

## Introducing a cluster based concept to portfolio selection from Colombo Stock Exchange (CSE)

M.S.M.S. Kumara<sup>1</sup>, U.P. Liyanage<sup>2</sup>

<sup>1,2</sup>Department of Statistics & Computer Science, Faculty of Science, University of Kelaniya, Sri Lanka  
<sanjayasamankumara92@gmail.com>

**Abstract**— *Portfolio management is a compelling concept in the financial sector for strategizing investments resulting maximum return with minimum risk. Normal procedure of selecting companies to the portfolios out of the registered companies in Colombo Stock Exchange (CSE) depends on the experience as well as the gut-feeling of the investor or the broker. Thereby, this selection depends on qualitative measures without much quantitative justification. This research objective is to introduce a quantitative criterion to select companies in an investment portfolio that result higher return compared to the selection made by qualitative measures, using clustering based algorithm. As per literature or the knowledge of collaborators, this technique was never used. Here, the daily share prices of 215 companies were collected for the period of five years, starting from 1<sup>st</sup> January 2012 and K-means clustering algorithm was used in clustering with Norm of the Variance of the Quarterly Share Prices (VQSP) as clustering variable. Based on the VQPS values, all companies were sorted, and the analysis was carried out by considering first 100%, 95%, 90%, 80%, 70% selected companies separately. There after the selected companies were clustered into 5 clusters and those clusters were used to build different portfolios by introducing a quantitative measure on the selection criterion. The return gained by the portfolios that were built-up based on company clusters was higher than that gained by the traditional selection criterion. Therefore, the research exhibits that the portfolio which are having higher return can be constructed by the selections from the clusters suitably.*

**Keywords**— **Portfolio Optimization, Clustering, Risk Measurements.**

### I. INTRODUCTION

The Colombo Stock Exchange (CSE) is the only stock exchange in Sri Lanka. It is one of the exchange markets in South Asia providing a fully automated trading platform. CSE has 296 companies representing 20

business sectors as at 3rd August 2015, with a market capitalization of Rs.3115.52 Billion. Normal procedure of selecting the portfolios from CSE is depended on the experience of the broker and the company reputation. So, often the selection process of the composition of the portfolio is qualitatively measured. Then the error could be high because there is no control for the measures. But taking a statistical approach by clustering the companies registered in CSE, businessman can select an optimum portfolio for the investment by giving a quantitative measure for the selection process.

Algorithms such as K-means, Self-Organizing Maps and Fuzzy C-means have been used for the company clustering purpose with reference to the literatures ((Kullmann, 2000), (Nanda, 2010), (Rashidi, 2007), (Zovko, 2007)). Here, K-means clustering algorithm has been used, since it has shown the best performance in stock data clustering. The daily share prices of the selected 215 companies registered in the CSE for the period of 1st of January 2012 to 30th December 2016 are considered in this study.

The portfolio selection process is traditionally done with the experience as well as the gut feeling of the broker and by considering other qualitative measures. Therefore, there is an impact of qualitative information on decision making, in selecting particular companies in portfolios. Towards minimizing the effect of quantitative information impact, the objectives of the study have been formulated as to recognize different classes(clusters) of companies where one can pick companies to form a portfolio and to build efficient portfolios by picking companies using those company classes (clusters).

### II. METHODOLOGY

Since the clustering of the companies is basically based on the risk level of each company, VQSP is defined as the risk measurement of the selected companies. Therefore, the data has been collected as mentioned in the introduction and each year has been grouped into four quarters (Table 1) and the VQSPs for twenty quarters in five-year period are calculated.

