Chuka University
info@chuka.ac.ke | 020 231 0512 / 020 231 0518

MODELLING KENYA MACROECONOMIC INDICATORS USING PRINCIPAL COMPONENT ANALYSIS

Morris Kateeti Mbaluka*[1]*, Dennis K. Muriithi, and Gladys G. Njoroge *Department of Physical Sciences, Chuka University, P.O. Box 109-60400, Chuka, Kenya* [1]*Corresponding Author Email: Morriskateeti@gmail.com; dkariuki@chuka.ac.ke*

How to cite:
Mbaluka, M. K., Muriithi, D. K. and Njoroge, G. G. (2021). Modelling Kenya macroeconomic indicators using principal component analysis. *In: Isutsa, D. K. (Ed.). Proceedings of the 7th International Research Conference held in Chuka University from 3rd to 4th December 2020, Chuka, Kenya, p. 565-573*

ABSTRACT
Kenya's economic growth has been lower than in other nations, especially the European nations for a long time. Macroeconomic indicators are the main factors that affect the economic growth of a country. The study sought to model Kenya macroeconomic indicators using principal component analysis. The study used PCA to capture data of 34 macroeconomic indicators for the period 1980 to 2019. The study aimed at applying PCA to reduce the dimensionality of the macroeconomic indicators and classify them into principal components. The study aimed at improving the way macroeconomic data has been handled in the past since several assumptions about the relationship between macroeconomic variables and economic growth have been made in the previous studies. The KMO statistics was found to be 0.720 and the p-value was 0.000. The KMO statistics indicated that the correlation matrix was appropriate for component analysis and the p-value depicted a significant difference. With reference to the correlation matrix, the variable were found to be closely correlated. Principal Component analysis was used to reduce the variables using varimax technique to principal components without compromising the variability of the original data. Only 8 variables (Principal Components) were retained since they explained 85% of the overall variations after scree plot, Kaiser Criterion and parallel analysis extraction approaches were utilized. The first component explained 28% of the total variance and was highly correlated with 10 macroeconomic indicators. Since the first principal component had the highest variance it was concluded that the monetary related macroeconomic indicators greatly impact economic growth in Kenya. Future researchers should consider having more diversified variables to help explain how economic growth is impacted by the all-round macroeconomic indicators.
Keywords: Principal Components, Gross Domestic Product, Eigenvalues, Eigenvectors, Correlation Matrix

INTRODUCTION
Pearson first described principal component analysis (PCA) in 1901, by preparing a fully functional method that generates a set of orthogonal axes placed in decreasing order and determining the main directions of the variability of samples (Pearson, 1901). Principal component analysis helps researchers to be able to reduce the number of possible variables and group them into factors (principal components). Some of these variables tend to be redundant, hence the importance of using PCA (Zou *et al*., 2006). Principal component analysis helps ensure that all variables that are related to each other are grouped together, especially if they measure the same construct.

When utilizing a large data set, it becomes increasingly difficult to interpret. Therefore, PCA is used to reduce the dimensionality of such data sets with minimal loss of information. The principal component analysis helps in the creation of uncorrelated variables that successfully maximize the variance. When dealing with many variables, say more than 20 variables, it becomes difficult to study these variables individually. For instance, fitting a regression model to a data set comprising of more than 20 variables would be much difficult to determine the effect of each of these variables (Corner, 2009). However, such a data set dimensionality can be reduced with minimal loss of information, hence creating new uncorrelated variables that are easier to study. Principal components are obtained from Eigenvalues calculated from the data set matrix. Therefore the principal components are defined by the data set at hand hence making PCA an adaptive data analysis technique.

Researchers have previously adopted PCA to carry out studies in different fields. For instance, a study by Esmaeili & Shokoohi (2011) applied PCA to assess the effects of oil prices on world food prices. The study's main objective was to apply PCA to investigate the movement of food prices and the macroeconomic index, especially the oil price. Additionally, PCA was meant to understand the influence of the macroeconomic index on food prices. The GDP, food production index, consumer price index, and crude oil prices were the macroeconomic variables that were studied for the period 1961 and 2005 around the world. Scree plots and the proportion of variance (Kaiser Criterion) were used in determining the optimal number of common factors. The correlation coefficient between the extracted principal components and macroeconomic index varied between 0.87 for the world GDP and 0.36 for the consumer price index. The study concluded that the food production index had the greatest influence on the macroeconomic index. Additionally, it was also concluded that the oil price index had an influence on the food production index.

