

Classification of Capital Markets by Using Cluster Analysis

M Mallikarjuna ^a, R Prabhakara Rao ^b

^a *Corresponding author: Research Scholar, Department of Economics, Sri Sathya Sai Institute of Higher Learning, Prasanthi Nilayam, AP, India.*

^b *Professor and Dean, Economics and Humanities, Sri Sathya Sai Institute of Higher Learning, Prasanthi Nilayam, AP, India.*

Abstract

Financial markets have become pivotal segments of an economy over the past three decades. Because they act as the facilitators between the investors and the industries by serving the dual purposes of channelizing the funds to the industries and fetching higher returns to the investors. However, many investors hesitate to invest their surplus funds in capital markets because of the dynamic nature and the inherent risk and uncertainty of the capital markets. The investors always aim to minimize the risk and maximize the returns by constructing optimum portfolio. After globalization, almost all the economies have liberalized their capital markets allowing the foreign investors to invest. In these circumstances, the investors can tap the opportunities that exist in different capital markets around the world. However, investing in randomly selected capital markets result in huge losses. So, one of the best ways to choose the markets is to look for the similarity in the characteristics of the markets, classify them into different categories, and select to invest in these markets. There are some international agencies such as International Monetary Fund, Dow Jones, Financial Times Stock Exchange, and Morgan Stanley Capital International, which classified the markets into different categories as developed, emerging and frontier markets by using some financial and socio-economic variables, however, this kind of classification cannot be verified by the investors since it is very complex to obtain the data for these variables. Also, there are very few studies, concerning the classification of capital markets based on financial variables. Therefore, we tried to classify the capital markets into different categories in much simple manner by using financial variables, which are available in open sources. We considered 60 capital markets around the world and employed clustering techniques to classify these capital markets as highly developed, advanced emerging, secondary emerging and frontier markets.

Keywords: Financial Markets, Classification, Portfolio Diversification, Clustering Methods

1. Introduction

During the last three decades, especially after globalization and liberalization of several economies, the world financial markets have witnessed a revolutionary change, and it showed a remarkable impact on the growth of the economies. This phenomenon was observed by many through their empirical research, for example, (Levine, 1997; Rajan and Zingales, 1998; Christopoulos and Tsionas, 2004; Gupta and Rao, 2014; Gupta and Rao, 2018) in different economies.

In addition to the integration of markets, the expansion of financial markets can be attributed to several other factors such as the awareness among the investors, data availability on market operations, innovations, development in information technologies, and the extensive use of communications etc. Also, with the advancement of these technologies, the market participants have direct access to all the information regarding market movements, opinions of experts, and the information of economic, political and social events, which in turn affect the financial markets. Due to these influences, there is a drastic decrease in operational cost in financial transactions as well as increase in handling large volumes of financial data all over the world. These changes along with macroeconomic conditions and policy decisions of each country opened new avenues for investments in different markets. This informational efficiency allowed the financial markets accessible to all agents who involved in the operations of the market. As a result of these changes, markets exposed to large swings and hence become more volatile. This makes the structure of financial markets more complex.

Moreover, the integration of markets coupled with automated trade facilities stimulated the investors' opportunities to diversify their investments and thereby increase the rates of return on investments making higher profits. This led to promotion of introducing more different financial products and instruments in the markets. In recent years the use of financial derivatives, such as forwards, options and swaps has increased in financial market operations. Because of these advances, it is imperative to comprehend the market structure and their characteristics of each country. However, understanding every market may not be feasible, nevertheless, identifying similar group of markets is essential for safe investments and risk minimization. In addition, this classification of markets may be useful to the several players in the financial markets such as banks, financial regulators, fund managers, agents, firms, policymakers of the Governments and other players of economic and financial fields in making their policy decisions appropriately.