**Table 1. Quarter Periods**

Quarter	Time period
Q1	1st January – 31th March
Q2	1st April – 30th June
Q3	1st July – 30th September
Q4	1st October - 31st December

Vectors consisting of twenty VQSPs are built for each of the selected companies. Finally, the Euclidean norm of the vector is calculated and it is used as the risk measurement of each company to cluster the companies.

$$\|V\|_{Euc} = \sqrt{\sum_{i=1}^{20} (V_i)^2}$$

where;

- $V_i$  :- The variance of the quarter share prices.
- $V$  :- The vector of 20 variances of the quarterly share prices.
- $\|V\|_{Euc}$  :- The Euclidian Norm of VQSP.

VQSP is used instead of using the whole variance throughout the five-year period with the consideration of the stochastic nature of the stock prices.

This suggests that it highlights better reflection of the variance. Also the Euclidean norm of VQSP is used to select one dimensional clustering variable. As mentioned in the introduction, K – Means clustering algorithm is used to cluster the companies and an appropriate value for the number of clusters (K) is found using the distortion curve method.

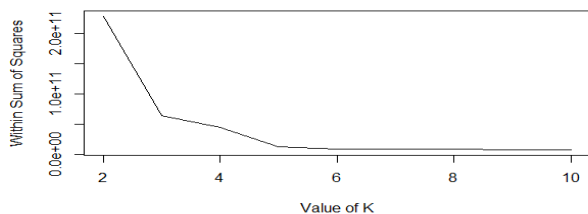


Figure 1. Distortion Curve (Elbow curve)

According to the distortion curve (Fig.1), the significant value for K is 5. The selected companies therefore are clustered into five clusters using the norm of VQSP.

### III. RESULTS

#### A. Clustering the selected companies using the norm of VQSP.

Based on the VQSP values, all companies are sorted and the analysis is carried out by considering five different percentages of companies under five categories (Table 2). In some companies, the VQPS values are very high and those values have affected the significance of the clustering. On behalf of that, five different categories are considered in the study.

Under the category 1, 215 companies have been clustered into 5 clusters and 204, 193, 172 and 150 companies are clustered into 5 clusters separately under the other four categories.

Table 2. Percentages of the norm of VQSP

Category	Percentage
1	100%
2	95%
3	90%
4	80%
5	70%

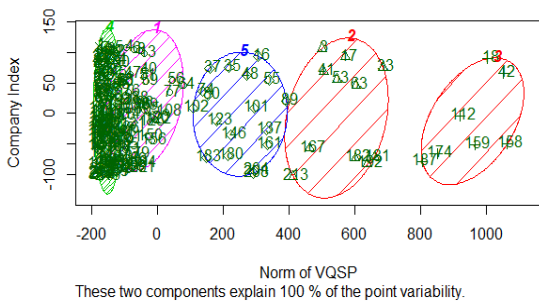
The companies that are removed, with high norm of VQSP are considered as one cluster with high risk. The coefficient of determination ( $R^2$  value of each category explains how well the companies are clustered based on the VQPS values. Therefore, the  $R^2$  - value is used as the measurement of significance of the clusters. The number of companies clustered, the number of companies removed, the number of companies in each cluster and the  $R^2$  - value for each category are tabulated in the table 3.

Table 3. Summary table of clustering using norm of VQSP

Category	No of companies clustered	Removed Companies	No of companies in each cluster	$R^2$ - Value
1	215	0	199,10,4,1,1	98.5%
2	204	11	158,24,5,11,6	96.6%
3	193	22	137,28,	96.8%

			12,7,9	
4	172	43	94,42, 18,11,7	97.3%
5	150	65	65,31, 25,15,14	95.1%

Cluster Segment Graph



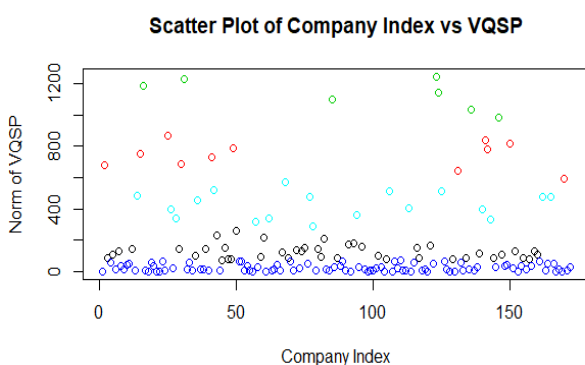
Although the  $\chi^2$  value of category 1 is higher than the other four categories, the number of companies in each cluster are not well classified. Out of the other four categories, category 4 has the highest  $\chi^2$  value 97.3%. Therefore, category 4, with 80% of the norm VQSP gives a significant clustering for the selected CSE companies.