However, oil prices had an indirect impact on food prices. Although the study was effective, it failed to use parallel analysis which has been found effective to help in making the decision of the number of components to be retained.

When performing principal component analysis, the challenge of over and under extraction of components usually arises hence the need to use several procedures when extracting components. A study by Njoroge *et al.* (2014) applied PCA to evaluate secondary school examination results. The study aimed at finding the principal components in terms of subjects that contribute to student's performance. Data for three years in school that were purposively selected from Nyanza, Nairobi, Rift Valley, and eastern provinces in Kenya was used. The SPSS software was used for analysis. Principal component analysis brought out the component loadings and the correlation structure between the different subjects in which one principal component was extracted. The results depicted that all subjects were highly correlated, and the first component had the highest variance. The principal component emerged to be English; hence, it was considered the subject that played the most significant role in the performance of the examination. This study used both Kaiser and Catelli scree plots extraction procedures. However, this study exhibits the problem of under or over-extraction since it did not go further to apply parallel analysis in the determination of the components to be retained.

Generally, most researchers find it challenging to determine the number of components that they should retain when applying PCA. This results in the problem of the under-extraction of over-extraction. To effectively use Keiser's criterion, Boligon *et al*., (2016), suggest that a researcher need to use a sample size of more than 250 observations and have an average commonality of more than 0.6 so that they can retain all factors with eigenvalue beyond 1. Field (2005) further makes a suggestion that a sample size of more than 300 observations makes scree plot an effective procedure of factor extraction. However, Granato *et al*. (2018) analyzed different studies that utilized various principal component extraction methods, including the Keiser criterion, scree plot, and parallel analysis. The comparison made by the study reviewed that parallel analysis was the most effective method in deciding the number of components a researcher should retain. Once the principal components are obtained, they must be renamed into new variables to be used in making inferences.

The principal component analysis is currently used in exploratory data analysis and in coming up with predictive models. It is done through the decomposition of the auto values of the covariance matrix. The results of the PCA are discussed in terms of principal components (Factor scores) (Brown, 2009). The main objective of PCA is to construct a linear combination of variables under study so as to explain the variance and covariance of a random vector that consists of random variables. The linear combination is what is known as the principal components.

Consider a random vector of interest $X' = (X_1, X_2, ..., X_N)$ with a covariance matrix $\sum$ and eigenvalues

$\lambda_1 \geq \lambda_2 \geq \lambda_3 \geq ... \geq \lambda_n \geq 0$

We have the linear combinations as follows;

$$\begin{aligned} Y_1 &= a_1' X = a_{11} X_1 + a_{12} X_2 + ... + a_{1p} X_p \\ Y_2 &= a_2' X = a_{21} X_1 + a_{22} X_2 + ... + a_{2p} X_p \\ &\vdots \\ Y_p &= a_p' X = a_{p1} X_1 + a_{p2} X_2 + ... + a_{pp} X_p \end{aligned} \tag{1}$$

We have $Var(Y_i) = a_i' \sum a_i$ and $Cov(Y_i Y_k) = a_i' \sum a_k$ where $i, k = 1,2,3,..., p$

The principal components $Y_1, Y_2, ..., Y_P$ should, therefore, capture as much information as possible.

Let $\sum$ be the covariance matrix with the eigenvalue eigenvector pairs $(\lambda_1, \ell_1), (\lambda_2, \ell_2), ..., (\lambda_p, \ell_p)$, and $\lambda_1 \geq \lambda_2 \geq \lambda_3 \geq ... \geq \lambda_n \geq 0$, then the th principal component is given by:

$$Y_1 = \ell_1' X = \ell_{11} X_1 + \ell_{12} X_2 + ... + \ell_{1p} X_p \tag{2}$$

for $i = 1,2,3,..., p$

It is interesting to note that the variance of the ^{h} principal component is the ^{h} eigenvalue.

