

## **Cluster Analysis of Macroeconomic Indices**

**MM Kembe<sup>1</sup>, AA Onoja<sup>2\*</sup>**

<sup>1</sup>Benue State University, Makurdi, Nigeria

<sup>2</sup>University of Jos, Jos, Nigeria

### **Research Article**

Received date: 13/12/2016

Accepted date: 09/02/2017

Published date: 15/02/2017

#### **\*For Correspondence**

AA Onoja, University of Jos, Jos Nigeria

**E-mail:** donmaston09@gmail.com

**Keywords:** Macroeconomic indicators, cluster analysis, gross domestic Product, Clusters Comparison

#### **ABSTRACT**

Macroeconomic indicators, help to shows the market values of shares transactions in a country, is of paramount concern on the issue relating to the economic development of nations and, thus, an analysis of its immediate causes is inevitable. There is also the view of development economists and policy makers that a well-developed macro economy is crucial for the mobilization of financial resources for long-term investment and thus constitutes one of the major pillars of economic growth. Cluster analysis classifies data into groups such that data in the same group are similar but significantly different from data in other clusters. In finance, these macroeconomic indices reflect the general performance of the Nigeria economy. The economy is the concern of how wealth is made, circulated, and meets its end. It centers on the pattern in which a country creates, disseminate, and consumes the tangible material commodities of life. However, the Gross Domestic Product (GDP) can be influenced by some of these macroeconomic indices. This research work is set out to identify, investigate, and classify the highlighted scenario to with; the relationship between macroeconomic indices in Nigeria and the economic growth. Some of the macroeconomic indices can be sort using Cluster analysis. Cluster Analysis is appropriate when you must classify some observed variables and wish to develop similarity between the numbers of artificial variables (called clusters) that will account for most of the similarities in the observed variables. Two-step clustering technique was adopted for the study. This procedure is an exploratory tool designed to reveal natural groupings (or clusters) within a dataset that would otherwise call for similarity. The Similarity measure of between the clusters formed was determined. Thirteen input variables were used to extract the groups via; (Macroeconomic factors: Gross Domestic Product, Public Finance, Fixed Asset, Private Investment, Net Export, Consumption, Net Savings, External Reserve, Net Import, and Balance of Payment, inflation rate, Forex Rate, and public debt). They Were classified based on the Gross Domestic Product of Nigeria from 1980 – 2010. The Cluster analysis results indicate that macroeconomic indicators tested have the significant influence on the Nigerian economy (Gross Domestic Product). From the ANOVA table, the cluster membership was used as a factor and tested against the Gross Domestic Product shows that it is significant ( $P = 0.000$ ) which lead to the rejection of the null hypothesis. It is recommendable that the Government of the day should pay more heels to our External Reserves, Fixed Assets, Consumption, Net Savings than any other macroeconomic factors in boosting a reliable Gross Domestic Product..

### **INTRODUCTION**

Macroeconomy is the center of investment throughout the world, Nigeria has not been an exempted. One of the head to head issues in economics and business over the years is whether macro economy as in its diverse nature causes economic

growth or whether it is a consequence of increased economic activity. Macroeconomic indicators, this explain how the market values of shares transactions in Nigeria, is of utmost concern to the economic growth of developing nations and, thus, an analysis of its possible determinants is discernible. Cluster Analysis as a primary classification method that can be used to classified multivariate data into similar clusters.

