International Journal of Business and Management Invention
ISSN (Online): 2319 – 8028, ISSN (Print): 2319 – 801X
www.ijbmi.org Volume 3 Issue 5‖ May. 2014 ‖ PP.01-10

Comparative Study of Demand Forecast Accuracy for Healthcare Products Using Linear and Non Linear Regression

Sushanta Sengupta[1], Ruma Dutta[2]
[1]Tata Consultancy Services Limited, Kolkata
[2]Netaji Subhash Engineering College, Kolkata

ABSTRACT : *Business forecasting generally follows time series variations with seasonal, cyclical, trend and random fluctuations in it. Most of the business demand forecasting techniques generally follow one or more of the combinations of the above variations. Similarly, demand functions for healthcare products also follow similar demand patterns. Everyday healthcare managers make decisions about service delivery without having the knowledge of what will happen in future. Forecasting techniques help them in planning for their future deliveries. In this paper comparisonof the forecast accuracy of demand of heath care products has been done using Linear and Non Linear regression functions. Two key factors have been identified specifically for demand forecast in health care domain. While analyzing demand forecast accuracy using Linear Regression function, five factors all total have been considered as predictor variables, with booking forecast considered as dependent variable. While analyzing the demand forecast using Non-linear regression functions, two key factors related to health care product demand have been considered to interpret the final outcome. Experimentally it is found that factors specific to healthcare domain contribute positively to demand forecast. While we used non-linear regression polynomial function, it was found that for fluctuating demand having lumpy and very small demand in history, the forecast accuracy was better than the what linear regression gave us. For steady demand although Linear regression function gave us better accuracy.*

KEYWORDS*: Healthcare, Mean Absolute Percentage Error, Multiple Linear Regression, Forecast Accuracy, Non-linear Regression, Chi Square distribution, Degrees of freedom, Model Adequacy.*

I. INTRODUCTION

Forecasting is the process of making statements about events whose actual outcomes have not yet been observed. One example is an estimation of some variable of interest at some specific future date. There are two kinds of forecast available, one is Qualitative method and other one is Quantitative method. Qualitative forecasting methods are subjective, based on the opinion and judgment of experts, consumers. This is applicable when historical data is not available. Examples of qualitative forecasting methods are Delphi method, market research etc. Quantitative forecasting model is used to forecast future data as a function of past data and this is applicable when the historical data is available. Examples of quantitative forecasting methods are last period demand, simple and weighted moving average, exponential smoothing etc. There is time series methods applied widely in Quantitative forecasting models. Autoregressive moving average (ARMA), Autoregressive Integrated Moving Average (ARIMA) like Box-Jenkins is some special cases for Time series methods.

There is another most significant forecasting method called causal/ econometric forecasting methods. This method identifies the most probable factors that might influence the variable that is being forecast. Causal method of forecasting technique includes Regression analysis. This includes a large group of methods for predicting future values of variable using information about other variable. This method includes both parametric (linear/non-linear) and non-parametric techniques.There is a term in forecasting which is called the Forecast accuracy. This is measured by the forecast error which is the difference between the actual value and forecast value for the corresponding period. The different methods used for measuring forecast accuracy are Mean Absolute Error (MAE), Mean Absolute PercentageError (MAPE), Mean Absolute Deviation (MAD), Percent Mean Absolute Deviation (PMAD), Mean Squared Error (MSE), Root Mean Squared Error (RMSE), Average Error etc. In business forecasting, forecasters and practitioners sometimes used different technology in the industry. Mostly used accuracy method in industry is MAPE.There is a wide application of forecasting in different kind of situations and in different industries. One of the most important applications of forecasting is in Supply chain management. Forecasting can be used in supply chain management to ensure that the right product is there at right time and at right place. Accurate forecasting will help the retailers to reduce excess inventory and in turn increasing the profit margin. Accurate forecasting will also help retail business to meet their customer demand.

The multiple regression model is very much suitable to predict for the forecast for better control of inventory in an industry. The more accurate forecast we can make out for demand of any product the fewer inventories we can keep in our stock. In turn, this says that, as the forecast value is closer to the actual demand of a product, the inventory turnover ratio will be more. That is keeping an optimum inventory we can satisfy most to our customer requirements. In this way, our inventory carrying cost will be optimum without compromising with the customer service level and all. It will in turn, help in increasing the total cost incurred by a company and hence the profitability of the business will also increase.

