International Journal of Innovative Mathematics, Statistics & Energy Policies 5(4):10-18, Oct.-Dec., 2017

 www.seahipaj.org ISSN: 2467-852X

Comparison between Multiple Linear Regression And Principal Component Regression Models In The Study of Dynamic Of Inflation On Food Items In Nigeria

Alabi Mudasiru Adebayo[1] & Issa Suleman[2]

[1]Department of Mathematics and Statistics,
Akanu Ibiam Federal Polytechnic Unwana, Afikpo, Ebonyi, Nigeria

[2]Department of Statistics, University of Ilorin, P.M.B. 1515, Ilorin, Nigeria.

Email: [1]alabi_ma@yahoo.com, [2]isstatisticsman01@gmail.com

ABSTRACT
The purpose of this study is to investigate an empirical determination of the variables that influence inflation of food in Nigeria. In Addition, the paper attempts to improve the predictive power of multiple linear regression models using principal components as input for predicting consumer price index for the next month. The result of multiple linear regression models indicated that in the presences of crude oil and Exchange rate, money supply is a poor predictor of consumer price index. Performance indicator such as Coefficient of Determination (R^2), Normalized Absolute Error (NAE), Root Mean Square Error (RMSE), and Coefficient of Variation (CV) were used to measure the accuracy of the models. The result of this analysis showed that, Principal component regression (PCR) model performs better than multiple linear regression (MLR) based on the performance indicators, because PCR had minimum NAE, RMSE, CV and the coefficient of determination (CD) has higher predicted accuracy than MLR. PCR also reducing their complexity and eliminating data co-linearity.
Keywords: Inflation of food; Correlation Matrix; Multicollinearity; Multiple linear regression; Principal Component Regression; Performance Indicator

1.0 INTRODUCTION

The maintenance of price stability is one of the macroeconomic challenges facing the Nigerian government in our economic history. Inflation is characterized by a fall in the value of the country's currency and a rise in her exchange rate with other nation's currencies. By definition, inflation is a persistent and appreciable rise in the general level of prices (Jhingan, 2002).

The persistent increase in general price level in Nigeria in the last two decades has posed a major challenge on monetary management yet a systematic macroeconomic account of the underlying shocks has attracted scant attention in the empirical literature. In addition, to achieve and maintain low inflation, Central Banks need to understand the dynamic nature of inflationary processes in their respective countries. These include the type of shocks that cause inflationary impulses and the nature of propagation mechanism. For the purpose of this research the following factors were considered; Commodity Price Index, Crude Oil Price (petroleum), Money Supply (currency in circulation) and Exchange Rate.

Consumer Price Index is a measure of changes in the purchasing-power of a currency and the rate of inflation. The consumer price index expresses the current prices of a basket of goods and services in terms of the prices during the same period in a previous year, to show effect of inflation on purchasing power. The consumer price index for food over the years in Nigeria constituted a larger proportion of the composite consumer price index and households in developing countries spend more on food relative to overall spending and therefore, food price inflation had played a bigger role in overall inflation.

Oil price fluctuation affects the economy through a number of channels. The financial market channel in particular, has gained the attention of authors while the debate around this issue is often to examine the effect of crude oil price fluctuation on the volatility of financial markets. For the purpose of this research crude oil Price (petroleum) was used.

Money Supply or Currency in Circulation was used in this work is the total amount of monetary assets available in an economy at a specific time. There are several ways to define "money," but standard measures usually include currency in circulation and demand deposits (depositors' easily accessed assets on the books of financial institution.

Exchange Rate in finance also known as a foreign-exchange rate between two currencies. It is the rate at which one currency will be exchange for another. It is also regarded as the value of one country's currency in terms of another currency. Exchange rates are determined in the foreign exchange market which is open to a wide range of different types of buyers and sellers and where currency trading is continuous: 24hours a day except weekends. This research work also used exchange rate of dollar (USD) as one of the considering factors in the dynamics of inflation on food items in Nigeria.

