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)
}
```

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

Analysis of Alembic Pharmaceuticals Limited (APLLTD.NS)

```
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
```

```
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
```

```
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
```

```
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
```

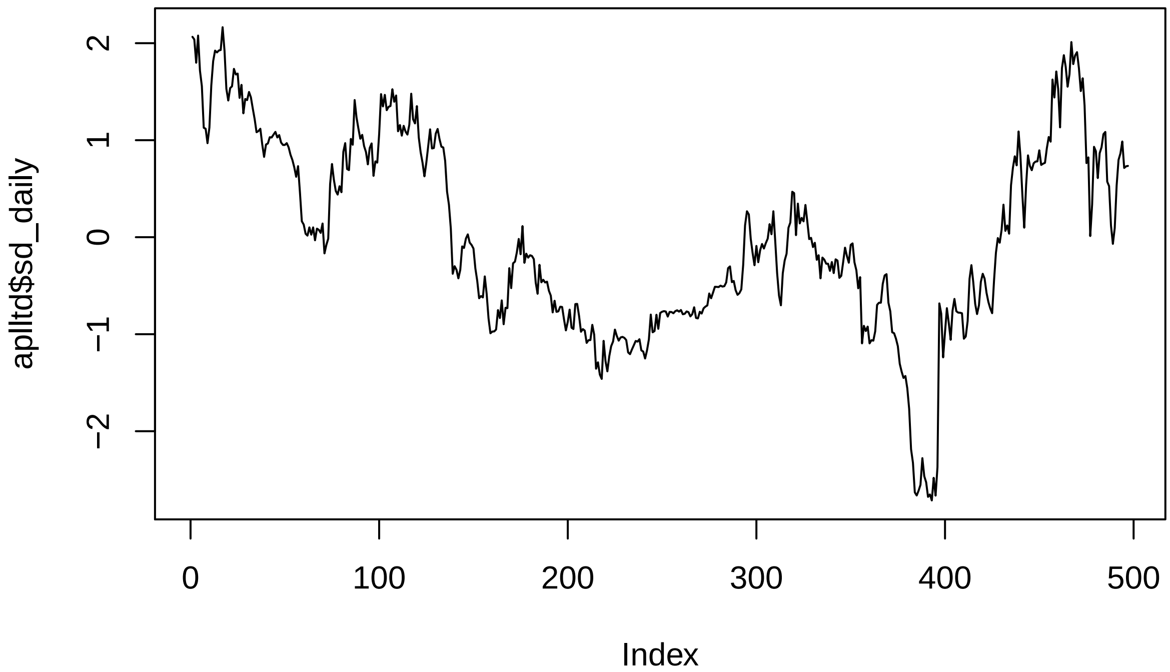
```
aplltd$sd_weekly = standardize(aplltd$raw_weekly)
```

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

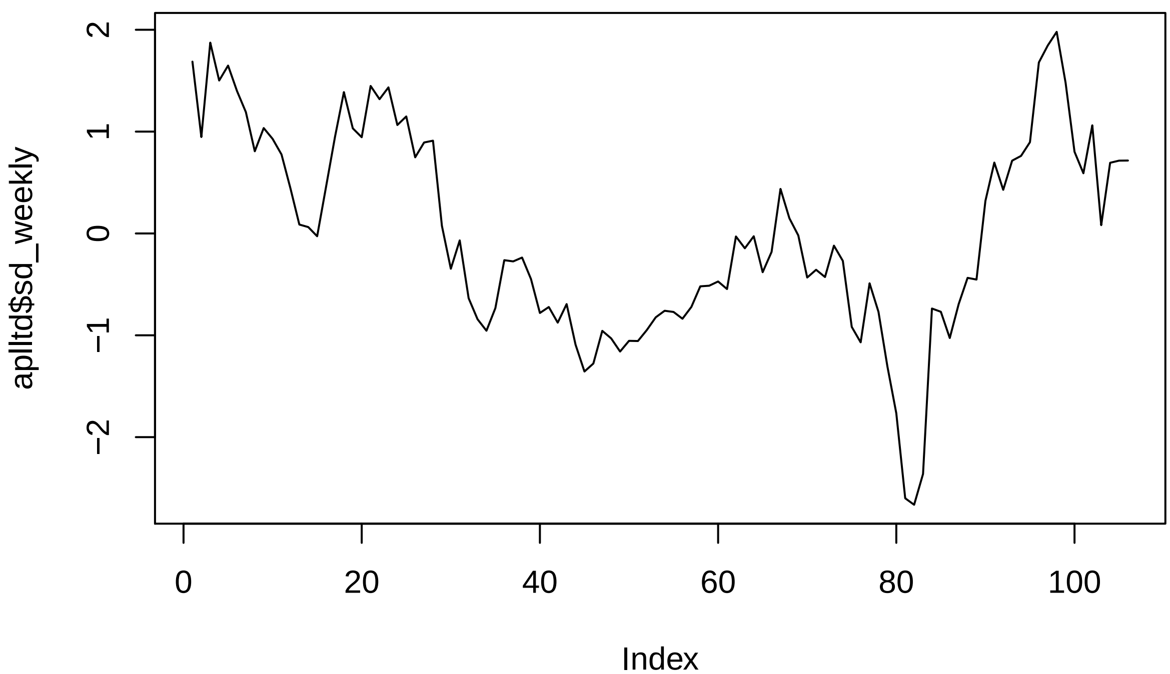
```
aplltd$sd_monthly = standardize(aplltd$raw_monthly)
```

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

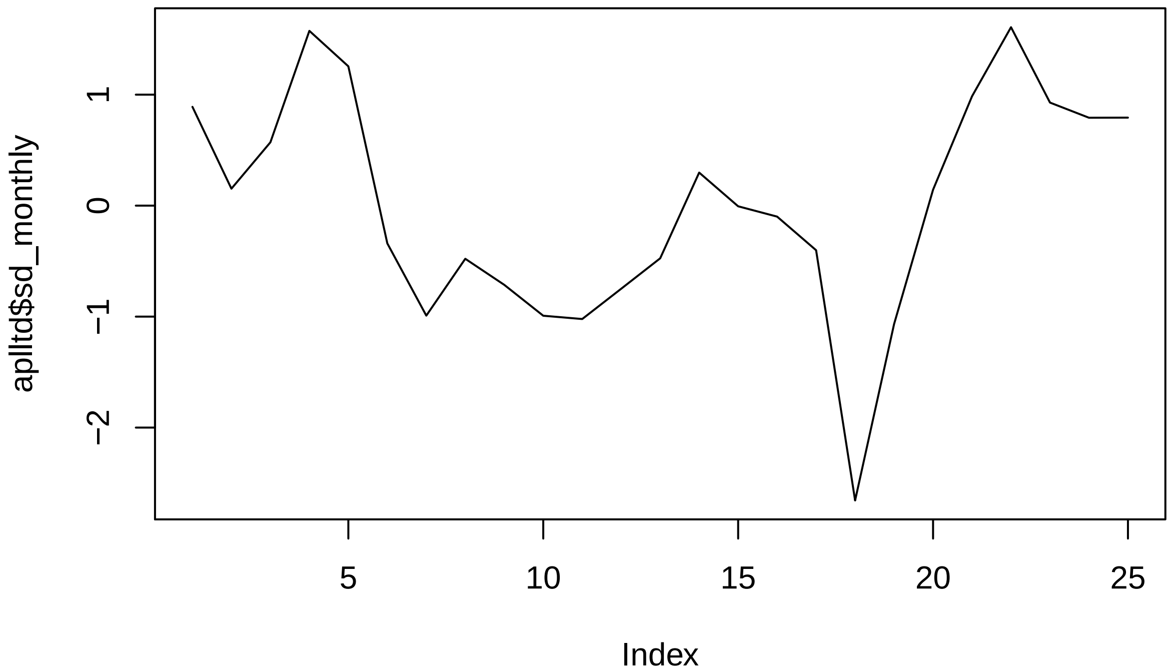
```
# plot all the data
plot(aplltd$sd_daily, type='l')
```



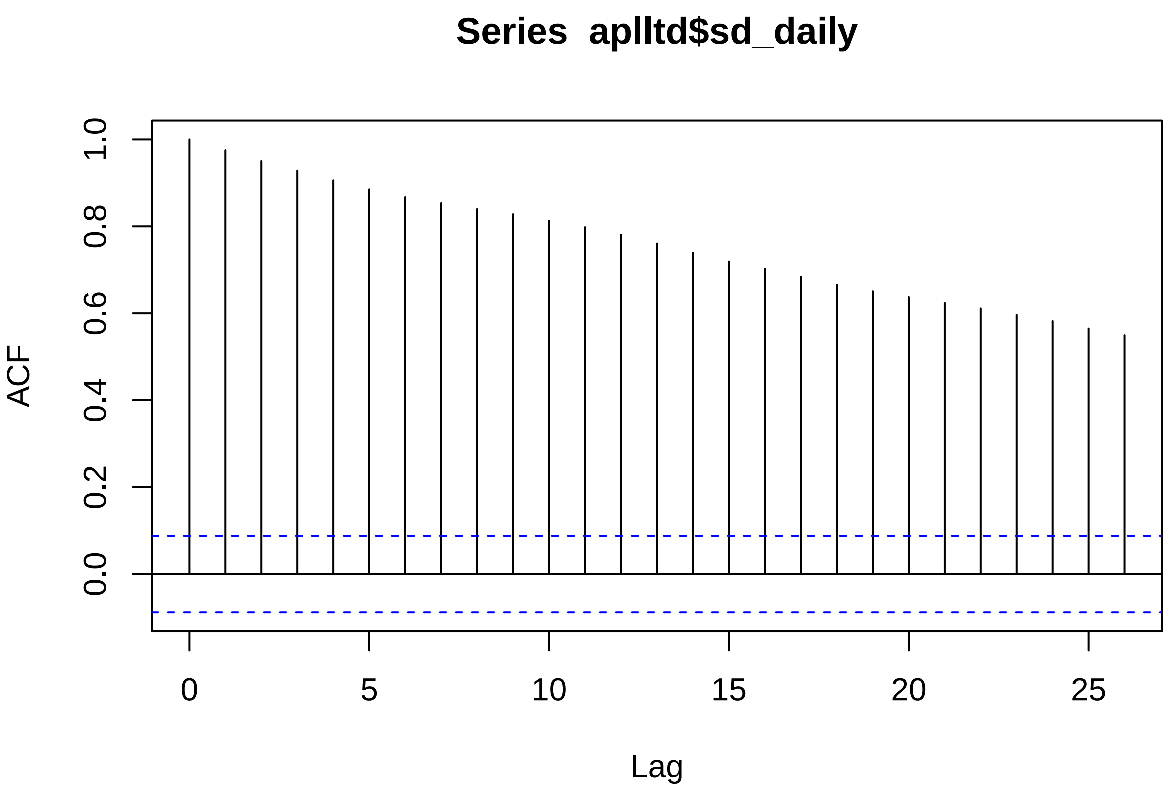

