Scholars Journal of Economics, Business and Management
e-ISSN 2348-5302
p-ISSN 2348-8875
Ogunleye TA *et al.*; Sch J Econ Bus Manag, 2014; 1(7):276-290

Econometric Modeling of Commercial Banks' Expenditure on the Sources of Profit Maximization in Nigeria

Ogunleye Timothy A.[1*], Olaleye Michael Olugbenga[2] and Solomon Adeoluwa Zaccheous[3]
[1*]Department of Budget, Planning, Research & Statistics, Ejigbo Local Government, State of Osun, Nigeria
[2,3]Department of Commerce and Economic Studies, Jomo Kenyatta University of Agriculture and Tech., Juja, Kenya.

**Corresponding Author: Ogunleye Timothy A; Email:* thompsondx@gmail.com

Abstract: This paper investigated the most profitable source(s) of profit maximization in commercial banks in Nigeria as a result of the available records from CBN. The data sets were based on the post-merger monthly income and expenditure of all the twenty-one licensed commercial banks in Nigeria for a period of eight years, equivalently ninety-six months. Commercial banks' sources of income were classified into five categories: interests on loans and advances, commission on turnover, electronic banking charges, commission on sales of third party institutions such as sales of institution forms, and finally income from other investments such as commission on treasury bills, money market instruments, capital projects financing and so on; all these are considered as exogenous variables for which commercial banks' expenditure is classified as endogenous variable. A number of assumptions underlying the use of econometric approach for modelling were confirmed using appropriate techniques. It was discovered that interests on loans and advances give the most profitable source of income while commission on sales of third party institutions happened to be the least means of profit. After proper variable scrutiny by stepwise selection procedure, a reliable econometric model was obtained for forecasting. The conclusions reached on this study resulted to a piece of very strong advice such that commercial banks should intensify efforts by all means to convince and encourage more customers on financial borrowings and, also increase the risks of financing capital projects with a view to maximizing profits in banking sector.
Keywords: Anderson-Darling Test, ANOVA, Box Plot, CBN, COT, Interests on Loans and Advances, Jarque-Bera Test, OLS, Profit Maximization, Stepwise Selection Procedure.

INTRODUCTION

An application of statistical technique on a set of mathematical equations of an economic related data with a view to estimating the parameters involved and drawing statistical inferences on the estimated parameters to facilitate correct decisions on economic phenomena is simply termed econometrics. In another dimension, we can still define the word 'econometrics as the social science in which the tools of economic theory, mathematics, and statistical inference are applied to the analysis of economic problems. Yes, it is an amalgam of economic theory, mathematical economics, economic statistics, and mathematical statistics [4].

The Nigerian banking industry, which is regulated by the Central Bank of Nigeria (CBN), is made up of deposit money banks referred to as commercial banks, development financial institutions and other financial institutions which include micro-finance banks, finance companies, bureau de change, discount houses and primary mortgage institutions. Specifically, banking industry consists of twenty-one commercial banks, five discount houses, five development finance institutions, fifty class 'A' bureau de change, five hundred and ninety-eight bureau de change, ninety-eight primary mortgage institutions, eighty-four finance companies and nine hundred and fourteen micro-finance banks. It is hoped that the present study, which is primarily focused on commercial banks alone, will model the commercial banks' expenditure on their sources of profit, will determine the amount of variation in the commercial banks' expenditure in connection with the available sources of income, and finally will determine the source with the highest profit.

Being a paper that discusses economic selection and modelling of sources of profit generation in banking sector, research revealed that sources of income in commercial banks are numerous; thus it has been classified into five major categories: interests on loans and advances, commission on turnover, electronic banking charges, commission on sales of third party institutions such as sales of institution forms, and finally income from other investments such as commission on treasury bills, money market instruments, etc; all these serve as exogenous variables while expenditure, which can be subdivided into various parts such as salaries, payment to service providers, annual replacement of equipment, expenses on recurrent expenditure, payment of dividend and bonus to shareholders as at when due, staff local and international trainings, and rent, is considered as endogenous variable. All these sources of expenditure are classified as one variable.

However, many people think that loan and advance are the same but are not. Though, both mean borrowings from one another, there are still lots of differences. Simply, loan means debt for personal or business purposes in which loan taker is responsible to return his taken money with interest; it is an amount borrowed from a bank or other institution that loans money. In this case, borrowers sign promissory note that states the terms and conditions of the loans and the length of time for repayment. Advance, on the other hand, is to get money from one another with mutual relationship. Though, in commercial banks, interests are charged on both loans and advances with a view to generating more income.

Moreover, COT, as abbreviated, means Commission On Turnover is another significant source of revenue in commercial banks. It is a charged levied on customers' withdrawals on their current and/or domiciliary accounts. Currently, most commercial banks in Nigeria are charging 0.5% per one thousand naira meaning that for every one thousand naira withdrawal, five naira is charged as COT. It is always calculated at the end of every month applying the percentage on customers' aggregate withdrawals from the first day to the last day of the month. The COT is a guaranteed income and as such is a very important source of bank revenue since customers mostly make withdrawals from their deposits. This gives the banks a two-way source of revenue as they (banks) charge interest on customers' deposits lent to other people and consequently charge on amount withdrawn by the customers themselves. It is an approved source of revenue for banks according to CBN guidelines. It should be noted that some commercial banks may be flexible with their COT charges depending on the turnover of customers' accounts.

Reasonably, one would expect some charges on any electronic alerts received either on phone as message or electronic mail. Some token amounts are charged in respect of any transaction alert. Anyway, this is optional to customers; this charge is made on customers' accounts be it savings, current and/or domiciliary accounts on formal request for electronic alerts by the customers. Also, some amounts are being charged as commission on the sales of third party institutions such as sales of West Africa Examination Council (WAEC) forms, Joint Admission Matriculation Board (JAMB) forms, National Examination Council (NECO) forms and National Board for Technical Education Board (NABTEB) forms as well as any other institution forms. Meanwhile, income realized from financing capital projects, charges from keeping valuables, commission on treasury bills, money market instruments, and so on are all examples of income from other investments, which is the fifth source of revenue in commercial banks.

METHODOLOGY

In any econometric research, four important stages are expected to pass through in order to conclude meaningfully. These stages are: specification of the model, estimation of model parameters, evaluation of estimates and finally, evaluation of the forecasting power of the estimated model. The first and the most important step to take in attempting the study of any relationship between variables is to express this relationship in mathematical form, that is, to specify the model with which the economic phenomenon will be explored empirically. It involves the determination of the endogenous variable (regressand) as well exogenous variables (regressors). It also involves the determination of the mathematical form of the model. Scatter diagrams may be drawn to determine the shape and nature of the model. These diagrams will give more understanding about the expected results [4].

After the model has been specified (formulated), we must proceed with its estimation. In other words, we must obtain numerical estimates of the coefficients of the model. The estimation of the model is a purely technical stage which requires knowledge of various econometric methods, their assumptions and the economic implications for the estimates of the parameters. After the estimation of the model, we will proceed with the evaluation of the results of the calculations, that is, with the determination of the reliability of these results. The evaluation consists of deciding whether the estimates of the parameters are theoretically meaningful and statistically satisfactory. For this purpose, we use various criteria which may be classified into two groups: statistical criteria (1st-order tests) and econometric criteria (2nd-order tests). Statistical criteria are determined by statistical theory and aim at the evaluation of the statistical reliability of the estimates of the parameters of the model while econometric criteria are set by the theory of econometrics and aim at the investigation of whether the assumptions of the econometric method employed are satisfied or not in any particular case.

