International Journal of Academic and Applied Research (IJAAR)
ISSN: 2643-9603
Vol. 3 Issue 10, October – 2019, Pages: 20-35

Generalized Two-Stage (OLS, LAD) Estimator

Musaphau Animashahun
Department of Mathematics and Statistics, Osun State College of Technology, Esa-Oke, Nigeria

Prof. Kayode Ayinde
Department of Statistics, Federal University of Technology, Akure, Nigeria

Abstract: *In any organization expenditure is generally believed to be related to income generated, among other factors. These factors need to be examined so as to know their effect and contribution on expenditure. Consequently, this work aims at modelling monthly expenditure and sales of a subsidiary of Nigerian National Petroleum Corporation's products using linear regression model. The data used in this study were monthly expenditure and sales of petro-chemical products for a period of ten years. The Dickey-Fuller test statistic was used to test for the stationarity of the data.. Other statistics tests used include the Jaque-Bera, Glejser/LM and Durbin-Watson test statistic for testing normality, homoscedasticity and independence of error terms respectively. The Variance Inflation Factor (VIF) was also used to examine the presence of multicollinearity. Arising from various linear regression assumptions' violation called for the use of Generalised Two-Stage(OLS, LAD) estimator. The results show that the original data is stationary. The OLS estimation results reveal the presence of multicollinearity, non-normality, heteroscedasticity and autocorellated error terms in the dataset. Attempting to improve the analysis using Generalised Two-Stage (OLS,LAD) shows that the results of the latter estimator are more efficient than the former, in terms of their significant variables. In conclusion, the results reveal that sales of Kerosene and Air Transport Kerosene(ATK) are positive and have significant contribution to NNPC monthly expenditure.*

INTRODUCTION

Regression analysis is a conceptually simple method for investigating functional relationship among variables. The term regression was first introduced by Galton (1877) while studying the relationship between the heights of fathers and sons. The term was introduced by him in the paper tagged "Regression towards Mediocrity in Hereditary Status". Regression analysis attempts to establish the 'nature of the relationship' between variables, the dependent and some set of independent variables. It provides a mechanism for prediction.

It is clear from the above definition that regression analysis is a statistical device with the help of which we are in position to estimate the unknown value(s) of one variable from known value(s) of another variable. The variables which are used to predict the variable of interest are called independent variable or explanatory variable and the variable we are trying to predict is called dependent variable or explained variable. In a nutshell, we denote dependent variable by Y and the set of independent variables by $X_1, X_2, ..., X_p$, where p denotes the number of predictor variables. The true relationship between Y and $X_1, X_2, ..., X_p$ can be approximated by regression model:

$$Y = f\left(x_1, x_2, ..., x_p\right) + U, \tag{1}$$

where U is the random error component with some assumptions.

Literature Review on Expenditure

Abby, Jaffrey and Ara (2006) established correlation between Bureau of Labour Statistics (BLS) petroleum-product Consumer Price Index (CPI) and changes in consumer spending on those products, as measured by the Consumer Expenditure Survey (CES). Odularu (2006) used regression analysis to model Nigerian economy using crude oil data. He discovered that oil sector contributed positively to Nigeria economy.

Abiodun and Solomon (2010) examined world price for Nigerian major agricultural commodities. He emphasized that Nigerian income and Nigerian past agricultural output were determinants of agricultural exports. Using Ordinary Least Squares regression estimator, the study recommended that priority should be accorded to the boosting of the current level of agricultural output. Mehmood and Sadiq (2010) examined the long run as well as short run relationship between the fiscal deficits. The results showed a negative relationship between government expenditure and poverty based on time series data from 1976 to 2010. Onwe (2012)

ISSN: 2643-9603
Vol. 3 Issue 10, October – 2019, Pages: 20-35

examined the impacts of changes of policy implementation in oil sector in Nigeria. The negative implications and positive effects of such policies were highlighted.

Falukasi and Awomuse (2011) assessed the determinants of demand functions for import in Nigeria using variables: Real Gross Domestic Product (RGDP), External Reserves (EXTR), Real Exchange Rate (REXCH) and Index of Openness (OPNS)as determinant factors. They used the statistical significance of the lagged error correction model, ECM(-1). Fukuda and Hiyoshi (2013) discovered the association of household expenditure and marital status with cardiovascular risk factors in Japanese adults. Olawuwo, Ogunleye and Olaleye (2013) modeled local government monthly expenditure on income. It was discovered that statutory allocation was recommended as principal determinant of local government expenditure in the South West Zone of Nigeria. Olukotun, James and Olorunfemi (2013) were able to model the effect of introduction of mobile phones and the students' expenditure pattern in Anyigba Community, using linear regression. A positive correlation was discovered. Ogunleye, Olaleye and Solomon (2014) modeled commercial banks' expenditure on sources of profit maximization in Nigeria, it was discovered that interest on loans and advances gave the most profitable sources of income.

Aim and Objectives of the Study

The aim of this work is to model monthly expenditure and sales of petroleum products of NNPC and diagnose the model in the light of its assumptions. The specific objectives are to detect whether or not the assumptions of:

(i) normality of error terms, Independence of explanatory variables, Independence of error terms and Homoscedasticity of error variance are violated.

(ii) when violated, we attempt to correct the violations.

Time Series Data and Stationarity Test

Time Series is a set of observations or values that a variable takes at different time. Such data may be collected at regular time interval, such as daily, weekly, monthly, quarterly, annually, quinquennially or decennially (Gujarati, 2007). In statistical inference of modeling time series data, the first thing one needs to do is ascertain the stationarity status of the data.

A time series is said to be stationary if its mean and variance are constant over time and the value of the covariance between the two time periods depends only on the lag between the two time periods and not the actual time at which the covariance is computed (Gujarati, 2007)

This can be expressed as follows:

Let Y_t be a stochastic time series with the following properties

Mean $$E[Y_t] = \mu \qquad (2)$$

Variance $$Var[Y_t] = E\left[(Y_t - \mu)^2\right] = \sigma^2 \qquad (3)$$

Covariance $$\gamma_k = E[(Y_t - \mu)(Y_{t+k} - \mu)], \qquad (4)$$

where γ_k is the covariance or autocovariance at lag k.

