American Journal of Economics 2015, 5(5): 508-512
DOI: 10.5923/j.economics.20150505.11

Principal Component Analysis of Nigeria Value of Major Imports

Usoro Anthony E.[1,*], Moffat I. U.[2]

[1]Department of Mathematics and Statistics, Akwa Ibom State University, Mkpat Enin, Nigeria
[2]Department of Mathematics and Statistics, University of Uyo, Uyo, Nigeria

Abstract A principal component analysis of value of major imports was carried out in this research work. From the analysis, the first principal component accounted for 98.89% of the total variation amongst the observed variables. The first principal component P_1 was then used as a predictor variable for subsequent analysis. The regression of the total value of major imports(Y) on the principal component (P_1) yielded $Y = 6479 + 1.04P_1$. Values obtained from the estimated model have shown reliability of the principal component approach. This paper recommends principal component analysis in a relationship between a response and multiple predictor variables to overcome the problem of multicollinearity in multiple regression analysis.

Keywords Principal component, Eigen values, Eigen vectors

1. Introduction

Principal component analysis (**PCA**) is appropriate when you have obtained measures on a number of observed variables and wish to develop a smaller number of artificial variables (called principal components) that will account for most of the variance in the observed variables. The principal components may then be used as predictor or criterion variables in subsequent analyses. Principal component analysis is a variable reduction procedure. It is useful when someone believes that there is redundancy in some of the variables. In this case, redundancy means, some of the variables are highly correlated with one another, possibly because they are measuring the same construct. Because of this redundancy, it is possible to reduce the observed variables into a smaller number of principal components (artificial variables) that will account for most of the variance in the observed variables, Kim and Mueller [2, 3, 7, 9].

Principal component analysis (PCA) is a covariance or correlation analysis between different factors. Covariance is always measured between two factors. So with three factors, covariance is measured between factors x and y, y and z, and x and z. If the there are more than three factors, the covariance values can be placed into a matrix, [4]. Principal component analysis is a mathematical procedure that uses an orthogonal transformation to convert a set of observations of possibly correlated variables into a set of values of uncorrelated variables called **principal components**, [5]. In principal component analysis, the number of principal components $P_1, P_2, P_3, \ldots, P_k$ extracted is always less than the number of original variables $X_1, X_2, X_3, \ldots, X_m$. The transformation is defined in such a way that the first principal component P_1 account for the highest of the total variability in a multiple relationship amongst variables, and that every principal component must account for higher variability than the succeeding components, [6]. The two major terms used in analysis of principal components by Pearson (1901) are **Eigenvectors** and **Eigenvalues**. Eigenvectors can be thought of as preferential direction of a data set, or in other words, main patterns in the data. Eigenvalues can be thought of as quantitative assessment of how much a component represents the data. The higher the eigenvalues of a component, the more representative it is of the data. Eigenvalues can also be representative of the level of explained variance as a percentage of total variance in the PCA.

In principal component analysis, the number of principal components is determined by the percentage of variation accounted for by the preceeding component. For instance, if the first principal component accounts for over three quarters of the total variation, subsequent principal components may not be feasible. The results of a PCA are usually discussed in terms of component scores (the transformed variable values corresponding to a particular data point) and loadings (the weight by which each standardized original variable should be multiplied to get the component score), [8]. Principal component analysis has been one of the most valuable results from applied linear algebra. It is used abundantly in all forms

* Corresponding author:
toskila2@yahoo.com (Usoro Anthony E.)
Published online at http://journal.sapub.org/economics

of analysis from neuroscience to computer graphics, because it is a simple non-parametric method of extracting relevant information from confusing data sets. It provides a road map to the reduction of a complex data set to a lower dimension in order to reveal the sometimes hidden simplified structure that often underlie it, [1]. [10] described principal component analysis as an analysis whose purpose is to introduce new variates (called principal components), which are linear combination of the original variables X's in a multiple relationship amongst the variables of interest, such that the principal components obtained are orthogonal. One of the reasons of adopting principal component method is to solve the problem of multicollinearity in a multiple regression model. [10] carried principal component analysis, using correlation matrix and extracted two principal components from the original five explanatory variables in a multiple relationship amongst the variables of interest. [4] used covariance matrix for the analysis of the principal component.

In this paper, we intend to analysis Nigeria's value of major imports through principal component method.

2. Statistical Method

This Section contains the procedure involve in the calculation of the eigenvalues in the principal component analysis.

