Analysis of Nigeria Gross Domestic Product Using Principal Component Analysis

Olayiwola, O.M[1], Agboluaje[2], S. AFantolaJ. O[2] and AmalareA.A[3]

[1]Department of Statistics, Federal University of Agriculture, Abeokuta, Nigeria.
[2]Department of Mathematics and Statistics, The Polytechnic, Ibadan, Nigeria.
[3]Department of Mathematics and Statistics, Lagos State Polytechnic, Ikorodu, Nigeria.

ABSTRACT

Nigeria is classified as a mixed economy emerging market, and has already reached middle income status according to the World Bank, with its abundant supply of natural resource, well developed financial, legal, communications, transport sectors and stock exchange which is the second largest in Africa. The main purpose of this research is to build a model that can capture the best variables that predict the Gross Domestic Product (GDP) of Nigeria.Correlation matrix was used to know the degree of relationship that exists between the pairs of predictors of GDP. The principal component analysis was employed to reduce the multidimensional data. Scree plot was used to determine the spread of the trend of the components and bi plot was used to determine the degree of closeness of Agriculture, oil Export, External Reserves, Exchange Rate, Transportation, Education, and Communication. There is a strong relationship between pairs of Agriculture, oil Export, External Reserves, Exchange Rate, Transportation, Education, and Communication. The proportion of variance accounted for by the first component is 92%. This implied that only component 1 is sufficient to explain GDP. The Scree plot showed that the best component is component 1. The bi plot showed that Agriculture, oil Export, External.Reserves, Exchange.Rate, Transportation, Education, and Communication are closely related and stand as good predictors of GDP.

Introduction

Nigeria is classified as a mixed economy emerging market, and has already reached middle income status according to the World Bank, with its abundant supply of natural resource, well developed financial, legal, communications, transport sectors and stock exchange which is the second largest in Africa. Nigeria is ranked 31st in the world in terms of GDP as of 2011. Nigeria is the United States' largest trading partner in Sub-Saharan Africa and supplies a fifth of its oil (11% of oil imports). It has the seventh largest trade surplus with the U.S. of any country worldwide. Nigeria is currently the 50th largest export market for U.S. goods and the 14th largest exporter of goods to the U.S. The United State is the country's largest foreign investor. The International Monetary Fund (IMF) projected economic growth of 9% in 2008 and 8.3% in 2009. The IMF further projects 8% growth for the Nigerian economy in 2011. According to Citigroup (2011), Nigeria will get the highest average GDP growth in the world between 2010 and 2050. Nigeria is one of two countries from among Global Growth Generators Countries. Previously, economic development had been hindered by years of military rule, corruption and mismanagement. The restoration of democracy and subsequent economic reforms have successfully put Nigeria back on track towards achieving its full economic potential, it is now the second largest economy in Africa, and the largest economy in the west Africa region.

During the oil boom of the 1970s, Nigeria accumulated a significant foreign debt to finance majorly infrastructural development, with the fall of oil prices during the 1980s oil glut. Nigeria struggled to keep up with its loan payments and eventually defaulted on its principal debt repayments limiting repayment to the interest portion of the loans. Arrears and penalty interest accumulated on the unpaid principal which increased the size of the debt.

However, after negotiations by Nigeria authorities, in October 2005, Nigeria and its Paris club creditors reached an agreement in which Nigeria repurchased its debt at a discount of approximately 60%. Nigeria used part of its oil profits to pay the residual 40% freeing up at least$1.1billion annually for poverty reduction programs. Nigeria made history in 2006 by becoming the first Africa country to completely pay off its debt (estimated $30billion) owed to the Paris club for a cash payment of roughly $12billion (USD). (www.nigerianeconomy.com)

After fifty-four years of political independence, the productive base of the Nigerian economy remains weak, narrow and externally-oriented with primary production activities of agriculture and mining and quarrying (including crude oil and gas) accounting for about 65 percent of the real gross output and over 80 percent of government revenues. In addition, primary production activities account for over 90 percent of foreign exchange earnings and 75 percent of employment. In contrast, secondary activities comprising manufacturing and building and construction, which traditionally have greater potential for broadening the productive base of the economy and generating sustainable foreign exchange earnings and government revenues account for a mere 4.14 percent and 2.0 percent of gross out put respectively. Services or tertiary activities which depend on wealth generated by the productive sectors for

Tele:
E-mail addresses: laniyimathew@yahoo.com

their operations comprise about 30 percent of gross output. Significantly, service activities have been expanding their influence in the economy over the last decade accounting for over 35 percent of the growth of the real gross domestic product (GDP).

