Prajwal Jagadish

Matriculation number: 31-04-067-20-2

Course of study: Msc in business intelligence and data science

Module: Business analytics

Topic: Simulation forecasting

Abstract:

Simulation forecasting is a wide study which leads to make assumptions on forecast and is validated with observable outcomes. Sales forecasting is one of a kind with respect to forecasting. The summary of sales forecasting is based on item-data and predicts the future sales with respect to quantities or sales value. In methodology, different forecasting methods are used to predict quarterly sales and evaluate performance based accuracy measures of each model. The sales forecasting is made for the betterment of future to predict prescribed outcomes. This process can help in making decisions for a particular company according to the outcomes provided by the forecast.

Introduction:

Forecasting is the process of estimating or predicting based on present or historic data, the analysis of forecasting is commonly based on trends. A forecast is successful, if the result explains various parameters like trend analysis, seasonality, or exponential smoothing. The forecasts are made to analyze the prediction and provide effective insights to implement better decisions for companies and to solve environmental issues. The following are the methods which are most commonly used to evaluate the forecast:

(i) Trend (linear or straight line): In this method, forecasting technique is based on linear regression of time series forecasting. This gives the best forecasting reliability when major factors influence in measuring forecast.

(ii) Growth (curved or curved line): Exponential regression technique of time series is used in growth forecasting methods. This gives the best forecasting reliability when major factors influence in measuring exponentially.

(iii) Auto regression (seasonal): The auto-correlation function is used in autoregression forecasting. This forecasting technique detects linear, non linear and seasonal fluctuations from historic data that tends to forecast trends into the future.

Forecasting is categorized into two different methods:

(i) Quantitative methods: To forecast demand, Quantitative methods uses mathematical models that rely on historical data, utilizes large amount of data and figures for effective integration.

(ii) Qualitative methods: This approach uses factors based on the expert's knowledge, experience, intuition, value systems and so on.

Therefore, forecasting regulates the comparison, evaluating with actual measures and review for modifications to ensure smooth predictions.

In organization activities time series forecasting have been widely used. At the organizational level many decisions of forecast is influenced. Sales forecasting measures product sales which is likely to be sold in a specified future period with influencing factors like price and market fluctuations. Accurate sales forecasting enables an organization to make further decisions like requirements for raw materials,

labour, price, and so on. Inaccurate forecast leads to opportunity loss, losses in cost of goods return. Forecasting sales is a global industrial practice, in which establishing objectives, implementing action-plans (budget, resources) and allocating them. Therefore, effective sales forecasting helps a company to take better decisions and implementing right strategies at right time.

Objectives and goals

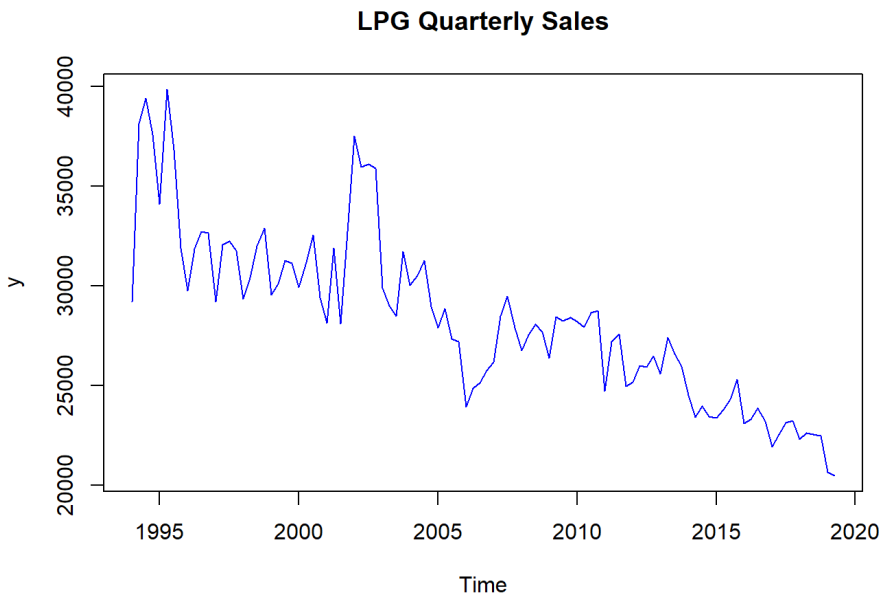
The objective of the paper is to evaluate performance from different methods used in forecasting quarterly sales. The different methods used are mean method, naïve method, seasonal naïve method, linear regression model, Arima models, seasonal decomposition and simple exponential smoothing. The performance of each method is measured by Relative squared error, Mean absolute scaled error, Root mean squared error, and Mean absolute percentage error. These error values define how well each method performs in forecasting sales. These forecasts can help in deciding policies, sales budget, production budget, extent of advertising, and guiding other business activities to achieve targets. However, with regard to evaluating performance and comparing each method fulfills the objective of forecasting quarterly sales.