In context to the healthcare industry, it is very critical to keep an optimum inventory of each product. There are mostly lifesaving drugs in healthcare industry which require immediate supply wherever the necessity comes. So the patients who are suffering from a critical disease might require one medicine, say within few hours of time without which his/ her life might be in danger. So we cannot keep the stock of that kind of drugs nil at any point in time. Again, we should not keep the stock of those too much, as in that case our inventory carrying cost/ holding cost will be more. This will in turn affect the total cost of the company and automatically the profitability for the business will get reduced/ affected. So in general, using multiple regression model in healthcare and many other products, not only the forecast accuracy will increase, but also the inventory cost will reduce largely causing the profitability for the business to enhance a lot without compromising with the customer satisfaction level.One more benefit we will get in accurate forecast is reduction of the cost of a product. As the inventory carrying cost is getting reduced by forecast accuracy, so it will impact on the total cost of the product. This in turn helps in reducing the sales price of the product although keeping more profit margin on it. As the sales price of the product reduces, the sale of the product manufactured by a certain company will get automatically increased as compared to the same product manufactured by some other company. This will in turn increase the market share of the company more as compared to the other companies in the same business. So the market price will be more competitive and sale will be more for the specific company products increasing their share values also in market. In healthcare industry cost reduction of a product is very much beneficial and desired for mass people.

We know that any human being will someday or the other suffer from any kind of diseases whatsoever. If we can reduce the market price of a healthcare product say a medicine as less as possible, most of the people in poor countries can afford to buy those drugs and can save their life. In gist we can say better forecast accuracy in healthcare industry has a bigger effect on the society as a whole.In this study, multiple linear and non-linear regression models have been used with different relevant parameters to predict the booking history of a few healthcare products. The aim was to relate the demand forecast with five predictor variables influencing the healthcare forecast in case of linear regression function. We are going to establish that two among the five parameters are having most significant contribution to the forecasted variable in that case. While doing the non-linear analysis, we took those two key factors and put them in a suitable non-linear regression polynomial model giving better forecast accuracy. Mean absolute percentage error is used to get the most accurate regression function for all the cases.

II. RELATED WORK

The existing literatures in the demand forecasting area generally focuses on forecasting solution for a particular type of demand function. The paper by S. Sengupta &R.Dutta on Identification of Demand Forecasting Model considering Key factors in the context of Healthcare products, 2014 focused only on Linear multiple regression model for demand forecasting. There are papers which talk about various techniques in managing the level of uncertainty (Bartezzaghiet. Al., 1995;Bartezzaghiet. Al., 1999; Synetoset. Al.,2005). Such papers focused on development of single algorithm or framework. The second group of papers focused on providing solution to intermittent or lumpy demand with a variety of tools (Ward, 1978; Wemmerlovet. Al., 1984; Wemmerlovet. Al., 1986). While some of the papers focused on the issues given a fixed context (Synetos, 2001) and some have mentioned importance of identification and description of demand function (Rafael, 2002). Most of these papers do not generally discuss any points about the proper identification of the demand function. The other group of papers focuses on the system that oversees the operation and how improvement to this system does help (Fildes, 1992). There are also papers on describing how the processes are affected by the lumpy demand (Ho, 1995). They do not classify the type of demand with respect to the individual characteristics. In the papers by Balis, Peppers, Kress and Synder in 1993, 1994forecasting actions are defined clearly and the consistency of forecasting elements is clear, but the model has intermittent character and the feedback is disregarded. Paper by Shim, Siegel and Liew in 1994 focuses strongly on the forecasting information through the prism of planning which can limit the number of factors of external environment.

In the paper of Cox and Karsten in 1989, 1990, clear objects of forecasting of demand are emphasized. This increases the possibility of making more accurate forecast due to detail analysis of external factors. However, too little attention was given to feedback and the final action of forecasting and control of forecasts are not considered. The paper of Adams in 1986 focuses on combination of combination of three forecasting levels which make possible to assess better influence of forecasting factors on forecast. However, internal mechanism of forecasting and its elements are not assessed. The consistency of assessing the accuracy of forecast is uncertain. The paper by Churchill, Ford and Walker in 1993 clearly states the variables and factors of forecasting. The model is also having a continuous character. There is uncertainty of selection of forecasting variables in noted and only ingoing variables are emphasized. The paper by Makridakis, Weelwright and Hyndman in 1998 focuses on detailed description of forecasting. The model is not complicated and easy to master. However, it has lack of forecasting actions consistency. The model also does not contain feedback and limited number of actions of forecasting is provided.