One of the objectives of any econometric research is to obtain good numerical estimates of the coefficients of economic relationships and to use them for the prediction of the values of economic variables. Therefore, forecasting is one of the prime aims of econometric research. Before using an estimated model for forecasting the value of the endogenous variable, we need to assess the predictive power of the model. It is conceivably possible that the model is economically meaningful and statistically and econometrically correct for the sampled period for which the model has been estimated, yet it may very well not be suitable for forecasting due, for example, to rapid change in the structural parameters of the relationship in the real world. The final stage of any applied econometric research is the investigation of the stability of the estimates, their sensitivity to changes in the size of the sample. We must establish whether the estimated function performs adequately outside the sample of data.

DATA GRAPHICS

The preliminary examination of most data is facilitated by the use of diagrams. Thus, diagrams prove nothing but bring outstanding features readily to the eyes; they are therefore no substitute for such critical tests as may be applied to the data, but are available in suggesting such tests, and in explaining the conclusions made on them. Several methods of data graphics are available in the literature but only two of them will be used in the present study.

Scatter Diagram

This is a diagram that shows the location of points (values) of both endogenous and exogenous variables on a rectangular coordinate system. Scatter Diagram enables us to study the data and gives more understanding about the result [5]. It is the first step (procedure) to be considered in analyzing a set of economic data using econometrics approach before embarking on any statistical test with a view to substantiating any conclusion made. The values of the exogenous variable are plotted against that of the endogenous variable. If all points in this scatter diagram seem to lie near a line, then the indication is that there is a linear relationship between the two variables involved. Otherwise, we can conclude that no linear relationship exists. With the aid of *R* statistical software, scatter diagrams are plotted accordingly.

Box Plot

Box plots are excellent tools for conveying location and variation information on data sets, particularly for detecting and illustrating location and variational changes between different groups of data [6]. During its application, response variables are plotted on vertical axis while factors of interest are plotted on horizontal axis. Means or medians as well as lower and upper quartiles are parameters required to put box plot in place thereby assumed that it is an effective tool for summarizing large quartiles of information and finally used for detecting outlier. Summarily, box plot is a very useful way to display data in which minimum, maximum, lower and upper quartiles including the median as well as outlying value(s) are obtained in a rectangular box aligned either horizontally or vertically. In this study, *R* statistical package was used to obtain the box plot.

Some Assumptions of Econometric Modelling

It is assumed that a linear regression model should be linear in both parameters and variables, exogenous variables are assumed to be non-stochastic, that is, values of regressors are considered fixed in repeated samples. Another important assumption in econometric modelling is such that expected value of the random disturbance term (residual) is zero, and also the same residual should equal variance (homoscedasticity). There should be neither multicollinearity nor autocorrelation problems between the disturbances, and finally there should be absence of outliers in the data sets [1].

Statistical Procedures for Confirmation of Normality Assumption

This is a very powerful assumption in econometric modelling. A normality check is a statistical process used to determine if a sample or a group of data fits a standard normal distribution; it can be performed either mathematically or graphically. Before the use of analysis of variance technique, residuals of the model are expected to behave well in such a way that values of the disturbances should not in any way far away from normal distribution. Two major approaches are considered in this study to verify normality assumption: **Anderson-Darling** Normality Test and **Jarque-Bera (JB)** Test of Normality.

Anderson-Darling Normality Test:

This test is used to test if a sample of data comes from a population with a specific distribution usually normal distribution. It is a modification of the Kolmogorov-Smirnov test and gives more weight to the tails than the Kolmogorov-Smirnov test. The Kolmogorov-Smirnov test is distribution-free test in the sense that the critical values do not depend on the specific distribution being tested but the Anderson-Darling test makes use of a particular distribution most especially normal distribution in calculating critical values. **A-squared** is the test statistic for Anderson-Darling Normality test; it is a measure of how closely a data set follows the normal distribution [2]. The null hypothesis for this test is that the data is normal. The statistic is stated as follows:

$$A^2 = -\frac{1}{n}\left[\sum_{i=1}^{n}(2i-1).\{\log z_i + \log(1-z_{n+1-i})\}\right] - n$$

For:

$$z_i = \Phi\left(\frac{x_{(i)}-\bar{x}}{s}\right), \quad \Phi(x) = \int_{-\infty}^{x}\frac{1}{\sqrt{2\pi}}\ell^{-\frac{1}{2}u^2 du},$$

where $x_{(1)} \le x_{(2)} \le \ldots \le x_{(n)}$ are the ordered observations, and s^2 is the sample variance. The null hypothesis of normality is rejected for large values of A^2 indicating that if we get **A-squared** that is fairly large, then we will get a small *p-value* and thus reject the null hypothesis. Hence, small **A-squared** value implies large *p-value*, meaning that null hypothesis cannot be rejected. This formular is embedded in *R*-statistical software.

Jarque-Bera (JB) Test of Normality:

The JB Test of Normality is an asymptotic, or large sample, test. It is based on the Ordinary Least Square (OLS) residuals. This test first computes the **Skewness** and **Kurtosis** measures of the Ordinary Least Square (OLS) residuals and uses the following test statistic:

$$JB = n\left[\frac{S^2}{6} + \frac{(K-3)^2}{24}\right],$$

where n is the sample size, S is the Skewness coefficient, and K is the Kurtosis coefficient. For a normally distributed variable, $S = 0$ and $K = 3$. Therefore, the JB Test of Normality is a test of joint hypothesis that the values of S and K should be approximately zero and three respectively. In that case, the value of the JB statistic should be approximately zero. Hence, under the null hypothesis that the residuals are normally distributed, Jarque and Bera showed that asymptotically the JB statistic follows the chi-square distribution with 2 degree of freedom. If the computed *p-value* of the JB statistic in an application is sufficiently low, which is likely to happen if the value of the statistic is very different from zero, one can reject the hypothesis that the residuals are normally distributed. But if the *p-value* is reasonably high, which is likely to happen if the value of the statistic is close to zero, we do not reject the normality assumption [9].

Statistical Procedure for Confirming Homoscedasticity Assumption

The term *'homo'* means 'the same' while *'hetero'* means *'different'* [8] therefore, homoscedasticity assumption means that the variance of each residual should be the same throughout. If the errors (residuals) fail to possess equal (but sometimes unknown) variance, the reliability of application of analysis of variance technique may be badly affected. Simply, the assumptions are such that residuals are normally, identically and independently distributed with mean zero and constant but unknown variance leading to the test that all samples came from populations with identical variances [11]. For this case, we employ spearman's rank correlation technique.

Spearman's Rank Correlation Test for Homogeneity of Variances:

This is another way to test whether or not the data set is homogeneous. The statistical procedures involved are such that we obtain the residuals and compute the spearman's correlation coefficients between the residuals and each of the explanatory variables. The simple interpretation is that if none of the coefficients is significantly high, it is an indication that the errors (disturbances) are homogeneous. But on the other way round, if any of these coefficients is significantly high, it shows the data sets are heterogeneous.