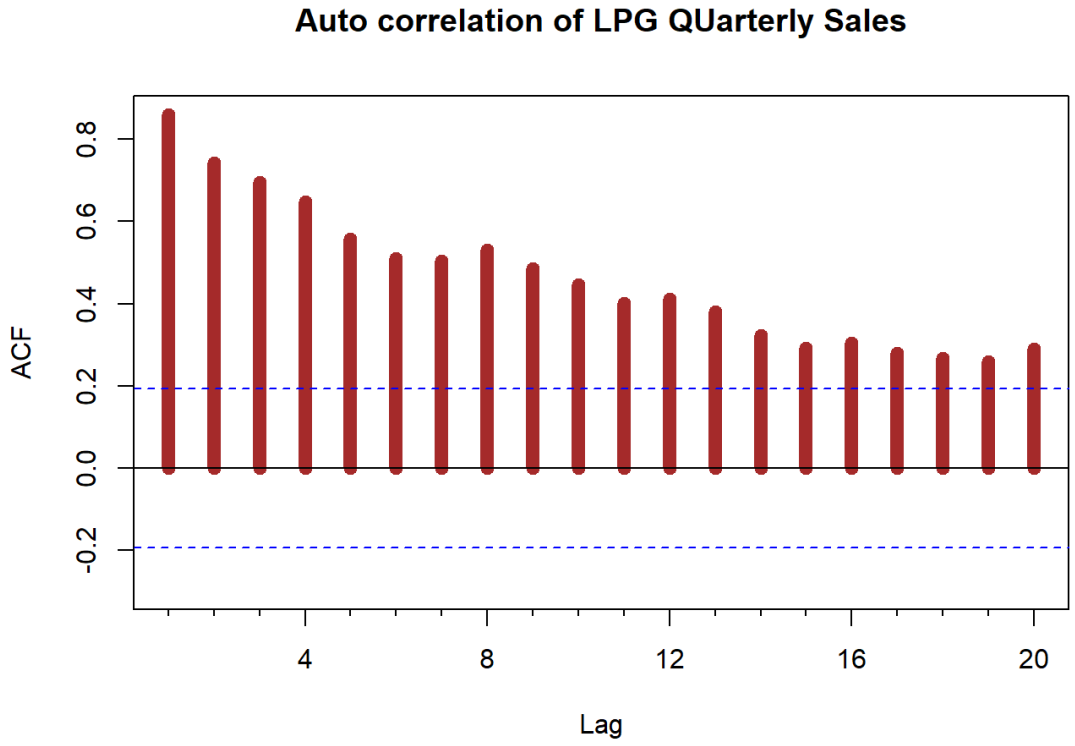
Methodology:

The data set is obtained via web search and contains quarterly sales data. The product is liquefied petroleum gas from Indian Oil Corporation. The data set contains 103 observations of sales data from 1994 to 2019.

The sales data is assigned into a data frame. The data frame is partitioned into training and testing data. In partitioning, function window() is used to set start and end of the time period given.

The below diagram explains about the quarterly sales about LPG. Initially the sales of LPG have strong growing trend, the rest of the period is diminishing except in the year 2002 and 2003. Therefore, the data exploratory stage is to analyze the pattern and to understand the trend and seasonality.

