International Journal of Science and Advanced Innovative Research
ISSN: 2536-7315 (Print) 2536-7323 (Online)
Volume 3, Number 3, September 2018
http://www.casirmediapublishing.com

PRINCIPAL COMPONENTS ANALYSIS OF NIGERIAN ECONOMIC VARIABLES

Eze-Emmanuel, Peace and Ette Harrison Etuk
Department of Statistic
Rivers State University, Port Harcourt
Email: adelepeace678@gmail.com and ettetuk@yahoo.com

ABSTRACT
Principal Components Analysis (PCA) of Nigeria economic variables was done to decide on the most important variables to be considered in determining those variables that have positive effect on Nigerian economy. It is important to determine the significant proportion of those variables that contributed to the Nigerian Gross Domestic Product because it helps reveal especially with regards to revenue generation by the government. This research work examined the performance of these variables using quarterly Nigeria Gross Domestic data from 1981Q1-2013Q4. The methods used are Principal Components Analysis (PCA) and factor analysis (FA) multivariate technique. Using Minitab 17 statistical software, the data were examined; summary statistics are as follows: covariance matrix, correlation matrix, standard deviation, Eigenvalues, Eigenvectors, transformation of the sample onto new subsample (Sorted Unrotated and Rotated Factors), computing the principal components and finally plotting the graphs. Our results showed that 24 economic variables from 31 variables have almost perfect (positive) effect on the economy and identify the sorted rotated factor loading as the better (appropriate) method, when using economic data variables.
Keywords: Eigenvalues, Eigenvectors, Transformation, Unrotated and Rotated Factors, Principal Components Analysis (PCA) and factor analysis (FA)

INTRODUCTION
PCA is a statistical procedure that uses an orthogonal transformation to convert a set of observations of possibly correlated variables into a set of values of linearly uncorrelated variables called principal components (or sometimes, principal modes of variation). This transformation is defined in such a way that the first principal Components has the largest possible variance (that is, accounts for as much of the variability in the data as possible), and each succeeding component in turn has the highest variance possible under the constraint that it is orthogonal to the preceding components. PCA is used as a tool for making predictive and exploratory data analysis. However, to calculate the PCs, the eigenvalue decomposition of the data covariance (or correlation) matrix is done, usually after the data matrix is normalized (using Z-scores) [Ildiko and Jerome, (1993); Bocuvxa, *et al.* (2005)].

According to Eric *et al.,* (2006), the results of a PCA are usually discussed in terms of Components scores (or factor scores where the transformed variable values correspond to a particular data point), and loadings (the weight by which each standardized original variable should be multiplied to get the Components score). PCA operation can be thought of as revealing the internal structure of the

data in a way that best explains the variance in the data. PCA is a multivariate dataset that can supply the user with a lower-dimensional picture, a projection of this object when viewed from its most informative viewpoint. This is done by using only the first few principal components so that the dimensionality of the transformed data is reduced. In addition, PCA is closely related to factor analysis. Factor analysis typically incorporates more domain specific assumptions about the underlying structure and solves eigenvectors of a slightly different matrix.

The goal of factor analysis is to reduce the redundancy among the variables by using a smaller number of factors. If the pattern of the high and low correlations in the correlation matrix is such that the variables in a particular subset have high correlations among themselves, but low correlations with all the other variables, then there may be a single underlying factor that gave rise to the variables in the subset. In addition, if the other variables can be similarly grouped into subsets with a like pattern of correlations, then a few factors can represent these groups of variables. In this case, the pattern in the correlation matrix corresponds directly to the factors. Also if the correlation matrix does not have such a simple pattern; factor analysis will still partition the variables into clusters. Factor analysis is related to principal components analysis in that both seek a simpler structure in a set of variables but they differ in many respects:

1. Principal Components are defined as linear combinations of the original variables. In factor analysis, the original variables are expressed as linear combinations of the factors.
2. In principal Components analysis, we explain a large part of the total variance of the variables, $\sum_{i=1}^{p} S_{ii}$, where S_{ii} are the variances. In factor analysis, we seek to account for the covariance or correlations among the variables.

Over the years, researchers have given attention to the subject of classification/factoring-out of variables into pre-determined groups or to reduce the redundancy among the variables by using a smaller number of factors. Complex problems and the results of bad decisions frequently force researchers to look for more objective ways to predict outcomes. That is why this research is interested in comparing two factor methods to determine the more suitable one when larger numbers of variables are involved. In addition with the use of principal components analysis and factor analysis, the most important of these variables can be factored out from Nigerian economic variables.

Aim and Objectives of the Study

The aim of this study is to use Principal Components and Factor Analysis to determine the most important Nigerian economic variables amongst others. The objectives of the study are;

i. To identify associating factors between the variables using Principal Components Analysis
ii. To determine the most important variables using the two factor methods techniques
iii. Compare the results obtained using sorted unrotated and rotated factor loadings of Factor Analysis.

It is essential to establish by research (or determine) the main or crucial factors among the economic variables that contribute more to Nigerian Gross Domestic Product (GDP). Hence this will greatly reduce the financial and physical stress the Government of Nigeria will face as the years go by. Principal Components and Factor Analysis will be able to highlight the most relevant factors among the considered factors.

The research work data is collected from National Bureau of Statistics (NBS), which is the Quarterly Gross Domestic product (NGDP) of Nigeria, from 1981 to 2013. There are 31 independent variables with 132 data points (or total observation) for each variable. The research work is based on Principal Components and Factor Analysis of Nigerian economic variables and it will not concern it-self much with how the data are generated since they are from a secondary source. Gerns *et al.,* (2014) observed that the role of HIV-1-specific antibody responses in HIV disease progression is complex and requires an analysis technique that examines grouping of responses. Principal Components Analysis (PCA) was found to reduce the data dimensionality by creating fewer composite variables that maximally account for variance in a dataset. Hence, to identify clusters of antibody responses involved in disease control, they investigated the association of HIV-1-specific antibody responses by protein microarray, and assessed their association with disease progression using PCA in a nested cohort design (Marshal *et al.,* 2005). Associations observed among collections of antibody responses paralleled protein-specific responses. PCA and protein microarray analyses highlighted a collection of HIV-specific antibody responses that together were associated with reduced disease progression, and may not have been identified by examining individual antibody responses.

Libin (2015) discussed the application of PCA to stock portfolio management using analysis of the Australian stock market from 2000 to 2014 data. He found that any combination of stocks cannot be used to depict the whole data set. The variance described by the first principal Components can serve as a primary pointer of financial crisis. The first ten principal Components were retained to

present the major risk sources in the stock market. Vukosi (2015) considered missing values or variables often resulting from data collection. Using data from a 2000 HIV survey data set and comparing 3 distinct data imputation models and identifying their merits by using accuracy measures, he concluded that the use of PCA improves the overall performance of the autoencoder network with accuracies of up to 97.4%.

METHODOLOGY

Principal Components Analysis (PCA)

Principal Components Analysis is a useful statistical technique that has many applications in fields such as face recognition and image compression, and is a common technique for finding patterns in data of high dimension. It is concerned with explaining the variance – covariance, standard deviation, eigen value and eigenvectors structure through a few linear combinations of the original variables. Principal Components are particular (algebraically) i.e. linear combinations of the p random variables $X_1, X_2, \ldots, X_P$. It depends solely on the covariance matrix or correlation matrix of $X_1, X_2, \ldots, X_P$.