Statistical Procedure for Confirmation of Orthogonality Assumption

One of the problems being encountered by the estimates of linear econometric model is the existence of intercorrelations among the exogenous variables. Simply, when there is a complete absence of linear relationship among the predictor variables, they are said to be orthogonal. Thus, orthogonality means complete absence of linear relationship among the exogenous variables. This problem is also known to be multicollinearity; a situation whereby two or more exogenous variables are strongly intercorrelated. Hence, interpretation of the multiple regression models depend implicitly on the assumption that the predictor variables are not strongly interrelated [1]. It is usual to interpret a regression coefficient as measuring the change in the response variable when the corresponding predictor variable is increased by one unit and all other predictor variables are held constant. This interpretation may not be valid if there are strong linear relationships among the predictor variables. It is always conceptually possible to increase the value of one variable in an estimated regression equation while holding the others constant. The seriousness of the effects of multicollinearity seems to depend on the degree of intercorrelation as well as on the overall correlation coefficient. Thus, one might suggest that the standard errors, the partial correlation coefficients and the total R^2 may be used to test for multicollinearity. Yet, none of these criteria by itself is a satisfactory indicator of multicollinearity because of the following reasons:

(i) Large standard errors do not always appear with multicollinearity. (Cobb-Douglas Production Function is very good evidence). Furthermore, large standard errors may arise for various reasons and not only because of the presence of linear relationships among the explanatory variables.

(ii) The intercorrelations of the explanatory variables need not be high for the values of regression coefficients and their standard errors to be affected badly, that is, $r_{x_i x_j}$ is not an adequate criterion by itself.

(iii) The overall R^2 may be high (relative to $r_{x_i x_j}$) and yet the results may be highly imprecise and insignificant with wrong signs and/or large standard errors [5].

However, a combination of all these criteria may help the detection of multicollinearity. In order to gain as much knowledge as possible as to the seriousness of multicollinearity, we suggest the application of Farrar-Glauber Test.

Farrar-Glauber Chi-squared Test:

A statistical test for multicollinearity has been developed by Farrar and Glauber in one of their books tagged '*Multicollinearity in Regression Analysis'* in 1967. It is really a set of three tests. The first test is *Chi-Square test* for the detection of the existence and severity of multicollinearity in a function. The second test is an *F test* for locating which variables are multicollinear while the third is a *t test* for finding out the pattern of multicollinearity, that is, for determining which variables are responsible for the appearance of multicollinear situations. Only the first test would be considered in this study because we are interested only to know if there exist collinear variables. Farrar-Glauber considered multicollinearity in a sample as a departure of the predictor variables from orthogonality. Therefore, a chi-square test for the presence and severity of multicollinearity in a function with several explanatory variables is outlined as follows:

$$\chi^2_{cal} = -\left\{(\mathrm{n}-1)-\frac{1}{6}(2p+5)\right\}.\mathrm{In}|\mathrm{R}|,$$

The null hypothesis is such that no multicollinearity exists, that is, exogenous variables are orthogonal.

Statistical Procedure for Confirmation of Existence of Serial Correlation Problem

One of the standard assumptions in the econometric modelling is that the error terms (residuals) should be uncorrelated. Correlation in the error term suggests that there is additional information in the data that has not been exploited in the current model. Autocorrelation problem arises when the errors fail to be independent of each other [7]. It is expected that errors (residuals) should be independent in any situation, but if this condition does not hold, we say there is problem of autocorrelation in the data sets. When the observations have a *natural* sequential order, the correlation is referred to as *autocorrelation.*

Durbin-Watson Statistic:

This is a test that is aimed at determining whether there is dependency among the successive values of the error term. The most reliable and mostly used test for detecting existence of autocorrelation is Durbin-Watson Test having the statistic:

$$\boldsymbol{d} = \frac{\sum_{t=2}^{n}(e_t - e_{t-1})^2}{\sum_{t=1}^{n} e_t^2} \quad \text{and/or} \quad r = \frac{\sum_{t=2}^{n} e_t e_{t-1}}{\sum_{t=1}^{n} e_t^2}$$

An approximate relationship between ***d*** and $\hat{r}$ is ***d*** $\cong 2(1-\hat{r})$ showing that ***d*** has a range of **0** to **4**. Recall that the statistic ***d*** is used for testing the null hypothesis $H_0 : r = 0$ against an alternative $H_1 : r > 0$. Note that when ***r*** ≅ **0**, the errors are uncorrelated. Hence, from above, we can obtain ***r*** from the relation: ***d*** $\cong 2(1-\hat{r})$ [3].

Computational Procedures for Estimation of Least Square Parameters

Since it has been established that the sources of income in commercial banks in Nigeria are grouped into five categories: interests on loans and advances (x_1), commission on turnover (x_2), electronic banking charges (x_3), commission on sales of third party institutions such as sales of institution forms (x_4), and finally income from other investments such as commission on treasury bills, money market instruments, and so on (x_5), for which commercial banks' expenditure is classified as endogenous variable, the appropriate linear model is stated as follows:

$$y_i = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + \beta_5 x_5 + \ell_i \quad \text{(a)}$$

In matrix form, we have the model as follows:

$$Y = X\beta + \ell \quad \text{(b)}$$

From (b), we have:

$$\ell = \mathrm{Y} - \mathrm{X}\beta$$

Remember that the optimization problem is minimization of Error Sum of Squares (SSE) given by:

$$SSE = \ell^1 \ell$$

$$SSE = (\mathrm{Y} - \mathrm{X}\beta)^1 (Y - X\beta)$$

$$SSE = (\mathrm{Y}^1 - X^1\beta^1)(Y - X\beta)$$

$$SSE = \mathrm{Y}^1 Y - Y^1 X\beta - X^1\beta^1 \mathrm{Y} + X^1\beta^1 X\beta$$

$$SSE = \mathrm{Y}^1 Y - 2X^1\beta^1 \mathrm{Y} + X^1\beta^1 X\beta \ \{\text{for } \mathrm{Y}^1 X\beta \equiv X^1\beta^1 \mathrm{Y}\}$$

Taking the partial derivative of SSE with respect to β:

$$\frac{\delta SSE}{\delta\beta} = -2X^1Y + \left[X^1X\beta + X^1X\beta^1\right]$$

$$\frac{\delta SSE}{\delta\beta} = -2X^1Y + 2X^1X\beta \;\left\{ for\; X^1X\beta \equiv X^1X\beta^1 \right\}$$

Equate the Partial Derivative to zero in order to minimize error:

$$-2X^1Y + 2X^1X\beta = 0$$

$$-2X^1Y = -2X^1X\beta$$

$$X^1Y = X^1X\beta \;\left[Normal\,Equation\right]$$

Therefore:

$$\left.\begin{aligned} \hat{\beta} &= \frac{\left(X^1Y\right)}{\left(X^1X\right)} \\ \hat{\beta} &= \left(X^1X\right)^{-1}.X^1Y \end{aligned}\right\} \quad \text{......} \quad (c)$$