So far, the classification of markets taken up by some agencies like International Monetary Fund (IMF), Morgan Stanley Capital International (MSCI) Financial Times Stock Exchange (FTSE), and Dow Jones (DJ) among others. In these classifications they considered a combination of quantitative as well as qualitative variables relating to economic and market characteristics of a country. Usually qualitative measurements will not reveal full information and these classifications may be of limited use (Mendes et al. 2018). Likewise, according to MSCI, if there is misclassification of a market, it is difficult to track the index and it will increase the risk and cost of the misclassification error. Moreover, these classifications are not updating dynamically according to the speed at which the changes taking place in the investment environment. This restricts the efficacy of the classification procedures.

Various studies have been undertaken to understand the characteristics of markets for example, Kohers, et al. (2006) have used the risk and return characteristics for the developed and emerging stock markets. Singh, (2008) studied the stock markets and economic growth in low and middle-income countries. Alrgibi et al. (2010) have conducted a study to identify the factors that discriminate the developed and emerging stock markets. Krichene, (2016) studied the differences between the mature and immature markets by analyzing the stylized facts of ten stock markets.

Recently, some researchers attempted to classify the markets by considering quantitative variables which enable us to recompute, reproduce and update the indices at any time as the information of the variables can be obtained very easily. For classification, one can employ some multivariate statistical techniques to provide empirical evidence. There exist very few studies in the literature such as Alrgibi et al. (2010),

Mendes et al. (2018) by using multivariate statistical method viz., discriminant analysis to classify the world markets. Neto et al., (2012) considered the stock market indices of the 34 countries and classified the countries based on two factors, i.e., volatility and B Coefficient by using ideal gas approach. More recently, Mendes et al. (2018) classified 40 countries based on quantitative variables as developed, emerging and frontier markets by using discriminant analysis.

However, the discriminant analysis requires an assumption of the variables under study follow a multivariate normal distribution which is a serious limitation as this method might not be efficient for the variables distributed non normally (Fraley and Raftery, 2002). Also, the discriminant analysis demands the apriori knowledge of variables. Hence it may lead to the problem of misclassification.

Since there are very few studies (Neto et al., 2012; Mendes et al., 2018) on the classification of financial markets based on quantitative financial variables. In this paper, we propose to use the quantitative variables related to the important market parameters viz., Access, Efficiency, Depth and Stability (World Bank, 2018) which captures the important features of the financial markets. Also, we employ most widely used unsupervised or clustering methods to classify the world major financial markets.

The rest of the paper is organized as follows. In section 2, we describe the data and methodology employed for the study, section 3 deals with the empirical results of the study and the conclusions are given in section 4.

2. Data and Methodology:

In this section, we describe the data and variables considered for the study and different methods employed to classify the capital markets.

2.1. Data:

In this study, we have taken annual data of the variables from the World Bank database from 2009 to 2016 of the capital market parameters viz., Access, Depth, Efficiency, and Stability for classification. The world markets considered in this study according to the availability and uniformity in data and the list is given in table 1. The variables of this study along with their notation are provided in table 2.

Table. 1. Capital Markets considered for Classification

Asia-Pacific (24)	Europe (21)	Africa (6)	North America (4)	South America (5)
1. Australia 2. China 3. Cyprus 4. India 5. Indonesia 6. Israel 7. Japan 8. Jordan 9. Kazakhstan 10. Kuwait 11. Lebanon 12. Malaysia 13. Mauritius 14. New Zealand 15. Oman 16. Pakistan 17. Philippines 18. Saudi Arabia 19. Singapore 20. Sri Lanka 21. South Korea 22. Thailand 23. UAE 24. Vietnam	1. Austria 2. Belgium 3. Croatia 4. France 5. Germany 6. Greece 7. Hungary 8. Italy 9. Ireland 10. Luxembourg 11. Malta 12. Netherlands 13. Norway 14. Poland 15. Portugal 16. Russia 17. Slovenia 18. Spain 19. Switzerland 20. Turkey 21. UK	1. Egypt 2. Morocco 3. Nigeria 4. South Africa 5. Tunisia 6. Kenya	1. Canada 2. Mexico 3. USA 4. Panama	1. Argentina 2. Brazil 3. Chile 4. Columbia 5. Peru

Table-2. Variables and notation

Factor	Variable	Notation
Access	Stock market capitalization	MCAP
Depth	Stock market capitalization to GDP	DEP
Efficiency	Stock market turnover ratio	TURNR
Stability	Stock price volatility	VOL

Source: World Bank database – 2018.