Testing for Stationarity

There are several methods for testing for stationarity. The followings are some of the methods:

(i) Graphical Method

Plotting time series data can give initial clue about the likely nature of the time series. If the time series plot shows an upward trend, it suggests that the mean of time series varies; which implies that time series is not stationary, otherwise stationary(Ogunleye et al., 2014).

(ii) Autocorrelation Function and Correlogram

A simple test of stationary is based on autocorrelation function (ACF). The ACF at lag *k*, denoted by ρ_k is defined as

$$\rho_k = \frac{\gamma_k}{\gamma_0} = \frac{Covariance\ at\ lag\ k}{Variance} \tag{5}$$

Plotting ρ_k against *k*, the graph so obtained is termed population correlogram.Whenever the correlogram of a time series hover around zero at various lags, the time series is said to be stationary, otherwise nonstationary (Gujarati, 2007).

(iii) <u>Unit Root Test</u>

A test of stationarity that has become widely popular is the Unit Root Test.

Given $Y_t = \rho Y_{t-1} + U_t \qquad -1 \le \rho \le 1$ (6)

Where U_t is a white noise error terms. If $\rho = 1$, that is unit root, meaning the stochastic process is non-stationary.

Subtracting Y_{t-1} from both sides of (6), it becomes

$$Y_t - Y_{t-1} = \rho Y_{t-1} - Y_{t-1} + U_t$$

$$= (\rho - 1)Y_{t-1} + U_t$$

$$\Delta Y_t = \delta Y_{t-1} + U_t \tag{7}$$

where $\delta = (\rho - 1)$ and Δ is first-difference operator.

So, if $\delta = 0$, $\rho = 1$, indicating that the time series under consideration is not stationary but its 1st differences stationary.

Dickey and Fuller (1979) under the null hypothesis indicated that $\delta = 0,$ the estimated t value of coefficient Y_{t-1} follows the Tau-statistic. They computed the critical values of the Tau-statistic on the basis of Monte Carlo simulation. The Tau-statistic or test is known as Dickey-Fuller (DF) Test (Gujarati, 2007).

<u>Classical Linear Regression Model and Its Assumptions</u>

Classical Linear Regression Model is a kind of model that shows relationship between dependent and independent variables. The Ordinary Least Square (OLS) method is one of most important ways for estimating the parameters of linear models, because of its simplicity and rationality. The results are obtained when specific assumptions of the model are not violated. However, when any of these assumptions are violated, the results of OLS estimator are affected, especially its efficiency.

Suppose there is a linear relationship between dependent variable Y_j, the explanatory variables X_{ij} and error terms U_j, $i = 1, 2, \ldots, p\ \ and\ j = 1, 2, \ldots, n.$

This relationship can be expressed in matrix form as follow:

$$Y = X\beta + U \tag{8}$$

where Y is an $(n \times 1)$ dimensional vector observations of dependent variables, X is the $[n \times (p+1)]$ matrix of explanatory variables, and β, is the $[(p+1) \times 1]$ vector of regression coefficients, U is the $(n \times 1)$ vector of errors with properties $E[U] = 0$, $E[UU'] = \sigma^2 I_n$ and I_n represents the $(n \times n)$ dimensional identity matrix.

The assumptions of Classical Linear Regression Model are:

(i) Error terms, U, have mean zero; E[U] = 0:

(ii) The error terms, U, have equal variance; i.e., they are homoscedastic, $Var[UU'] = \sigma^2$:

(iii) Error terms, U, are normally distributed; U ~N(0, $\sigma^2 I_n$) ;

(iv) There should not be intercorrelation between the explanatory variables; there should be no problem of multicollinearity; $Cor\lfloor X_i, X_j \rfloor = 0$, $(i \neq j)$;

(v) Correlation of error terms and explanatory variables should be zero, $Cor[U, X] = 0$;

(vi) The error terms of different observations $(U_i, U_j), (i \neq j)$, are independent; $Cor\lfloor U_i, U_j \rfloor = 0$. Hence, there is no autocorrelation problem ;

(vii) The regression model is linear ;

(viii) X values are fixed in repeated sampling; and

(ix) Number of observation, n, must be greater than the number of parameter to be estimated (Dimitrius and Stephen, 2011).

<u>Statistics to Check Assumption of Classical Linear Regression Model</u>

The following are some of the statistics to check the assumptions of Linear Regression Model.

<u>Testing for Normality</u>

The followings are some statistics for testing for normality of a set of data.

(i) <u>Anderson-Darling Normality Test:</u>

The 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.It is a measure of how closely a data set follows the normal distribution. The null hypothesis for this test is that the data set is normal. The statistic is stated as follows:

$$A^2 = -\frac{1}{n}\left[\sum_{i=1}^{n}(2i-1).\{\log z_i + \log(1 - z_{n+1-i})\}\right] - n \tag{9}$$

where

$z_i = \Phi\left(\frac{x_{(i)} - \bar{x}}{s}\right)$, $\Phi(u) = \int_{-\infty}^{u} \frac{1}{\sqrt{2\pi}} \ell^{-\frac{1}{2}u^2} du$, $x_{(1)} \le x_{(2)} \le \ldots \le x_{(n)}$ 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(Adejumo, Olawuwo and Ojo, 2014)

(ii) Jarque-Bera Normality Test

Jarque and Bera (1987) established a statistic to test the normality of observations. The statistic is based on the skewness and kurtosis of the residuals which are calculated using the sample moments. The sample skewness and kutosis coefficients can be calculated by

$$S = \frac{\hat{\mu}_3}{\hat{\mu}_2^{3/2}} \quad \text{(Skewness)} \tag{10}$$

$$K = \frac{\hat{\mu}_4}{\hat{\mu}_2^2} \quad \text{(Kurtosis)} \tag{11}$$

where, μ_j is the j^{th} order central sample moment, $\hat{\mu}_j = \frac{1}{T}\sum\left(\hat{\varepsilon}_t - \bar{\varepsilon}\right)^j$ with $\bar{\varepsilon} = \frac{1}{T}\sum \hat{\varepsilon}_t$. If there is interception in the model, the Jarque-Bera test statistic for null hypothesis that the observations (residuals) are normally distributed is given as

$$JB = T\left[\frac{S^2}{6} + \frac{(K-3)^2}{24}\right] \tag{12}$$

The Jarque-Bera test has a chi-squared distribution with 2 degrees of freedom asymptotically. The null hypothesis is rejected if the computed chi-squared value exceeds the criticalchi-squared value.