Consider the linear combinations

$$\begin{aligned} Y_1 &= L_1X = L_{11}X_1 + L_{21}X_2 + \ldots + L_{P1}X_P \\ Y_2 &= L_2X = L_{12}X_1 + L_{22}X_2 + \ldots + L_{P2}X_P \\ &\vdots \\ Y_P &= L_PX = L_{1P}X_1 + L_{2P}X_2 + \ldots + L_{PP}X_P \end{aligned} \qquad (3.1)$$

where $L_1, L_2, \ldots, L_P$ are row vector and X is column vector, such that $L_1 = (L_{11}, L_{21}, \ldots, L_{P1})$, $L_2 = (L_{12}, L_{22}, \ldots, L_{P2})$, . . ., $L_P = (L_{1P}, L_{2P}, \ldots, L_{PP})$.

Principal Components are those uncorrelated linear combinations Y (i.e. Y is column vector) whose variances are as large as possible. Each component is a linear combination of the *p* variables. The first Components accounts or makes up for the largest possible amount of variance while the second largest amount of variance is accounted or made up for by the second Components, formed from the variance remaining after that related with the first Components has been removed, etc. The restriction (assumption upon which) principal Components are isolated or taken out is that they are orthogonal. They may be geometrically viewed as being dimensions in *m*-dimensional space where each dimension is perpendicular to the other dimension (Ye *et al.,* 2002).

Each of the *m* variable's variances is standardized to one. To see how much more (or less) variance it represents than does a single variable, each factor's eigenvalue may be compared to 1. There is $p \times 1 = m$ variance to distribute with *m* variables. The principal Components extraction will produce *m* Components which in the aggregate account for all of the variances in the *p* variables. That is, the sum of the *m* eigenvalues will be equal to *m*, the number of variables. The proportion of variance accounted for by one Component equals its eigenvalue divided by *p*.

Variance – Covariance Matrix

We recall that covariance is always measured between 2 dimensions. There is usually more than one covariance measurement that may be calculated if we have a data set with more than 2 dimensions. For 3 dimensional data set,(dimension a, b, c), we have: Cov(a, b), Cov(a, c) and Cov(b, c).
Suppose we have a random vector X,

$$\underline{X} = \begin{pmatrix} X_1 \\ X_2 \\ . \\ . \\ . \\ X_p \end{pmatrix} \qquad (3.2)$$

Then the population variance-covariance matrix (*Abdi and Williams, 2010*) is

$$\sigma_i^2 = \frac{\sum_{i=1}\left(X_i - \bar{X}\right)^2}{n-1} = \frac{\sum X_1^2 - n\bar{X}_1^2}{n-1} \qquad (3.3)$$

$$\text{Cov}(X_i, X_j) = \sigma_{ij} = \frac{\sum_{i=1}^{n}\sum_{j=1}^{n}\left(X_i - \bar{X}_i\right)\left(X_j - \bar{X}_j\right)}{n-1} \qquad (3.4)$$

Eigenvalues and Eigenvectors

Many applications of matrices to technological problems involve $A.X = \lambda X$; where $A = [a_{ij}]$ is a square matrix and λ is a number (scalar). Clearly $X = 0$ is a solution for any value of λ and is not normally useful. For non-trivial solution, i.e $X \neq 0$, the values of λ are called the eigenvalues, character values or latent roots (polynomial) of matrix A and the corresponding solutions of the given equations $A.X = \lambda X$ are called the eigen vectors or characteristic vectors of A.
From the linear combination of X_i. i.e

$$\ell_{ij} = \begin{pmatrix} \ell_{i1} \\ \ell_{i2} \\ \ell_{i3} \\ . \\ . \\ . \\ \ell_{ip} \end{pmatrix} \qquad (3.5)$$

i.e

$$AX - \lambda X = 0 \text{ , } (3.6a)$$

where the x's are the eigenvectors and

$$|A-\lambda I| = 0 \qquad (3.6b)$$

λ's are the eigen-values.

Hence the coefficients ℓ_{ij} are collected into the vector.

$$\ell_{ij} = \begin{pmatrix} \ell_{i1} \\ \ell_{i2} \\ \ell_{i3} \\ . \\ . \\ . \\ \ell_{ip} \end{pmatrix} \text{ and } \lambda_i = (\lambda_1, \lambda_2, \ldots, \lambda_p)$$

Are eigenvalues and Eigen-vectors.

First Principal Components Analysis (PCA 1)

Y_1, the first principal Components is the linear combination of x-variables that has maximum variance (among all linear combinations), being that it accounts or make up for as much variation in the data as possible. Specifically we will define coefficient $\ell_{11}, \ell_{12}, \ldots, \ell_{1p}$ for that Components in such a way that its variance is maximized, subject to the constraint that the sum of the squared coefficients is equal to one. This constraint is required so that a unique value may be obtained.

More formally, select $\ell_{11}, \ell_{12}, \ldots, \ell_{1p}$ that maximizes

$$\text{Var}(Y_1) = \sum_{i=1}^{p}\sum_{j=1}^{p} \ell_{1i}\ell_{1j}\sigma_{ij} = \ell_1' \sum \ell_1 \qquad (3.7)$$

Subject to the constraint that

$$\ell_1'\ell_1 = \sum_{j=1}^{p} \ell_{1j}^2 = 1$$

Second Principal Components (PCA 2)

Y_2, the second principal Components is the linear combination of x- variables that accounts for as much of the remaining variations as possible, with the

constraint that the correlation between the first and second Components is 0. Select $\ell_{21}, \ell_{22}, \ldots, \ell_{2p}$ that maximizes the variance of this new Component.

$$\text{Var}(Y_2) = \sum_{i=1}^{p}\sum_{j=1}^{p} \ell_{2i}\ell_{2j}\sigma_{ij} = \ell'_2 \sum \ell_2 \qquad (3.8)$$

Subject to the constraint that the sum of squared coefficients add up to one

$$\ell'_2\ell_2 = \sum_{j=1}^{p} \ell_{2j}^2 = 1$$

Along with the additional constraint that these two Components will be uncorrelated with one another,

$$\text{Cov}(Y_1, Y_2) = \sum_{i=1}^{p}\sum_{j=1}^{p} \ell_{1i}\ell_{2j}\sigma_{ij} = \ell'_1 \sum \ell_2 = 0 \qquad (3.9)$$

All subsequent principal Components have same property, they are linear combinations that account for as much of the remaining variations as possible and they are not correlated with other principal Components.

Proportion

Proportion of variance explained by the k[th] principal component is

$$\frac{\lambda_k}{\lambda_1 + \lambda_2 + \ldots + \lambda_p} \times 100\% \qquad (3.10a)$$

where, λ_k is the k[th] eigenvalues.

Cumulative Proportion

Cumulative proportion of variance explained by the first k[th] principal components is

$$\frac{\lambda_1 + \lambda_2 + \ldots + \lambda_k}{\lambda_1 + \lambda_2 + \ldots + \lambda_p} \qquad (3.10b)$$

Factor Analysis

In factor analysis we represent the variables $y_1, y_2, \ldots, y_p$ as linear combinations of a few random variables $f_1, f_2, \ldots, f_m$ $(m < p)$ called *factors*. The factors are underlying constructs variables that "generate" the y's. If the random sample $y_1, y_2, \ldots, y_n$ from a homogeneous population with mean vector μ and covariance matrix $\sum$.