2.1. Source of Data

The source of data for this research work is from page 198 of Central Bank of Nigeria Statistical Bulletin Volume 21, December, 2010. Particularly, Value of major imports is our study interest. The total value of major imports is aggregation of the following components: Food & Live Animal Beverages & Tobacco, Crude Materials Inedible, Mineral Fuels, Animal and Vegetable, Oil & Fat, Chemicals, manufactured goods, Machinery & Transport Equipment, Miscellaneous Manufactured Goods and Miscellaneous Transactions. In this paper, we have conveniently removed the last two components, since their details are not reflected in the bulletin.

2.2. Description of Variables

The variables used for the analysis are defined as follows:

Total Value of Major Imports (Y)

Food & Live Animal (X_1)

Beverages & Tobacco (X_2)

Crude Materials Inedible (X_3)

Mineral Fuels (X_4)

Animal and Vegetable (X_5)

Oil & Fat (X_6)

Chemicals (X_7)

Manufactured Goods, Machinery & Transport Equipment (X_8)

2.3. Correlation Matrix

The correlation matrix for each pair of variables is given below:

	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8
X_1	r_{11}							
X_2	r_{21}	r_{22}						
X_3	r_{31}	r_{32}	r_{33}					
X_4	r_{41}	r_{42}	r_{43}	r_{44}				
X_5	r_{51}	r_{52}	r_{53}	r_{54}	r_{55}			
X_6	r_{61}	r_{62}	r_{63}	r_{64}	r_{65}	r_{66}		
X_7	r_{71}	r_{72}	r_{73}	r_{74}	r_{75}	r_{76}	r_{77}	
X_8	r_{81}	r_{82}	r_{83}	r_{84}	r_{85}	r_{86}	r_{87}	$r_{88.}$

2.4. Principal Loadings for the Principal Component

[10] expressed loadings for the first, second, third ,…, kth principal components as:

$L_{1j}, L_{2j}, L_{3j}, \ldots, L_{kj}$ (j = 1, 2, 3, …, 8).

$L_{1j} = r_{.j}/\sqrt{r..}$, j = 1, 2, 3, …, 8.

$r_{.j}$ are the column sums from the correlation coefficient for a particular principal component. r.. is the sum of all the correlation coefficients for a particular principal component.

2.5. Eigenvectors for the Principal Component

The percentage of variation accounted for by "k" number of principal components is expressed by Eigenvectors, whose values are obtained from the principal loadings. The eigenvectors are λ_i (i = 1, 2, …, k). $\lambda_1, \lambda_2, \lambda_3, \ldots, \lambda_k$ are the eigen vectors for the 'k' number of principal components. Therefore, the percentage of variation accounted for by a principal component, say P_1 is $P_1\% = \lambda_1/k$ percent. K is the number of X's in original data. This expresses the proportion of the total variation accounted for by each principal component, and also gives an idea as to whether subsequent principal components are significant or negligible in the analysis.

3. Estimation and Analysis

Here, we apply the aforementioned statistical procedure for the extraction of the principal loadings and components.

3.1. Matrix of Correlation Coefficients

	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8
X_1	1.00							
X_2	0.967	1.00						
X_3	0.973	0.998	1.00					
X_4	0.956	0.986	0.991	1.00				
X_5	0.956	0.996	0.997	0.992	1.00			

X_6	0.975	0.996	0.998	0.989	0.996
	1.00				
X_7	0.976	0.977	0.999	0.987	0.996
	0.999	1.00			
X_8	0.985	0.994	0.997	0.982	0.990
	0.997	0.998	1.00		

The column sums of the correlation coefficients are $r._1 = 7.788$, $r._2 = 7.914$, $r._3 = 7.953$, $r._4 = 7.883$, $r._5 = 7.923$, $r._6 = 7.95$, $r._7 = 7.932$, $r._8, = 7.943$. The sum of the column totals of the correlation coefficients is $r.. = 63.286$. $\sqrt{r..} = \sqrt{63.286} = 7.955$.