$Var(Y_i) = \ell_i' \Sigma \ell_i = \lambda_i$ and $Cov(Y_i Y_k) = \ell_i' \Sigma \ell_k = 0$ where $i = 1,2,3,..., p$ and $i \neq k$

The principal components are linear combinations of the random variables. They are uncorrelated and have variances equal to the eigenvalues of Σ (the covariance matrix), and their development does not require any distributional assumption about multivariate normality (Lever *et al.*, 2017).

The proportion of the total variance explained by the ^{h} principal component can be written as:

$$\frac{\lambda 1}{\lambda 1 + \lambda 2 + ... + \lambda p} \tag{3}$$

Where is the eigenvalue of the ^{h} principal component. For instance, if the first principal components can explain most of the variations in the population covariance, then variables can replace the original variables with little loss of information (Lever *et al.*, 2017).

The main objective of PCA is to find common factors known as Principal components in the form of linear combinations of variables under study hence ending up ranking them according to their importance. In PCA, the extractions of PC can be made using either the original multivariate data set or using the covariance matrix if the original data set is not available (Field, 2016). In deriving PC, the correlation matrix was used, instead of the covariance matrix since different variables in the data set are measured using different units and thus might have different variances. Using the correlation matrix is equivalent to standardizing the variables to zero mean and unit standard deviation. Obtaining the principal component involved decomposing the covariance matrix of the random vector of interest. After the transformation of the random vector, the covariance matrix relative to the transformed vectors was used to determine the components. In this case, the principal components were determined from the originally standardized variables' covariance matrix. These are equal to the extracting the principal components using the correlation matrix of the original variables. This, this study applied PCA to model Kenya macroeconomic indicators. As opposed to other studies that have used PCA before, this study used Kaiser Criterion, scree plots, and parallel analyses for extracting and making decisions on retaining principal components. The main objective of the study was to apply PCA to reduce the dimensionality of Kenya Macroeconomic variables and classify them into principal component based on communalities. The analysis of data was done using SPSS and R Software.

METHODOLOGY

Research Design and Data Collection

The study adopted correlational research design. Correlational research design was used in the efforts of determining the kind of relationship that naturally occurs between the variables under study. The correlational research design helped figure out the variables related to each other in any way (Marczyk et al., 2005).

This study aims at examining 34 macroeconomic variables. The data set matrix is a $N \times P$ matrix where N is the macroeconomic variables. The study used secondary data, which was collected from the Kenya National Bureau of Statistics (KNBS) and the World Bank websites. The data was be stored in excel sheets then later imported to statistical software for analysis.

RESULTS AND DISCUSSION

Keiser-Meyer-Olkin

The Kaiser-Meyer Olkin (KMO) and Bartlett's Test help in measuring sample adequacy hence helping examine the appropriateness of principal component analysis. An approximate chi-square, degrees of freedom and significance level are utilized in helping explain the adequacy.

Table 1: KMO and Bartlett's Test

Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		**.720**
Bartlett's Test of Sphericity	Approx. Chi-Square	1972.683
	df	561
	Sig.	.000

Null Hypothesis: The inter-correlation matrix of the variables is not different from an identity matrix.
Alternate Hypothesis: The inter-correlation matrix of the variables is different from an identity matrix.

Test Results
$\chi2 = 1972.683$; df = 561; $p<0.0001$

Statistical Decision
According to the test results, the inter-correlation matrix of the variables is significantly different from an identity matrix. This implies that the sample inter-correlation matrix did not come from a population in which the inter-correlation matrix is and identity matrix. Principal Component Analysis is usually recommended for analysis if the Bartlett's test of sphericity is statistically significant that is p-value is less than 0.05 and the KMO statistics exceeds 0.6. According to the rule of thumb, a Kaiser-Meyer-Olkin Measure of Sampling Adequacy that is greater than 0.7 is considered to be a good indication that PCA is useful for the variables under study. According to the results in table 1. The KMO statistics was 0.720 which is a good indication that the correlations matrix is for component analysis. The Bartlett's Test of Sphericity was used to test the difference between the correlation matrix for variables and the identity matrix. The Bartlett's Test of Sphericity obtained for the data was 1972.683 and p-value was 0.000 which was an indication of a significant difference hence implying that the correlation matrix for the measured variables was significantly different from the identity matrix hence remaining consistent with the factorable assumption of the matrix. Therefore, the Bartlett's Test in the above table is precisely sufficient for the data under study.