## Review of Literature

There is a view by development economists and policy makers that well developed macroeconomic indicators are serious for the mobilization of financial resources for long-term investment and thus constitute one of the major pillars of economic growth. The fact is that almost all the macroeconomic indices that come to African landmark are channeled to South Africa, ignoring Nigeria and other African countries with hardly any private inflow. Hondroyannis reported a bi-directional relationship between macroeconomic indices and economic development. Ewah, appraised the role of macroeconomic indices utility on economic growth in Nigeria and opined that the capital has the potential to induce growth but has not contributed rigorously to the economic growth due to low macroeconomic indices, low absorption capacity inadequate liquidity, misappropriation of funds among others. Okereke, posits that the cheap source of funds from the capital market remain a vital element in maintaining development of the economy. Cluster Analysis is a ubiquitous technique for multivariate data analysis and quality processing<sup>[1-5]</sup>. This research work discusses Cluster Analysis as a primary classification method, how many components to be in a group and have similarity on the macro economy. The main thrust of this research work is to classify and find similarities between the macroeconomic indices such as Gross Domestic Product, Public Finance, Fixed Asset, Private Investment Export, Consumption, Net Savings, External Reserve, Net Import, and Balance of Payment using Nigerian's data. The data for the research is a secondary data sourced from the internet via; the Central Bank of Nigeria's Record. Cluster analysis is a powerful multivariate statistical tool that classifies data into groups (Clusters), in many objects, based on their similarities. Like discriminant analysis, each object has multiple characteristics, which can be expressed as a random vector  $X = (X_1, X_2, \dots, X_p)$  with values that differ from component to component. The prime aim of cluster analysis is to identify similar objects by their similarities. Cluster analysis group same objects into clusters, so that the element within a group are very alike and objects from different groups are significantly different in their characteristics. Cluster Analysis becomes necessary when you must classify some observed variables and wish to develop similarity between the numbers of artificial variables (called clusters) that will account for most of the similarities in the observed variables. In finance, these macroeconomic indices reflect the general performance of the Nigeria economy. The economy is about how wealth emanates, circulated, and consumed. It concentrates on how a country produces, distributes, and consumes the tangible material commodities of life. It is also about how the income from these activities linked between those that contribute towards them. Macroeconomic indicators also known as macroeconomic indices are economic statistics that are released periodically by government agencies and private organization. These signs enhance the economic performance of a country or region, and, therefore, can have a significant effect on the Gross Domestic Product. However, the Gross Domestic Product (GDP) can be influenced by some of these macroeconomic indices. Some of which are Public Finance, Fixed Assets, Private Investment, Forex Rate, Inflation Rate, Public Debt, Interest Rate, Net Export, Consumption, Net Savings, External Reserve, Net import, and Balance of Payment. This research work is set out to identify, investigate, and classify the highlighted scenario to with; the relationship between macroeconomic indices in Nigeria economic performance. Some of the macroeconomic indices can be organize using Cluster Analysis. Indeed, despite the tremendous efforts by various stakeholders aimed at strengthening and deepening the macroeconomic indices, the Nigerian macroeconomic indices still faced with a lot of challenges. Lack of confidence on the part of investors, low level of market capitalization, leadership crisis, lack of clarity and transparency on the part of the regulatory bodies. Consequently, the Nigerian macroeconomic indices are unable to discharge its fundamental role in mobilizing long-term funds for the industrial sector and the economy in general effectively. The broad objective of this study examined the activities and classification of the Nigerian macroeconomic indices. The goals of the study are via; to examines and classify the operations of the Nigerian macroeconomic indices<sup>[6-10]</sup>. To evaluate the similarities between the macroeconomic indices about the economic growth and development in Nigeria, to make recommendations concerning how the operations of macroeconomic indices could be improved to boost economic growth and development of Nigeria Arising from the preceding. This study sets to provide answers to these questions: What is the classification of the macroeconomic indices about economic growth in Nigeria? There any similarities on the classification of the operations of the Nigerian macroeconomic indices in the Nigerian economy? How could the macroeconomic indices through its crucial role stimulate economic growth and development in Nigeria? In a quest to achieve the above objectives, the following hypotheses are formulated for testing:  $H_{01}$ : There is no significant similarity between macroeconomic indices factors and the growth and development of the Nigerian economy.  $H_{11}$ : There is a significant similarity between macroeconomic indices and the growth and development of the Nigerian economy.

## METHODS

### The Two-Step Cluster Analysis

This procedure is an exploratory tool designed to reveal natural groupings (or clusters) within a dataset that cannot be detected. The method used by this algorithm has many desirable characteristics that make it different from other forms of clustering methods:

- 1. Dealing with categorical and continuous variables:** By assuming variables to be independent, a joint multinomial-normal distribution can be placed on categorical and continuous variables.
- 2. Automatic selection of some clusters:** By comparing the values of a model choice criterion across different clustering solutions, the procedure can automatically determine the optimal number of clusters.
- 3. Scalability:** By constructing a group features (CF) tree that summarizes the records, the Two-Step algorithm allows you to analyze large data files.

## Distance and Similarity Determination

Cluster analysis group same objects into the similar cluster, and group unlike objects into different groups. Invariably, one need to define what is "same" and how to quantify similarity before clustering occurs. In the clustering technique, one often uses "distance" to depict how close each two of objects are. There is different distance determination and the most frequently used distance measures in cluster analysis as discussed.

### Euclidean distance

The Euclidean distance between any two objects, that is, the distance between object  $i$  and object  $k$ ,

$$d_{ik} = \sqrt{\sum_{j=1}^N (x_{ij} - x_{ki})^2} \quad (1)$$

Most references call this Euclidean distance "rule distance".

### Standardized Euclidean distance

Clearly, by using the Euclidean distance defined in Equation (1), the distance will be dominated by the few variables with larger numerical scales. If we want each variable to have an equal importance in deciding the distances between objects, then we need to standardize each variable so that each of them has about the same numerical scale. The standardization of each variable can do by subtracting the mean and dividing by the standard deviation, which is like standardizing normally distributed data. Specifically, we will transform each raw data  $x_{ij}$  into standardized data,

$$Z_{ij} \text{ given by } z_{ij} = \frac{(x_{ij} - x_{j\text{bar}})}{S_j} \quad (2)$$

$$\text{Where: } x_{j\text{bar}} = \sum_{k=1}^N \frac{x_{kj}}{N} \text{ and } S_j = \sqrt{\sum_{k=1}^N \frac{(x_{kj} - x_{j\text{bar}})^2}{N-1}}$$

then the standardized Euclidean distance between any two objects  $l$  and  $k$  is

$$d_{lk} = \sqrt{\sum_{j=1}^p (z_{lj} - z_{kj})^2} \quad (3)$$

Standardized Euclidean distance is also called the Pearson distance.