(iii) The Lilliefors Test

The test statistic is

$$LF = Sup[F^*(x) - S(x)] \tag{13}$$

where $F^*(x)$ is the Standard Normal distribution function

$S(x)$ is the empirical distribution function(Lilliefors, 1969).

(iv) The Shapro-Wilk Test

The statistic is

$$W = \frac{\left(\sum_{i=1}^{n} a_i x_{(i)}\right)^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2} \tag{14}$$

where $x_{(i)}$ is the i^{th} order statistic,(smallest number in the sample);

$\bar{x}$ is the sample mean and a_i are constants given by Shapiro and Wilk (1965).

Testing for Autocorrelation

The following is a method for detecting autocorrelation

Durbin-Watson Test for Autocorrelation:

International Journal of Academic and Applied Research (IJAAR)
ISSN: 2643-9603
Vol. 3 Issue 10, October – 2019, Pages: 20-35

Of all the available tests for the existence of serial correlation in a dataset, Durbin-Watson statistic is the most popular and to certain degree reliable. The procedure is as follows:

Step I: Define the hypothesis as follow:

$H_0 : \delta = 0$ (Meaning that there is no first order autocorrelation)

$H_1 : \delta \neq 0$ (Meaning that first order autocorrelation exists)

Step II: Fit in the estimated parameters into the model as follows:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_p X_p + U$$

Step III: Calculate $\hat{Y}$ from the values of the explanatory variables (X_s)

Step IV: List the values of observation under $Y \text{ and } \hat{Y}$

Step V: Calculate $U_t = Y - \hat{Y}$ and also obtain U_t^2 , and list them out

Step VI: Calculate $U_t - U_{t-1}$ and also obtain $(U_t - U_{t-1})^2$

Step VII: Calculate $d = \frac{\sum_{j=2}^{n}(U_j - U_{j-1})^2}{\sum_{j=1}^{n} U_j^2}$

Step VIII: By the help of computed ***d*** above, use the mathematical relationship $d \cong 2(1-\rho)$ to obtain $\hat{\rho}$. The interpretation is that if $\hat{\rho}$ is approximately zero, then autocorrelation does not exist in the dataset or better still, its presence can be tolerated.

In summary, an approximate relationship between ***d*** and ρ is $d \cong 2(1-\rho)$ showing that ***d*** has a range of **0** to **4** respectively. Since $\hat{\rho}$ provides an estimate of δ , it is clear that ***d*** is close to **2** when $\hat{\delta}$ = **0** and ***d*** is near to **0** when $\hat{\delta}$ = **1**. The closer the samples value of ***d*** to **2**, the stronger the evidence that there is no autocorrelation present in the error. Evidence of autocorrelation is indicated by the deviation of ***d*** from **2**. In the case of small samples, the following decision rules are designed by Durbin and Watson:

* If $d < d_L$, we reject the null hypothesis of no autocorrelation and accept that there is positive autocorrelation of order one (1^{st} *order*).

* If $d > (4 - d_L)$, we reject the null hypothesis of no autocorrelation and accept that there is negative autocorrelation of order one (1^{st} *order*).

* If $d_U < d < (4 - d_U)$, we accept the null hypothesis of no autocorrelation.

* If $d_L < d < d_U$, the test is inconclusive.

(Ogunleye et al.,2014)

Testing for Multicollinearity

The following are methods for testing for multicollinearity.

(i) Auxiliary Regressions

Since multicollinearity arises because one or more of the regressors are exact or approximately linear combinations of the other regressors, one way of finding out which *x* variable is related to other *x* variables is to regress each x_i on the remaining *x* variables

ISSN: 2643-9603
Vol. 3 Issue 10, October – 2019, Pages: 20-35

and compute the corresponding R^2, which we designate as R_i^2; each one of these regressions is called **auxiliary regression.** The regression is an auxiliary to the main regression of *y* on the *x's*. The relationship between *F*and R_i^2 is established as follows:

$$F_i = \frac{R_i^2/(p-2)}{(1-R_i^2)/(n-p+1)} \sim F_{(p-2),(n-p+1);\alpha} \tag{15}$$

It should be noted that '*n*' stands for sample size, '*p*' stands for the number of explanatory variables, and R_i^2 is the coefficient of determination in the regression of variable x_i. The decision rule is such that if the computed F_i is greater than the corresponding critical value ($F_{tab.}$) at a chosen significant level, then the particular x_i is collinear with other *x's*; otherwise it is not.

Instead of formally testing all auxiliary R^2 value, one may adopt Klien's rule of thumb, which suggests that that multicollinearity may be a troublesome problem only if R_i^2 obtained from an auxiliary regression is greater than the overall R^2 (the one obtained from the regression of *y*on all the regressors)**(**Klien, 1962).

(ii) <u>Eigen Values and Condition Index</u>

This is another way to suspect the existence of multicollinearity in a data set. When Eigenvalues are obtained, we can compute the condition index (CI) by first calculating '*k*' as follows:

$$k = \frac{Maximum\, eigenvalue}{Minimum\, eigenvalue}$$

Then, we can now calculate the **condition index** as follows:

$$CI = \sqrt{\frac{Maximum \quad eigenvalue}{Minimum \quad eigenvalue}} = \sqrt{k}$$

It has been reported in some literature that if ***k* is between 100 and 1000,** there is moderate to strong multicollinearity but if **it exceeds 1000**, there is **severe** multicollinearity. Alternatively, if CI **is between 10 and 30,** there is moderate to strong multicollinearity but if CI **exceeds 30**, there is **severe** multicollinearity(Gujarati, 2007).