MCAP: Total value of all listed shares in the stock market

DEP: $\frac{\text{Stock market capitalization}}{\text{GDP}} * 100$

TURNR: $\frac{\text{Total value traded}}{\text{Stock market capitalization}}$

VOL: The mean value of 360-day volatility in stock prices

2.2. Data Standardization: Variables should be scaled, i.e., standardized before measuring the inter-observation dissimilarities. Especially, when variables are measured in different scales. The goal of

standardization is to make the variables comparable. Generally, variables are scaled to have Mean zero and Standard deviation one.

When scaling variables, the data can be transformed as:

$$\frac{x_i - \text{mean}_i}{SD_i} \quad 1$$

Where SD_i is standard deviation of x_i

2.3. Distance Measures:

The choice of distance measures is a critical step in clustering. These methods measure the similarity/dissimilarity between two elements (x, y) and this distance will influence the shape of the clusters. The distance measure we employed is Canberra Distance (Lance and Williams, 1967).

Canberra Distance measures is defined as:

$$d_{can}(x, y) = \sum_{i=1}^n \frac{|x_i - y_i|}{|x_i + y_i|} \quad 2$$

Where x and y are two vectors of length n

2.4. Determination of optimal number of clusters:

Determining the optimal number of clusters for a data set is a fundamental issue in clustering. There are some methods available in the literature namely, Elbow, Silhouette, and Gap statistic method. However, the user can specify the number of clusters according to his/her requirement. In this paper, we considered the Elbow method as it is widely used in for determining the optimal number of clusters (Kassambara, 2017).

2.4.1. Elbow method

The Elbow method considers the total Within Sum of Squares (WSS) as a function of the number of clusters. The optimal number of clusters can be obtained by computing any clustering algorithm for different values of k, for instance, varying k from 1 to 10, then for each k, Within Sum of Squares (WSS) is calculated and these values are plotted to find the optimal number by looking at the bend in the plot that indicates the appropriate number of clusters.

2.5. Methods for Clustering

2.5.1. K- Means Clustering

K-means clustering is one of the most widely used unsupervised machine-learning algorithm for classifying the given data set into a set of k clusters or groups, where k denotes the number of clusters pre-specified by the analyst. This algorithm classifies the data into multiple clusters such that the clusters have high intra-class similarity, i.e., the objects in the same cluster are as similar as possible and low inter-class similarity, i.e., objects from different clusters are as dissimilar as possible. In K-Means clustering, 'Centroid' represents each cluster, which corresponds to the mean of the points assigned to the cluster.

The clusters are defined in such a way that, the total intra-cluster variation or within-cluster-variation is minimized. Hartigan-Wong algorithm (1979), which defines the total within-cluster variation as the sum of squared Euclidean distances between the objects and the corresponding centroid is the standard algorithm used in this method. This is defined as:

$$W(C_k) = \sum_{x_i \in C_k} (x_i - \mu_k)^2 \quad 3$$

Where, x_i design a data point pertaining to the cluster C_k , and μ_k denotes the mean value of the objects assigned to the cluster C_k . Each observation, x_i is assigned to a particular cluster such that the Sum of Squares (SS) distance of the observation to their assigned cluster centers μ_k is a minimum.

The total within-cluster variation is defined as:

$$tot.withinss = \sum_{k=1}^k W(C_k) = \sum_{k=1}^k \sum_{x_i \in C_k} (x_i - \mu_k)^2 \quad 4$$

The ‘total within-cluster sum of square’ measures the compactness of the clustering, where the smaller value indicates high compactness and lesser variation. While using the k-means clustering, the first step is to indicate the number of clusters, k. The algorithm initiates the clustering by randomly selecting k objects from the dataset to serve as the initial centres for the clusters, which are called centroids or cluster means. Then, the objects would be assigned to the clusters, which is called ‘cluster assignment step,’ where objects are assigned to their closest centroid, where the closeness is defined by using the Euclidean distance between the object and the mean of the cluster. After assigning the clusters, the algorithm updates the cluster by computing the new mean value of each cluster, which is called ‘centroid update.’ All the objects are reassigned again using the updated cluster means. After the centres have been recalculated, each object is checked again to see if it might be closer to a different cluster. All the objects are reassigned again using the updated cluster means. The ‘cluster assignment’ and ‘centroid update’ steps are repeated iteratively until the cluster assignments stop changing or until convergence is achieved, i.e., the clusters formed in the current iteration are the same as those obtained in the previous iteration.