### Similarity

The similarity is a measure of similarity between two objects or two clusters. Its maximum value is 1 (or 100%), and its minimum value is 0. The larger the similarity value, the more similar the two objects or clusters are. Higher similarity value indicates smaller distance in objects. Specifically, given two objects and, a similarity measure between and is denoted by which satisfies the following conditions:

- $0 \leq S_{rs} \leq 1$
- $S_{rs} = 1$  if and only  $x_r = x_s$
- $S_{rs} = S_{sr}$

One commonly used similarity measure is defined as follows:  $S_{rs} = 1 - \frac{d_{rs}}{d_{\max}}$  (4)

Where  $d_{\max}$  is the maximum distance value in the distance matrix  $D$ . Another commonly used similarity measure is the Pearson product moment correlation defined as

$$q_{rs} = \frac{\sum_{j=1}^p (x_{ij} - x_{j\text{bar}})(x_{sj} - x_{s\text{bar}})}{\left[ \sum_{j=1}^p (x_{ij} - x_{j\text{bar}})^2 \sum_{j=1}^p (x_{sj} - x_{s\text{bar}})^2 \right]^{1/2}}$$

### Clustering criterion

To determine how the automatic clustering algorithm determines the number of clusters the clustering criterion can be used. Either the Bayesian Information Criterion (BIC) or the Akaike Information Criterion (AIC) can be specified.

## Data

The method works with both continuous and categorical variables. Cases represent objects to be classified, and the variables represent attributes upon which the clustering is based.

## Case order

Cluster features tree and the final solution may depend on the order of instances. To minimize order effects, randomly order the cases. One may want to obtain several different solutions with cases sorted in different random orders to verify the stability of a given solution. In situations where this is difficult due to enormous file sizes, multiple runs with a sample of cases arranged in various random orders might be substituted.

## Assumptions

The likelihood distance measure implies that variables in the cluster model are independent. Moreso, an individual continuous variable is assumed to have a normal (Gaussian) distribution, and each categorical variable is considered to have a multinomial distribution. Empirical internal testing indicates that the procedure is robust to violations of both the assumption of independence and the distributional assumptions, but you should try to be aware of how well these assumptions are satisfied. However, the data used for this project were collected from the central bank statistical bulletin from the National Bureau of Statistics [11-17].

## Analysis of data

The data for this research work were analyzed using the Predictive Analytical Software (PASW) IBM Version 20.

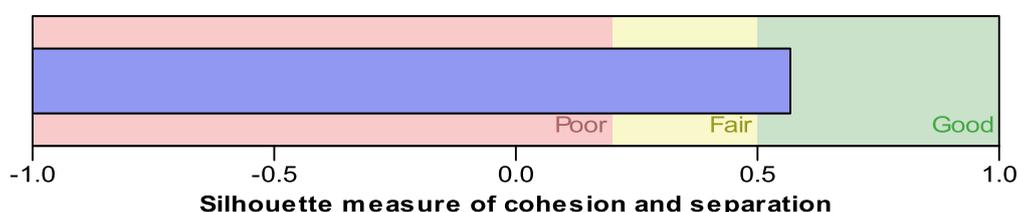
## RESULTS AND DISCUSSION

From the Model Summary table above, the two-step algorithm nonhierarchical method was adopted. 13 input variables were used to extract the clusters via; Macroeconomic factors: Gross Domestic Product, Public Finance, Fixed Asset, Private Investment, Net Export, Consumption, Net Savings, External Reserve, Net Import, and Balance of Payment, inflation rate, Forex Rate, and Public Debt. The data were classified based on the Gross Domestic Product of Nigeria from 1980 – 2010. The Cluster Quality served as those set of indicators used to check how consonant or right these clusters are in the classification (**Figure 1**). The Cluster Sizes shown in a pie-chart from the smallest to the largest (**Figure 2**). The smallest group is having five members and occupied about 16.1%, the largest cluster had about ten members and held 32.3%, the ratio of sizes of clusters from the largest clusters to smallest groups is 2.0 which means that no cluster in the cluster set is more than 2 times the other clusters. The Cluster table showed the sizes of the groups in percentages. The input variables arranged in Predictor of importance Order via; External Reserves, Fixed Assets, Consumption, Net Savings, Net import, Public Finance, Forex Rate, Public Debt, Private Investment, Balance of Payments, Interest Rate, Inflation Rate, and Net Export. The Evaluation Fields is the Gross Domestic Product (**Cluster 1**). The Predictor importance graph shows the macroeconomic indicator effects on the Gross Domestic Product from the most important (External Reserve) to the least important (Private Investment) indices (**Graph 1**).