```
plot(aplltd$sd_weekly, type='l')
```



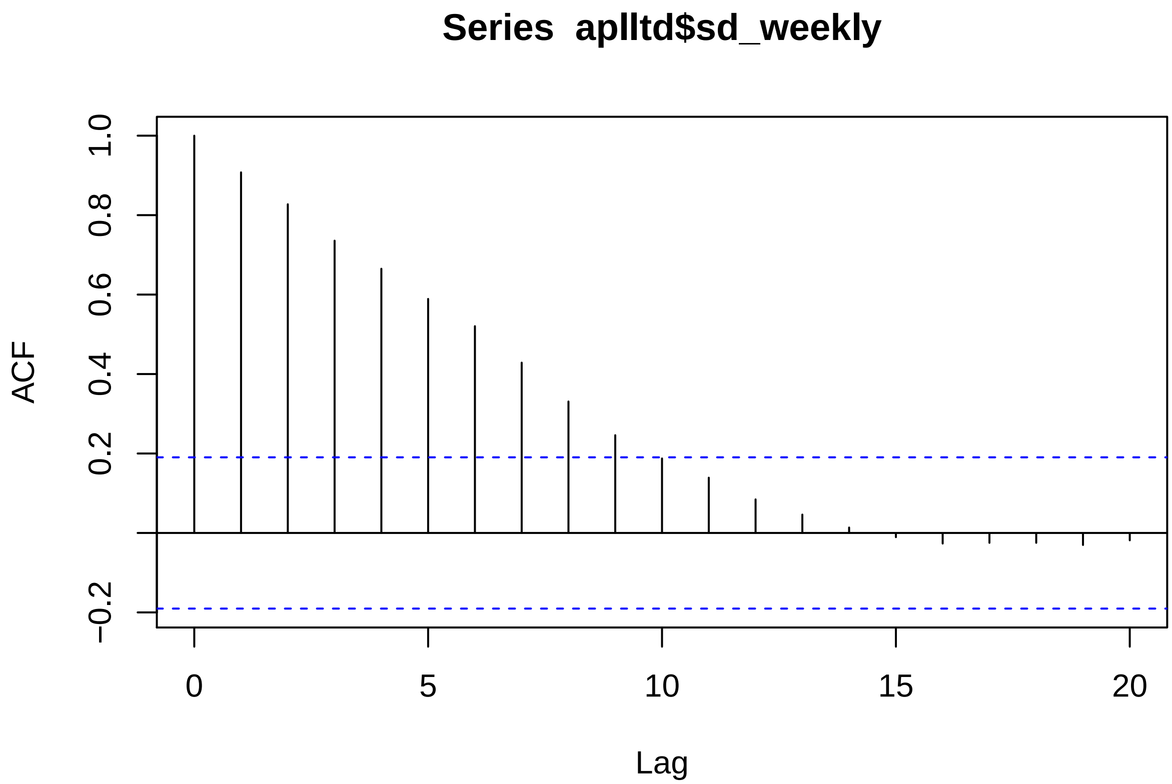