III. PROPOSED WORK

In a regression analysis we generally study the relationship or the regression function between one dependent variable y and several other independent variables represented as x_i. Regression function also involves a set of unknown parameters called b_i. If a regression function is linear with respect to the parameters, we call it as linear regression model. If the function is not linear in the parameters, then it is called non-linear regression model.Linear regression models with more than one independent variable are called multiple linear regression models. If the number of independent variable is one, then we call it as simple linear regression model.

The following notations can be used by us:

y = Dependent variable (Predicted by the regression model)

yt= Dependent variable (Actual value)

p = number of independent variables/ number of coefficients

$x_i(i=1,2\ldots,p)$ = ith independent variable from total set of p variables

$b_i(i=1,2\ldots,p)$ = ith coefficient corresponding to x_i

b_0= intercept or constant

k=p+1 is the total number of parameters including intercept/ constant

n = number of observations (experimental data points)

i = 1,2,…p is independent variable's index

The general formula for multiple linear regression is given as below:

$y = b_0 + b_1x_1 + b_2x_2 + \ldots\ldots + b_px_p$(1)

There are different Non-linear Regression models available. Some popular non-linear regression models are as below:

- Exponential Model: $y = ae^{bx}$(2)
- Power Model: $y = ax^b$(3)
- Saturation Growth Model: $y = ax/(b+x)$ (4)
- Polynomial Model: $y = a_0 + a_1x + a_2x^2 + \ldots\ldots + a_mx^m$ (5)

In case of Multiple Linear Regression, the primary goal was to determine the best set of parameters biso that the model predicts the actual values of the dependent variables as much accurate as possible. That means predicted values y should be close to actual values y_t. In case of Non-linear regression equation we have chosen the Polynomial model in this case and one independent key factors for health care domain. Hence, the primary objective was to determine the best set of parameter $a_i$and also the exact polynomial which will give best forecast accuracy. Whether all the independent variables in our model are significant or nothad been checked. In turn we established which one of the independent/ predictor variables is most significant in predicting most accurately in our work in linear scenario. In non-linear scenario, we used the most significant factor from linear analysis and fit the best polynomial function in order to search for getting a better accuracy over the linear model.MAPE (Mean absolute percentage error) have been used in order to find out the most accurate regression function by varying our independent variables. In this paper, a set of month-wise booking history data for previous one year from some healthcare industry for a few healthcare products has been considered. On the basis of the booking history data and some other independent variables, month-wise forecast for current one year has been done.

The dependent and independent variables for our work are as follows:

y = Month-wise Predicted booking history for the year 2013

y_t= Month-wise actual booking history for the year 2013

x_1 = Month-wise actual booking history for the year 2012

x_2 = ASP$(Average selling price in dollars)

x_3 = Advertisement Cost in dollars
x_4 = Percentage of people w.r.t the total population having age more than 65 years (Above 65)
x_5 = Percentage of people w.r.t the total population doing physical exercise regularly (Exercise)
In the linear regression analysis of our work, all the independent/ predictor variables together have been varied and the parameters for which the forecast accuracy becomes maximum, has been found out. To be more precise, we have varied the last two variables, whichare"Above 65" and "Exercise" and looked for which one of these two has the most impact on the predicted forecast value. To do that, these two variables together have been varied first. As a next step we made both of the variables as constant w.r.t the months. At last one variable say "Above 65"has been varied keeping the variable "Exercise" as constant and vice versa. While doing that, in each case the residual output part of regression analysis has been noted. Then Percent error for each of the observations for that specific item and combination of independent parameter values has been calculated. The percent error has been calculated by dividing the residuals with the actual booking history values. Then we have calculated Mean absolute error (MAE) for the specific set of regression coefficients. The minimum value of these MAE gave us the maximum accuracy on the predicted forecast value. When we did the exercise with few healthcare products, it was observed that the factor "Above 65" and "Exercise" had most significant impact in the demand forecast value. The maximum forecast accuracy value was achieved by varying these two decisive predictor variables for healthcare forecasting.