### Model Summary

<b>Algorithm</b>	TwoStep
<b>Inputs</b>	13
<b>Clusters</b>	4

### Cluster Quality



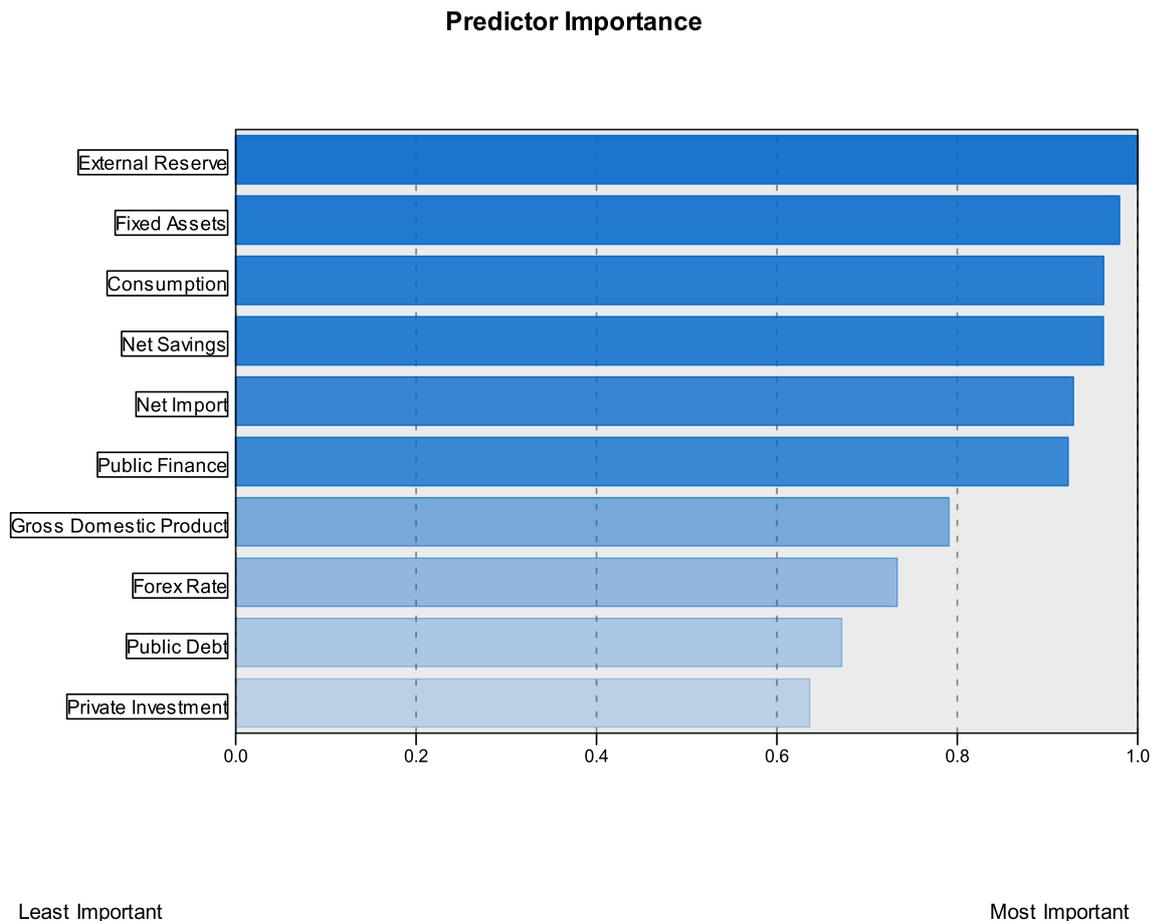
**Figure 1.** Model summary and cluster quality.

## Clusters

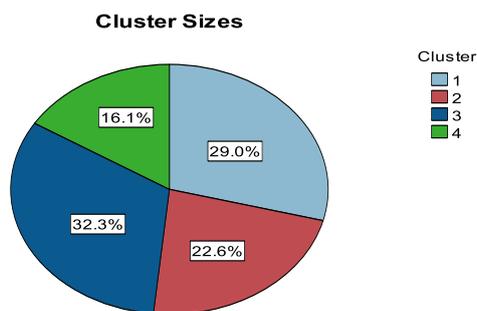
Input (Predictor) Importance  
 ■ 1.0 ■ 0.8 ■ 0.6 ■ 0.4 ■ 0.2 ■ 0.0

Cluster Label	1	2	3	4
Description				
Size	29.0% (9)	22.6% (7)	32.3% (10)	16.1% (5)
Inputs	External Reserve 3.34 Fixed Assets 6.50 Consumption 4.76 Net Savings 4.76 Net Import 4.02 Public Finance 4.16 Forex Rate 0.07 Public Debt 4.61 Private Investment 3.59 Balance of Payments 2.89 Interest Rate 1.10 Inflation Rate 1.16 Net Export 5.45	External Reserve 4.49 Fixed Assets 7.41 Consumption 5.61 Net Savings 5.61 Net Import 5.08 Public Finance 5.01 Forex Rate 1.15 Public Debt 5.78 Private Investment 4.12 Balance of Payments 4.33 Interest Rate 1.41 Inflation Rate 1.53 Net Export 5.76	External Reserve 5.92 Fixed Assets 8.47 Consumption 6.60 Net Savings 6.60 Net Import 6.09 Public Finance 5.91 Forex Rate 1.85 Public Debt 6.47 Private Investment 5.31 Balance of Payments 5.06 Interest Rate 1.36 Inflation Rate 1.10 Net Export 5.81	External Reserve 6.80 Fixed Assets 9.19 Consumption 7.26 Net Savings 7.23 Net Import 6.62 Public Finance 7.22 Forex Rate 2.13 Public Debt 6.70 Private Investment 6.13 Balance of Payments 5.28 Interest Rate 1.18 Inflation Rate 0.78 Net Export 5.86
Evaluation Fields	Gross Domestic Product	Gross Domestic Product	Gross Domestic Product	Gross Domestic Product

**Cluster 1.** Cluster showing input predictor values.



**Graph 1.** Graph showing predictor importance (ranging from least to the most important).



<b>Size of Smallest Cluster</b>	5 (16.1%)
<b>Size of Largest Cluster</b>	10 (32.3%)
<b>Ratio of Sizes: Largest Cluster to Smallest Cluster</b>	2.00