Data Collection Procedure

The data were obtained from central bank of Nigeria and statistical bulletin, spanning from the year 1981 to 2012

Literature Review

According to Jolliffe (2002) it is generally accepted that PCA was first described by KarlPearson in 1901. In his article " On lines and planes of closest fit to systems of points in space,"Pearson (1901) discusses the graphical representation of data and lines that best represent thedata. He concludes that "The best-fitting straight line to a system of points coincides in direction with the maximum axis of the correlation ellipsoid". He also states that the analysis used in his paper can be applied to multiple variables.However, PCA was not widely used until the development of computers. It is not really feasibleto do PCA by hand when number of variables is greater than four, but it is exactly for larger amount of variables that PCA is really useful, so the full potential of PCA could not be used until after the spreading of computers (Jolliffe, 2002).

According to Jolliffe (2002) significant contributions to the development of PCA were made byHotelling (1933) and Girshick (1936; 1939) before the expansion in the interest towards PCA. In1960s. as the interest in PCA rose, important contributors were Anderson (1963) with a theoretical discussion, Rao (1964) with numerous new ideas concerning uses, interpretations andextensions of PCA, Gower (1966) with discussion about links between PCA and other statisticaltechniques and Jeffers (1967) with a practical application in two case studies.

Methodology

Multivariate Principal Component Analysis(PCA)

Suppose that $X=(x_{(1)}\ x_2.\ldots\ldots.x_p)^1$ is a random vector with mean μ and covariance matrix ∑. Then the principal components of X denoted by$\gamma_{(1,)}\ldots\ldots\gamma_p$, satisfy the following conditions.

$\gamma_{(1,)}\ \gamma_2.\ldots\ldots\gamma_p$, are mutually uncorrelated .

$Var(\gamma_1)\geq Var(\gamma_2)\ldots\ldots Var(\gamma_p)$

$\gamma_j=\alpha_{ij}\ x_1+\alpha_2j\ x_2+\cdots+\alpha_{pj}x_p=a_j^1\ x$

Where $\alpha_j=\alpha_{ij}\ x_1+\alpha_2j\ x_2+\cdots+\alpha_{pj}x_p)^1$ is a vector of constant variance satisfying $//\alpha_j\ /^2=\sum_{(k=1)}^p a_{kj}^2=1$, for j=$\gamma_{(1,)}\ldots\ldots,\gamma_j$

Derivation of $\gamma_{(1,)}$ from a linear combination, then

$$Var(\gamma_1)=Var(a_j^1 x)=a_j^1\sum a_j \quad j=1,\ldots\ldots\ldots\ldots,p(1) \qquad From(1)$$

$$Var(\gamma_1)=a_j^1\sum a_1$$

The idea is to select a_1 in such a way that $Var(\gamma_1)$ is as large as possible, subject to the constraint $a_1^i\ a_1=1$

This is a standard problem in constrained optimization and be solved using the method of Lagrange multipliers .To use this method, from the lagrangian:

$$l_1=a^1\sum a_1-\partial(a^1 a-1) \quad (2)$$

The required a_1 is the value of a, that is a stationary point of (2)

Now define

$\nabla_a=(\partial/(\partial\alpha_1),\partial/(\partial\alpha_2),\ldots\ldots,\partial/(\partial\alpha_p))^1$

It may be shown that

$$\nabla_a(a_1^i\sum a)=2\sum a$$

$$\nabla_a(a_1^i a)=2a$$

A stationary point of (2) must satisfy:

$\nabla_a(L_1(a))=0$

Since

$$\nabla_a(L_1(a))=a\left(a_1^i\sum\propto a\right)-\partial\nabla_a(a_1^i a-1)$$

$$=2\sum a-2\partial a$$

it follows thata_1^i satisfies

$$2\sum a-2\partial a=0$$

That is

$$(\sum 1-\lambda I_p)\alpha_1=0\ (3)$$

A non- trivial solution ($\propto\neq0$) to the above exist if and only if

$$\left|\sum-\lambda I_p\right|=0$$

Where |.| is the determinate operator.

