- ARMA trend is followed in case of Daily data, Data can be made stationary after considering first order difference

Weekly Series

```
acf(glaxo$diff1_weekly)
```




```r
pacf(glaxo$diff1_weekly)
```

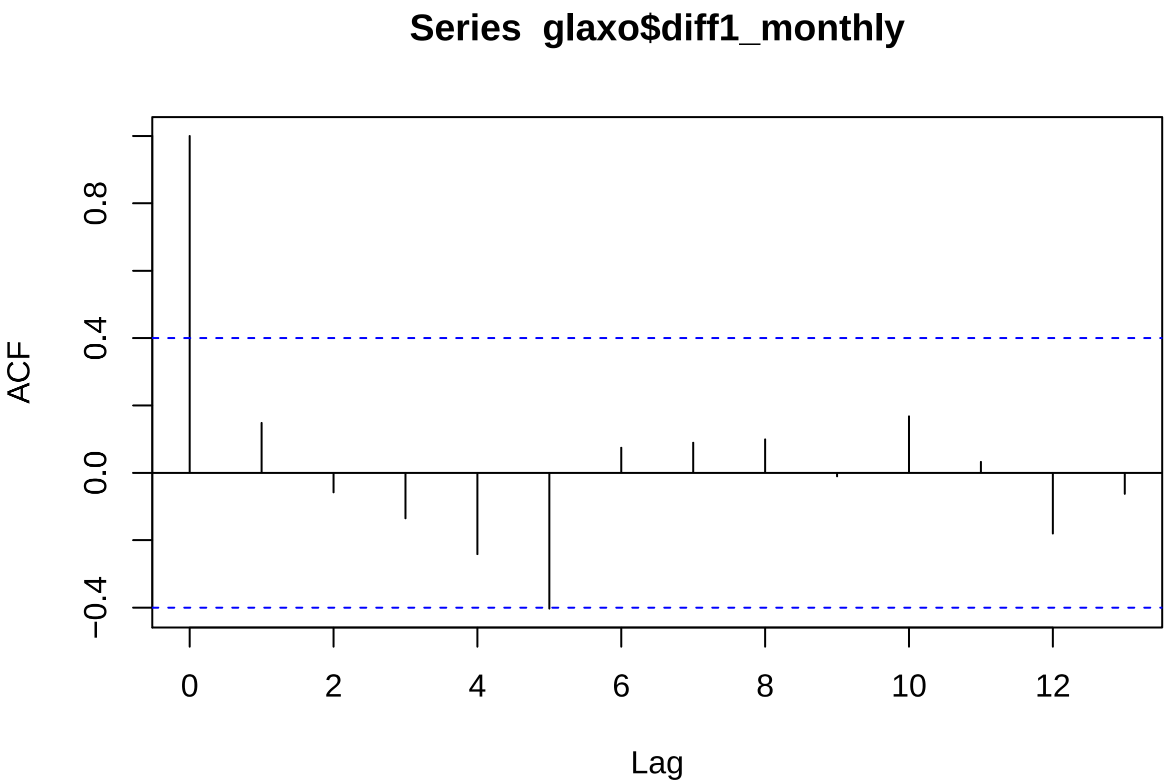



```r
# As ACF and PACF are inbetween significant bands, this shows that signal
# is random (i.e there exists no trend to model) and stationary
```

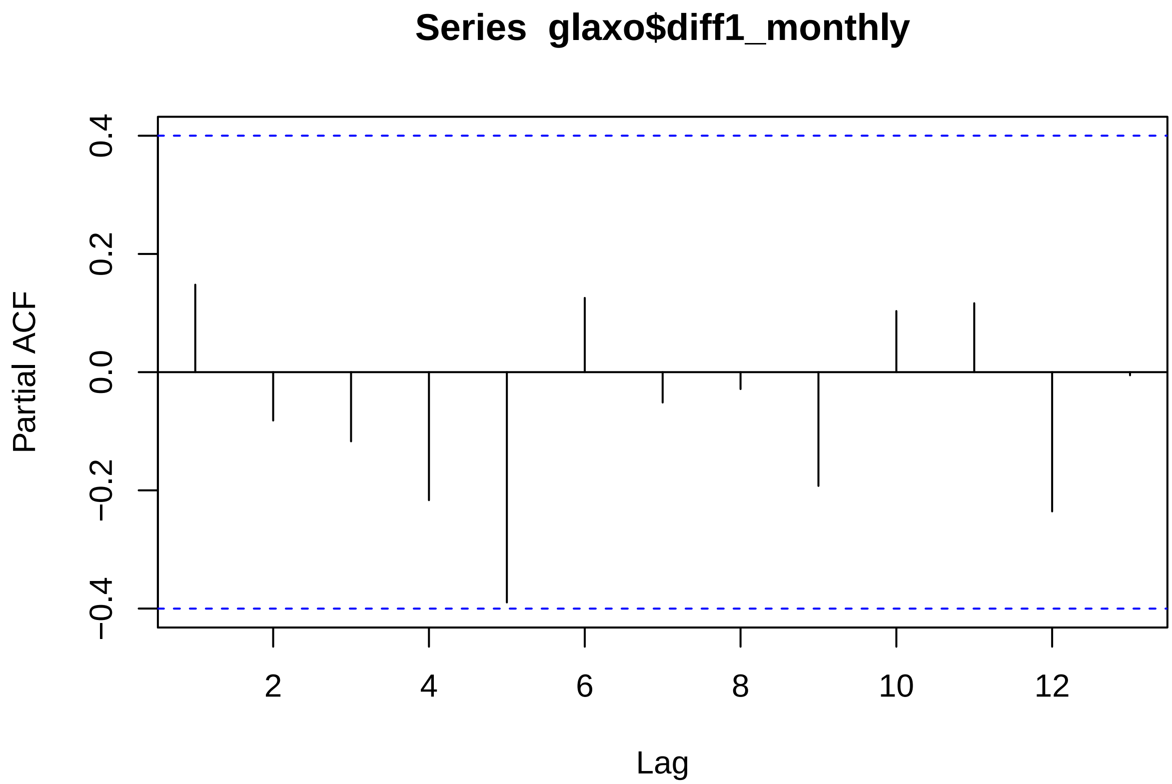
- As it can be seen that acf at any lag is bounded by significant limits, which means that the stationary series obtained after taking first order difference is white noise.

Monthly Series

```
acf(glaxo$diff1_monthly)
```




```
pacf(glaxo$diff1_monthly)
```




```
# As ACF and PACF are inbetween significant bands, this shows that signal
# is random (i.e there exists no trend to model) and stationary
```

- As it can be seen that acf at any lag is bounded by significant limits, which means that the stationary series obtained after taking first order difference is white noise.

Glenmark Pharmaceuticals Limited (GLENMARK.NS)

```
glenmark = NULL
glenmark$raw_daily = loadCSVData('../Data/GLENMARK/GLENMARK.NS_daily.csv')
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   500.6   581.3   616.3   675.8   853.8   959.3
```

```
glenmark$raw_weekly = loadCSVData('../Data/GLENMARK/GLENMARK.NS_weekly.csv')
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   500.6   582.7   618.6   674.5   749.0   953.2
```

```
glenmark$raw_monthly = loadCSVData('../Data/GLENMARK/GLENMARK.NS_monthly.csv')
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   526.3   583.0   621.5   663.9   665.7   925.5
```

```
glenmark$sd_daily = standardize(glenmark$raw_daily)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## -1.2750 -0.6876 -0.4329  0.0000  1.2950  2.0630
```

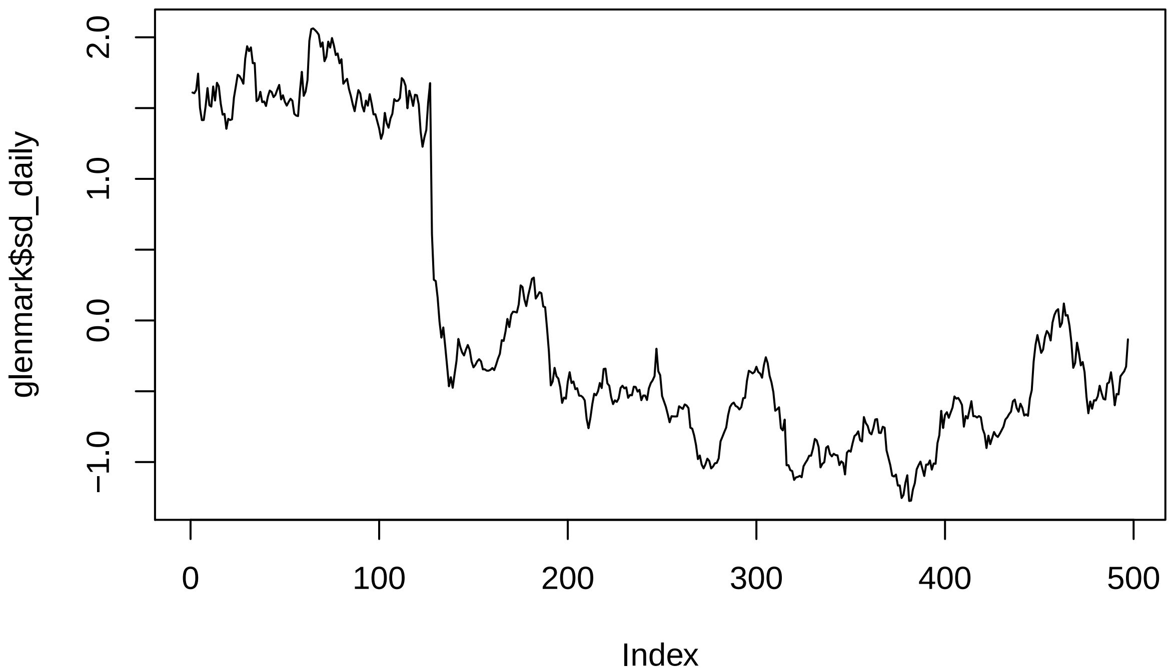
```r
glenmark$sd_weekly = standardize(glenmark$raw_weekly)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## -1.2750 -0.6731 -0.4098  0.0000  0.5461  2.0430
```

