African Journal of Mathematics and Statistics Studies

ISSN: 2689-5323

Volume 8, Issue 2, 2025 (pp. 16-39)

www.abjournals.org

EXAMINING THE INFLUENCE OF BASKET OF COMMODITIES ON CONSUMER PRICE INDEX USING STEPWISE REGRESSION ANALYSIS

Moffat Victoria Ekpedeme[1] and Iseh Matthew Joshua[2]

[1]Department of Statistics, Akwa Ibom State Polytechnic, Ikot Osurua, Nigeria.
Email: vicekpedeme001@gmail.com

[2]Department of Statistics, Akwa Ibom State University, Mkpat Enin, Nigeria.
Email: matthewiseh@aksu.edu.ng

Cite this article:

Moffat, V. E., Iseh, M. J. (2025), Examining the Influence of Basket of Commodities on Consumer Price Index using Stepwise Regression Analysis. African Journal of Mathematics and Statistics Studies 8(2), 16-39. DOI: 10.52589/AJMSS-ZE5LA6TN

Manuscript History

Received: 13 Feb 2025

Accepted: 17 Mar 2025

Published: 15 Apr 2025

ABSTRACT: *This study adopts the concept of stepwise regression analysis with the aim of generating a model to examine the influence of basket of commodities on consumer price index (CPI) in both rural and urban areas in Nigeria. Data used is obtained from the CBN data bank. The result of analysis shows that X_2(housing, water, electricity, gas and other fuel), X_3(food and non alcoholic beverages), X_5 (alcoholic beverage, tobacco and kola) X_7(clothing and footwear), X_8(health), X_9(furnishing and household equipment maintenance), X_{11}(restaurant and hotels) and X_{12}(education) are the variables that mostly affect the urban area. Also, for the Rural area, it is observed that X_2(housing, water, electricity, gas and other fuel), X_3(food and non alcoholic beverages), X_5 (alcoholic beverage, tobacco and kola), X_7(clothing and footwear), X_8(health), X_9(furnishing and household equipment maintenance), X_4(miscellaneous goods and services), and X_6(transport) are the variables that contribute significantly. Therefore, the independent variables that have contributed significantly and common to both urban and rural areas are X_2, X_3, X_5, X_7, X_8, and X_9. However, it was noticed that, among the indicators in the basket of commodities, X_{11}(restaurant and hotels) and X_{12}(education) have effect only on the urban area while X_4(miscellaneous goods and services) and X_6(transport) were significant in the rural areas only. On this note, the choice of stepwise regression model has really paid off by distinguishing the cause and effect of basket of commodities on CPI for rural and urban areas.*

KEYWORDS: Basket of commodities, Consumer price index, Stepwise regression, Rural area, Urban area.

INTRODUCTION

One important metric for monitoring inflation and the cost of living in a nation's economy is the Consumer Price Index (CPI). Despite the abundance of natural resources, Nigerian consumers have endured economic hardship over the years. Many Nigerian customers can no longer afford the high cost of living and commodity prices (Edo et al., 2020b). Many customers are unable to purchase regular meals throughout the day because of the severe inflation brought on by harsh economic policies, corruption, misuse of public funds, and poor management of the citizens' collective commonwealth (Ashakah, 2022).

The average change over time in the prices that urban and rural consumers pay for a market basket of goods and services is measured by the Consumer Price Index (CPI). It is a crucial economic metric that aids in monitoring inflation and a population's cost of living (Adekoya, 2021). To take into consideration geographical variations in economic conditions and spending habits, the CPI is sometimes separated into urban and rural components. The CPI is determined by the Bureau of Labor Statistics (BLS) using a weighted average of prices for a selection of goods and services. One of the most widely used indicators of inflation and deflation is the CPI. The U.S. Bureau of Labor Statistics states that the CPI report employs a different survey methodology, price samples, and index weights than the producer price index (PPI), which tracks changes in the prices received by American producers of goods and services.

Economic theory, sampling, and other statistical techniques are used in the construction of the CPI, which uses data from other surveys to produce a weighted measure of average price changes in the Nigerian economy. The weighting is done to capture the importance of the selected commodities in the entire index. The production of the CPI requires the skills of economists, statisticians, computer scientists, data collectors, and others. The selection of the market basket of goods and services is crucial to the construction of the price index. The calculation of the CPI is done each month using price data from 10,534 informants spread throughout the nation. The market items currently consist of 740 regularly priced goods and services. The gathering of prices for each item (740 commodities and services) from establishments in each sector (rural or urban) for each state is the initial step in calculating the CPI. Next, prices are averaged throughout the state for each item by sector. The next step is to determine each commodity's basic index using the average price. To determine a relative price, the price of each commodity in the current year is contrasted with the price of a base year.

Nigeria is known for its urban and rural areas with varying economic characteristics and consumption patterns. The CPI in Nigeria, like in other regions, is constructed to monitor price changes in a basket of goods and services that the average consumer in urban and rural areas typically purchases. The CPI is calculated based on a “basket of commodities” or a collection of goods and services that represent what a typical consumer in Nigeria buys during a specific period. This basket includes items such as food, clothing, housing, transportation, education, and healthcare, among others. In rural areas of Nigeria, the basket might have a different composition compared to urban areas because of differing lifestyles and economic conditions. The selected goods and services include recreation and culture, housing, water, electricity, gas and other fuel, food and non-alcoholic beverage, miscellaneous goods and services, alcoholic beverage, tobacco and kola, transport, clothing and footwear, health, furnishing and household equipment maintenance, communication, restaurant and hotels, and education.

Article DOI: 10.52589/AJMSS-ZE5LA6TN
DOI URL: https://doi.org/10.52589/AJMSS-ZE5LA6TN

The idea of using the CPI as a measurement tool was first introduced during World War I when an astronomical surge in commodity prices led to the introduction of an index to calculate the cost-of-living adjustments. Sasu (2022) opined that the CPI examines movements in the value of a currency and measures changes in the price level of consumer goods and services purchased by households. Fernando (2022) posited that the CPI measures the monthly changes in the prices paid by consumers; it serves as the major measure of inflation rates. Inflation is calculated by the rate at which the basket price changes over some time. According to AVA Trade (2022), period reports of the CPI help consumers to analyze changes in prices of individual commodities and assist economists to put in place strategies to minimize the effects of inflation.

Unlike the simple linear regression, multivariate regression is a method used to measure the degree at which more than one independent variable (predictors) and more than one dependent variable (responses), are linearly related (Rencher, 1995). The method is broadly used to predict the behavior of the response variables associated with changes in the predictor variables, once a desired degree of relation has been established (Altman et al., 2015). Most regression models are described in terms of the way the outcome variable is modeled; in linear regression, the outcome is continuous. Logistic regression has a dichotomous outcome as adopted by Iseh and Udoh (2022) to model the risk of infant mortality using logistic regression, and survival analysis involves a time to event outcome. Statistically speaking, multivariate analysis refers to statistical models that have two or more dependent or outcome variables (Van, 2004), and multivariate analysis refers to statistical models in which there are multivariate independent or response variables. This type of statistical model can be used to attempt to assess the relationship between a number of variables; one can assess independent relationships while adjusting for potential confounders (Katz, 2003; Obozinski et al., 2011; Chang & Welsh, 2023). Multivariate regression tries to find out a formula that can explain how factors in variables respond simultaneously to changes in others (Gunst, 1980; Davis et al., 1995).

The indoor radon concentration was the response variable in Akerblom's (1986) multivariate stepwise regression analysis, whereas the housing and geological characteristics were the predictor factors. Screening models that can be used to forecast indoor radon levels are the regression equations that are produced. Only two reliable variables contributed statistically significantly to the explanation of the variation in indoor radon levels out of the numerous predictor variables that were included in the regression analysis. The multivariate stepwise regression model is used in this study to streamline the factors that significantly contribute to the basket of commodities in Nigeria's rural and urban areas in order to determine the factors that actually affect the country's consumer price index.

Statement of the Problem

Nigerian consumers over the years have suffered long periods of economic hardship amidst the presence of many natural resources. There is a huge gulf in the living conditions of Nigerians when compared to massive natural endowments that remain economic assets in Nigeria. The current economic hardship in Nigeria can be attributed to failed economic policies, mismanagement of the vast natural resources, corruption in high places, and the sleaze of public funds. There is a problem when it comes to addressing the discrepancy between consumer price index (CPI) and basket of commodities in Nigeria. The divergence results from various factors, including the evolving consumption patterns and the local economic conditions which might

Article DOI: 10.52589/AJMSS-ZE5LA6TN
DOI URL: https://doi.org/10.52589/AJMSS-ZE5LA6TN

lead to inflation**.** Multivariate regression analysis would be used to know what basket of commodities really affects CPI in rural and urban areas.