Under category 4, 172 companies have been selected and other 43 companies, which have high norm of VQSP, are removed. The selected 172 companies are clustered into five clusters with the cluster sizes, 42, 11, 7, 18 and 94 (Table 4).

Table 4. Cluster centres and no of companies in each cluster for 80% of the norm of VQSP

Cluster No	Cluster means	No of companies
5	21.56244	94
1	124.25249	42
4	425.02798	18
2	743.03326	11
3	1132.01257	7

When we consider the scatter plot (Fig. 3) of company



index Vs the norm of VQSP of 80% companies, it shows that the companies are well clustered into 5 clusters. The five clusters that are separated, can be shown by cluster segment graph (Fig. 4). Further, it highlights how well the 5 clusters are separated. Furthermore, the removed 43 companies under category 4 are also considered as a one cluster (as cluster 6). Finally, 6 clusters are defined and those 6 clusters can be considered as, companies which have very low risk level up to companies which have the least risk level up to companies which have the highest risk level. The risk levels of the companies are varying in the order of cluster 5, cluster 1, cluster 4, cluster 2, cluster 3 and cluster 6 respectively.

Figure 2. Scatter plot of company index vs 80% of norm of VQSP

Figure 3. Cluster segments of the companies cluster under 80% of the norm of VQSP

Further, the companies in each cluster are analyzed by considering the twenty business sectors in CSE (Table 5).

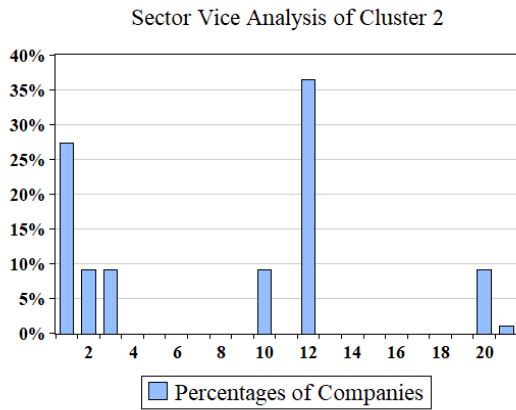
Table 5. The 20 business sectors in

Sector Name	Sector Index
Bank Finance And Insurance	1
Beverage Food And Tobacco	2
Chemicals And Pharmaceuticals	3
Constructions And Engineering	4
Diversified Holidngs	5
Footwear And Textile	6
Health Care	7
Hotels And Travels	8
Infromation Technology	9
Investment Trust	10
Land And Property	11
Manufacturing	12
Motors	13
Oil Plams	14
Plantations	15
Power And Energy	16
Services	17
Stores Supplies	18