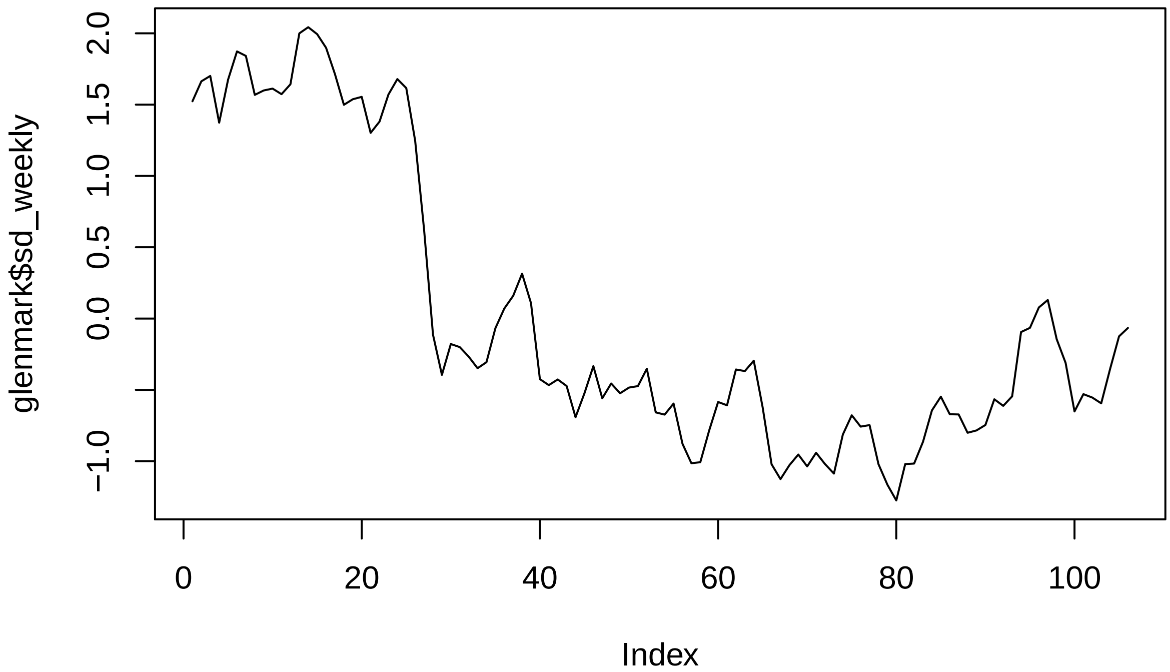
```r
glenmark$sd_monthly = standardize(glenmark$raw_monthly)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## -1.1160 -0.6554 -0.3437  0.0000  0.0143  2.1210
```

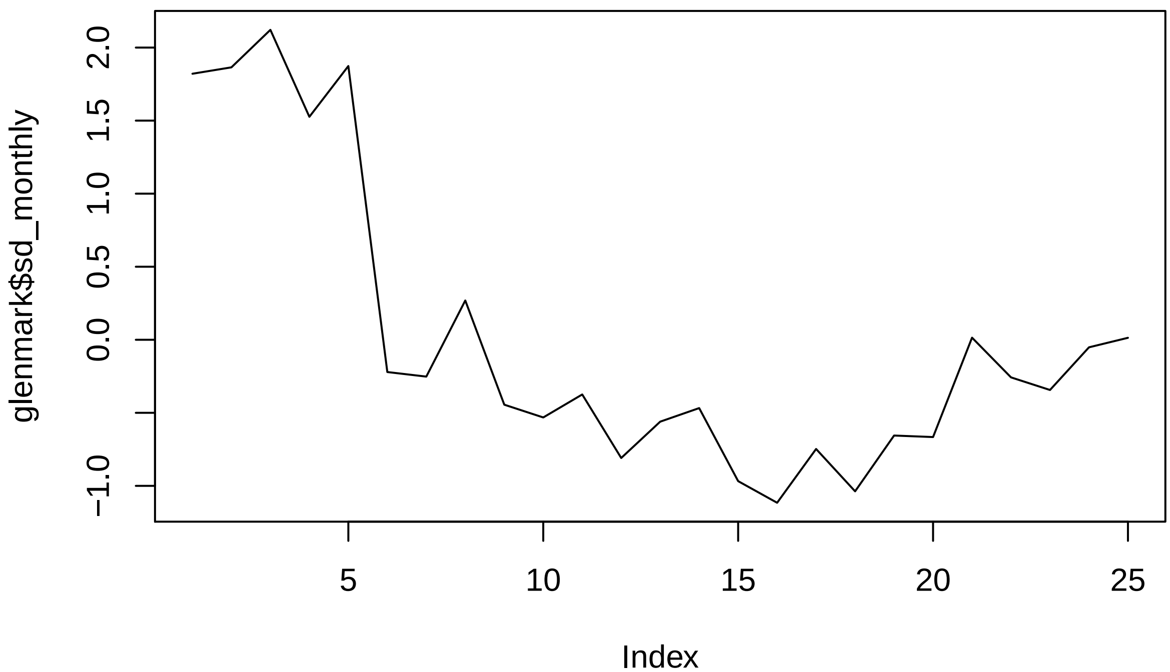
```r
# plot all the data
plot(glenmark$sd_daily, type='l')
```




```r
plot(glenmark$sd_weekly, type='l')
```

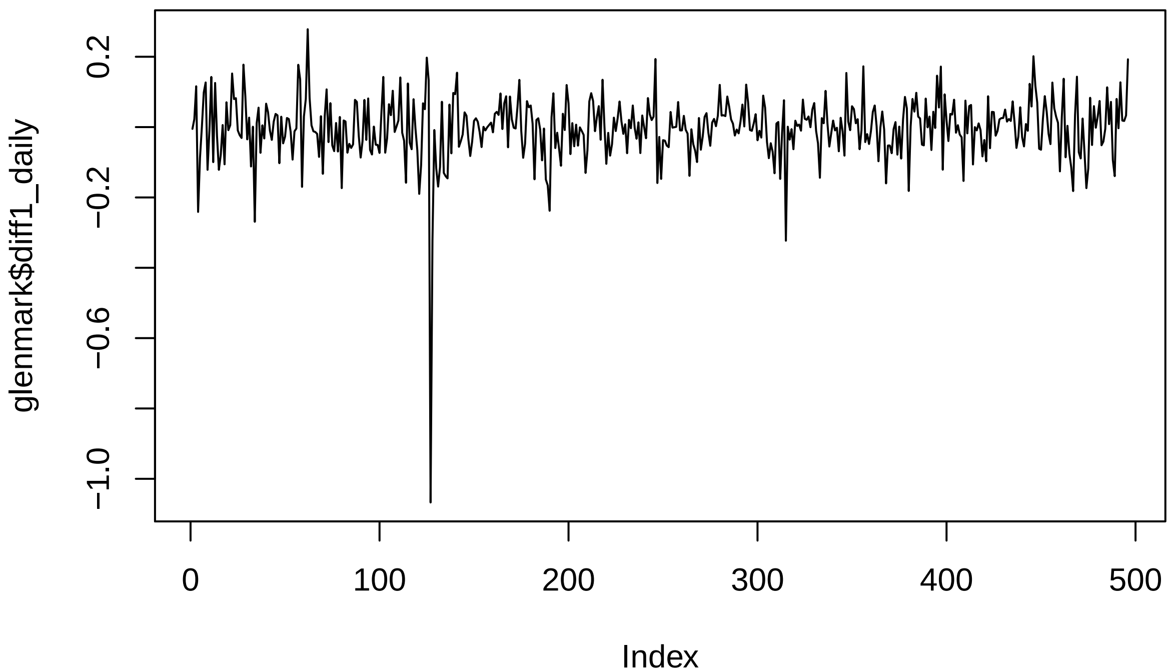



```
plot(glenmark$sd_monthly, type='l')
```