3.2. Loadings for the First Principal Component and Its Percentage of Variation

The loadings for the first principal component are estimated as follows:

$l_{1j} = r._j/\sqrt{r..}$

$l_{11} = 0.9790$, $l_{12} = 0.9948$, $l_{13} = 0.9917$, $l_{14} = 0.9909$, $l_{15} = 0.9960$, $l_{16} = 0.9994$,

$l_{17} = 0.9971$, $l_{18} = 0.9985$. The principal component is

$P_1 = L_{11}X_1 + L_{12}X_2 + L_{13}X_3 + L_{14}X_4 + L_{15}X_5 + L_{16}X_6\, L_{17}X_7 + L_{18}X_8$. From the estimated loadings, the principal component model is obtained as,

$$P_1 = 0.9790X_1 + 0.9948X_2 + 0.9917X_3 + 0.9909X_4 + 0.9960X_5 + 0.9994X_6 + 0.9971X_7 + 0.9985X_8. \quad (1)$$

The eigenvalue for the first principal component is $\lambda_1 = \sum l^2_{1j} = 0.9584 + 0.9896 + 0.9994 + 0.9819 + 0.9920 + 0.9988 + 0.9942 + 0.9970 = 7.9113$.

Therefore, the percentage of variation accounted for by the first principal component, P_1 is $P_1\% = (7.9113/8)100\% = 98.89\%$.

From the percentage of variation accounted for by the first principal component, the analysis terminates at the first principal component, because the first principal component accounts for almost the total variation amongst the variables. The values of P_1 as obtained from 'model 1' are given in table 1.

3.3. Regression of total Value of Import on the Principal Component

The final analysis in this paper is the regression of the total value of import on the principal component. As mentioned in the introductory part of this paper, P_1 is going to act as a predictor variable in a simple regression analysis. This is an alternative approach to multiple regression analysis of the total value of import on the X's. Therefore, the regression of Y on P_1 yields the following model,

$$Y = 6479 + 1.04P_1. \quad (2)$$

The estimates from 'model 2' are shown in table 1. The graph of original values and estimates of total value imports are plotted in figures 1 and 2.

Table 1. Table of principal component, original and estimates of total value of imports

S/N	Y	P_1	Y EST	S/N	Y	P_1	Y EST
1981	12840	11788	18752	1996	562627	535997	564541
1982	10771	9971	16860	1997	845717	805240	844867
1983	8904	8411	15237	1998	837419	799629	839026
1984	7178	6848	13609	1999	862516	821944	862259
1985	7063	6790	13549	2000	985022	939631	984791
1986	5984	5697	12410	2001	1358180	1293572	1353303
1987	17862	17112	24296	2002	1512695	1402399	1466610
1988	21446	20402	27721	2003	2080235	1952495	2039351
1989	30860	29511	37205	2004	1987045	1856809	1939726
1990	45718	43246	51506	2005	2800856	2693849	2811224
1991	89488	86797	96850	2006	3108519	2989770	3119327
1992	143151	137258	149387	2007	3911953	3762512	3923880
1993	165629	158262	171256	2008	5189803	4992580	5204585
1994	162789	153963	166781	2009	5102534	4883115	5090613
1995	755128	719720	755827	2010	8005374	7696637	8019957