Telecommunications	19
Trading	20

Figure 7. Sector Vice Analysis of Cluster 4

Figure 4. Sector Vice Analysis of Cluster 1



Sector Vice Analysis of Cluster 1

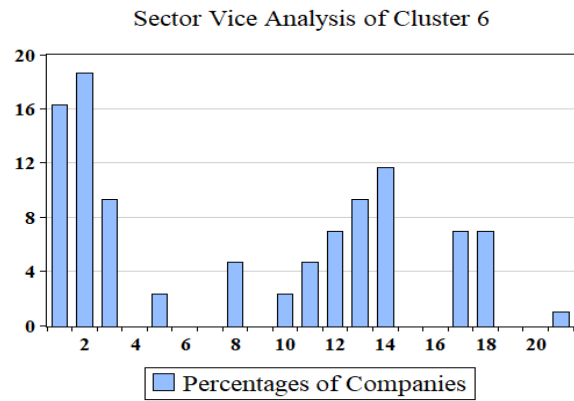


Figure 5. Sector Vice Analysis of Cluster 2

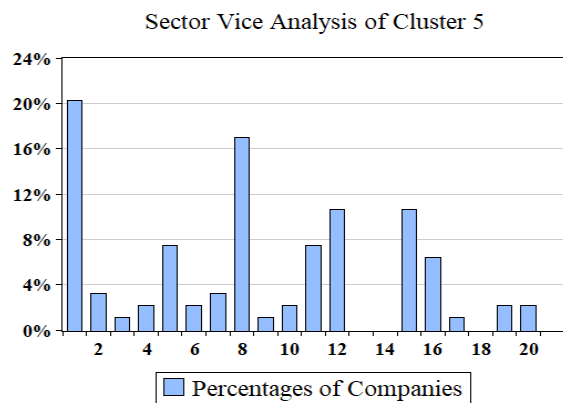


Figure 8. Sector Vice Analysis of Cluster 5

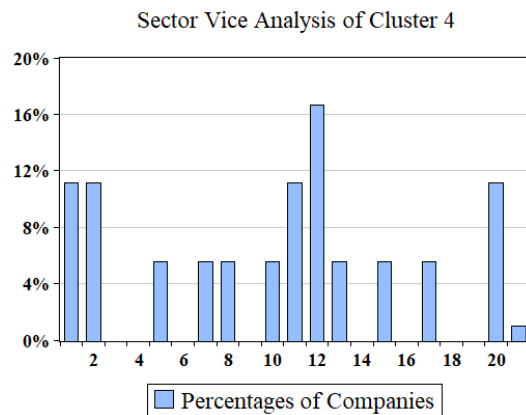


Figure 6. Sector Vice Analysis of Cluster 3

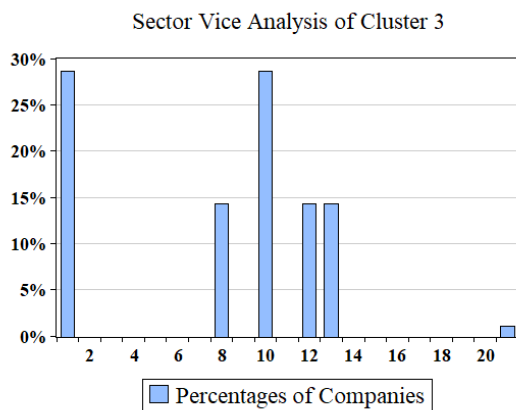


Figure 9. Sector Vice Analysis of Cluster 6

In cluster 1, highest percentage of 21% per each companies belong to sector 1 (“bank finance and insurance”) and sector 12 (“manufacturing”) (Fig. 4). When considering the cluster 2, 36% of companies belong to sector 12 (“manufacturing”) (Fig. 5). 29% per each companies in the cluster 3, belong to the sector 1 (“bank finance and insurance”) and sector 10 (“investment trust”) respectively (Fig. 6). Further in cluster 4, 17% of companies belong to sector 12 (“manufacturing”) (Fig. 7) and 20%, the highest percentage of companies in cluster

5, belong to sector 1 (“bank finance and insurance”) (Fig. 8). When the cluster 6, which is with the companies that are having high risk level is considered, 19% of companies belong to sector 2 (“beverage food and tobacco”) (Fig. 9).

**A. Portfolio selection**

When considering norm of VQSP under the category (4), 80% of the norm of VQSP gives more significant company clustering for the selected companies in CSE. Hence portfolio selection can be done with the clusters under the category 4, with 80% of the norm of VQSP. Four portfolios are built using the traditional selection (Table 6) and also another four portfolios are built by selecting companies from the six clusters (Table 7) which are defined under the 80% of the norm of VQSP.