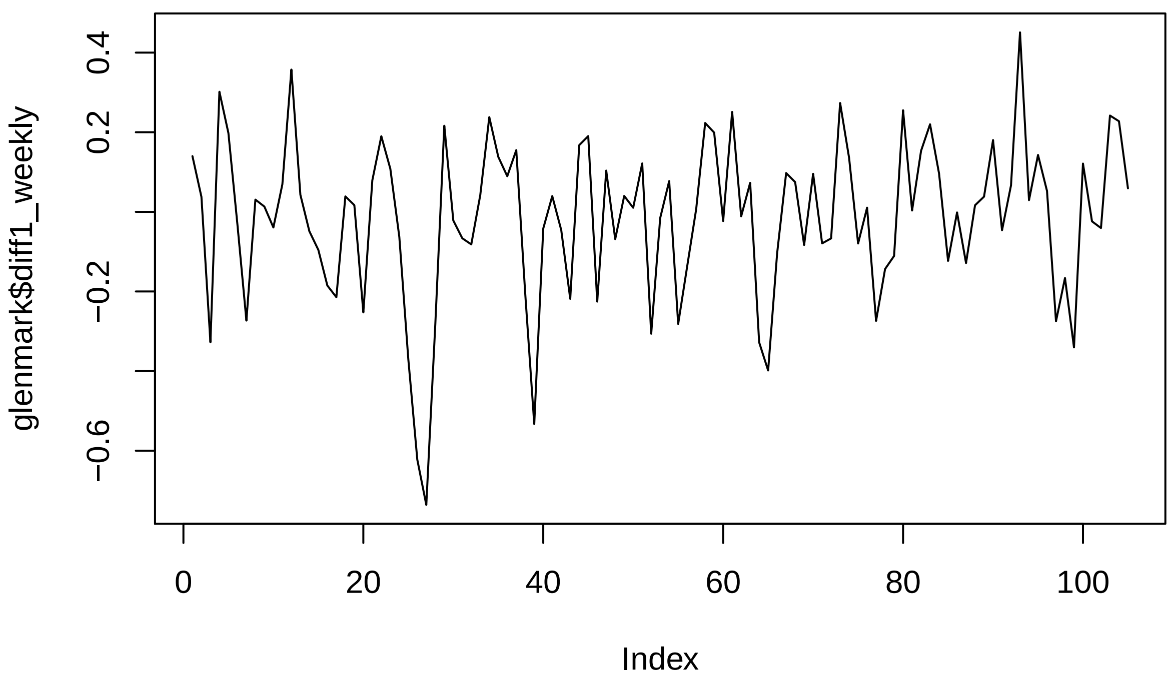



```
# Based on plots we can clearly see that given data is
# non stationary (data has some integrating effect, which should estimated)
# which can be converted to stationary by taking difference of order = n

glenmark$diff1_daily = diff(glenmark$sd_daily)
glenmark$diff1_weekly = diff(glenmark$sd_weekly)
glenmark$diff1_monthly = diff(glenmark$sd_monthly)

# Plotting differenced data
```

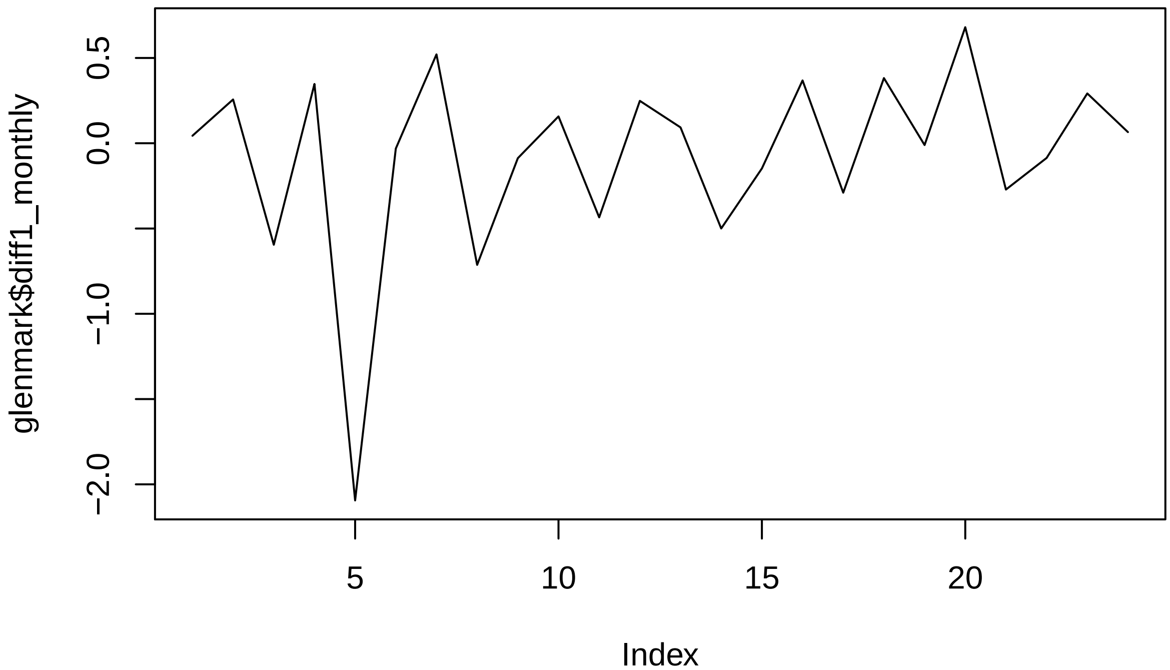
```r
plot(glenmark$diff1_daily, type='l')
```




```r
plot(glenmark$diff1_weekly, type='l')
```




```r
plot(glenmark$diff1_monthly, type='l')
```




```r
# Differenced plots seems to be stationary this can also be verified by summary of the data
print(summary(glenmark$diff1_daily))
```

```
##       Min.    1st Qu.     Median       Mean    3rd Qu.       Max.
## -1.0670000 -0.0477400  0.0007278 -0.0035140  0.0420100  0.2783000
```

```r
print(summary(glenmark$diff1_weekly))
```

```
##     Min.  1st Qu.   Median     Mean  3rd Qu.     Max.
## -0.73610 -0.10410  0.01063 -0.01514  0.10810  0.45090
```

```r
print(summary(glenmark$diff1_monthly))
```

```
##     Min.  1st Qu.   Median     Mean  3rd Qu.     Max.
## -2.09400 -0.27620  0.01662 -0.07531  0.26540  0.68030
```

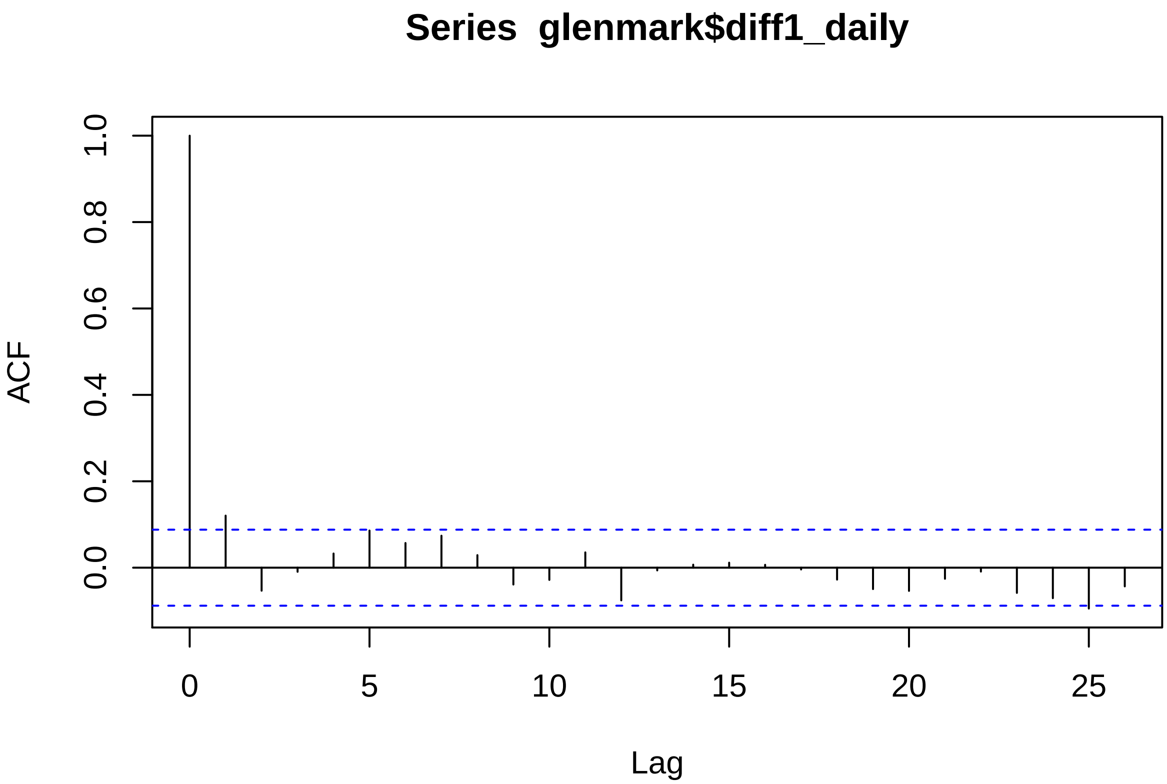
- Actual Raw data is non stationary
- By analysing the data it's clear that some integrating effect exists, which can be removed by taking first order difference of the signal
- Stationary component of the data is estimated using first order difference of the time series
- Resulting plot after removing integrating effect are stationary