2.5.2. K-Medoids or Partitioning Around Medoids (PAM)

The K-Medoids or the PAM algorithm is a clustering approach similar to K-Means clustering for classifying the data set into k clusters or groups. In K-Medoids clustering, the term ‘Medoid’ refers to an object within the cluster for which the average dissimilarity between it and all the other objects of the cluster is minimal. Medoid corresponds to the most central point in the cluster. K-medoid algorithm is one of the robust alternatives to k-means clustering, and the algorithm is relatively less sensitive to outliers and noise because it uses medoids instead of means as cluster centres. The use of means in clustering is highly sensitive to outliers, and this can affect the assignment of observations to appropriate clusters.

The PAM algorithm is based on the exploration for k medoids or representative points among the objects of the data set. After finding a set of k medoids, clusters are formed by assigning each object to the nearest medoid. Then, the objective function is computed by swapping each selected medoid m and each non-medoid data point. This objective function corresponds to the sum of the dissimilarities of all the objects to their nearest medoids. The quality of the clustering can be improved by swapping selected objects (medoids) and non-selected objects (non-medoids). The ‘swap’ step is very crucial in k medoids method. The goal is to find k representative objects which minimize the sum of the dissimilarities of the observations to their closest representative object. The swap step is carried out, if the objective function can be reduced by interchanging a selected object with an unselected object. This step is continued until the objective function can no longer be decreased.

2.5.3. Agglomerative Clustering

The agglomerative clustering, also known as AGNES (Agglomerative Nesting) is one of the most widely used types of Hierarchical clustering, that is used to group the objects based on their similarity. The algorithm treats each object as a single cluster, and pairs of such clusters are merged successively until all the clusters have been merged into one big bluster that contains all the objects. The result of this algorithm is a tree-based representation of the objects called ‘Dendrogram.’ Agglomerative clustering works in a “bottom-up” manner, i.e., each object is initially considered as a single-element cluster (leaf),

and at each step of the algorithm, the two clusters that are the most similar are combined into a new bigger cluster (nodes). This procedure is iterated until all points are a member of just one single big cluster (root).

2.6. Cluster Validation Statistics

In this section, we describe the most widely used clustering validation indices. These indices are designed to evaluate the goodness-of-fit of clustering algorithms for the data set. The validation measures reflect the compactness and the separation of the cluster partitions. Where, the compactness measures the closeness among the objects in the same cluster, which generally measured based on distance measures such as the cluster-wise within average/median distances between observations. Separation measures how well separated a cluster from other clusters, which is measured by the distance between the centres of the clusters.

2.6.1. Connectivity:

It corresponds to what extent items are placed in the same cluster as their nearest neighbours in the data space. The connectivity has a value between 0 and infinity and should be minimized. Generally, most of the indices used for internal clustering validation combine compactness and separation measures as follow:

$$Index = \frac{\alpha * Separation}{\beta * Compactness} \quad 5$$

Where α and β are weights.

2.6.2. Silhouette coefficient

The Silhouette coefficient measures how well an observation is clustered, and it estimates the average distance between clusters. For each observation i , the silhouette width S_i is calculated calculating the average dissimilarity a_i between i and all other points of the cluster to which i belongs, then, for all other clusters C , to which i does not belong, calculate the average dissimilarity $d(i,C)$ of i to all observations of C . The smallest of these $d(i,C)$ is defined as $b_i = \min_C d(i,C)$. The value of b_i can be seen as the dissimilarity between i and its “neighbour” cluster, i.e., the nearest one to which it does not belong. Later, the Silhouette width of the observation i is defined by the formula:

$$S_i = \frac{b_i - a_i}{\max(a_i, b_i)} \quad 6$$

Where, a_i is the average dissimilarity between i and all other points of the cluster to which i belongs and b_i is the dissimilarity between i and its neighbouring cluster. The range of S_i is -1 to 1, where the value near to 1 indicating the values being well clustered, a small value around 0 indicates that the observations lie between two clusters and the observations with negative S_i indicate that the observations are probably placed in wrong cluster.

