Time Series Case Study

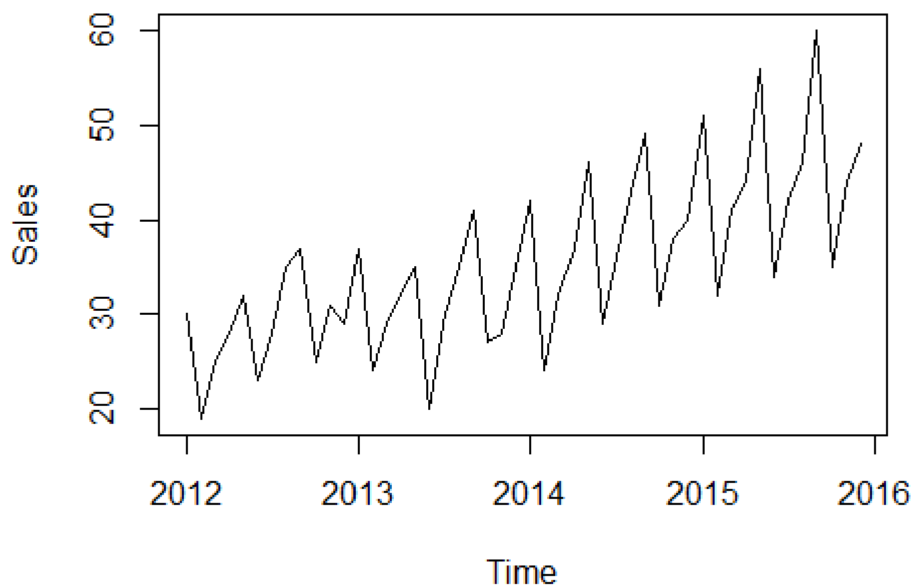
proschool
An IMS Initiative

Contents

Data

Load the dataset into R and plot the data series. It is always good to plot the dataset as a first step so that we could get have some understanding of the components in the series. Convert the series into a time series data. Since the series is a monthly data starting from 2012 we set the frequency to 12 and also set the start period for the series.

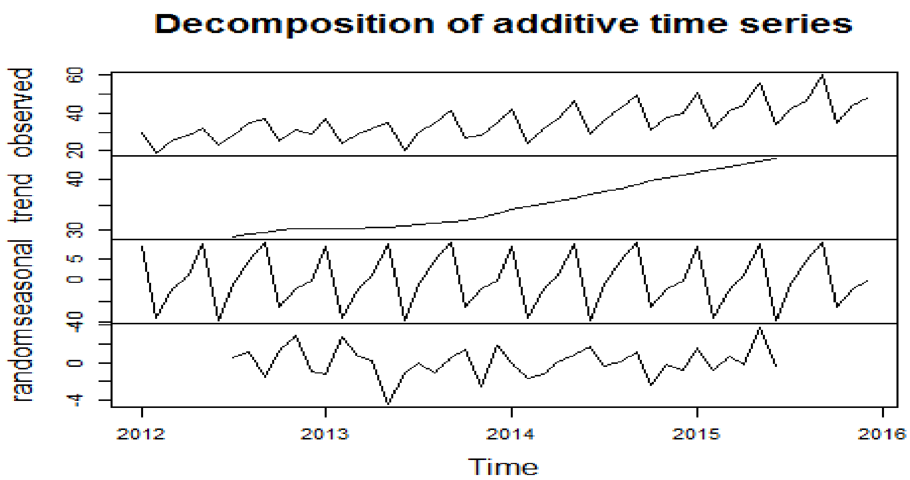
```
SalesData = read.csv("Sales.csv")
SalesDataTS = ts(SalesData, frequency=12, start=c(2012,1))
plot(SalesDataTS)
```



Decomposing the series

To help us find the components in the series we could use the decompose function. Plot the object obtained after applying decompose.

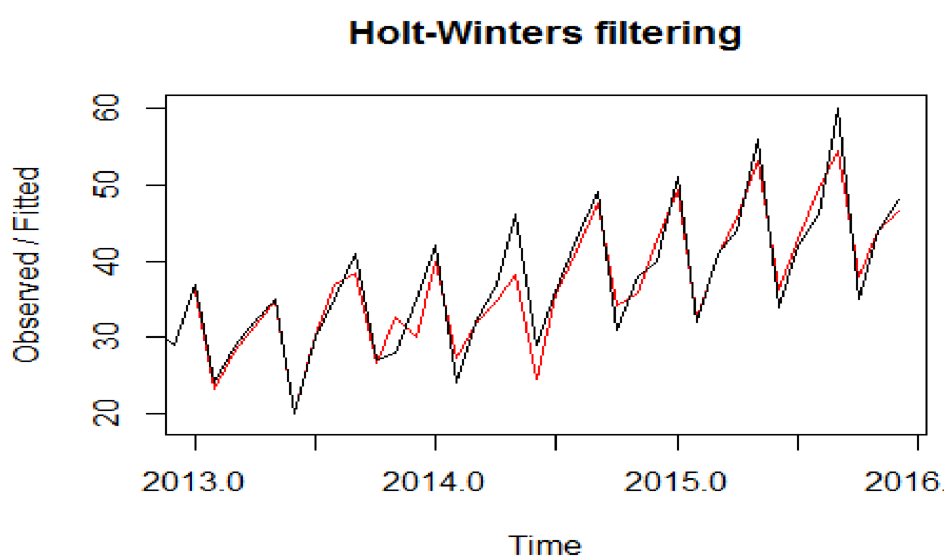
```
Decomp = decompose(SalesDataTS)
plot(Decomp)
```