Equivalently:

$$\begin{bmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \\ \hat{\beta}_2 \\ \hat{\beta}_3 \\ \hat{\beta}_4 \\ \hat{\beta}_5 \end{bmatrix} = \begin{bmatrix} n & \sum x_1 & \sum x_2 & \sum x_3 & \sum x_4 & \sum x_5 \\ \sum x_1 & \sum x_1^2 & \sum x_1x_2 & \sum x_1x_3 & \sum x_1x_4 & \sum x_1x_5 \\ \sum x_2 & \sum x_1x_2 & \sum x_2^2 & \sum x_2x_3 & \sum x_2x_4 & \sum x_2x_5 \\ \sum x_3 & \sum x_1x_3 & \sum x_2x_3 & \sum x_3^2 & \sum x_3x_4 & \sum x_2x_5 \\ \sum x_4 & \sum x_1x_4 & \sum x_2x_4 & \sum x_3x_4 & \sum x_4^2 & \sum x_4x_5 \\ \sum x_5 & \sum x_1x_5 & \sum x_2x_5 & \sum x_3x_5 & \sum x_4x_5 & \sum x_5^2 \end{bmatrix}^{-1} . \begin{bmatrix} \sum y \\ \sum x_1y \\ \sum x_2y \\ \sum x_3y \\ \sum x_4y \\ \sum x_5y \end{bmatrix}$$

Statistical Procedure for Overall Test of Econometric Model using ANOVA Technique

ANOVA, as the name implies, is the acronym for Analysis of Variance. The aim of this method is to split the total variation of a variable (around its mean) into components which may be attributed to specific (additive) causes. To simplify the analysis, we will assume that there is only one systematic factor which influences the variable being studied. Any variation not accounted for by this (explanatory) factor is assumed to be random (or chance) variation, due to various random happenings. When we have a series of values of endogenous variable and the corresponding values of the exogenous variables, therefore, the Analysis of Variance (ANOVA) method concentrates on the values of the endogenous and studies their variation. However, whatever the relationship of the data being studied, the Method of Analysis of Variance (ANOVA) reduces the estimation of the two variances, and the comparison of these variances in order to establish whether the difference between them is statistically significant, or whether it is due to chance, in which case we conclude that there is no real difference between the variance-estimates [11].

Hence, the comparison of any two variances is implemented by the *F-statistic* and the *F-tables*. We simply define *F-statistic* as the ratio of any two independent estimates of variances, which can be obtained from sample data. Each estimate involves some loss of degrees of freedom. If we have any two independent variance estimates with v_1 *and* v_2 degrees of freedom respectively, their ratio has the *F-distribution* with v_1 *and* v_2 degrees of freedom. For this reason, *F* is called the variance ratio. The test aims at finding out whether the exogenous variables (x_1, x_2, …, x_p) do actually have any significant influence on the endogenous variable. Formally the test of the overall significance of the regression implies testing the null hypothesis:

$$H_0 : \beta_i = 0 \quad \text{Vs} \quad H_1 : \beta_i \neq 0, \text{for at least one i}$$

If the null hypothesis is true, that is, if all the true parameters are zero, then there is no linear relationship between endogenous and exogenous variables, equivalently we can say that the model is not statistically significant. If otherwise, then there is linear relationship between endogenous and exogenous variables, equivalently we can say that at least one of the model parameters is statistically significant. The test of the overall significance can be carried out using the

Analysis of Variance (**ANOVA)** table demonstrating all the calculations as follows:

Table-1: ANOVA Table (Theoretical)

Source of Variation	Degree of Freedom	Sum of Squares	Mean Square	F_{cal}
Regression	$p-1$	$SSR = (\hat{\beta}_0\ \hat{\beta}_1\ \hat{\beta}_2\ \hat{\beta}_3\ \hat{\beta}_4\ \hat{\beta}_5) \cdot \begin{bmatrix} \sum y \\ \sum x_1 y \\ \sum x_2 y \\ \sum x_3 y \\ \sum x_4 y \\ \sum x_5 y \end{bmatrix} - n\left[\sum y / n\right]^2$	$MSR = \frac{SSR}{p-1}$	$F_{cal} = \frac{MSR}{MSE}$
Error	$n-p$	$SSE = \sum y^2 - (\hat{\beta}_0\ \hat{\beta}_1\ \hat{\beta}_2\ \hat{\beta}_3\ \hat{\beta}_4\ \hat{\beta}_5) \cdot \begin{bmatrix} \sum y \\ \sum x_1 y \\ \sum x_2 y \\ \sum x_3 y \\ \sum x_4 y \\ \sum x_5 y \end{bmatrix}$	$MSE = \frac{SSE}{n-p}$	
Total	$n-1$	$SST_0 = \sum y^2 - n\left[\frac{\sum y}{n}\right]^2$	-	

***Note that *p* is the number of all variables involved in the study, which is 6.**

The decision rule is that we compare the calculated *F* value with that obtain from statistical table, and reject the null hypothesis if the calculated *F* value is greater than the tabulated *F* value or equivalently, we reject the null hypothesis if *p-value* is less than the chosen level of significance usually 0.05.

Statistical Procedure for Computation of Confidence Interval

Although hypothesis testing is a useful procedure to test for the significance of the parameters, it sometimes does not tell the entire story [4]. It is often preferable to provide an interval within which the value of the parameter(s) in question would be expected to lie. These interval statements are called confidence intervals. To construct confidence intervals for the regression parameters, we also need to assume that the *residuals* are normally distributed with mean zero and constant but unknown variance. Consequent upon these assumptions, the $100(1-\alpha)\%$ confidence interval for β_i is given by:

$$\Rightarrow \hat{\beta}_i \pm t_{\alpha/2,(n-p)} \cdot \sqrt{Var(\hat{\beta}_i)}$$

Statistical Procedures for Computation of Coefficients of Correlation and Determination

There are various methods for measuring the relationship existing between/among economic variables out of which include correlation analysis. The term correlation may be defined as the degree of relationship existing between two or more variables. However, the degree of relationship existing between two variables only is simply called simple correlation; the degree of relationship connecting three or more variables is called multiple correlation while the degree of relationship between two variables only keeping one or more variables constant is simply known as partial correlation. But we shall discuss only the first two correlation coefficients.

Simple Correlation

As earlier defined, simple correlation is the degree of relationship that exists between only two variables where one is dependent and the other one is independent. This form of correlation may be linear; when all points of dependent and independent variables on a scatter diagram seem to cluster near a straight line, or non-linear; when all points seem to lie near a curve. There are some assumptions associated with simple correlation which include the following: interval scale of measurement is assumed; linearity in relationship always exists; all variables involved must be normally distributed and the level of relationship must be the same throughout the range of the data. Two variables, however, may have a positive correlation, a negative correlation, or they may be uncorrelated (zero or spurious correlation). This holds for both linear and non-linear correlations. However, two variables are said to be positively correlated if they tend to change together in the same direction, that is, if they tend to increase or decrease together. In calculation, the coefficient of positively simple correlated variables fall within the range of $0 \prec r \leq 1$. The closer the value of *r* to one, the stronger the relationship becomes and, the vice-versa. Therefore, Karl Pearson's/Product Moment Correlation Coefficient between the two variables x and y can be obtained by the formular given as follows:

$$r_{xy} = \frac{n\sum x_i y_j - \sum x_i \sum y_j}{\sqrt{\left[n\sum x_i^2 - \left(\sum x_i\right)^2\right]\left[n\sum y_j^2 - \left(\sum y_j\right)^2\right]}}$$

In the present study, Karl Pearson's/Product Moment Correlation will be used to find the degree of relationship between the two variables involved ignoring the rank correlation method. This is because Karl Pearson's/Product Moment Correlation Coefficients are parametric indicating that the assumptions for its application are stronger and more reliable.