In Non-Linear regression analysis of our work, we have considered the most critical factors affecting the health care demand as an input. Using those variables we had fit one polynomial function. We started with degree 2 of the independent variable x and increased the degree gradually to 10 in some cases. In some cases we also used the degree of ½ for independent variable x with proper judgment looking at the actual demand history values and polynomial coefficients. What we saw is that when using fifth or sixth degree of the dependent variable the forecast accuracy coming as minimum in most of the cases. For few of the cases it was giving a better forecast accuracy in comparison to the accuracy of linear case.
The calculation of MAE / MAPE can be depicted as below:
MAE = Summation of [Absolute {(Actual-Forecast)/Actual}]/n (6)
MAPE = Summation of [Absolute {(Actual-Forecast)/Actual}]x100%/n (7)
Where absolute function denotes the absolute value irrespective of the sign (+/-) of the calculated value and n denotes the number of observations.
There are other methods of forecast accuracy measurement like Mean Squared Error (MSE). MSE is the average of the squared forecast errors. The other common methods of forecast accuracy measurement are Mean absolute deviation (MAD), Cumulative error, Average error, Tracking Signal(TS) etc.

IV. EXPERIMENTAL RESULT

In case of linear one, we have varied five parameters in different combinations for each of the healthcare products and calculated the MAE for the forecast values w.r.t the actual demand values for each month. There were predictor variables like "Booking History", "ASP", "Advertisement Cost", "Above 65" and "Exercise". For each of the regression functions obtained by varying the different predictor variables, we arrived at a set of MAE among which we identified the minimum value of the MAE for each item. The output result from our experiment has been shown in a concise form in the below table.

Table -4.1: Linear Regression Output: Item wise MAE values for different predictor variables for arriving at Maximum forecast accuracy

Item Number	Parameters Varied in Regression function	Mean Absolute Error (MAE)
1001-090	Booking History, ASP, Advertisement Cost, Exercise (%)	0.245
1005-110	Booking History, ASP, Advertisement Cost, Above 65(%)	0.074
10300214	Booking History, ASP, Advertisement Cost, Above 65(%)	0.89
10311822	Booking History, ASP, Advertisement Cost, Exercise(%)	5.11
10312745	Booking History, ASP, Advertisement Cost, Exercise(%)	7.069
10318862	Booking History, ASP, Advertisement Cost, Above 65(%)	0.585

From the above set of experimental results, we observed that when one of the healthcare specific important predictor variables (in this case this is either "Above 65" or "Exercise") has been varied, the forecast accuracy becomes maximum i.e. the error in forecast is minimum.

To be more precise, if we closely observe the output, it can be also said that for most of the cases when the critical predictor variable "Above 65" is varied keeping other key predictor variable "Exercise" constant, the forecast accuracy becomes maximum. Here "Above 65" denotes the percentage of total population having age more than 65 and "Exercise" denotes the percentage of total population doing physical exercise every day. For few of the cases it also happened that when the critical variable "Exercise"is varied keeping the other one "Above 65" constant, the forecast accuracy became maximum. The abnormally high MAE (Mean Absolute Error) like 5.111 has come due to some lumpy demands along with some very small actual demands for the specific product concerned.Since regression function has limitation in predicting very accurately for such a fluctuating demand, hence the MAE value has come out to be very high in this case.

Table -4.2: Non-Linear Regression Output: Item wise MAE values for different degrees of Polynomial function

Item Number	Parameter Used	Polynomial Used	Mean Absolute Error(MAE)
1001-090	Exercise(%)(X1)	$Yt=A1+A2*X1^{2}$	0.308578
		$Yt=A1+A2*X1^{3}$	0.307209
		$Yt=A1+A2*X1^{4}$	0.305141
		$Yt=A1+A2*X1^{5}$	0.302441
		$Yt=A1+A2*X1^{6}$	0.47588
1005-110	Above 65(%)(X1)	$Yt=A1+A2*X1^{2}$	0.268
		$Yt=A1+A2*X1^{3}$	0.2641
		$Yt=A1+A2*X1^{4}$	0.2604
		$Yt=A1+A2*X1^{5}$	0.2568
		$Yt=A1+A2*X1^{6}$	0.2536
		$Yt=A1+A2*X1^{7}$	0.2641
		$Yt=A1+A2*X1^{10}$	0.2914
10300214	Above 65(%)(X1)	$Yt=A1+A2*X1^{0.5}$	0.9529
		$Yt=A1+A2*X1^{2}$	0.9448
		$Yt=A1+A2*X1^{3}$	0.8397
		$Yt=A1+A2*X1^{4}$	0.7972
		$Yt=A1+A2*X1^{5}$	0.7855
		$Yt=A1+A2*X1^{6}$	0.9529
10311822	Exercise(%)(X1)	$Yt=A1+A2*X1^{2}$	6.767
		$Yt=A1+A2*X1^{3}$	6.2967
		$Yt=A1+A2*X1^{4}$	5.5323
		$Yt=A1+A2*X1^{5}$	4.92
		$Yt=A1+A2*X1^{6}$	4.4529
		$Yt=A1+A2*X1^{7}$	8.3147
10312745	Above 65(%)(X1)	$Yt=A1+A2*X1^{2}$	3.41
		$Yt=A1+A2*X1^{3}$	1.968
		$Yt=A1+A2*X1^{4}$	1.634
		$Yt=A1+A2*X1^{5}$	1.435
		$Yt=A1+A2*X1^{6}$	1.388
10318862	Above 65(%)(X1)	$Yt=A1+A2*X1^{2}$	0.669
		$Yt=A1+A2*X1^{3}$	0.731
		$Yt=A1+A2*X1^{4}$	0.745
		$Yt=A1+A2*X1^{0.5}$	0.701
		$Yt=A1+A2*X1^{5}$	0.729