PCA is mostly used for reducing the multiple dimensions associated to multiple linear regression which create new variables called the principal component (PCs) that are orthogonal and uncorrelated to each other. The first PC explains the largest fraction of the original data variability and second PC explains larger fraction than third PC and so on (Abdul-Wahab et al., 2005; Wang and Xiao, 2004; Sousa et al., 2007). Varimax rotation is mostly used to obtain the rotated factor loadings that represent the contribution of each variable to a specific principal component. Principal component regression (PCR) is a method that combines linear regression and PCA. PCR establishes a relationship between the output variable (y) and the selected PC of the input variables (x_i).

In addition, many previous researchers such as Kothai *et al.*, (2008) performed principal component analysis using varimax rotation to identify five major sources contributing to coarse and fine particulate mass. Morandi *et al.* (1991) compared between factor analysis/multiple regression and principal component analysis/regression models, the results indicated that the number and type of the sources resolved by the two approaches were similar. Pires *et al.*(2008) focused on the determination of the parameters that influence the concentration of troposphere ozone using PCA.

The aim of this study is to compare the predictive power of multiple linear regression and principal component regression models for examine the nature of relationship between Crude oil price, Money supply, exchange rate and consumer price index as dynamics of inflation on food in Nigeria.

2.0 Methodology

2.1 Site Description

The data collection for this study was collected from Central Bank of Nigeria website (www.cbn.gov.ng), National Bureau of Statistics website (www.nigerianstat.gov.ng) and International Monetary Fund website (www.imf.org).

2.2 Multiple Linear Regression

Multiple linear regression is one of the modeling techniques to investigate the relationship between a dependent variable and several independent variables. This is a generalisation of the simple linear regression model. In the multiple linear regression model, the error term denoted by ε is assumed to be normally distributed with mean 0 and variance δ^2 (which is a constant). ε is also assumed to be uncorrelated. We assume that the multiple linear regression models have k independent variables and there n are observations. Thus the regression model can be written as (Kovac-Andric *et al*., 2009).

The regression equation used was:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \varepsilon \text{ With } i = 1, 2, \ldots, n \quad (1)$$

Where b_i are the regression coefficients, X_i are independent variables and ε is error associated with the regression. To estimate the value of the parameters, the least squares method was used.

Where: Y = Consumer price index

β_0 = Constant

X_1 = Crude oil price

X_2= Money Supply

X_3= Exchange Rate; $\varepsilon = error\_term$

2.2 Principal Component Analysis

Principal component analysis was used to find a small set of linear combinations of the covariates which are uncorrelated with each other. This will avoid the multicollinearity problem. Besides, it can ensure that the linear combinations chosen have maximal variance. Application of principal component analysis (PCA) in regression has long been introduced by Kendall (1957) in his book on Multivariate Analysis. Jeffers (1967) suggested for regression model to achieve an easier and more stable computation, a whole new set of uncorrelated ordered variables that is the principal components (PCs) be introduced (Lam *et al*., 2010).

2.2 Keiser Meyer Olkin's and Bartlett's test of Sampling Adequacy and measuring the Homogeneity of variance across variables for inflation of food item data.

The Kaiser-Meyer Olkin (KMO) measure is an index for comparing the megnitudes of the observed correlation coefficients to the magnitudes of the partial correlation coefficients. KMO statistic is given by

$$KMO = \frac{\sum_{i\neq j}\sum r_{ij}^2}{\sum_{i\neq j}\sum r_{ij}^2 + \sum_{i\neq j}\sum a_{ij}^2}\text{..........................}(2)$$

Where r_{ij} is the simple correlation coefficient and a_{ij} is the partial correlation coefficient between i and j.