Trend Estimation

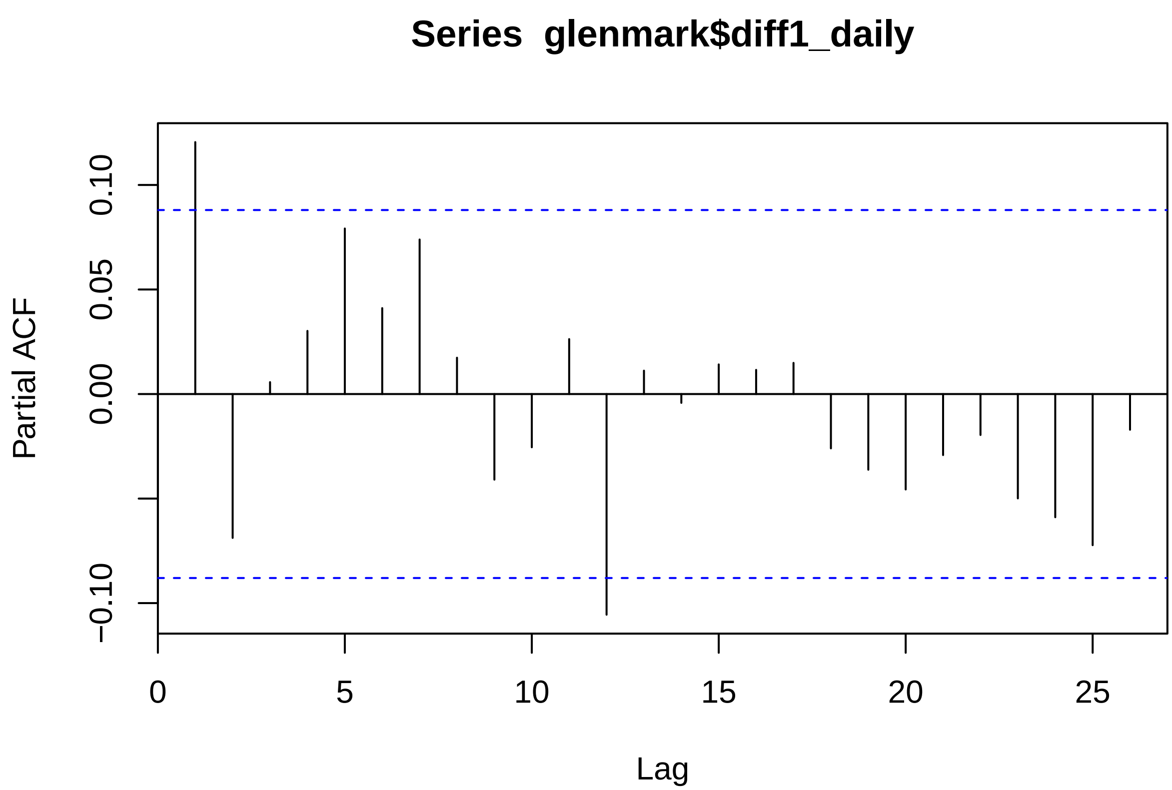
In this section we'll try to seperate predictable part of the series from random white noise. This is acheved using ACF PACF analysis.

Daily Series

```r
acf(glenmark$diff1_daily)
```



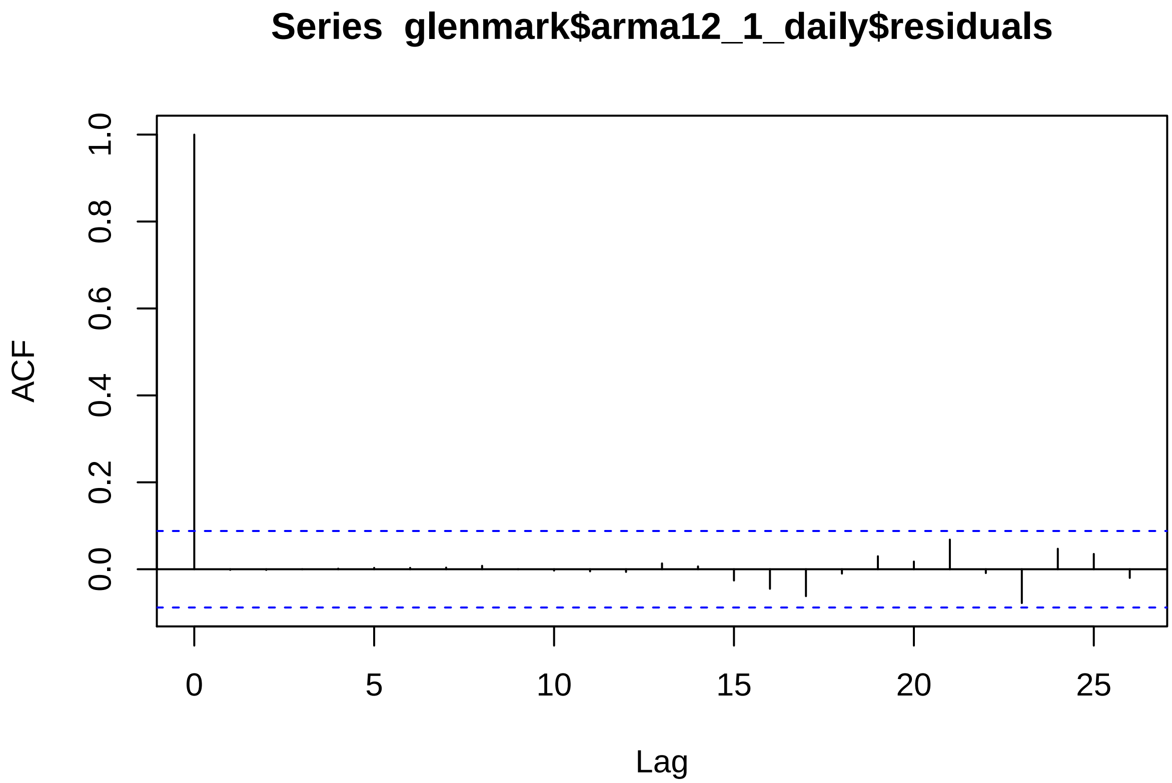

```r
pacf(glenmark$diff1_daily)
```



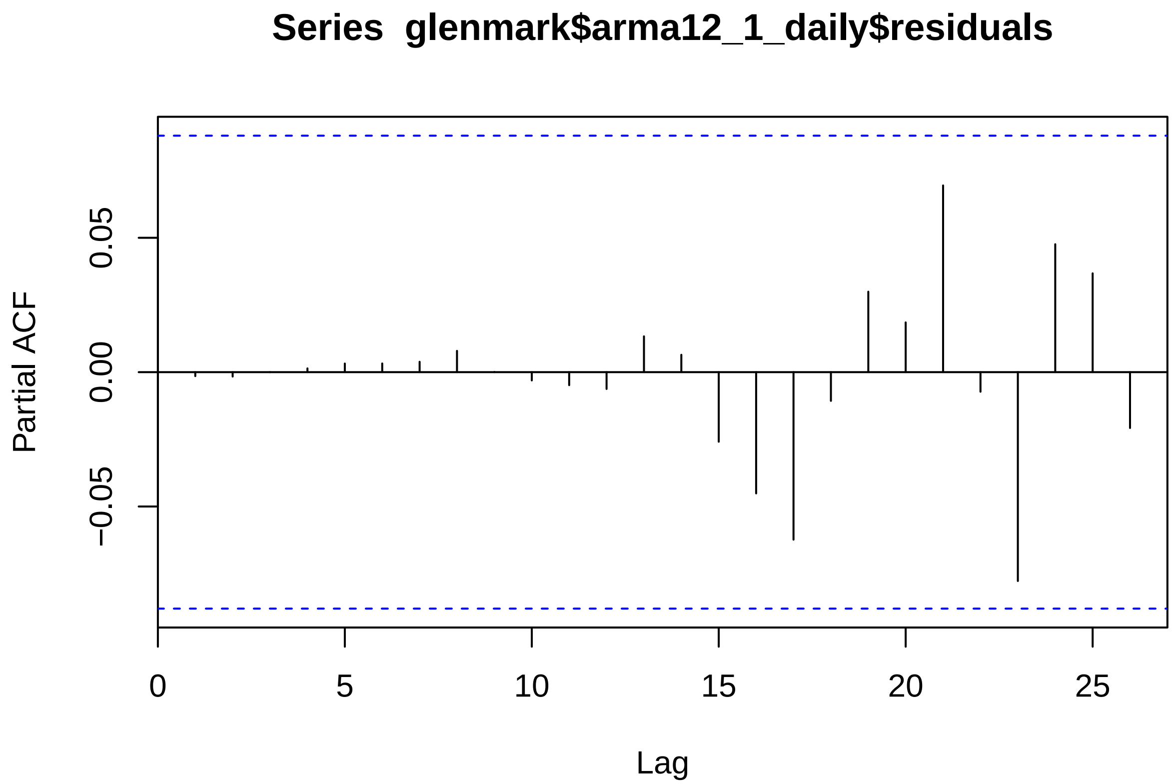

```r
# As ACF and PACF are inbetween significant bands, this shows that signal
# is random (i.e there exists no trend to model) and stationary
```

- As it can be seen that acf and pacf at lag = 1 and 12 respectively exceeds significant bounds, which means Autoregressive model ARMA(12, 1) can fit the data

```
glenmark$arma12_1_daily = arima(glaxo$diff1_daily, order=c(12,0,1))

# ACF of residuals should be white if model captures entire information
acf(glenmark$arma12_1_daily$residuals)
```



