IDENTIFICATION OF DEMAND FORECASTING MODEL CONSIDERING KEY FACTORS IN THE CONTEXT OF HEALTHCARE PRODUCTS

Sushanta Sengupta¹, Ruma Datta²

¹Tata Consultancy Services Limited, Kolkata
²Netaji 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 healthcare domain the most commonly used forecasting techniques are Judgmental forecast, time series approach, techniques of averaging, Naïve forecasts, exponential smoothing forecasts, linear regression etc.

In this paper two factors have been introduced specifically for demand forecast in healthcare domain. Five factors, all total have been considered as predictor variables with booking forecast considered as dependent variable. The model considered here is multiple linear regression. Experimentally it is found that factors specific to healthcare domain contribute positively to demand forecast. It was found that one of these critical factors is having highest impact on getting the most accurate demand forecast. The overall approach on business forecasting on few healthcare products helped us in getting the most appropriate multiple regression equation providing the best result in forecasting. However, if the demand pattern is very fluctuating in nature having few lumpy demands along with some small demands varying widely, then the multiple regression technique needs to consider those demands as outliers and exclude them while creating the forecast function. So this model will be better useful when the demand pattern is having less lumpy demands or fewer fluctuations in it.

Keywords: Healthcare, Mean Absolute Percentage Error, Multiple Linear Regression, Forecast Accuracy.

1. INTRODUCTION

Forecasting is the process of making statements about events whose actual outcomes have not been yet 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. Example 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 a 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 Percentage Error (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 less inventory 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 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 like India 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 regression model has been used with different relevant parameters to predict the booking history of few healthcare products. We have tried to relate the demand forecast with five predictor variables influencing the healthcare forecast. We also tried to establish that two among the five parameters are having most significant contribution to the forecasted variable. Mean absolute percentage error is used to get the most accurate regression function for them.

2. RELATED WORK
The existing literatures in the demand forecasting area generally focuses on forecasting solution for a particular type of demand function. There are other papers which talk about various techniques in managing the level of uncertainty (Bartezzaghi et. Al., 1995; Bartezzaghi et. Al., 1999; Synetos et. 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; Wemmerlov et. Al., 1984; Wemmerlov et. 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, 1994 forecasting 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, Wheelwright 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.

3. 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 variables are called multiple linear regression model. 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)
- \( y_t \) = Dependent variable (Actual value)
- \( p \) = number of independent variables/ number of coefficients
- \( x_{i} (i=1,2,...,p) \) = ith independent variable from total set of \( p \) variables
- \( b_{i} (i=1,2,...,p) \) = 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 is given as below:

\[
y = b_0 + b_1 x_1 + b_2 x_2 + ... + b_p x_p
\]

The primary goal of this work is to determine the best set of parameters \( b_i \) so 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 \( y_t \). We have to check if all the independent variables in our model are significant or not. In turn we will try to establish which one of the independent/ predictor variables is most significant in predicting most accurately in our work. 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, we have considered a set of month-wise booking history data for previous one year from some healthcare industry for few healthcare products. On the basis of the booking history data and some other independent variables, we tried to do month-wise forecast for current one year.

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
- \( x_2 \) = ASPS (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 this 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, which are “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 percentage error (MAPE) for the specific set of regression coefficients. The minimum value of these MAPE 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. The calculation of MAPE can be depicted as below:

\[
MAPE = \frac{\text{Summation of } \left| \frac{(Actual - Forecast)}{Actual} \right| \times 100}{n}
\]

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.

4. EXPERIMENTAL RESULT

We have varied five parameters in different combinations for each of the healthcare products and calculated the MAPE 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 MAPE among which we identified the minimum value of the MAPE for each item. The output result from our experiment has been shown in a concise form in the below table.