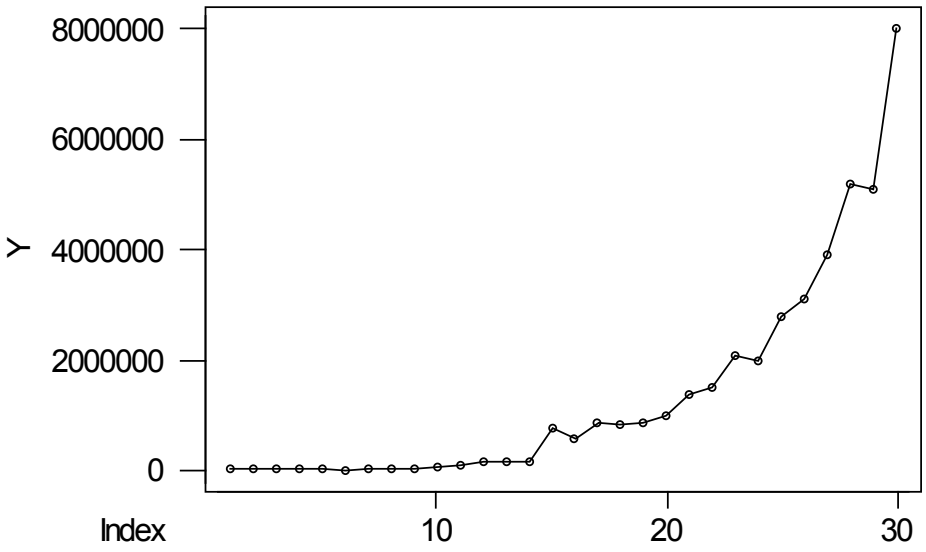

Figure 1. Gtaph of original values

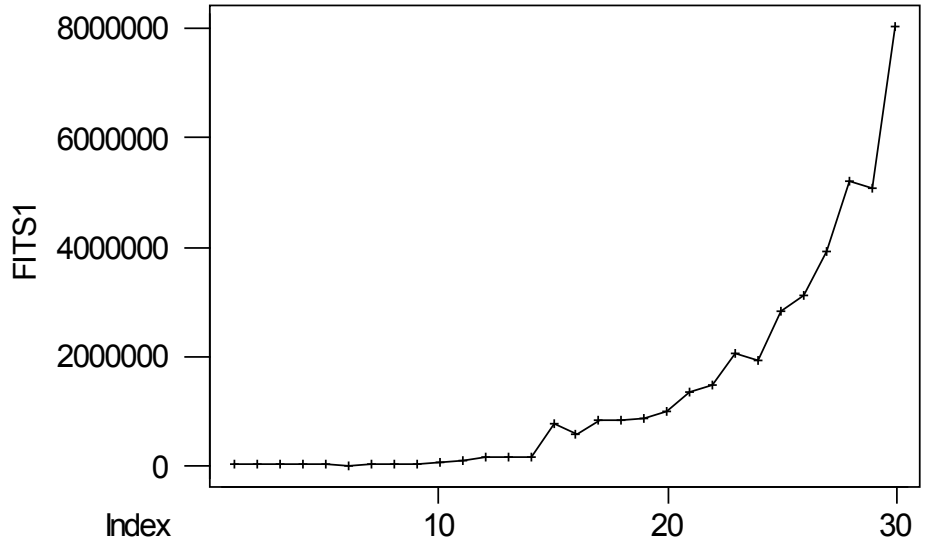

Figure 2. Gtaph of the estimates

4. Conclusions

Principal component analysis is a powerful tool for reducing a number of observed variables into a smaller number of artificial variables that account for most of the variance in a given data set. It is an explanatory tool to uncover unknown trends in a given set of data. The number of principal components extracted from the observed variables is determined by the percentage of total variation accounted for by the first principal component. The analysis carried out in this paper has revealed that the first principal component P_1 has accounted for almost the total variation amongst the observed variables. This further explains the fact that subsequent principal components are insignificant and negligible. Principal component method is useful, especially, when the explanatory variables are strongly correlated. Strong correlation amongst pairs of variables is synonymous with multicollinearity. So with principal component analysis, problem of multicollinearity in a multiple relationship between response and predictor variables is addressed.

REFERENCES

[1] Jonathon Shlens (2005): A tutorial on Principal Component Analysis. Systems Neurobiology Laboratory, Salk Insitute for Biological Studies La Jolla, CA 92037.

[2] Kim, J. O. & Mueller, C. W. (1978a). *Introduction to factor analysis: What it is and how to do it.* Beverly Hills, CA: Sage.

[3] Kim, J. O. & Mueller, C. W. (1978b). *Factor analysis: Statistical methods and practical issues.* Beverly Hills, CA: Sage.

[4] Lindsey Smith (2002): A tutorial on Principal Component Analysis. http:/www.cs.otago.ac.nz/cosc453/studenttutorials/

principal components.pdf.

[5] Pearson, K. (1901). On Lines and Planes of Closest Fit to Systems of Points in Space. *Philosophical Magazine* 2 (6): 559–572.

[6] Rao, C.R. (1964): The use and interpretation of Principal Component Analysis in Applied Research. *Sankhya* A 26, 329-358.

[7] Rummel, R. J. (1970). *Applied factor analysis.* Evanston, IL: Northwestern University Press.

[8] Shaw, P.J.A. (2003) *Multivariate statistics for the Environmental Sciences*, Hodder-Arnold. ISBN 0-3408-0763-6.

[9] Stevens, J. (1986). *Applied multivariate statistics for the social sciences.* Hillsdale, NJ: Lawrence Erlbaum Associates.

[10] Usoro, A. E and Ibiok, E. U (2008): Application of principal component analysis to regression coefficients estimation as an alternative to ordinary least squares method. *International Journal of Natural and Applied Sciences*, 4(2):248-250.