Aim and Objectives of the Study

The study is aimed to generate a multivariate regression model using stepwise approach to examine the influence of basket of commodities on consumer price index (CPI) in both rural and urban areas.

The objective of the study includes to:

i. fit multivariate regression model for both rural and urban areas on CPI using stepwise regression technique; and

ii. find the effect that the basket of selected commodities has on CPI in rural and urban areas.

METHODOLOGY

The Multivariate Regression Model

The multivariate regression model is the model where multivariate refers to the dependent variables and multiple patterns to the independent variable. In this case, several y's are measured corresponding to each set of x's. each of $y_{1,}y_2,\ldots,y_m$ is to be predicted by all of $x_{1,}x_2,\ldots,x_q$

Multivariate regression is a statistical analysis technique used to quantify the relationship between multiple independent (predictors) and dependent (response) variables in a dataset. The goal is to determine how each independent variable contributes to the variability in the dependent variable, while considering the interrelationships among the predictors.

Then n observed values of the vector of y's can be listed as rows in the matrix.

$$Y = [y_{1'}\ y_{2'}\ \vdots\ y_{m'}\ ]$$

The n values of $x_{1,}x_2,\ldots,x_q$ can be placed in a matrix same as the x matrix in the multiple regression formulation

$$X = \left[1\ x_{11}\ \ x_{12}\ 1\ x_{21}\ \ x_{22}\ \ \cdots\ x_{1q}\ x_{2q}\ \ \vdots \ddots \vdots\ 1\ x_{n1}\ \ x_{n2}\ \ \cdots\ x_{nq}\ \right]$$

Since each of the $y's$ will depend on the $x's$ in its own way, each column of y will need different $\beta's$. Thus a column of $\beta's$ for each column of y, will be a matrix $\beta = \beta_{1,}\ \beta_2,\ldots,\beta_p$. The multivariate model is therefore

$$Y = X\beta + E$$

where Y is nxm, X is $n \times (q+1)$ and β is $(q+1)m$. the notation E (the equivalent of ϵ) is the error term nxm.

Article DOI: 10.52589/AJMSS-ZE5LA6TN
DOI URL: https://doi.org/10.52589/AJMSS-ZE5LA6TN

The model for the first column of Y is where X is fixed from sample to sample

$$[y_{11}\ y_{21}\ y_{31}\ \vdots\ y_{n1}\ ] = [1\ x_{11}\ x_{12}\ldots x_{1q}\ 1\ x_{21}\ x_{22}\ldots x_{2q}\ 1\ \vdots\ 1\ x_{31}\ \vdots\ x_{n1}\ x_{32}\ldots x_{3q}\ \vdots\ \vdots\ x_{n2}\ x_{nq}\ ][\beta_{01}\ \beta_{11}\ \beta_{21}\ \vdots\ \beta_{q1}\ ] + [\epsilon_{11}\ \epsilon_{21}\ \epsilon_{31}\ \vdots\ \epsilon_{n1}\ ]$$

$$y_1' = \beta_{01} + \beta_{11}X_1' + \beta_{21}X_2' + \ldots + \beta_{q1}X_q' + \epsilon_1'$$

The model for the second column of Y is

$$[y_{12}\ y_{22}\ y_{32}\ \vdots\ y_{n2}\ ] = [1\ x_{11}\ x_{12}\ldots x_{1q}\ 1\ x_{21}\ x_{22}\ldots x_{2q}\ 1\ \vdots\ 1\ x_{31}\ \vdots\ x_{n1}\ x_{32}\ldots x_{3q}\ \vdots\ \vdots\ x_{n2}\ x_{nq}\ ][\beta_{02}\ \beta_{12}\ \beta_{22}\ \vdots\ \beta_{q2}\ ] + [\epsilon_{12}\ \epsilon_{22}\ \epsilon_{32}\ \vdots\ \epsilon_{n2}\ ]$$

$$y_2' = \beta_{02} + \beta_{12}X_1' + \beta_{22}X_2' + \ldots + \beta_{q2}X_q' + \epsilon_2'$$

The model for the m^{th} column of Y is

$$[y_{1m}\ y_{2m}\ y_{3m}\ \vdots\ y_{nm}\ ] = [1\ x_{11}\ x_{12}\ldots x_{1q}\ 1\ x_{21}\ x_{22}\ldots x_{2q}\ 1\ \vdots\ 1\ x_{31}\ \vdots\ x_{n1}\ x_{32}\ldots x_{3q}\ \vdots\ \vdots\ x_{n2}\ x_{nq}\ ][\beta_{0m}\ \beta_{1m}\ \beta_{2m}\ \vdots\ \beta_{qm}\ ] + [\epsilon_{1m}\ \epsilon_{2m}\ \epsilon_{3m}\ \vdots\ \epsilon_{nm}\ ]$$

$$y_m' = \beta_{0m} + \beta_{1m}X_1' + \beta_{2m}X_2' + \ldots + \beta_{qm}X_q' + \epsilon_m'$$

Least Squares Estimation in the Multivariate Model

The assumptions that leads to good estimates are

$E(Y) = X\beta\ or\ E(E) = 0$

$Cov(y_i) = \Sigma\ \ for\ all\ i = 1,2,\ldots,n$

$Cov(y_i, y_j) = 0\ for\ all\ i \neq j$

The least squares estimators of the multivariate case

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

where $\hat{\beta}$ is the least squares estimator for β because it minimized

$E = E'E, a\ matrix$

$E = E'E - (Y - X\hat{\beta})'(Y - X\beta)$

Article DOI: 10.52589/AJMSS-ZE5LA6TN
DOI URL: https://doi.org/10.52589/AJMSS-ZE5LA6TN

African Journal of Mathematics and Statistics Studies
ISSN: 2689-5323
Volume 8, Issue 2, 2025 (pp. 16-39)
www.abjournals.org

In the estimate $\beta = (X'X)^{-1}\,(X'Y)$, the matrix product $(X'X)^{-1}X'$ is multiplied by each column of Y therefore the estimate is given as

$$\hat{\beta} = (X'X)^{-1}\,X'(y_{1,}\,y_{2,}\;y_3\,,\ldots,\;y_p)$$

$$= [(X'X)^{-1}\,X'y_{1,}\,(X'X)^{-1}\,X'y_{2,}\,(X'X)^{-1}\,X'y_3\;,\ldots,\big((X'X)^{-1}\,X'y_{p,}\big)]$$

$$= [\widehat{\beta}_1, \hat{\beta}_2, \hat{\beta}_3, \ldots, \hat{\beta}_p]$$

Stepwise Regression Analysis

Stepwise regression is a step-by-step iterative construction of a regression model that involves the selection of independent variables to be used in a final model. It involves adding or removing potential explanatory variables in succession and testing for statistical significance after each iteration.

The steps are:

1. The model with no explanatory variable (null model; $\hat{y} = 0$) from the correlation coefficient between the dependent variable and each of the independent variables, the absolute value of the independent variable with the correlation coefficient that is the highest, and the significance value P< 0.05 will be added to the null model (call model 1)

2. From the excluded model, the independent variable with the highest absolute correlation coefficient value and with significance value less than 0.05 is added to the model 1 (call model 2).

3. Repeat 2, if the significance value of the independent variables in (2) are greater than 0.05, stop. Hence, the model is the final model.