Communalities
The i^{th} communality can be said to be the sum of the square of the loadings of the i^{th} variables on the n common factors. Communality is used to measure the total percentage of the variance in any given variable explained by all components jointly.

Table 2: Communalities

Variable	Initial	Extraction
Broad money (% of GDP)	1.000	.944
Communications, computer, etc. (% of service imports, BoP)	1.000	.695
Current health expenditure (% of GDP)	1.000	.914
Domestic credit to private sector (% of GDP)	1.000	.925
Domestic general government health expenditure (% of GDP)	1.000	.807
Electric power consumption (kWh per capita)	1.000	.885
Exports of goods and services (% of GDP)	1.000	.924
Government expenditure on education, total (% of GDP)	1.000	.698
Gross capital formation (% of GDP)	1.000	.850
Gross domestic savings (% of GDP)	1.000	.933
Gross fixed capital formation (% of GDP)	1.000	.872
Gross savings (% of GDP)	1.000	.825
Households and NPISHs Final consumption expenditure per capita growth (annual %)	1.000	.773
ICT goods imports (% total goods imports)	1.000	.905
Imports of goods and services (% of GDP)	1.000	.866
Inflation, consumer prices (annual %)	1.000	.856
Insurance and financial services (% of service imports, BoP)	1.000	.661
Labor force participation rate, total (% of total population ages 15-64)	1.000	.948
Lending interest rate (%)	1.000	.905
Life expectancy at birth, total (years)	1.000	.961
Military expenditure (% of GDP)	1.000	.851
Population, total	1.000	.955
Official exchange rate (LCU per US$, period average)	1.000	.944
Foreign direct investment, net inflows (BoP, current US$)	1.000	.765
Unemployment, total (% of total labor force)	1.000	.842
Total fisheries production (metric tons)	1.000	.908
Tariff rate, applied, weighted mean, all products (%)	1.000	.707
Manufacturing, value added (% of GDP)	1.000	.845
Air transport, freight (million ton-km)	1.000	.889
Air transport, passengers carried	1.000	.936
Aquaculture production (metric tons)	1.000	.919
Cereal yield (kg per hectare)	1.000	.654
Access to electricity (% of population)	1.000	.906
Remittance inflows to GDP (%)	1.000	.726

Extraction Method: Principal Component Analysis

Table 2 depicts that the communalities of each macroeconomic indicator is greater than 0.7 and implication that all the indicators have a similar pattern hence highly correlated. The high correlation indicates that all the indicators highly influence economic growth.

Scree Plot and Parallel Analysis

A scree plot is a line plot of the eigenvalues of the principal components. It is used in determining the number of principal components to retain. It gives a precise visualization of the magnitude of variability with each one of the principal components. On the other hand, parallel analysis help in determining the number of variables to be retained. It is useful since it has been found to be the most accurate.