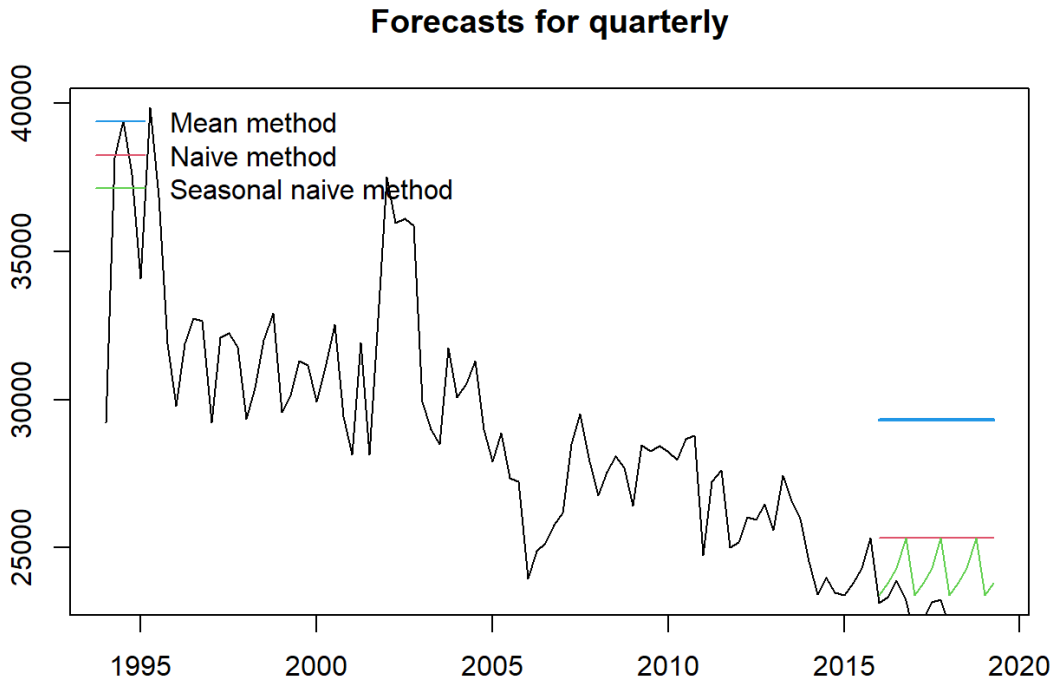

LPG Quarterly Sales
y
20000
25000
30000
35000
40000
1995
2000
2005
2010
2015
2020
Time


Auto correlation of LPG QUarterly Sales
ACF
-0.2
0.0
0.2
0.4
0.6
0.8
4
8
12
16
20
Lag

The Auto correlation function is used to analyze the seasonal trends is significant or not. The above graph indicates good seasonal patterns as it is greater than the blue dotted lines. This means the trend pattern shows there is increase in sales in every quarter when compared to neigbhours, even though the trend pattern is diminishing when the lag increases.

There are different methods in forecasting and the following are evaluated with error assessment:

1. The first method used in evaluating forecasting for quarterly sales is Mean method, naïve method and seasonal naïve method.
   - **The mean method:** In a time series, assuming every data is equally useful to predict all future values. If Trend and seasonality components are not present in the time series, this method holds significantly good. In mean method, it represents with single number (point forecast).
   - **Naïve method:** This method is the simplest method of forecasting as this also does not contain T and S components. With regard to daily basis or weekly basis this method is more effective.
   - **Seasonal naïve method:** This method accounts for seasonality by setting each prediction value to be equal to the last observed value of the same season.



- From the above graph, the package 'forecast' is used to forecast model in Mean method, naïve method, and seasonal naïve method.

Forecast point (blue line) is the mean of the probability distribution that represents an estimate of the average number of contributions. The forecasted test data of mean and naïve methods do not consider trend and seasonality. Seasonal naïve method describes there is slight increase of sales in first quarter for following three years as it follows season patterns with regard to forecasting the test data. Therefore, the gap of mean method is increasing when the quarterly sales decreases.

2. Linear trend method:

This method performs predictive analysis that examines regression estimates. In other words, regression estimates the relationship between one dependent variable and on or more independent variable.

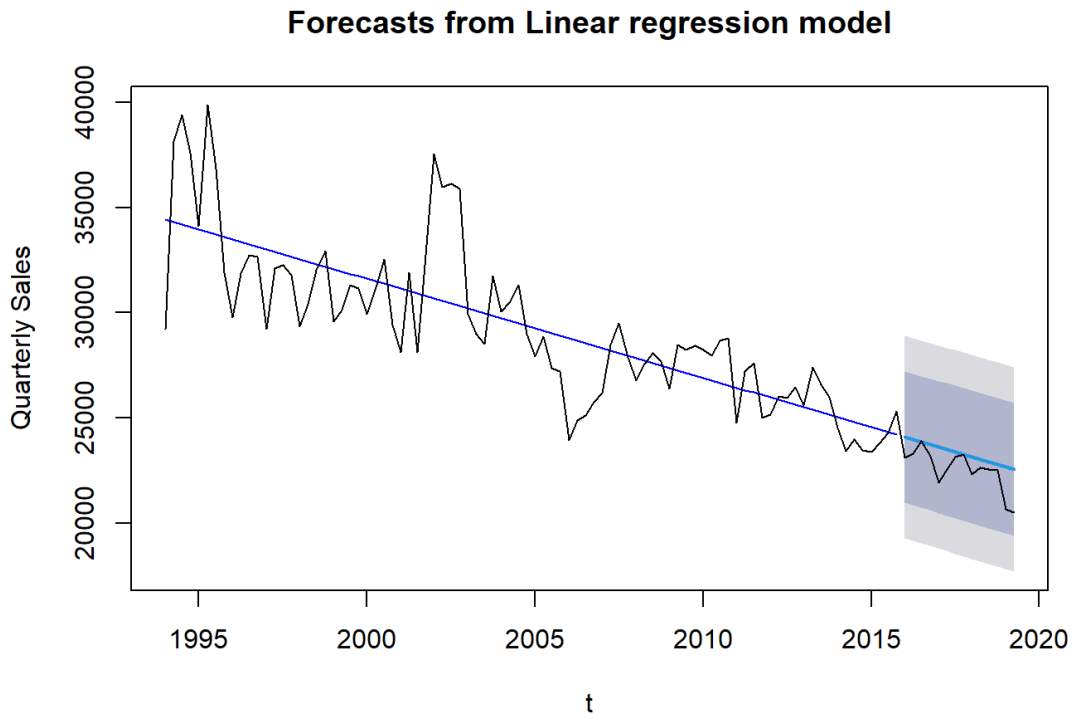
The formula for linear regression is:

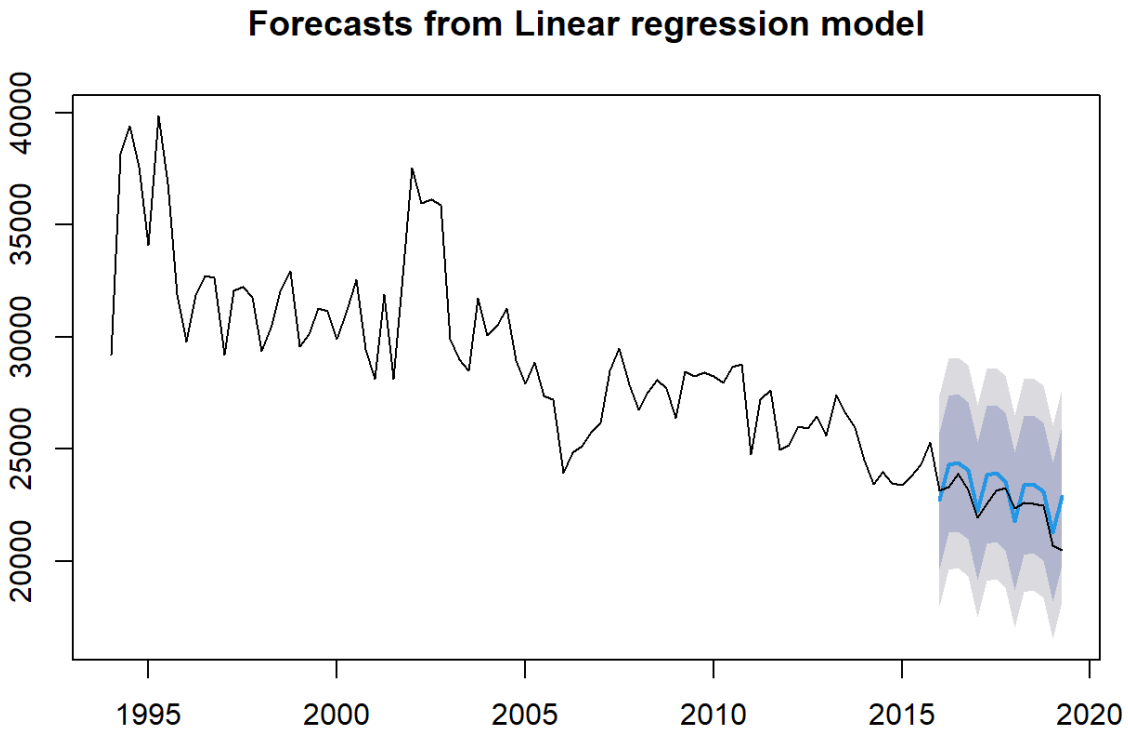
$Y = mx + c$

Where x- score of the independent variable

m- Regression coefficient

c: constant





In linear trend method, the straight line is perfectly drawn when the trends are decreasing. Here, the predictions of quarterly sales from the year 2016 are below the line. From the seasonal trends it is almost overlapping with test data. There is a slight increase in every quarter of the sales in test data with respect to seasonality.

From the below output of the summary() function describes the forecast quarterly sales based linear trend method. The Mape error is 5.77. This determines how well the model is suited to predict the forecast. Mape is 5% which means that the average difference between the forecasted value and the actual value is 5%.