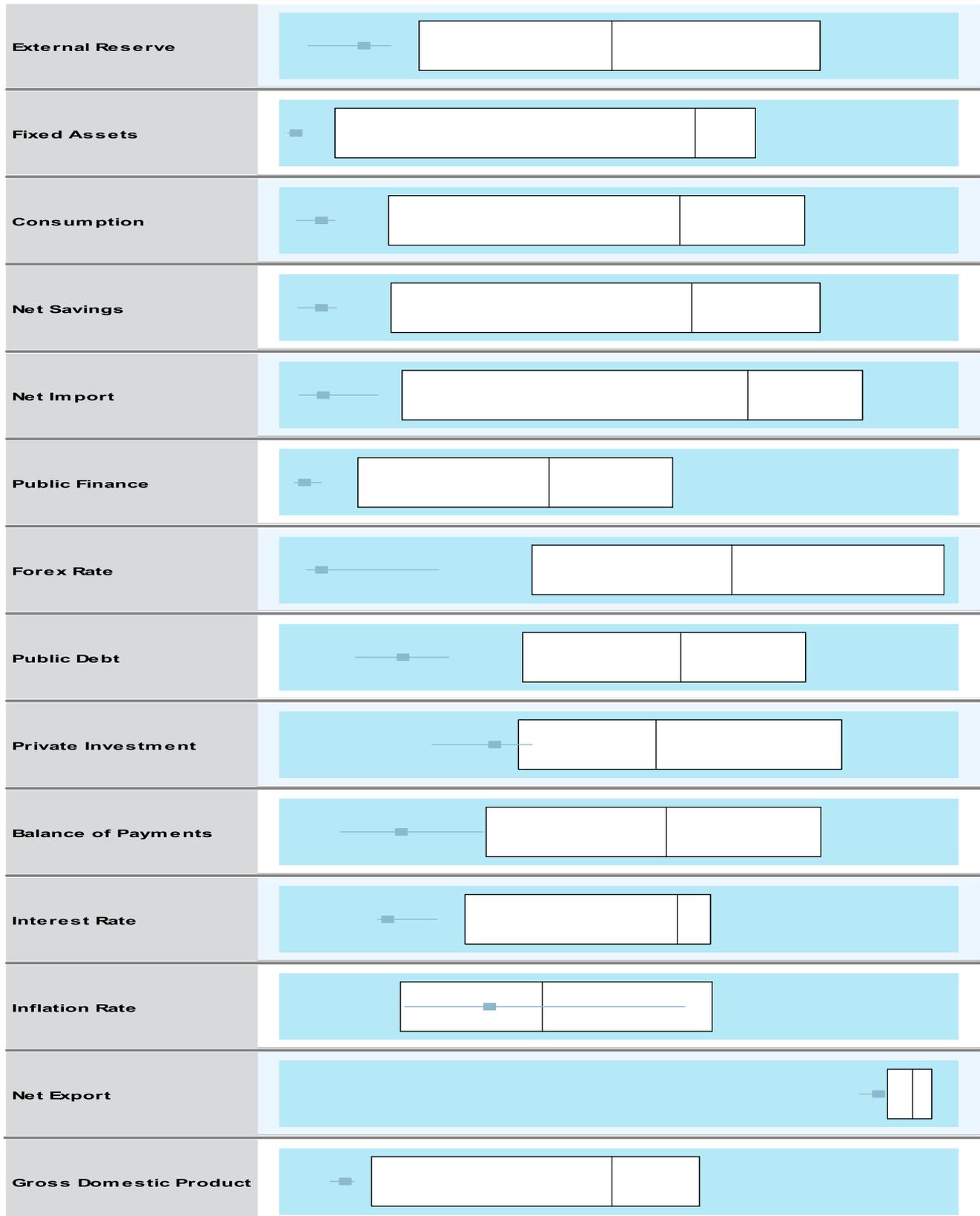
**Figure 2.** Pie diagram showing cluster sizes.

### Cluster Comparisons

The Cluster Comparison graph is a box plot used in comparing the clusters in the cause of classifications. The line at the middle represents the median while the upper and lower part of the table represent the mean at one standard deviation. In cluster one; it can see that the External Reserve, Fixed Assets, Consumption, Net Savings, Public Finance, Forex Rate, Public Debt, Private Investment, Balance of Payments, Interest Rate are all indigent these have resulted to relatively inflation rate and a poor Gross