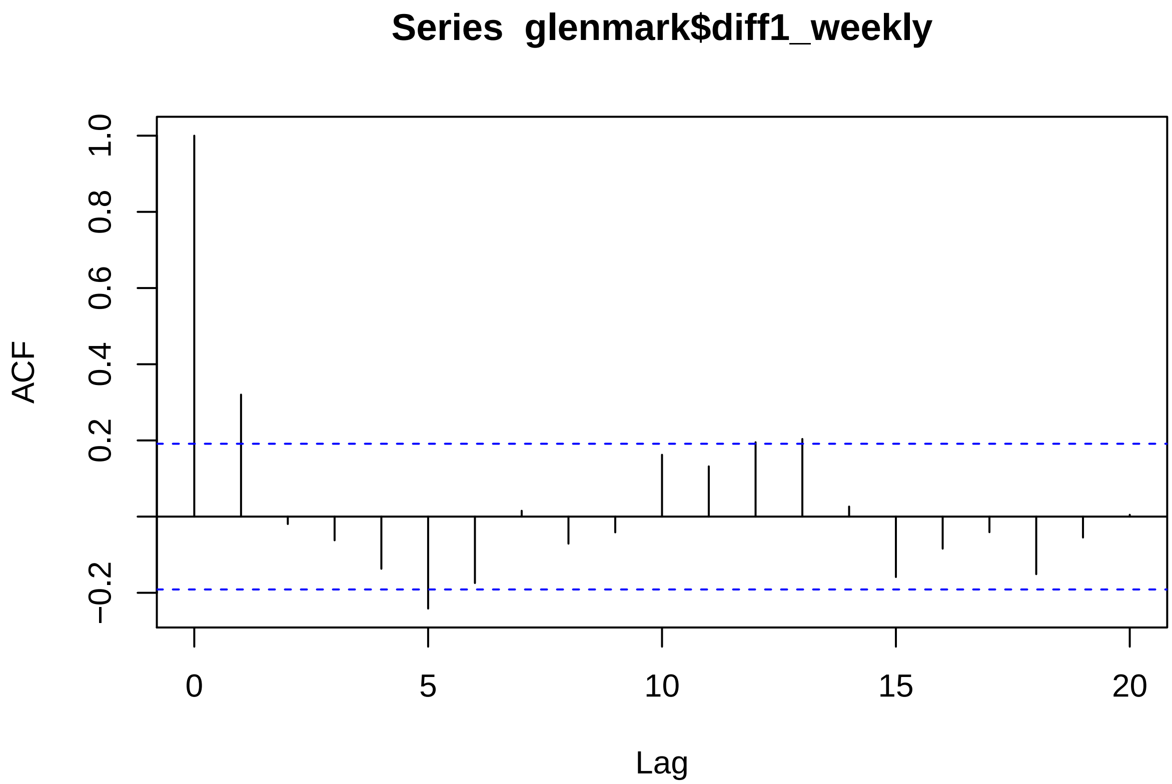

```
pacf(glenmark$arma12_1_daily$residuals)
```



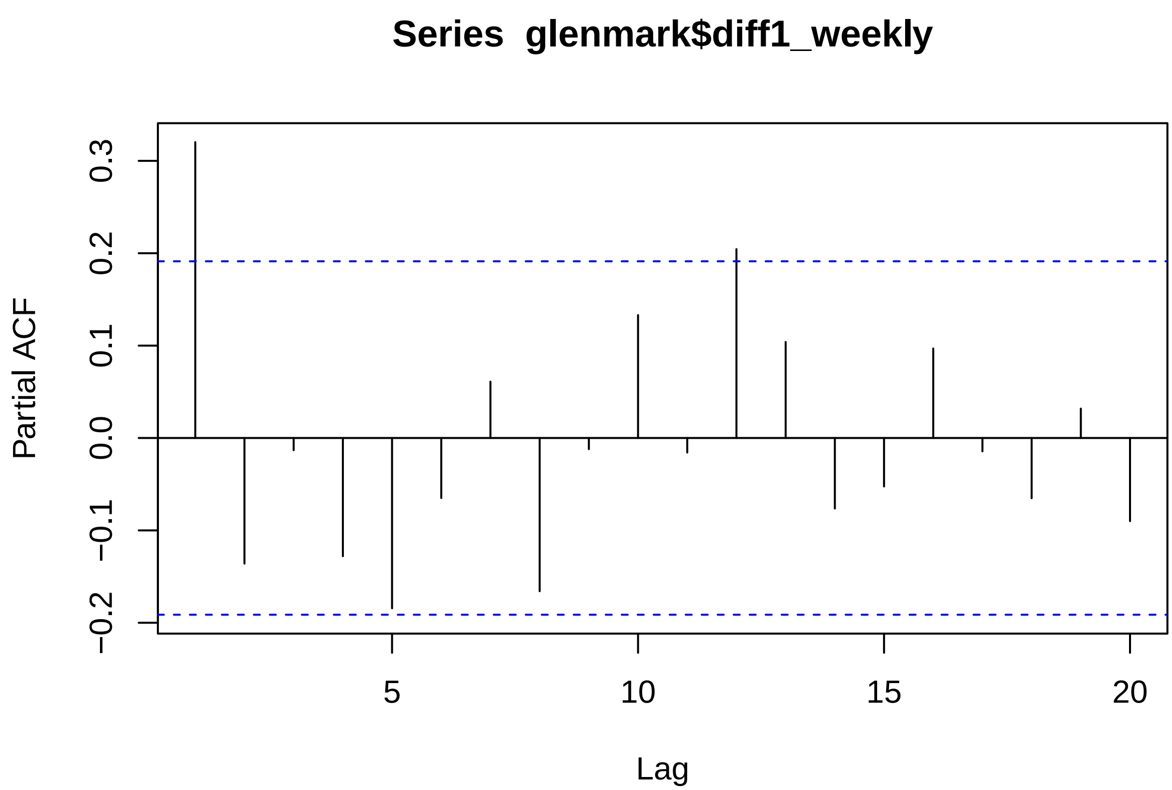
- By fitting ARMA(12,1) model predictable component of the data is exploited, residuals obtained forms white noise
- ARMA trend is followed in case of Daily data, Data can be made stationary after considering first order difference

Weekly Series

```
acf(glenmark$diff1_weekly)
```



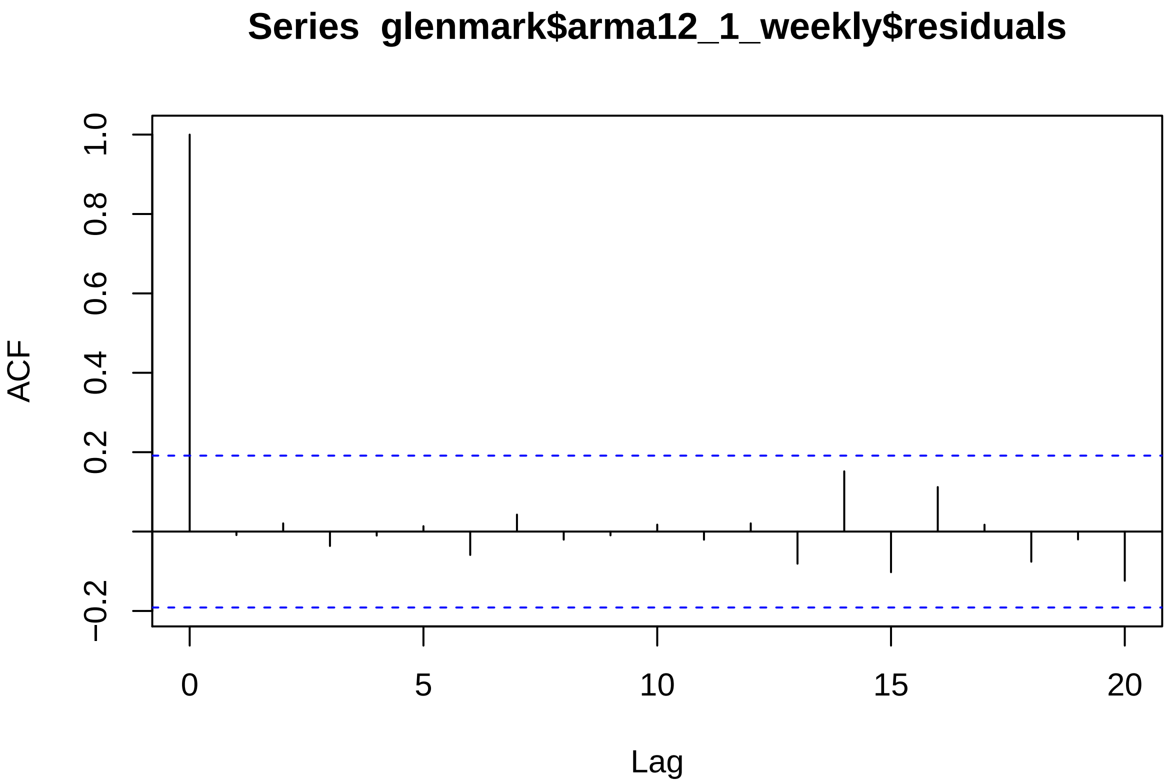

```r
pacf(glenmark$diff1_weekly)
```



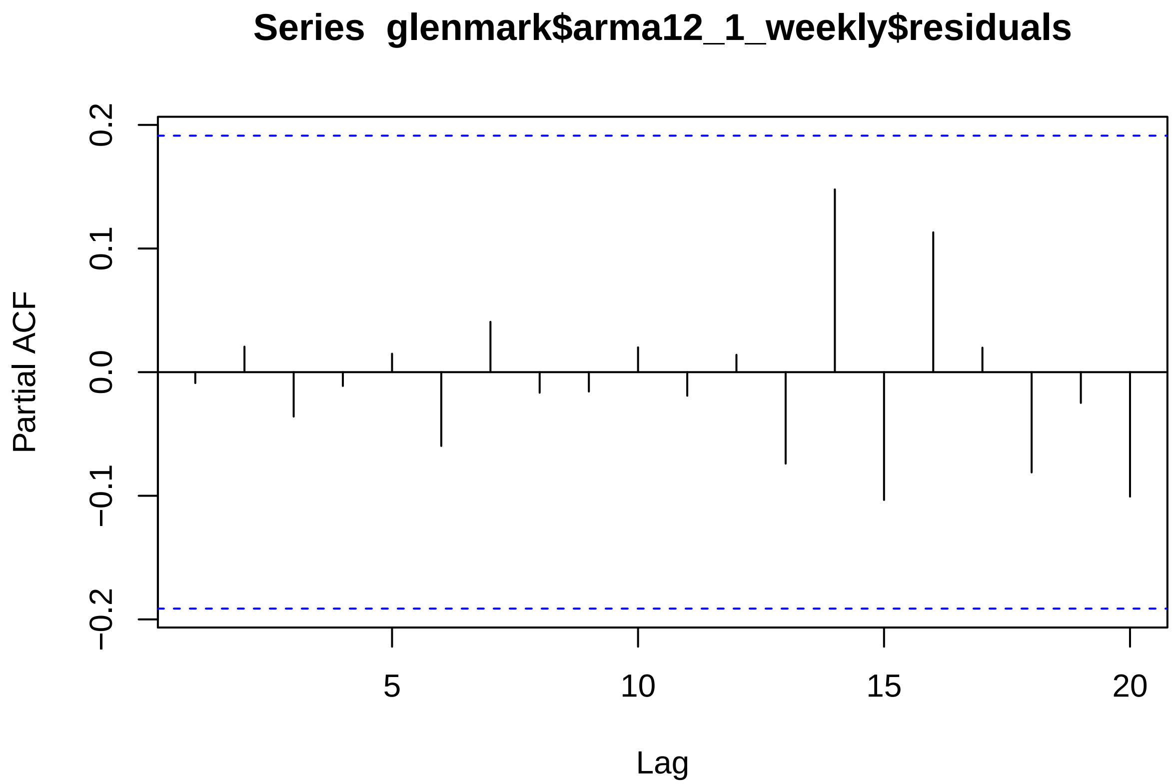

```r
# As ACF and PACF are inbetween significant bands, this shows that signal
# is random (i.e there exists no trend to model) and stationary
```