The plot clearly indicates that there is level, trend and seasonal components in the sales. Hence we would go for HoltWinters method.

Modelling

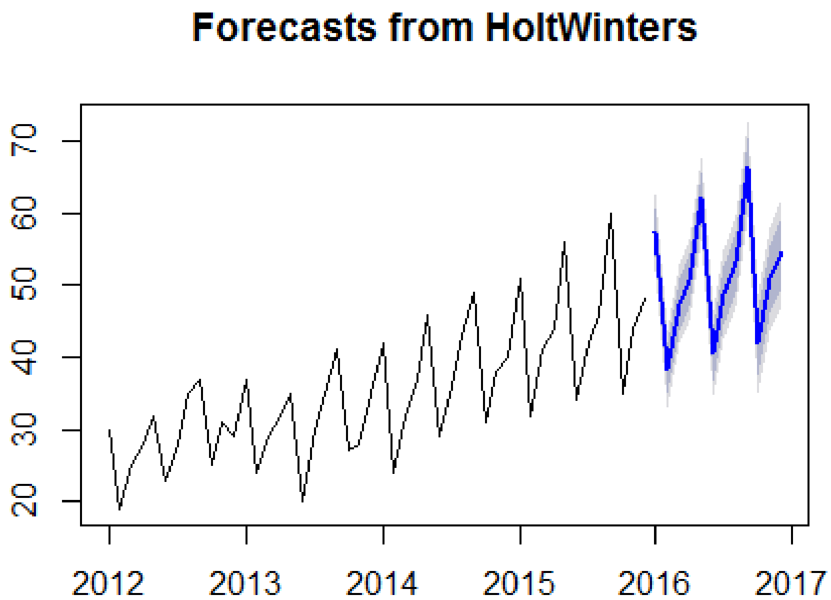
```
SalesDataHW =  HoltWinters(SalesDataTS)
plot(SalesDataHW)
```



The plot suggests that HoltWinters was able to give us decent fit indicating that we could use the model for further forecasting. For forecasting we would use the forecast.HoltWinters function and set the forecast period to 12 since we need to forecast for the next year.

Forecasting

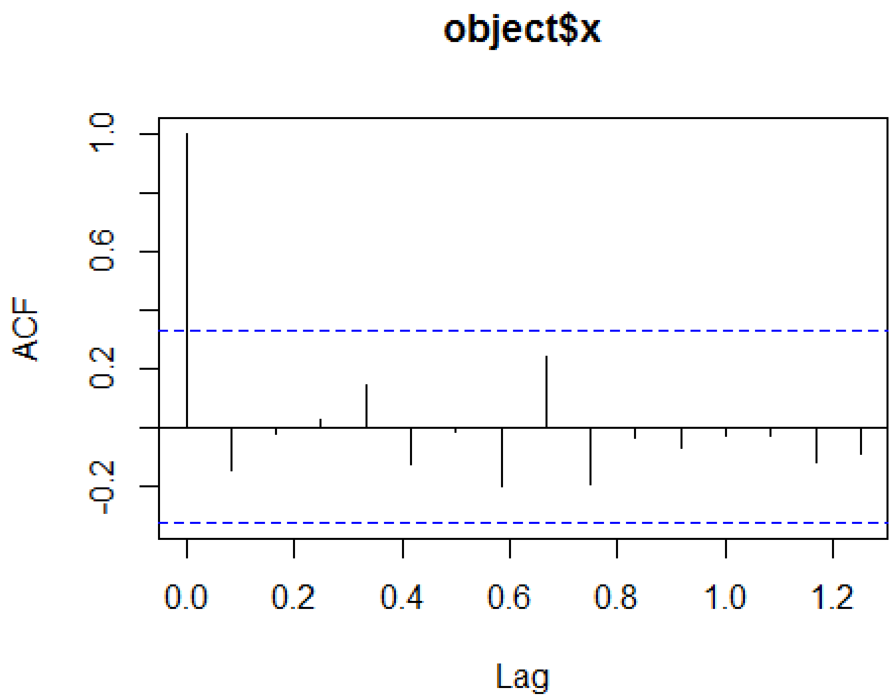
```
library(forecast)

SalesDataFC = forecast.HoltWinters(SalesDataHW, h=12)
plot(SalesDataFC)
```