Eigenvalues of the covariance matrix are used by PCA to obtain the independent axes under Gaussian assumption. Eigenvalues generated from PCA, also known as characteristic values, are special scalars associated with a linear equation. The Eigenvalues equation for standardized matrix is of the form

$$|C-\lambda I| = 0\text{...}(3)$$

Where C is the correlation of the standardized data, λ is the eigenvalue, and I is the identity matrix. Variable weights are assigned with equation

$$|C-\lambda I|W = 0\text{...}(4)$$

Where by W denote the matrix of weights. Varimax rotation is often used in investigation to see how groupings of items measure the same concept (Zhang *et al.*, 2008; Ozbay *et al.,* 2011). Contribution of each variable in a particular principal component is represented by factor loadings from the varimax rotation (Ozbay et al., 2011). The PCs are obtained by multiplying standardized data matrix by previously calculated weights, *W*. Then, Bartlett's test of sphericity, as expressed in equation (4), is used to justify applicability of PCA to the data sets in this study. It examines the null hypothesis that the variables in the population are uncorrelated. (Ozbay *et al.,* 2011)

$$\chi_k^2 = \left[n-k-\frac{2(p-k)+7+\dfrac{2}{p-k}}{6}+\sum_{j=1}^{k}\left(\frac{\bar{\lambda}}{\lambda_j}\right)^2 \right] x\left[-\ln\prod\nolimits_{j=k+1}^{p}\lambda_j+(p-k)\ln\bar{\lambda}\right]\text{.............}(5)$$

Where p *the number of components* λ_i denotes the eigenvalues for $k_j th$ component, and n is the number of observations. The $\bar{\lambda}$ can be defined by equation below (sousa et.al., 2007)

$$\bar{\lambda} = \sum_{j=k+1}^{p}\frac{\lambda_j}{p-k}\text{...}(6)$$

2.4 Measure of models performance

Performance indicators were used to evaluate the goodness of fit for the MLR and Principal Component Regression (PCR) to determine which method is appropriate to represent the dynamic of inflation on food item in Nigeria. Performance indicators that are used to determine the best method for predicting consumer price index are normalized absolute error (NAE), root mean square error (RMSE), coefficient of variation, and coefficient of determination (R^2). The equations used are reported by Lu (2003).

3.0 Result and Discussion

This study investigates an empirical determination of the variables that influence inflation of food in Nigeria. This section deals with Regression analysis and Principal Component Regression techniques, and interpretation. The specific variables discussed in this section include: Consumer Price Index (CPI), Crude Oil Price, Money Supply and Exchange Rate.

3.1. Residual Plots for Consumer Price Index (CPI)

The regression computer program outputs for residual plots of *Consumer Price Index* are given in Figure 1 below. The interpretations of the graphs were also discussed.