(iii)<u>Tolerance and Variance Inflation Factors (VIF)</u>

As R_i^2, the coefficient of determination in the regression of variable x_i on the remaining regressors in the model increases toward unity, that is, as the collinearity of x_i with the other regressors increases, Variance Inflation Factor (VIF_i) also increases and in the limit it can be infinite. VIF_i simply measures how the variance of a particular regressor x_i inflates or increases in association with other regressors. Some authors use VIF_i as an indicator of multicollinearity. The larger the value of VIF_i, the more 'troublesome' or collinear the variable x_i. It has been reported in some literaturethat if VIF_i of a particular variable is less than 10, that variable is said to be collinear. VIF_i can be computed using the following relation.

$$VIF_i = \frac{1}{1-R_i^2} \tag{16}$$

Of course, one could use 'Tolerance Factor' *TF* as a measure of multicollinearity in view of its intimate connection with VIF_i. The closer TF_i to zero, the greater the degree of multicollinearity of that variable with other regressors. On the other hand, the closer TF_i is to 1, the greater the evidence of that x_i is **not** collinear with the other regressors. Some believe that if the value of TF_i is between 0.10 and 0.20, then there is presence of multicollinearity. TF_i can be computed using the following relation:

$$TF_i = 1-R_i^2 \tag{17}$$

International Journal of Academic and Applied Research (IJAAR)
ISSN: 2643-9603
Vol. 3 Issue 10, October – 2019, Pages: 20-35

(Farrar and Glauber, 1967)

(iv) Chi-Squared Test for Multicollinearity

It should be noted that multicollinearity is a modeling error; it is a condition of deficient data. One of the three stage-test of Farrar and Glauber (1967) for the detection of existence of multicollinearity is the use of chi-square test. The test is tailored as followed:

Hypothesis:

H_0 : No multicollinearity

Vs

H_1 : Multicollinearity exists

Test statistic:

$$\chi^2_{cal} = -\left\{(n-1) - \frac{1}{6}(2p+5)\right\}.\,\text{In}\,|R| \quad \sim \quad \chi^2_{p(p-1)/2;\alpha} \tag{18}$$

Where $|R|$ is the determinant of correlation matrix of all the predictor variables and p is the number of all the predictor variables.

Decision Rule: Reject the null hypothesis if χ^2_{cal} is greater than $\chi^2_{p(p-1)/2;\alpha}$. Otherwise, do not reject.

Testing for Heteroscedasticity

These are some of the various methods that could be used for testing heteroscedasticity.

(i) White General Hetetroscedasticity Test

It is a test that does not rely on the normality assumption and is easy to implement. Consider the following three variable regression model:

$$Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + U_i \tag{19}$$

The White test procedures are as follows:

Step 1:- Given the data, we estimate (19) and obtain the residuals, $\widehat{U}_i$,

Step 2:- We then run the following (auxiliary) regression

$$\widehat{U}_i^2 = \alpha_1 + \alpha_2 X_{2i} + \alpha_3 X_{3i} + \alpha_4 X_{4i} + \alpha_5 X_{5i} + \alpha_6 X_{6i} + V_i \tag{20}$$

That is the squared residuals from the original regression are regressed on the original X variables and product(s) of the regressors. Higher power of regressors can also be introduced. Obtain the R^2 from this (auxiliary) regression.

Step 3:- Under the null hypothesis that there is no heteroscedasticity, it can be shown `that sample size (n) times the R^2 obtained from auxiliary regression asymptotically follows the Chi-Square distribution with degree of freedom equal to the number of regressor in the auxiliary regression. That is,

$n.R^2 \sim \chi^2$ (Asymptotically)

Step 4:- If the Chi-Square value obtains in (20) exceeds the critical Chi-Square value at the chosen level of significance, the conclusion is that there is heteroscedasticity, otherwise there is none (Gujarati, 2007).

(ii) Glejser Heteroscedasticity Test

Gujarati (2007) discussed Glejser's suggestion about heteroscedasticity test by regressing the absolute value of residual, $\widehat{U}_i$, from OLS regression on X variable that is thought to be closely associated with σ_i^2. The following functional forms are used:

$$|\widehat{U}_i| = \beta_1 + \beta_2 \sqrt{X}_i + V_i \tag{21}$$

$$|\widehat{U}_i| = \beta_1 + \beta_2 \frac{1}{X_i} + V_i \tag{22}$$

$$|\widehat{U}_i| = \beta_1 + \beta_2 \frac{1}{\sqrt{X_i}} + V_i \tag{23}$$

where V_i are the error terms.

It was discovered that for large samples, the proceeding models gives generally satisfactory result in detecting heteroscedasticity.

Estimation Methods under Violations of Classical Linear Regression (CLR) Model

The following are some remedial measures to handle violations of CLR model.

Non-Normality of Error Term

The followings are some of the methods that could be used to handle non-normality

(i) Least Absolute Deviation (LAD)

The ordinary least square estimator is optimal when the disturbance in the equation is normally distributed. But when the disturbance is not normally distributed, other estimators are better. If the distribution is known, the efficient estimator is maximum likelihood with the correct distribution function. However, in many cases, one may suspect that ones data distribution is "fat-tailed" or contains outliers, without knowing exactly its form. In this setting, the LAD estimator, which minimizes the sum of absolute deviations of the residuals, may be more efficient.

Least Absolute Deviation is also known as Least Absolute Value (LAV). LAD minimizes the sum of the absolute value of residual with respect to the coefficient vector, B:

$$\text{Min}\sum_{i=1}^{n}|y_i - X_i B| \tag{24}$$

The property of the LAD estimator is that there are K residuals that are exactly zero. LAD is robust to an outlier in Y-direction. However, LAD estimator does not protect against outlying X (Judge, Carter, William, Helmut and Tsoung-Chao, 1998).