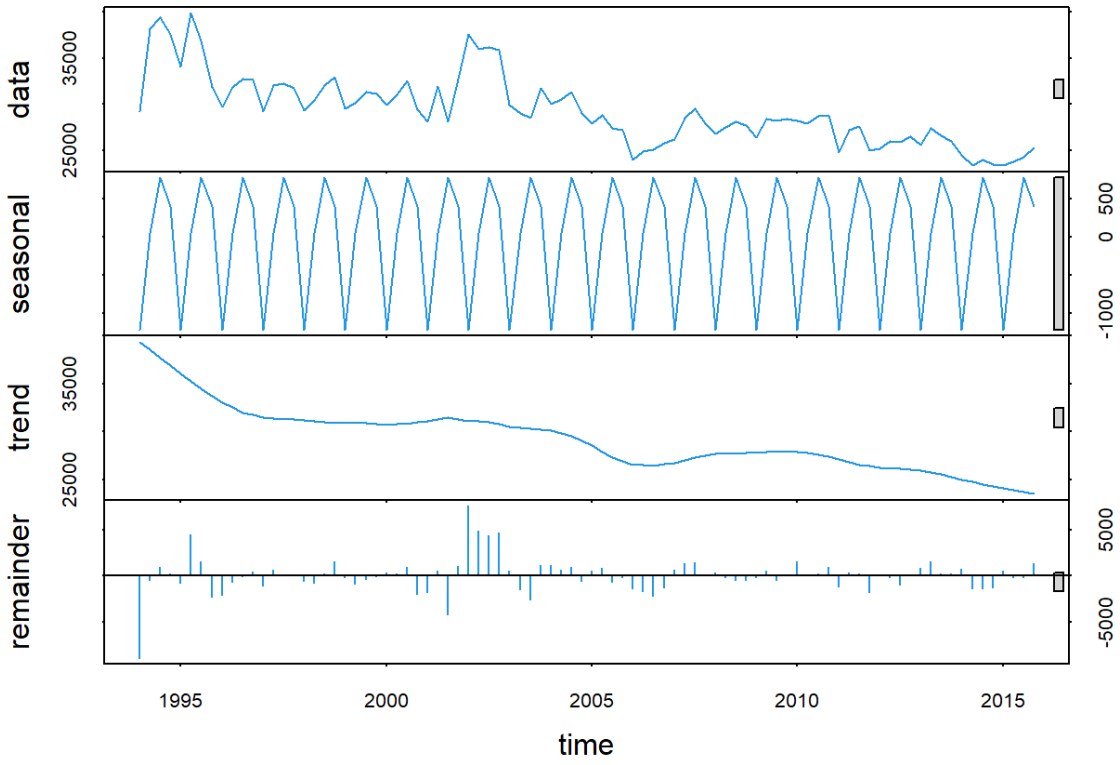
```
##
## Forecast method: Linear regression model
##
## Model Information:
##
## Call:
## tslm(formula = train ~ trend)
##
## Coefficients:
## (Intercept)        trend
##     34548.7       -117.5
##
##
## Error measures:
##                       ME     RMSE      MAE       MPE     MAPE      MASE
## Training set 1.654636e-13 2338.818 1744.816 -0.557401 5.779324 0.7881249
##                   ACF1
## Training set 0.5129561
##
## Forecasts:
##         Point Forecast    Lo 80    Hi 80    Lo 95    Hi 95
## 2016 Q1       24093.43 20968.14 27218.71 19282.75 28904.10
## 2016 Q2       23975.95 20848.30 27103.60 19161.64 28790.27
## 2016 Q3       23858.48 20728.41 26988.54 19040.44 28676.51
## 2016 Q4       23741.00 20608.47 26873.53 18919.17 28562.84
## 2017 Q1       23623.53 20488.48 26758.58 18797.82 28449.24
## 2017 Q2       23506.05 20368.43 26643.67 18676.39 28335.72
## 2017 Q3       23388.58 20248.34 26528.82 18554.88 28222.27
## 2017 Q4       23271.10 20128.20 26414.01 18433.30 28108.91
## 2018 Q1       23153.63 20008.00 26299.26 18311.64 27995.62
## 2018 Q2       23036.15 19887.76 26184.55 18189.90 27882.41
```

3. **Time series decomposition method:** The decomposition method is the most appropriate method, if the level of time series does not vary with the seasonal fluctuations, or the variation around the trend-cycle. Multiplicative decompositions are also more appropriate when the variation in the seasonal pattern or the variation around the trend-cycle develop to be proportional with respect to the level of time series. In this method the data needs to be transformed until the variation of the series develops to be stable.

   The formula for additive decomposition is

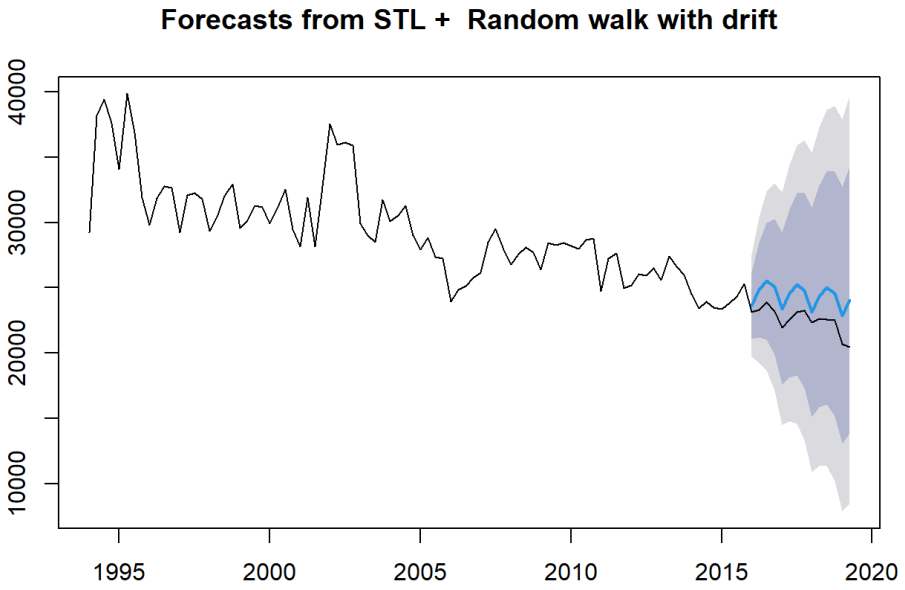
$$Y_t = S_t + T_t + R_t,$$

   Where y is the data, S stands for seasonal component. T is for the trend. In case of multiplicative decomposition the same formula is multiplied. $Y_t = S_t \times T_t \times R_t$



In this paper, STL model is used to decompose the time series data. From the above visualization there are many factors to analyze. Firstly from the function forecast, the method random drift (trend) walk is used to know how the trend moves with respect to drift. The seasonal component shows a stronger seasonality. In trend component the sales forecast is diminishing from the initial point. The top feature called data is showing an accurate measure with respect to the quarterly sales.