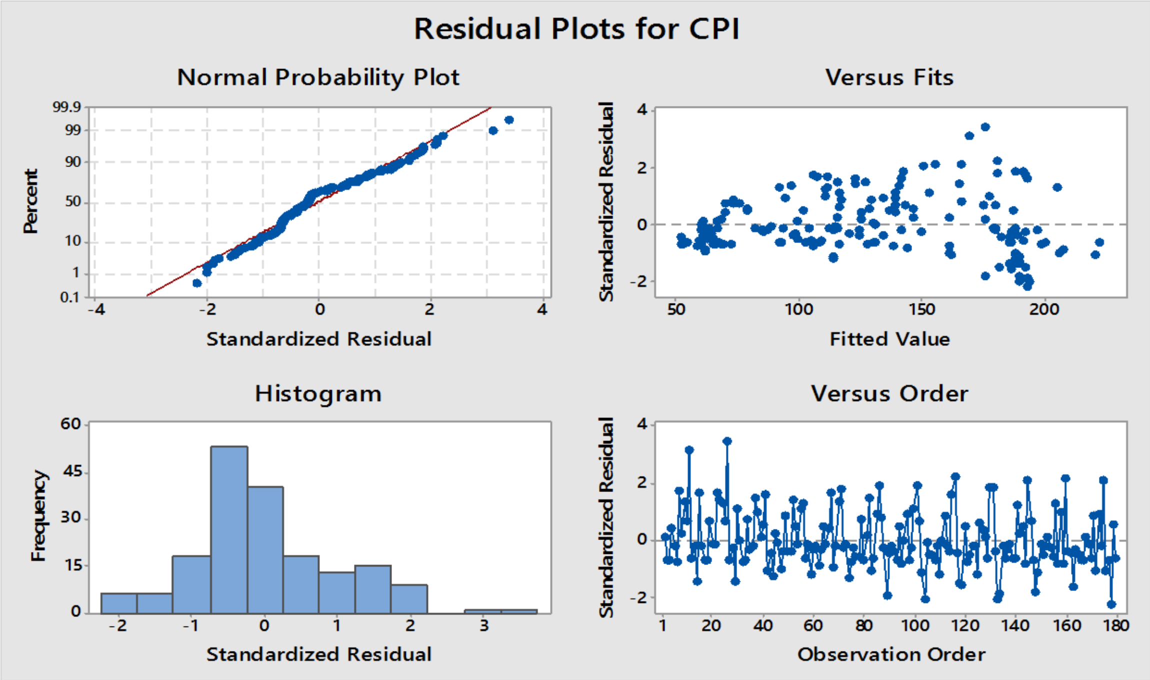

Figure 1 Residual plots for Consumer price Index

Interpreting the Graphs (Figure 1)

(1) From the normal probability plot, we observe that there exists an approximately linear pattern. This indicates the consistency of the data with a normal distribution. The outliers are indicated by the points in the upper-right corner of the plot.

(2) From the plot of residuals versus the fitted values, it is evident that the residuals get smaller, that is, closer to the reference line, as the fitted values increase. This may indicate that the residuals have non-constant variance, (see, Draper and Smith (1998), among others, for details).

(3) The histogram of the residuals indicates that no outliers exist in the data.

(4) The plot for residuals versus order is also provided in Figure 1. It is defined as a plot of all residuals in the order that the data was collected. It is used to find non-random error, especially of time-related effects. A clustering of residuals with the same sign indicates a positive correlation, whereas a negative correlation is indicated by rapid changes in the signs of consecutive residuals.

3.2 Descriptive Statistics and Distribution of Variables

Table 3.1 presents descriptive statistics of the independent variables used in estimating the multiple regression model as well as Principal component regression models. The statistics covers mean values, standard deviation, and a two-sample t-test statistic to compare the means of Consumer Price Index and other variables (Crude oil price, money supply and Exchange rate). The null hypothesis (H_0) in this test is that: "there is no statistical difference between the Consumer Price Index and other variables (Crude oil price, money supply and Exchange rate)".

It is clear from the table that the Consumer Price Index has significant differences in their mean values in terms of money supply and Exchange rate and in term of Crude oil price is not.

Table 3.1 profile Analysis of Means and Standard Deviations of Consumer Price Index, Crude oil price, Money Supply and Exchange Rate

	Consumer Price Index Mean =127.2; Std. Dev. = 48.3				
Variables	Mean	Std. Dev.	Mean Diff.	t-value	p-value
Crude oil price	125.6	48.3	1.60	1.92	0.057
Money Supply	946275	464257	-946148	-27.34	0.000
Exchange Rate	140.1	20.8	-12.89	-3.29	0.001

Note: p-values are meant for testing the null hypothesis that there is no statistical difference between the Consumer Price Index and other variables (Crude oil price, money supply and Exchange rate)