Statement of Hypothesis

$H_0\colon \beta_i = 0 \qquad H_1\colon \beta_i \neq 0$

Decision/conclusion: At α% level of significance, if $\Lambda \leq \Lambda\alpha, p, q, n-q-1$ reject the null hypothesis and conclude that $\beta_i \neq 0$

Test of Significance of Association between Y's and X's

The most widely used measures of association between two sets of variables are the canonical correlations. Canonical correlations analysis is used to determine the amount of (linear) relationship between two sets of variables. Assuming that the two set of variables are $y' = (y_{1,}y_2\,,\ldots,y_m)$ and $x' = (x_{1,}x_2,\ldots,x_q)$, we consider the hypothesis of independence, $H_0\colon \sum_{xy} = 0$. If $\sum_{xy} = 0$, the covariance of every y_i with every x_i is zero and all corresponding correlations likewise zero. Hence, under H_0, there is no linear relationship between the $y's$ and $x's$ and H_0 is equivalent to the statement that all canonical correlations $r_{1,}r_2\,,\ldots,r_n$ are non-significant. Furthermore, H_0 is equivalent to the overall regression hypothesis. $H_0\colon \beta_i = 0$ which also relates to all the $y's$ and to all the $x's$. The significance $r_{1,}r_2\,,\ldots,r_n$ is tested by;

Article DOI: 10.52589/AJMSS-ZE5LA6TN
DOI URL: https://doi.org/10.52589/AJMSS-ZE5LA6TN

$$\Lambda = \frac{|S|}{|S_{yy}|\,|S_{xx}|}$$

where;

$$S = \left(S_{yy}\ S_{yx}\ S_{xy}\ S_{xx}\right)$$

Statement of Hypothesis

H_0; no significant association between $y's$ and the $x's$

H_1; there is significant association between $y's$ and the $x's$

Decision/Conclusion: At α% level of significance. If $\Lambda \leq \Lambda\alpha, p, q, n-q-1$ reject the null hypothesis and conclude that there is significant association between $y's$ and $x's$.

Assumptions of Multivariate Regression Model

1. *Linearity*: The relationship between the independent variables and the dependent variable is assumed to be linear. This means the change in the dependent variable is proportional to changes in the independent variables.

2. *Endogeneity*: There should be no correlation between the error term and any of the independent variables. Endogeneity can cause a bias to the estimates and affect the validity of the model.

3. *Multicollinearity*: The independent variables should not be highly correlated with each other. Multicollinearity can make it difficult to determine the individual effect of each variable and can lead to unstable estimates.

4. *Homoscedasticity (Constant Variance)*: The error term should have a constant variance across all levels of the independent variables. Heteroscedasticity (unequal variances) can affect the accuracy of the regression coefficients and significance tests.

5. O*mitted variables or relevant variables*: All relevant variables should be included in the model, and irrelevant variables should be excluded. Omitted variables can introduce bias into the estimates.

6. *Large Sample Size*: Ideally, a large sample size is assumed to obtain reliable estimates and inferential statistics.

Article DOI: 10.52589/AJMSS-ZE5LA6TN
DOI URL: https://doi.org/10.52589/AJMSS-ZE5LA6TN

African Journal of Mathematics and Statistics Studies
ISSN: 2689-5323
Volume 8, Issue 2, 2025 (pp. 16-39)
www.abjournals.org

DATA ANALYSIS AND RESULT

Source of Data Collection

The data is obtained from the Central Bank of Nigeria Statistical website on consumer's price index for urban rural areas from 2015 to 2023.

Variable description

Some symbols will be used to represent the variables;

Dependent variables

Y_1 represent urban CPI

Y_2 represent rural CPI

Independent variables

X_1 = Recreation and culture

X_2 = Housing, Water, Electricity, Gas and other fuel

X_3 = Food and Non - Alcoholic Beverage

X_4 = Miscellaneous Goods and Services

X_5 = Alcoholic Beverage, Tobacco and Kola

X_6 =Transport

X_7 = Clothing and footwear

X_8 = Health

X_9 = Furnishing and Household equipment maintenance

X_{10} = Communication

X_{11} = Restaurant and Hotels

X_{12} = Education

Article DOI: 10.52589/AJMSS-ZE5LA6TN
DOI URL: https://doi.org/10.52589/AJMSS-ZE5LA6TN

African Journal of Mathematics and Statistics Studies
ISSN: 2689-5323
Volume 8, Issue 2, 2025 (pp. 16-39)
www.abjournals.org

MODEL FORMULATION

The multivariate stepwise regression model approach was applied because the response variables are continuous variables. The SPSS software was used in obtaining the model and also to carry out a stepwise analysis whereby the non-significant independent variable(s) was excluded to obtain a model with significant independent variables.

Multivariate Regression Model

Table 1: Coefficient (Parameters) of Model 1 for Y_1 (urban)

Coefficients[a]								
Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	Collinearity Statistics	
		B	Std. Error	Beta			Tolerance	VIF
1	(Constant)	-14.737	6.978		-2.112	.037		
	X_1	-.032	.077	-.020	-.414	.680	.006	172.174
	X_2	.288	.039	.329	7.381	.000	.007	142.033
	X_3	.333	.105	.373	3.167	.002	.001	995.186
	X_4	-.115	.057	-.105	-2.017	.047	.005	192.578
	X_5	.495	.100	.301	4.948	.000	.004	265.440
	X_6	-.036	.051	-.033	-.707	.481	.006	159.753
	X_7	.176	.078	.162	2.248	.027	.003	374.042
	X_8	-.176	.066	-.147	-2.680	.009	.005	214.421
	X_9	-.196	.079	-.184	-2.491	.015	.003	390.826
	X_{10}	.042	.048	.017	.859	.393	.034	29.238
	X_{11}	.083	.041	.066	2.021	.046	.013	75.429
	X_{12}	.280	.082	.249	3.405	.001	.003	383.257
a. Dependent Variable: Y_1(URBAN)								

The model for CPI urban (y_1) as a multiple regression is given as:

$$y_1 = \beta_{01} + \beta_{11}x_1 + \beta_{21}x_2 + \beta_{31}x_3 + \beta_{41}x_4 + \beta_{51}x_5 + \beta_{61}x_6 + \beta_{71}x_7 + \beta_{81}x_8 + \beta_{91}x_9 + \beta_{10\,1}x_{10} + \beta_{11\,1}x_{11} + \beta_{12\,1}x_{12}$$

$$\hat{y}_1 = -14.737 - 0.032x_1 + 0.288x_2 + 0.333x_3 - 0.115x_4 + 0.495x_5 - 0.036x_6 + 0.176x_7 - 0.176x_8 - 0.196x_9 + 0.042x_{10} + 0.083x_{11} + 0.280x_{12}$$

Article DOI: 10.52589/AJMSS-ZE5LA6TN
DOI URL: https://doi.org/10.52589/AJMSS-ZE5LA6TN

African Journal of Mathematics and Statistics Studies
ISSN: 2689-5323
Volume 8, Issue 2, 2025 (pp. 16-39)
www.abjournals.org

Table 2: Coefficient (Parameters) Model 2 Y_2 (RURAL)

Coefficients[a]								
Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	Collinearity Statistics	
		B	Std. Error	Beta			Tolerance	VIF
1	(Constant)	9.883	6.570		1.504	.136		
	X1	-.068	.073	-.039	-.932	.354	.006	172.174
	X2	.108	.037	.112	2.929	.004	.007	142.033
	X3	.704	.099	.721	7.115	.000	.001	995.186
	X4	.156	.054	.129	2.901	.005	.005	192.578
	X5	-.194	.094	-.108	-2.064	.042	.004	265.440
	X6	.201	.048	.171	4.215	.000	.006	159.753
	X7	-.336	.074	-.283	-4.559	.000	.003	374.042
	X8	.230	.062	.174	3.709	.000	.005	214.421
	X9	.261	.074	.224	3.527	.001	.003	390.826
	X10	-.055	.046	-.021	-1.202	.233	.034	29.238
	X11	-.002	.039	-.002	-.061	.952	.013	75.429
	X12	-.097	.078	-.079	-1.255	.213	.003	383.257
a. Dependent Variable: Y2(RURAL)								

The model for CPI rural (y2) as a multiple regression is given as:

$$y_2 = \beta_{02} + \beta_{12}x_1 + \beta_{22}x_2 + \beta_{32}x_3 + \beta_{42}x_4 + \beta_{52}x_5 + \beta_{62}x_6 + \beta_{72}x_7 + \beta_{82}x_8 + \beta_{92}x_9 + \beta_{10\,2}x_{10} + \beta_{11\,2}x_{11} + \beta_{12\,2}x_{12}$$

$$\hat{y}_2 = 9.883 - 0.068x_1 + 0.108x_2 + 0.704x_3 + 0.156x_4 - 0.194x_5 + 0.201x_6 - 0.336x_7 + 0.230x_8 + 0.261x_9 - 0.055x_{10} - 0.002x_{11} - 0.097x_{12}$$

The multiple regression models of CPI for the Urban and Rural are as follows;

$$\hat{y}_1 = -14.737 - 0.032x_1 + 0.288x_2 + 0.333x_3 - 0.115x_4 + 0.495x_5 - 0.036x_6 + 0.176x_7 - 0.176x_8 - 0.196x_9 + 0.042x_{10} + 0.083x_{11} + 0.280x_{12}$$

$$\hat{y}_2 = 9.883 - 0.068x_1 + 0.108x_2 + 0.704x_3 + 0.156x_4 - 0.194x_5 + 0.201x_6 - 0.336x_7 + 0.230x_8 + 0.261x_9 - 0.055x_{10} - 0.002x_{11} - 0.097x_{12}$$

Stepwise Regression Analysis

The forward selection of stepwise will be used in carrying out the analysis where to include any independent variable; the correlation between the independent and dependent variable is considered. The independent variable with the highest correlation coefficient with the dependent variable will be selected to the null model.