The factor analysis model expresses each variable as a linear combination of underlying common factors $f_1, f_2, \ldots, f_m$, with an accompanying error term to account for that part of the variable that is unique (not in common with the other variables).

Thus, for $y_1, y_2, \ldots, y_p$ in any observation vector y, the model is given as

$$
\begin{aligned}
y_1 - \mu_1 &= \lambda_{11} f_1 + \lambda_{12} f_2 + \ldots + \lambda_{1m} f_m + \varepsilon_1 \\
y_2 - \mu_2 &= \lambda_{21} f_1 + \lambda_{22} f_2 + \ldots + \lambda_{2m} f_m + \varepsilon_2 \\
&\vdots \\
y_p - \mu_{p1} &= \lambda_{p1} f_1 + \lambda_{p2} f_2 + \ldots + \lambda_{pm} f_m + \varepsilon_p
\end{aligned}
\tag{3.11}
$$

where

Coefficients λ_{ij} are loadings and serve as weights, which shows each y individually depends on the f's (eigenvalues).

Note: *m* should be substantially smaller than *p*; otherwise we have not achieved a parsimonious description of the variables as functions of a few underlying factors.

A simple expression for the variance of y's is

$$\text{var}(y_i) = \lambda_{i1}^2 + \lambda_{i2}^2 + \ldots + \lambda_{im}^2 + \psi_i \tag{3.12}$$

Thus, the emphasis in factor analysis is on modeling the covariances or correlations among the *y*'s. Model (3.11) can be written in matrix notation as

$$y - \mu = \wedge f + \varepsilon \tag{3.13}$$

where

$y = (y_1, y_2, \ldots, y_p)'$, $\mu = (\mu_1, \mu_2, \ldots, \mu_p)'$, $f = (f_1, f_2, \ldots, f_m)'$,

$\varepsilon = (\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_p)'$ *and*

$$\wedge = \begin{pmatrix} \lambda_{11} & \lambda_{12} & \ldots & \lambda_{1m} \\ \lambda_{21} & \lambda_{22} & \ldots & \lambda_{1m} \\ & \vdots & & \\ \lambda_{p1} & \lambda_{p2} & \ldots & \lambda_{pm} \end{pmatrix}$$

The assumptions can be expressed as follow

If $E(f_j) = 0$, $j = 1, 2, \ldots, m$

(1) $E(f) = 0$ (3.14)

$\text{var}(f_j) = 1$, $j = 1, 2, \ldots, m$ and $\text{cov}(f_j, f_k) = 0$, $j \neq k$, therefore

(2) $\text{cov}(f) = \text{I}$ (3.15)

$E(\varepsilon_i) = 0$, $i = 1, 2, \ldots, p$

(3) $E(\varepsilon) = 0$ (3.16)

$\text{var}(\varepsilon_i) = \psi_i$, $i = 1, 2, \ldots, p$ and $\text{cov}(\varepsilon_i, \varepsilon_k) = 0$, $i \neq k$, therefore

$$(4)\quad \operatorname{cov}(\varepsilon) = \psi = \begin{pmatrix} \psi_1 & 0 & \ldots & 0 \\ 0 & \psi_2 & \ldots & 0 \\ & & \cdot & \\ & & \cdot & \\ & & \cdot & \\ 0 & 0 & \ldots & \psi_p \end{pmatrix} \qquad (3.17)$$

and $\operatorname{cov}(\varepsilon_i\ ,\ f_j) = 0$ *for all* i *and* j

However, Equation (3.11) can be written as

$$\begin{aligned} \Sigma &= \operatorname{cov}(y) = \operatorname{cov}(\wedge f + \varepsilon) \\ &= \wedge \operatorname{cov}(f) \wedge' + \psi \\ &= \wedge \mathrm{I} \wedge' + \psi \\ \Sigma &= \wedge \wedge' + \psi \end{aligned} \qquad (3.18)$$

Since μ does not affect variances and covariances of y.

Note: $\wedge$ has only a few columns, say two or three

In general,

$$\operatorname{cov}(y_i\ ,\ f_j) = \lambda_{ij}\ ;\quad i = 1,2,\ldots, p \text{ and } j = 1,2,\ldots,m \qquad (3.19)$$

Since λ_{ij} is the $(ij)^{th}$ element of $\wedge$, Equation (3.19) can be writing as

$$\operatorname{cov}(y\ ,\ f) = \wedge \qquad (3.20)$$

If standardized variables are used, Equation (3.18) is replaced by

$$\rho = \wedge \wedge' + \psi \qquad (3.21)$$

and the loadings become correlations:

$$\operatorname{cov}(y_i\ ,\ f_j) = \lambda_{ij} \qquad (3.22)$$

In partitioning variance of y_i into components due to the common factor in Equation (3.12), called the communality, and a Component unique to y_i called the specific variance:

$$\begin{aligned} \sigma_{ij} = \operatorname{var}(y_i) &= \left(\lambda_{i1}^2 + \lambda_{i2}^2 + \ldots + \lambda_{im}^2\right) + \psi_i \\ &= h_i^2 + \psi_i \\ &= communality + specific\ \operatorname{var} iance \end{aligned} \qquad (3.23)$$

where

$Communality = h_i^2 = \lambda_{i1}^2 + \lambda_{i2}^2 + \ldots + \lambda_{im}^2$ is also called common variance

$Specific\ \operatorname{var} iance = \psi_i$ is also called Specificity, unique variance, or residual variance.

The two factor analysis methods considered were Principal Components Extraction Method (PCEM) and Maximum Likelihood Method (MLM), while MLM is divided into two parts: (a) Sorted and unrotated factor loadings of the factor analysis between the variables using Maximum Likelihood; (b) Sorted and a Rotation of the factor analysis between the variables.

Comparison of MLM Methods

The level of reliability of the Pearson correlation for classification using the characterizations reported by [Sim and Wright (2005); Ogoke*et al.* (2003)]. These characterizations range from 0.00 to 0.20 (Slight), 0.21 to 0.40 (Fair), 0.41 to 0.60 (Moderate), 0.61 to 0.80 (Substantial), 0.81 to 1.00 (Almost Perfect). These characterizations range was used to classify the variables.

Data Analysis

Correlation Analysis of the Variables

Correlation analysis was carried out on the thirty-one variables considered for the study using Micro-excel software. Four variables (metal ores, cement, rail transport and pipeline, water transport)out of the thirty-one variables showed correlation less than absolute of ±0.5. The four variables which showed low correlation were removed and the remaining twenty-seven variables which showed correlation above 0.5 were used for further analysis. The twenty-seven variables which showed a correlation above 0.5, Public administration and Livestock has the highest correlation coefficient of 0.99 followed by water and livestock, real estate and livestock, education and livestock, other services and livestock, health services and livestock, post and road transport, other services and road transport, post and road transport which showed a correlation coefficient of 0.98. This result showed that there is sufficient evidence at $\alpha = 0.01$ that the correlations are not zero. To understand the underlying data structure and/or form a smaller number of uncorrelated variables, principal component analysis was used.

