Priya Laxman

Seasonality Identification

```
asx<-read_excel("C:/Docs/Forecasting/Assignment 1/ASX_data.xlsx")
View(asx)
class(asx)
```

```
## [1] "tbl_df"     "tbl"        "data.frame"
```

```
head(asx)
```

```
## # A tibble: 6 × 2
##        Month  price
##       <dttm>  <dbl>
## 1 2003-01-01 2935.4
## 2 2003-02-01 2778.4
## 3 2003-03-01 2848.6
## 4 2003-04-01 2970.9
## 5 2003-05-01 2979.8
## 6 2003-06-01 2999.7
```

```
#convert to time series object
asx.ts<-ts(asx$price, start = c(2003,1), frequency=12) #starting month is Jan 2003
View(asx.ts)
class(asx.ts) #now it is a ts object
```

```
## [1] "ts"
```

```
head(asx.ts)
```

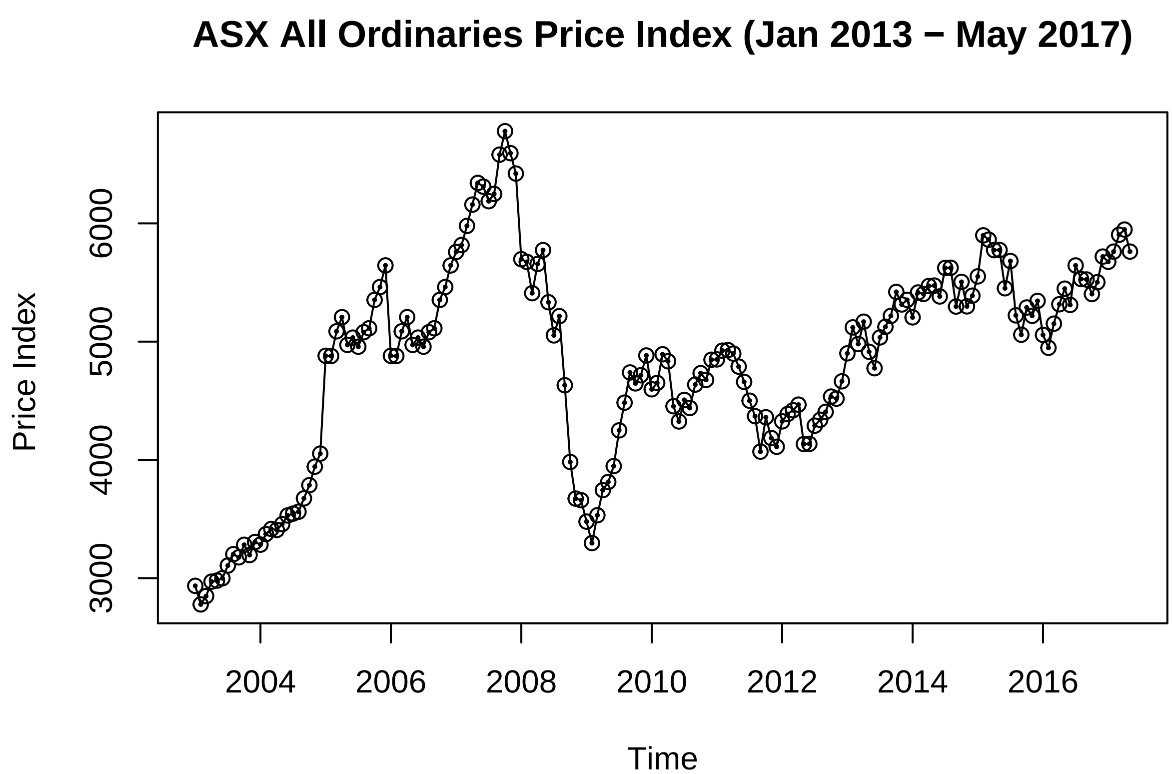
```
##         Jan    Feb    Mar    Apr    May    Jun
## 2003 2935.4 2778.4 2848.6 2970.9 2979.8 2999.7
```

```
asx.ts
```

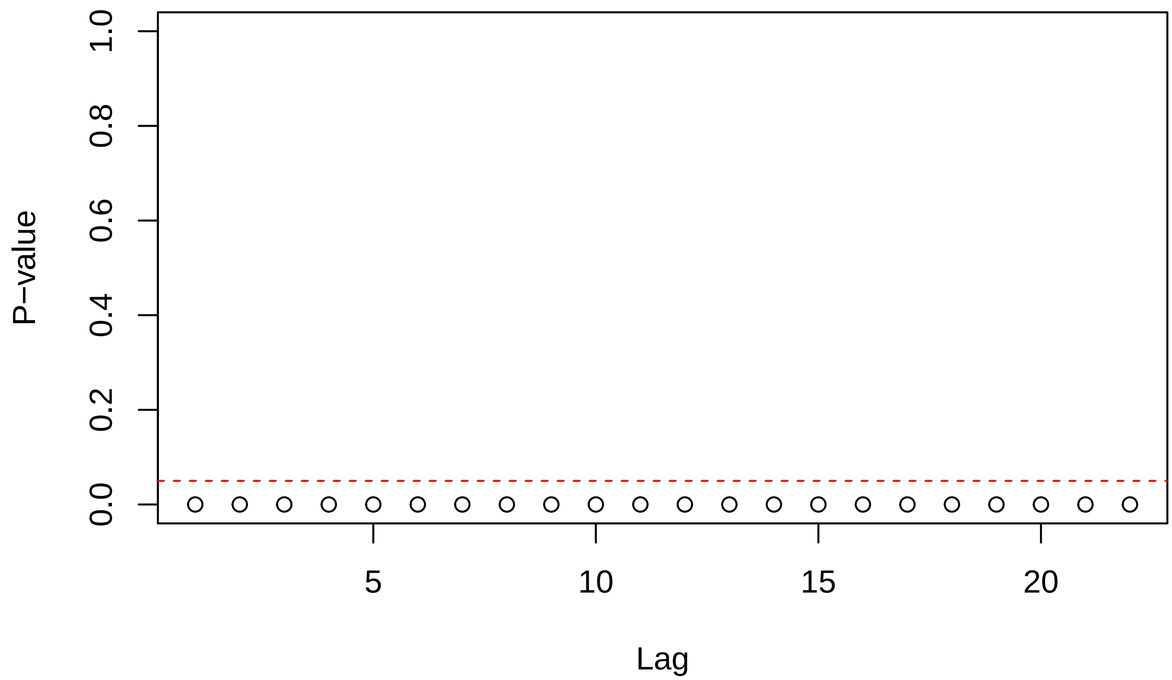
```
##          Jan     Feb     Mar     Apr     May     Jun     Jul     Aug
## 2003 2935.40 2778.40 2848.60 2970.90 2979.80 2999.70 3106.70 3202.90
## 2004 3283.60 3372.50 3416.40 3407.70 3456.90 3530.30 3546.10 3561.90
## 2005 4880.20 4878.40 5087.20 5207.00 4972.30 5034.00 4957.10 5079.80
## 2006 4880.20 4878.40 5087.20 5207.00 4972.30 5034.00 4957.10 5079.80
## 2007 5757.70 5816.50 5978.80 6158.30 6341.80 6310.60 6187.50 6248.30
## 2008 5697.00 5674.70 5409.70 5656.90 5773.90 5332.80 5052.60 5215.50
## 2009 3478.10 3296.90 3532.30 3744.70 3813.30 3947.80 4249.50 4484.10
## 2010 4596.90 4651.10 4893.10 4833.90 4453.60 4324.80 4507.40 4438.80
## 2011 4849.90 4923.60 4928.60 4899.00 4788.90 4659.80 4500.50 4369.90
```

```
## 2012 4325.70 4388.10 4420.00 4467.20 4133.70 4135.50 4289.40 4339.00
## 2013 4901.00 5120.40 4979.90 5168.60 4914.00 4775.40 5035.70 5125.30
## 2014 5205.10 5415.40 5403.00 5470.80 5473.80 5382.00 5623.10 5624.60
## 2015 5551.60 5898.50 5861.90 5773.67 5774.90 5451.20 5681.70 5222.10
## 2016 5056.60 4947.90 5151.80 5316.00 5447.80 5310.40 5644.00 5529.40
## 2017 5675.00 5761.00 5903.80 5947.60 5761.30
##          Sep     Oct     Nov     Dec
## 2003 3176.20 3282.40 3195.70 3306.00
## 2004 3674.70 3786.30 3942.80 4053.10
## 2005 5113.00 5352.90 5461.60 5644.30
## 2006 5113.00 5352.90 5461.60 5644.30
## 2007 6580.90 6779.10 6593.60 6421.00
## 2008 4631.30 3982.70 3672.70 3659.30
## 2009 4739.30 4646.90 4715.50 4882.70
## 2010 4636.90 4733.40 4676.40 4846.90
## 2011 4070.10 4360.50 4184.67 4111.00
## 2012 4406.30 4535.40 4518.00 4664.60
## 2013 5217.70 5420.30 5314.30 5353.10
## 2014 5296.80 5505.02 5298.10 5388.60
## 2015 5058.60 5288.60 5218.20 5344.60
## 2016 5525.20 5402.40 5502.40 5719.10
## 2017
```

```r
plot(asx.ts,ylab='Price Index ',xlab='Time',type='o', main="ASX All Ordinaries Price Index (Jan 2013 - 
points(y=asx.ts,x=time(asx.ts), cex=0.2)
```



```r
#Mc Leod Li Test for conditional heteroscedasticity
McLeod.Li.test(y=asx.ts)
```