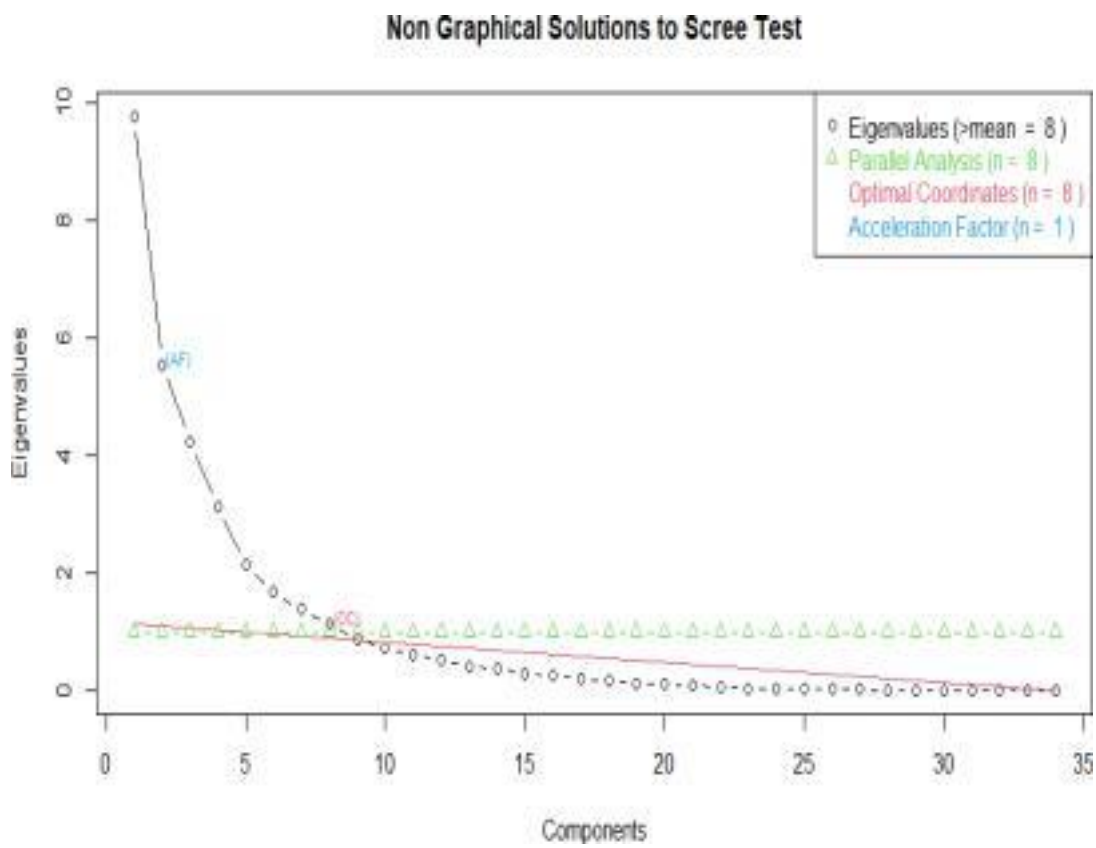

Figure 1: **Scree plot**

The Scree test was carried out to help visually analyze the Eigen values for point of inflection. The components to be retained according to the scree plot were the observations above the point of inflection. The visual representation above appears to have a decrease in downward slope after the ninth principal component an indication that the 8 preceding principal components can be precisely summarized to be representatives of the variables in totality. As indicated in the scree plot, the utilization of Eigen values extracted 8 principal components from the data. According to Kaiser's Eigen value >1 rule, the principal components with eigenvalues exceeding 1 should be the only retained. However, to help avoid over extractions and under extraction, parallel analysis was carried out. Parallel analysis has been pointed out by various researchers as the most effective method of retaining principle components. Using parallel analysis, 8 variables were retained for further analysis.

Total Variance Explained

Total variance explained indicates the ratio between the variance of principal component and the total variance. The total column usually give the eigenvalue or the amount of variance within the original variables that is accounted by each of the principal components.

Table 3: Total Variance Explained

Component	Initial Eigenvalues			Extraction Sums of Squared Loadings			Rotation Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	9.768	28.730	28.730	9.768	28.730	28.730	7.097	20.873	20.873
2	5.545	16.307	45.038	5.545	16.307	45.038	6.908	20.317	41.190
3	4.242	12.478	57.515	4.242	12.478	57.515	3.272	9.623	50.813
4	3.120	9.177	66.692	3.120	9.177	66.692	2.883	8.479	59.292
5	2.136	6.283	72.974	2.136	6.283	72.974	2.595	7.632	66.924
6	1.667	4.904	77.879	1.667	4.904	77.879	2.324	6.834	73.758
7	1.384	4.072	81.951	1.384	4.072	81.951	2.066	6.076	79.834
8	1.132	3.329	85.280	1.132	3.329	85.280	1.851	5.445	85.280
9	.867	2.550	87.830						
10	.714	2.099	89.929						
11	.620	1.824	91.754						
12	.519	1.528	93.281						
13	.414	1.218	94.499						
14	.369	1.086	95.585						
15	.292	.858	96.443						
16	.265	.779	97.222						
17	.201	.592	97.815						
18	.174	.512	98.327						
19	.127	.374	98.700						
20	.104	.306	99.007						
21	.088	.258	99.265						
22	.061	.180	99.444						
23	.045	.131	99.576						
24	.040	.118	99.693						
25	.033	.096	99.789						
26	.030	.087	99.876						
27	.017	.051	99.927						
28	.011	.033	99.961						
29	.007	.019	99.980						
30	.003	.008	99.988						
31	.002	.007	99.995						
32	.001	.003	99.998						
33	.000	.001	100.000						
34	6.676E-005	.000	100.000						

Extraction Method: Principal Component Analysis.