The gray bar on the right plays an important role in analyzing each panel features. The seasonal component has the longest bar which means the seasonal component includes smallest variance (between -1000 to 500). The other components have similar gray bar and the variance is large. These parameters describe which forecasting method can be selected. This analysis can be very important in considering trend pattern and analyze all the factors in one diagram. Therefore, seasonal component works better than all other models with respect to gray bar.



From the above diagram, the model uses STL and random walk with drift. The forecast from STL random walk is greater than usual quarterly sales, which is a good sign for the company to consider. The plot describes that the forecast follows seasonal component and there is increase in quarterly sales. Forecasts of STL objects have shown greater seasonality when compared to linear regression model. The below are the accuracy measures from time series decomposition model. When compared to the error score of Mape the model performs well, when compared to linear model. However, this method holds good for forecasting LPG quarterly sales. Mape score is 4.3 which indicate average of actual and forecast values.

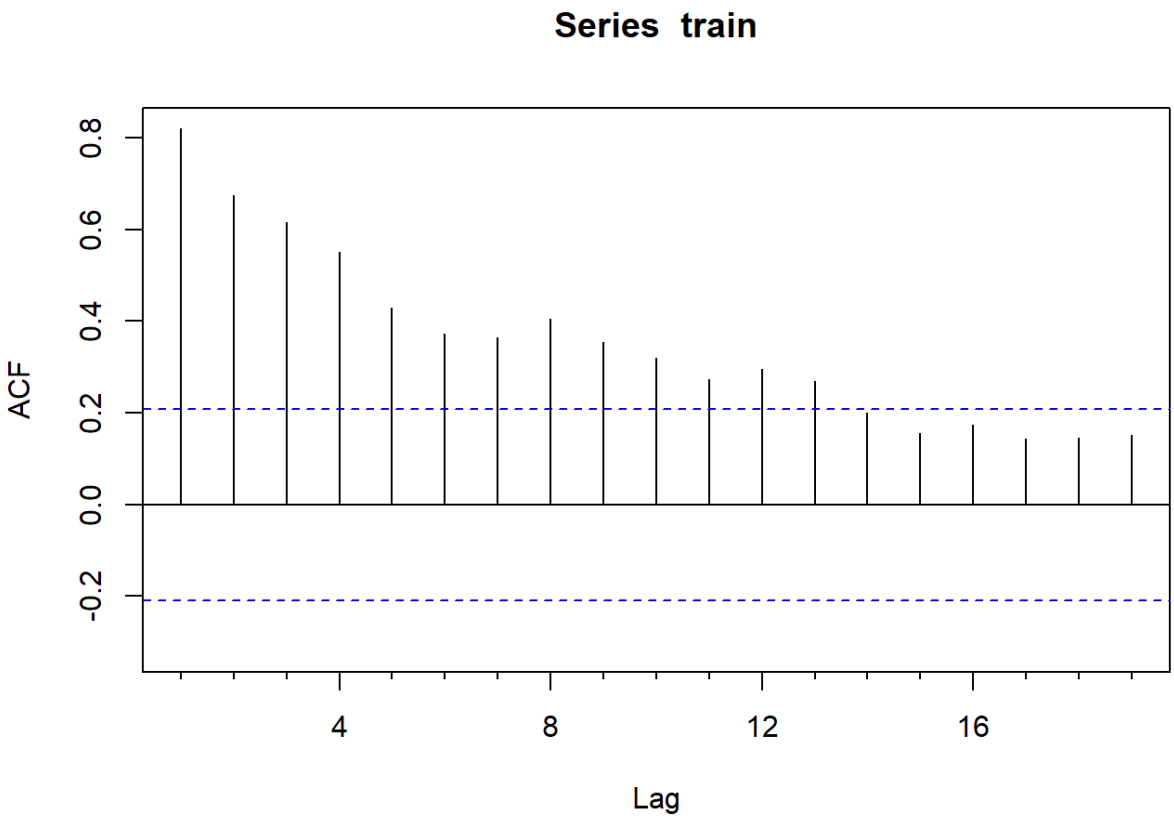
```
Forecast method: STL +  Random walk with drift

Model Information:
Call: rwf(y = x, h = h, drift = TRUE, level = level)

Drift: -63.5552  (se 213.3033)
Residual sd: 1989.5608

Error measures:
                         ME     RMSE      MAE        MPE     MAPE      MASE
Training set -8.364446e-14 1978.093 1327.908 -0.1876156 4.387341 0.5998095
                     ACF1
Training set -0.08596377
```