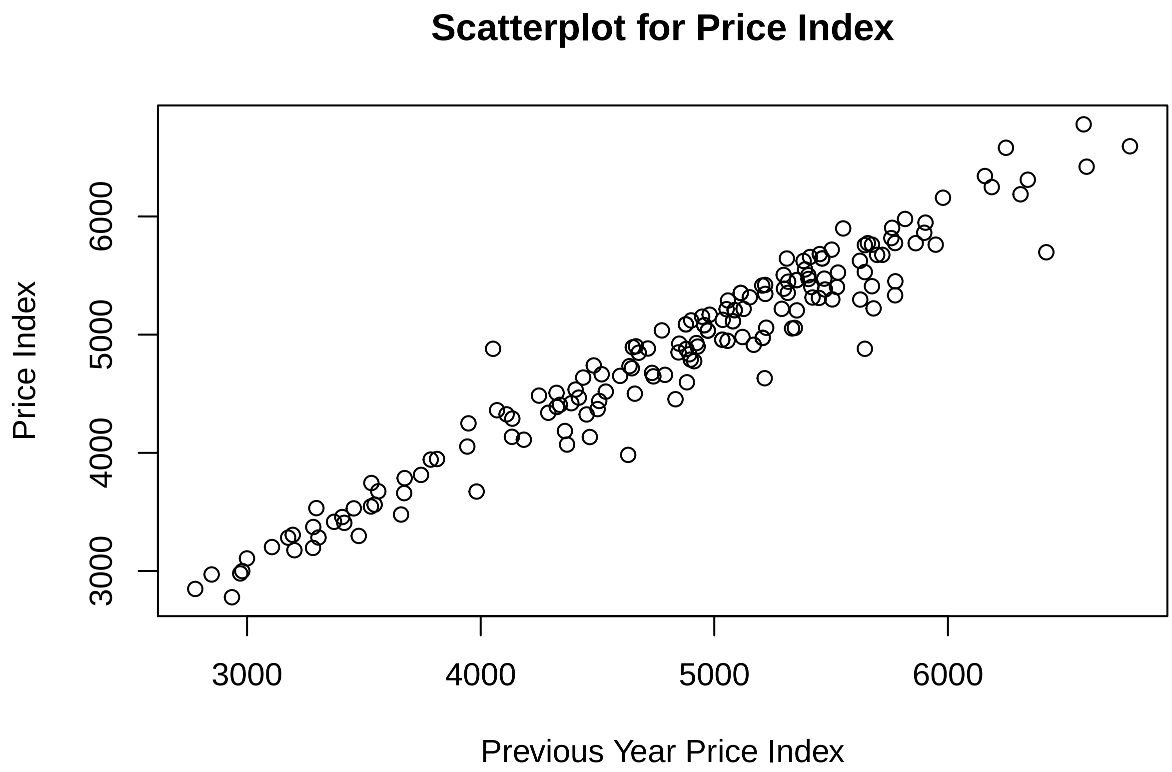


Inferences:

1) This time series appears to have an upward trend caused by the changing mean level with an upward drift early on and decreasing mean level post 2008. The level of the time series increases post 2004 and stabilises to a higher level compared to when it started from 2010 onwards.

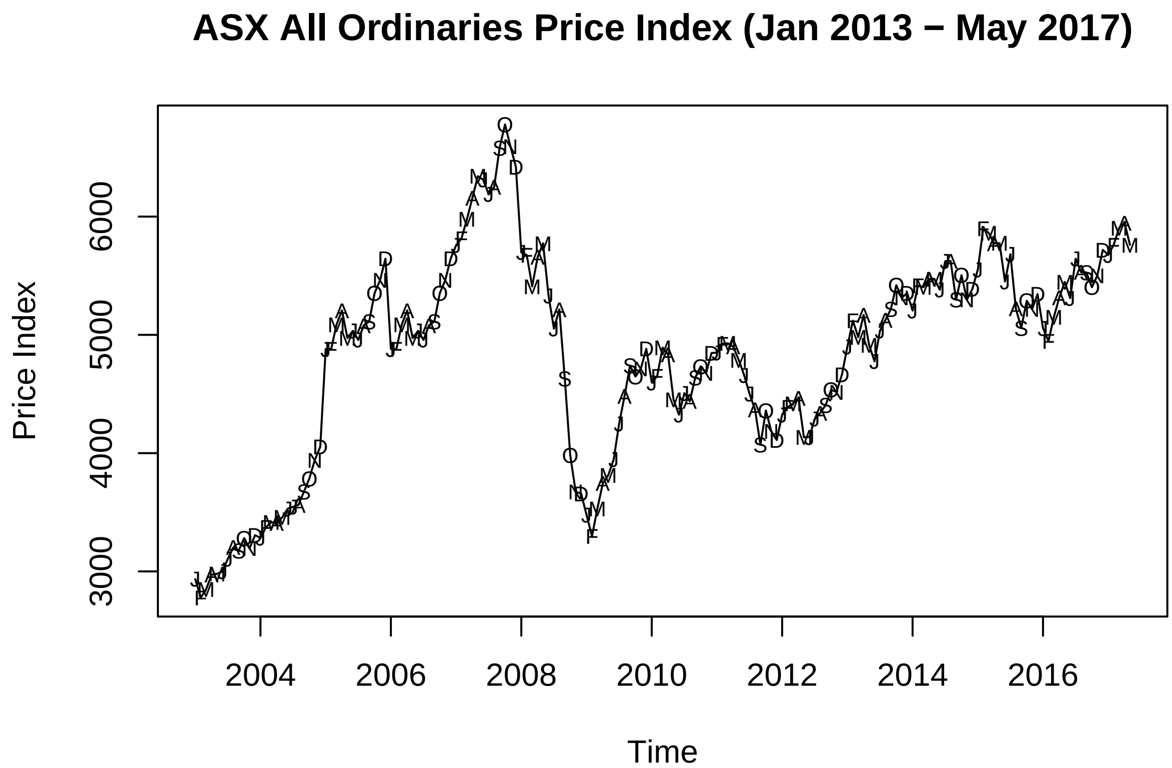
2) There seems to be no discernable seasonality in the series.
3) From the Mc Leod Li Test, hence we reject the null hypothesis stating the absence of conditional heteroscedasticity and there seems to be changing variance in the time series. Hence a transformation of the series is needed.

```r
#Scatterplot for Price Index
plot(y=asx.ts,x=zlag(asx.ts),ylab='Price Index',xlab='Previous Year Price Index', main="Scatterplot for
```