**Cluster Comparison**

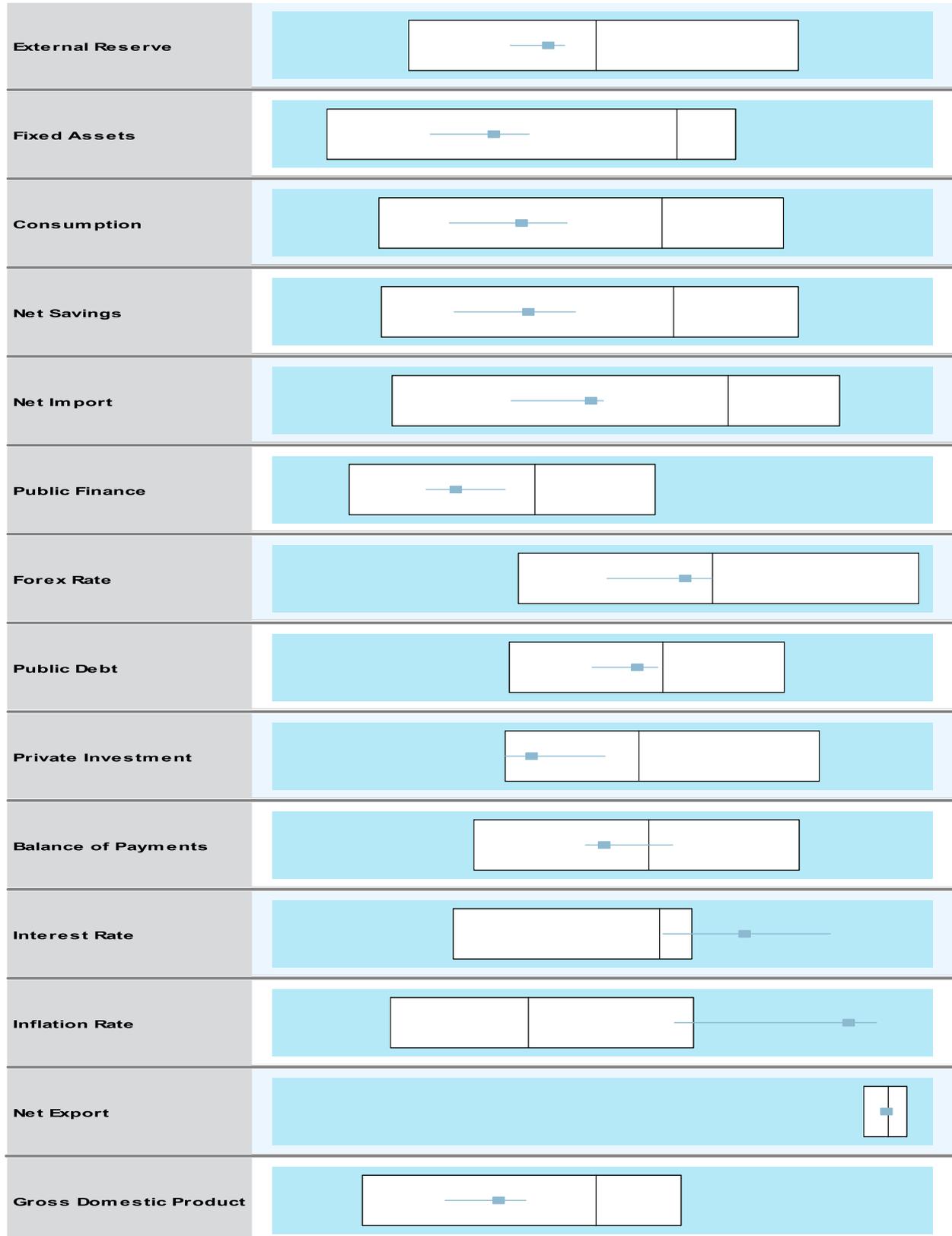
■ 1



**Cluster comparison 1**

**Cluster Comparison**

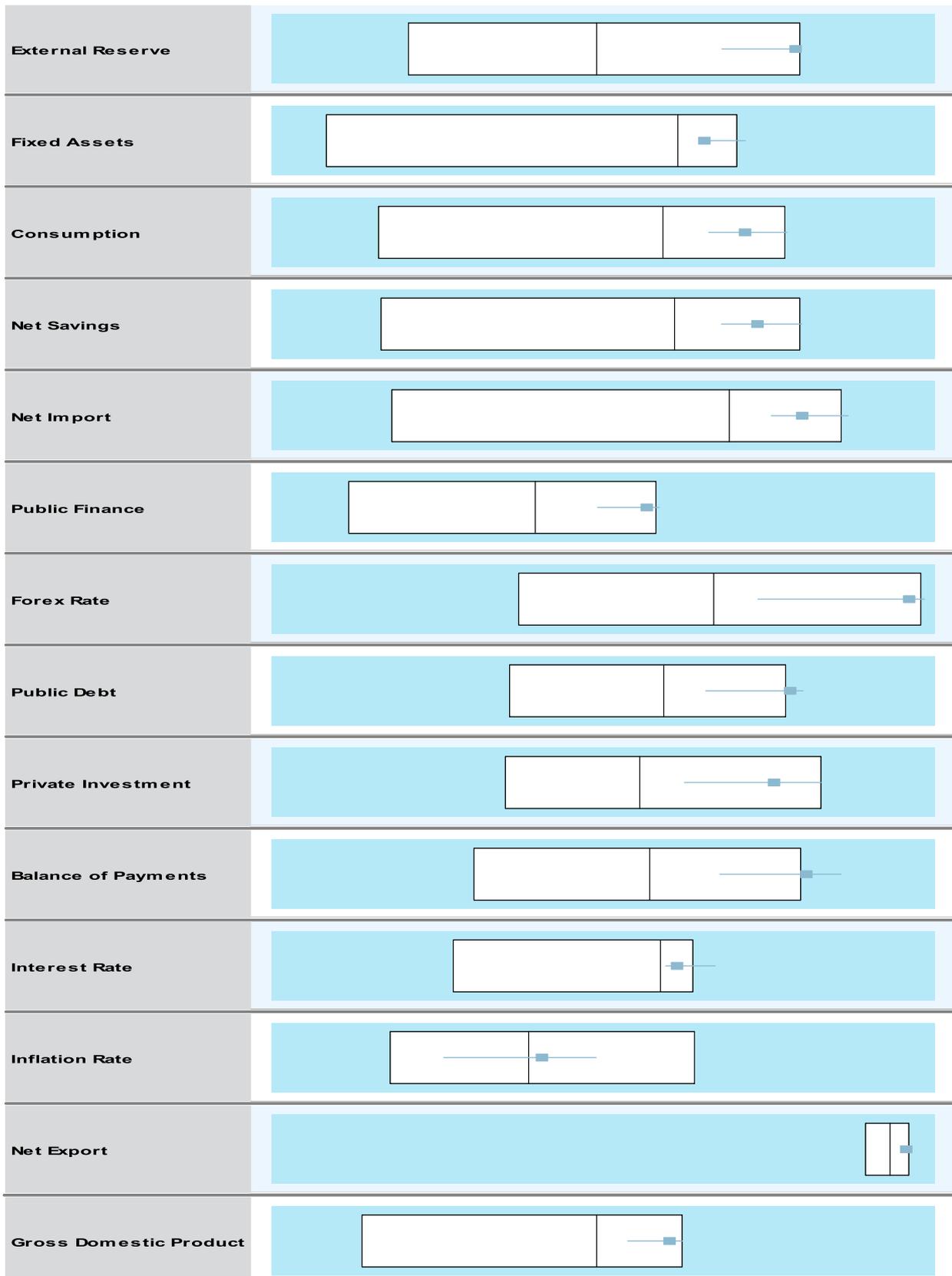
■ 2



**Cluster comparison 2**

**Cluster Comparison**

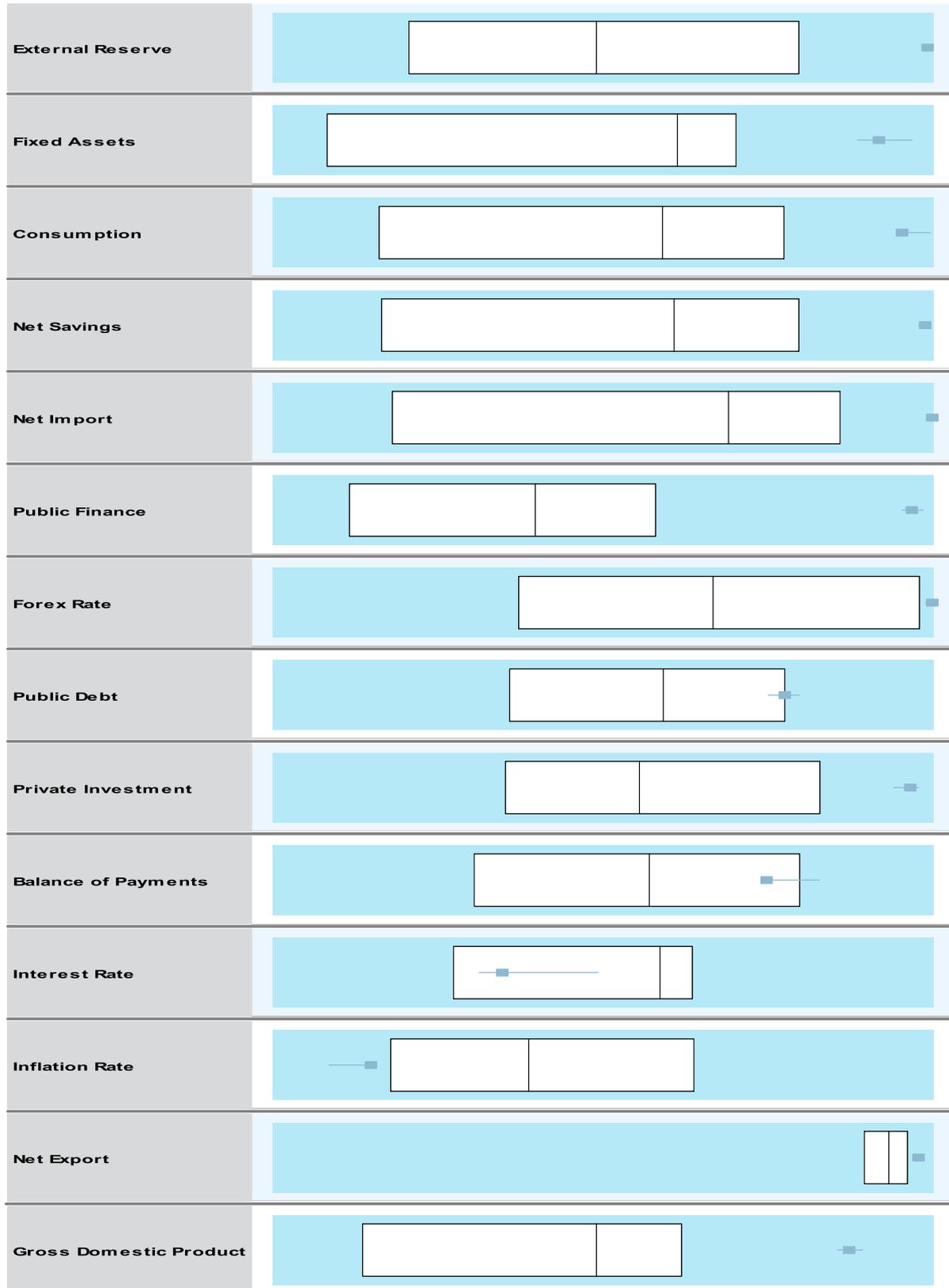
■ 3



**Cluster comparison 3**

**Cluster Comparison**

■ 4



**Cluster comparison 4**

Domestic Product. In Cluster two; External Reserve, Fixed Assets, Consumption, Net Savings, Public Finance, Forex Rate, Public Debt, Private Investment, and Balance of Payments, are all moderate compared to the first cluster. Interest Rate and Inflation Rate are high though there is an improvement on the Gross Domestic Product compare to the previous Cluster. In Cluster three; External Reserve, Fixed Assets, Consumption, Net Savings, Public Finance, Forex Rate, Public Debt, Private Investment, Balance of Payments, Interest Rate are all very high, and this have resulted in a high Inflation Rate and as well the Gross Domestic Product. In Cluster four the External Reserve, Fixed Assets, Consumption, Net Savings, Public Finance, Forex Rate, Public Debt, Private Investment, Balance of Payments, are all extremely high except for the Interest Rate which is moderate. This in-turn resulted in low Inflation Rate and high Net Export as well as the Gross Domestic Product. From the ANOVA Table above it can also be seen that p-value calculated is less than the p-value tabulated; that is  $p = 0.000 < 0.005$ , thus; reject the null hypothesis and conclude that there is a significant similarity between macroeconomic indices and the growth and development of the Nigerian economy (Table 1).

**Table 1.** ANOVA table showing the gross domestic product between different groups.

Gross Domestic Product					