Thus λ must be an eigenvalue of ∑ with αi being its corresponding eigenvector.

Assume, for the moment, that the eigen-values of ∑ ,$\lambda_1\ldots\ldots\lambda_{(p,)}$ are all distinct, that is

$\lambda_1>\lambda_2>\cdots>\lambda_p\geq0$

Observed that

$$Var(Y)=Var(a_j^1 x)=a_j^1\sum a_j=a_j^1(\lambda I_p)$$

Using (3), which is equal to $a_j^1\ \lambda I_p=\partial$ since $a_1^i\ a_1=1$

It is clear that Var(Y)= ∂ will take its largest value at $\partial=\lambda_1$, since this is the value of the largest eigenvalue, with a_1 being the eigenvector corresponding to λ_1

The second PC

$\alpha_2' x$,, maximizes

$$\alpha_2' \Sigma \alpha_2$$

subject to the uncorrelated matrix with , or equivalently subject to

$$\text{cov}(\alpha_1' x, \alpha_2' x)\text{'}$$

where cov(x,y) denotes the covariance between random variable x and y.

$$\alpha_2' \Sigma \alpha_2 - \lambda \alpha_2' \alpha_2 - \phi \alpha_1 = 0 \qquad \text{But}$$

$$\text{cov}(\alpha_1' x, \alpha_2' x) = \alpha_1' \Sigma \alpha_2 = \alpha_2' \Sigma \alpha_1 = \alpha_2' \Sigma \alpha_1 = \alpha_2' \lambda_1 \alpha_1' = \lambda_1 \alpha_2' \alpha_1$$

Thus, any of the equations

$$\alpha_1' \Sigma \alpha_2 = 0, \;\; \alpha_2' \Sigma \alpha_1 = 0,$$

$$\alpha_1' \alpha_2 = 0 \;, \;\; \alpha_2' \alpha_1 = 0$$

Could be used to specify zero correlation between   and . Choosing the last of these equations, and noting that a normalization constraint is necessary, the quantity to be maximized is

$$\alpha_2' \Sigma \alpha_2 - \lambda(\alpha_2' \alpha_2 - 1) - \phi \alpha_2' \alpha_1,$$

where λ and ϕ are LaGrange Multipliers.
Differentiating with respect to   gives

$$\Sigma \alpha_2 - \lambda \alpha_2 - \phi \alpha_1 = 0$$

And multiplying this equation on the left by gives

$$\alpha_2' \Sigma \alpha_2 - \lambda \alpha_2' \alpha_2 - \phi \alpha_1 = 0$$

Which since the first two terms are zero and   reduces to . Therefore   is once more eigen-value of $\sum$ and   the corresponding eigenvector.

The Scree Test

With the Scree test, we plot the eigenvalues associated with each component and look for a"break" between the components with relatively large eigenvalues and those with smalleigenvalues. The components that appear before the break are assumed to be meaningful and areretained for rotation; those appearing after the break are assumed to be unimportant and are notretained.Sometimes a Scree plot will display several large breaks. When this is the case, we should lookfor the last big break before the eigenvalues begin to level off. Only the components that appearbefore this last large break should be retained.The Scree test can be expected to provide reasonably accurate results provided the sample islarge (over 200) and most of the variable communalities are large. However, this criterion has itsown weaknesses as well, most notably the ambiguity that is often displayed by Scree plots under

typical research conditions: Very often, it is difficult to determine exactly where in the Scree plota break exists, or even if a break exists at all.Why do they call it a "Scree" test? The word "Scree" refers to the loose rubble that lies at thebase of a cliff. When performing a Scree test, we normally hope that the Scree plot will take theform of a cliff: At the top will be the eigenvalues for the few meaningful components, followedby a break (the edge of the cliff). At the bottom of the cliff will lie the Scree: eigenvalues for thetrivial components.