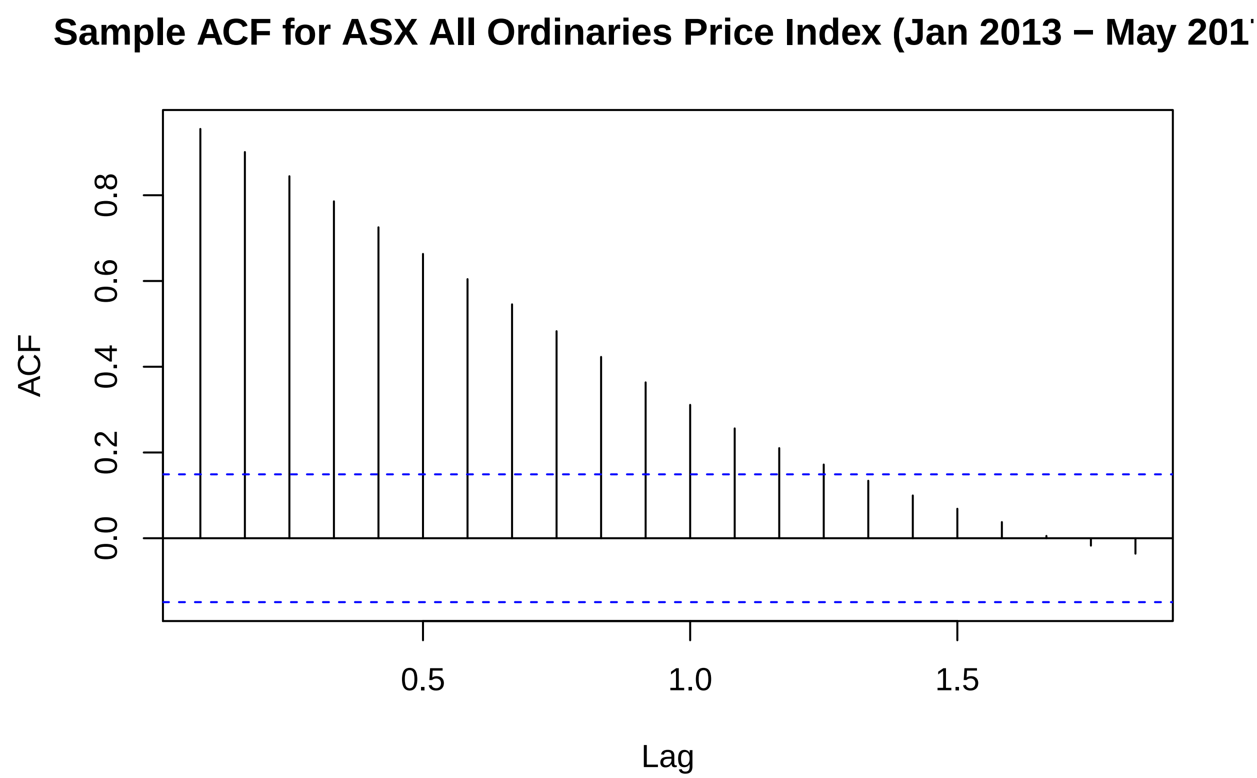


As seen from the scatterplot, there is an apparent upward trend - low values tend to be followed in the next year by low values, middle-sized values tend to be followed by middle-sized values, and high values tend to be followed by high values.

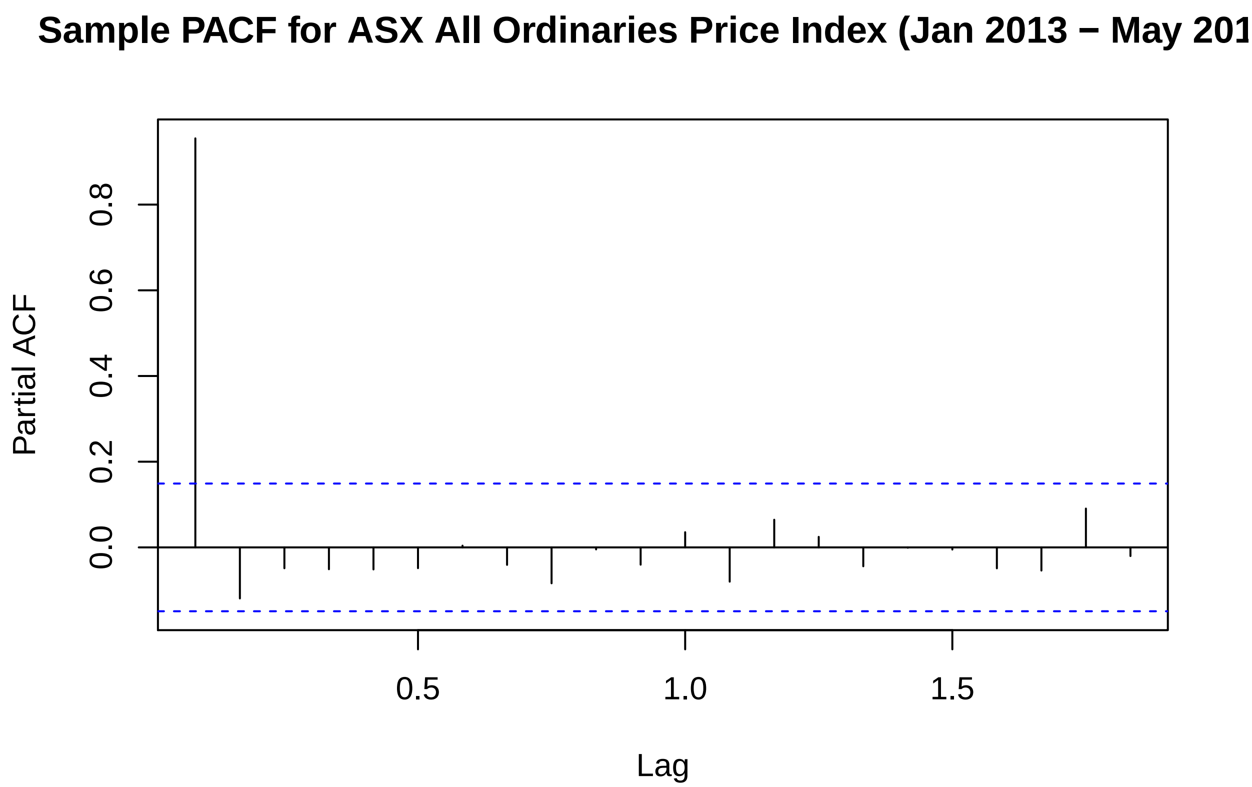
```
plot(asx.ts,ylab='Price Index',xlab='Time', main = "ASX All Ordinaries Price Index (Jan 2013 - May 2017)
points(y=asx.ts,x=time(asx.ts), pch=as.vector(season(asx.ts)), cex=0.7)
```




```r
acf(asx.ts, main="Sample ACF for ASX All Ordinaries Price Index (Jan 2013 - May 2017)")
```




```r
pacf(asx.ts, main="Sample PACF for ASX All Ordinaries Price Index (Jan 2013 - May 2017)")
```



Because we have a slowly decreasing pattern in ACF and a very high PACF value at the first lag, we infer that there is a trend and the series is non-stationary.The data is correlated at lag 1. To observe further serial correlation characteristics of these series we need to treat the trend and then display sample ACF and PACF again. There is no wave pattern in the ACF which further confirms that there is no seasonality.

Stationarity Identification

First we determine the lag length.

```
#KPSS test for stationarity
kpss.test(asx.ts) # p = 0.01, series is not stationary
```

```
## Warning in kpss.test(asx.ts): p-value smaller than printed p-value
```

```
##
##  KPSS Test for Level Stationarity
##
## data:  asx.ts
## KPSS Level = 1.0515, Truncation lag parameter = 3, p-value = 0.01
```

```
k = ar(asx.ts)$order
print(k)
```

```
## [1] 2
```

```
#Lag length is 2.
```

```r
adf.test(asx.ts, k = k)
```

```
##
##  Augmented Dickey-Fuller Test
##
## data:  asx.ts
## Dickey-Fuller = -2.4741, Lag order = 2, p-value = 0.3783
## alternative hypothesis: stationary
```

The p-value is more than default confidence level with alpha=0.05, hence we fail to reject the null hypothesis of non-stationarity.

A second approach to determine the value of lag length is carried out below:

```r
k = trunc(12*((length(asx.ts)/100)^(1/4)))
print(k)
```

```
## [1] 13
```

```r
adf.test(asx.ts, k = k)
```

```
##
##  Augmented Dickey-Fuller Test
##
## data:  asx.ts
## Dickey-Fuller = -2.7217, Lag order = 13, p-value = 0.2749
## alternative hypothesis: stationary
```

Third approach - default lag length

```r
adf.test(asx.ts)
```

```
##
##  Augmented Dickey-Fuller Test
##
## data:  asx.ts
## Dickey-Fuller = -2.7628, Lag order = 5, p-value = 0.2577
## alternative hypothesis: stationary
```

From all versions of the test, we can conclude that the series is nonstationary at 5% level of significance.