The results in table 3 depicts that the first component explains 28.7% of the total variance. The second component explains approximately 16.3% of the overall variance. The third component, fourth component, fifth component, sixth component, seventh component and eighth component explain about 12.5%, 9.2%, 6.3%, 4.9%, 4.1%, and 3.3% respectively of the total variation. The table depict that the first components explains the largest variation and a descending trend is established as we move from one component to the other. The study applied orthogonal Varimax technique to produce the uncorrelated factor structures. The study found that the summarized overall variation in the original set of data variables per the 8 retained components was approximately 85.3%.

Rotated Component Matrix

The rotated component matrix helps in determining what the components represent. It contains the estimates of the correlations between each of the variables and the estimated principal components.

Table 4: Rotated Component Matrix

	Component							
	1	2	3	4	5	6	7	8
Broad money (% of GDP)		.883						
Communications, computer, etc. (% of service imports, BoP)		.765						
Current health expenditure (% of GDP)				.655				.569
Domestic credit to private sector (% of GDP)	.543	.746						
Domestic general government health expenditure (% of GDP)	.625		.473					
Electric power consumption (kWh per capita)		.584		.472				
Exports of goods and services (% of GDP)	-.514					.465	.469	
Government expenditure on education, total (% of GDP)	-.478		-.410					
Gross capital formation (% of GDP)	.446	-.692						
Gross domestic savings (% of GDP)		-.799						
Gross fixed capital formation (% of GDP)	.535					.603		
Gross savings (% of GDP)			.469				.428	
Households and NPISHs Final consumption expenditure per capita growth (annual %)							-.763	
ICT goods imports (% total goods imports)								.926
Imports of goods and services (% of GDP)				.687		.428		
Inflation, consumer prices (annual %)							.809	
Insurance and financial services (% of service imports, BoP)					.691			
Labor force participation rate, total (% of total population ages 15-64) (modeled ILO estimate)			.792	-.450				
Lending interest rate (%)	-.590		.597					
Life expectancy at birth, total (years)	.909							
Military expenditure (% of GDP)		-.901						
Population, total	.473	.833						
Official exchange rate (LCU per US$, period average)		.927						
Foreign direct investment, net inflows (BoP, current US$)	.745	.400						
Unemployment, total (% of total labor force) (modeled ILO estimate)					.862			
Total fisheries production (metric tons)			.873					
Tariff rate, applied, weighted mean, all products (%)			.402	-.461	.441			
Manufacturing, value added (% of GDP)				.851				
Air transport, freight (million ton-km)	.695		-.438					
Air transport, passengers carried	.916							
Aquaculture production (metric tons)	.862							
Cereal yield (kg per hectare)						.801		
Access to electricity (% of population)	.818							
Remittance inflows to GDP (%)		.572						

a. Extraction Method: Principal Component Analysis.
b. Rotation Method: Varimax with Kaiser Normalization.
c. Rotation converged in 15 iterations.

The results in table 4 depicts that component 1 is highly correlated with 10 original variables (Broad money, Communications, computer etc.(% of service imports), Domestic credit to private sector, Electric power consumption, Gross capital formation, Gross domestic savings, Military expenditure, Official exchange rate, Population total and Remittance inflows to GDP). This component mostly resemble the monetary economy.

Component 2 was powerfully correlated with 7 of the original variables (Access to electricity, Air transport freight, Air transport, Aquaculture production, foreign direct investment, Lending interest rate, Life expectancy at birth). This component closely resembles the investment factor of the economy.

Component 3 was highly correlated with 4 of the original variables (Domestic general government health expenditure, Government expenditure on education, Labor force participation rate, Total fisheries production). This component resembles the expenditure segment of the economy.