Principal Components Analysis (PCA) analysis

Principal component analysis was employed on the twenty-seven variables using Minitab 17 statistical software (which showed a correlation above absolute of ±0.5). Note that the correlation matrix was used to standardize the measurements because they are not measured with the same scale.

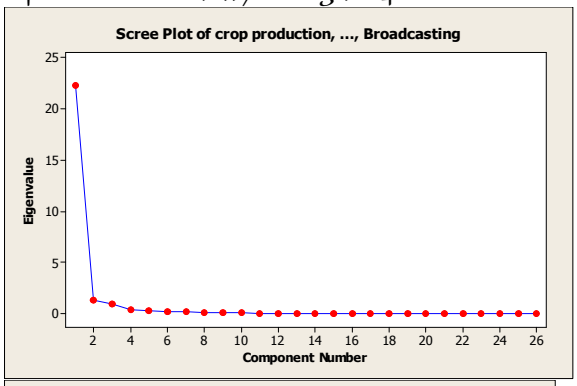
The first principal component has variance (eigenvalue) 23.298 and accounts for 86.3% of the total variance. Also, the coefficients listed under PC1 showed negative values for all the calculated PCs (summarized in Table 4.1);

Table 4.1: The First-Three Principal Component

Variable	PC_1	PC_2	PC_3
Crop production	-0.188	-0.007	-0.343
Livestock	-0.204	**0.081**	**0.021**
Forestry	-0.197	**0.043**	**0.214**
Fishing	-0.201	**0.085**	-0.052
CP&NG	-0.151	**0.48**	-0.269
Quarrying &O M	-0.167	-0.391	-0.004
Oil Refining	-0.202	**0.08**	-0.005
Other manuf.	-0.095	-0.462	-0.654
B&C	-0.198	-0.094	**0.152**
W&RT	-0.200	-0.093	**0.022**
Road trans.	-0.203	**0.039**	**0.007**
Air trans	-0.177	-0.292	**0.216**
O T S	-0.198	**0.136**	-0.127
Telecom.	-0.19	-0.249	**0.132**
Post	-0.205	-0.059	-0.028
Electricity	-0.184	-0.055	-0.333
Water	-0.205	-0.006	**0.058**
Hotels& rest.	-0.200	-0.119	**0.198**
Financial inst.	-0.183	**0.351**	-0.152
Insurance	-0.204	-0.008	**0.09**
Real estate	-0.206	-0.029	**0.041**
business S	-0.185	**0.135**	**0.024**
Pub Admin.	-0.202	**0.174**	**0.015**
Education	-0.204	-0.002	**0.123**
Health	-0.202	0	**0.145**
Other Services	-0.204	**0.053**	-0.001
Broadcasting	-0.205	**0.005**	**0.066**

It should be noted that the interpretation of the principal components is subjective; however, obvious patterns emerge quite often. The first principal component represents an overall population size because the eigenvalue obtained is showed 86.3% variation and the coefficients of these terms have the same sign (negative) and are not close to zero. The second principal component has variance of 1.300 and accounts for 4.8% of the data variability. This component could be thought of as contrasting, because the coefficients of some terms have the different sign and Health Services is zero. *Note:* 14 variables showed negative while 12 variables showed positive effect (see Figure 4.2). The third principal component has variance of 0.985 and accounts for 3.6% of the data variability. This component is also contrasting, because the coefficients of 11 terms have the same sign (negative) while 16 variables showed positive effect.

Together, the first two and the first three principal components represent 91.1% and 94.7%, respectfully, of the total variability (see Figure 4.1). Thus, most of the data structure can be captured in two or three underlying dimensions. The remaining principal components account for a very small proportion of the variability and are probably unimportant. The Scree plot provides this information visually in Figure 4.1.



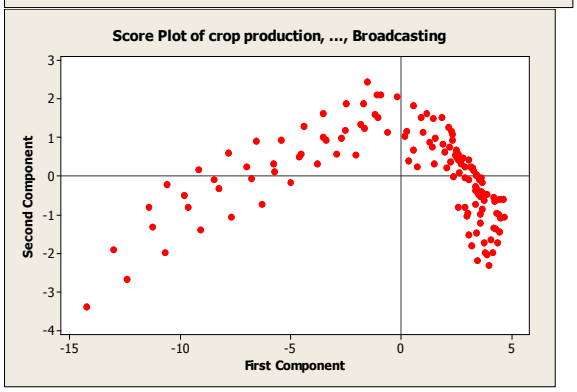
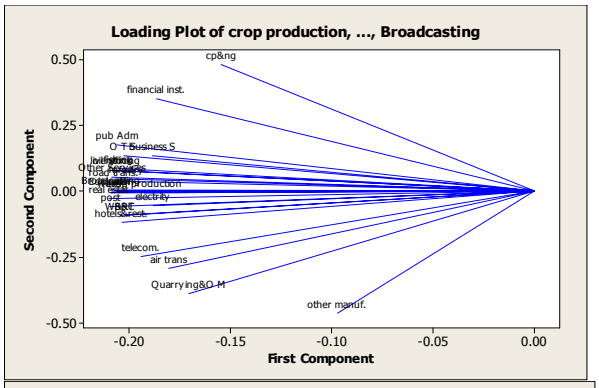


Figure 4.1: Scree plot of the entire principal components**Figure 4.2:** Score plot for the first-two principal components



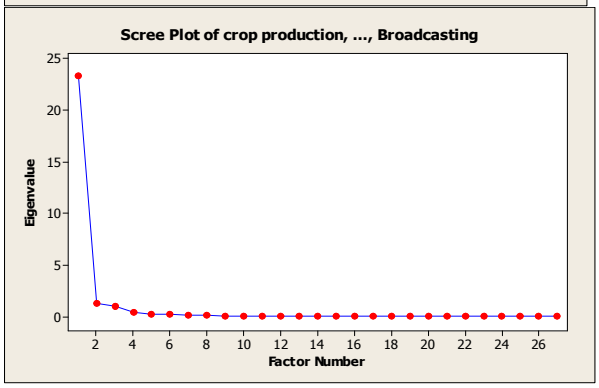


Figure 4.3: Loading plot for the first-two principal components**Figure 4.4:** Scree Plot of crop production, ..., Broadcasting

Next, like principal components analysis, to summarize the data covariance structure in a few dimensions of the data, we used factor analysis. However, the emphasis in factor analysis is the identification of underlying "factors" that might explain the dimensions associated with large data variability. The two factor analysis methods considered as discussed in chapter three are Principal Components and Maximum likelihood method.

Factor Analysis

Factor Analysis Using Principal Components Extraction Method

To investigate what "factors" might explain most of the variability. We used the principal components extraction method and examine an eigenvalues (scree) plot in order to help decide the number of factors to considered; summarized in Table 4.2.

Table 4.2: Factor Analysis using Principal Components Extraction method (Unrotated Factor Loadings and Communalities)

Variable	Factor1	Factor2	Factor3	Factor4	Factor5	Communality
Variance	23.298	1.300	0.985	0.401	0.236	26.221
% Var	0.863	0.048	0.036	0.015	0.009	0.971

In Table 4.2, five factors describe these data perfectly, but the goal is to reduce the number of factors needed to explain the variability in the data. Examine the results line of percentage Variance in Figure 4.4 and Table 4.2 (or the eigenvalues plot). The proportion of variability explained by the last two factors is minimal (0.015 and 0.009, respectively) and they can be eliminated as being unimportant. The first two factors together represent 91.1% of the variability while three factors explain 94.7% of the variability. The question is whether to use two or three factors. The next step might be to perform separate factor analyses with two and three factors and examine the communalities to see how individual variables are represented. If there were two or more variables not well represented by the more parsimonious two factor model, then select a model with three or more factors.