PP test

```r
PP.test(asx.ts, lshort = TRUE)
```

```
##
##  Phillips-Perron Unit Root Test
##
## data:  asx.ts
## Dickey-Fuller = -2.331, Truncation lag parameter = 4, p-value =
## 0.438
```

From the PP test, we conclude that the series is non-stationary with a p-value greater that 0.05 at 95% level of significance. This indicates that there is a higher order of autocorrelation. The test is robust with respect to unspecified autocorrelation and heteroscedasticity in the time series.

Detrending

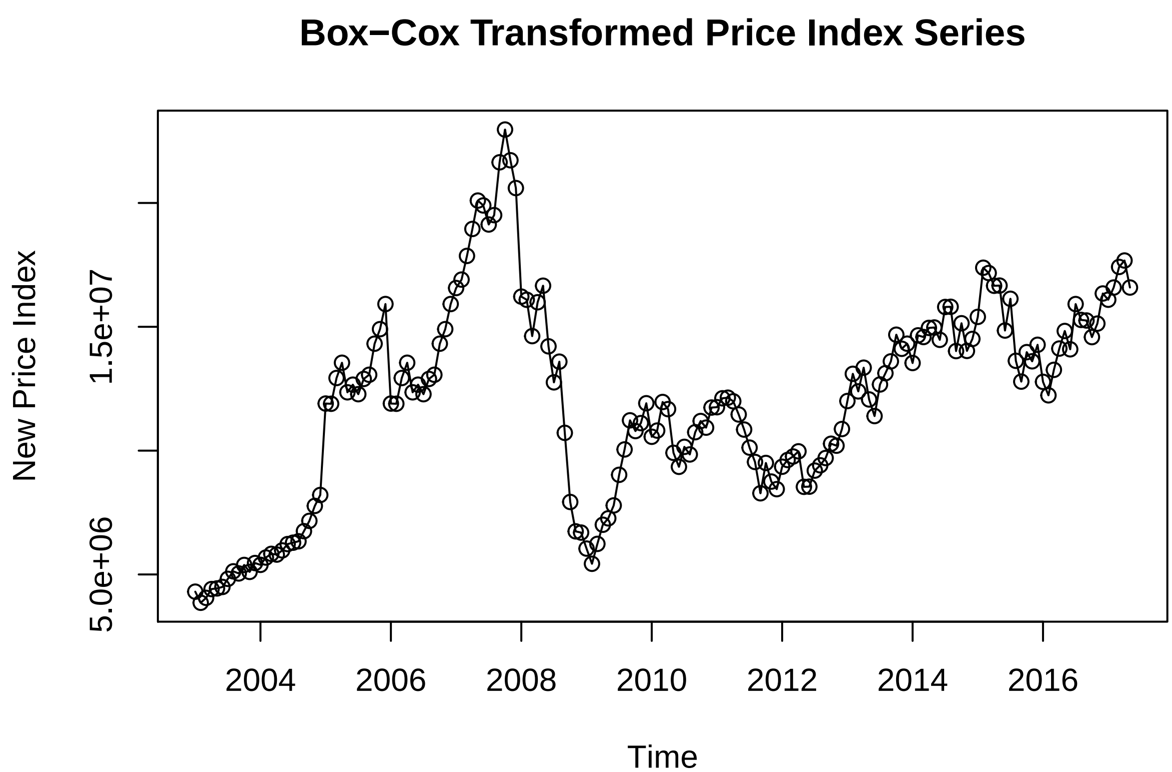
Since there exists a trend and changing variance, we transform the series first and then detrend the series using differencing to make the process stationary.

Detrending data using Box Cox transformation:

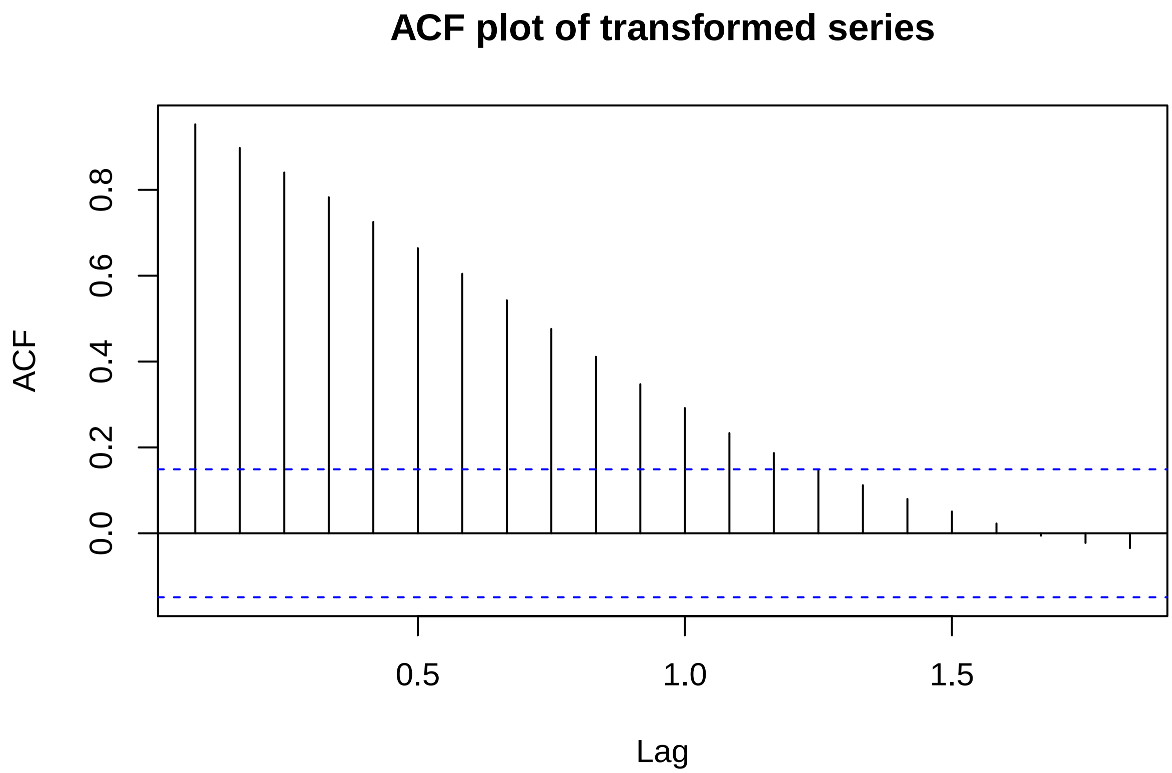
```
lambda = BoxCox.lambda(asx.ts)
lambda
```

```
## [1] 1.999924
```