Proportion Of Variance Accounted For

A criterion in solving the number of factors problem involves retaining a component if itaccounts for a specified proportion (or percentage) of variance in the data set. For example, wemay decide to retain any component that accounts for at least 5% or 10% of the total variance.This proportion can be calculated with a simple formula:

proportion=(Eigenvalue for the component of interest)/(Total eigenvalues of the correlation matrix)

In principal component analysis, the "total eigenvalues of the correlation matrix" is equal to thetotal number of variables being analyzed (because each variable contributes one unit of varianceto the analysis). An alternative criterion is to retain enough components so that the cumulativepercent of variance accounted for is equal to some minimal value.When researchers use the "cumulative percent of variance accounted for" as the criterion forsolving the number-of-components problem, they usually retain enough components so that thecumulative percent of variance accounted for at least 70% (and sometimes 80%).If we were to use 70% or 80% as the "critical value" for determining the number of componentsto retain, we would retain components 1 and 2 in the present analysis.The proportion of variance criterion has a number of positive features. For example, in mostcases, we would not want to retain a group of components that, combined, account for only aminority of the variance in the data set (say, 30%). Nonetheless, the critical values discussedearlier (10% for individual components and 70%-80% for the combined components) are

obviously arbitrary. Because of these and related problems, this approach has sometimes beencriticized for its subjectivity.

Egenvalue and Egenvector

For every square matrix A, a scalar λ and a nonzero vector x can be found such that

$$Ax = \lambda x. \tag{1}$$

In (1), λ is called an eigenvalue of A, and x is an eigenvector of Acorresponding to λ. To find λ and x, we write (1) as

$$(A - \lambda I)x = 0 \tag{2}$$

If $|A - \lambda I| = 0$, then $(A - \lambda I)$ has an inverse and $x = 0$ is the only solution. Hence, in order to obtain nontrivial solutions, we set $|A - \lambda I| = 0$ to find values of λ that can be substituted into (2.105) to find corresponding values of x. Recall, before defining the rank of a matrix, we first introduce the notion of linear independence and dependence. A set of vectors a1, a2, . . . , anis said to be linearlydependent if constants c1, c2, . . . , cn(not all zero) can be found such thatc1a1+ c2a2+• • •+cnan= 0.(3)If no constants c1, c2, . . . ,cncan be found satisfying (2.69), the set of vectors issaid to be linearly independent.

Table 1.Correlation Matrix Of The Selected Indicator On Nigeria Gdp

	Agric	OilExp	Exter.Reserves	Exch rates	Transport	Education	Communication
Agric	1.000	0.987	0.882	0.8557	0.973	0.994	0.953
OilExp		1.000	0.891	0.836	0.973	0.979	0.945
Exter.Reserves			1.000	0.764	0.909	0.846	0.919
Exch rates				1.000	0.913	0.869	0.720
Transport					1.000	0.963	0.908
Education						1.000	0.932
Communication							1.000

Table 2.Principal Component Analysis

	Comp.1	Comp.2	Comp.3	Comp.4	Comp.5	Comp.6	Comp.7
Standard deviation	2.538	0.574	0.416	0.187	0.100	0.090	0.047
Proportion of Variance	0.920	0.047	0.025	0.005	0.001	0.012	0.0003
Cumulative Proportion	0.920	0.967	0.992	0.997	0.998	0.9997	1.0000

Table 3. Significant Loading Of The Principal Components

	COMPONENT 1
Agric	-0.390
OilExp	-0.389
Exter.Reserves	-0.364
Exch rates	-0.349

Alternatively,(3) require that the columns of $A - \lambda I$be linearly dependent. Thus in$(A - \lambda I)x = 0$, the matrix $A - \lambda I$must be singular in order to find a solution vectorxthat is not 0.The equation $|A - \lambda I| = 0$ is called the characteristic equation. If A is n × n,

the characteristic equation will have n roots; that is, A will have n eigenvalues $\lambda 1, \lambda 2, . . . , \lambda n$. The λ's will not necessarily all be distinct or all nonzero. However, if A arises from computations on real (continuous) data and is nonsingular, the λ'swillall be distinct (with probability 1). After finding $\lambda 1, \lambda 2, . . . , \lambda n$, the accompanying eigenvectorsx1, x2, . . . , xncan be found using (2).If we multiply both sides of (2) by a scalar k , we obtain$(A - \lambda I)kx = k0 = 0.$ (4)Thus if x is an eigenvector of A, kxis also an eigenvector, and eigenvectors are unique only up to multiplication by a scalar. Theefore,vectors ,kxwhich satisfy (4) are called the eigenvectors or characteristic vectors.