Multiple Correlation

In Statistics, Multiple Correlation is a linear relationship among more than two variables. It is measured by the Coefficient of Multiple Determination, denoted by R^2, which is a measure of the fit of a linear regression. The square root of coefficient of multiple coefficient of determination gives multiple correlation coefficient. Multiple correlation is the type of correlation (degree of relationship) between the endogenous variable and a set of exogenous variables. It has similar characteristics as that of the simple correlation in terms of interpretation of signs and magnitudes. The only difference is that it measures the degree of relationship between endogenous variable and more than one exogenous variables whereas simple correlation measures the degree of relationship between two variables **only**. It's denoted by the capital letter 'R' and computed as follows:

$$R = \sqrt{1 - \frac{SSE}{SST_0}}$$

Computation of Variance Inflation Factor (VIF)

This measures how the variance of a particular regressor (exogenous variable) inflates or increases in association with other regressors. The interpretation is that if the value of Variance Inflation Factor falls between 5 and 10, then there is presence of multicollinearity in data sets. Hence, the computational procedure is stated as follows:

$$VIF = \frac{1}{1 - R^2}$$

Statistical Procedures for the Selection of the Best Econometric Model

In building econometric model, it is possible that the full model might not be the best model; though it may be significant but not adequate. It may be too expensive to manage or there may be presence of high intercorrelated independent variables which may only add little to the predictive power of the model. Hence, the problem then is how to reduce the number of exogenous variables to be used in the final model as reasonable as possible. Several methods are available in the literature to perform the task of selection of the best equation but only one of these methods shall be discussed.

Stepwise Selection Procedure

This procedure is essentially a forward selection procedure but with the addition that at each stage there is possibility of deleting a variable. That is, a variable that entered in an earlier step may be deleted at a later stage. The calculations made for inclusion and deletion of variables are as for forward selection and backward elimination respectively. The first step is to start with the regression equation containing **no** variables in the model. After this step, as for forward selection, select the variable with highest correlation with the endogenous variable as the first variable to enter the regression equation and test the significance of the resulting equation. If not significant, then the model with only the constant term is the best. Hence, the third step is to select the next variable with the highest correlation with the endogenous variable as the variable to enter and obtain the new regression equation and test for the significance of the regression coefficient of the new variable. If not significant then the previous equation is the best. Finally, test for significance of all the regression coefficients. If any of them is not significant, then drop the corresponding and go back to the preceding step. Thus, these procedures are embedded in the *R* statistical software [10].

EMPIRICAL RESULTS

In this section, we shall make use of all the formulae and procedures highlighted in the preceding subheading and discuss the results of the analyses one after the other. The computations of results of analyses shall be done with the aid of *R* statistical software/package with a view to reducing the tedium of calculations and, also to having accurate and reliable results.

Justification for Data Graphics

We have discussed extensively in the previous section what data graphics is all about. In this subheading, we wish to present reasons why data graphics is necessary in the present research. It is intended to show preview of the kind of relationships that exist between endogenous variable and each of the exogenous variables; this is done by plotting scatter diagrams and, also to check whether or not outlier is present in the data sets with the help of box plot.

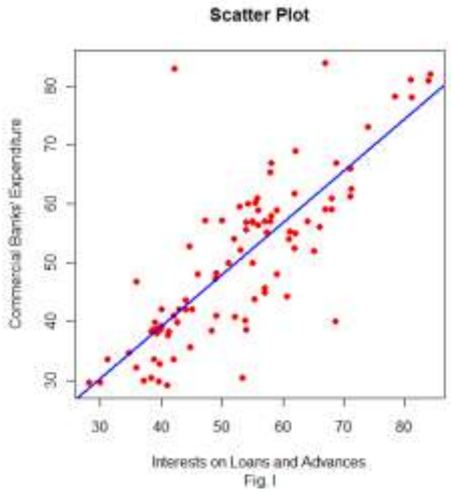

Fig-1: Scatter Diagram showing Relationship between Commercial Banks' Expenditure and Interests on Loans and Advances.

Interpretation of Figure 1: Since most of the points seem to lie near the straight line, it is an indication that there is linear relationship between commercial banks' expenditure and interests realized from loans and advances within the ninety-six months period. The diagram shows that there exists a positive correlation between the two variables involved.

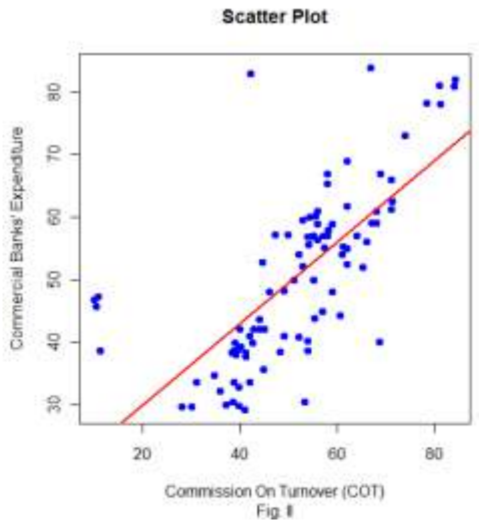

Fig-2: Scatter Diagram showing Relationship between Commercial Banks' Expenditure and Commission On Turnover (COT).

Interpretation of Figure 2: From the figure above, it shows that there is a linear relationship between commercial banks' expenditure and commission on turnover; this fact is established as a result of nearness of the points to the straight line drawn. The preview of this diagram indicates that a positive correlation exists.

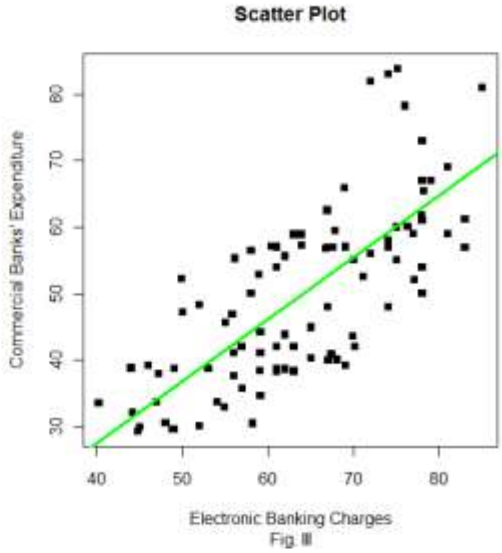

Fig-3: Scatter Diagram showing Relationship between Commercial Banks' Expenditure and electronic banking charges.

Interpretation of Figure 3: It can be depicted from the figure above that linear relationship exists between commercial banks' expenditure and electronic banking charges. The implication of this fact is that electronic banking charges has positive contribution to the expenditure incurred in commercial banks in Nigeria.

