International Journal of Basic Science and Technology
August, Volume 7, Number 2, Pages 142 - 149

ISSN 2488-8648
http://www.ijbst.fuotuoke.edu.ng/ 142

Improving Prediction Accuracy using a Combined Two Multiple Regression Equations in a Real-Time System

[1]Agholor, S., [2]Agholor, O. A. and [3]Akinyemi, I. O.

[1]Department of Computer Science, [2]Registry Department
Federal College of Education, Abeokuta, Nigeria
[3]Mountain Top University, Ibafo, Nigeria

Article Information

Article # 02018
Received: 18th May 2021
Revision: 29th July 2021
Acceptance:29th July 2021
Published:11th August 2021

Key Words
Estimated Demand, Forecasting Model, Forecasted Demand, Mono Equation, Multiple Regression Equation

Abstract

There have been widely publicized media reports on goods confiscated by government agencies because the goods are expired or manufactured below standard. Preliminary investigation revealed that when production far exceeds demand, it can lead to having expired goods in the market and when manufacturers are in a hurry to meet up consumers' demand, it can lead to production of sub-standard goods. Forecasting models were developed to solve these problems. One of the commonest forecasting models is the one that uses mono equation of simple or multiple regression analysis. This type of model, still gives wide variations between the actual and forecasted demand due to its inability to accommodate the contributions of other competitors to the total market demand for the product. To correct this, a combined two multiple regression equations model called dForecast 2019 was developed. The model was implemented using JAVA and Microsoft Visual Studio.NET. In addition to forecasting, dForecast 2019 can be used for sensitivity analysis. Data collected for a period of sixty months was used to test the proposed model and the existing mono equation model. Both were then used for forecasting consumers' demand. The result showed that the mean absolute percentage error (MAPE) for the proposed model is 7.51%, while that of the existing model is 10.06%. This implies that the proposed model has higher prediction accuracy

***Corresponding Author**: **[1]Agholor, S.; sunday.agholor@gmail.com**

Introduction

Manufacturers are moving from a production push environment, which is largely focused on production efficiency, to a consumer pull environment, which is focused on meeting expected consumer demand (Agholor, 2004). This has led to an increased interest in demand forecasting processes as a means to better understand future consumer demand. Forecasting a product demand is crucial to any supplier, manufacturer or retailer. Forecasts of future demand will determine the quantities that should be purchased, produced and shipped. Demand Forecasts are necessary since the basic operations process, that is, moving from the suppliers of raw materials to finished goods in the customers' hands, takes time. Most firms cannot simply wait for demand to emerge and then react to it. Instead, they must anticipate and plan for future demand so that they can react immediately to customer orders as they occur. In other words, most manufacturers "make to stock" rather than "make to order". Forecasting, therefore, needs to be viewed as part of the overall management process.

A forecast is a statement of likely events in the future (Agholor, 2004). Predicting the future is vital in all facets of human endeavour. Nayal (2015) defines forecasting as a science of predicting the future. In the context of this study, a forecast is an estimate of the level of demand to be expected for a product or for several products for a period of time in the future. Accurate demand forecasts lead to efficient operations and high levels of customer service, while inaccurate forecasts will inevitably lead to the inefficient, high costs of operations and/or poor levels of customer services. Statistical models which predict future demand based on historical sales volume were the first attempts to automate this otherwise highly human-dependent sector of the supply chain. Existing studies relied on a mono equation multiple regression model whose prediction accuracy needs improvement. In this work, a combined two multiple regression equations is proposed to increase the prediction accuracy.

Literature Review

Why Forecasting Customers' Demand?

The advantages of forecasting customers' demand for a product or service cannot be overemphasized, especially in this industrial revolution age. The

benefits of forecasting consumers' demand for a product according to Magretta (1998) are:

1) Effective demand forecasting will help businesses reach their goals;
2) Customer's needs will be met by having goods available when they want them;
3) Company's revenue will increase since lost sales are minimized; and
4) Business cost structure will improve due to more efficient use of capital.

In addition to the benefits mentioned above, the other advantages as itemized by Barnett (1998) are:
1) Improved customer service levels;
2) Fewer backorders and lost sales;
3) Less inventory investment in safety stock;
4) Improved production planning processes; and
5) Earlier recognition of marketplace trends.