```
plot(aplltd$sd_monthly, type='l')
```



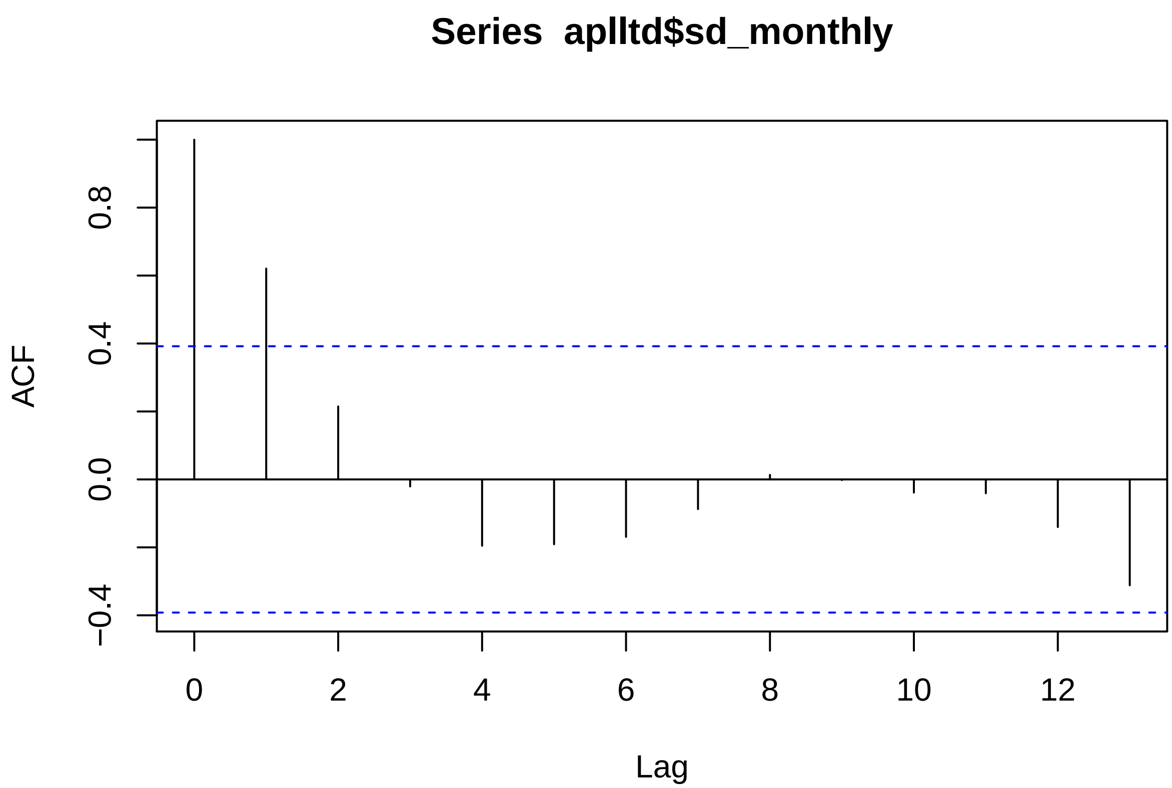

```r
# ACF analysis for sd data
acf(aplltd$sd_daily)
```



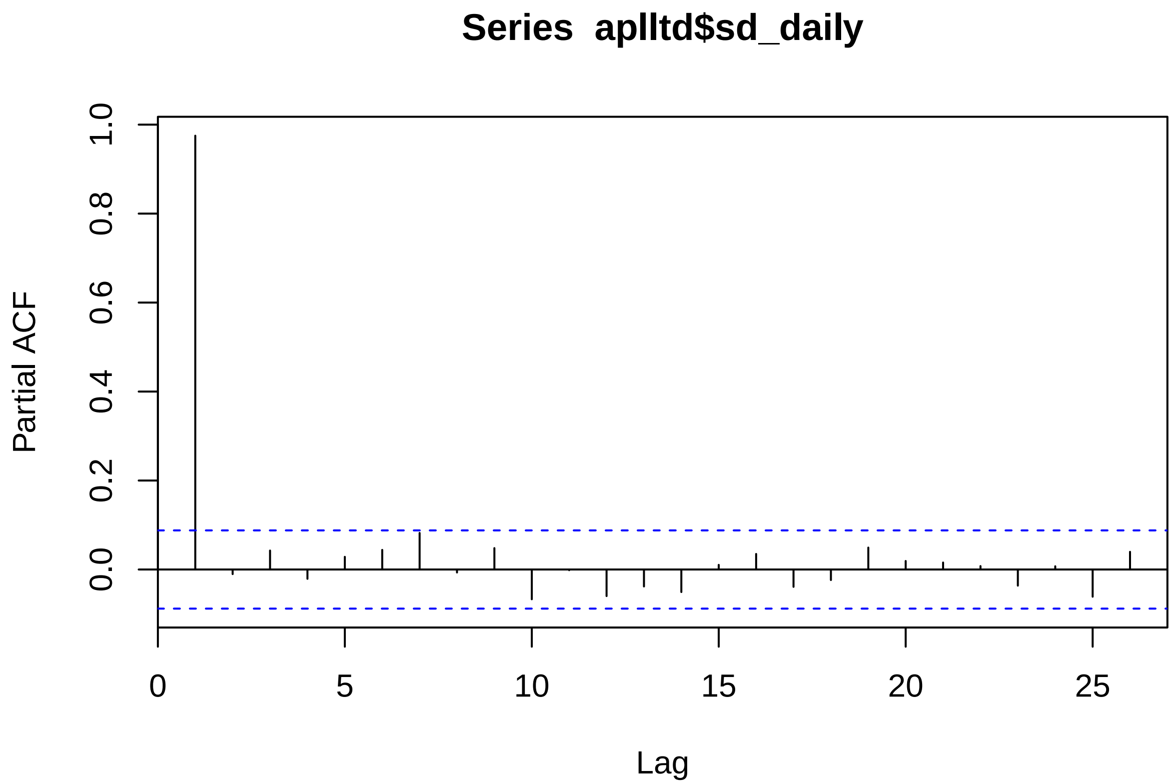