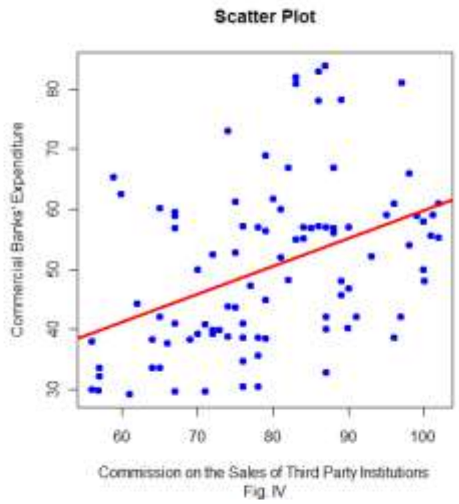

Fig-4: Scatter Diagram showing Relationship between Commercial Banks' Expenditure and Commission on the Sales of Third Party Institutions

Interpretation of Figure 4: Since most of the points are scattered around the straight line, it is an indication that the level of linear relationship between commercial banks' expenditure and commission on the sales of third party institutions is **very low** over the period of ninety-six months. The diagram shows that the degree of correlation is nothing to write home about.

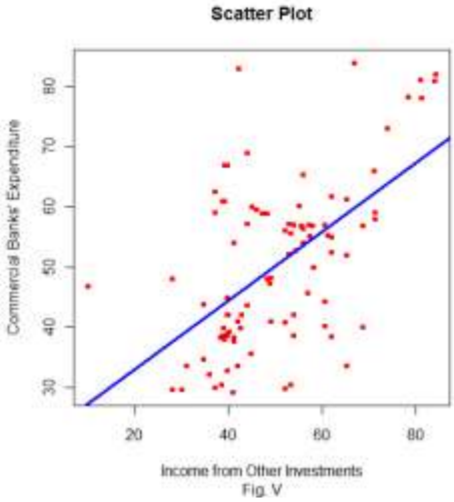

Fig-5: Scatter Diagram showing Relationship between Commercial Banks' Expenditure and Income from Other Investments.

Interpretation of Figure-5: From the look of things, we can deduce that there is a positive link between commercial banks' expenditure and income from other investments. The implication of this statement is that income realized from other investments such as capital projects' financing, commission on treasury bills, money market instruments and so on, determines the level of expenditure in commercial banks in Nigeria.

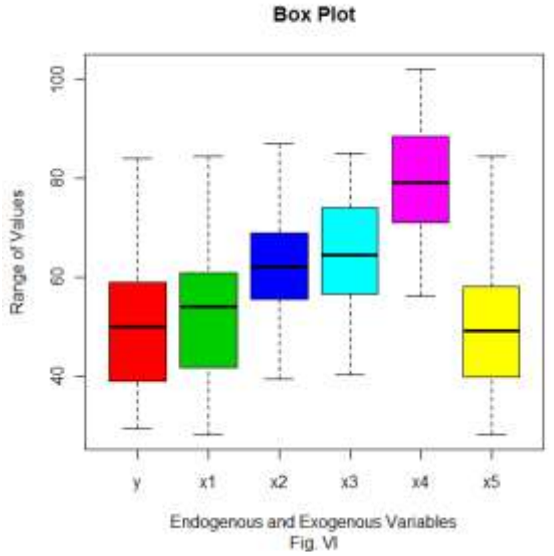

Fig-6: Box Plot for detecting the presence of outliers within the variables involved.

Interpretation of Figure-6: From all indications, reasonably one can read and observe that there is no outlier in the data sets used for this study. Outliers, in Statistics, refer to relatively small or large values which are considered to be different from, and not belong to, the main body of data [12]. It is very obvious that there exists no outlying observation in the data sets, and this guaranteed the reliability of the results of analyses.

Verification of Normality Assumption on the Data Sets

As discussed previously, two important approaches were used to test for normality assumption on the data sets: Anderson-Darling and Jarque-Bera Normality Tests. The table below shows the results of these analyses when *R* statistical software was used.

Table-2: Table of Application of Anderson-Darling and Jarque-Bera Tests of Normality

Required Information	Anderson-Darling Test	Jarque-Bera Test	Remarks
Sample Size	$n = 96$	$n = 96$	Selected number of months to obtain the data for the study
Null Hypothesis	Data sets are normal	Data sets are normal	Conventional
Test Statistic	$A^2 = 1.764$	$S = 0.0032$, $K = 2.9986$, $JB = 0.00017 \cong 0$	These were obtained with the aid of *R* statistical software (CRAN)
P-Value	*P*-Value = 0.168	*P*-Value = 0.235	Generated by computer; embedded in *R* statistical software
Decisions	Normality assumption holds	No departure from normality	The rule is such that we should reject the null hypothesis if *P*-Value is less than the significance level

Brief Explanation: From the Table-2 above, it is crystal clear that the data sets were normally distributed. In short, both Anderson-Darling and Jarque-Bera Tests supported the decision to uphold the normality assumption.

Verification of Homoscedasticity Assumption on the Data Sets

A spearman's rank correlation test was conducted on the residuals, and as such the following table gives the results after the application of R statistical software. The procedure is that spearman's rank correlation coefficients were obtained between the residuals and each of the exogenous variables. And, when none of the coefficient is significant high, we can say that error variances are the same.

Table-3: Spearman's Rank Correlation Test for Homogenous Assumption

Correlation Coefficients Between each of the Exogenous Variables and the Residuals ($r_{x_i e_i}$)				
r_1	r_2	r_3	r_4	r_5
-0.0029	-0.0067	-0.0052	0.0049	-0.0034

Interpretation: Since none of the correlation coefficients between residuals and each of the predictor variables is significantly high, it shows that the disturbances (errors) are homoscedastic. Therefore, homoscedasticity assumption is **not** violated in the data sets.

Confirmation of Orthogonality Assumption on the Data Sets

As we have considered only one of the three tests of Farrar-Glauber Chi-squared Test, our intention in this situation is just to know whether multicollinearity problem exists in the data sets. From the application of the formular, the following table was generated:

Table-4: Chi-squared Test of Orthogonality of Exogenous Variables

Required Information	**Farrar-Glauber Chi-squared Test**	**Remarks**
Null Hypothesis	Exogenous variables are orthogonal	Conventional
Test Statistic	$\chi^2_{cal} = 8.56$	This was obtained with the aid of *R* statistical software (CRAN)
P-Value	*P*-Value = 0.273	Generated by computer; embedded in *R* statistical software
Decisions	No departure from orthogonality of exogenous variables	The rule is such that we should reject the null hypothesis if *P*- Value < 0.05

Brief Explanation: From table-4 above, it can be depicted that there is complete absence of intercorrelations among the exogenous variables. Therefore, the assumption of orthogonality of exogenous variables is not violated.

Verification of Autocorrelation Problem on the Data Sets

It is hoped that Durbin-Watson test will be used to explore if there exists a problem of autocorrelations (serial correlations) in the data sets for this study. Sincerely, we have discussed Durbin-Watson statistic extensively in one of our previous sections. Now, the results of the application of this statistic are displayed in the following table with the aid of *R* statistical software:

Table-5: Durbin-Watson Statistic for testing Autocorrelations within the data sets

Required Information	**Durbin-Watson Test**	**Remarks**
Null Hypothesis	No serial correlations / No problem of autocorrelation	Conventional
Test Statistic	$d = 1.965$, $r = 0.002$	This was obtained with the aid of *R* statistical software (CRAN)
P-Value	*P*-Value = 0.126	Generated by computer; embedded in *R* statistical software
Decisions	No serial correlations	The rule is such that we should reject the null hypothesis if *P*- Value < 0.05

Brief Explanation: We can deduce from the table-5 above that since the values of *d* and *r* are approximately 2 and 0 respectively, and still *p-value* exceeds the value of significance level (0.05), the null hypothesis should not be rejected indicating that autocorrelation problem can be tolerated.