(ii) M-Estimator

The class of M-Estimator was introduced by Huber in 1964 (Marquardt and Snee, 1975). M-Estimator, T_n, is defined as a solution of the minimization problem.

$$\sum_{i=1}^{n} \rho(X_i, \theta) := \min \theta \quad \text{with respect to } \theta \in \Theta \tag{25}$$

$$E_{P_n}[\rho(X, \theta)] = \min \theta \in \Theta \tag{26}$$

where $\rho(.\ ,\ .)$ is a properly chosen function. The class M-Estimator cover also the maximal likelihood estimator of parameter θ in the parametric model $\rho = \{\rho_\theta, \theta \in \Theta\}$; if $f(x, \theta)$ is the density function of ρ_θ then MLE is a solution of the maximization

$$\sum_{i=1}^{n}\left(-\log f(x_i, \theta)\right) = min \quad \theta \in \Theta \tag{27}$$

If ρ in (25) is differentiable in θ with continuous derivative $\varphi(.\ ,\ \theta) = \frac{\partial}{\partial\theta}\left(\rho(.\ ,\ \theta)\right)$, then T_n is a root or roots of the equation

$$\sum_{i=1}^{n} \varphi(X_i\ ,\ \theta) = 0,\ \ \theta \in \Theta \tag{28}$$

Hence,

$$\frac{1}{n}\sum_{i=1}^{n} \varphi(X_i, T_n) = E_{\rho_r}[(X, T_n)] = 0,\ \ T_n \in \Theta \tag{29}$$

From (26) and (29) that the M-functional corresponding to T_n is defined as the solution of the minimization of $f(x, \theta)$.

Non-Homoscedasity of Error Term

The estimator below is used to handle non-homoscedasity.

(i) Weighted Least Square

Consider the two variable case

$$Y_i = \beta_1 + \beta_2 X_i + U_i \tag{30}$$

where the error terms are heteroscedastistic, that is $V(U_i) = \sigma_i^2, \quad i = 1, \dots, n$ we will assume all the other classical assumptions still hold.
Dividing each observation of the linear model by σ_i, we get

$$\frac{Y_i}{\sigma_i} = \beta_1 \frac{1}{\sigma_i} + \beta_2 \frac{X_i}{\sigma_i} + \frac{U_i}{\sigma_i}$$

$$Y_i^* = \beta_1 X_{1i}^* + \beta_2 X_{2i}^* + \dots + \beta_k X_{ki}^* + U_i^* \tag{31}$$

Note that $V(U_i^*) = V\left(\frac{U_i}{\sigma_i}\right) = 1$, then the residuals of this transformed model are homoscedastic. Hence, if we know σ_i^2, the BLUE is simply the OLS with the transformed model is

$$\hat{\beta}_{2\ wls} = \frac{\sum_{i=1}^{n} X_i^* Y_i^*}{\sum_i^n X_i^{*2}}$$

$$\hat{\beta}_{1\ wls} = \bar{Y}^* - \hat{\beta}_{2\ wls} \bar{X}^*$$

with $Y_i^* = {}^{Y_i}/_{\sigma_i}, \quad X_i^* = {}^{X_i}/_{\sigma_i}$.

The name weighted least squares comes from the fact that the estimator can be obtained by solving the following minimization problem

$$min \sum_{i=1}^{n} \frac{1}{\sigma_i^2} e_i^2 \tag{32}$$

Suppose heteroscedasticity is present in the form of an unknown function of the regressors which can be approximated by a quadratic relationship. Alin and Riccardo (2010) provided an heteroscedastity correction which offers the possibility of consistent standard errors and more efficiently parameter estimates as compared with OLS.
The procedure involves (a) OLS estimation of the model of interest, followed by (b) an auxiliary regression to generate an estimate of the error variance, then finally (c) weighted least squares, using as weight the reciprocal of the estimated variance.
In the auxiliary regression (b) we regress the log of the squared residuals from the first OLS on the original regressors and their squares. The log transformation is performed to ensure that the estimated variances are non-negative. Call the fitted values from this regression u*. The weight series for the final WLS is then formed as 1/exp(u*).

Lack Independence of Explanatory Variables (Multicollinearity Problem)

Some of the estimators which could be used to correct Multicollinearity problem are:

(i) Ridge Regression Estimator

One of the most popular estimators proposed by Hoerl and Kennard (1970) is the Ridge Estimator defined as

$$\hat{\beta}_R = (X'X + K)^{-1}.X'Y \tag{33}$$

$$\hat{\beta}_R = (I_p + K(X'X)^{-1})^{-1} \hat{\beta}_{OLS} \tag{34}$$

where K is a constant $(0 \le K \le 1)$.

ISSN: 2643-9603
Vol. 3 Issue 10, October – 2019, Pages: 20-35

equation (2.39) is called Ordinary Ridge Regression, ORR, estimator. Nonetheless, when K is a diagonal matrix, the equation (2.34) becomes Generalized Ridge Regression, GRR,estimator.

(ii) Liu Estimator

Liu (1993) proposed an estimator similar in form but different from ridge estimator of Hoerl and Kennard (1970).

The Liu Estimator,$\hat{\beta}_{LIU}$, is given thus

$$\hat{\beta}_{LIU} = \left(X'X + I_p\right)^{-1}(X'X + dI_p)\hat{\beta}_{OLS} \qquad (35)$$

where $d \in (-\infty, \infty)$.

Autocorrelation

The following estimators are used to correct autocorrelation in dataset:

(i) The Generalised Two-Stage Estimator

When the assumption of independence of error terms of Classical Linear Regression Model is violated, there is problem of autocorrelation. Suppose the error terms follow AR(1), the model becomes:

$$Y_t = \beta_0 + \beta_1 X_{1t} + \cdots + \beta_p X_{pt} + U_t \qquad (36)$$

where $U_t = \rho U_{t-1} + \varepsilon_t$, $V(U_t) = \sigma^2\Omega$

In this case, the estimator $\hat{\beta}$ of β is given as:

$$\hat{\beta} = (X'\Omega Y)^{-1}X'\Omega' Y \qquad (37)$$

$$V(\hat{\beta}) = \sigma^2(X'\Omega^{-1}X)^{-1} \qquad (38)$$

This is called Generalised Least Square estimator. Since Ω is not always known, therefore, it has to be estimated. There are different methods of doing this. If model is estimated using OLS estimator and $\hat{\rho}$ is further estimated using its residual and it is then used to transform the data, the resulting estimator is called Generalised Least Square estimator. An example of this is the Cochrane-Orcutt estimator, which is an iterative method of estimation (Ayinde, 2006).

(ii) Feasible Generalised Least Square Estimator (Cochrane-Orcutt)

The steps involved are as follows:

1. Estimating of ρ : This is accomplished by noting that the autoregressive errors process can be viewed as a regression through the origin

$$\in_t = \rho \in_{t-1} + U_t \qquad (39)$$

where $\in_t$ are the response variable and $\in_{t-1}$ the predictor variable and U_t the error term.