```r
acf(aplltd$sd_weekly)
```



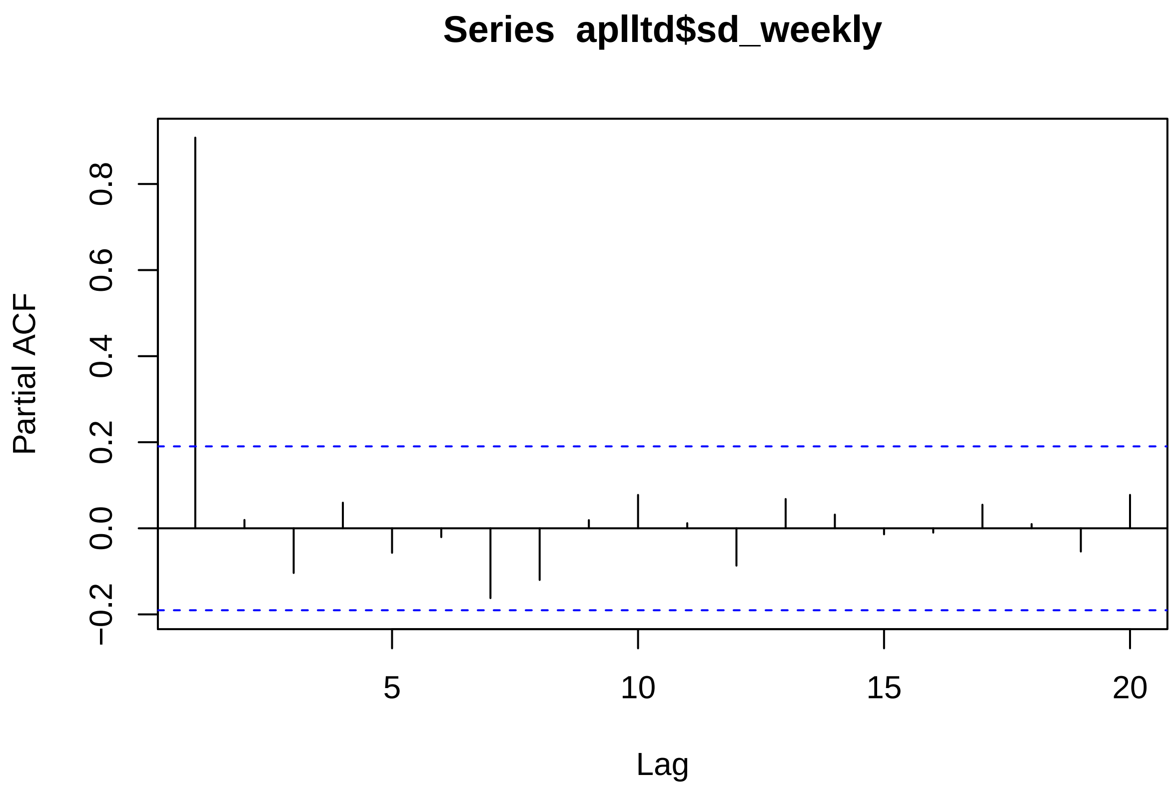

```
acf(aplltd$sd_monthly)
```



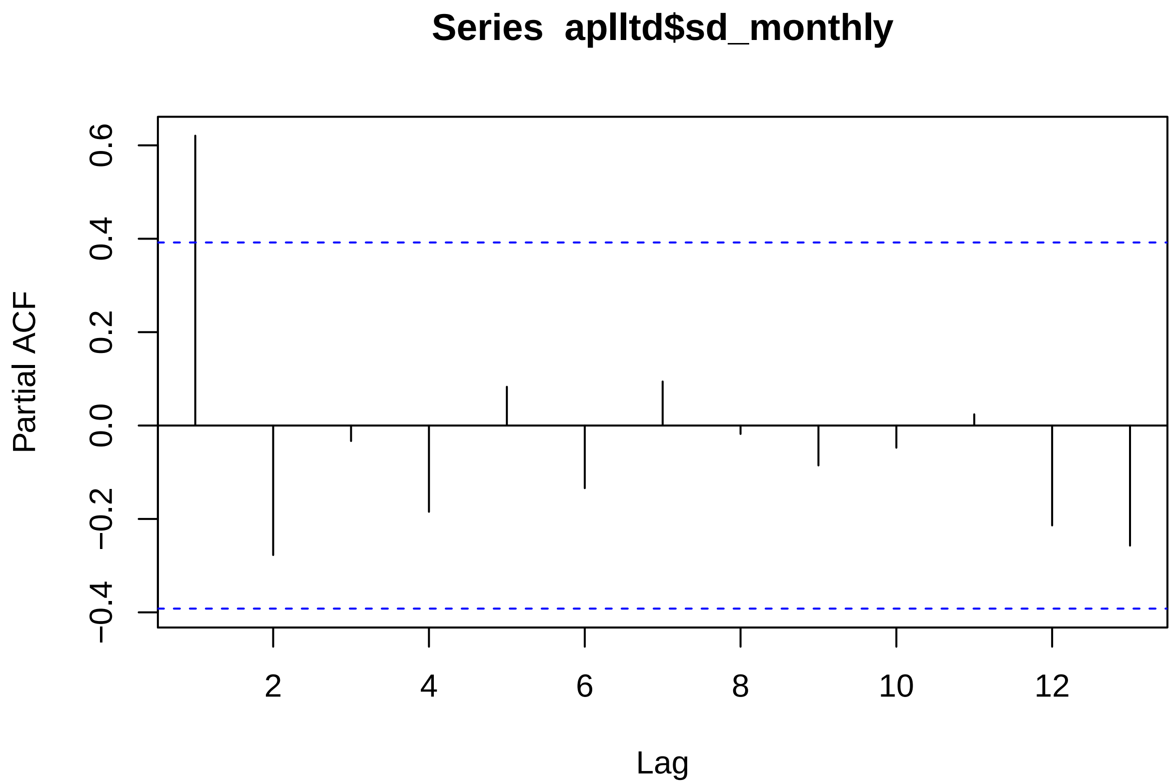

```
# PACF analysis for sd data
pacf(aplltd$sd_daily)
```




```
pacf(aplltd$sd_weekly)
```



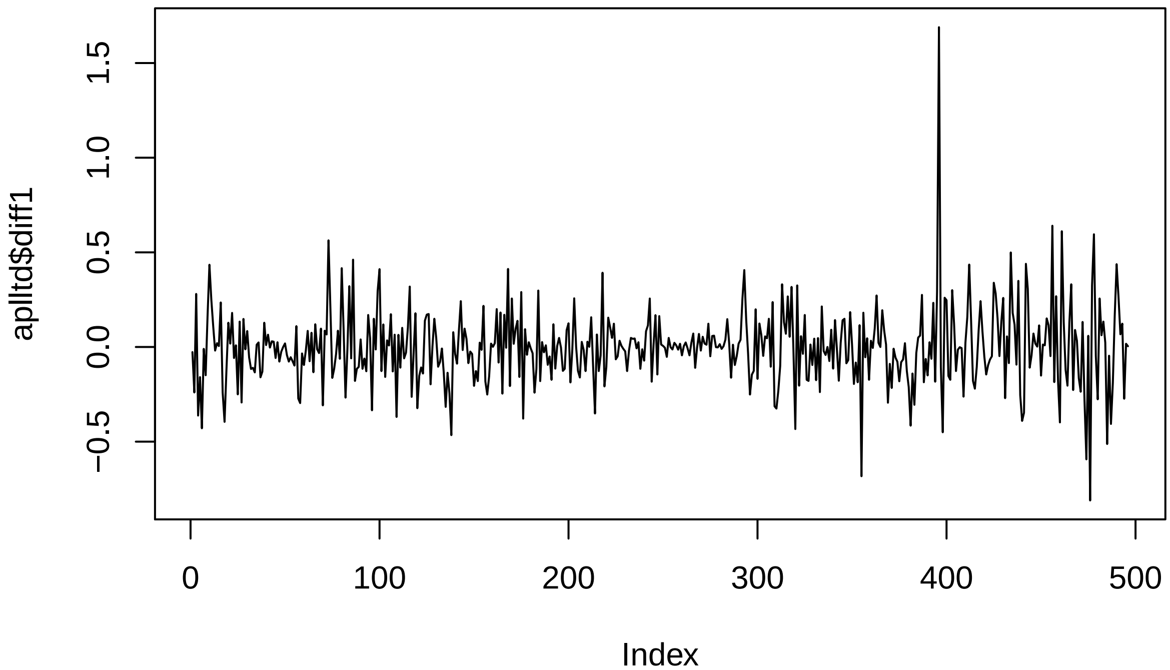

```
pacf(aplltd$sd_monthly)
```



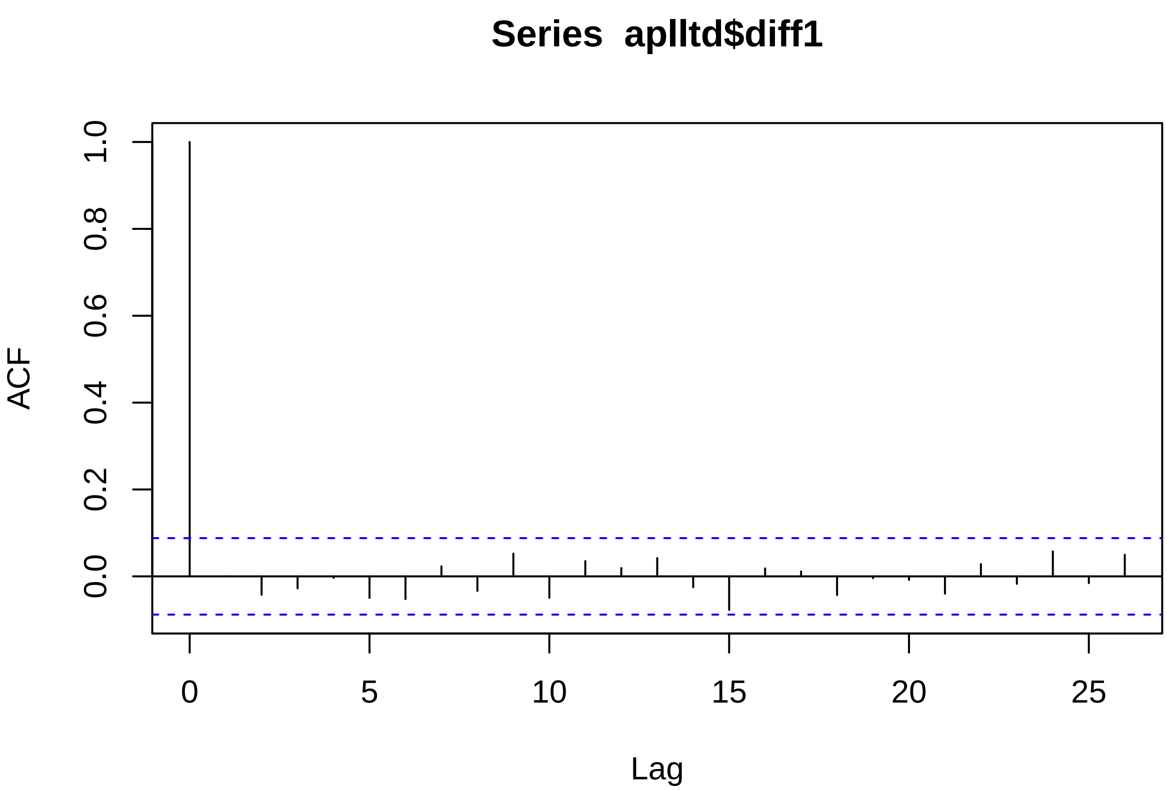