Article DOI: 10.52589/AJMSS-ZE5LA6TN
DOI URL: https://doi.org/10.52589/AJMSS-ZE5LA6TN

Stepwise Regression Analysis for Model 1

The null model is given as:

$$\hat{y} = \beta_0$$

Table 3: Coefficients for the Stepwise Analysis under CPI Urban (Model 1)

Coefficients[a]								
Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	Collinearity Statistics	
		B	Std. Error	Beta			Tolerance	VIF
1	(Constant)	-10.230	1.004		-10.187	.000		
	X7	1.079	.008	.997	132.465	.000	1.000	1.000
2	(Constant)	.339	1.238		.274	.785		
	X7	.575	.049	.531	11.738	.000	.013	74.229
	X3	.418	.040	.469	10.354	.000	.013	74.229
3	(Constant)	5.504	1.387		3.969	.000		
	X7	.300	.063	.277	4.746	.000	.006	164.873
	X3	.434	.035	.487	12.377	.000	.013	74.668
	X2	.209	.036	.238	5.870	.000	.013	79.547
4	(Constant)	8.915	1.570		5.678	.000		
	X7	.320	.059	.296	5.391	.000	.006	166.170
	X3	.612	.057	.686	10.814	.000	.005	222.046
	X2	.224	.034	.256	6.683	.000	.012	80.677
	X9	-.251	.065	-.236	-3.865	.000	.005	206.118
5	(Constant)	2.457	2.722		.903	.369		
	X7	.226	.066	.209	3.413	.001	.005	221.294
	X3	.580	.056	.650	10.388	.000	.004	231.613
	X2	.215	.033	.245	6.596	.000	.012	81.520
	X9	-.206	.065	-.193	-3.174	.002	.005	219.510
	X5	.151	.053	.092	2.857	.005	.016	61.545
6	(Constant)	6.292	3.081		2.042	.044		
	X7	.304	.072	.281	4.221	.000	.004	274.723
	X3	.641	.060	.719	10.698	.000	.004	280.932
	X2	.216	.032	.246	6.798	.000	.012	81.536
	X9	-.198	.063	-.186	-3.129	.002	.005	220.050
	X5	.139	.052	.085	2.675	.009	.016	62.144
	X8	-.171	.070	-.142	-2.449	.016	.005	208.359
7	(Constant)	-2.611	4.527		-.577	.566		
	X7	.251	.073	.232	3.447	.001	.003	297.819
	X3	.516	.075	.578	6.844	.000	.002	471.197
	X2	.231	.031	.263	7.361	.000	.012	84.306
	X9	-.295	.072	-.278	-4.114	.000	.003	300.449
	X5	.304	.081	.185	3.767	.000	.006	159.034

Article DOI: 10.52589/AJMSS-ZE5LA6TN
DOI URL: https://doi.org/10.52589/AJMSS-ZE5LA6TN

	X8	-.185	.068	-.154	-2.729	.008	.005	209.767
	X12	.201	.077	.178	2.620	.010	.003	305.253
8	(Constant)	-12.652	5.601		-2.259	.026		
	X7	.201	.072	.186	2.786	.006	.003	315.928
	X3	.312	.102	.350	3.065	.003	.001	924.388
	X2	.254	.031	.289	8.114	.000	.011	90.287
	X9	-.206	.076	-.194	-2.717	.008	.003	361.225
	X5	.427	.089	.260	4.804	.000	.005	207.800
	X8	-.193	.065	-.160	-2.943	.004	.005	210.102
	X12	.215	.074	.191	2.908	.005	.003	306.717
	X11	.108	.038	.086	2.861	.005	.016	63.697
a. Dependent Variable: Y1(URBAN)								

From Table 3, the correlation coefficient value between (Y_1 and X_3) and (Y_1 and X_7) is higher; therefore, either X_3 or X_7 will be introduced to the null model. The stepwise model under urban is as follows:

$$\hat{y}_1 = -10.230 + 1.079x_7 \qquad \text{model 1.1}$$

Table 4: Excluded Variables (Urban)

					Partial correlation	Collinearity statistics		
Model	model	Beta in	t	sig		Tolerance	vif	Minimum tolerance
1	X1	.354[b]	5.919	.000	.509	.012	84.303	.012
	X2	.200[b]	3.108	.002	.297	.013	79.078	.013
	X3	.469[b]	10.354	.000	.719	.013	74.229	.013
	X4	.263[b]	4.232	.000	.390	.013	79.337	.013
	X5	.129[b]	2.304	.023	.224	.017	57.504	.017
	X6	.118[b]	1.952	.054	.192	.015	65.897	.015
	X8	.291[b]	3.833	.000	.358	.009	115.681	.009
	X9	.304[b]	5.503	.000	.482	.014	69.309	.014
	X10	-.031[b]	-.838	.404	-.084	.041	24.321	.041
	X11	.147[b]	6.032	.000	.517	.071	14.174	.071
	X12	.195[b]	5.329	.000	.470	.033	30.075	.033
2	X1	.099[c]	1.683	.095	.167	.008	126.261	.008
	X2	.238[c]	5.870	.000	.508	.013	79.547	.006
	X4	-.044[c]	-.757	.451	-.076	.008	124.055	.008
	X5	.131[c]	3.465	.001	.329	.017	57.506	.008
	X6	.115[c]	2.792	.006	.270	.015	65.899	.007
	X8	-.169[c]	-2.292	.024	-.224	.005	205.131	.005
	X9	-.188[c]	-2.579	.011	-.251	.005	203.231	.005
	X10	-.010[c]	-.387	.700	-.039	.041	24.473	.010
	X11	-.016[c]	-.534	.595	-.054	.031	32.639	.006
	X12	-.131[c]	-2.764	.007	-.268	.012	86.082	.005
3	X1	.068[d]	1.329	.187	.133	.008	127.643	.005
	X4	-.155[d]	-3.010	.003	-.291	.007	138.029	.006

Article DOI: 10.52589/AJMSS-ZE5LA6TN
DOI URL: https://doi.org/10.52589/AJMSS-ZE5LA6TN

	X2	.117[d]	3.594	.001	.341	.017	57.790	.005
	X6	-.018[d]	-.401	.690	-.040	.010	100.585	.006
	X8	-.174[d]	-2.757	.007	-.268	.005	205.165	.005
	X9	-.236[d]	-3.865	.000	-.364	.005	206.118	.005
	X10	-.026[d]	-1.143	.256	-.115	.040	24.811	.005
	X11	.029[d]	1.059	.292	.106	.028	35.384	.006
	X12	-.105[d]	-2.544	.013	-.249	.011	87.086	.005
4	X1	.052[e]	1.073	.286	.108	.008	128.675	.004
	X4	-.093[e]	-1.716	.089	-.172	.006	163.766	.004
	X5	.092[e]	2.857	.005	.279	.016	61.545	.004
	X6	.009[e]	.199	.842	.020	.010	103.307	.004
	X8	-.157[e]	-2.643	.010	-.259	.005	206.349	.004
	X10	-.017[e]	-.805	.423	-.081	.040	25.104	.004
	X11	.001[e]	.040	.968	.004	.026	38.450	.002
	X12	-.037[e]	-.791	.431	-.080	.009	117.338	.004
5	X1	.008[f]	.155	.877	.016	.007	145.211	.004
	X4	-.098[f]	-1.890	.062	-.189	.006	163.985	.004
	X6	-.053[f]	-1.153	.252	-.117	.008	127.470	.004
	X8	-.142[f]	-2.449	.016	-.243	.005	208.359	.004
	X10	-.006[f]	-.277	.782	-.028	.038	26.115	.004
	X11	.077[f]	2.408	.018	.239	.016	63.319	.002
	X12	.163[f]	2.327	.022	.231	.003	303.203	.002
6	X1	.002[g]	.042	.967	.004	.007	145.544	.003
	X4	-.083[g]	-1.617	.109	-.164	.006	166.891	.004
	X6	-.048[g]	-1.059	.293	-.108	.008	127.797	.003
	X10	-.005[g]	-.266	.791	-.027	.038	26.116	.003
	X11	.080[g]	2.568	.012	.255	.016	63.393	.001
	X12	.178[g]	2.620	.010	.260	.003	305.253	.002
7	X1	-.030[h]	-.612	.542	-.063	.006	154.739	.002
	X4	-.114[h]	-2.276	.025	-.229	.006	173.994	.002
	X6	-.076[h]	-1.702	.092	-.173	.007	133.839	.002
	X10	.007[h]	.359	.721	.037	.036	27.696	.002
	X11	.086[h]	2.861	.005	.283	.016	63.697	.001
8	X1	-.006[i]	-.130	.897	-.013	.006	159.768	.001
	X4	-.095[i]	-1.913	.059	-.195	.006	178.350	.001
	X6	-.037[i]	-.802	.425	-.083	.007	152.209	.001
	X10	.007[i]	.377	.707	.039	.036	27.696	.001