Diagnostics

Finally we need to perform the fit diagnostics which would help in checking for the assumptions.

```
acf(SalesDataFC$residuals)
```

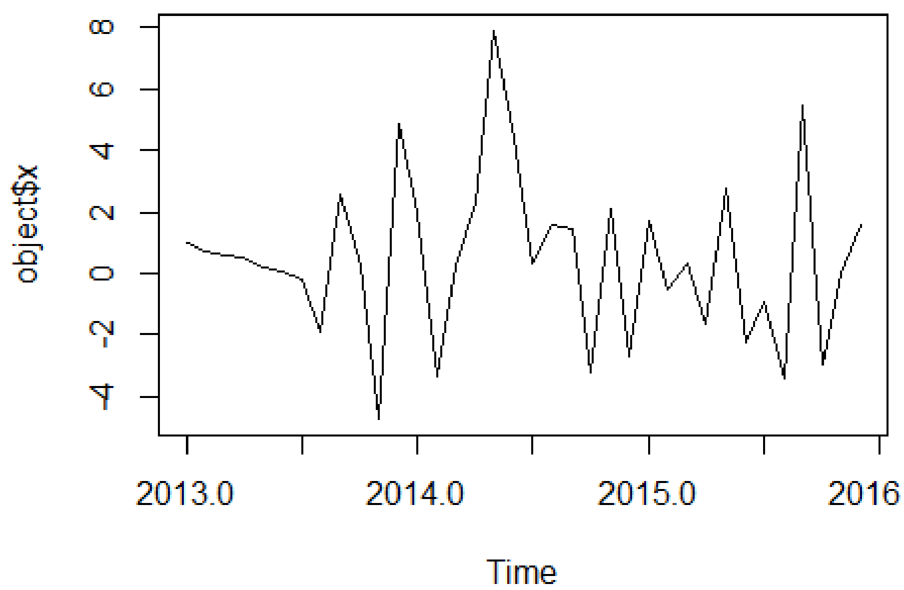


Since the acf plot indicates that the lags are within significance bounds indicating the errors are not autocorrelated.

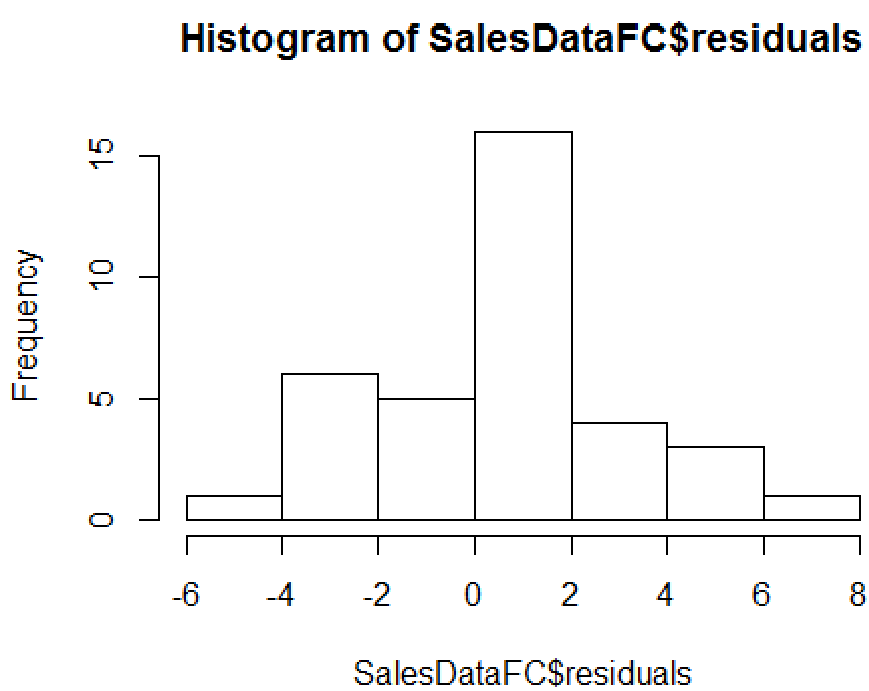
```
Box.test(SalesDataFC$residuals, lag=20, type="Ljung-Box")

##
##  Box-Ljung test
##
## data:  SalesDataFC$residuals
## X-squared = 13.854, df = 20, p-value = 0.8378
```

Since the p-value for Ljung-Box test is 0.8378, indicating that there is little evidence of non-zero autocorrelations at lags 1-20

```
plot(SalesDataFC$residuals)
```




```
hist(SalesDataFC$residuals)
```



The line plot indicates that the errors are randomly distributed and the histogram indicates that mean of the error distribution is close to 0.