- By analysing acf plot it's clear that daily data has some integrating effect which can be removed by taking first order difference of the signal.
- Weekly Data can be modelled using Moving Average model with lag=9
- Monthly Data can be modelled using MA model with lag = 1

Daily Series

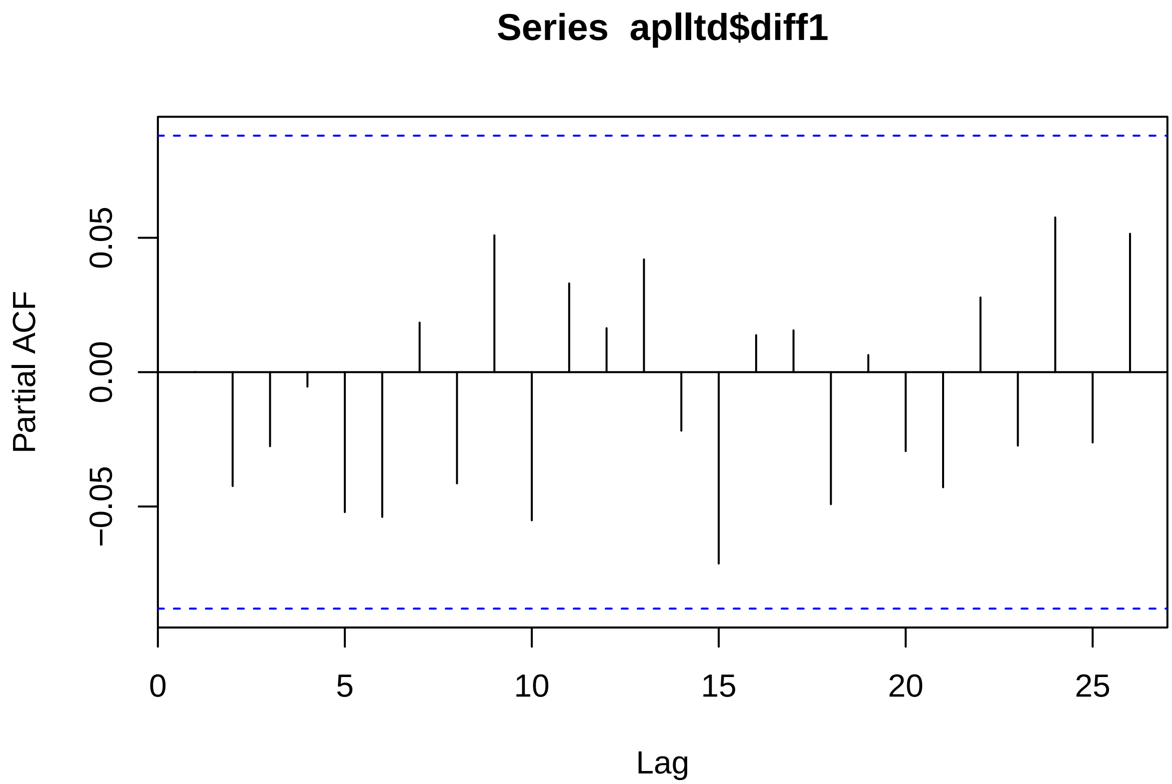
```
aplltd$diff1 = diff(aplltd$sd_daily)
plot(aplltd$diff1, type='l')
```