a. Dependent Variable: Y1(URBAN)
b. Predictors in the Model: (Constant), X7
c. Predictors in the Model: (Constant), X7, X3
d. Predictors in the Model: (Constant), X7, X3, X2
e. Predictors in the Model: (Constant), X7, X3, X2, X9
f. Predictors in the Model: (Constant), X7, X3, X2, X9, X5
g. Predictors in the Model: (Constant), X7, X3, X2, X9, X5, X8
h. Predictors in the Model: (Constant), X7, X3, X2, X9, X5, X8, X12
i. Predictors in the Model: (Constant), X7, X3, X2, X9, X5, X8, X12, X11

From the excluded variables in Table 4, X_3 has the highest correlation coefficient value, therefore X_3 is added to the model 1.1 and the stepwise model is

Article DOI: 10.52589/AJMSS-ZE5LA6TN
DOI URL: https://doi.org/10.52589/AJMSS-ZE5LA6TN

$\hat{y}_1 = 0.339 + 0.575x_3 + 0.418x_7$ model 1.2

Again, from Table 4, X_2 has the highest correlation coefficient value, therefore, X_2 is added to the model 1.2 as

$\hat{y}_1 = 5.507 + 0.209x_2 + 0.434x_3 + 0.3x_7$ model 1.3

The excluded variables in Table 4 shows that X_9 has the highest correlation coefficient value, therefore, X_9 is added to model 1.3 as

$\hat{y}_1 = 8.915 + 0.224x_2 + 0.612x_3 + 0.3x_7 - 0.251x_9$ model 1.4

From the excluded variables in Table 4, X_5 has the highest correlation coefficient value, therefore, X_5 is added to model 1.4, then

$\hat{y}_1 = 2.457 + 0.215x_2 + 0.58x_3 + 0.151x_5 + 0.226x_7 - 0.206x_9$ model 1.5

Also, the excluded variables in Table 4 shows that X_8 has the highest correlation coefficient value, therefore, X_8 is added to model 1.5, hence

$\hat{y}_1 = 6.292 + 0.216x_2 + 0.641x_3 + 0.139x_5 + 0.304x_7 - 0.171x_8 - 0.198x_9$ model 1.6

From the excluded variables in Table 4, X_{12} has the highest correlation coefficient value, therefore, X_{12} is added to model 1.6 as

$\hat{y}_1 = -2.611 + 0.21x_2 + 0.516x_3 + 0.304x_5 + 0.2514x_7 - 0.185x_8$
$- 0.295x_9 + 0.201x_{12}$ model 1.7

From the excluded variables in Table 4, X_{11} has the highest correlation coefficient value, therefore, X_{11} is added to model 1.7 as

$\hat{y}_1 = -12.652 + 0.254x_2 + 0.312x_3 + 0.427x_5 + 0.201x_7 - 0.193x_8$
$- 0.206x_9 + 0.108x_{11} + 0.215x_{12}$ model 1.8

From the excluded variables in Table 4; since the significant level of all the variables are greater than 0.05, therefore, no variable is selected. Hence, the stepwise analysis ends and the final model is model 1.8.

DISCUSSION FOR FINAL MODEL CPI (URBAN)

It is observed that X_2 (housing, water, electricity, gas and other fuel), X_3 (food and non alcoholic beverages), X_5 (alcoholic beverage, tobacco and kola), X_7 (clothing and footwear), X_8 (health), X_9 (furnishing and household equipment maintenance), X_{11} (restaurant and hotels) and X_{12} (education) are the variables that most affect the dependent variable y_1 and are significant.

Article DOI: 10.52589/AJMSS-ZE5LA6TN
DOI URL: https://doi.org/10.52589/AJMSS-ZE5LA6TN

Table 5: Coefficients for the Stepwise Analysis under CPI Rural (Model 2)

Coefficients[a]											
Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	Correlations			Collinearity Statistics	
		B	Std. Error	Beta			Zero-order	Partial	Part	Tolerance	VIF
1	(Constant)	3.250	.650		5.002	.000					
	X3	.975	.005	.999	186.262	.000	.999	.999	.999	1.000	1.000
2	(Constant)	-4.069	1.302		-3.124	.002					
	X3	.704	.044	.721	16.073	.000	.999	.849	.074	.010	96.135
	X4	.337	.054	.279	6.211	.000	.996	.528	.028	.010	96.135
3	(Constant)	-6.320	1.391		-4.544	.000					
	X3	.550	.061	.563	9.087	.000	.999	.674	.039	.005	204.051
	X4	.241	.058	.199	4.133	.000	.996	.384	.018	.008	123.409
	X9	.277	.079	.238	3.511	.001	.998	.333	.015	.004	244.127
4	(Constant)	-6.360	1.342		-4.741	.000					
	X3	.552	.058	.565	9.452	.000	.999	.691	.040	.005	204.073
	X4	.154	.064	.127	2.415	.018	.996	.237	.010	.006	158.483
	X9	.259	.076	.223	3.396	.001	.998	.324	.014	.004	245.707
	X6	.101	.035	.086	2.904	.005	.989	.282	.012	.020	50.368
5	(Constant)	-3.654	1.289		-2.834	.006					
	X3	.649	.055	.665	11.864	.000	.999	.769	.044	.004	229.619
	X4	.203	.057	.168	3.562	.001	.996	.340	.013	.006	162.757
	X9	.256	.067	.220	3.798	.000	.998	.360	.014	.004	245.724
	X6	.216	.038	.184	5.748	.000	.989	.504	.021	.013	74.857
	X7	-.279	.052	-.236	-5.337	.000	.992	-.476	-.020	.007	142.575

Article DOI: 10.52589/AJMSS-ZE5LA6TN
DOI URL: https://doi.org/10.52589/AJMSS-ZE5LA6TN

ISSN: 2689-5323

Volume 8, Issue 2, 2025 (pp. 16-39)

6	(Constant)	-8.062	1.736		-4.645	.000					
	X3	.571	.056	.585	10.172	.000	.999	.720	.036	.004	270.865
	X4	.179	.054	.148	3.287	.001	.996	.318	.011	.006	165.432
	X9	.251	.064	.216	3.942	.000	.998	.373	.014	.004	245.843
	X6	.220	.036	.187	6.199	.000	.989	.535	.022	.013	74.937
	X7	-.373	.056	-.315	-6.662	.000	.992	-.562	-.023	.005	182.942
	X8	.238	.067	.180	3.565	.001	.996	.342	.012	.005	209.741
7	(Constant)	-4.593	1.854		-2.477	.015					
	X3	.585	.053	.599	11.133	.000	.999	.752	.036	.004	272.246
	X4	.145	.052	.120	2.815	.006	.996	.277	.009	.006	170.284
	X9	.266	.060	.228	4.448	.000	.998	.415	.015	.004	246.780
	X6	.138	.039	.117	3.502	.001	.989	.338	.011	.009	105.669
	X7	-.460	.057	-.388	-8.077	.000	.992	-.638	-.026	.005	216.995
	X8	.242	.062	.184	3.891	.000	.996	.371	.013	.005	209.825
	X2	.136	.035	.141	3.864	.000	.989	.369	.013	.008	124.978
8	(Constant)	-.269	2.797		-.096	.924					
	X3	.617	.054	.632	11.426	.000	.999	.762	.037	.003	296.637
	X4	.143	.051	.118	2.825	.006	.996	.280	.009	.006	170.337
	X9	.227	.062	.195	3.687	.000	.998	.355	.012	.004	271.946
	X6	.176	.043	.150	4.093	.000	.989	.389	.013	.008	130.435
	X7	-.401	.063	-.338	-6.348	.000	.992	-.548	-.020	.004	275.432
	X8	.229	.062	.174	3.714	.000	.996	.358	.012	.005	212.265
	X2	.123	.035	.128	3.497	.001	.989	.339	.011	.008	129.156
	X5	-.103	.051	-.057	-2.039	.044	.983	-.206	-.007	.013	76.900
a. Dependent Variable: Y2(RURAL)											