2.6.3. Dunn index

The Dunn index can be computed by calculating the distance between each of the objects on the cluster and the objects in the other clusters.

Dunn index can be defined as:

$$D = \frac{\min. separation}{\max. diameter} \quad 7$$

Where, min. separation is the minimum of pairwise distance as the inter-cluster separation and max. diameter is the maximal intra-cluster distance as the intra-cluster compactness.

Higher the Dunn Index, better the clustering results. Because, if the data set contains well-separated clusters, the diameter of the clusters is expected to be small and the distance between the clusters is expected to be large.

2.6.4. Stability measures:

Another version of validation measures available in the literature is called stability measures that evaluate the consistency of the results of clustering methods by comparing the clustering results after each column is removed at a time. These measures include The Average Proportion of Non-overlap (APN), The average distance between means (ADM) and the figure of merit (FOM). The APN measures the average proportion of observations that are not placed in the same cluster based on the entire data set and later with a single column removed. As the name suggests, the ADM measures the average distance between the centres of clusters for observations that are placed in the same cluster. The FOM measures the average intra-cluster variance of the deleted column. The range of APN, ADM, and FOM is from 0 to 1, where the smaller value indicates highly consistent clustering results.

3. Empirical results:

In this section, we present the empirical results of the study. First, in order to find the optimal number of clusters we employed the Elbow method and shown in figure-1. From this, we can observe that the optimal number of clusters (k) is four.

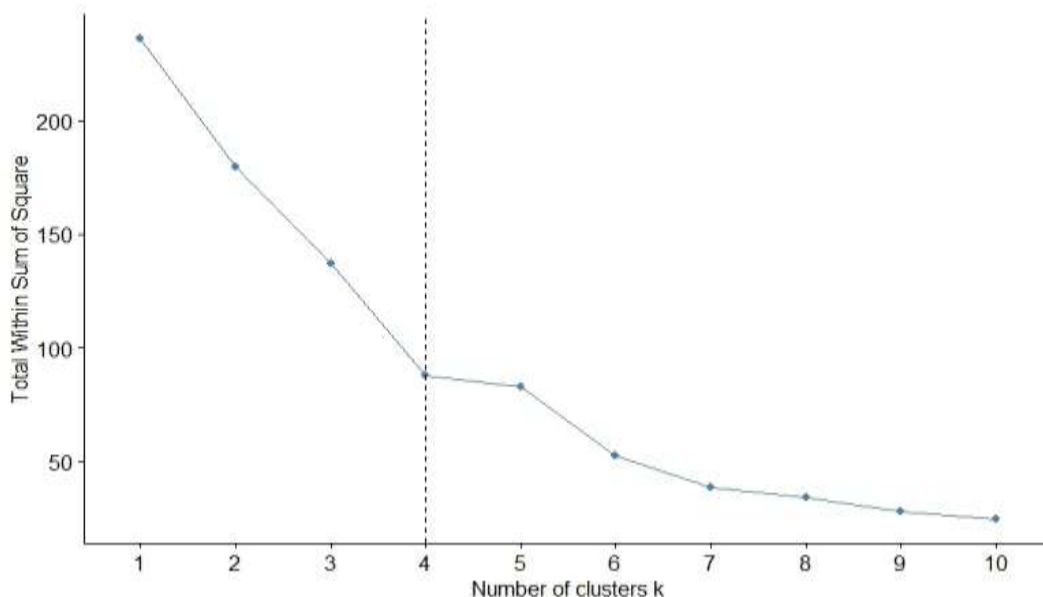


Figure -1. Determination of Number of Clusters

After choosing the optimal number of clusters, we considered three widely used classification techniques K-Means, Partitioning Around Medoids (PAM), and Hierarchical clustering methods for the classification of financial markets. In order to check the goodness-of-fit of these methods three the validation techniques Silhouette Method, Dunn Index and Connectivity Index were calculated and given in table 3.