3.3 CORRELATION ANALYSIS

Below in Table 3.2 is a Pearson correlation matrix for all the variables used in estimating the models. Correlation analysis is a possible way of assessing the strength of a group of independent variables as against the dependent variable. It also offers a general idea of the inter relationship between the regressors prior to estimation. It provides an overview about possible multicollinearity problems. From the correlation matrix, all the predictor variables recorded their expected relationship to Consumer Price Index. The Crude oil price, Money supply and Exchange rate show a positive expected relationship with the Consumer price index. All independents variables have statistical significant correlation with consumer price index at 5 percent (α =0.05) significance level. To test for the presence of any multicollinearity problem, we observed that some predictor variables are correlated with other predictors. Various techniques have been developed to identify predictor variables that are highly collinear, and for possible solutions to the problem of multicollinearity, (see, Montgomery and Peck (1982), Draper and Smith (1998), Tamhane and Dunlop (2000), and McClave and Sincich (2006), among others, for details). We can examine the variance inflation factors (VIF), which measure how much the variance of an estimated regression coefficient increases if the predictor variables are correlated. Following Montgomery and Peck (1982), if the VIF is 5 - 10, the regression coefficients are poorly estimated. From output of multiple linear regression based on the raw data, it shows that there is multicollinearity problem with variance inflation factor (VIF) greater than 10, as shown in Table 3.2.

Table 3.2 Correlation Matrix for the Consumer Price Index and other variables (Crude oil price, money supply and Exchange rate) and Significant of correlation

	Consumer Price Index	Crude Oil Price	Money Supply	Exchange Rate	Variance Inflation Factor (VIF)
Consumer Price Index p-value/sig. value	1.00				
Crude oil price p-value/sig. value	0.993** 0.000	1.00			3.60
Money Supply p-value/sig. value	0.759** 0.000	0.713** 0.000	1.00		10.52
Exchange rate p-value/sig. value	0.396** 0.000	0.342** 0.000	0.835** 0.000	1.00	5.85

**P<0.001

3.4 Multiple linear Regressions

Table 3.3 multiple linear regression based on raw data

Predictors variables	Coefficients	Std. Error	t-value	p-value/sig. value
Constant	27.560	4.289	6.425	0.000
Crude oil price	0.758	0.011	71.791	0.000
Money supply	1.414E-05	2.25E-06	6.295	0.000
Exchange Rate	-0.064	0.037	-1.729	0.090
R-squared	0.992	Mean square Error	18.511	
Adjusted R-squared	0.992	Root Mean Square Error	4.302	
F-statistic	7473.48	Coefficient of Variation	0.0338(3.38%)	
Prob(F-statistic)	0.000	Normalization Abs. Error	0.0272	

CPI = 27.56 + 0.7581*COP + 0.000014*MOS - 0.0638*EXR

Where CPI is consumer price Index, COP is Crude Oil Price, MOS is Money Supply, and EXR is Exchange Rate

3.4.1 Interpreting the Results

(i). From Table 3.3, we observe that F-statistic (ANOVA) is 7473.48 with Prob(F-statistic) is 0.000. This implies that the model estimated by the regression procedure is significant at α-level of 0.05. Thus at least one of the regression coefficients is different from zero.

(ii). The p-values for the estimated coefficients of Crude Oil Price (COP) and Money Supply(MOS), are respectively 0.000 and 0.000, indicating that they are significantly related to Consumer price Index (CPI). The p-value for Exchange Rate is 0.090, indicating that it is probably not related to Consumer Price index (CPI) at α-level of 0.05.

(iii). The R-square and Adjusted R-square Statistic: There are several useful criteria for measuring the goodness of fit of the multiple regression models. One such criterion is to determine the square of the

multiple correlation coefficient R^2 (also called the coefficient of multiple determinations), (see, Mendenhall, et al (1993), and Draper and Smith (1998), among others). The R^2 value in the regression output indicates that only 99.2 % of the total variation of the Consumer Price Index values about their mean can be explained by the predictor variables used in the model. The adjusted R^2 value (R^2a) indicates that only 99.2 % of the total variation of the Consumer Price Index values about their mean can be explained by the predictor variables used in the model. As the values of R^2 and R^2a are not very different, it appears that at least one of the predictor variables contributes information for the prediction of Consumer Price Index. Thus both values indicate that the model fits the data well.
(iv). The Mean square Error (MSE), Root Mean square Error (RMSE) and Coefficient of variation (CV) are 18.511, 4.302, and 0.0338 (3.38%) respectively. this statistic indicates that the smaller it is the better, that is, the more precise will be the predictions.