2. Fitting the transformed model

Using the estimator of ρ , we next obtain the transformed variable Y'_t and X'_t in

$\{Y'_t = Y_t - \rho Y_{t-1}$, $X'_t = X_t - \rho X_{t-1}\}$ and use OLS with these transformed variables to yield the fitted regression function in

$$\widehat{Y'} = B'_0 + B'_1 X' \qquad (40)$$

3. Using Durbin-Watson to test for need to iterate. If transformed model are uncorrelated, the iteration stops (Gujarati, 2007).

Estimation under Joint Violations of Linear Regression Model

This method is used for estimation when the dataset has more than one violation. It involves combination of robust methods. The followings are some of those methods for estimation under joint violation of linear regression model:

Generalised Two-Stage [OLS, LAD] Estimator

The estimator is developed following the technique of Generalized Two Stage estimator to handle both problem of autocorrelation and non-normality of error terms. The procedure is same as that Generalised Two-Stage except that the LAD estimator is used to analyze the data at the second stage.
The methodology are as follows:

1st Stage:- (i) Run the LAD estimator of the model parameters and obtain the Durbin-Watson Statistic, $d = \frac{\sum_{t=2}^{n}(U_t - U_{t-1})^2}{\sum_{t=1}^{n} U_t}$, using the residual of the LAD estimator.

(ii) Estimate $\hat{\rho}$ from (i) using the relation $d = 2(1-\hat{\rho})$. Alternatively, $\hat{\rho}$ obtained from relation $d = \frac{\sum_{t=2}^{n} \hat{U}_t \hat{U}_{t-1}}{\sum_{t=2}^{n} \hat{U}_t^2}$ could be used to transform the data where, $\widehat{U}_t$ are the residual as a result of OLS estimator.

2nd Stage:- (i) Use the $\hat{\rho}$ in 1st Stage to transform the data.

(iii) Use the LAD estimator to estimate the parameters of the transformed variable model.

Result of Stationary Test

Unit Root Test

Table 1 shows the result of the unit root test for stationarity in all variables using the Augmented Dickey Fuller (ADF). All the variables are stationary at 1% level of significance.

Table 1: Unit Root Test

VARIABLE		ADF VALUE	P-VALUE	LAG	REMARK
Expenditure	Y	-5.5751	0.0000	1	Stationary
PMS	X_1	-4.9998	0.0004	0	Stationary
Diesel	X_2	-4.3576	0.0025	1	Stationary
Fuel Oil	X_3	-3.8736	0.0131	1	Stationary
Kerosene	X_4	-7.2990	0.0000	1	Stationary
Asphalt	X_5	-3.7374	0.0036	1	Stationary
Benzene (LAB)	X_6	-3.7244	0.0038	1	Stationary
Base Oil	X_7	-5.4171	0.0000	1	Stationary
Gas	X_8	-6.0651	0.0000	1	Stationary
ATK	X_9	-4.5547	0.0001	1	Stationary

Source: Computer Output

Handling Autocorrelation Multicillinearity, heteroscedaticity and Non-Normality of Error Terms Using Generalised Two-Stage (OLS, LAD) Estimator

The Generalized Two-Stage (OLS,LAD) estimator is used. The results are presented in Table 2. From Table 2, it can be seen that the following were corrected: autocorrelation (Durbin-Watson P-Value > 0.05), heteroscedasticity (Glejser het. P-Value > 0.05) and multicollinearity (all VIF's < 10) and problem of non-normality has been corrected by LAD estimator used. Thus, the effect of kerosene, base oil and ATK is significant on the monthly expenditure, using a data collected from subsidiary of Nigerian National Petroleum Corporation.

Table 2: Generalised Two Stage (OLS, LAD) Estimator Output

VARIABLE		EST. OF REG. COEF.	STAND ERROR	T RATIO	P-VALUE	VIF
	C	22.9258	3.16045	7.2539	0.000	
PMS	X_1T	0.06239	0.04475	1.3942	0.166	1.432
Diesel	X_2T	-0.0052	0.02977	-0.2190	0.827	4.287
Fuel Oil	X_3T	0.01621	0.02935	0.55056	0.583	4.228
Kerosene	X_4T	0.04388	0.00965	4.54564	0.000	1.165
Asphalt	X_5T	0.01519	0.01610	0.94370	0.347	5.628
Benzene (LAB)	X_6T	-0.02509	0.01609	-1.5595	0.122	6.445
Base Oil	X_7T	-0.02033	0.01048	-1.9400	0.096	1.321
Gas	X_8T	0.0000	0.00737	-1.2788	0.204	1.186
ATK	X_9TS	0.75070	0.01774	42.0003	0.000	1.149

R-squared	**0.8942**	**DW Statistic**	**1.8292**	**Glejser het. test**	**7.2251**
Adj R-squared	**0.8854**	**DW P-value**	**0.7724**	**Glejser P-value**	**0.6140**
				SBIC	**372.32**

Source: Computer Output

Conclusion

In this work, monthly expenditure of NNPC, Kaduna has been modeled on the sales of Premium Motor Spirit (PMS), X_1; Automotive Gas Oil (AGO) also known as Diesel, X_2; Fuel Oils, X_3; Kerosene for household use,X_4;Asphalt also known Bitumen,X_5; Linear Alkyl Benzene (LAB), X_6; Base Oils, X_7; Liquefied Petroleum Gas (LPG)$X_8$and Air Transport Kerosene (ATK), X_9, using linear regression model. Various violations of the model assumptions including non-normality of error terms, autocorrelation, multicollinearity and heteroscedasticity are evident in the OLS regression output. All violations detected in the dataset had been corrected and Generalised Two-Stage (OLS, LAD) estimator was used to handle afore-mentioned OLS violations. Results reveal that the sales of kerosene and ATK have positive and significant effect on monthly expenditure while that of base oil is negative and significant.

Consequently, the sale of kerosene and ATK has been observed to determine monthly expenditure.