Table 4.3: First Two and Three Factors Communalities

Variable	three factors Communality	first two factors Communality
crop production	0.942	0.826
live stock	0.979	0.979
Forestry	0.952	0.907
Fishing	0.953	0.950
cp&ng	0.901	0.829
Quarrying&O M	0.850	0.850
oil refining	0.957	0.957
other manuf.	0.908	0.486
B&C	0.946	0.923
W&RT	0.945	0.944
road trans.	0.967	0.967
air trans	0.885	0.839
O T S	0.957	0.942
telecom.	0.942	0.925
Post	0.984	0.984
Electricity	0.900	0.791
Water	0.983	0.979
Hotels & rest.	0.987	0.948
financial inst.	0.962	0.939
Insurance	0.973	0.965
real estate	0.988	0.987
business S	0.819	0.819
pub Adm	0.987	0.987
Education	0.987	0.972
Health	0.972	0.951
Other Services	0.976	0.976
Broadcasting	0.981	0.977

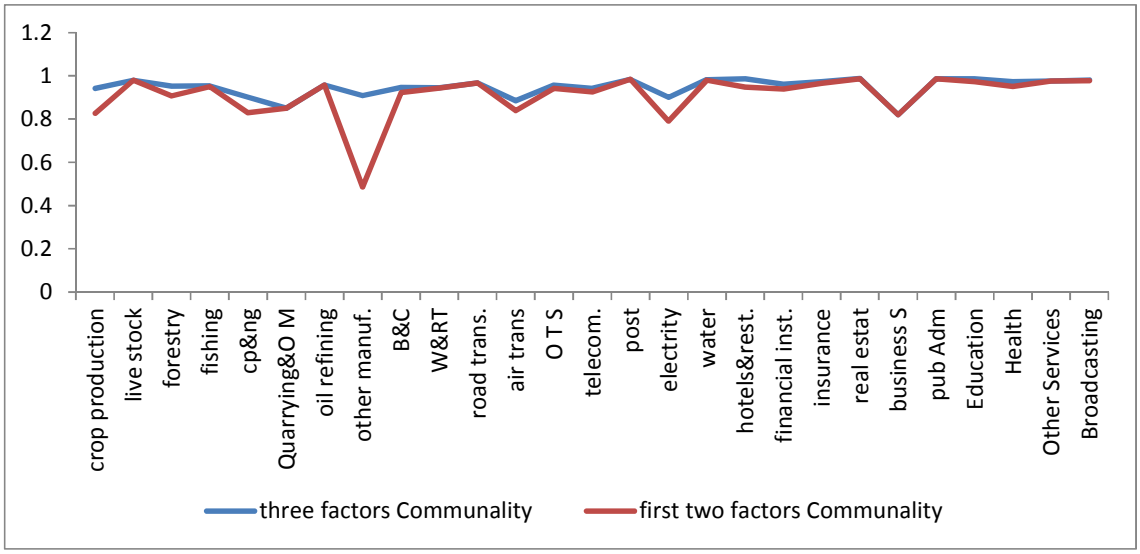


Figure 4.5: Comparison of the first two and three factors Communalities correlation

In Figure 4.5, only "other manufacturing" seem not to be well represented by the more parsimonious first two factor model, which seem not significant to reject the two factor model. Furthermore, a rotation (or varimax rotation) of loadings extracted by the maximum likelihood extraction method with a selection of two factors will be performed to interpret the factors.

Factor Analysis using Maximum likelihood Method (MLM)

The section is divided into two parts: (a) Sorted and unrotated factor Loadings and Communalities of the factor analysis between the variables using Maximum Likelihood; (b) Sorted and a Rotation of the factor analysis between the variables using Maximum Likelihood. The results obtained are summarized in three tables of loadings and communalities: unrotated, rotated, and sorted and rotated. The results of the three tables of loadings and communalities factors explain 89.9% of the data variability and the communality values indicate that all variables are well represented by these two factors analysis.

The computation of (a) and (b) are summarized in Table 4.4 and the variables level of reliability (Pearson correlation) classification are in Table 4.5 and 4.6 below using factor loadings and communalities;

International Journal of Science and Advanced Innovative Research
ISSN: 2536-7315 (Print) 2536-7323 (Online)
Volume 3, Number 3, September 2018
http://www.casirmediapublishing.com

Table 4.4: Sorted Unrotated and Rotated Factor Loadings and Communalities

	Sorted Unrotated Factor Loadings and Communalities				Sorted Rotated Factor Loadings and Communalities		
Variable	Factor 1	Factor 2	Communality	Variable	Factor 1	Factor 2	Communality
real estate	0.995	0.043	0.992	Telecom.	0.906	-0.371	0.958
Water	0.993	0.028	0.987	hotels& rest.	0.859	-0.491	0.98
Education	0.992	0.056	0.987	air trans	0.854	-0.318	0.83
Broadcasting	0.988	-0.006	0.976	Quarrying&O M	0.82	-0.272	0.746
live stock	0.987	-0.084	0.982	B&C	0.812	-0.528	0.938
Insurance	0.986	0.029	0.973	W&RT	0.796	-0.554	0.94
Other Services	0.985	-0.079	0.976	Health	0.782	-0.6	0.972
Post	0.984	0.025	0.968	Education	0.779	-0.617	0.987
Health	0.983	0.07	0.972	real estate	0.773	-0.628	0.992
road trans.	0.98	-0.053	0.962	Water	0.762	-0.638	0.987
pub Admin.	0.974	-0.18	0.981	Insurance	0.757	-0.632	0.973
oil refining	0.971	-0.114	0.956	Post	0.752	-0.634	0.968
hotels& rest.	0.969	0.202	0.98	Forestry	0.744	-0.608	0.923
Fishing	0.965	-0.144	0.952	Broadcasting	0.735	-0.66	0.976
W&RT	0.963	0.113	0.94	road trans.	0.698	-0.69	0.962
Forestry	0.96	0.039	0.923	financial inst.	0.384	-0.897	0.953
B&C	0.958	0.144	0.938	cp&ng	0.161	-0.894	0.824
O T S	0.948	-0.224	0.949	O T S	0.561	-0.797	0.949
telecom.	0.924	0.323	0.958	pub Admin.	0.609	-0.781	0.981
business S	0.893	-0.115	0.81	Fishing	0.626	-0.748	0.952
crop production	0.891	-0.138	0.813	oil refining	0.651	-0.73	0.956
financial inst.	0.883	-0.417	0.953	live stock	0.683	-0.718	0.982

Electrity	0.864	-0.117	0.76	Other Services	0.684	-0.713	0.976
air trans	0.85	0.328	0.83	crop production	0.575	-0.694	0.813
Quarrying&O M	0.794	0.341	0.746	business S	0.591	-0.678	0.81
cp&ng	0.713	-0.562	0.824	Electricity	0.569	-0.661	0.76
other manuf.	0.42	0.129	0.193	other manuf.	0.4	-0.183	0.193
Variance	23.181	1.101	24.282	Variance	13.142	11.14	24.282
% Var	0.859	0.041	0.899	% Var	0.487	0.413	0.899