Discussion Of Results

It was observed from table 1 that Agriculture, oil Export, External.Reserves, Exchange.Rate, Transportation, Education, and Communication are strongly correlated with each other. From table 2, the proportion of variance accounted for by the first component, which is a linear combination of Agriculture, oil Export, External.Reserves, Exchange.Rate, Transportation, Education, and Communication is 92%. This implied that only component 1 is sufficient to explain GDP.

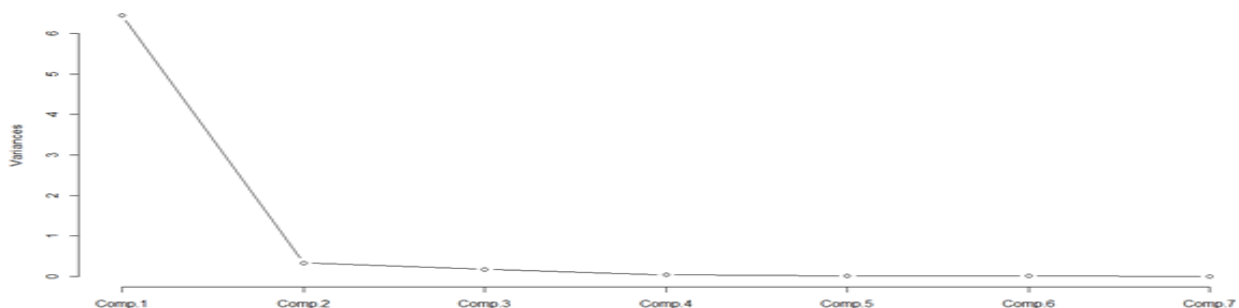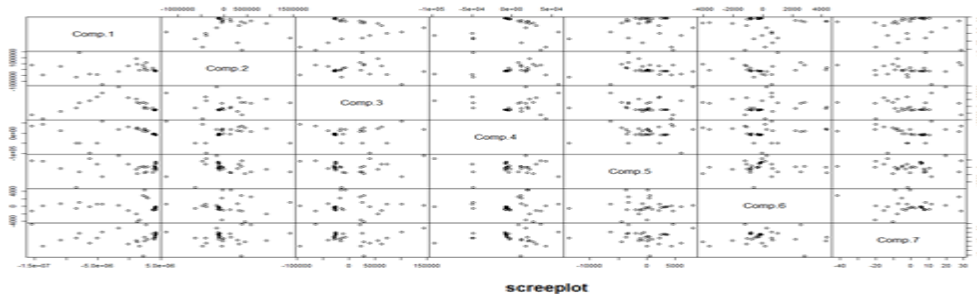**Figure1. A Pairs Plot From A Principal Component Analysis Of 7 Variables**

Figure 2.Scree Plot From A Principal Component Analysist.

Figure 1 shows the pair plot for the principal component analysis of the seven variables. It shows the degree of spread of the variables. From the Scree plot in figure 2, it showed that spread of the trend of the components. The best components are often greater or equals to 1. Hence, first component is chosen. Therefore, it reduced the seven components to one component. From figure 3, It is clearly shown that the first is equal or greater than 1. So, it is the best component for the principal component analysis. The bi plot in
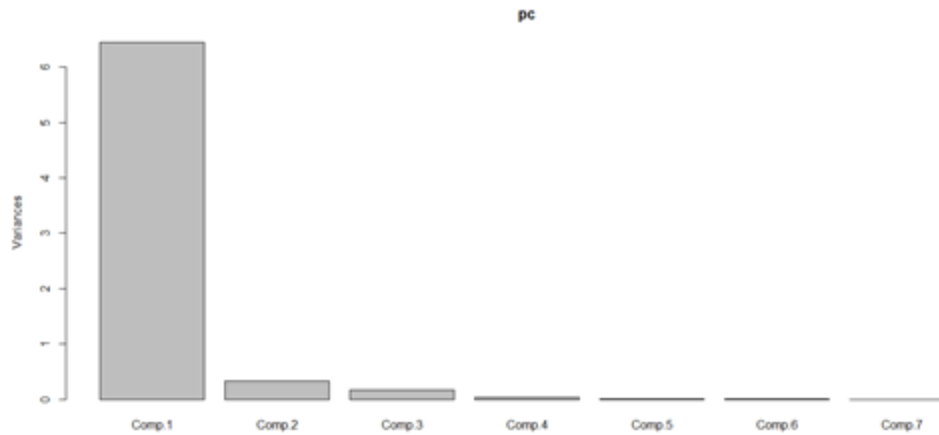