**Table 6. Selected portfolios under traditional methods**

Portfolio 1	Portfolio 2	Portfolio 3	Portfolio 4
CFL	DIPD	SHOT	ABAN
IDL	REG	LITE	BRWN
CCS	TKYO	JINS	RENU
BBH	TRAN	BIL	STAF
ACL	KOTA	TYRE	CSD
TESS	LALU	SING	NTB

**Table 7. Selected portfolios under clustering methods**

Portfolio 5	Portfolio 6	Portfolio 7	Portfolio 8
CCS	IDL	HUNT	PARA
LIOC	BOPL	KDL	BALA
KOTA	COCO	KAPI	HUNA
NDB	SIL	UAL	CTHR
BRWN	CERA	SINI	WAPO
CFI	REG	SMOT	SFS

The profit or the loss is calculated for a unit share for the 1st quarter of 2017 from 2nd January to 31st March (Table 8). It highlights a significant different between the two selection methods. 75% of the portfolios selected under traditional selection shows a loss whereas a profit for the investments can be obtained by the 100% of the

portfolios selected under clustering method. Hence it reveals that the clusters help to build efficient portfolios.

**Table 8. Profit or Loss of the selected portfolios**

Portfolio	Profit/Loss (%)
1	7%
2	-14%
3	-8%
4	-1%
5	4%
6	3%
7	1%
8	8%

**CONCLUSIONS**

Norm of the variances of the quarterly share prices (VQSP) reflects the risk of the registered companies in CSE well. Additionally, it can be used as an accurate clustering variable to cluster the registered companies in CSE based on the risk levels. K-means clustering algorithm gives significant clusters under 80% norm of VQSP with 97.3% of  $\chi^2$  value. Finally, six clusters are defined according to the risk levels and the company compositions in each cluster are identified. Thereby, the investor can get a better insight about the companies. Furthermore, the standard principle which is used to build portfolios is maximize the profit while minimizing the risk. When there are different company compositions with different risk levels, the investor can manage his selection rather than the traditional method. The sample portfolios selected from the defined clusters give more profit than the portfolios selected using the traditional methods.

Additionally, the compositions of the clusters are considered under the 20 business sectors of CSE. When the cluster having low risk level companies is considered, high percentage of companies belong to the “bank finance and insurance” sector while high percentage of the companies of the cluster, which has a very high risk level companies, belong to “beverage food and tobacco” sector. Furthermore, it shows the composition of the companies according to business sectors are varying with the risk levels of the companies.

#### ACKNOWLEDGEMENT

I would like to express my sincere gratitude to Colombo Stock Exchange (CSE), for providing stock data. Also I would like to extend my gratitude to all the lecturers and temporary staff of the Department of Statistics & Computer Science for their immense support.

#### REFERENCES

Anon., n.d. *Colombo Stock Exchange*. [Online] Available at: <https://www.cse.lk/home/cseDail> [Accessed 2017].

Kullmann, L. K. J. & M. R. N., 2000. Identification of clusters of companies in stock indices via Potts super-paramagnetic transitions. *Physica A: Statistical Mechanics and its Applications*, 287(3-4), pp. 412-419.

Nanda, S. R. M. B. & T. M. K., 2010. Clustering Indian stock market data for portfolio management. *Expert Systems with Applications*, 37(12), pp. 8793-8798.

Rashidi, P. A. M. & A. J., 2007. Modified k-means algorithm for clustering stock market companies. pp. 201-211.

Zovko, I. I. & F. J. D., 2007. Correlations and clustering in the trading of members of the London Stock Exchange. 965(1), pp. 287-299.

Zovko, I.I. and Farmer, J.D., 2007, December. Correlations and clustering in the trading of members of the London Stock Exchange. In *AIP Conference Proceedings* (Vol. 965, No. 1, pp. 287-299). AIP.