TABLE -1: Item wise MAPE values for different predictor variables

<table>
<thead>
<tr>
<th>Item Number</th>
<th>Parameters varied in regression function</th>
<th>Mean Absolute Percentage Error (MAPE)/100</th>
</tr>
</thead>
<tbody>
<tr>
<td>1001-090</td>
<td>Booking History, ASP, Advertisement Cost, Exercise</td>
<td>0.245</td>
</tr>
<tr>
<td></td>
<td>Booking History, ASP, Advertisement Cost</td>
<td>0.319</td>
</tr>
<tr>
<td></td>
<td>Booking History, ASP, Advertisement Cost, Above 65</td>
<td>0.312</td>
</tr>
<tr>
<td>1005-110</td>
<td>Booking History, ASP, Advertisement Cost, Above 65</td>
<td>0.074</td>
</tr>
<tr>
<td></td>
<td>Booking History, ASP, Advertisement Cost, Above 65, Exercise</td>
<td>0.082</td>
</tr>
<tr>
<td></td>
<td>Booking History, ASP, Advertisement Cost</td>
<td>0.371</td>
</tr>
<tr>
<td>10300214</td>
<td>Booking History, ASP, Advertisement Cost, Above 65</td>
<td>0.4</td>
</tr>
<tr>
<td></td>
<td>Booking History, ASP, Advertisement Cost</td>
<td>0.557</td>
</tr>
<tr>
<td></td>
<td>Booking History, ASP, Advertisement Cost, Exercise</td>
<td>0.448</td>
</tr>
<tr>
<td>10311822</td>
<td>Booking History, ASP, Advertisement Cost, Exercise</td>
<td>5.111</td>
</tr>
<tr>
<td></td>
<td>Booking History, ASP, Advertisement Cost, Above 65, Exercise</td>
<td>6.559</td>
</tr>
<tr>
<td></td>
<td>Booking History, ASP, Advertisement Cost</td>
<td>5.328</td>
</tr>
<tr>
<td>10312745</td>
<td>Booking History, ASP, Advertisement Cost, Above 65</td>
<td>0.487</td>
</tr>
<tr>
<td></td>
<td>Booking History, ASP, Advertisement Cost, Above 65, Exercise</td>
<td>0.512</td>
</tr>
<tr>
<td></td>
<td>Booking History, ASP, Advertisement Cost</td>
<td>0.513</td>
</tr>
<tr>
<td>10318862</td>
<td>Booking History, ASP, Advertisement Cost, Above 65</td>
<td>0.585</td>
</tr>
<tr>
<td></td>
<td>Booking History, ASP, Advertisement Cost</td>
<td>0.641</td>
</tr>
<tr>
<td></td>
<td>Booking History, ASP, Advertisement Cost, Above 65, Exercise</td>
<td>0.623</td>
</tr>
</tbody>
</table>

From the above set of experimental results, we observed that when we vary one of the healthcare specific important predictor variable (in this case this is either “Above 65” or “Exercise”) the forecast accuracy becomes maximum that is 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 we vary the critical predictor variable “Above 65” 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 we varied the critical variable “Exercise” 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 demands, hence the MAE value has come out to be very high in this case.

5. CONCLUSION

In this paper, we did try to find out the most critical single external factor (in this case it is the percentage of people more than 65 years of age) influencing 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. We have attempted to show how the five predictor variables influence the demand forecast for some simulated products from healthcare. We identify the relationship to real demand data so as to correctly identify the demand function and determined the most appropriate demand regression function with the smallest Mean Absolute Percentage Error. We have 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

AUTHOR

Mr. Sushanta Sengupta: Author, presently he is working as Functional Consultant in ERP (Enterprise Resource Planning) in Tata Consultancy Services Limited. He has also worked in other manufacturing and Engineering industries in various roles. He has completed his B.E. in Electrical from Jadavpur University and MBA in Systems. He is a Six Sigma Green Belt holder and holds various other Certifications in ERP and oracle application product.

Dr. Ruma Dutta: is currently an Associate Professor of Netaji Subhash Engineering College, Kolkata. She had done her PhD. (Engg) from Jadavpur University. Previously, she had done her M.E. and B.E. from Bengal Engineering College (now BESU). She has research interest on Web Mining, Data Mining, Cloud Computing and MIS.