Regression Analysis

Regression is the study of the relationship among variables, a principal purpose of which is to predict, or estimate the value of one variable from known or assumed values of other variables related to it (Agholor, 2004). The regression analysis is a statistical tool that aims to explore the strength of the relationship between one dependent variable which is the response variable and one or many other changing variables known as the independent or the explanatory variables (Suri, 2006). It is a powerful technique that uses the values from historical data of one or more variables to develop a model that helps in predicting the value of the dependent variable (Suri, 2006). The regression analysis has different types according to the nature of the relationship between the dependent and the independent variables (Frost, 2015). However, two main types according to Frost (2015) exist. They are:
1) Linear Regression (Simple or Multiple); and
2) Non-linear Regression (Simple or Multiple).

Simple Linear Regression generates a forecast by linking one independent variable to the demand for a product. A typical simple linear regression model is given by the equation:

$Y=a+bX$……………………………………….(1)

where:
Y is the value of the dependent variable to be predicted or explained;
a is the constant that equals to the value of Y when X=0;
b is the coefficient of X. It is the slope of the regression line that can explain how much Y will change for each one-unit change in X; and
X is the value of the independent variable.

Models of this type do not incorporate the effects of other variables on demand. In an attempt to solve the problems created by simple linear regression, the multiple linear regression which generates a forecast by linking two or more independent variables to the demand for a product or a service was developed. A typical multiple linear regression model is given by the equation:

$Y=a+b_1X_1+ b_2X_2…+b_nX_n$ ……………………..(2)

where:
Y is the value of the response or the dependent variable to be predicted or explained;
a is the constant or the intercept;
b_1, b_2, …, bn are the regression coefficients; and
X_1, X_1,…,Xn are the independent variables.

The linear regression becomes non-linear when the variable(s) is/are raised to a certain power. According to McClave and Benson (1998), the method of fitting non-linear regression is identical to that of fitting linear regression, namely the method of least squares. However, simple or multiple regression models, whether linear or non-linear are mono equation approach to forecasting, and forecasts produced by mono equation regression models do not accommodate the contributions of other competitors in the market, which affects its prediction accuracy.

Linearity Test

According to Balance (2012), scholars agree that the linearity test is the most important test when fitting a regression model as it directly relates to the bias of the results of the whole analysis since it is used to ascertain whether the dependent variable is a linear function of the independent variables. If it is violated, all the estimates from the regression and its statistical output may be biased resulting in serious error in the predicted values (Balance, 2012).

Review of Some Research Works Based on Multiple Regression Analysis

A search in the literature showed that existing studies made use of multiple regression methods to develop forecasting models. As stated earlier, multiple regression models are mono equation approach to developing a forecasting model. To the best of our knowledge, there is no research work on forecasting using regression analysis that made use of more than one multiple regression equation. We present below some of the existing works found in the literature.

Wang *et al.* (2018) constructed a mathematical model using regression analysis to forecast the monthly electricity sales in Guizhou Province of China. The mono equation model developed is:

$Ei=0.417+1.304Ti+0.1961Ii$……………………(3)

where:
Ei is the forecasted value;

Ti is the monthly average temperature; and
Ii is the monthly economic growth index.

Chayalakshmi *et al*. (2018) fitted a multiple linear regression model for predicting boiler efficiency. The data for this work was obtained from JK Cements Pvt. Ltd., Muddapur, Lokapur taluk, Bagaikot district. The mono equation multiple linear regression model is:

$$BE=95.072-(0.714XL_1)-(0.894XL_2)-(0.714XL_3) \qquad (4)$$

where:
BE is Boiler Efficiency;
L_1 is loss due to dry flue gas;
L_2 is loss due to hydrogen in fuel; and
L_3 is loss due to moisture in fuel.

Dhakal (2018) developed a multiple regression model that can be used in rice production forecasting in Nepa. The mono equation multiple linear regression model is:

$$Production=-1619+5.26X(HA)-0.239X(RP)+0.321X(PH) \qquad (5)$$

where:
HA is Harvested Area;
RP is Rural Population; and
PH is Price at Harvest.

To ensure that organizations prepare and project well ahead of its manpower need, Aref and Sabah (2015) constructed a regression equation for forecasting the manpower needs of Ferdowsi University of Mashhad, Iran. The mono regression equation is:

$$Y=603.681+0.002(X1)+0.01(X2)+0.023(X3)-0.046(X4) \qquad (6)$$

where:
Y is the forecasted value (manpower needed);
X1 is the number of Bachelor degree students;
X2 is the number of Masters degree students;
X3 is the number of PhD students; and
X4 is the number of scientific papers.