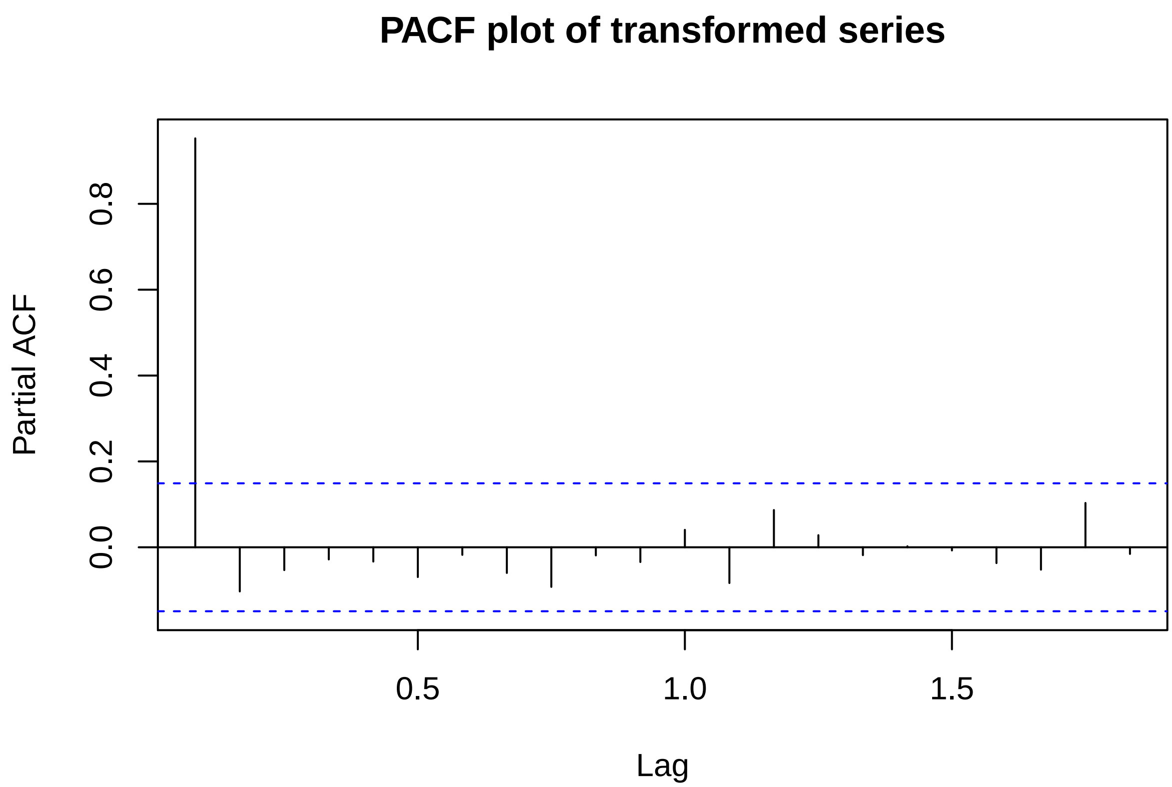
```
#lambda = 2

#Transforming the series using the value of lambda:

bc.asx.ts = BoxCox(asx.ts, lambda = lambda)
plot(bc.asx.ts,ylab='New Price Index',xlab='Time',type='o', main="Box-Cox Transformed Price Index Series
```




```
acf(bc.asx.ts, main="ACF plot of transformed series")
```



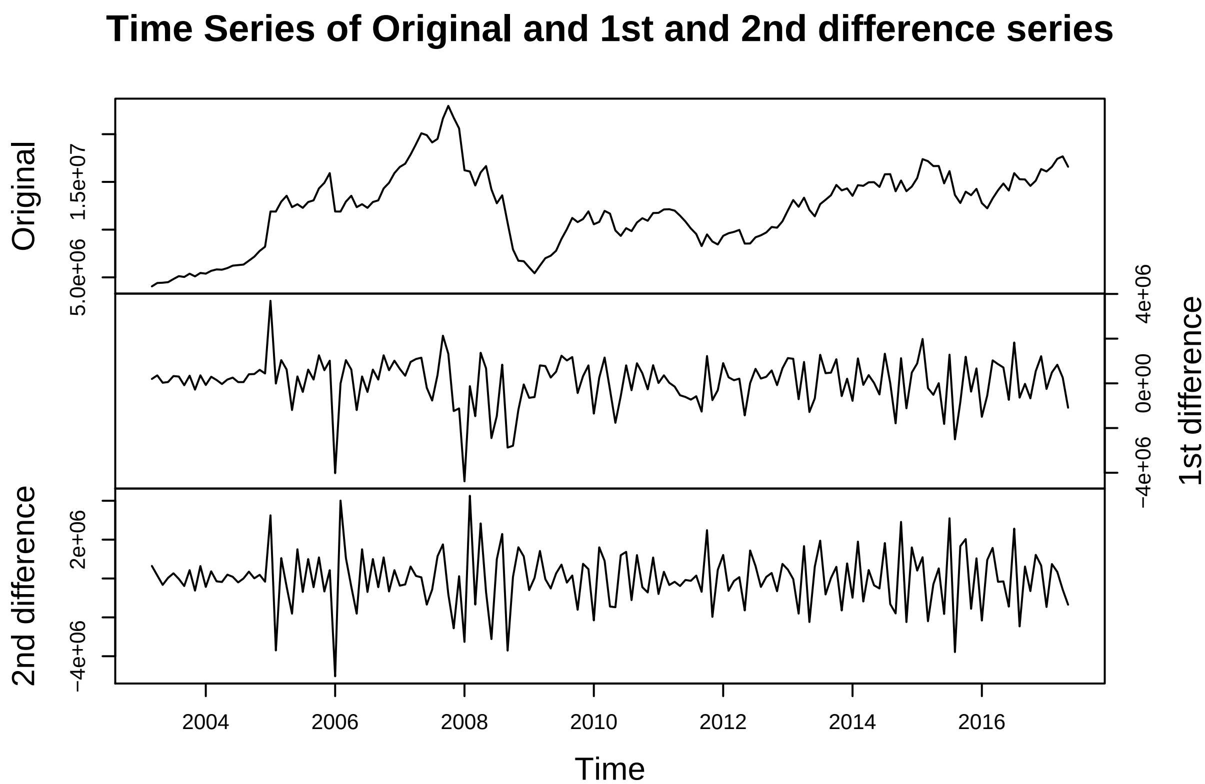

```
pacf(bc.asx.ts, main="PACF plot of transformed series")
```



The ACF and PACF plot of the transformed series still shows a trend exists and tranformation is not enough to make the data stationary. Hence we proceed to ordinary differencing of the non-stationary series.

Differencing

```
# Create an intersection of original, first and second difference of the series
asx2 = ts.intersect(bc.asx.ts , diff(bc.asx.ts) , diff(bc.asx.ts , differences = 2))
colnames(asx2) =  c("Original","1st difference","2nd difference")
plot(asx2 , yax.flip=T, main="Time Series of Original and 1st and 2nd difference series")
```



First difference appears to have detrended the series and made observations bounce around the mean level satisfactorily

```
#First Difference

diff1=diff(bc.asx.ts)
adf.test(diff1) #p-value = 0.01
```

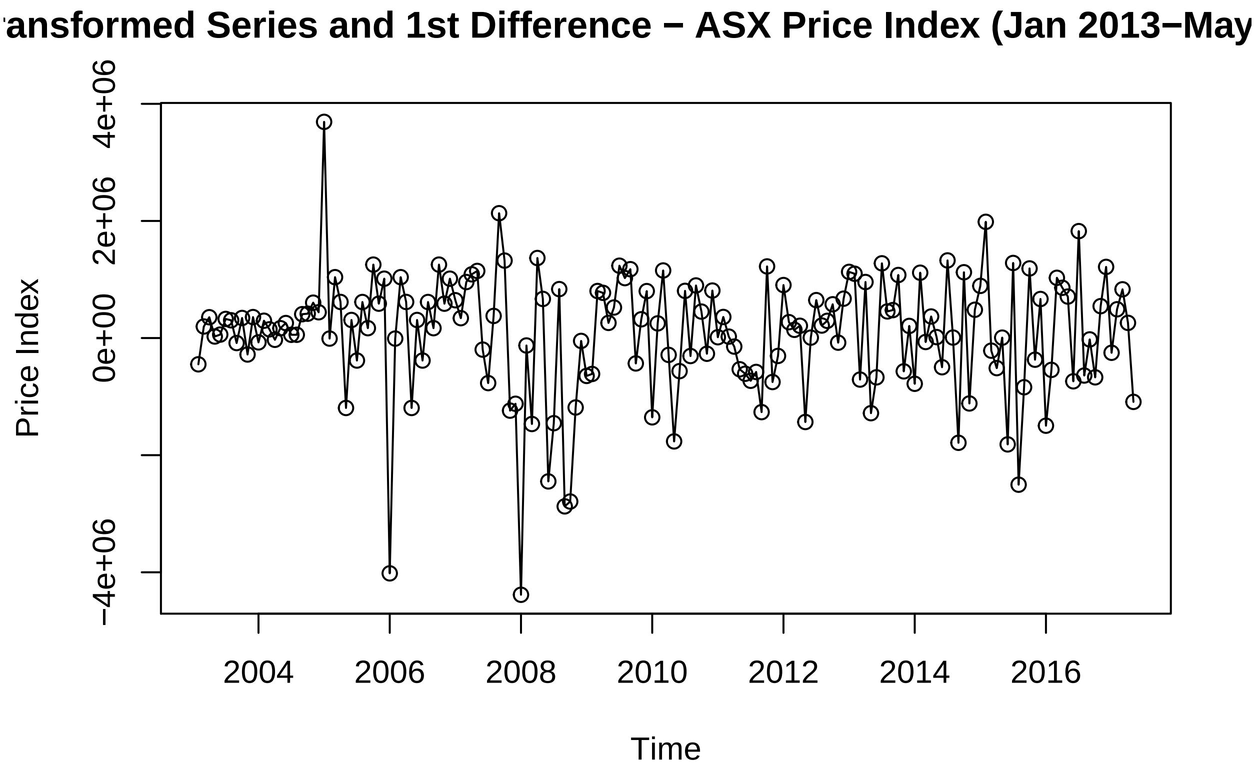
```
## Warning in adf.test(diff1): p-value smaller than printed p-value

##
##  Augmented Dickey-Fuller Test
##
## data:  diff1
## Dickey-Fuller = -4.6901, Lag order = 5, p-value = 0.01
## alternative hypothesis: stationary
```

```
PP.test(diff1, lshort = TRUE) #p-value = 0.01
```

```
##
##  Phillips-Perron Unit Root Test
##
## data:  diff1
## Dickey-Fuller = -12.223, Truncation lag parameter = 4, p-value =
## 0.01
```

