A CLUSTERING BASED APPROACH TO STUDY THE IMPACT OF COVID -19 PANDEMIC ON COLOMBO STOCK EXCHANGE

M.G.Avanthi Saumyamala

208603T

Degree of Master of Science in Financial Mathematics

Department of Mathematics
Faculty of Engineering

University of Moratuwa
Sri Lanka

May 2023

A CLUSTERING BASED APPROACH TO STUDY THE IMPACT OF COVID -19 PANDEMIC ON COLOMBO STOCK EXCHANGE

M.G.Avanthi Saumyamala

208603T

Dissertation submitted in partial fulfillment of the requirements for the degree
Degree of Master of Science in Financial Mathematics

Department of Mathematics
Faculty of Engineering

University of Moratuwa
Sri Lanka

May 2023

DECLARATION

I declare that this is my own work and this Dissertation does not incorporate without acknowledgement any material previously submitted for a Degree or Diploma in any other University or Institute of higher learning and to the best of my knowledge and belief it does not contain any material previously published or written by another person except where the acknowledgement is made in the text. I retain the right to use this content in whole or part in future works (such as articles or books).

Signature: Date:

The supervisor should certify the Dissertation with the following declaration.

The above candidate has carried out research for the Degree of Master of Science in Financial Mathematics Dissertation under my supervision. I confirm that the declaration made above by the student is true and correct.

Name of Supervisor: Dr.H.A.S.G.Dharmarathna

Signature of the Supervisor: Date:

DEDICATION

This dissertation is dedicated to my loving baby Aveen for missing Amma in great times from his precious childhood on behalf of my education.

ACKNOWLEDGEMENT

I am using this opportunity to express my gratitude to everyone who supported me directly and indirectly in completing this research successfully. Firstly, I would like to express my sincere gratitude to my supervisor Dr. H.A.S.G.Dharmarathna for sharing knowledge and guidance to conduct this research successfully. Beside my supervisor, I would like to thank the coordinator of MSc in Financial Mathematics degree, Dr.Miuran Dencil for the continuous support, kindness and friendly cooperation throughout the study period. I am also grateful for all my lecturers including head of the department Mr.Rohana Dissanayaka for the knowledge they disseminated. Further I am thankful to the Mr. Nishantha Hewavithana, head of the department of Research and Development in Colombo stock exchange for providing valuable dataset. Also I'm grateful for my friends, Sachini Karunaratna and Irushi Jayathunga for the support given throughout the study period. Last but not the least, I would not be able to make this a success without my family. I am forever grateful for my strength Amma & Thaththa, my loving husband & baby, Akka, Ayya, my husband's parents, Nangi and Malli for supporting me spiritually throughout the research and my entire life. All my achievements are dedicated for you.

ABSTRACT

Investor's main objective is to maximize the portfolio returns while minimizing the risk. Unsystematic risk inherent to the industry can be minimized by diversification and systematic risk is hard to avoid. Main reason for the systematic risk is economic recession and covid pandemic is the most recent recession affected Sri Lankan economy. The primary objective of this research is to study the impact of covid-19 pandemic on risk and return of all listed companies in Colombo stock exchange (CSE) using clustering-based approach.

For this study stock market data for 3-year period from 1st January 2019 to 31st December 2021 were used. The first dataset of All share price index (ASPI) which represent the behavior of all the listed companies was used to identify the different periods. Using the background study and performing structural breakpoint tests the entire period was divided in to 5 meaningful periods. The second data set with daily open and close price of 289 companies were used for the clustering. Expected return and variance of each company in each period were used as the inputs for the clustering. K means clustering was used to achieve the main objective of clustering. When clustering the basic idea was to achieve meaningful clustering based on risk return tradeoff to facilitate better investment decision. Then the composition of clusters from period to period were analyzed with relevant to industry classification.

The cluster analysis shows that first and second period in year 2019 before the covid pandemic can be divided in to 4 clusters. During covid first wave market can be clearly separated in to 4 clusters, during covid second wave 6 clusters and during the third wave 3 clusters. Some companies have performed unusually with high returns and high variance during covid first wave and third wave. Further by monitoring the shift of clusters from period to period set of companies were suggested as the high performing companies despite the covid pandemic for the Risk seeking investors. Another set of companies which obtain high returns during the pandemic by tolerating moderate level of risk were suggested for the Risk neutral investors. The remaining set of companies are suitable for the Risk Averse investors. Finally, the study concludes that there is an impact of the covid -19 pandemic for the number of listed entities in the stock market, composition of clusters and risk return level of individual companies at different stages of a Recession.