A regression analysis to forecast the demand for new single family houses in USA was developed by Nayal (2015). The mono equation model is thus:

$$HS=-8579583+1904*PC-895621*UR+9187261*P1+2890891*P2 \qquad (7)$$

where:
HS is Home Sales;
PC is Personal Consumption;
UR is Unemployment Rate;
P1 is Period1; and
P2 is Period2.

The model below was a research work of Agholor (2015) in which he focused on the demand for pure water manufactured by XYZ company in Abeokuta. Data for this work was collected for a period of 6 years. The mono equation model thus formed is:

.

$$Y=40626+4.89X_1+76.52X_2-2.93X_3 \qquad (8)$$

where:
Y is the forecasted sales of paints;
X_1 is the price of the paint;
X_2 is the advertising budget; and
X_3 is the distribution cost.

Gan and Ahmad (2011) fitted a multiple linear regression model to forecast the balance of trade in Sabah, Malaysia. The mono equation multiple regression model is:

$$Y=-164.67+0.001X+0.001X_2-0.001X_3-0.003X_4+0.002X_5 \qquad (9)$$

where:
Y is the balance of trade;
X is exports of palm oil;
X_2 is exports of crude petroleum;
X_3 is imports of petroleum products;
X_4 is imports of motor cars, completely built-up; and
X_5 is exports of plywood plain

Power companies must be able to predict the peak power at their various stations in order to operate effectively. The peak power load is the maximum amount of power that must be generated each day to meet demand. In modeling the daily power load of a power company located in the Southern part of USA, Price (2000) fitted a non-linear regression model on a random sample of 25 summer days. The mono equation model thus formed is:

$$Y=385.05-8.29X+0.06X^2 \qquad (10)$$

where:
Y is the daily power demand and
X is the peak power.

The models under review are multiple regression models (linear/non-linear) and have helped to solve the problems associated with the simple regression model. However, they are all mono equation multiple regression models with some drawbacks. For instance, it failed to:
a) Incorporate the contributions of other competitors to the total market demand for the product;
b) Accommodate the market share of the company under focus;
c) Accommodate the prices of the products from other competitors in the market; and
d) Incorporate the power of the advertising costs of other competitors in the market, among others.

Methodology

Theoretical Framework

Assuming there are three (3) companies, X, Y, and Z manufacturing a product, say chewing gum, with

three different brand names representing the product from each of the companies, then each company will produce a certain quantity and send it to the market. The total consumers' demand is the aggregate of the products of the three companies, while the demand for each company's product is the company's market share. Thus, taking into consideration the contributions of various companies making the products under focus and the market share of the company of our interest, then we can make a forecast for the company of our interest from the relation:

ED=TCD*MS……(11)

where:

ED is the estimated (forecasted) demand.

TCD is the Total Consumers' Demand which is the aggregate of the products bought by the consumers from the various companies manufacturing the same product under focus.

MS is the Market Share of the company of our interest/focus.

Since the contributions of the various companies making the products can be obtained, we first build a multiple regression model for the aggregate of the products bought by the consumers as follows:

$TCD= a+b_1X_1+b_2X_2+…+b_nX_n$……(12)

where:

TCD is Total Consumers' Demand.

$X_1, X_2, ...X_n$ are independent (explanatory) variables which include average industry price of the product, average industry advertising cost, average industry cost of promotion sales, the average cost of product design, inflation rate, taxation, among others.

$b_1, b_2, …b_n$ are the coefficients of the independent variables.

Next, we consider the Market Share of the Company under study. Thus we have:

MS=RD/N……(13)

where:

MS is the Market Share, which is the ability of the company under focus to compete effectively with other companies producing the same product.

RD is the Relative Demand which is the consumers' demand for the product of the company under focus compared to the average industry demand.

N is the number of companies producing the products that add up to give TCD.

But RD can be obtained by building another multiple regression model as shown in equation (14).

$RD=b+c_1Y_1+c_2Y_2+….+c_nY_n$……...(14)

where:

RD is as explained in equation (13).