Recommendation

One of the principal objectives of any organization is to reduce the cost of production. Expenditure is a part of cost of production, which needs to be minimized. Hence, the sales of kerosene, base oil and ATK should be looked into, since it is observed that they give significant contribution to the Corporation's Monthly expenditure.

Contribution to Knowledge

Development of Generalised Two-Stage (OLS, LAD) estimator to handle autocorreation and non-normality of error terms/outliers

References

Abby, L. D.; Jeffrey, A. H.; Ara M.; Khatchadourian, R. T. U. and Melissa, C. W. (2006). Monthly Labour Review, United State on Petroleum Prices and Expenditure. A publication of US Labour Statistics. 2, 34-38

Abiodun, O. F. and Solomon, A. O. (2010): Determination of World Price for Nigerian Major Agricultural Commodities. Journal of Economic Theory: 4(4), 84-92.

Adejumo, A. O.; Ogunleye, T. A.; Olawuwo, S. and Ojo, T. O. (2014).Comparison of Classical Least Squares (CLS), Ridge and Principal Component Methods of Regression Analysis using Gynaecological Data.IOSR Journal of Mathematics. 9(6), 61 - 74.

Al-Hassan, Y. M. (2009).A Monte Carlo Comparison between Ridge and Principal Components Regression Methods.Applied Mathematical Sciences. 3(42), 2058 – 2098.

Anderson, T. W. and Darling, D. A. (1952): Asymptotic Theory of Certain "goodness of fit Criteria Based on Stochastic Process". Annals of Mathematical Statistics 23, 193-212

Alin, C. and Roccardo J. L. (2010): Gretl User's Guide. Gnu Regression, Econometrics and Time-Series Library. 72, 45-67

Alheety, M. I. and Kibria, B. M. (2009): On the Liu and Almost Unbiased Liu Estimators in the Presence of Multicollinearity with Heteroscedastic or Correlated Error. Surveys in Mathematics and its Applications. 4. 155-167.

Ayinde, K. (2006). A comparative study of the performances of the OLS and some GLS estimators when stochastic regressors are both Stochastic and Collinear, West Afr. J. Biomaths. 2, 56-57

Anatolyev, S. (2003). Durbin-Watson Statistic and Random Individual Effect. Cambridge University Press, Economics Theory. 19, 53-58

Anderson, T. W. (1962): "On the Distribution of the Two-Sample Cramer–von Mises Criterion" (PDF). The Annals of Mathematical Statistics (Institute of Mathematical Statistics). 33(3), 1148–1159.

Brown P. J. (1994).Measurement, Regression and Calibration.Oxford.

Bolton, S. and Dons, C. (2004). Pharmaceutical Statistics: Practical and Clinical Applications, Fourth Edition, Revised and Expanded Edition. 135, 345-386

Christopher, B., Victor, H. J. A., Fausto, R. M. and Robert, W. M. (2013). Landscape Genetics of Leaf-Toad Geckos in Tropical Dry Forest in Northern Mexico. American Bio-jounal. 86, 56-58.

Coin, A., Perissinotto, E., Enzi, G., Zamboni, M. and Inelmen, E. M., (2007) Predictor of Low Bone Mineral Dendity in the Elderly; the Role of Dietary Intake, Nutrition. 63, 802-809

Cochrane, D. and Orcutt, G. H. (1949).Application of Least Squares Regression to Relationships Containing Auto-Correlated Error Terms.Journal of the American Statistical Association. 44 (245), 32-61.

Development Core Team (2012): 'R: A Language and Environment for Statistical Computing, Vienna'. R Foundation for Statistical Computing.

Dimitrius, Asterios and Stephen, G. Hall (2011): "Applied Econometrics", 2nd Edition, Macmillan

Dickey, D. A. and Fuller, W. A. (1979): Distribution of the Estimator for Autoregressive Time Series with a Unit Root. Journal of American Statistical Association 74(366), 427-431.

Drag , J. S., Drag, B. T., and Nils, G. K. (2007). A Caution Note on Use of Kolmogorov-Smirnov Test for Normality. America Metrological Society. 35, 32-38.

Draper, N. R. and Smith, H. (1998): "Applied Regression Analysis". Wiley Series in Probability and Statistics.

Draper, N. R. and Van Nostrada (1977a,b): "Ridge Regression and James Sten Estimates, Review and Comment". Technical Report. 501.

El-Dereny, M. and Rashwan, N. I. (2011): Solving Multicollinearity Problem using Ridge Regression Model. International Journal of Contemporary Mathematical Sciences.. 6(12), 585-600.

Everitt, B. S. (2002). The Cambridge Dictionary of Statistics, Second Edition, Cambridge University Press.

Farrar, D. E. and Glauber, R.R. (1967). Multicollinearity in Regression Analysis: The Problem Revisited. Review of Economics and Statistics. 49, 92-107.

Falukasi, B. and Awomuse, B. O. (2011): Determinants of Import in Nigeria: Application of Error Correlation Model: Centrepoint Humanities Edition. 14 (1), 55-72

Fisher, J. (2006): Income Imputation and the Analysis of Expenditure Data in the Consumer Expenditure Survey. BLS Working Papers. 394.

Fukuda, Y. and Hiyoshi, A. (2013):Associations of Household Expenditure and Marital Status With Cadiovascular Risk Factors in Japanese Adults: Analysis of National Representative Surveys. J Epidemol. 23(1), 21-27.

Galton, F. (1877) Regression towards Mediocrity in Hereditary Status. Paper of Nature Philosophy.

George G.Judge, R. Carter Hill, William E. Griffiths, Helmut Lutkepohl, and Tsoung-Chao Lee (1982). Introduction to the Theory and Practice of Econometrics, JohnWiley and Sons, New York. p. 621.

Gujarati, D. N. (2007): "Basic Econometrics". Fourth Edition, McGraw – Hill.

Guy, B. Mia, H. and Anja, S., (2004). A Robustification of Jarque-Bera test of Normality COMPST Symposuim.

Hoerl, A. E. and Kennard, Robert W. (1970): Ridge Regression: Application to Non-Orthogonal Problems". Technometrics, 12(1), 69 – 82.

Hoerl, A. E; Kennard, R. W. and Baldwin J. (1975): " Ridge Regression: Biased Estimation to Non-Orthogonal Problems". Technometrics. 12(4), 56 – 67.