Article DOI: 10.52589/AJMSS-ZE5LA6TN
DOI URL: https://doi.org/10.52589/AJMSS-ZE5LA6TN

African Journal of Mathematics and Statistics Studies
ISSN: 2689-5323
Volume 8, Issue 2, 2025 (pp. 16-39)

Table 6: Excluded Variables (Rural)

Excluded Variables[a]								
Model		Beta In	t	Sig.	Partial Correlation	Collinearity Statistics		
						Tolerance	VIF	Minimum Tolerance
1	X1	.019[b]	.362	.718	.036	.011	93.188	.011
	X2	.146[b]	5.067	.000	.452	.028	35.813	.028
	X4	.279[b]	6.211	.000	.528	.010	96.135	.010
	X5	.000[b]	-.006	.996	-.001	.031	32.473	.031
	X6	.158[b]	5.680	.000	.494	.028	35.348	.028
	X7	.051[b]	1.099	.274	.109	.013	74.229	.013
	X8	.184[b]	3.177	.002	.303	.008	127.318	.008
	X9	.370[b]	5.735	.000	.498	.005	190.175	.005
	X10	.032[b]	1.461	.147	.145	.058	17.350	.058
	X11	-.095[b]	-3.650	.000	-.343	.038	26.229	.038
	X12	.075[b]	1.591	.115	.157	.013	79.122	.013
2	X1	-.036[c]	-.794	.429	-.080	.010	96.837	.006
	X2	.076[c]	2.393	.019	.234	.020	50.532	.007
	X5	-.051[c]	-1.891	.062	-.187	.028	35.385	.010
	X6	.094[c]	3.030	.003	.291	.020	50.044	.007
	X7	-.084[c]	-1.911	.059	-.189	.010	95.786	.008
	X8	.069[c]	1.233	.220	.123	.007	150.379	.006
	X9	.238[c]	3.511	.001	.333	.004	244.127	.004
	X10	-.016[c]	-.762	.448	-.076	.049	20.412	.009
	X11	-.036[c]	-1.370	.174	-.136	.030	33.047	.005
	X12	.060[c]	1.484	.141	.148	.013	79.415	.006
3	X1	-.032[d]	-.737	.463	-.074	.010	96.917	.004
	X2	.074[d]	2.478	.015	.243	.020	50.543	.004
	X5	-.037[d]	-1.417	.160	-.142	.027	36.374	.004
	X6	.086[d]	2.904	.005	.282	.020	50.368	.004
	X7	-.090[d]	-2.169	.033	-.214	.010	95.934	.004
	X8	.062[d]	1.169	.245	.117	.007	150.591	.004
	X10	-.020[d]	-1.036	.303	-.104	.049	20.498	.004
	X11	-.019[d]	-.735	.464	-.074	.029	34.492	.003
	X12	.007[d]	.164	.870	.017	.011	94.553	.003
4	X1	-.143[e]	-3.002	.003	-.292	.007	141.160	.003
	X2	.027[e]	.629	.531	.064	.010	101.696	.004
	X5	-.135[e]	-4.606	.000	-.424	.017	59.625	.004
	X7	-.236[e]	-5.337	.000	-.476	.007	142.575	.004
	X8	.022[e]	.410	.683	.042	.006	163.460	.004
	X10	-.036[e]	-1.855	.067	-.185	.046	21.777	.004
	X11	.034[e]	1.139	.257	.115	.019	52.119	.002
	X12	.079[e]	1.740	.085	.174	.008	120.732	.003
5	X1	-.091[f]	-2.046	.043	-.204	.007	150.289	.003
	X2	.138[f]	3.535	.001	.339	.008	124.928	.004

Article DOI: 10.52589/AJMSS-ZE5LA6TN
DOI URL: https://doi.org/10.52589/AJMSS-ZE5LA6TN

	X5	-.085[f]	-2.778	.007	-.273	.014	73.639	.004
	X8	.180[f]	3.565	.001	.342	.005	209.741	.004
	X10	-.001[f]	-.077	.939	-.008	.039	25.340	.004
	X11	.044[f]	1.658	.101	.167	.019	52.346	.002
	X12	.039[f]	.937	.351	.095	.008	125.392	.003
6	X1	-.082[g]	-1.928	.057	-.194	.007	150.938	.003
	X2	.141[g]	3.864	.000	.369	.008	124.978	.004
	X5	-.075[g]	-2.565	.012	-.254	.013	74.412	.003
	X10	-.002[g]	-.114	.909	-.012	.039	25.342	.004
	X11	.034[g]	1.348	.181	.137	.019	53.034	.002
	X12	.023[g]	.579	.564	.059	.008	127.114	.003
7	X1	-.066[h]	-1.650	.102	-.168	.007	152.646	.003
	X5	-.057[h]	-2.039	.044	-.206	.013	76.900	.003
	X10	-.007[h]	-.450	.654	-.046	.039	25.525	.004
	X11	.024[h]	1.002	.319	.103	.019	53.734	.002
	X12	.031[h]	.827	.410	.085	.008	127.473	.003
8	X1	-.052[i]	-1.279	.204	-.131	.006	159.235	.003
	X10	-.015[i]	-.881	.380	-.091	.038	26.576	.003
	X11	-.001[i]	-.048	.962	-.005	.014	73.600	.001
	X12	-.071[i]	-1.219	.226	-.125	.003	330.103	.002
a. Dependent Variable: Y2(RURAL)								
b. Predictors in the Model: (Constant), X3								
c. Predictors in the Model: (Constant), X3, X4								
d. Predictors in the Model: (Constant), X3, X4, X9								
e. Predictors in the Model: (Constant), X3, X4, X9, X6								
f. Predictors in the Model: (Constant), X3, X4, X9, X6, X7								
g. Predictors in the Model: (Constant), X3, X4, X9, X6, X7, X8								
h. Predictors in the Model: (Constant), X3, X4, X9, X6, X7, X8, X2								
i. Predictors in the Model: (Constant), X3, X4, X9, X6, X7, X8, X2, X5								

X_3 is added to the null model from Table 5, hence the stepwise model is given as

$$\hat{y}_2 = 3.250 + 0.975x_3 \qquad \text{model 2.1}$$

From Table 6, X_4 has the highest correlation coefficient value, therefore, X_4 is added to model 2.1 as

$$\hat{y}_2 = -4.069 + 0.704x_3 + 0.337x_4 \qquad \text{model 2.2}$$

In Table 6, X_9 has the highest correlation coefficient value, therefore, X_9 is added to model 2.2 as

$$\hat{y}_2 = -6.320 + 0.550x_3 + 0.241x_4 + 0.277x_9 \qquad \text{model 2.3}$$

From the excluded variables of Table 6, X_6 has the highest correlation coefficient value, therefore, X_6 is added to model 2.3 as

Article DOI: 10.52589/AJMSS-ZE5LA6TN
DOI URL: https://doi.org/10.52589/AJMSS-ZE5LA6TN

$\hat{y}_2 = -6.360 + 0.552x_3 + 0.154x_4 + 0.101x_6 + 0.259x_9$ model 2.4

Table 6 shows that X_7 has the highest correlation coefficient value, therefore, X_7 is added to model 2.4 as

$\hat{y}_2 = -3.654 + 0.649x_3 + 0.203x_4 + 0.216x_6 - 0.299x_7 + 0.256x_9$ model 2.5

The excluded variables Table 6 has X_8 with the highest correlation coefficient value, therefore, X_8 is added to model 2.5

The model is given as

$\hat{y}_2 = -8.062 + 0.571x_3 + 0.179x_4 + 0.220x_6 - 0.373x_7 + 0.238x_8 + 0.251x_9$ model 2.6

[

Again, Table 6 shows that X_2 has the highest correlation coefficient value, therefore, X_2 is added to model 2.6 as

$\hat{y}_2 = -4.593 + 0.136x_2 + 0.583x_3 + 0.145x_4 + 0.138x_6 - 0.242x_8 + 0.266x_9$ model 2.7

From Table 6, X_5 has the highest correlation coefficient value, therefore, X_5 is added to model 2.7 as

$\hat{y}_2 = -0.269 + 0.123x_2 + 0.617x_3 + 0.143x_4 - 0.103x_5 + 0.176x_6 - 0.401x_7 + 0.229x_8 + 0.227x_9$ model 2.8

From Table 6, since all the significant level values are greater than α= 0.05, therefore no variable will be selected again after model 2.8. Hence model 2.8 is the final model.