Variables	Sum of Squares	Df	Mean Square	F	Sig.
Between Groups	33.419	3	11.140	88.605	0.000
Within Groups	3.395	27	0.126	--	--
Total	36.814	30	--	--	--

## CONCLUSION AND RECOMMENDATIONS

The Cluster analysis results indicate that macroeconomic indicators via; External Reserves, Fixed Assets, Consumption, Net Savings, Net import, Public Finance, Forex Rate, Public Debt, Private Investment, Balance of Payments, Interest Rate, Inflation Rate, and Net Export have significant effects on the Nigerian economy. This result supported by Poon and Taylor and Tursoy, Gonsel and Rjoub, which found that when macroeconomic indicators classified; there is a significant relationship between the macroeconomic indices and the Gross Domestic Product of a country. However, each cluster may significantly affect different sector in a different manner. That is macroeconomic indices that classified into the same group may significantly affect one area positively, but may significantly influence the other area negatively. It is also worth noting from the ANOVA table that all the cluster membership when used as a factor and tested against the Gross Domestic Product, shows that it is significant ( $P = 0.000$ ) which lead to the rejection of the null hypothesis. In recommendation; it is advisable that the Government of the day should pay more heels to our External Reserves, Fixed Assets, Consumption, and Net Savings than any other macroeconomic indicators in boosting a reliable Gross Domestic Product especially in the realization of the 2020 vision.

## REFERENCES

1. Bartlett MS. Test of significance in factor analysis. The British J Psychology 1950; 3: 77-85.
2. Cardima J and Joluffe IT. Loadiye and correlations in the interpretation of principal components. J Appli Statist 1995; 22: 203-214.