Hoerl, A. E' and Kennard, R. W. (1976): Ridge Regression: Application Biased Estimation to Non- Orthogonal Problems. Technometrics. 12(3), 49 – 58.

Jarque, C. M. and Bera, A. K. (1987). A Test for Normality of Observations and Regression Residual. International Statistical Review: 55, 163-172.

Jesmin, A. (2014). Bootstapped Durbin-Watson Test for autocorrelation for Small Sample. ABC Journal of Advanced Researce. 3(2), 167-184

Judge, G. R.; Carter H., William E. G.; Helmut L. and Tsoung-Chao, L. (1998): Introduction to Theory and Practice of Econometric. John Wiley & sons New York (Second Edition)

Jonathan F. (2006): Income Imputation and Analysis of Expenditure Data in Consumer Expenditure Survey. Bureau of Labour Statistics Working Paper 394.

Johnston, J. (1988): Economic Methods 4th Edition MicGrawHil.

Kmenta, J. (1986). Elements of Econometrics, 2nd ed., Macmillan, New York, p. 431.

Krutchkoff, R.C. (1970): Probability and Statistical Inference. Gordon B. Reach Science Publication Ltd., 12 Bloomsbury Way London W.C.I.

Kumar, T. K. (1975). Multicollinearity in Regression Analysis, Review of Economics and Statistics. 57. 366-368.

Kleiber, C. and Kramer, W. (2004). Finite Sample of Durbin-Watson Test against Functionally Integrated Disturbances. Techniche Universitat Dormuud, Sonderforschungberach. 15, 471-475

Klien, L. R. (1962). An Introduction to Econometrics, Pretence-Hall, Englewood Cliffs, N.J., p. 101.

Lewis, J. (2014). Income, Expenditure and Personal Well-being, 2011/12. Office for National Statistics

Lilliefors, H. (1969). On the Kolmogorov- Smirnov Test for Normality with Mean and Variance Unknown. Journal of the American Statistical Association. 64, 399-402

Liu, K. (1993). A New Class of Biased Estimate in Linear Regression.Communications in Statistics-Theory and Methods. 22(2), 1657-1734

Makolm, S. L. (1992). The Gender Wage Gap in the 1990's. A Publication of University of New South Wales

Marquardt, D. W. and Snee, R. D. (1975): Ridge Regression in Practice. American Statistician, Vol. 29(1), pp. 3–20.

Mehmood, R. and Sadiq, S. (2010). The Relationship between Goverment Expenditure and Poverty: A cointegration analysis. Romanian Journal of Fiscal Police. 1(1), 29-37.

Nigerian National Petroleum Corporations (NNPC) (2014). Official website.

O'Hagan, J. and McCabe, B. (1975). Test for the severity of Multicollinearity in Regression Analysis: A Comment, Review of Economics and Statistics. 57, 368-370.

Odularu G. O. (2006) : Crude Oil and the Nigerian Economic Performance. A Publication of Department of Economics and Development Studies, Covenant University, Ogun State, Nigeria.

Olawuwo, S.; Ogunleye, T.A. and Olaleye, M.O. (2013):Economic Analysis and Modeling of Local Governments' Monthly Expenditure on Income in the South West Zone of Nigeria.IJRRAS Journal. 17(2), 342-367.

Ogunleye, T.A., Olaleye, M.O. and Solomon, A.Z. (2014).Econometric modelling of commercial banks' expenditure on the sources of profit maximization in Nigeria.Scholars Academic and Scientific Publishers, Sch J. Econs Bus Management. 1(7), 276-290.

Ojo, T. O., Ogunleye, T. A., Olawuwo, S. and Adeleke, M. O. (2014). '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. 9(6), 52-60.

Olukotun, G. A.; James, O. O. and Olorunfemi, A. (2013). The Intoduction of GSM Services in Anyigba Community and its impact on Students' Expenditure Pattern. Global Journal of Management and Business Research Finance. 13(8), 2249-4588.

Olowa, O. W. and Olowa, O. A. (2015). Effect of Remittance on Financial Sector Development in Nigeria. Department of Agricultural Education, Federal College of Education (Technical), Akoka, Nigeria.

Onwe, O. J. (2012): Economic Implications of Petroleum Policies in Nigeria: An Overview America International Journal of Contemporary Research. 2(5), 2761-2856

Pasha, G.R. and Shah, M.A. (2004).Application of Ridge Regression to Multicollinear Data.J. Res. Sci.,15:97-106.

Rawlings, J. O. (1988): Applied Regression Analysis, Wadsworth Books, California (ARA).

Riess, H., Clowes, P., Kroidl, I., Kowuov, D. O., Nsojo, A., Mangri, C., Schule, S. A., Mansmann, U., Geldmacher, C., Mhina, L., Hoelscher, M. and Saathoff, E. (2013). Hookworm Infection and Environmental Factors in Mbeya Region, Tanzania: A Cross-Sectional Population-Based Study: Neglected Tropical Diseases- A Peer-Reviewed Open Access Journal. 10, 1371.

Rousseeuw, P. J. (1964). Least Median of Squares Regression. Journal of the American Statistical Association. 79(388), 871-880.

Samprit, C. and Alli, S. H. (2006): "Regression by Example". A John Wiley and Sons Inc. Publication, 4thEdition.

Searle, S.R. (1971). Linear Models.New York, John Wiley and Sons.

Shapiro, S. S. and Wilk, M. B. (1965): An Analysis of Variance Test for Normality (Complete Sample). Biometrika. 52, 3-4.

Van Mises, R. E. (1928): WahrscheinlichkeitStatistik und Wahrheit, Julius Pringer

White, H. (1980): An heteroscedaticity –Consistent Covariance Matrix Estimator and a Direct Test for Heteroscedaticity. Econometrica. 48(4), 815-838.

Wichers, C. R. (1975). The Detection of Multicollinearity: A Comment. Review of Economics and Statistics. 57, 365-366.

Zoran, B. (2014).Does Economic Factor Inpact Economic Growth? Decomposing the Effects of Bosnia and Herzegovinia ACTA ECONOMICA. 12(2), 34-51.