$Y_1, Y_2, …Y_n$ are the independent variables that affect the company under focus, which include relative price of the product, which is the company's price when compared to the average industry price, relative advertising, which is the company's advertising cost when compared to average industry advertising cost, customers' loyalty which refers to consumers' previous month's purchase, consumers' income, among others.

$c_1, c_2,…c_n$ are the coefficients of the independent variables.

Substituting (14) in (13), we have:

$MS=(b+c_1Y_1+c_2Y_2+….+c_nY_n)/N$……(15)

Thus, the proposed model will be a combination of the two multiple regression equations, that is, equations (12) and (15). Recall that the estimated demand is given in equation (11), reproduced here for clarity.

ED=TCD*MS……(11)

Substituting the two equations, that is, equation (12) for TCD and equation (15) for MS in equation (11), we obtain the forecasting model, which is a combination of two equations-(12) and (15). Hence, the proposed model is:

$ED=(a+b_1X_1+b_2X_2+…+b_nX_n)*((b+c_1Y_1+c_2Y_2+ C_3Y_3….+c_nY_n)/N)$……(16)

From equation (16), we have moved away from the mono multiple regression equation to a combined two multiple regression equations.

For comparison, the data collected from company "A" alone was used to develop a multiple regression model that can estimate the costumers' demand for company "A" products in line with existing methods as discussed in the literature review. Thus, the multiple regression equation developed is:

$FD=d+e_1Z_1+ e_2Z_2+…+e_nZ_n$……(17)

where:

FD is the forecasted demand using the mono equation multiple regression model.

d is the constant.

$Z_1, Z_2, …Z_n$ are the independent variables.

$e_1, e_2, … e_n$ are the coefficients of the independent variables.

Sensitivity Analysis

Equation (16) is used for estimating the consumers' demand for a product. From the prediction accuracy of equation (16) the actual demand for the consumers can be calculated. Thus, the gross income can be computed from the relation:

$GI = \frac{PA*ED*PP}{100}$ ……(18) and

NI=GI-TPC……...(19)

where: GI is Gross Income;

PA is Prediction Accuracy;
ED is Estimated Demand;
PP is the Price of the Product.
NI is Net Income;
TPC is the Total Production Cost.

The combination of equations (18) and (19) forms a very useful tool for sensitivity analysis. In this work, equations (16), (18) and (19) are collectively referred to as dForecast 2019.

Statistical Treatment

The data used to implement the theoretical framework were collected from ten (10) manufacturing industries in Ogun State of Nigeria for a period of 60 months (5 years), from 2014 to 2018. However, company "A" is the company of interest whose RD was developed. Thus, N in this study is ten (10). A test regression was run using SPSS with all independent variables included to establish the validity or otherwise of each variable and to enable us fit a mathematical representation of TCD, RD and FD. A linearity test was conducted by exploring the variables graphically in scatterplots to ascertain if a linear model is appropriate for describing the relationships. In addition to the scatterplots, the line of best fit was plotted between dependent and independent variables. The results from these two linearity tests showed that a linear relationship exists between the dependent and independent variables.

Implementation

The proposed model, *dForecast 2019* was implemented using JAVA and Microsoft Visual Studio.NET. The Mean Absolute Percentage Error (MAPE), which is a measure of prediction accuracy of a forecasting model was used to evaluate the prediction accuracy of the model in comparison with the existing mono equation multiple regression model.

Pilot Test

The pilot test was carried out in conjunction with the Marketing Manager of Company "A". The monthly estimated demand obtained from both the proposed model and that of existing model were compared with the actual sales obtained from the Marketing Department of Company "A". The results were used for further analyses.

Forecasted Results from the Proposed Model (dForecast 2019)

Table 1 shows the results from the proposed Model. From table 1, the prediction error showed that the proposed model is unbiased. A forecasting model is said to be biased when it has the persistent tendency to err in the same direction, that is, to consistently over-predict or under-predict demand. Further analysis with mean absolute percent error (MAPE) showed a percentage prediction error of 7.51%. This implies that the proposed model has 92.49% prediction accuracy.