Discussion for Final Model CPI (Rural)

It is observed that X_2(housing, water, electricity, gas and other fuel), X_3 (food and non alcoholic beverages), X_5 (alcoholic beverage, tobacco and kola) X_7 (clothing and footwear), X_8(health), X_9 (furnishing and household equipment maintenance), and X_4 (miscellaneous goods and services), and x_6 (transport) are the variables that most affect the dependent variable y_2 and are significant.

From the stepwise regression analysis, the final models for the urban and rural are given as;

$\hat{y}_1 = -12.652 + 0.254x_2 + 0.312x_3 + 0.427x_5 + 0.201x_7 - 0.193x_8$

$- 0.206x_9 + 0.108x_{11} + 0.215x_{12}$

$\hat{y}_2 = -0.269 + 0.123x_2 + 0.617x_3 + 0.143x_4 - 0.103x_5 + 0.176x_6 - 0.401x_7 + 0.229x_8 + 0.227x_9$

Article DOI: 10.52589/AJMSS-ZE5LA6TN
DOI URL: https://doi.org/10.52589/AJMSS-ZE5LA6TN

From the two models, it is observed that the variables $x_2, x_3, x_5, x_7, x_8, x_9$ are significant in the models ($\hat{y}_1$ *and* $\hat{y}_2$) and then x_{11} *and* x_{12} are significant in $\hat{y}_1$model while x_4 and x_6 are significant in $\hat{y}_2$.

Table 7: Multivariate Model with the Intersected Independent Variables

Parameter Estimates							
Dependent Variable	Parameter	B	Std. Error	t	Sig.	95% Confidence Interval	
						Lower Bound	Upper Bound
Y1(URBAN)	Intercept	-.002	.117	-.015	.988	-.235	.231
	X3	.558	.044	12.564	.000	.470	.646
	X2	.193	.030	6.427	.000	.133	.252
	X5	.220	.034	6.469	.000	.152	.287
	X7	.263	.070	3.762	.000	.124	.401
	X8	-.098	.061	-1.620	.109	-.218	.022
	X9	-.145	.058	-2.482	.015	-.260	-.029
Y2(RURAL)	Intercept	-.003	.114	-.028	.978	-.229	.223
	X3	.609	.043	14.119	.000	.524	.695
	X2	.239	.029	8.200	.000	.181	.297
	X5	-.029	.033	-.880	.381	-.095	.036
	X7	-.400	.068	-5.893	.000	-.535	-.265
	X8	.243	.059	4.133	.000	.127	.360
	X9	.346	.057	6.114	.000	.234	.458

$$\hat{y}_1 = -0.002 + 0.193x_2 + 0.558x_3 + 0.220x_5 + 0.263x_7 - 0.098x_8 - 0.145x_9$$

$$\hat{y}_2 = -0.003 + 0.239x_2 + 0.609x_3 - 0.029x_5 - 0.0.400x_7 + 0.243x_8 + 0.346x_9$$

Test of significance of Association between the $y's$ and $x's$

Statement of hypothesis

H_0: No significant association between the $y's$ and $x's$

H_1: There is no significant association between the $y's$ and $x's$

$$\Lambda_I = \frac{|S|}{|S_{yy}|\,|S_{xx}|}$$

$S_{yy} = (Var\ y_1\ Cov\ y_1y_2\ Cov\ y_1y_2\ Var\ y_2)$ = (829.1323 903.4091 903.4091 994.4154)

Article DOI: 10.52589/AJMSS-ZE5LA6TN
DOI URL: https://doi.org/10.52589/AJMSS-ZE5LA6TN

$$|S_{yy}| = |(829.1323 * 994.4154) - (903.4091 * 903.4091)| = 8353.9258$$

$$S_{xx} = \begin{bmatrix} Var\ x_1 & Cov\ x_1x_2 & Cov\ x_2x_3 & \cdots & Cov\ x_1x_{12} \\ Cov\ x_2x_1 & Var\ x_2 & Cov\ x_2x_3 & \cdots & Cov\ x_2x_{12} \\ Cov\ x_3x_1 & Cov\ x_3x_2 & Var\ x_3 & \cdots & Cov\ x_3x_{12} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ Cov\ x_{11}x_1 & Cov\ x_{11}x_2 & Cov\ x_{11}x_3 & \cdots & Cov\ x_{11}x_{12} \\ Cov\ x_{12}x_2 & Cov\ x_{12}x_2 & Cov\ x_{12}x_3 & \cdots & Var\ x_{12} \end{bmatrix}$$

=

3328	591.1513	585.7394	471.3089	316.1865	484.7333	482.3663	433.5258	490.1604	213.1177	400.9947	459.2235
591.1513	1075.86	1044.284	846.3972	566.7003	875.2744	867.2125	776.446	876.291	385.0674	709.9673	817.0142
585.7394	1044.284	1042.75	837.3022	557.0399	854.3852	853.4065	769.924	872.1589	377.1367	720.99428	820.21673
471.3089	846.3972	837.3022	679.3997	449.6338	692.5534	689.1579	620.438	702.7756	305.8047	573.5725	659.0771
316.1865	566.7003	557.0399	449.6338	307.027	465.2864	462.1574	413.512	465.8818	203.4708	376.4565	431.5376
484.7333	875.2744	854.3852	692.5534	465.2864	720.4277	720.4276	634.9676	717.3956	314.2343	577.0399	668.197
482.3663	867.2125	853.4065	689.1579	413.512	720.4276	707.9818	634.1555	715.3284	313.4679	583.9911	668.7573
433.5258	776.446	769.924	620.438	465.8818	634.9676	634.1555	572.9807	644.8602	280.6437	530.1569	605.2182
490.1604	876.291	872.1589	702.7756	203.4708	717.3956	715.3284	644.8602	733.3318	316.7698	599.7726	688.1815
213.1177	385.0674	377.1367	305.8047	376.4565	314.2343	313.4679	280.6437	316.7698	144.7432	258.2884	295.7102
400.9946	709.9673	720.99428	573.5725	431.5376	577.0399	583.9911	530.1569	599.7726	258.2884	518.2807	571.5353
459.2235	817.0142	820.21673	659.0771	431.5376	668.197	668.7573	605.2182	688.1815	295.7102	571.5353	653.4326

$|S_{xx}| = 7403319614438.79$

$S_{yx} =$

$[522.7166\ 522.7190\ 522.7182\ 522.7179\ 522.7176\ 522.7182\ 522.7183\quad 522.7179\ 522.7175\ 522.716$

$S_{xy} = S_{yx}^T$

$|S| = |S_{yy}\ S_{yx}\ S_{xy}\ S_{xx}|$ = merging all the matrices gives a 14 X14 matrix

Article DOI: 10.52589/AJMSS-ZE5LA6TN
DOI URL: https://doi.org/10.52589/AJMSS-ZE5LA6TN