- As it can be seen that acf and pacf at lag = 1 and 12 respectively exceeds significant bounds, which means Autoregressive model ARMA(12, 1) can fit the data

```
glenmark$arma12_1_weekly = arima(glaxo$diff1_weekly, order=c(12,0,1))

# ACF of residuals should be white if model captures entire information
acf(glenmark$arma12_1_weekly$residuals)
```



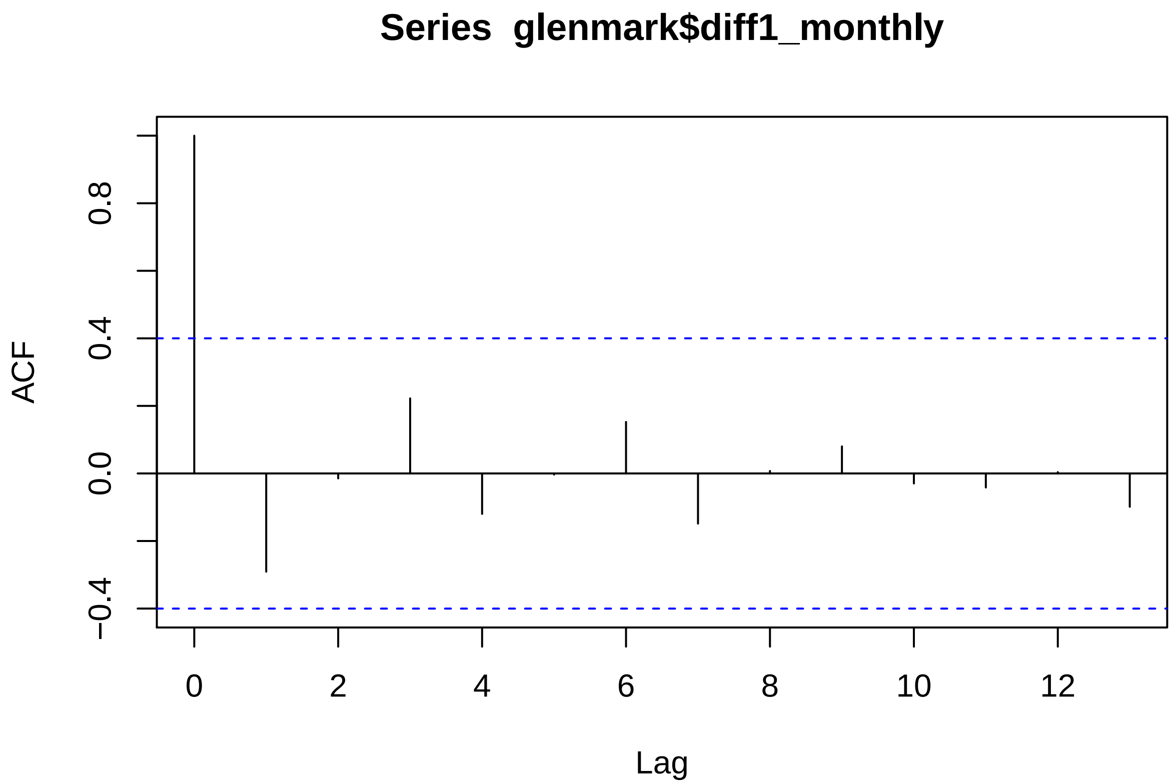

```
pacf(glenmark$arma12_1_weekly$residuals)
```



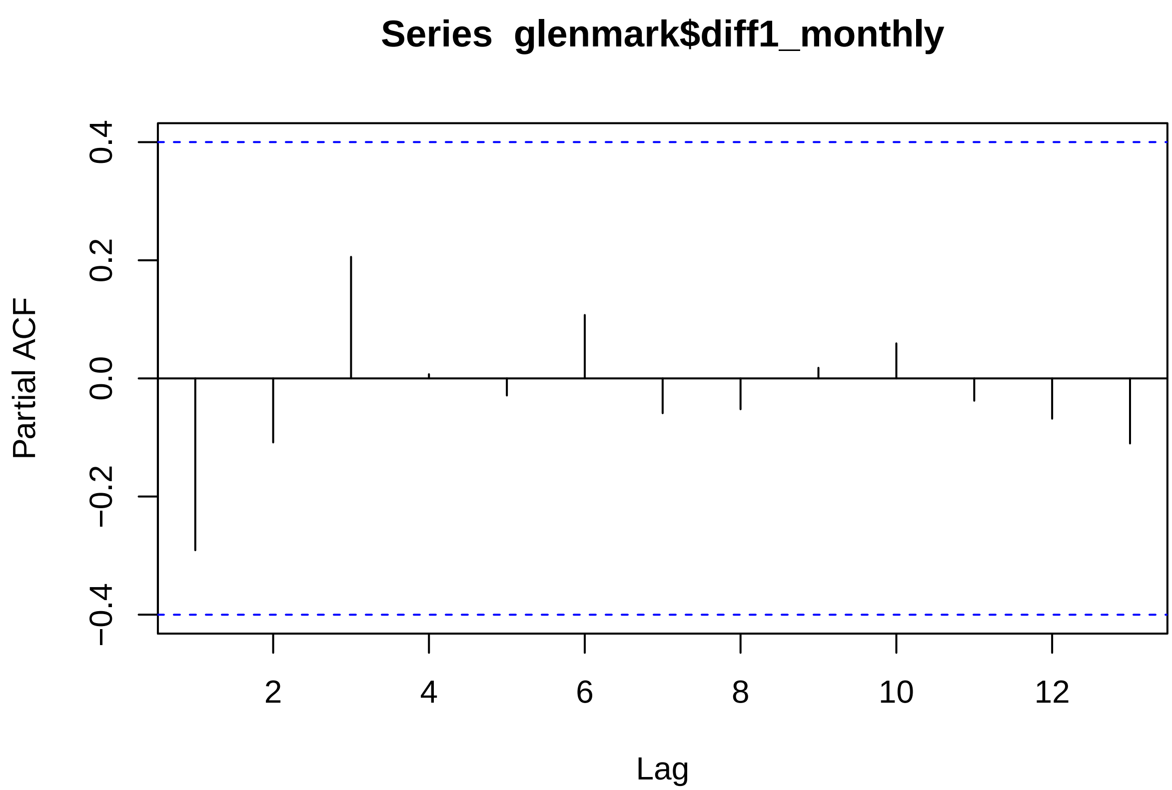
- By fitting ARMA(12,1) model predictable component of the data is exploited, residuals obtained forms white noise
- ARMA trend is followed in case of Daily data, Data can be made stationary after considering first order difference

Monthly Series

```
acf(glenmark$diff1_monthly)
```




```r
pacf(glenmark$diff1_monthly)
```




```r
# As ACF and PACF are inbetween significant bands, this shows that signal
# is random (i.e there exists no trend to model) and stationary
```

- As it can be seen that acf at any lag is bounded by significant limits, which means that the stationary series obtained after taking first order difference is white noise.