Estimation of Ordinary Least Square Parameters
Recall that:

$$\begin{bmatrix}\hat{\beta}_0\\ \hat{\beta}_1\\ \hat{\beta}_2\\ \hat{\beta}_3\\ \hat{\beta}_4\\ \hat{\beta}_5\end{bmatrix} = \begin{bmatrix} n & \sum x_1 & \sum x_2 & \sum x_3 & \sum x_4 & \sum x_5 \\ \sum x_1 & \sum x_1^2 & \sum x_1x_2 & \sum x_1x_3 & \sum x_1x_4 & \sum x_1x_5 \\ \sum x_2 & \sum x_1x_2 & \sum x_2^2 & \sum x_2x_3 & \sum x_2x_4 & \sum x_2x_5 \\ \sum x_3 & \sum x_1x_3 & \sum x_2x_3 & \sum x_3^2 & \sum x_3x_4 & \sum x_2x_5 \\ \sum x_4 & \sum x_1x_4 & \sum x_2x_4 & \sum x_3x_4 & \sum x_4^2 & \sum x_4x_5 \\ \sum x_5 & \sum x_1x_5 & \sum x_2x_5 & \sum x_3x_5 & \sum x_4x_5 & \sum x_5^2 \end{bmatrix}^{-1} \cdot \begin{bmatrix}\sum y\\ \sum x_1y\\ \sum x_2y\\ \sum x_3y\\ \sum x_4y\\ \sum x_5y\end{bmatrix}$$

Putting all necessary summations, we obtain the following estimates with the aid of *R* statistical software:

$$\begin{bmatrix}\hat{\beta}_0\\ \hat{\beta}_1\\ \hat{\beta}_2\\ \hat{\beta}_3\\ \hat{\beta}_4\\ \hat{\beta}_5\end{bmatrix} = \begin{bmatrix}16.67869\\ 0.52512\\ 0.08244\\ 0.41978\\ 0.03852\\ 0.07908\end{bmatrix}$$

Indicating that the initial estimated regression model is:

$$y = 16.67869 + 0.52512x_1 + 0.08244x_2 + 0.41978x_3 + 0.03852x_4 + 0.07908x_5$$

Analysis of Variance Technique for Testing Significance of the Initial Model

After careful application of all necessary formulae on how to obtain sum of square error and regression sum of square as well as mean square error and mean square regression, we came out with the following table:

Table-6: Results of Analysis of Variance for Overall Test of the estimated Model

Source of Variation	Degree of Freedom	Sum of Squares	Mean Square	F_{ratio}	$F_{5,90;0.05}$
Regression	5	983.46	196.692	2.04	2.32
Residual	90	8674.90	96.388		

Interpretation: Estimated values in the tabl-6 above were obtained from the data sets and analyzed using *R* statistical package. Recall that the null hypothesis is such that the estimated initial model is not significant, and since the calculated *F* value is less than the tabulated *F* value, it indicates that the null hypothesis should not be rejected and we conclude that the estimated initial model is not adequate, or precisely we conclude that at least one of the parameters is not significant.

Confidence Intervals, Correlations, Multiple Coefficient of Determination and Variance Inflation Factor

Having applied all necessary formulae and being careful with all precautions guiding the use of the formulae discussed in our previous sections, we arrived at the following summary of estimates as being analyzed with the aid of R statistical software given in the table below:

Table-7: Computation of Confidence Intervals, Simple and Multiple Correlation Coefficients, Multiple Coefficient of Determination and Variance Inflation Factor

Confidence Intervals (95%)	Simple Correlations	Multiple Correlation	Multiple Coefficient of Determination	Variance Inflation Factor
$0.062 \le \beta_1 \le 0.124$	$r_{x_1y} = 0.88$	$R = 0.89$	$R^2 \cong 0.79$	$VIF = (1-R^2)^{-1} = 4.81$
$0.541 \le \beta_2 \le 0.987$	$r_{x_2y} = 0.79$			
$1.516 \le \beta_3 \le 2.671$	$r_{x_3y} = 0.60$			
$-0.971 \le \beta_4 \le 1.384$	$r_{x_4y} = 0.27$			
$0.713 \le \beta_5 \le 1.930$	$r_{x_5y} = 0.85$			

Interpretations: When the estimates of confidence intervals (also known as confidence limits) are computed, the simple interpretation is that if the confidence limit passes through the origin (zero), the variable involved in this scenario is **not** statistically significant at a chosen level of significance. Results from table-7 show that only $-0.971 \leq \beta_4 \leq 1.384$ passes through zero meaning that commission on the sales of third party institutions such as commission on treasury bills, money market instruments, sales of institution forms, etc does not have much contribution to the expenditure made in commercial banks in Nigeria while other sources of revenue contribute significantly.

Basically, one of the objectives of this study is to determine the source of income in commercial banks that produce the highest profit. Therefore, from the same table VII, it is obvious that interest on loans and advances is the highest source of profit while commission on the sales of third party institutions happens to be the least source. The implication of this is that commercial banks shouldn't stop giving out loans and advances to interested customers, and should also be encouraged in financing capital projects because these two sources are the major determinants of profit maximization in commercial banks. In addition, COT also has about 79% positive contribution to profit maximization. This fact is obtained from the computation of simple correlation between the sources of income and the corresponding expenditure.

From the same table VII, however, the estimated value for multiple correlation indicates that all the sources of income covered in the present study has 89% positive relationship with expenditure meaning that the model formulated is approximately 89% reliable in all situations of its usage. This can further be explained such that the probability (chance) of any forecast made using the formulated model is 0.89; that is, it is 89% certain that the prediction made with the formulated model would be true and reliable. This chance is stronger and better since it is above average of 100%.

Coefficient of Determination computed in table VII shows that the coefficient is 0.79 indicating existence of about 21% variations in the expenditure of the commercial banks which cannot be accounted for by the sources of income under consideration. The implication of this estimate is that there are some other sources of income that are included in the present research or equivalently, there are some variations which are approximately 21% that cannot be explained by all the five sources of income mentioned in the present study. Invariably, only 79% of banks' expenditure can be explained by the profits realized from all the sources of revenue under study.

It has been discussed in the previous subheading that Variance Inflation Factor (VIF) can be interpreted such that it is an indicator for detecting the presence of collinearity within the exogenous variables of the data sets provided that the value of VIF falls between 5 and 10. However, since the value of VIF does not fall between the range of 5 and 10, it is a clear indication that exogenous variables are not intercorrelated and, therefore the problem of multicollinearity does not exists.