```r
# plot visually seems to be random stationary signal which can be
# therotically proved by plotting acf and pacf of signal
acf(aplltd$diff1)
```




```r
pacf(aplltd$diff1)
```




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

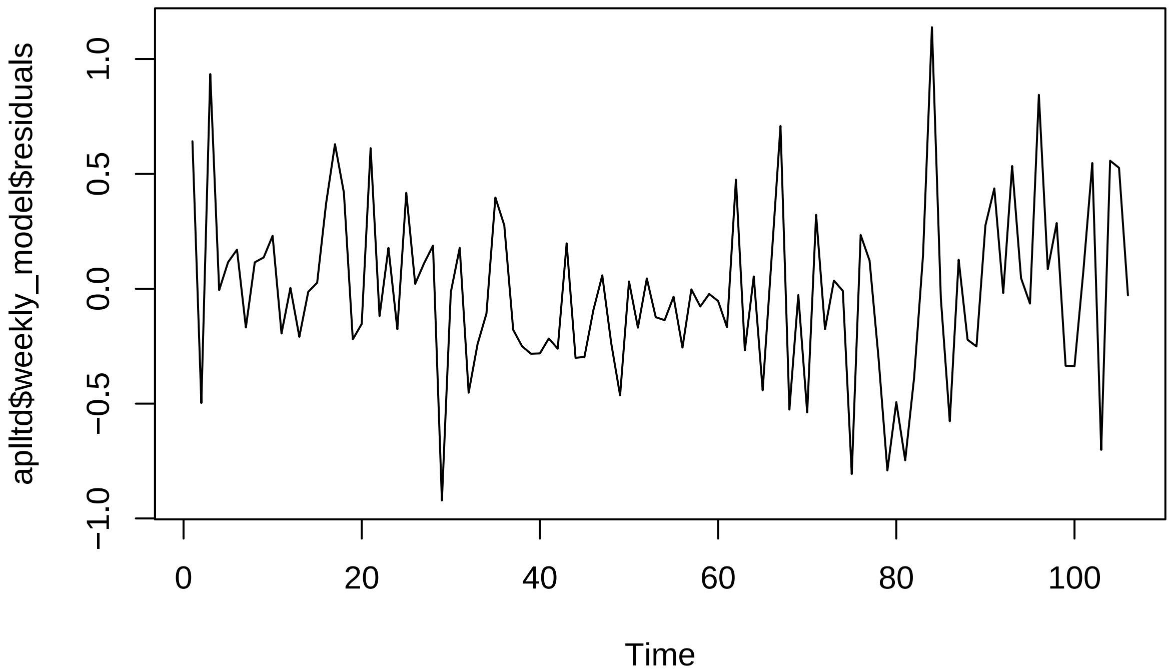
Daily data has no trend to exploit, Signal can be transformed to stationary signal by taking first order difference, which is illustrated in above plots.

Weekly Series

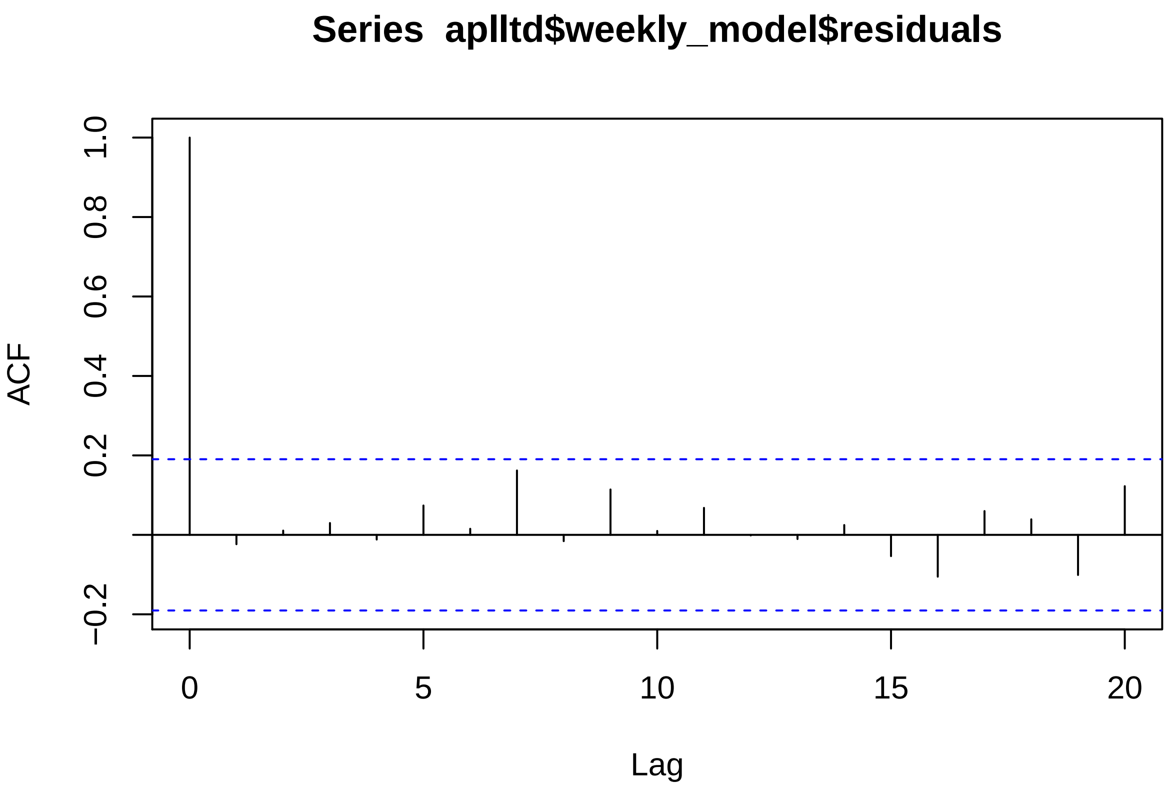
```
aplltd$weekly_model = arima (aplltd$sd_weekly, order = c(0,0,9), include.mean = T)
print(summary(aplltd$weekly_model))
```