In case of Non-linear one, a polynomial function is used between the forecast variable and predictor variable and considered the most significant predictor variable from the linear regression analysis. Different polynomial function is used by varying the degree of the predictor variable for each item. In this way we achieved a better forecast accuracy figure. The output result of our experiment on non-linear analysis has been provided in above Table-2.
In the following Table -4.3, an item wise value of linear and non-linear regression outputs are compared.

Table -4.3: Item wise comparison of MAE values for Linear and Non-Linear Regression output

Item Number	Mean Absolute Error (MAE) for Linear Regression	Mean Absolute Error (MAE) for Non-Linear Regression
1001-090	0.245	0.302
1005-110	0.074	0.253
10300214	0.89	0.785
10311822	5.11	4.452
10312745	7.069	1.388
10318862	0.585	0.669

V. ANALYSIS OF RESULT

While analyzing the result and to establish the significance of the sample data to conclude the result on the population as a whole, Statistical Test of Significance is used. The outcome of the test of significance method gave us the goodness of fit of our assumption that for steady demand, linear regression method provides better accuracy and for fluctuating demand having lumpy and very small demands in it, non-linear regression gives better accuracy for the entire population.Pearson's Chi-squared distribution for goodness of fit has been used to conclude on the same. The formulation of the statistical test of significance for goodness of fit using Pearson's Chi-squared distribution is as follows:
Let H_0(Null Hypothesis) – For steady historical demand Linear Regression function provides better accuracy than the Non-linear one and for fluctuating demand Non-linear regression function gives better accuracy than Linear one.
H1 (Alternate Hypothesis) – For steady historical demand Linear Regression function does not provide better accuracy than the Non-linear one and for fluctuating demand Non-linear regression function does not give better accuracy than Linear one.
In our case, there are two categories, Steady Demand and Fluctuating Demand. Hence the degree of freedom is one. So mathematically:
n = Number of categories which is 2 in our case
Degree of freedom = n-1 = 1
The value of the test-statistic is

$$\chi 2 = \sum_{i=1}^{n} (O_i - E_i)^2 / E_i \qquad (8)$$

Where
$\chi 2$= Pearson's cumulative test statistic
O_i= an observed frequency
E_i= an expected (theoretical) frequency
n = Number of categories
There were 2 samples where the demand was steady and for 4 samples the demand was fluctuating. In reality we saw for 3 of the cases linear regression gave better forecast accuracy and for other 3 non-linear regression gave better forecast accuracy. So the sample data can be represented in a table like below:

Table -5.1: Category wise frequency distribution of maximum accuracy outcome of sample