829.1323	**903.4091**	**522.7166**	**522.719**	**522.7182**	**522.7179**	**522.7176**	**522.7182**	**522.7183**	**522.7179**	**522.7175**	**522.7161**	**522.71695**	**522.71674**
903.4091	**994.4154**	**522.3166**	**528.6564**	**526.5171**	**525.7486**	**524.9705**	**526.6329**	**526.8085**	**525.7978**	**524.8674**	**521.2099**	**523.33175**	**522.7848**
522.7166	**522.3166**	**332.5938**	**591.1513**	**585.7394**	**471.3089**	**316.1865**	**484.7333**	**482.3663**	**433.5258**	**490.1604**	**213.1177**	**400.9947**	**459.22356**
522.719	**528.6564**	**591.1513**	**1075.86**	**1044.284**	**846.3972**	**566.7003**	**875.2744**	**867.2125**	**776.446**	**876.291**	**385.0674**	**709.9673**	**817.0142**
522.7182	**526.5171**	**585.7394**	**1044.284**	**1042.75**	**837.3022**	**557.0399**	**854.3852**	**853.4065**	**769.924**	**872.1589**	**377.1367**	**720.99428**	**820.21673**
522.7179	**525.7486**	**471.3089**	**846.3972**	**837.3022**	**679.3997**	**449.6338**	**692.5534**	**689.1579**	**620.438**	**702.7756**	**305.8047**	**573.5725**	**659.0771**
522.7176	**524.9705**	**316.1865**	**566.7003**	**557.0399**	**449.6338**	**307.027**	**465.2864**	**462.1574**	**413.512**	**465.8818**	**203.4708**	**376.4565**	**431.5376**
522.7182	**526.6329**	**484.7333**	**875.2744**	**854.3852**	**692.5534**	**465.2864**	**720.4277**	**720.4276**	**634.9676**	**717.3956**	**314.2343**	**577.0399**	**668.197**
522.7183	**526.8085**	**482.3663**	**867.2125**	**853.4065**	**689.1579**	**413.512**	**720.4276**	**707.9818**	**634.1555**	**715.3284**	**313.4679**	**583.9911**	**668.7573**
522.7179	**525.7978**	**433.5258**	**776.446**	**769.924**	**620.438**	**465.8818**	**634.9676**	**634.1555**	**572.9807**	**644.8602**	**280.6437**	**530.1569**	**605.2182**
522.7175	**524.8674**	**490.1604**	**876.291**	**872.1589**	**702.7756**	**203.4708**	**717.3956**	**715.3284**	**644.8602**	**733.3318**	**316.7698**	**599.7726**	**688.1815**
522.7161	**521.2099**	**213.1177**	**385.0674**	**377.1367**	**305.8047**	**376.4565**	**314.2343**	**313.4679**	**280.6437**	**316.7698**	**144.7432**	**258.2884**	**295.7102**
522.7169518	**523.331752**	**400.9947**	**709.9673**	**720.99428**	**573.5725**	**431.5376**	**577.0399**	**583.9911**	**530.1569**	**599.7726**	**258.2884**	**518.2807**	**571.5353**
522.7167404	**522.784797**	**459.22356**	**817.0142**	**820.21673**	**659.0771**	**431.5376**	**668.197**	**668.7573**	**605.2182**	**688.1815**	**295.7102**	**571.5353**	**653.4326**

$$|S| = 5768151795259350000.0$$

$$|S_{yy}| = 8353.9258$$

$$|S_{xx}| = 7403319614438.79$$

$$\Lambda_I = \frac{5768151795259350000.0}{8353.9258 * 7403319614438.79} = 93.262652$$

Article DOI: 10.52589/AJMSS-ZE5LA6TN
DOI URL: https://doi.org/10.52589/AJMSS-ZE5LA6TN

$\Lambda_{m,q,n-1} = \Lambda_{2,12,(103-1-12)} = \Lambda_{2,12,90,0.05} = 3.24$

m = number of dependent variables

q = $V_H = number\ of\ independent\ variables$

$V_E = n - 1 - q$

DECISION/CONCLUSION:

At 5% level of significance, $\Lambda_I > \Lambda_{m,q,n-1-q}$ $(93.2652 > 3.24)$, we accept H_0. Hence, there is no significant association between the y's and the x's.

Conclusion

This research work focused on fitting a multivariate regression model to examine the influence of basket of commodities on consumer price index (CPI) in both rural and urban areas in Nigeria for a period of eight years (2015-2023). The data was obtained from CBN data bank. The CPI covers both rural and urban areas in Nigeria for this period (2015-2023). Twelve indicators were considered to constitute the basket of commodities. This include X_1(Recreation and culture), X_2(Housing, Water, Electricity, Gas and other fuel), X_3(Food and Non - Alcoholic Beverage), X_4(Miscellaneous Goods and Services), X_5(Alcoholic Beverage, Tobacco and Kola), X_6(Transport), X_7(Clothing and footwear), X_8(Health), X_9(Furnishing and Household equipment maintenance), X_{10}(Communication), X_{11}(Restaurant and Hotels), and X_{12}(Education). The research was subjected to multivariate regression using the SPSS software. In addition, stepwise regression analysis was carried out whereby the non significant independent variables were excluded to obtain models with significant independent variables.

Housing, water, electricity, gas, and other fuels; food and nonalcoholic beverages; alcoholic beverages, tobacco, and kola; footwear and clothing; health; maintenance of household equipment and furnishings; dining establishments and lodgings; and education are found to be the factors that most impact urban areas. Housing, water, electricity, gas, and other fuels; food and nonalcoholic beverages; alcoholic beverages, tobacco, and kola; footwear and clothing; health; maintenance of household equipment and furnishings; miscellaneous goods and services; and transportation are the factors that are found to have the greatest impact in rural areas. Consequently, the independent variables that are prevalent in both rural and urban settings can be thought of as having an impact on the dependent variable.

The following factors were found to be important in both rural and urban areas: housing, water, electricity, gas, and other fuel; food and nonalcoholic beverages; alcoholic beverages, tobacco, and kola; clothing and footwear; health; furnishing and maintenance of household equipment; education has an impact only in urban areas, while miscellaneous goods and services and transportation have an impact only in rural areas.

Article DOI: 10.52589/AJMSS-ZE5LA6TN
DOI URL: https://doi.org/10.52589/AJMSS-ZE5LA6TN

REFERENCES

Adekoya, F. (2021). Inflation, rising input costs spell trouble for manufacturers, consumers. Retrieved @ Https://Guardian.Ng/Category/Business-Services/Industry.

Akerblom, G . (1986). Investigation and mapping of radon risk areas. Swedish Geological Co., Lulea, Sweden, Report TRAP 86036.

Altman, N. and Krzywinski, M. (2015). Simple linear regression. *Nature Methods*, **12** (11): 999–1000.

Ashakah, F. (2022). Nigeria's economy needs a major repair. Business Day Newspaper.

AVATrade (2022). Consumer price index. Retrieved @ Https://Www.Avatrade.Com.

Chang, L. and Welsh, A. H. (2023) Robust Multivariate Lasso Regression with Covariance Estimation. *Journal of Computational and Graphical Statistics, Vol* 32(3):961-973.

Davis, R. O., Drobnis, E. Z. and Overstreet, J. W. (1995). Application of multivariate cluster, discriminant function, and stepwise regression analyses to variable selection and predictive modeling of sperm cryosurvival. Supported by grant RO1 HD-25907 *National Institutes of Health, Bethesda, Maryland.* Vol 63(5):1051-1057.

Edo, O. C., Okafor, A. and Justice, A, E. (2020b). Corporate taxes and foreign direct investments: an impact analysis. *Public Policy and Administration Research*, 10(9): 51-62. DOI: 10.7176/PPAR/10-9-07.

Fernando, J. (2022). Consumer price index (CPI) explained: What it is and How it is Used. Investopedia:Https://Www.Investopedia.Com/Contributors/53746/

Gunst, R. F. and Mason, R. L. (1980). Regression analysis and its application. A data-oriented approach. New York, Marcel Dekker, Inc.

Iseh, M. J. and Udoh, N. S. (2022). Modelling the risk index of infant mortality in Nigeria. Abacus (Mathematics Science Series) Vol. 49(2):350-373.

Katipamula, S., Reddy, T. A., and Claridge, D. E. (1995). Effect of time resolution on statistical modeling of cooling energy use in large commercial buildings. In *ASHRAE Transactions*, Vol. 101(2):172-185.

Katz M. H. (2003). Multivariable analysis: a primer for readers of medical research. *Ann Intern Med*. 2003;138(8):644–650.

Obozinski, G. Wainwright, M. J. and Jordan, M. I. (2011). Support union recovery in high-dimensional multivariate regression. *The Annals of Statistics*, Vol 39(1), 1-47.

Rencher, A. C. (1995). Methods of Multivariate Analysis. New York, John Wiley & Sons, Inc.

Sasu, D. D. (2022).CPI in Nigeria 2022, By Sector. Statista Corporate Solutions.

U.S. Bureau of Labor Statistics. "Consumer Price Index." https://www.bls.gov/cpi/

Van, B. G. (2004). Biostatistics: A Methodology for the Health Sciences. Hoboken, NJ: Wiley-Interscience;

Article DOI: 10.52589/AJMSS-ZE5LA6TN
DOI URL: https://doi.org/10.52589/AJMSS-ZE5LA6TN