Table 1: Results from the proposed Model (dForecast 2019)

Month	Year	Actual Demand (AD)	Estimated Demand (ED)	Prediction Error (AD-ED)	Absolute Percentage Error
Jan	2019	6,682	7,217	-535	8.00%
Feb	2019	6,721	6,249	472	7.02%
Mar	2019	7,996	7,516	480	6.00%
Apr	2019	8,231	8,889	-658	7.99%
May	2019	8,446	7,688	758	8.97%
Jun	2019	8,202	9,108	-906	11.05%
Jul	2019	7,708	8,479	-771	10.00%
Aug	2019	8,102	8,916	814	10.05%
Sep	2019	8,894	8,271	623	7.00%
Oct	2019	9,300	8,835	465	5.00%
Nov	2019	10,633	10,207	426	4.01%
Dec	2019	10,230	9,718	512	5.00%
Mean Absolute Percentage Error (MAPE)					7.51%

Forecasted Results from the existing mono equation Model

The result from the existing mono equation model is presented in table 2. From table 2, the prediction error showed that the existing mono equation model is unbiased, while the mean absolute percentage error is 10.06%. This implies that the existing mono equation multiple regression model has 89.94% prediction accuracy.

Table 2: Results from the existing Model

Month	Year	Actual Demand (AD)	Forecasted Demand (FD)	Prediction Error (AD-FD)	Absolute Percentage Error
Jan	2019	6,682	5,880	802	12.00%
Feb	2019	6,721	5,981	740	11.01%
Mar	2019	7,996	8,796	-800	10,01%
Apr	2019	8,231	7,408	823	10.00%
May	2019	8,446	7,458	988	11.70%
Jun	2019	8,202	9,200	-998	12.17%
Jul	2019	7,708	6,712	996	12.92%
Aug	2019	8,102	9,006	-904	11.16%
Sep	2019	8,894	7,912	982	11.04%
Oct	2019	9,300	8,308	992	10.67%
Nov	2019	10,633	9,773	860	8.09%
Dec	2019	10,230	9,231	999	9.77%
Mean Absolute Percentage Error (MAPE)					10.06%

From the two analyses, the proposed model with 92.49% prediction accuracy performs better than the existing mono equation model with 89.94% prediction accuracy. We therefore conclude that the proposed model is a good forecasting model with better prediction accuracy.

Results from Sensitivity Analysis

In addition to forecasting the consumers' demand, the proposed model can be used to carry out sensitivity analysis. In this case, we can compute the net income from gross income. Since the prediction accuracy is 92.49%, it then follows that the actual demand, that is, goods bought for the month under focus will be 92.49% of ED. Here, we presented two cases of sensitivity analyses, that is Case A and Case B.

Case A

Here we have two scenarios: Scenario 1 is where the price of the product is N400.00 and cost of advertisement is N120,000.00, while other costs such as labour, raw materials, transportation, etc is N1,105,000.00. In scenario 2, the price of the product is reduced to N370.00, while keeping the production cost of other variables the same as scenario 1. The consumers' demand for the product for the previous month is 8,712, while that of the total industry demand is 60,051. If the average industry price is N390, while the average industry advertisement cost is N100,000.00, what effect will it have on the profitability of the company? The output/screenshot from dForecast 2019 is shown in figure 1.