Table 4.5: Comparison of Sorted Unrotated and Rotated Communalities using Level of Reliability

Level of Reliability	Correlation	MLMSorted Unrotated Communality	MLM Sorted Rotated Communality	Remarks
Almost perfect	0.81-1.00	real estate	Telecom.	
		Water	Hotels & rest.	
		Education	air trans	
		Broadcasting	B&C	
		live stock	W&RT	
		Insurance	Health	
		Other Services	Education	
		Post	real estate	
		Health	Water	
		road trans.	Insurance	
		pub Admin.	Post	
		oil refining	Forestry	
		Hotels & rest.	Broadcasting	
		Fishing	road trans.	
		W&RT	financial inst.	
		Forestry	Cp&ng	
		B&C	O T S	
		O T S	pub Admin.	
		Telecom.	Fishing	
		business S	oil refining	
		crop production	live stock	
		financial inst.	Other Services	
		air trans	crop production	
		Cp&ng	business S	
Substantial	0.61-0.80	Quarrying &O M	Quarrying &O M	
		Electricity	Electricity	
Moderate	0.41-0.60	Nil	Nil	
Fair	0.21-0.40	Nil	Nil	
Slight	0.00-0.20	other manuf.	other manuf.	

Table 4.6: Comparison of Sorted Unrotated and Rotated Factors

Level of Reliability	Correlation	MLM Sorted Unrotated Factor 1	MLM Sorted Rotated Factor 1	Remarks
Almost perfect	0.81-1.00	real estate Water Education Broadcasting live stock Insurance Other Services Post Health road trans. pub Admin. oil refining Hotels & rest. Fishing W&RT Forestry B&C O T S Telecom. business S crop production financial inst. Electricity Air trans	Telecom. Hotels &rest. air trans Quarrying &O M B&C	
Substantial	0.61-0.80	Quarrying &O M Cp&ng	W&RT Health Education real estate Water Insurance Post Forestry Broadcasting road trans. pub Admin. Fishing oil refining live stock Other Services	

Moderate	0.41-0.60	Nil	O T S crop production business S Electricity
Fair	0.21-0.40	Nil	financial inst. other manuf.
Slight	0.00-0.20	other manuf.	cp&ng

In Table 4.4, the sorted unrotated factor loadings and communalities have a large positive loading on real estate of 0.995 in factor 1 and the loadings variance between factor 1 and 2 are 0.859 and 0.041, respectively. While, sorted rotated factor loadings and communalities have a large positive loading on telecommunication of 0.906 in factor 1 and the loadings variance between factor 1 and 2 are 0.487 and 0.413, respectively.

The percent of total variability represented by the factors does not change without rotation, but after rotating, these factors are more evenly balanced in the percent of variability that they represent factor 1 and 2 are 48.7% and 41.3%, respectfully. Table 4.5 showed 24 variables with almost perfect level of reliability, while three variables are not (Hint: Substantial: Quarrying &OM and Electricity, then Slight: other manufacturing) both Sorted Unrotated and Rotated. Table 4.6 showed 24 variables with almost perfect level of reliability for Sorted Unrotated, while five variables for Sorted Rotated.

Comparison of the two factors analysis (Sorted Unrotated and Rotated Factor Loadings and Communalities)

Table 4.4 was rearranging alphabetically in Table 4.7 for comparison of the sorted unrotated and rotated factor loadings and communalities.

Table 4.7: Alphabetical Rearrangement of the Nigerian Economic Variables

	Sorted Unrotated Factor Loadings and Communalities				Sorted Rotated Factor Loadings and Communalities		
Variable	Factor1	Factor2	Communality	Variable	Factor1	Factor2	Communality
air trans	0.85	0.328	0.83	air trans	0.854	-0.318	0.83
B&C	0.958	0.144	0.938	B&C	0.812	-0.528	0.938
Broadcasting	0.988	-0.006	0.976	Broadcasting	0.735	-0.66	0.976
business S	0.893	-0.115	0.81	business S	0.591	-0.678	0.81
cp&ng	0.713	-0.562	0.824	cp&ng	0.161	-0.894	0.824
crop production	0.891	-0.138	0.813	crop production	0.575	-0.694	0.813
Education	0.992	0.056	0.987	Education	0.779	-0.617	0.987
Electricity	0.864	-0.117	0.76	Electricity	0.569	-0.661	0.76
financial inst.	0.883	-0.417	0.953	financial inst.	0.384	-0.897	0.953
Fishing	0.965	-0.144	0.952	Fishing	0.626	-0.748	0.952
Forestry	0.96	0.039	0.923	Forestry	0.744	-0.608	0.923
Health	0.983	0.07	0.972	Health	0.782	-0.6	0.972
hotels&rest.	0.969	0.202	0.98	hotels&rest.	0.859	-0.491	0.98
Insurance	0.986	0.029	0.973	Insurance	0.757	-0.632	0.973
live stock	0.987	-0.084	0.982	live stock	0.683	-0.718	0.982
O T S	0.948	-0.224	0.949	O T S	0.561	-0.797	0.949

oil refining	0.971	-0.114	0.956	oil refining	0.651	-0.73	0.956
other manuf.	0.42	0.129	0.193	other manuf.	0.4	-0.183	0.193
Other Services	0.985	-0.079	0.976	Other Services	0.684	-0.713	0.976
Post	0.984	0.025	0.968	Post	0.752	-0.634	0.968
pub Adm	0.974	-0.18	0.981	pub Adm	0.609	-0.781	0.981
Quarrying&OM	0.794	0.341	0.746	Quarrying&OM	0.82	-0.272	0.746
real estate	0.995	0.043	0.992	real estate	0.773	-0.628	0.992
road trans.	0.98	-0.053	0.962	road trans.	0.698	-0.69	0.962
telecom.	0.924	0.323	0.958	telecom.	0.906	-0.371	0.958
W&RT	0.963	0.113	0.94	W&RT	0.796	-0.554	0.94
Water	0.993	0.028	0.987	Water	0.762	-0.638	0.987
Variance	23.181	1.101	24.282	Variance	13.142	11.14	24.282
% Var	0.859	0.041	0.899	% Var	0.487	0.413	0.899