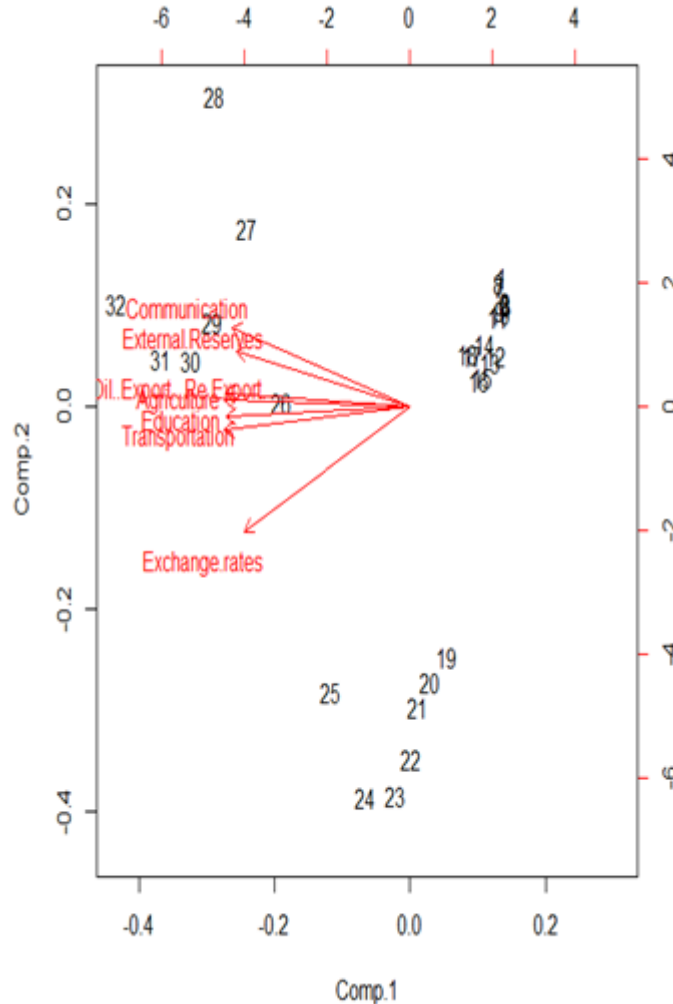**Figure 3. Bar Plot From A Principal Component Analysis**

Figure 4.A Bi plot From A Principal Component

figure 4 shows the degree of closeness of the Agriculture, oil Export, External.Reserves, Exchange.Rate, Transportation, Education, and Communication. It was observed that the Agriculture, oil Export, External.Reserves, Exchange.Rate, Transportation, Education, and Communication are closely related and have strong degree of relationship. From table 3, it was also observed that the best components to be chosen is the first component and this lead to the formulation of the PCA model below:comp 1= $0.390 \times$ Agriculture$-0.389 \times$oil..Export…Re.Export$-0.364 \times$External.Reserves$-0.3249 \times$Exchange.Rate$-0.390 \times$Transportation$-0.387 \times$Education$-0.0.375 \times$Communication

Conclusion

There is a strong relationship between pairs of Agriculture, oil Export, External Reserves, Exchange Rate, Transportation, Education, and Communication. The proportion of variance accounted for by the first component is 92%. This implied that only

component I is sufficient to explain GDP. The Scree plot showed that the best component is component 1. The bi plot showed that Agriculture, oil Export, External.Reserves, Exchange.Rate, Transportation, Education, and Communication are closely related and stand as good predictors of GDP.

Refrences

D. J. Bartholomew and M. Knott. Latent Variable Models and Factor Analysis, volume 7 of Kendall's Library of Satistics. Oxford University Press, New York, 1999.