Data Correlation

Based on above analysis it can be seen that time series with daily frequency has predictable trend in this section, correlation between multiple companies are estiamtes to find the best trading pairs.

```
data = c()
data$aplltd = aplltd$sd_daily
data$auropharma = auropharma$sd_daily
data$glaxo = glaxo$sd_daily
data$glenmark = glenmark$sd_daily
data$sunpharma = sunpharma$sd_daily

print(cor(data.frame(data)))
```

```
##                aplltd auropharma     glaxo  glenmark sunpharma
## aplltd     1.0000000  0.2614837 0.5011108 0.6687222 0.8423672
## auropharma 0.2614837  1.0000000 0.2538927 0.2542017 0.2439262
## glaxo      0.5011108  0.2538927 1.0000000 0.2393000 0.5234630
## glenmark   0.6687222  0.2542017 0.2393000 1.0000000 0.8175574
## sunpharma  0.8423672  0.2439262 0.5234630 0.8175574 1.0000000
```

Based on correlation matrix it can clearly oberved that best pairs to consider are:

- APLLTD & SUNPHARMA (corr = 0.8423)
- GLENMARK & SUNPHARMA (corr = 0.8175)

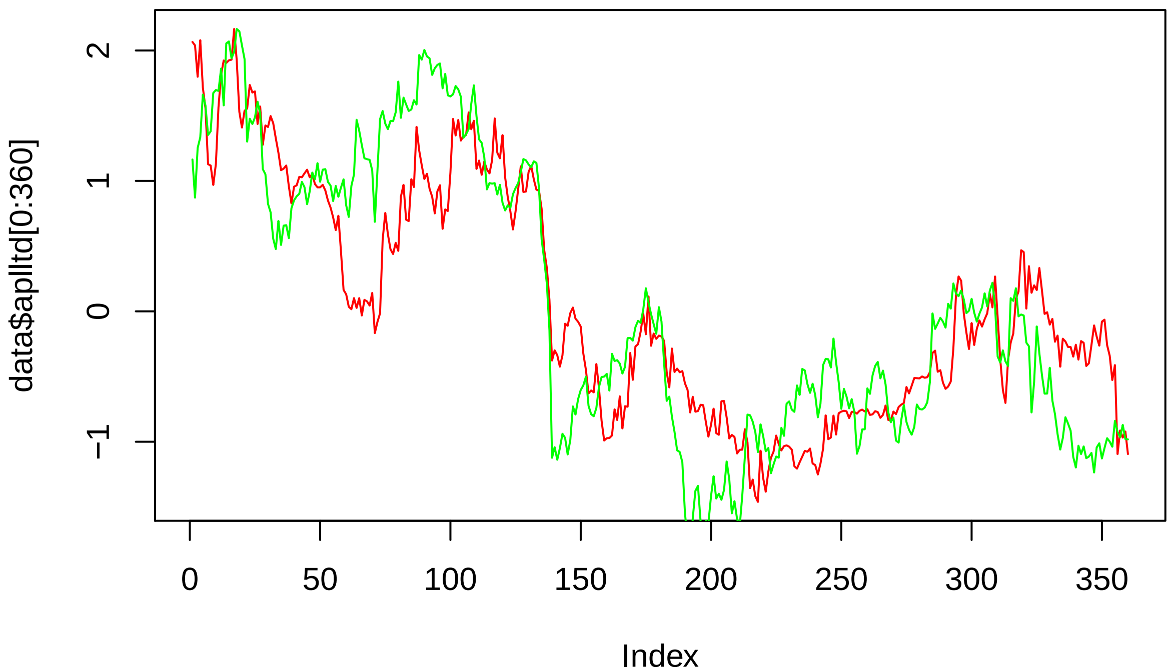
```
findCrossOverIndex <- function(data, threshold, window = 10){
  index_data = (data < -1*threshold) + (data > threshold)
  index = NULL
  count = 0
  prev = FALSE
  for (ii in seq(1, length(index_data))){
    if (index_data[ii] && count == 0){
      prev = TRUE
      count = count + 1
    }
    else if (index_data[ii] && prev){
      count = count + 1
    }
    else{
      if (count > window){
        index = append(index, c(ii, count))
      }
      count = 0
      prev = FALSE
    }
  }
  return(index)
}
```

```
findProfit = function(comp1, comp2, idx, lower = 360){
  profit = 0
  for (ii in seq(1, floor(length(idx)/2))){
    pidx = lower + idx[2*ii - 1] - floor(idx[2*ii]/2)
    print (c(pidx, idx[2*ii-1], floor(idx[2*ii]/2)))
    profit = profit + abs(comp1[pidx] - comp2[pidx])
  }
  return (profit)
}
```

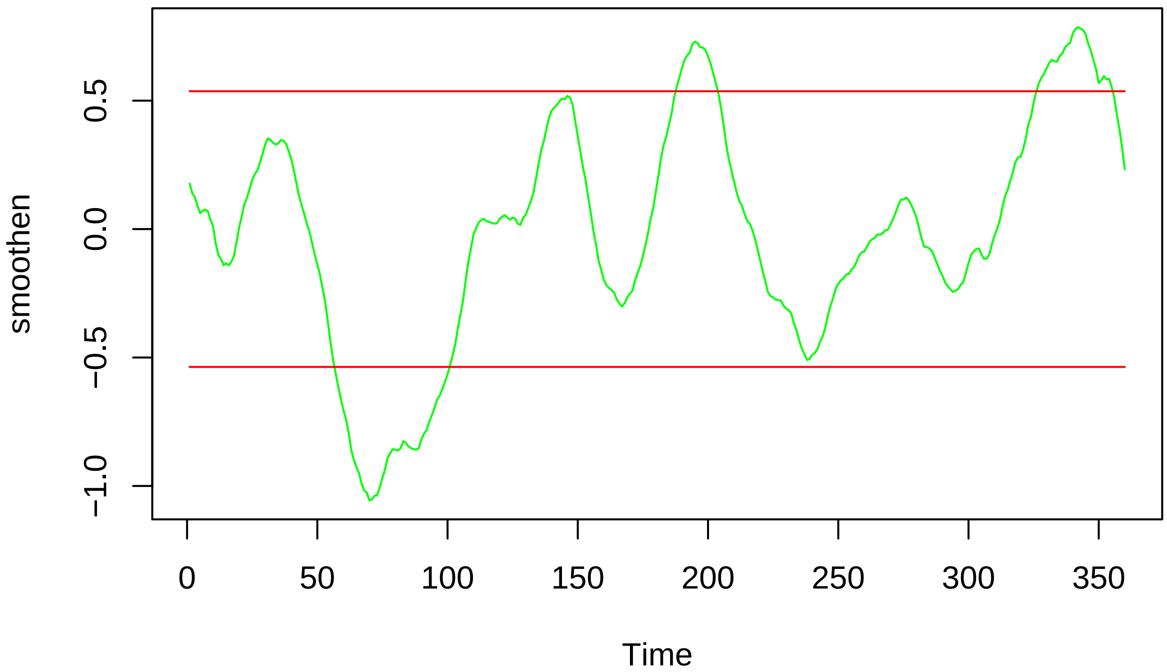
APLLTD & SUNPHARMA

Threshold estimation

```
plot(data$aplltd[0:360], type='l', col='red')
lines(data$sunpharma[0:360], type='l', col='green')
```

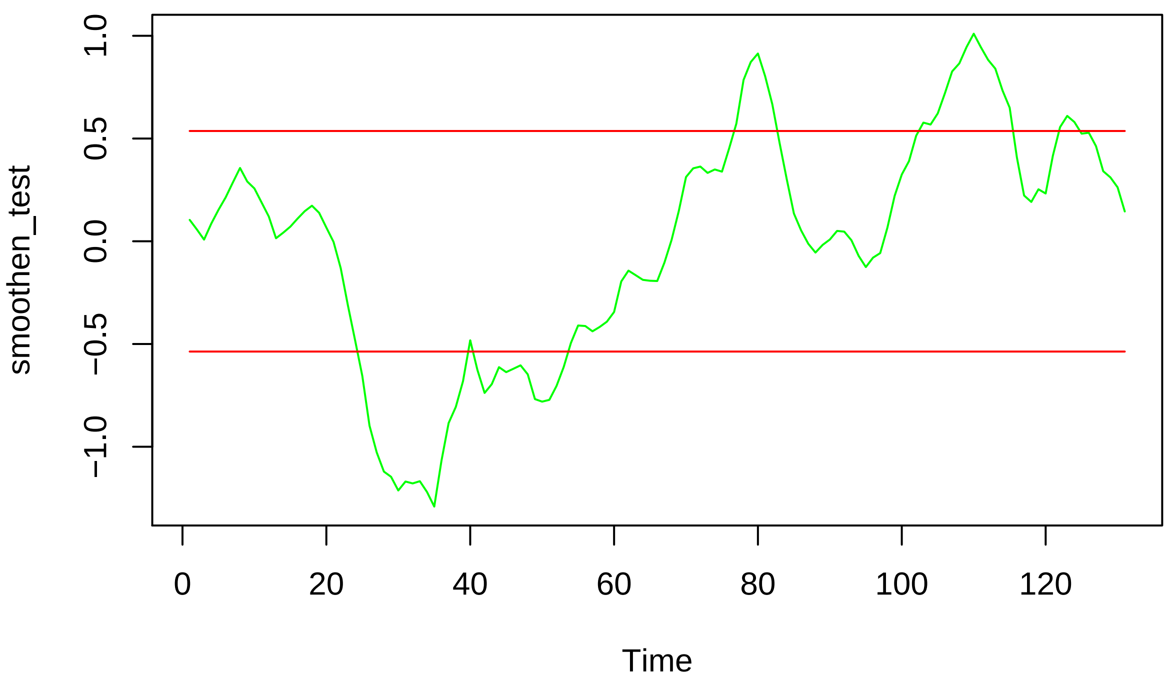



```
spread_p1_train = data$aplltd[0:360] - data$sunpharma[0:360]
threshold = 1.0*sd(spread_p1_train)

smoothen = filter(spread_p1_train, rep(1/20, 20), circular = T)
plot(smoothen, type='l', col='green')
lines(rep(threshold, length(smoothen)) , col='red')
lines(rep(-1*threshold, length(smoothen)) , col='red')
```



Testing

```
spread_p1_test = data$aplltd[360:490] - data$sunpharma[360:490]

smoothen_test = filter(spread_p1_test, rep(1/5, 5), circular = T)
plot(smoothen_test, type='l', col='green')
lines(rep(threshold, length(smoothen_test)) , col='red')
lines(rep(-1*threshold, length(smoothen_test)) , col='red')
```



```r
# net profit estimate per stock
index = findCrossOverIndex(smoothen_test, threshold, window = 10)
profit = findProfit(aplltd$raw_daily, sunpharma$raw_daily, index)
```

```
## [1] 393  40   7
## [1] 408  54   6
## [1] 470 116   6
```

```r
print(paste0("[INFO] Number of trading instances: ", floor(length(index)/2)))
```

```
## [1] "[INFO] Number of trading instances: 3"
```