Component 4 was highly correlated with 3 of the original variables (Exports of goods and services, Gross savings, Inflation consumer prices). This component resembles the trade and openness of the economy.

Component 5 was highly correlated with 3 of the original variables (Imports of goods and services, Insurance and financial services, Unemployment total). This component resembles the labor economy.

Component 6 was highly correlated with 3 of the original variables (Current health expenditure, Manufacturing value added, Tariff rate) which is a representation of open economy with government activities.

Component 7 had only 1 of the original variables that is ICT goods import a representation of economy technology advancement.

Component 8 was highly correlated with 3 of the original variables (Cereal yield, Gross fixed capital formation, Households and NPISHs Final consumption expenditure) a representation of consumption factor of the economy.

CONCLUSION

From the above analysis, all the 34 macroeconomic indicators have been found to be highly correlated. The variation in the data can be attributed to the first 8 components that collectively explain up to 85% of the total variation. However, from the loadings, components 1 has the greatest variation of 28% depicting that the monetary factors have greatest impact on Kenya Economic growth. Additionally, the second component explains up to 16% of the total variations depicting that the investment factors also have a greater impact on Kenya economic growth. The analysis depicts that ICT and consumption factors have less impact on the economic growth.

RECOMMENDATION

This study recommends that Government budgeting strategies should be reevaluated so as to ensure that the monetary factors are motivated further to help in further economic growth. Additionally policy makers and advisors together with the most of the financial institutions should work closely together to help the monetary economy take a step further hence increasing the country's economic growth. Future researchers should consider having more diversified variables to help explain how economic growth is impacted by the all-round macroeconomic indicators.

ACKNOWLEDGEMENT

My sincere gratitude goes to my supervisors: Dr. Dennis K. Muriithi and Dr. Gladys G. Njoroge (Faculty of Science, Engineering and Technology, Chuka University) for their inspiration and guidance throughout this research.

DECLARATION OF COMPETING INTERESTS

The author declares no competing interests.

AUTHORS' CONTRIBUTION

All authors have read and agreed to the paper.

REFERENCES

Boligon, A. A., Vicente, I. S., Vaz, R. Z., Campos, G. S., Souza, F. R. P., Carvalheiro, R., & Albuquerque, L. G. (2016). Principal component analysis of breeding values for growth and reproductive traits and genetic association with adult size in beef cattle. Journal of Animal Science, 94(12), 5014-5022.

Brown, J. (2009). Choosing the right number of components or factors in PCA and EFA. JALT Testing & Evaluation SIG Newsletter, 13(2).

Corner, S. (2009). Choosing the right type of rotation in PCA and EFA. JALT testing & evaluation SIG newsletter, 13(3), 20-25.

Esmaeili, A., & Shokoohi, Z. (2011). Assessing the effect of oil price on world food prices: Application of principal component analysis. Energy Policy, 39(2), 1022-1025.

Field, A. (2016). An adventure in statistics: The reality enigma. Sage.

Field, A. P. (2005). Discovering statistics using SPSS (2nd edition). London: Sage.

Granato, D., Santos, J. S., Escher, G. B., Ferreira, B. L., & Maggio, R. M. (2018). Use of principal component analysis (PCA) and hierarchical cluster analysis (HCA) for multivariate association between bioactive compounds and functional properties in foods: A critical perspective. Trends in Food Science & Technology, 72, 83-90.

Lever, J., Krzywinski, M., & Altman, N. (2017). Points of significance: Principal component analysis.

Marczyk, G., DeMatteo, D., & Festinger, D. (2005). Essentials of research design and methodology. John Wiley & Sons Inc.

Njoroge, E. W., Njoroge, G. G., & Muriithi, D. K. (2014). Evaluating Secondary School Examination Results: Application of Principal Component Analysis. Journal of Statistical and Econometric Methods, 3(2), 31-46.

Pearson, K. (1901). Principal components analysis. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, *6*(2), 559.

Zou, H., Hastie, T., & Tibshirani, R. (2006). Sparse principal component analysis. *Journal of computational and graphical statistics*, *15*(2), 265-286.

****