Anderson, M. J., (2001), Permutation Tests for Univariate or Multivariate Analysis of Variance and Regression, Canadian Journal of Fisheries and Aquatic Sciences, 58, pp. 626–639.

Benko, M., H• ardle, W. &Kneip, A. (2009), `Common functional principal components', The Annals of Statistics 37(1), 1{34.

Berk, R. A. (2008), Statistical Learning from a Regression Perspective, Springer, New York.

Jolliffe, I. T. (2002).Principal component analysis (Second ed.): Springer.

Leger, L., & Leone, V. (2008). Changes in the risk structure of stock returns: Consumer confidence and the dotcom bubble. Review of Financial Economics, 17(3), 228-244.

Malava, A. (2006). Principal component analysis on term structure of interest rates. Unpublished Independent Research Project in Applied Mathematics Helsinki University of Technology Department of Engineering Physics and Mathematics

Adenipekun, O. (2006, May 25) Stakeholders' Move to Block Gaps in National Education Policy. Vanguard Newspaper (p.35).

Adepoju, A.&Fabiyi, A. (2007) Universal Basic Education in Nigeria: Challenges and Prospects.Http://uaps2007.princetonedu/download.aspx?submissionid=70830(Retrieved 08-04-11)

Burra, N. (1990) Child Labour in India: Poverty,Exploitation and Vested Interest (69-103) International Child Labour Seminar. Amsterdam.

Charles, J. O., Ikoh, M. O., Iyamba, E. I., & Charles, A. O. (2005) Human Development, Child Welfare and Addiction: Social work Perspective. Lagos: Serenity Printing and Publishing Co.

Dickson, K. A. (1985). "Education Culture and Development". In Emeyonu, E. N., Okoli, E. Ali, S. N. (ed). Education Culture and Development Africa (pp 46-51). New York: PWPA.

Federal Republic of Nigeria (1998).National Policy on Education (3rd Edition).

Federal Republic of Nigeria (2004) National Policy on Education (4th edition) NERDC press.

Federal republic of Nigeria 2000, Implementation Guideline for the Universal Basic Education (UBE) Programme Federal Ministry of Education Abuja Nigeria.

Adubi A. A. and F. Okunmadewa (1999) "Price, Exchange Rate Volatility and Nigeria's Agricultural Trade Flows: A Dynamic Analysis". AERC

Research Paper 87 African Economic Research Consortium, Nairobi March 1999.

Ayogu, Melvin (1999) "Empirical Studies of Nigeria's Foreign Exchange Parallel Market: Speculative Efficiency and Noisy Trading" Paper No 69 Africa Economic Research Consortium.

Bacchetta, P. and van Wincoop, E. (2003), "Why do Consumer Prices React Less than ImportPrices to Exchange Rates?" Journal of European Economic Association, 1, 662-670.

George, O. (2007), External Reserves Management in Nigeria. The Bullion Publication of the CBN, Vol. 31 No. 2. April – June.

IMF (2004), Guidelines for Foreign Exchange Reserve Management, IMF, Washington D.C.

International Monetary Fund (2001).Draft Guidelines for Foreign Exchange Reserves Management.

International Monetary Fund, (2003).Guidelines for Foreign Exchange Reserves Management.

Obaseki, P. J. (2007). Sources, Currency Composition and uses of External Reserves.CBNBullion.Vol.32 (2).

Obaseki, P.J. (2007), Foreign Exchange Management in Nigeria. Past, Present and Future.CBN Economic and Financial Review.Vol.29, No. 1.

Methods of Multivariate Analysis Second Edition ALVIN C. RENCHER Brigham Young University

Abu, AbdullahiUsman, 2010, Government Expenditure And Economic Growth In Nigeria, 1970-2008: A Disaggregated Analysis, Department of Economics, University of Abuja, PMB 117 Bureau of public affairs, Nigeria village case studies

The Review Of The Nigeria Economy(2010)

Econometrics in RGrant V. Farnsworth(2008)

Central Bank of Nigeria Website (2012).www.cbn.com

Gomez, K.A. and Gomez, A.A. (1984).Statistical Procedures for Agricultural Research (2nded.). New York: John Wiley & Sons.