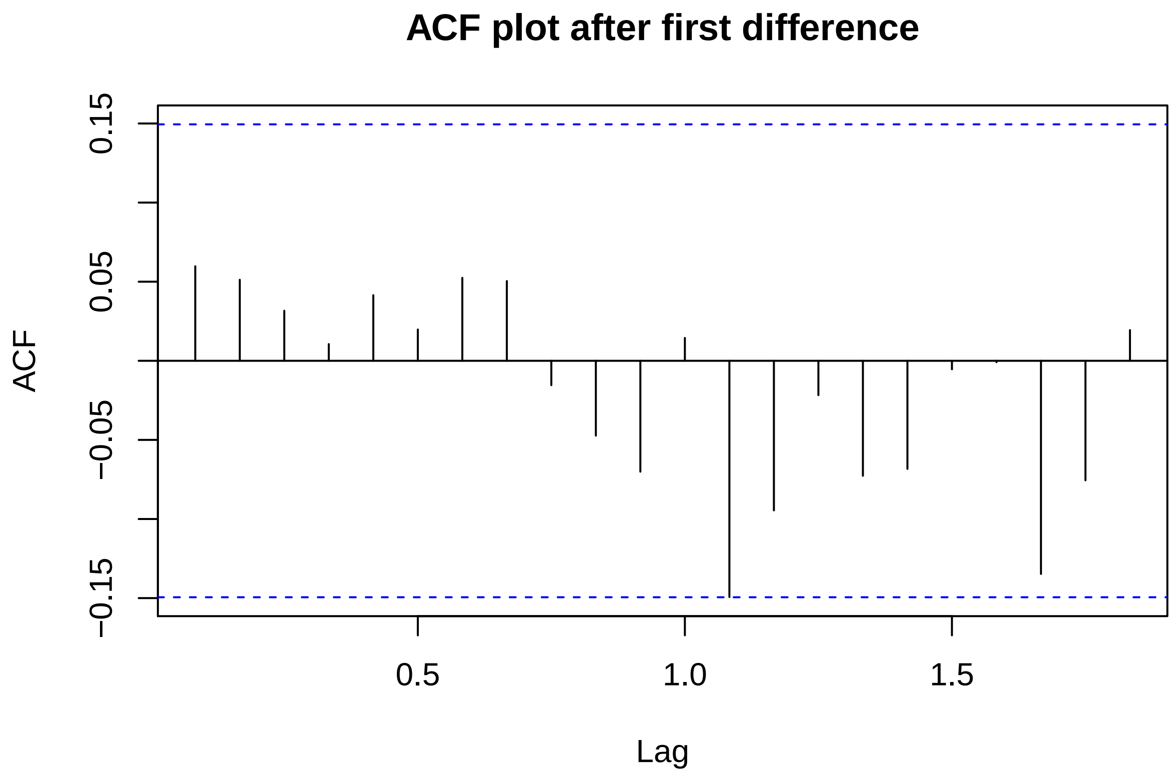
p-value is really small (lesser than 0.01) for both the tests and we reject the null hypothesis of non-stationarity. Hence data is stationary after 1st difference at 5% level of significance.

```r
plot(diff1,ylab='Price Index',xlab='Time',type='o', main = "Transformed Series and 1st Difference - ASX
```

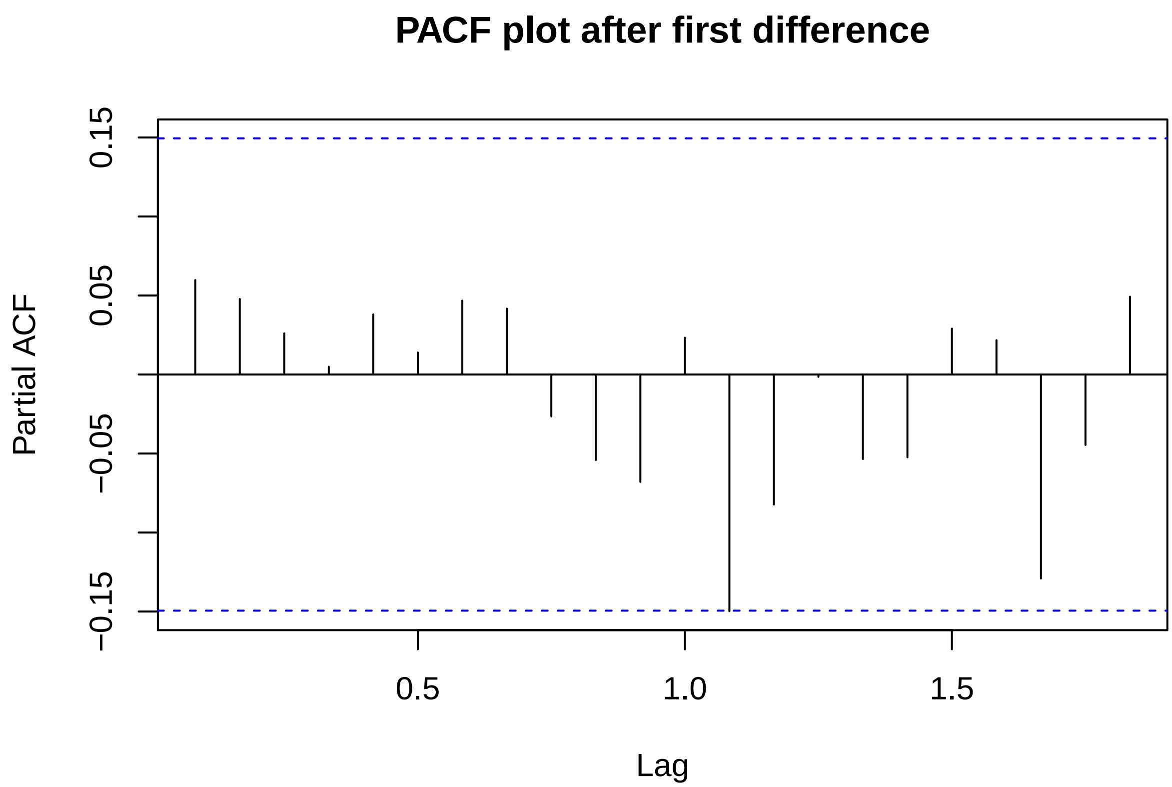


As seen from the plot, the series is more centered around the mean and appears stationary after the first difference

```r
#ACF and PACF of transformed data

acf(diff1, main="ACF plot after first difference")
```




```r
pacf(diff1, main="PACF plot after first difference")
```