```
##           Length Class  Mode
## coef       10    -none- numeric
## sigma2      1    -none- numeric
## var.coef  100    -none- numeric
## mask       10    -none- logical
## loglik      1    -none- numeric
## aic         1    -none- numeric
## arma        7    -none- numeric
## residuals 106    ts     numeric
## call        4    -none- call
## series      1    -none- character
## code        1    -none- numeric
## n.cond      1    -none- numeric
## nobs        1    -none- numeric
## model      10    -none- list
```

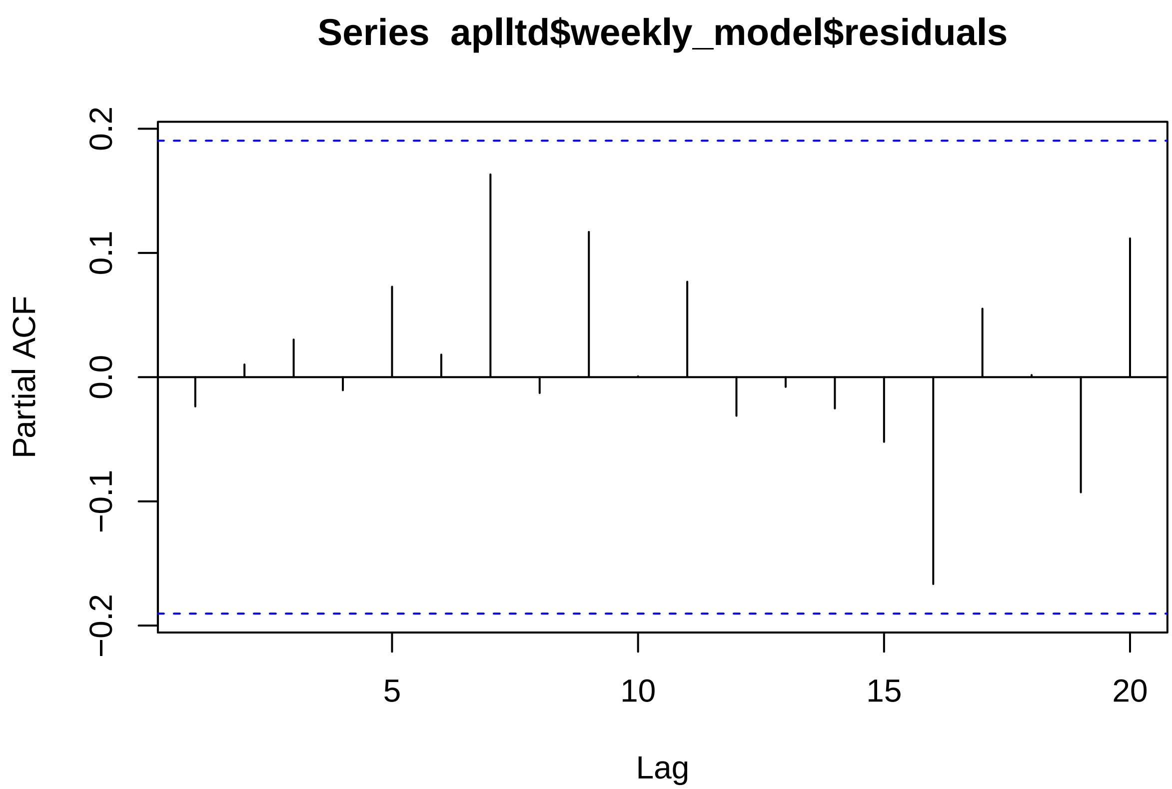
```
# residual analysis
plot(aplltd$weekly_model$residuals)
```




```r
# plot visually seems to be random stationary signal which can be
# therotically proved by plotting acf and pacf of signal
acf(aplltd$weekly_model$residuals)
```