Table-3. Validation Statistics of Clustering Methods

Clustering Method / Index	Connectivity	Dunn	Silhouette	Number of Clusters
Hierarchical	10.6448	0.2075	0.3863	4
K-Means	26.8397	0.0580	0.3198	4
PAM	39.6512	0.0389	0.1958	4

From table 3, we can see that all the three measures of validation suggest that the Hierarchical clustering method is suitable for classification. Also to see the consistency of clustering three stability ADM, APN and FOM were calculated and given table 4. It is found to be ADM and APN are minimum for Hierarchical clustering and FOM is for K-Means clustering method.

Table-4. Stability Measures of Clustering Methods

Stability Measure	Score	Clustering Method	Number of Clusters
ADM	0.4892224	Hierarchical	4
APN	0.1387879	Hierarchical	4
FOM	0.9519100	K - Means	4

From the validation and the stability measures, we have chosen the Hierarchical clustering method for classification of financial markets. The dendrogram and the cluster plot obtained and provided in figure -2 and figure-3 respectively.

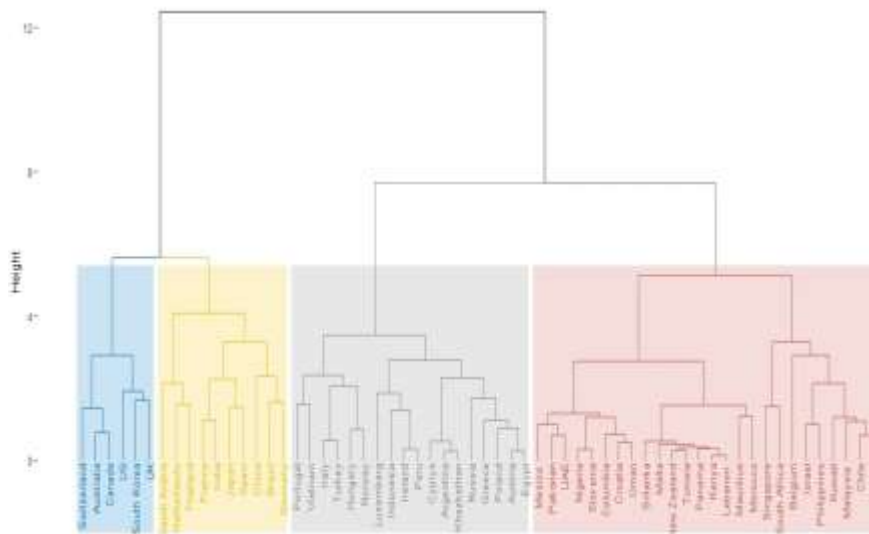


Figure -2. Cluster Dendrogram of Capital Markets

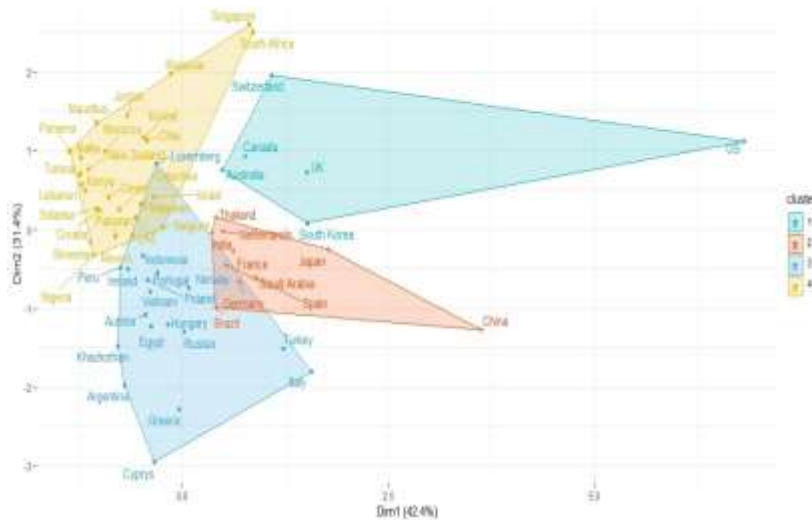


Figure -3. Cluster Plot of Capital Markets

From the cluster dendrogram in figure 2 and cluster plot in figure 3, we can observe that all the sixty markets considered for the study have been classified into four clusters, and the list is given in table 5. From these figures, it is evident that these markets are classified as four clusters. The first cluster consists of 6 countries, the second 10 countries, the third 18 countries, and the fourth cluster has 26 countries.

Table. 5. Classification of Capital Markets

Cluster 1 (6)	Cluster 2 (10)	Cluster 3 (18)	Cluster 4 (26)	
1. Australia	1. Brazil	1. Argentina	1. Belgium	19. Panama
2. Canada	2. China	2. Austria	2. Chile	20. Philippines
3. South Korea	3. France	3. Cyprus	3. Columbia	21. Sri Lanka
4. Switzerland	4. Germany	4. Egypt	4. Croatia	22. UAE
5. UK	5. India	5. Greece	5. Israel	23. Singapore
6. US	6. Japan	6. Hungary	6. Jordan	24. Slovenia
	7. Netherlands	7. Italy	7. Kenya	25. South Africa
	8. Saudi Arabia	8. Indonesia	8. Kuwait	26. Tunisia
	9. Spain	9. Ireland	9. Lebanon	
	10. Thailand	10. Kazakhstan	10. Malaysia	
		11. Luxemburg	11. Malta	
		12. Norway	12. Mauritius	
		13. Peru	13. Mexico	
		14. Poland	14. Morocco	
		15. Portugal	15. New Zealand	
		16. Russia	16. Nigeria	
		17. Turkey	17. Oman	
		18. Vietnam	18. Pakistan	