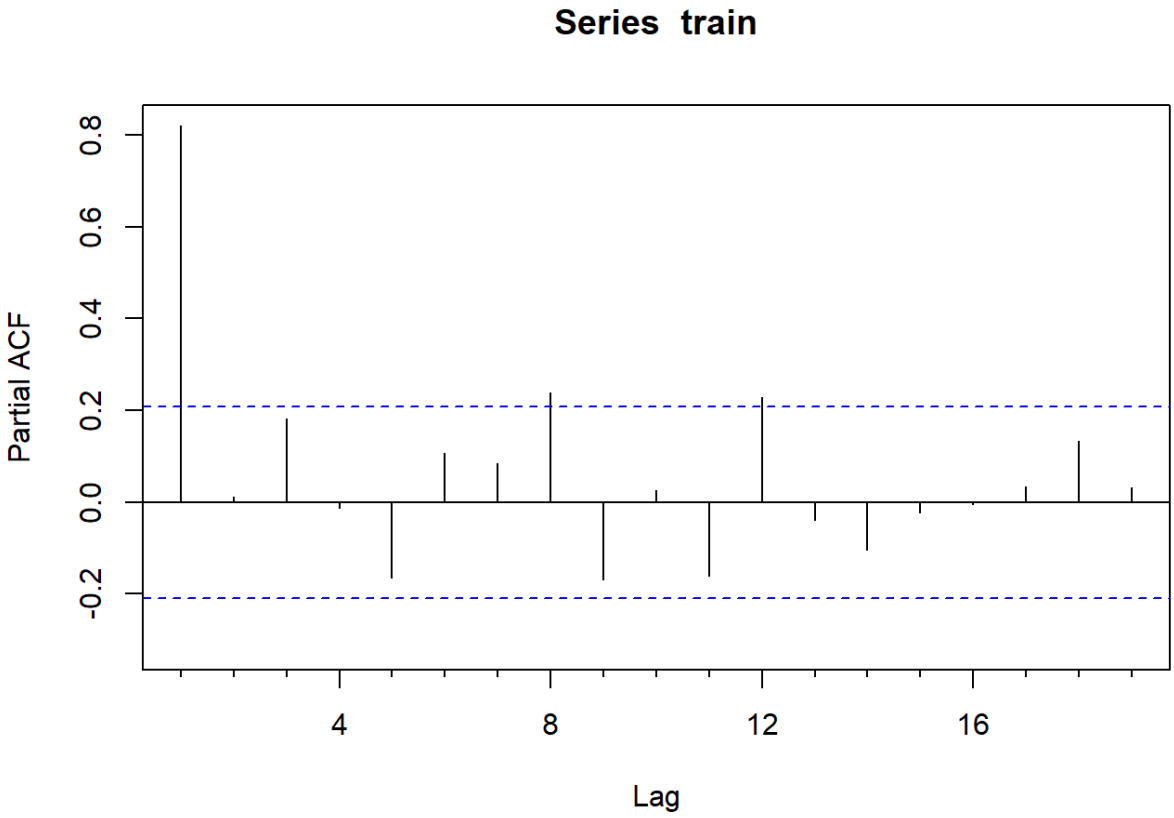
4. ARIMA model of time series forecasting:

Auto regressive integrated moving average has two models namely Seasonal and Non-seasonal which are used for forecasting. In non-seasonal method there are three variables to be considered: P- periods to lag, this helps in adjusting line which is being fitted to forecast the series. D – Differencing, This is used to transform a time series into stationary (series without trend or seasonality), and is important step in preparing data to be used in an ARIMA model. Q- This variable is the lag error component, where this not explained by a trend or seasonality.

Auto correlation function plot: This function refers to the plot used to describe the correlation between the points, according to the lag unit. So, according to the past values this determines how correlated each value in a time series. We use AR model when there is positive correlation at lag 1. When there is negative correlation we use MA model.

A partial auto correlation function plot is the relationship between observations in a time series and provides summary of the relationship.