Selection of the Best Model using Stepwise Method

As discussed, Stepwise Selection Procedure of the best model is very sound in scrutinizing variables that are most relevant and significant to formulate models that are adequate in the long run; this is because it comprises both backward elimination and forward selection methods. Its beauty is embedded in the *R* statistical package [10]. Having applied all necessary steps and be guided by all precautions involved in the usage of the method, the following final model is generated with the aid of *R* statistical package:

$$y = 29.81265 + 16.48523x_1 - 13.59241x_2 + 8.44269x_3 + 14.99846x_5$$

Finally, above model is recommended for forecasting as a result of its validity as well as reliability that is embedded in it.

Summary of Findings

The data sets, classified as interests on loans and advances, commission on turnover, electronic banking charges, commission on sales of third party institutions such as sales of institution forms, and finally income from other investments such as commission on treasury bills, money market instruments, and so on, which are termed as exogenous variables independent of commercial banks' expenditure, were sourced from Central Bank of Nigeria (CBN). Scatter diagrams were drawn to preview the kind of relationship that exists between the expenditure of commercial banks and their income; it was discovered that only commission on sales of third party institutions such as sales of institution forms was not as significant as other sources of income. Box Plot was later sketched to detect the presence of outliers in the data sets but no outlier was found. Anderson-Darling and Jarque-Bera tests were used to investigate normality assumption. We realized that the data sets were all normal. Chi-squared and Spearman's Rank Correlation tests were also implored to verify orthogonality and homoscedasticity assumptions. Having applied these two procedures, the data sets were free from collinearity and error variances were all the same. With the application of Durbin-Watson statistic, we discovered that no serial correlation problem indicating that residuals behave well.

From the computation of Analysis of Variance table, we found out that our initial model is not adequate, that is, one of the model parameters is not significant, and after variable scrutiny by Stepwise Selection Procedure, it was discovered that one of the exogenous variables: commission on sales of third party institutions such as sales of institution forms did not significantly contribute to the expenditure of commercial banks in Nigeria. And as such, its source of profit is not reliable. Results obtained from computation of confidence intervals showed that only one exogenous variable, which is commission on sales of third party institutions such as sales of institution form, etc, seem to have been weakened in terms of relationship with commercial banks' expenditure.

CONCLUSIONS

So far, we were able to conclude that income realized from giving loans and advances to customers is the most profitable source and approximately 88% reliable, followed by income from other investments such as financing capital projects, commission on treasury bills, money market instruments, and so on, which is approximately 85% reliable, and thus commission on sales of third party institutions such as sales of institution forms was the least profitable source of income. We strongly advise that commercial banks in Nigeria should find all means to fast-track loans' applications of their customers with a view to granting the approval as quick as possible.

Commercial banks in Nigeria should not rely on income realized as commission from the sales of third party institutions because from the present study, it was revealed that the source is not reliable and, commercial banks may be disallowed from acting as agents to institutions by issuance of circulars from Central Bank of Nigeria (CBN). We strongly advise that commercial banks in Nigeria should intensify efforts by all means to convince and encourage more customers on financial borrowings and, also increase the risks of financing capital projects with a view to maximizing profits in banking sector.

Finally, it is reasonable to conclude that for any commercial bank to make forecast within the scope of their sources of income and expenditure,

$$y = 29.81265 + 16.48523x_1 - 13.59241x_2 + 8.44269x_3 + 14.99846x_5$$

should be used as the final model indicating that only four sources of income: interests on loans and advances, commission on turnover, electronic banking charges and income realized from other investments like part-taking in financing capital projects, charges from keeping valuables, etc, are both significant and adequate to make prediction.

ACKNOWLEDGEMENTS

We would like to appreciate the following people: Dr Adejumo Adebowale Olusola (Senior Lecturer in Statistics Department of the University of Ilorin, Nigeria); Pastor R. M. Alabison (Dean, School of Pure and Applied Sciences, Federal Polytechnic Ede, Nigeria); Dr 'Tunji Abiala (Mathematics Department, Federal University of Agriculture, Abeokuta, Ogun State, Nigeria); Dr Kolapo Ige (Senior Lecturer in Economics Department of Joseph Ayo Babalola University, Ikeji-Arakeji, Osun State); Dr. Agboola, Jacob Olusola (Rector, Osun State Polytechnic, Iree, Nigeria); Mrs. Olaniyan Olabisi Dorcas (Due Process Officer, Ejigbo Local Government, State of Osun, Nigeria); Dr Isamotu Rafiu (A Medical Practitioner in Dublin who doubles as Special Assistant to the Governor on Health Matters, State Government of Osun, Nigeria); Mr. Ogunleye Samson Oluseyi; Ogunleye 'Segun Ayobami; Mr Adewusi Adegboyega Adelakin (Director, Department of Budget, Planning, Research and Statistics, Ejigbo Local Government, State of Osun, Nigeria); Mr Babawale George (Admin. and General Services Department, Ifelodun Local Government, Osun State, Nigeria); Prof. Babatunde Lateef Adeleke (Dean, Faculty of Science, University of Ilorin, Nigeria); Pastor Afolabi Samuel Olufemi (Head of Accountancy Department, Federal Polytechnic Ede, Nigeria); Comrade Oyedeji 'Femi Samson (Executive Chairman, NULGE, Ifelodun Local Government Chapter, Ikirun, Nigeria); Adegbenro Abraham Bukola (CEO, Omoluabi Consults, Ede, State of Osun, Nigeria). I (Ogunleye Timothy A.) wish to specially thank Mrs. Adebayo Morufat Olanrewaju (Nee Oseni) from PHC Department, Ejigbo Local Government, State of Osun, Nigeria for all her timely supports.

REFERENCES

1. Brown PJ; Measurement, Regression and Calibration, Oxford, 1994.
2. Everitt BS; The Cambridge Dictionary of Statistics, Second Edition, Cambridge University Press. 2002.
3. Drapper NR, Smith H; Applied Regression, Wiley Series in Probability and Statistics, 1998.
4. Gujarati; Basic Econometrics, Fourth Edition, The McGraw-Hill Companies, 2004.
5. Timothy OA, Olawuwo S, Olaleye OM; Economic Analysis and Modelling of Local Government Expenditure on Income in the South West Zone of Nigeria. 2013; 17(2):200-206.
6. Thompson OO, Timothy OA, Olawuwo S, Adeleke MO; Experimental Design Analysis on the Efficiency of Grain Yields of Cowpea Varieties in the three Senatorial Zones of Osun state, Nigeria. IOSR Journal of Mathematics, 2014;9(6):52-60.

7. Olawuwo S, Timothy OA, Thompson OO, Adebowale AO; Comparison of Classical Least Squares (CLS), Ridge and Principal Components Methods of Regression Analyses using Gynaecological Data', IOSR Journal of Mathematics, 2014; 9(6):61-74.
8. Oxford Advanced Learners' Dictionary of Current English by A. S. Hornby, seventh edition.
9. Rawlings JO; Applied Regression Analysis, Wadsworth Books, California (ARA), 2014.
10. R statistical package from Development Core Team, *R*: A Language and Environment for Statistical Computing, Vienna, Austria, *R* Foundation for Statistical Computing, 2012.
11. Chatterjee S, Ali SH; Regression Analysis by Example, Fourth Edition, John Wiley and Sons, Inc., 2006.
12. Sanford B, Charles D; Pharmaceutical Statistics: Practical and Clinical Applications, Fourth Edition, Revised and Expanded Edition, Volume 135, 2004.