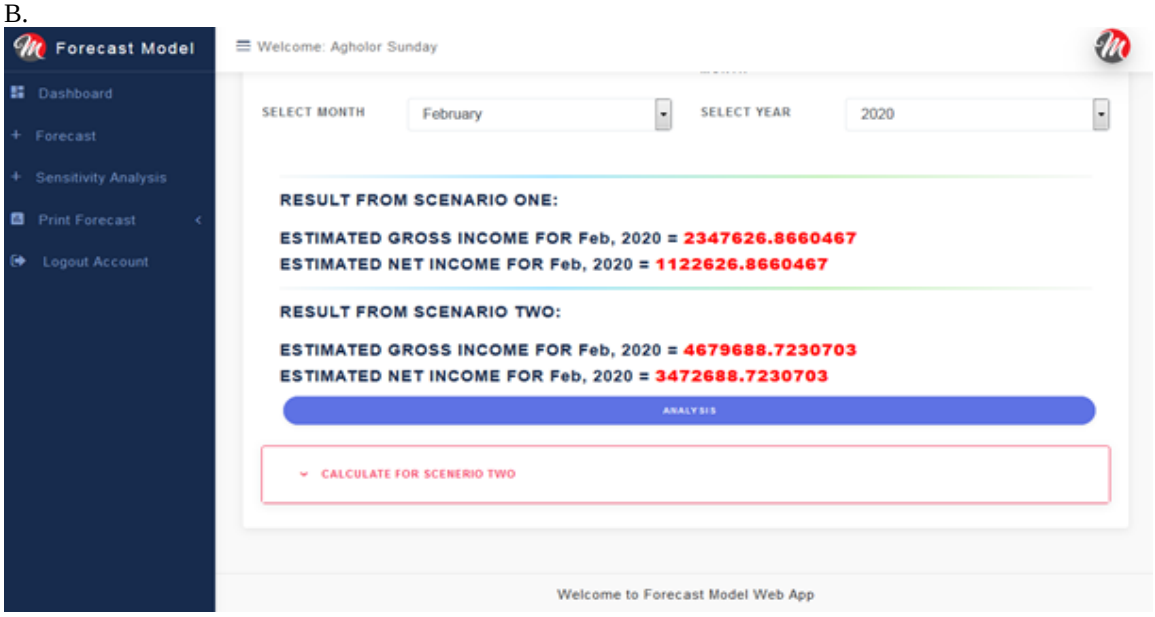

Figure 1: Case A Output/Screenshot from dForecast 2019

Comparing the result from Scenario 1 and Scenario 2 as presented in figure 1, it shows that reducing the price by N30.00 increases the consumers' demand for the product which resulted in Net Income of N1,122,626.86 and N3,472,688.72 for Scenario 1 and Scenario 2 respectively.

Decision
The Company is advised to take the option of reducing the price from N400.00 to N370.00 as this will bring more profitability to the company.

Case B
Again, we have two scenarios. Scenario 1 is where the price of the product is N400.00 and the cost of advertisement is N120,000.00, while other costs such as labour, raw materials, transportation, etc is N1,105,000.00. In scenario 2, the price of the product and the cost of advertisement are reduced to N370.00 and N102,000.00, while keeping the production cost of other variables the same as scenario 1. The consumers' demand for the product for the previous month is 8,712, while that of the total industry demand is 60,051. If the average industry price is N390, while the average industry advertisement cost is N100,000.00, what effect will it have on the profitability of the company? The output/screenshot from *dForecast* 2019 is shown in figure 2.

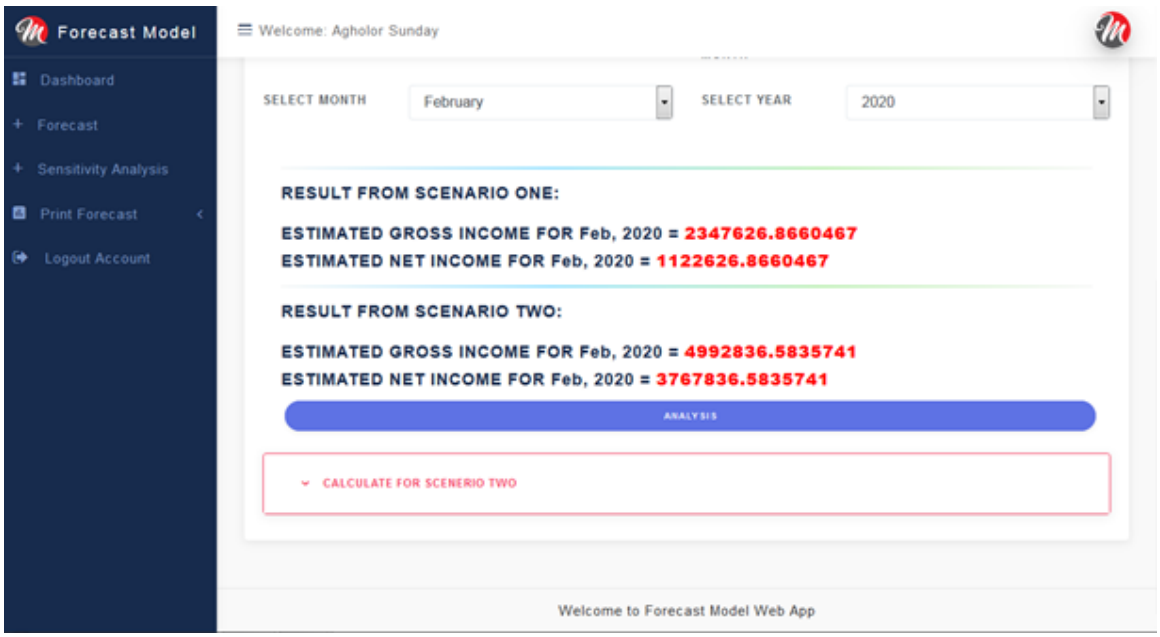

Figure 2: Case B Output/Screenshot from dForecast 2019

From figure 2, a reduction of the price from N400.00 to N370.00 and cost of an advertisement from N120,000.00 to N102,000.00 resulted in the net income of N1,122,626.00 and N3,767,836.58 for Scenario 1 and Scenario 2 respectively.