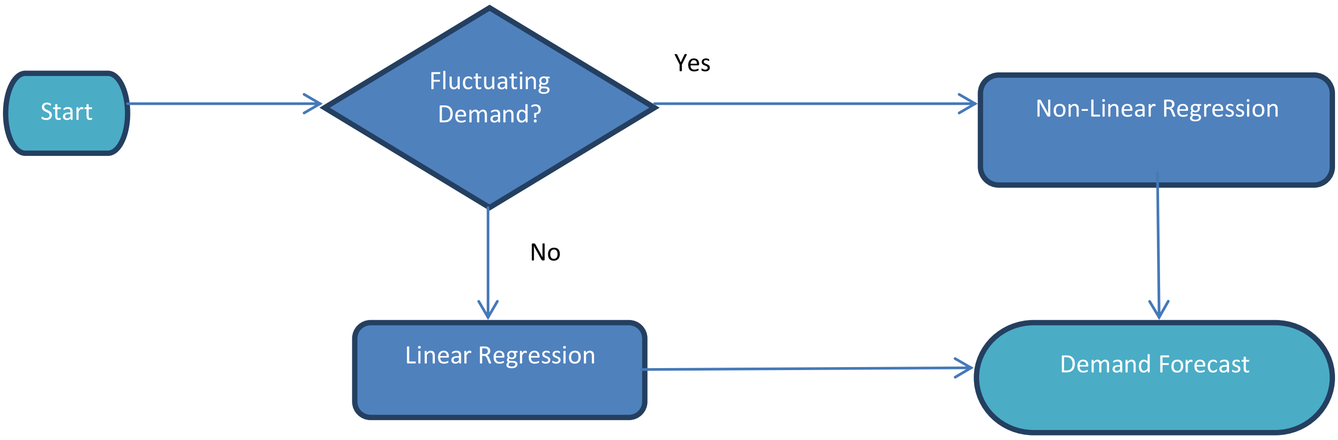
Category	Observed Frequency (O_i)	Expected Frequency (E_i)
Linear Regression	3	2
Non-linear Regression	3	4
Total Frequency	**6**	**6**

So, in our case, the Chi-squared value of test-statistic is calculated as below:

$$\boldsymbol{\chi 2 = (3-2)^2 / 2 + (3-4)^2 / 4 = 0.75}$$

Now, Chi-squared value for 95% confidence levelwith 1 degree of freedom from the table is $\chi 2_{0.05}$ = **3.841**, which is more than the Chi-squared value 0.75 obtained from our test sample statistic.In other words, Chi-squared value for the test-statistic is well below the acceptable region value obtained from the Chi-squared table at 95% confidence level. Hence we can conclude that our assumption or the null hypothesis cannot be rejected. So, the null hypothesis that for steady historical demand Linear Regression function provides better accuracy than the Non-linear one and for fluctuating demand Non-linear regression function gives better accuracy than Linear one in population is accepted and in conformity with our sample statistic output.

Flow Diagram – 1:



Model Adequacy: In order to check for the regression model adequacy for the best fit regression equations which gave us maximum forecast accuracy of the forecast figures w.r.t. the actual demand we checked few parameters:

- **Coefficient of determination**: This value comes out from the Regression Statistics output (R Square). This is computed as the ratio of variation in y accounted for by the regression model and the total variation in y. This is interpreted as the proportion of total variation of y that can be explained by the regression. R Square is a number between 0 and 1 and larger the value of R Square the better fit the model has. If R Square is 1, the model has a perfect fit. In our case, for the best fit model, R Square value for one of the sample item was 0.9879 for best fit curve and which clearly indicates the best fit regression model is adequate.
- **Residual Analysis**: The residual values are the difference between the observed value and the fitted values. Residuals are interpreted as estimates of random errors. For model adequacy, the residual values should be normally distributed about a mean zero, homoscedasticity. When we plot the residuals against fitted value, there should be no discernible pattern such as megaphone or runs of positives and negatives while checkingthe homoscedasticity of errors. In our all best fit models, the residuals are normally distributed with mean zero and residuals against fitted values do not show a discernible pattern.
- **Standard Residuals**: 95% of the values of the scaled residuals should be within [-2,2]. When we plotted our scaled or standard residuals against the fitted values, the values were within [-2,2] only for most of the records.
- **Residual as a function of x**: If residuals as a function of x show non-linearity that indicates that the model is not adequate.

Table -5.2: Regression Statistics Summary output of a best fit model showing R Square value:

Multiple R	0.993568487
R Square	0.987178339
Adjusted R Square	0.836994533
Standard Error	43.29908249
Observations	12

Graph – 1: Plot of Residuals against fitted values for a best fit regression data showing no discernible pattern:

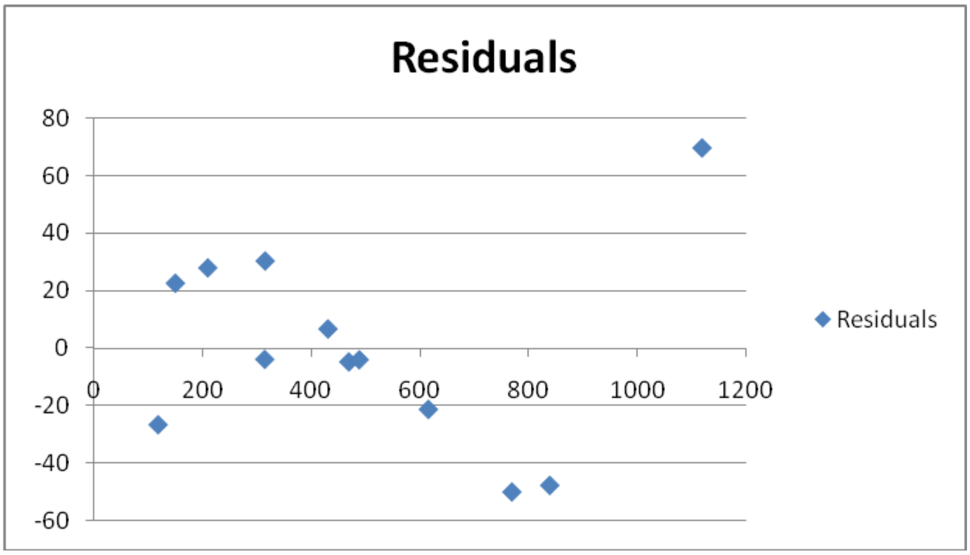


Table -5.3: Standard Residuals plotting against the fitted/ Predicted variable for a best fit regression data showing the range of Standard/ Scaled Residual values within [-2,2] for most cases:

Observation	*Predicted Booking history(2013)*	*Residuals*	*Standard Residuals*
1	468.6900001	-4.690000112	-0.141819439
2	838.6745193	-47.6745193	-1.441614799
3	487.9284626	-3.928462558	-0.118791544
4	768.9525321	-49.95253213	-1.510498913
5	313.791694	-3.791694039	-0.114655844
6	615.2203188	-21.2203188	-0.641674548
7	117.5433611	-26.54336112	-0.802636351
8	149.3009076	22.69909236	0.686390717
9	430.2190789	6.780921125	0.205046142
10	1119.195012	69.80498753	2.11081107
11	208.9421099	28.05789008	0.848433716
12	314.542003	30.45799696	0.921009793

Graph – 2: Plot of Residuals against the predictor variable “Above 65” for a best fit regression data shows linearity:

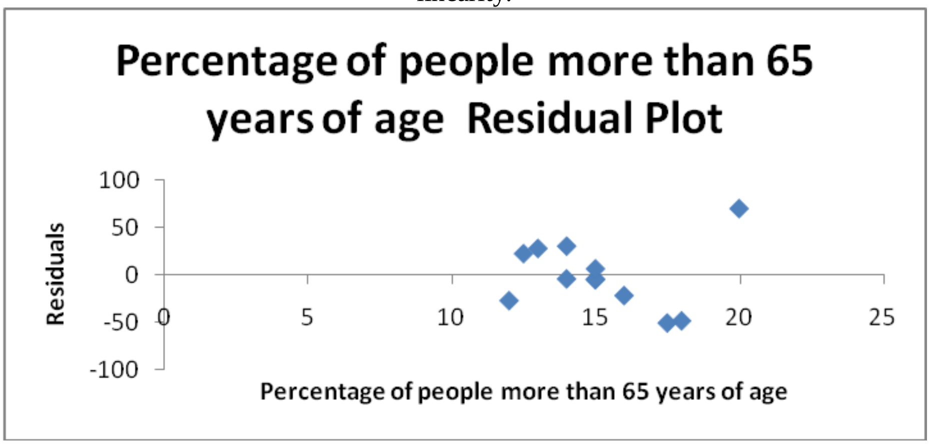
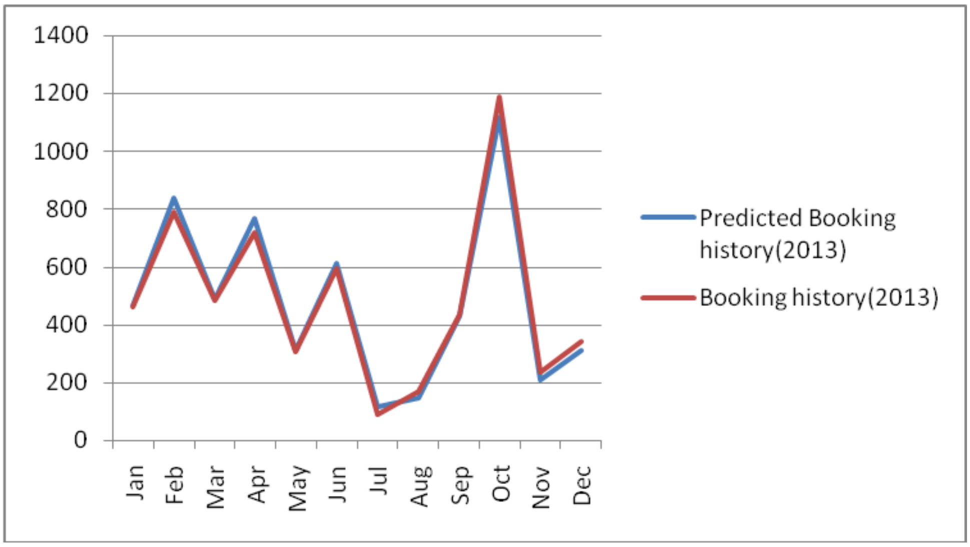