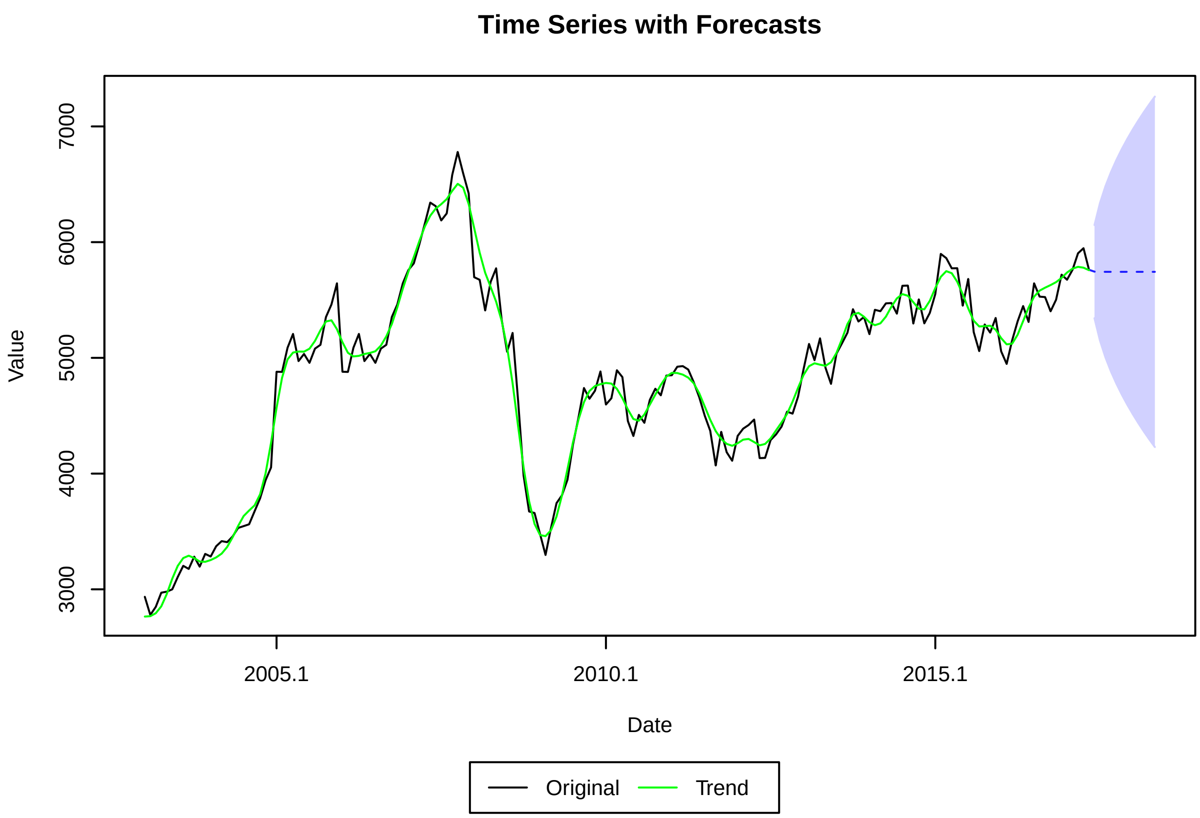
As can be seen from PACF and ACF there are no significant lags and the data is stationary now

Decomposition of the Time Series

Now we will apply decomposition methods to separate the effect of trend and seasonality. This will let us infer the patterns of trend and seasonality separately.

a. x-12 method

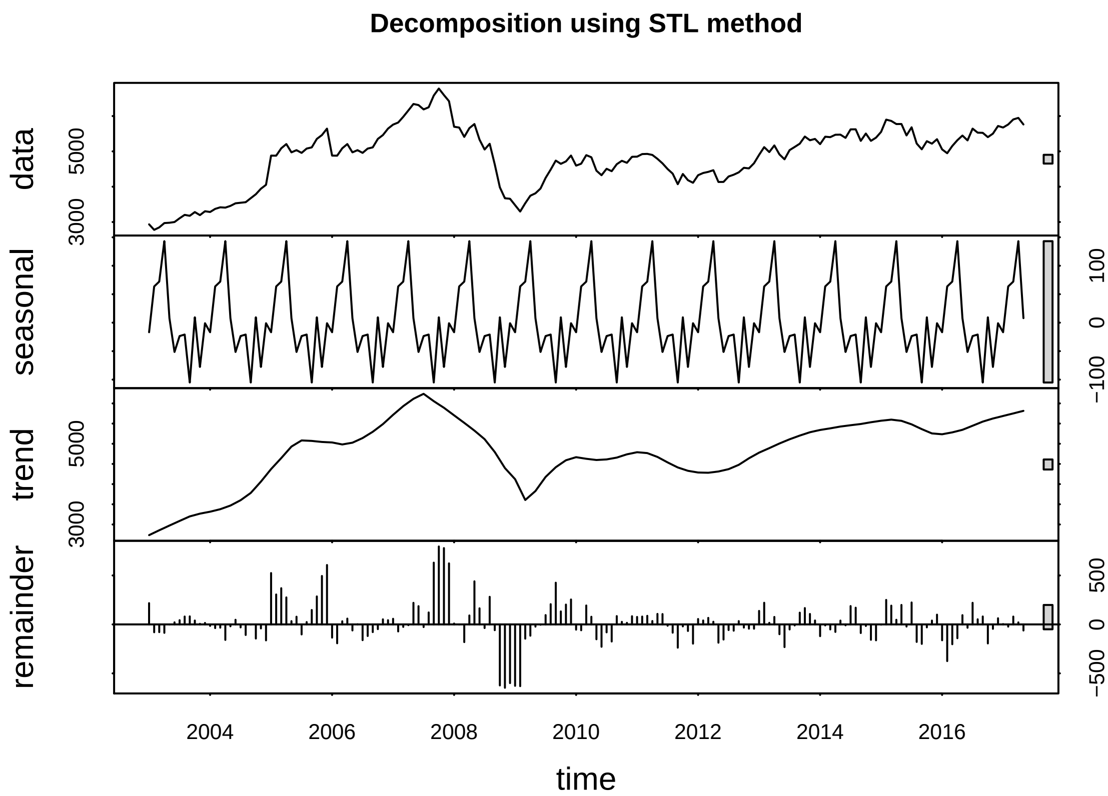
```
fit.x12 = x12(asx.ts)
plot(fit.x12 , sa=FALSE , trend=TRUE, forecast=TRUE, main="Decomposition plot using x12 method")
```



There is an observable trend effect and we observe some deviances from the trend component in the seasonally adjusted series.

b. STL method

```
fit.asx <- stl(window(asx.ts,start=c(2003,1), end =c(2017,5)), t.window=15, s.window="periodic", robust=
plot(fit.asx, main="Decomposition using STL method")
```

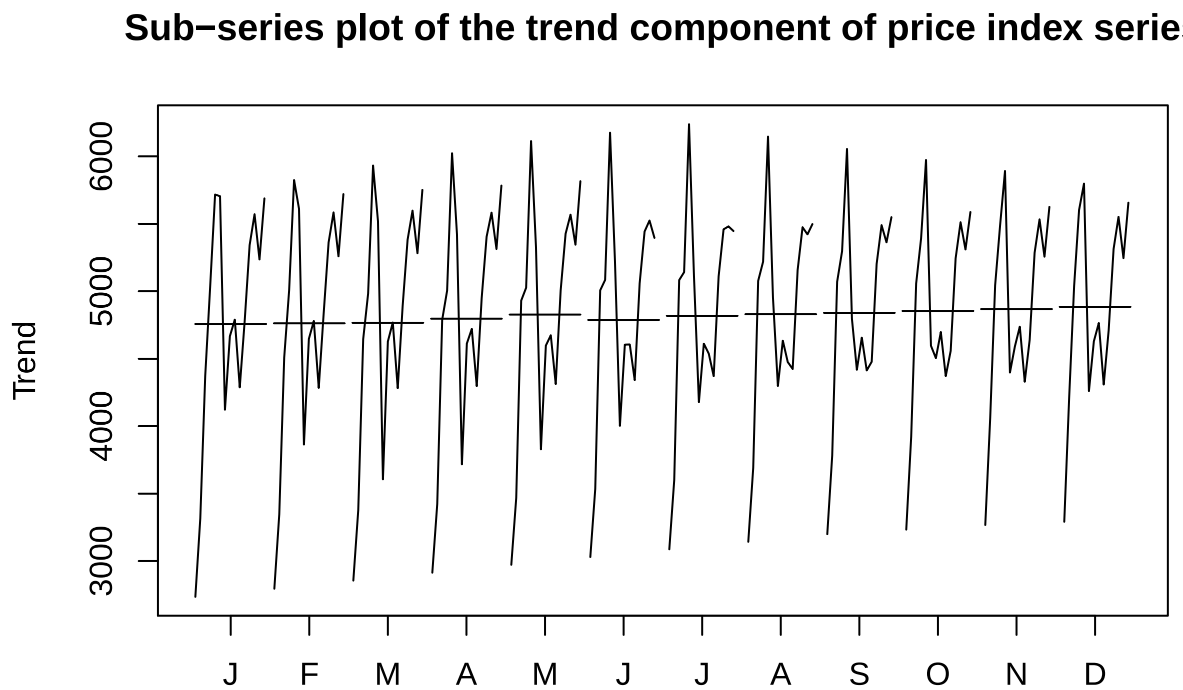


Inferences:

1) In the remainder part, there are some significant fluctuations and changing variance is present.
2) Since we have ruled out existence of seasonality, we dig deeper into the trend component. The trend effect shows increasing variance and then the trend smoothens out 2010 onwards.