Decision*:* The Company is advised to embark on Scenario 2 with high profitability.

However, on the overall, the net income of Scenario 2 of Case A is N3,472,688.72, while that of Scenario 2 of Case B is N3,767,836.58. Hence, comparing the two cases, Case B scenario 2 is a better option for the company to embark upon.

Recommendations
Following the successes recorded by the model in terms of prediction accuracy and its usage in sensitivity analysis, the paper recommends the adoption of the model in future development of forecasting model using multiple regression analysis. Furthermore, the software (*dForecast 2019*) is recommended for companies that are interested in forecasting for their business growth and profitability.

Conclusion and Future Work
In this work, we took a departure from the much-used mono multiple linear regression equation to a combined two multiple linear regression equations. The result obtained from this novelty showed that the combined two multiple linear regression equations is a more promising model for forecasting consumers' demand for a product in a competitive market. In addition, it is a very useful tool for carrying out sensitivity analysis, which hitherto could not be done with the existing model.

Based on the great roles effective forecasting plays in all sectors of the economy, it is strongly recommended that more researches be carried out in this area through a multi-disciplinary approach. Finally, to increase the prediction accuracy, we suggest that more independent variables such as the income of buyers, inflation rate and exchange rate among others be added to the model.

References

Agholor, S. (2004). Design of Computer-Based Integrated Forecasting Model. Unpublished M.Sc Dissertation submitted to Computer Science Department, Federal College of Agriculture, Abeokuta, Nigeria

Agholor, S. (2015). Modelling and Forecasting the Demand for Pure Water in Abeokuta Metropolis, School of Science Seminar Series 3, Federal College of Education, Abeokuta, Nigeria

Aref, M. and Sabah, M. (2015). Manpower Planning for Demand Forecasting of Faculty Members using Trend Analysis and Regression, International Journal of Academic Research in Business and Social Sciences, Vol. 5, No. 2

Balance, D. (2012). Assumptions in multiple Regression, retrieved on 23/01/2018 from www.dianneballanceportfolio.com/uploads/1/2/8/2/

Barnett, F. W. (1998). Four Steps to Forecast Total Market Demand, Harvard Business Review, July-August

Chayalakshmi, C. L., Jangamshetti, D. S., and Sonoli, S. (2018). Multiple Linear Regression Analysis for Prediction of Boiler losses and Boiler Efficiency. In International Journal of Instrumentation and Control Systems, Vol. 8, No. 2, pp. 1-8.

Dhakal, C. P. (2018). Multiple Regression Model Fitted for Rice Production Forecasting in Nepal: A Case of Time Series Data. In Napalese Journal of Statistics, Vol. 2, pp. 89-98.

Frost, J. (2015). What is the F-test of overall significance in Regression Analysis? The Minitab Blog Website, retrieved on 23/01/2018 from www.blog.minitab.com/blog/adverture...

Gan, S. O. and Ahmad, S. (2011). Multiple Linear Regression to Forecast Balance of Trade. In Journal of Fundamental Sciences, Vol. 7, No. 2, pp. 150-155.

Magretta, J. (1998). The Power of Virtual Integration: An interview with Dell Computers, Harvard Business Review, March-April

McClave, J. T. and Benson, P. G. (1998). Statistics for Business and Economics, USA, Dellen Publishing Co.

Nayal, L. (2015). Regression Analysis to Forecast the Demand for new single family houses in USA. Culminating Projects in Mechanical and Manufacturing Engineering Department, Paper 15, St. Cloud State University

Price, S. (2000). A Handbook for Linear Regression, New York, Wiley.

Suri, A. (2006). A Linear Regression Model for Predicting Market Demand tool for Engineering Managers and Top Management

Wang, J., Liu, J., Li, T., Yin, S. and He, X. (2018). The Research of Regression Method for Forecasting Monthly Electricity Sales Considering Coupled Multi-factor, proceedings of IOP Conference Series: Earth and Environmental Sciences