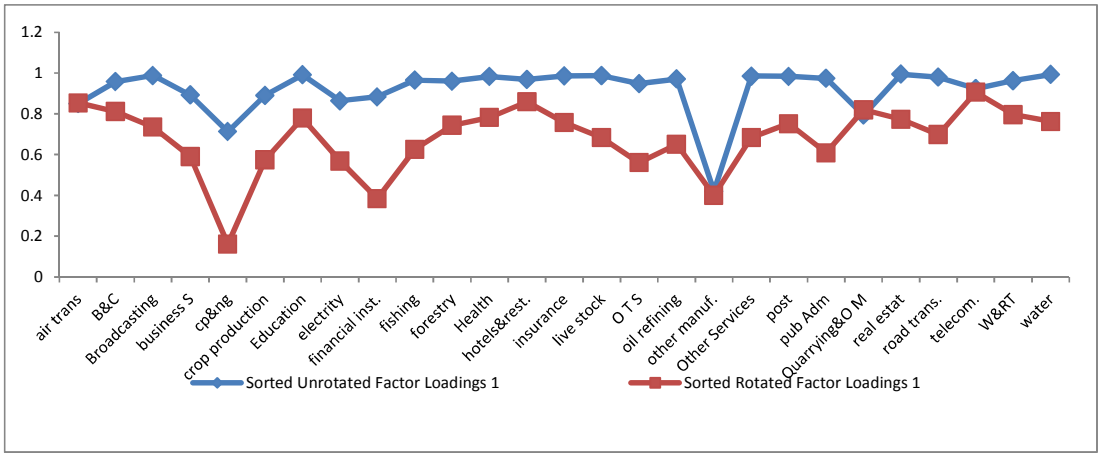


Figure 4.6: Comparison of factors 1 sorted unrotatedand rotated

Figure 4.6 showed that the sorted unrotated and rotated factor loadings 1 are exhibiting the same pattern. However, sorted rotated factor1 loading have the minimum variance of 13.142(48.7%) variation of the variables while sorted unrotated factor 1 showed 23.181(85.9%) variation of the variables. The result of the sorted rotated factor 1 loading seems to be better (appropriate) than the sorted unrotated factor 1 loading.

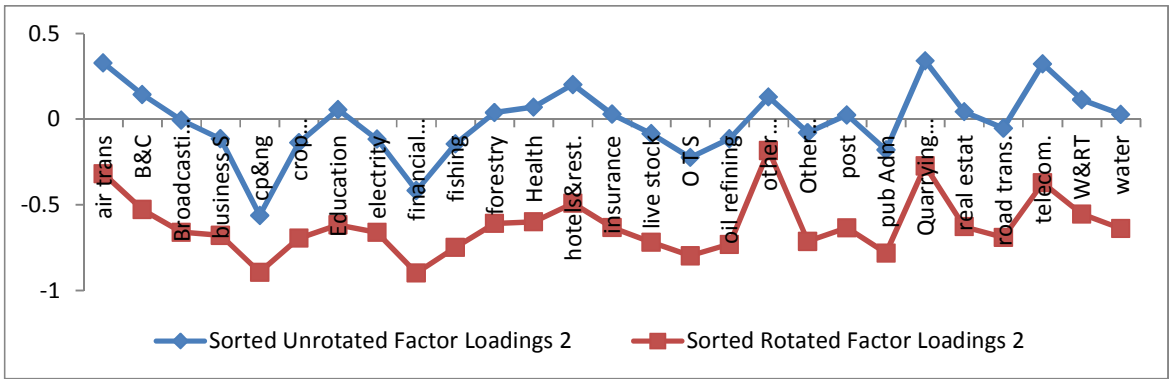


Figure 4.7: Comparison of factors 2 sorted unrotated and rotated

Figure 4.7 showed that the sorted unrotated and rotated factor loadings 2 are exhibiting the same pattern.

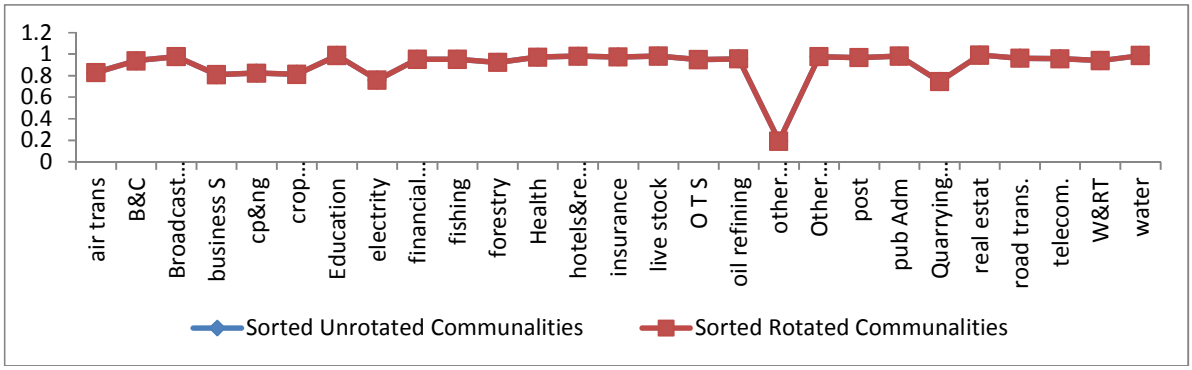


Figure 4.8: Comparison of sorted unrotated and rotated communalities

Figure 4.8 showed that the sorted unrotated and rotated communalities exhibited the same pattern.

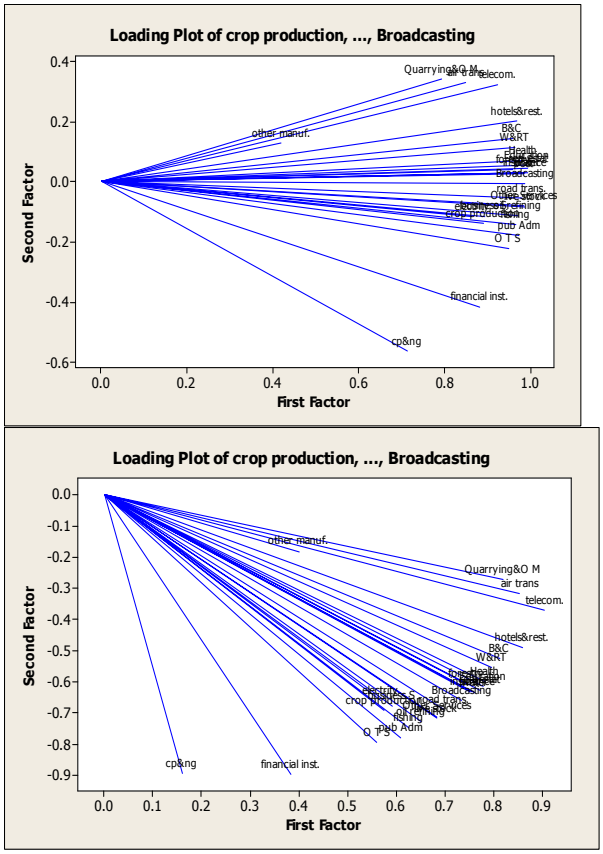


Figure 4.9: Loading Plot of the two factors (sorted unrotated)**Figure 4.10:** Loading Plot of the two factors (sorted rotated)

Figure 4.9 the sorted unrotatedloading plot showed both positive and negative effects by the Nigeria economic variables between the two factors considered,

while the sorted rotated loading plot in Figure 4.10 showed only negative effects by the Nigeria economic variables between the two factors considered.

CONCLUSION

The result of the correlation coefficients obtained showed the four variables which are uncorrelated (or less than absolute of ±0.5) and were removed, then remaining twenty-seven variables were used for further analysis. The first two and the first three principal components represent 91.1% and 94.7%, respectfully, of the total variability by Nigeria economic variables. Thus, most of the data structure can be captured in two or three underlying dimensions. The remaining principal components account for a very small proportion of the variability and are probably unimportant. Then, comparing first two and the first three principal components plots, only other manufacturing seem not to be well represented by the more parsimonious two factor model, which seem not significant to reject the two factor model. Furthermore, a rotation (or varimax rotation) of loadings extracted by the maximum likelihood extraction method with a selection of two factors was performed to interpret the factors. The results showed real estate as the most important variable when sorted and unrotated, while telecommunication as the most important variable when sorted and rotated. This research identified significant difference between the loading factors and no difference between the communalities, when Nigeria economic variables are sorted unrotated and rotated.