3. Cattell RB. The screen test for the number of factors multivariate behavioural research 1966; 1: 245-276.
4. Cooley WW and Lohnes PR. Multivariate data analysis. John Willy & Sons. Inc New York 1971.
5. Haiping L, et al. MPCA: Multilinear principal components analysis of tensor objects. Mineral Networks 2008; 19: 18-39.
6. Hill T and Twewicki P. Methods and applications. Statsoft Tulsa. 2007.
7. Horn JL. A Rationale and test for the number festus in factor analysis. Psychometrika 1965; 30: 179-185.
8. Horn JL and Engstorm R. Cattell's scree test about Bartlett's Chi-square test and other observation on the number of factors problem. Multivariate Behavioural Research 1979; 14: 283-300.
9. Hottelling H. Analysis of a complex of statistical variables into principal components. J Educ Psycholo 1933; 24: 417-441.
10. Hubbard R and Allen SJ. An empirical comparison of alternative methods for principal component extraction. J Busi Rese. 1987; 15: 173-190.
11. Humphreys LG and Muntanelia RG. An investigation of the parallel analysis criterion for determining the number of common factors. Multivariate Behavioural Research, 1975; 10: 193-205.
12. Hyverinen A, et al. Independent component analysis. New York: Wiley. 2001.
13. Johiffer IT, et al. A modified principal component technique based on the LASSO. J Computation and Graphical Statistics 2003; 3: 531-547.
14. Joluffe IT. Principal component Analysis. (2<sup>nd</sup> ed). New York: Springer-verlag. 2012.
15. Yang K and Trewn J. Multivariate statistical methods and quality Management. 2004.
16. Pearson K. On lines and planes of closet fit to systems of points in space. Philosophical Magazine Series 1961; 6: 519-572
17. Kcho Y. The basics of linear principal components analysis. Principal Component Analysis. 2012.