3.4.2. Tests of Significance for Individual Parameters

Table 3.4: Using multiple linear regressions outputs, the analysis of t-statistic values for different parameters (β_i's) are given below:

Null Hypothesis	t(178, 0.975)*	t	Inference	Conclusion
Ho: $\beta_1=0$	1.960	71.791	Reject Ho	In the presence of Money Supply and Exchange Rate, crude oil price is a good predictor of consumer price Index.
Ho: $\beta_2=0$	1.960	1.729	Fail to reject Ho	In the presence of Crude oil price and Exchange Rate, Money Supply is a poor predictor of Consumer Price Index.
Ho: $\beta_3=0$	1.960	6.295	Reject Ho	In the presence of Crude oil price and Money Supply, Exchange Rate is a good predictor of Consumer Price Index.

***For t(178, 0.975), we can use t(∞, 0.975) = 1.96 or interpolate in the *t* – table**

3.5: Principal component Regression

The prediction of dependent variable in the regression model is highly affected from the multicollinearity between the predictor variables. In the presence of multicollinearity, the standard errors of the parameter estimates could be quite high, resulting in unstable parameter estimates. In this case, using this MLR analysis to investigate the relationships between two data sets may not give reliable results. The correlations between the independent variables are in the range of 0.342-0.835. One of the approaches to avoid this problem is PCA. Apart from omitting the correlations of predictor variables, PCA simplifies the complexity of relationship between them. Two important tests for verifying the applicability of PCA were used: i) KMO test for sampling adequacy and ii) Bartlett's test of sphericity for testing all correlations are zero or for testing the null hypothesis that the correlation matrix is an identity matrix. Kaiser (1974) recommends accepting values greater than 0.5, Table 3.5 showed the results of KMO and Bartlett's sphericity test. Overall Kaiser's measure of sampling adequacy is 0.607>0.6, indicated that sample size is enough to apply the PCA. The value of Bartlett's sphericity test is 438.911 and it also implied that PCA is applicable to our data set (P<0.0001) and therefore factor analysis is appropriate for this data.

Table 3.5: KMO Statistics for Sampling Adequate and Bartlett's test for Homogeneity

Test	DF	Approx. Chi-Square	P-value
Keiser-Meyer-Olkin Measure of Sampling Adequate	-	-	.607
Bartlett's Test of Sphericity	3	438.911	0.000

Table 3.6: Total Variance Explained

Component	Initial Eigenvalue			Extraction sums of Squared loadings			Rotation sums of Squared loadings		
	Total	% of variance	Cumulative %	Total	% of variance	Cumulative %	Total	% of variance	Cumulative %
1	2.281	76.040	76.040	2.281	76.040	76.040	1.632	54.389	54.389
2	0.663	22.110	98.150	0.663	22.110	98.150	1.313	43.761	98.150
3	0.055	1.850	100.00						

Table 3.6 lists the eigenvalues associated with each linear component (factor) before extraction, after extraction and after rotation. Before extraction, SPSS has identified two (2) linear components within the data set. The eigenvalues associated with each factor represent the variance explained by the particular linear component and also displays their eigenvalue in term of the percentage of variance explained (so, Factor 1 explains 76.040% of total variance). PCA extracts all Factors with eigenvalues greater than 0.5, the cumulative variance explained by two principal components is 98.15%. The eigenvalues associated with these factors are again displayed in the label extraction sums of squared loading. In the final part of the Table 3.6, the eigenvalues of the factors after rotation are displayed. Rotation has the effect of optimizing the factor structure and one consequence for these data is that the relative importance of the three factors is equalized. Before rotation. Factor 1 accounted for considerably more variance than the remaining other factor (54.389% compared to 43.761%), however after extraction it accounts for only 54.389% of variance compared to 43.761%.