Monthly trend component:

```
monthplot(fit.asx,choice = "trend", main="Sub-series plot of the trend component of price index series"
```

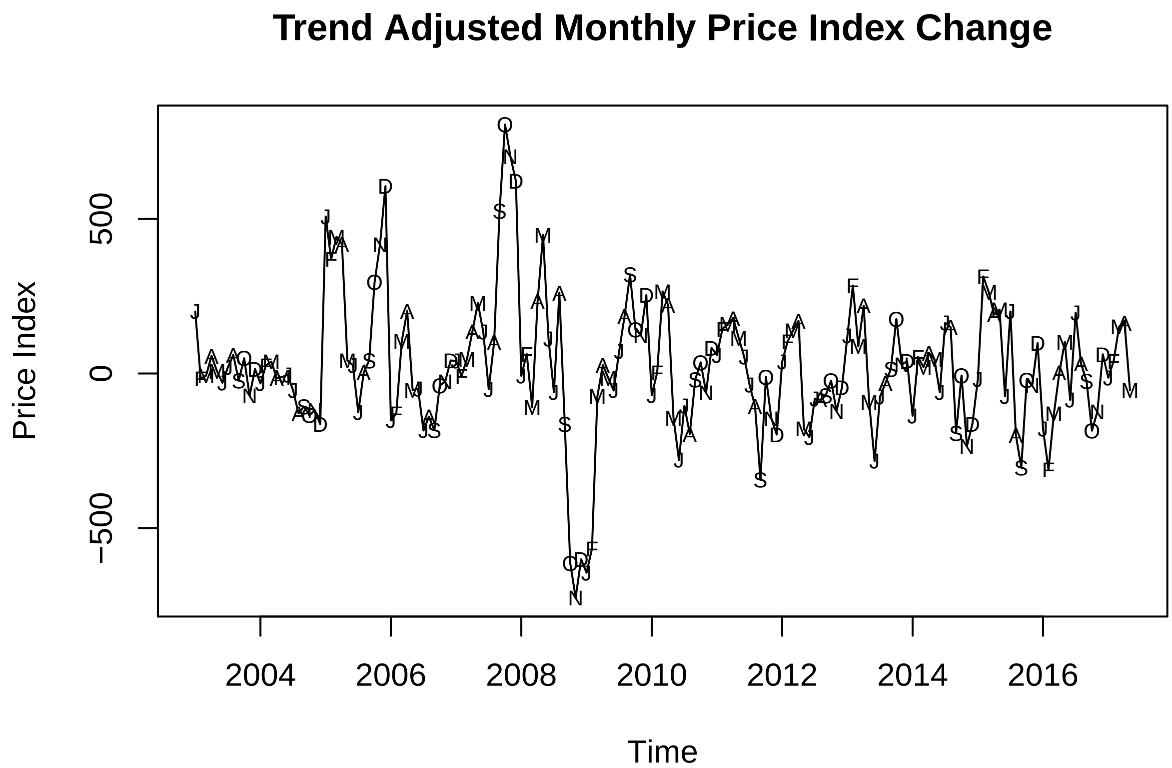


Inferences:

The mean levels are fluctuating in the plot and we infer that there is a trend effect.

After the decomposition, we can adjust for trend as well. For this, we need to subtract the trend component from the original series.

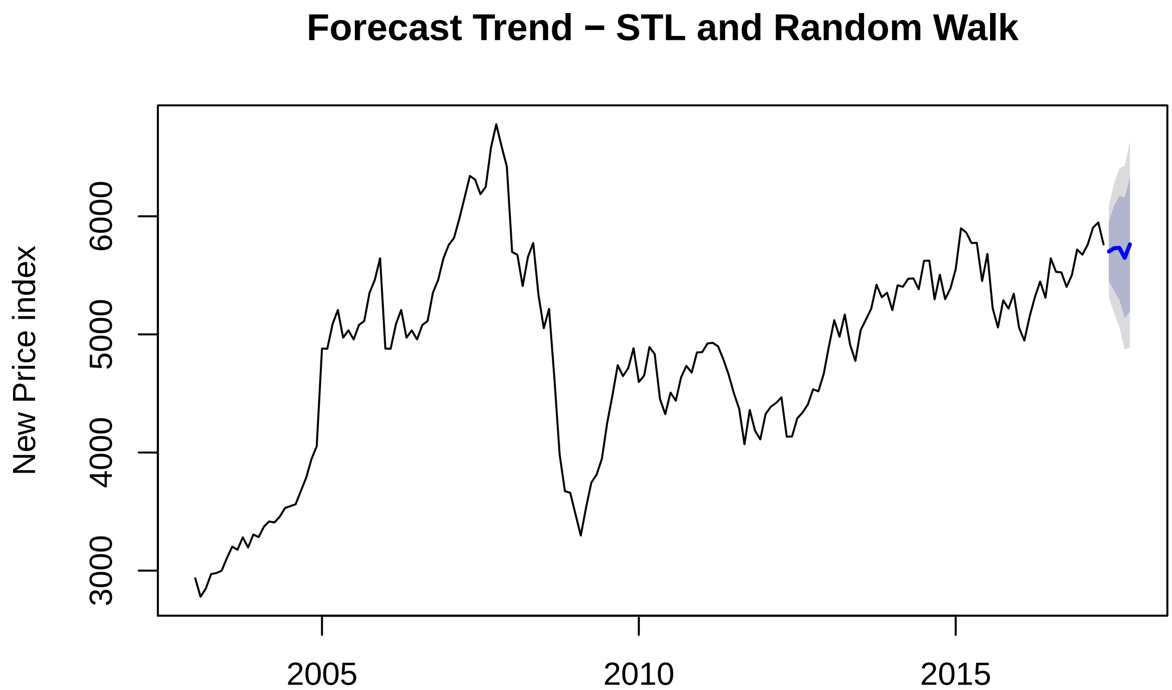
```
fit.asx.trend = fit.asx$time.series[,"trend"] # Extract the trend component from the output
asx.trend.adjusted = asx.ts - fit.asx.trend
plot(asx.trend.adjusted, ylab='Price Index',xlab='Time', main = "Trend Adjusted Monthly Price Index Char
points(y=asx.trend.adjusted,x=time(asx.trend.adjusted),pch=as.vector(season(asx.trend.adjusted)),cex=0.
```



We can see that the series has been detrended with the data fluctuating around the mean level. This results in filtering the effect of trend from the original series.

Forecasting with Decomposition using Naive Method

```
forecasts <- forecast(fit.asx, method="naive", h=5)
plot(forecasts, ylab="New Price index", main="Forecast Trend - STL and Random Walk")
```



As can be seen from forecasting the values of the next 5 months using the Naive method on the STL decomposed time series, the following values are obtained:

```
forecasts
```

```
##          Point Forecast    Lo 80    Hi 80    Lo 95    Hi 95
## Jun 2017       5701.956 5447.356 5956.556 5312.579 6091.333
## Jul 2017       5729.848 5369.790 6089.907 5179.186 6280.510
## Aug 2017       5732.540 5291.560 6173.520 5058.120 6406.960
## Sep 2017       5648.289 5139.089 6157.489 4869.536 6427.042
## Oct 2017       5762.766 5193.464 6332.069 4892.094 6633.439
```

References

1) AlignAlytics
2) Time Series Analysis With Applications in R - Jonathan D. Cryer and Kung-Sik Chan, 2nd Ed
3) Forecasting with Exponential Smoothing - The State Space Approach - Rob J. Hyndman, Anne B. Koehler, J. Keith Ord and Ralph D. Snyder