In the first cluster, we have, Australia, Canada, South Korea, Switzerland, UK, and the US. These markets can be considered as ‘**Highly Developed Markets.**’ The distance between the US capital markets and all other markets is very high, indicating US markets are way ahead of all other markets in terms of size, efficiency, depth.

In cluster 2, there are ten countries, namely, Brazil, China, France, Germany, India, Japan, Netherlands, Saudi Arabia, Spain, and Thailand. These markets can be considered as ‘**Advanced Emerging Markets.**’ There are 18 countries in cluster 3, namely, Argentina, Austria, Cyprus, Egypt, Greece, Hungary, Italy, Indonesia, Ireland, Kazakhstan, Luxemburg, Norway, Peru, Poland, Portugal, Russia, Turkey, and Vietnam. These capital markets can be considered as ‘**Secondary Emerging Markets.**’ In cluster 4, there are 26 countries, namely, Croatia, Chile, Columbia, Belgium, Israel, Jordan, Kenya, Kuwait, Lebanon, Malaysia, Malta, Mauritius, Mexico, Morocco, New Zealand, Nigeria, Oman, Pakistan, Panama, Philippines, Sri Lanka, UAE, Singapore, Slovenia, South Africa and Tunisia. These markets can be considered as ‘**Frontier and Under-Developed Markets.**’

Out of five countries from Africa considered for the study, four markets are lying in the cluster of frontier and underdeveloped markets. Past studies suggested that the underdeveloped markets would not help much in economic growth and it would be better to utilize the human and material resource in developing and improving the banking sector rather than the stock market (Singh, 2008). Also investing in emerging markets would fetch higher returns compared to developed markets (Kohers et al., 2006). Therefore, investors can maximize their returns by investing the markets, which lie in cluster 2 and cluster 3, i.e., advanced emerging markets and secondary emerging markets.

4. Summary and Conclusions:

With the liberalization and globalization of financial markets all over the world, the investors found new avenues to invest their surplus funds and diversify their portfolio. There are several studies, which revealed that investments in emerging markets would fetch higher returns compared to developed and underdeveloped markets. However, there is no clear classification of markets into different categories as developed, emerging and under-developed markets. Classification of capital markets based on financial attributes would help the investors and other players in the market in their decision-making process. Though there exist some classifications by international agencies, they have some limitations.