Table -5.4: Month-wise fitted data values against the actual demand history for a best fit model:

Observation	*Predicted Booking history(2013)*	Booking history(2013)
Jan	468.6900001	464
Feb	838.6745193	791
Mar	487.9284626	484
Apr	768.9525321	719
May	313.791694	310
Jun	615.2203188	594
Jul	117.5433611	91
Aug	149.3009076	172
Sep	430.2190789	437
Oct	1119.195012	1189
Nov	208.9421099	237
Dec	314.542003	345

Graph – 3: Plot of Month-wise fitted data values against the actual demand history for a best fit model:



VI. CONCLUSION

In this paper, using linear regression analysisthe most critical single external factor (in this case it is the percentage of people more than 65 years of age) was found out which influence the health care product demand forecasting. So the demand forecast accuracy for the healthcare products under investigation can be closely controlled and monitored by this factor. It has also been shown how the five predictor variables influence the demand forecast for some simulated products from healthcare. The relationship to real demand data was identified, so as to correctly identify the demand function and determine the most appropriate demand regression function with the smallest Mean Absolute Percentage Error. Next, we took the most significant predictor variable as an input to the Non-Linear regression analysis from the linear regression output. Then it has been investigated to identify for which degree of polynomial in non-linear case the forecast accuracy is coming maximum. It was seen that for a steady demand history figures the linear regression function gave more accuracy than the non-linear one. Again it was established that for very fluctuating demand history having few lumpy and small demands, the non-linear function gave better accuracy than the linear regression.It was also implemented that a good forecast does not only depend on the forecasting technique used but also on correct identification of demand function. The paper also provided a simple approach to identify the correct demand function without having a complex calculations or evaluation technique.

REFERENCES

[1]. S. Sengupta, R. Dutta, "Identification of Demand Forecasting Model considering Key factors in the context of Healthcare products", "International Journal on Application or Innovation in Engineering & Management", 2014

[2]. F.G. Adams,"The business Forecasting Revolution", Oxford: Oxford University Press, 1986.

[3]. M.L. Valentine,C.A. Dauten,"Business Cycles and Forecasting". Cincinnati: South – Western Publishing Company, 1993.

[4]. S.A. DeLurgio,"Forecasting Principles and Applications", Boston: McGraw-Hill, 1998.

[5]. K.CGreen.,J.S. Armstrong ,"Structured analogies for forecasting" , "International Journal of Forecasting", 2007.

[6]. Hanke J.E., " Business Forecasting", London: Simon & Schuster, Inc, 1992.

[7]. P. Herbig,J.Michewig, J.E.Golden, "Differences in forecasting Behavior between industrial product firms and Consumer product firms", "Journal of Business and Industrial Marketing", 1994.

[8]. S. Makridakis,S.C.Wheelwright, R.J.Hyndman, "Forecasting: Methods and applications", New York: John Willey & Sons, 1998.

[9]. G.J. Kress,J.Synder."Forecasting and Market Analysis Techniques", London: Quorum Books, Inc., 1994.

[10]. V. Pilinkien,"Selection of Market Demand Forecast methods: Criteria and application" Engineering Economics, 2008.

[11]. R. Daves, R.Fildes, M.Lawrence, K.Ord,"The past and the future of Forecasting Research", International Journal of Forecasting , 1994.

[12]. J.K.Shim, J.G.Siegel, C.J.Liew.,"Strategic Business Forecasting", Chicago: Probus Publishing Company, 1994.