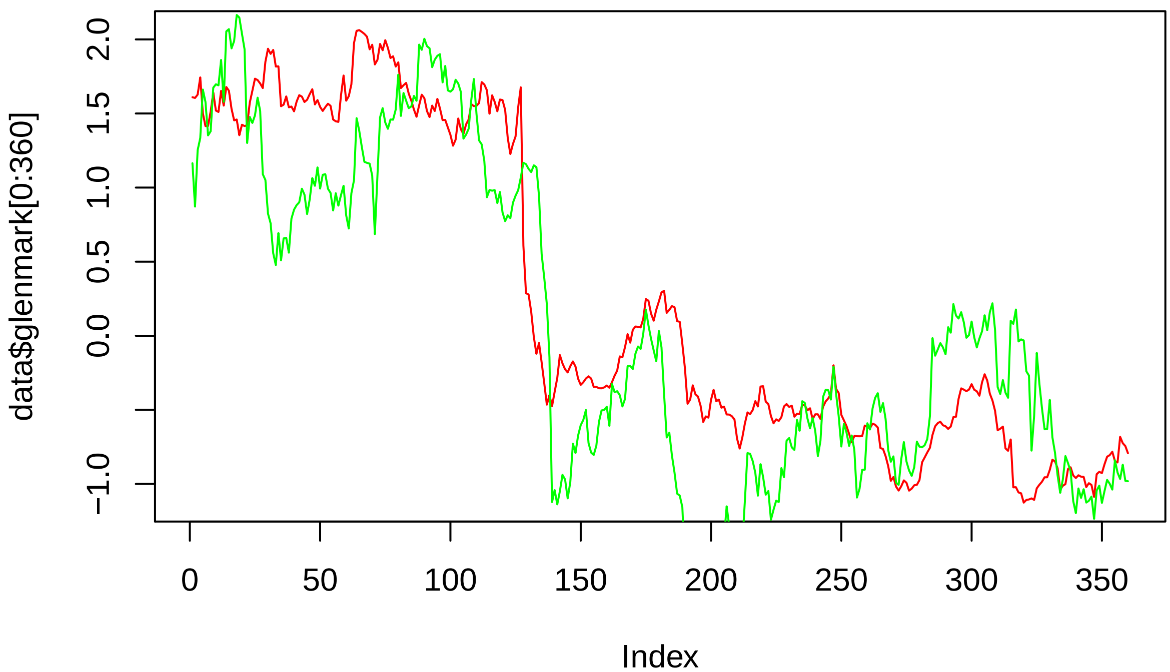
```r
print(paste0("[INFO] Profit made per stock by following the statistic is given by: Rs. ", profit))
```

```
## [1] "[INFO] Profit made per stock by following the statistic is given by: Rs. 122.650054"
```

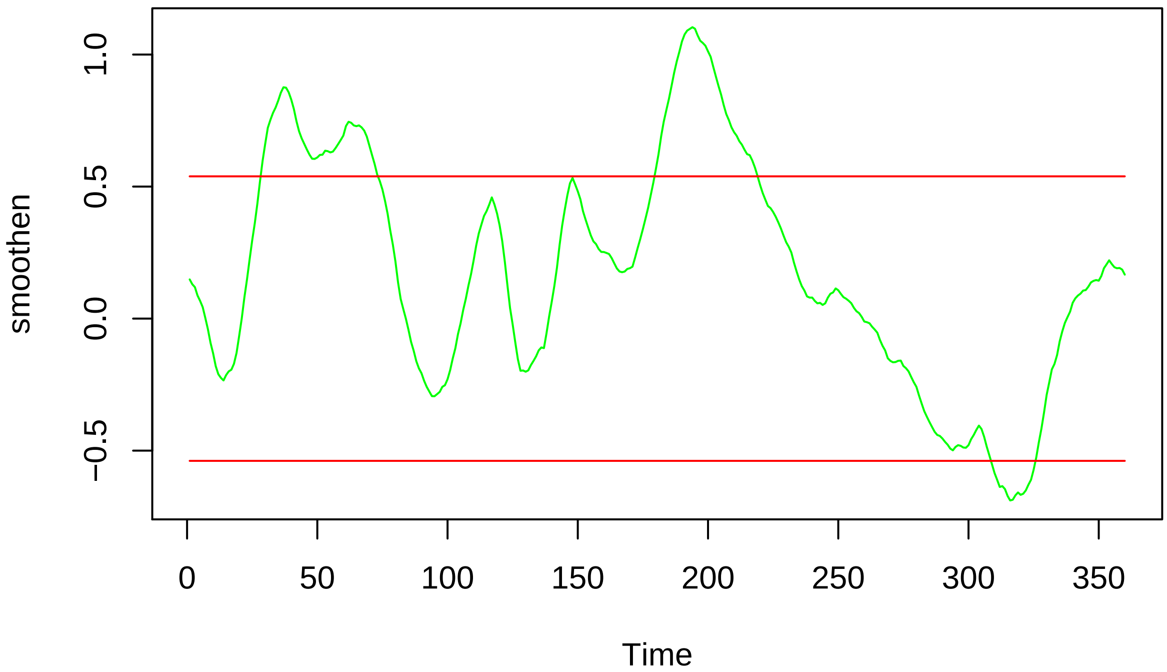
GLENMARK & SUNPHARMA

Threshold estimation

```r
plot(data$glenmark[0:360], type='l', col='red')
lines(data$sunpharma[0:360], type='l', col='green')
```

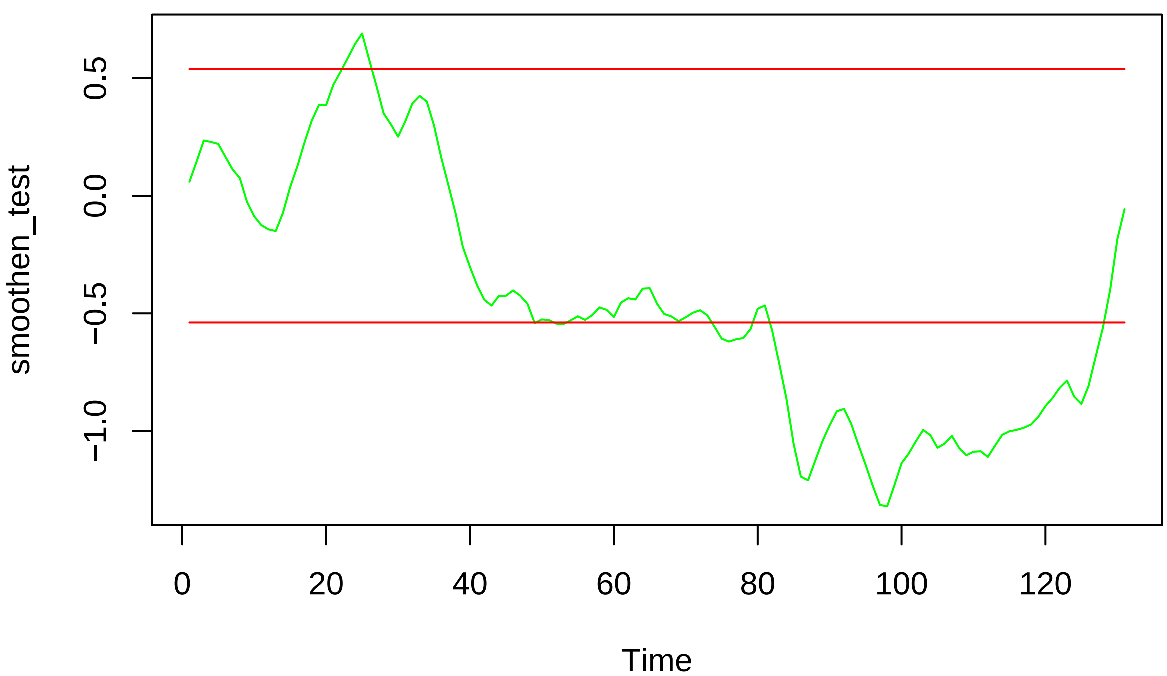



```r
spread_p1_train = data$glenmark[0:360] - data$sunpharma[0:360]
threshold = 1.0*sd(spread_p1_train)

smoothen = filter(spread_p1_train, rep(1/20, 20), circular = T)
plot(smoothen, type='l', col='green')
lines(rep(threshold, length(smoothen)) , col='red')
lines(rep(-1*threshold, length(smoothen)) , col='red')
```



Testing

```
spread_p1_test = data$glenmark[360:490] - data$sunpharma[360:490]

smoothen_test = filter(spread_p1_test, rep(1/5, 5), circular = T)
plot(smoothen_test, type='l', col='green')
lines(rep(threshold, length(smoothen_test)) , col='red')
lines(rep(-1*threshold, length(smoothen_test)) , col='red')
```



```r
# net profit estimate per stock
index = findCrossOverIndex(smoothen_test, threshold, window = 5)
profit = findProfit(aplltd$raw_daily, sunpharma$raw_daily, index)
```

```
## [1] 437  80   3
## [1] 466 129  23
```

```r
print(paste0("[INFO] Number of trading instances: ", floor(length(index)/2)))
```

```
## [1] "[INFO] Number of trading instances: 2"
```

```r
print(paste0("[INFO] Profit made per stock by following the statistic is given by: Rs. ", profit))
```

```
## [1] "[INFO] Profit made per stock by following the statistic is given by: Rs. 30.399964"
```