CONTRIBUTION

This research was able

1) to identify significant differences between the factor loadings andno difference between the communalities of the Nigeria Economic variables data considered.
2) to identify the variation pattern of Nigeria economic variables usingsorted unrotated and rotated factor loadings methods.
3) to identify that economic variables were better under the sorted rotated factor loading thereby identify the sorted rotated as the better (appropriate) method, when using economic data variables.
4) to identify24 Nigeria economic variables with almost perfect level of reliability.

REFERENCES

Abdi, H.,and Williams, L. J. (2010). "Principal Components analysis". WileyInterdisciplinary Reviews: *Computational Statistics.* 2 (4), 433–459.

Abhishek, B. (2012). Impact of Principal Components Analysis in the Application of image Processing. *International Journal of Advanced*

Research in Computer Science and Software Engineering Vol 2, Issue 1.

Arash a., (2005). Color Image Processing Using Principal Component Analysis. Master's Thesis, Sharif University of Technology, Mathematics science Department, Tehran, Iran.

Arash, A and Shohreh, K. (2008).Colour PCA Eigen Images and their Application to Compression and watermarking, Volume 26, issue 7 pages 0260-8858.

Bocuvxa L., Vacek O. and Jehlicka J. (2005).Principal Component Analysis as a Tool to Indicative the Origin of Potentially Toxic Elements in Soils.Geoderma; 128, pp. 289-300.

Brook, J. S., Whiteman, M. and Nomura, C. *(1988).*Personality, Family, and Ecological Influences on Adolescent Drug Use: A Developmental Analysis.*Journal of Chemical Dependency Treatment.*1,123-161.

Chang, Y., Cesarman, E., Pessin, M. S., Lee, F., Culpepper, J., Knowles, D. M.and Moore, P. S. (2013). *Identification of herpes virus-like DNA sequences in AIDS-associated Kaposi's sarcoma. Science;* 266,1865–1869.

Cheng, S.-C., and Hsia, S.-C., *(2003).*Fast Algorithm's for Color Image Processing by Principal Component Analysis. *Journal of Visual Communication and Image Representation.*14, 184-203.

Dragovic, S. and Onjia, A. (2006).Classification of Soil Samples According to their Geographic Origin Using Gamma-ray Spectrometry and Principal Component Analysis.*Journal of Environmental Radioactivity.*84,pp. 150-158.

Ekpo, A. H. and Umoh, O.J. (2012).Overview of the Nigerian Economic Growth and Development.EghosaOsagie (1992). Edited, Structural Adjustment Programme in the Nigeria Economy*, National Institute for Policy and Strategic Studies, Kuru, Jos Nigeria.*71-104.

Eric, B., Trevor, H., Debashis, P. and Robert, T. (2006)."Prediction by Supervised Principal Components".*Journal of the American Statistical Association.*101 (473), 119–137.

Gerns, S. H. L., Richardson, B. A., Singa, B., Naulikha, J., Prindle, V. C., Diaz-Ochoa, V. E., Felgner, P. L., Camerini, D., Horton, H., John-Stewart G. and Judd, L. W. (2014). Use of Principal Components Analysis and Protein Microarray to Explore the Association of HIV-1-Specific IgG Responses with Disease Progression, AIDS Research and Human Retroviruses: Vol. 30, Number 1, 39-44.

Golub, G. H and Van, Loan C. F. (1988).Matrix Computations. Baltimore, Maryland: Johns Hopkins University Press

Harman, N. N. (1960). Applied Factor Analysis. *Journal of Education Psychology* 312-320

Hotelling, H. (1933). Analysis of a complex of statistical variables into principal Components. *Journal of Educational Psychology*, 24, 417–441, and 498–520.

Jaisheel, M., Fulufhelo, V. N. and Tshilidzi, M. (2009).Missing Data Estimation using Principle Components Analysis and Auto associative Neural Networks. Systemic, Cybernetics and Informatics; 7(3),1690-4534.

Jolliffe, I. T. (2002).Principal Components Analysis, (2nd edition) Springer-Verlag. ISBN 978-0-387-95442-4.

Kamolchanok, P., Kanokporn, S., Natdhera, S. and Daoroong, S. (2012).Principal Components Analysis for the characterization in the application of some soil properties; International journal of Environment and Ecological Engineering.Vol. 6, No. 5.

Khaled, L., Rao, V. andVemuri, I.(2004). An Application of Principal Components Analysis to the DETECTION and Visualization of Computer Network Attack; Annals of Telecommunication 61(1), 218-234.

Landis, J. R. and Koch, G. G. (1977). The Measurement of Observer Agreement for categorical data. Biometrics. 33, 159-174

Libin, Y. (2015). An application of Principal Components Analysis to Stock Portfolio Management.Department of Economics and Financial University of Canterbury.

Lindsay, I. S. (2002). A Tutorial on Principal Components Analysis, Journal of Business and Management Science Vol. 2, No. 1, 10-20.

Ildiko E. F. and Jerome H. F. (1993)."A Statistical View of Some Chemometrics Regression Tools".Technometrics.35 (2), 109–135.

Maike, R. (2016). Factor Analysis: A Short Introduction, http://www.theanalysisfactor.com/factor-analysis-1-introduction/.Retrieved at: 10/23/2016.

Marshal N., Faust M. andHendler T., (2005) The Role of the Right Hemisphere in Processing Nonsalient Metaphorical Meaning: Application of Principal Components Analysis to FMRI data. Neuropsychologia. 43(14), 2084-100.

Noko, E. J. (2016). Economic recession in Nigeria: Causes and Solution, Published by educLn.com 2016 http://educacinfo.com/economic-recession- Nigeria .

Ogoke U. P., Nduka, E. C., Biu, E. O. and Ibeachu, C. (2003).A Comparative Study of Foot Measurements Using Receiver Operating Characteristics (ROC) Approach.*Scientia Africana Journal of Pure and Applied Sciences.* Volume 12(1), pp 76-88.

Pearson, K. (1901). "On Lines and Planes of Closest Fit to Systems of Points in Space" (PDF). Philosophical Magazine.2 (11), 559–572.

Rafael, Do, E. S. (2012). Principal Components Analysis Applied to digital image compression; 10(2): 135-139.

Santanu P. and Jaya, S (2008)."Rice Disease Identification using Pattern Recognition Techniques" 11th International Conference, - ICCTT, pp. 420-423.

The Pennsylvania State University (2007). STA 505, Applied Multivariate Statistical Analysis.Lecture Notes. Department of Statistics, Philadephia.https:onlinecourses.science.psu.edu/sta505/node/79

Vukosi, N. M. (2015). Autoennodes, Principal Components Analysis and Support Vector Regression for Data Imputation.Journal of Research Gate; School of Electrical and Information Engineering, University of the Witwatersrand.Johannesbury, South Africa.

Ye N., Emran S., Chen Q. and Vilbert S., (2002). Multivariate Statistical Analysis of Adult Trails for Host Based Institution Detection. IEEE Transaction on Computers.Vol. 51, No. 7.

-