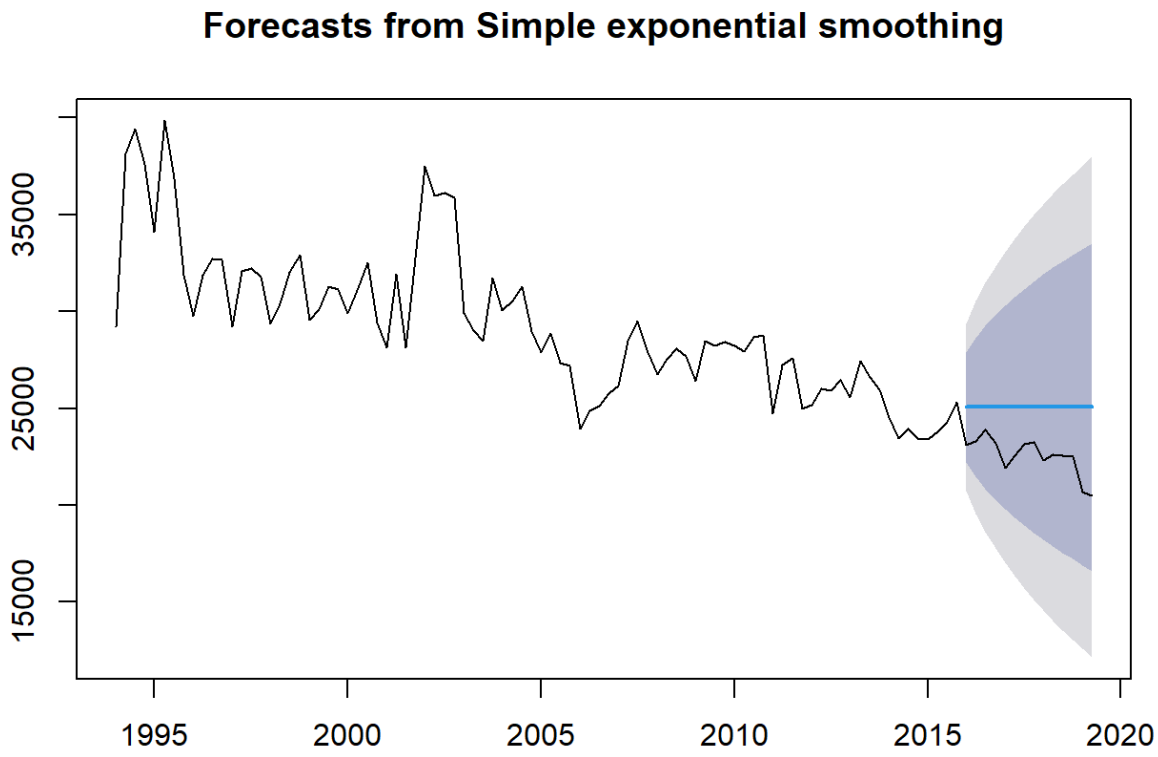


From the function Acf() and pacf() we get the above diagram. In Acf the lines are decreasing throughout the time series period, which after lag 14 the line are not reaching the blue dotted lines, so we can say it is not statistically significant. By comparing these models we can analyze whether it is autoregressive or moving averages. When Acf tails off and Pacf cuts off after at lag(p) then this model is called AR(1). The quarterly sales of LPG are not significant in this forecast. Therefore, the Mape value of this model is 5.01, by this we can conclude that this model is not effective as time series decomposition.

5. Simple exponential smoothing:

The smoothing models attempt to smooth the fluctuations in time series by smoothing or averaging. Exponential smoothing is one of the statistical methods commonly used for forecasting. In this forecasting method weighted averages of past observations are taken to produce forecasts. In this framework forecasts are generated quickly and for a wide range of time series that have great influence and leading application for industries.

This method is the simplest method and is suitable for forecasting data with no clear trend or seasonal pattern. Simple exponential smoothing calculates on weighted averages (the average of previous level and current observation). The weight of each parameter, or decrease in weight is always determined by smoothing parameter which is called as **alpha (α)**.



From the above diagram, the simple exponential smoothing has only one component level, which is visible as a straight line. As mentioned earlier, this method is suitable for forecasting with no trend or seasonal pattern. With respect to the dataset the trend is diminishing from the start, so the plot is almost similar to Mean or naïve method.

From the below error assessment alpha value is 0.77, that means when the alpha value is close to one indicates in fast learning based on giving more weights according to the recent observations. Practically, this method impacts on most recent values because the quarterly sales in the past are decreasing as seen in the trend and seasonality. Therefore, the Mape value of this method is also as competitive as ARIMA model which is 5.1%.

```
##
## Forecast method: Simple exponential smoothing
##
## Model Information:
## Simple exponential smoothing
##
## Call:
##  ses(y = train, h = h, initial = "simple")
##
##   Smoothing parameters:
##     alpha = 0.7797
##
##   Initial states:
##     l = 29226.3
##
##   sigma:  2208.7
## Error measures:
##                      ME    RMSE      MAE        MPE     MAPE      MASE       ACF1
## Training set -60.64427 2208.7 1543.169 -0.5223736 5.100971 0.6970421 0.05078305
##
```

Conclusion:

LPG quarterly sales forecasting was a challenging process because the quarterly sales were decreasing every year. So, initially data exploration step was important to analyze the sales data. With this regard, different methods were used to evaluate the performance. The performance of the quarterly sales stood out in Time series decomposition model according to mape value and trend analysis. The Linear regression model showed good seasonality when compared to other models. Arima and exponential smoothing were average and showed competitive error value assessments. By this analysis LPG Company can implement effective decisions based on models and error assessment. This helps in taking strategic moves with regard budgeting and advertisement cost. However, LPG quarterly sales forecasting can improve sales with above analysis in forecasting.