In this study, we used essential financial features such as access, depth, efficiency, and stability with the data available in open sources to classify the world major capital markets. This study employed widely used clustering techniques namely, K - Means, K- Medoids and Hierarchical Clustering methods in classification of markets. This study finds four clusters, and the number of markets in each cluster is six, ten, eighteen and twenty-six and termed as ‘Highly Developed Markets’ ‘Advanced Emerging Markets’ ‘Secondary Emerging Markets’ and ‘Frontier and Under-Developed Markets’ respectively.

References

- Alrgibi, G., Ariff, M., & Murray, L. (2010). What factors discriminate developed and emerging capital markets. *Applied economic letters*, 17 (3): 1293-1298.
- Beck, T. and Levine, R. (2004). Stock markets, banks and growth: Panel evidence. *Journal of Banking and Finance*, (3): 423–442
- Bruner, R. F., Li, W., Kritzman, M., Myrgren, S., & Page, S. (2008). Market integration in developed and emerging markets: Evidence from the CAPM. *Emerging Markets Review*, (2): 89-103.
- Christopoulos, D. K. and Tsionas, E. G., (2004). Financial development and economic growth: evidence from panel unit root and cointegration tests. *Journal of Development Economics*, 73 (1): 55-74.

- El-Wassal, K. A. (2013). The Development of Stock Markets: In Search of a Theory. *International Journal of Economics and Financial Issues*, 3 (3): 606-624.
- Fraley, C. & Raftery, A. E. (2002). Model-Based Clustering, Discriminant Analysis, and Density Estimation. *Journal of the American Statistical Association*, 97 (458): 611 - 631.
- Gupta, K. S. & Rao, R. P. (2014). An Empirical Analysis of Stock Markets and Economic Growth After Liberalization in India. *Finance India*, 28 (2): 499-506.
- Guptha, K. S. & Rao, R. P. (2018). The causal relationship between financial development and economic growth: an experience with BRICS economies. *Journal of Social and Economic Development*, 20 (2) 308–326.
- Hartigan, J. A. and Wong, M. A. (1979). Algorithm AS 136: A K-Means Clustering Algorithm. *Journal of the Royal Statistical Society Series C (Applied Statistics)*, 28 (1): 100-108.
- Ho, S.-Y., & Iyke, B. N. (2017). Determinants of stock market development: a review of the literature. *Studies in Economics and Finance*, 34 (1): 143-164.
- Kassambara, A. (2017). *Practical Guide to Cluster Analysis in R, Statistical Tools for High-Throughput Data Analysis (STHDA)*, London.
- Kohers, G., Kohers, N., & Kohers, T. (2006). The risk and return characteristics of developed and emerging stock markets: the recent evidence. *Applied Economics Letters*, 13 (11): 737-743.
- Krichene, H. (2016). Immature versus mature stock markets' properties: univariate and multivariate stylized facts analysis. Paper presented at Doctoral days MASE, 12 September, Carthege Tunisia, available at <https://bit.ly/2H2elp6> (accessed 20 November 2018).
- Lance, G. & Williams, W. (1967). Mixed-Data Classificatory Programs I - Agglomerative Systems. *Australian Computer Journal*, 1 (1): 15-20.
- Levine, R., (1997). Financial Development and Economic Growth: Views and Agenda. *Journal of Economic Literature*, 35 (2): 688-726.
- Mendes, B. V., & Martins, R. A. (2018). Determinants of stock market classifications. *Applied Economics Letters*, 25(17): 1244-1249.
- Neto, P. S., Cavalcanti, G. D., Madeiro, F., & Ferreira, T. A. (2013). An ideal gas approach to classify countries using. *Physica A*, 392: 177-183.
- Rajan, R. G. & Zingales, L. (1998). Financial Dependence and Growth. *The American Economic Review*, 88(3): 559-586.
- Singh, A. (2008). *Stock Markets in Low and Middle-Income Countries*. Working paper, Centre for Business Research, University of Cambridge, Cambridge,