Table 3.7: Loadings of variables in selected principal components

Variables	Loadings after varimax rotation Component	
	1	2
Crude oil price		0.975
Money supply	0.791	
Exchange Rate	0.933	

Rotated matrix rotation using varimax rotation with Kaiser Normalization is shown in Table 3.7. This matrix contains the loading of each variable onto each factor where values less than 0.4 are suppressed from the output. The first factor seems to all relate to Money Supply and Exchange Rate parameters while Second factor from Crude oil price parameter.

Table 3.8 Component Score Coefficient Matrix

	Component	
	PC1	**PC2**
Crude oil price	-0.338	0.947
Money supply	0.380	0.216
Exchange Rate	0.781	-0.367

$$PC1 = -.338 * Crude\_oil\_price + .380 * Money\_Supply + .781 * Exchange\_rate$$

$$PC2 = 0.947 * Crude\_oil\_price + 0.216 * Money\_\sup ply - 0.367 * Exchange\_Rate$$

Table 3.9: Multiple linear Regression based on Principal Component scores Coefficient

Predictor	Coefficient	Std. Error	t	p-value	VIF
Constant	127.245	0.373	340.790	0.000	
PC1	8.385	0.248	33.823	0.000	1.000
PC2	56.967	0.460	123.884	0.000	1.000
R-squares(R^2)	0.994	Mean Square Error	16.215		
Adj. R-Square(R^2)	0.994	Root Mean Square Error	4.026		
F-statistic	8245.633	Coefficient of variation	0.0325(3.25%)		
Prob(F-statistic)	0.000	Normalization Abs..Error	0.0262		
Dubin-Watson(DW)	1.425				

Multiple linear regression analysis was repeated by using principal component analysis as inputs. Here coefficient of determination (R^2) is 0.994. The Value for variance Inflation Factor (VIF) for the independent variables is 1 indicating no multicollinearity problem. Durbin Watson statistic shows that the model does not have any first order autocorrelation problem (DW=1.425). The residual analysis shows that the residuals are distributed normally with zero mean and constant variance. Two main factors from PCA were used as independent variables and the following model was obtained. The principal component scores of selected PCs (PC1-PC2) are used as predictor variables for MLR analysis. The results revealed that multicollinearity was removed and PC1 and PC2 were found to be statistically significant, as shown in Table 3.9.

The final model can be written as:-

$$Consumer\_price\_Index = 127.245 + 8.385 * PC1 + 56.967 * PC2$$

3.6 Comparison of performance Between PCR and MLR

Performance indicators were used to compare between MLR and PCR for inflation on food item in Nigeria. Table 3.10 shows the performance indicator values. The value of the accuracy measure is Coefficient of Determination. The accuracy measure for PCR is higher than for MLR. The values of the error measures namely Normalized Absolute Error, Root Mean Square Error and Coefficient of variation are smaller for PCR than for MLR. This shows PCR gives better result than MLR based on accuracy measures and error measures. So, PCR should provide a better prediction than MLR.

Table 3.10: Performance Indicator between MLR and PCR models

Performance indicators	MLR	PCR
Coefficient of Determinant (R^2)	0.992	**0.994**
Normalized Absolute Error	0.0272	**0.0262**
Root Mean Square Error	4.302	**4.026**
Coefficient of variation	0.0338	**0.0325**

4.0 CONCLUSION/SUGGESTIONS/FINDINGS

In this paper, multiple linear regression was used to predict the next month of Consumer price Index using predictors such as Crude oil price (COP), Money supply (MOS), and Exchange rate (EXR). Two different approaches were used, considering original data and principal component as inputs. The result showed that the used of principal component as input provides a more accurate result than original data because it reduced the number of inputs and therefore decreased the model complexity. Besides that, the use of principal component (PC) based models was considered more efficient, due to elimination of collinerity problem and reduction of the number of predictor variables.

The quality and reliability of the models were evaluated through performance indicators (Coefficient of determinant (R^2), Normalized Absolute Error (NAE), Root Mean Square Error (RMSE), and Coefficient of Variation (CV)). Assessment of model performance indicated that principal component regression can predict particulate matter better than multiple regressions. Similar conclusions were found by previous studies (Ul-Saufie *et. al.*, 2011; Sousa *et al*., 2007; Ozbay *et al.*, 2011). However models adequacy checked by various statistical methods showed that the developed multiple regression models can also be used for prediction of consumer price index.