```r
pacf(aplltd$weekly_model$residuals)
```




```r
# As ACF and PACF are inbetween significant bands shows that residual signal stationary
```

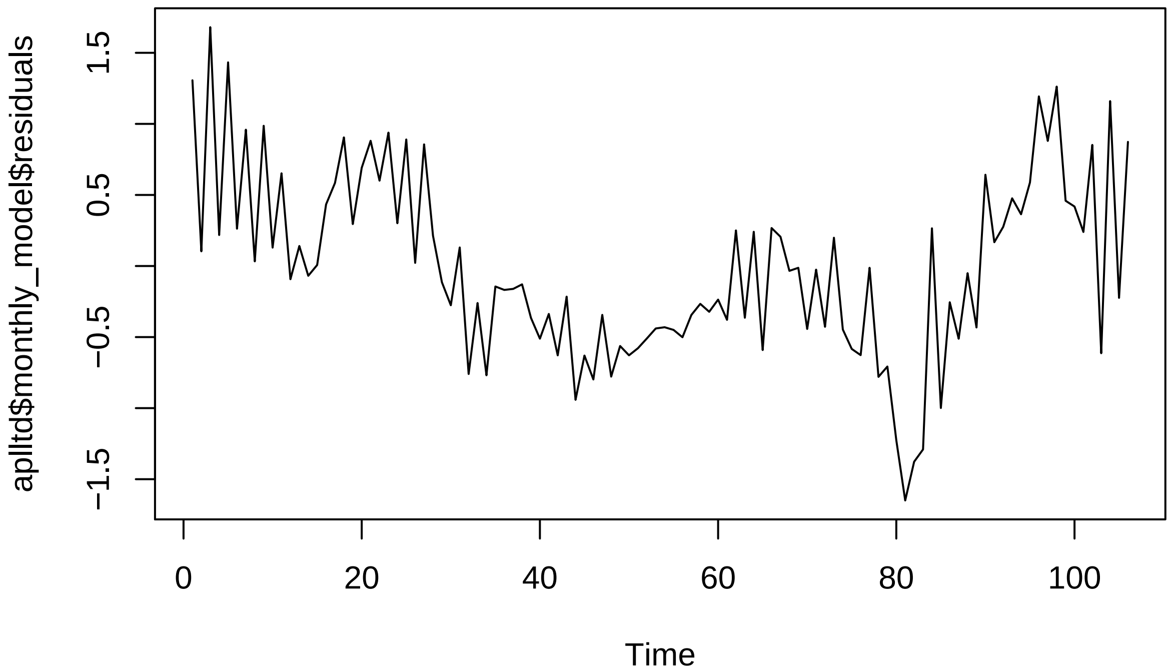
Weekly data has moving average (lag=9) trend which can be exploited, MA residuals form a stationary signal

Monthly Series

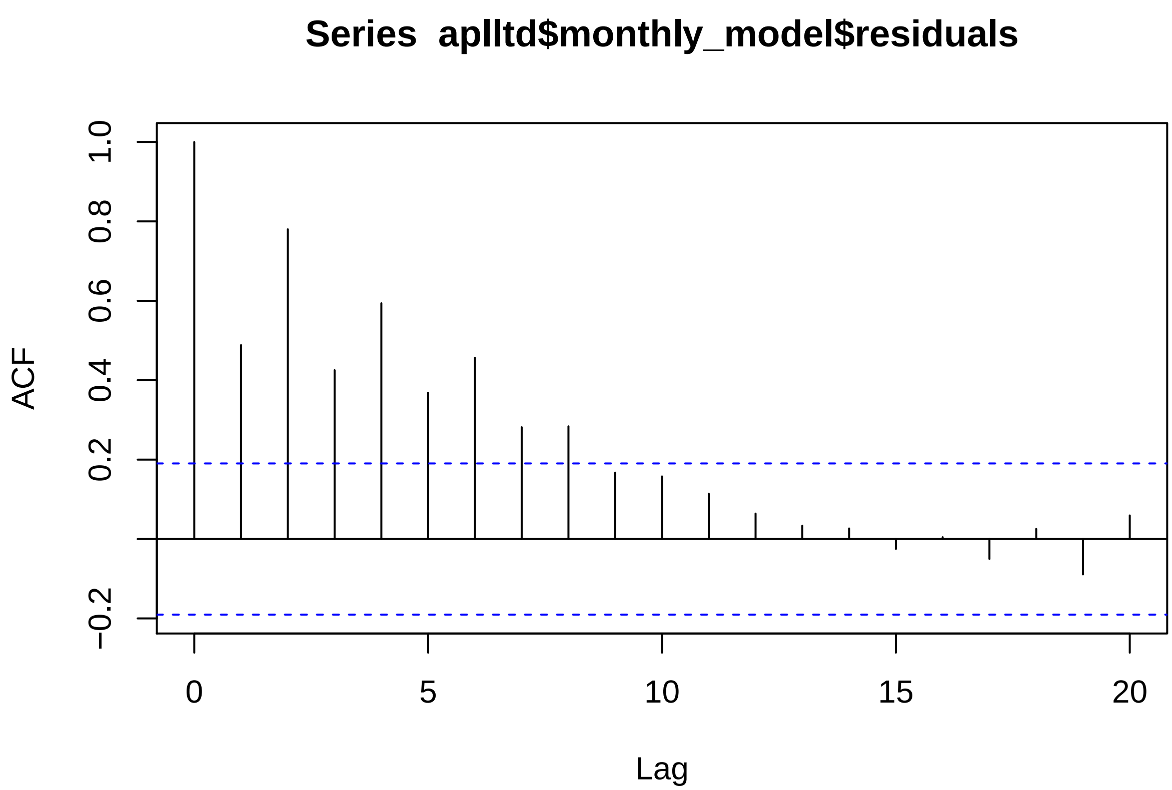
```r
aplltd$monthly_model = arima (aplltd$sd_weekly, order = c(0,0,1), include.mean = T)
print(summary(aplltd$monthly_model))
```

```
##           Length Class  Mode
## coef        2    -none- numeric
## sigma2      1    -none- numeric
## var.coef    4    -none- numeric
## mask        2    -none- logical
## loglik      1    -none- numeric
## aic         1    -none- numeric
## arma        7    -none- numeric
## residuals 106    ts     numeric
## call        4    -none- call
## series      1    -none- character
## code        1    -none- numeric
## n.cond      1    -none- numeric
## nobs        1    -none- numeric
## model      10    -none- list
```

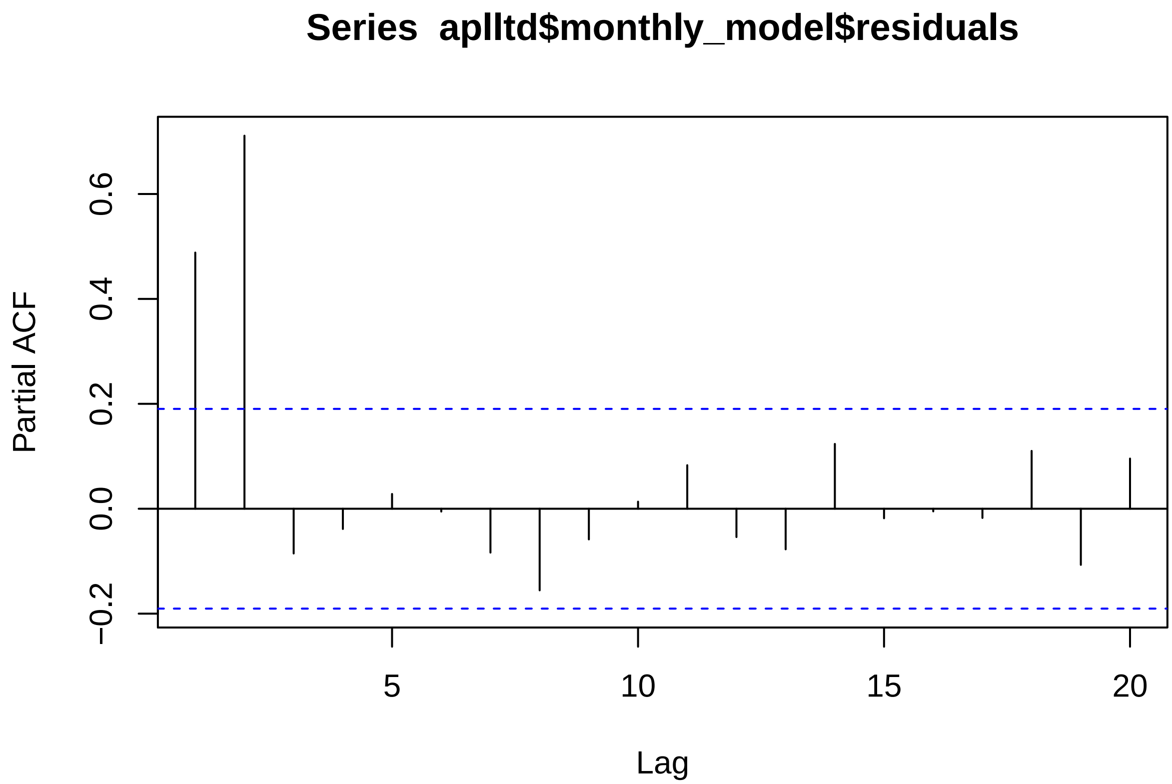
```r
# residual analysis
plot(aplltd$monthly_model$residuals)
```




```r
# plot visually seems to be random stationary signal which can be
# therotically proved by plotting acf and pacf of signal
acf(aplltd$monthly_model$residuals)
```




```r
pacf(aplltd$monthly_model$residuals)
```




```
# As ACF and PACF are inbetween significant bands shows that residual signal stationary
```

Monthly data has moving average (lag=1) trend which can be exploited, MA residuals form a stationary signal