REFERENCES

Abdul-Wahab S.A., Bakheit C.S. and Al-Alawi S.M., (2005). *Principal component and multiple regression analysis in modelling of ground-level ozone and factors affecting its concentrations, Environmental Modelling & Software,* 20 (10), pp 1263–1271.

Draper, N. R., and Harry S. (1998). *Applied Regression Analysis (3rd edition).* New York: John Wiley & Sons, INC.

Jeffers J. N.R (1967). *Two case studies in the applicationof principal component analysis. Applied Statistics,* 16, pp 225-236

Kendall M.G (1957).*A course in multivariate Analysis,* London, Griffin

Kothai P.,Saradhi, Prathibha1 P., Philip K. Hopke, Pandit G. G., Puranik V.D. (2008).*Source Apportionment of Coarse and Fine Particulate Matter at Navi Mumbai, India . Aerosol and Air Quality Research*, 8(4), pp 423-436.

Kovac-Andric, E., Brana, J., Gvozdic V., (2009*). Impact of meteorological dactors on ozone concentrations modeled by btime series and multivariate statistical methods. Ecological. Infomatics, 4, pp 117-122.*

Lam, K.C., Tau, T., M.C.K., (2010*). A Material supplier selection model for property developers using fuzzy principal component analysis. Automation in Construction,* 19, pp 608-618

Lu, H. C. (2003*). Estimating the Emission Source Reduction of PM10 in Central Taiwan. Journal of Chemosphere,* 54, pp 805-814.

McClave, J. T., and Sincich, T. (2006). *Statistics (10th edition).* Upper Saddle River, NJ: Pearson Prentice Hall.

Mendenhall, W., James E. R., and Robert J. B. (1993). *Statistics for Management and Economics (7th edition).* Belmont, CA: Duxbury Press.

Montgomery, D. C., and Peck, E. A. (1982). *Introduction to Linear Regression Analysis.* New York: John Wiley & Sons, INC.

Ozbay, Bilge, Keskin, Gulsen Aydin, Dogruparnak, Senay Cetin, Ayberk, Savas (2011*). Multivariate methods for ground-level ozone modeling, Atmospheric reseach, doi.10.1016/j.atmosres.*2011.06.005

Piresa J.C.M., Sousaa S.I.V., Pereiraa M.C., Alvim-Ferraza M.C.M. and Martins F.G. (2008), *Management of air quality monitoring using principal component and cluster analysis—Part I: SO2 and PM10 Atmospheric Environment*, 42(6), February 2008, pp 1249-1260

Sousa, S.I.V., Martins, F.G., Alvim-Ferraz M.C.M. and Pereira M.C. (2007*). Multiple linear regression and artificial neural networks based on principal components to predict ozone concentrations, Environmental Modelling & Software,* 22, pp 97–103.

S. Wang and F. Xiao, (2004*). AHU sensor fault diagnosis using principal component analysis method, Energy and Buildings*, 36(2), pp 147–160.

Tamhane, A. C., and Dunlop, D. D. (2000). *Statistics and Data Analysis: From Elementary to Intermediate (1st edition).* Upper Saddle River, NJ: Pearson Prentice Hall.

Ul-Saufie A.Z, Yahya A.S, Ramli N.A (2011). *Improving multiple linear regression model using principal component analysis for predicting PM10 concentration in Seberang Prai, Pulau* Pinang International Journal of Environmental Sciences Volume 2 No.2. pp 415-422

Zhang, C., Wu L., Luo, Y., Zhang, H., Christie, P., (2008). *Identifying sources of soil inorganic pollutants on regional scale using a multivariate statistical approach: Role of pollutant migration and soil physicochemical properties. Environmental Pollution*, 151, pp 470-476