Keywords: K-means clustering, Structural breakpoints, Stock clustering, Covid-19

TABLE OF CONTENTS

Declaration of the Candidate & Supervisor i

Dedication ii

Acknowledgement iii

Abstract iv

Table of Contents v

List of Figures viii

List of Tables xi

List of Abbreviations xi

1 Introduction 1

1.1 Background of the Study 1

1.2 Problem Statement 2

1.3 Objectives 3

1.4 Significance of the study 3

1.5 Chapter Summary 4

2 Literature Review 5

2.1 Modern Portfolio Theory (MPT) and Diversification 5

2.2 Cluster Analysis of Stocks 5

2.3 Impact of Covid -19 pandemic on stock market 7

2.4 Structural Breakpoints 7

3 Methodology 8

3.1 Proposed Methodology 8

3.2 Data Collection 8

3.3 Identification of Clustering Variables 9

3.3.1 Holding period Return 9

3.3.2 Expected Return 9

3.3.3 Risk 10

3.4 Period Identification 10

3.4.1 By the Background Study 10
3.4.2 Structural Changes 13
3.5 Cluster Analysis 15
3.6 K- Mean Clustering 16
3.7 Identifying Optimal number of Clusters 17
3.8 Discriminant Analysis 22
3.9 Tools and Techniques 22
4 Data Cleaning and Processing 23
4.1 Summary of the Data set 23
4.2 Data Cleaning 24
4.3 Data Processing 24
4.3.1 Adjusting for stock splits and dividend issues 25
4.3.2 Calculating Log Returns 29
5 Preliminary Analysis 31
5.1 Time series Plot 31
5.2 Structural Break Tests 32
6 Advanced Analysis 42
6.1 Identification of Clustering Variables 42
6.2 Period 1 Clustering 44
6.2.1 Clustering with K=2 48
6.2.2 Clustering with K=4 52
6.3 Period 2 Clustering 56
6.3.1 Clustering with K=2 59
6.3.2 Clustering with K=4 61
6.4 Period 3 Clustering 65
6.4.1 Clustering with K=2 68
6.4.2 Clustering with K=4 71
6.5 Period 4 Clustering 76
6.5.1 Clustering with k=3 78
6.5.2 Clustering with k=6 80
6.6 Period 5 Clustering 82

6.6.1 Clustering with K=2 85
6.6.2 Clustering with k=3 87
7 Discusion and Conclusion 91
7.1 Discussion 91
7.2 Conclusion 93
7.3 Limitations 94
7.4 Future work 94
References 95
APPENDICES 96

LIST OF FIGURES

Figure	Description	Page
Figure 3.1	Methodology of the Research	8
Figure 3.2	Covid Report of Epidemiology Unit	11
Figure 3.3	Distribution of Covid confirmed cases in First wave	11
Figure 3.4	Distribution of Covid confirmed cases in Second wave	12
Figure 3.5	Distribution of Covid confirmed cases in Third wave	12
Figure 3.6	CSE All Share Price Index	13
Figure 3.7	Clustering Techniques	15
Figure 3.8	Basic Risk Matrix	18
Figure 3.9	Further classified Risk Matrix	19
Figure 3.10	Utility function	20
Figure 3.11	Risk Return Trade-off	21
Figure 3.12	Efficient Frontier Line	21
Figure 4.1	ASPI data set	23
Figure 4.2	Stock Price data set	24
Figure 4.3	Calculating Stock Split factor	25
Figure 4.4	Unusually High Stock Split Factor values	26
Figure 4.5	Stock Split of company ACL at 2nd march 2021	26
Figure 4.6	Stock Split Announcement	27
Figure 4.7	Dividend Issue Announcement	28
Figure 4.8	Adjusting Stock splits	28
Figure 4.9	Calculating Dividend adjustment factor	29
Figure 4.10	Adjusting the dividend issue	29
Figure 4.11	Calculating daily holding period return	30
Figure 5.1	Time series plot of ASPI	31
Figure 5.2	Time series plot of %change in ASPI	32
Figure 5.3	OLS Cumulative Sum	33
Figure 5.4	Output of OLS-CUSUM test	34
Figure 5.5	Moving Sum	35
Figure 5.6	Output of OLS based MOSUM Test	36
Figure 5.7	Output of the F Test	37
Figure 5.8	Output of F statistic	38
Figure 5.9	Output of the Bai and Perron Test	39
Figure 5.10	Partitioning ASPI	40
Figure 6.1	Output of Elbow method	44

Figure 6.2	Output of Average Silhouette method	45
Figure 6.3	Output of Gap statistic	46
Figure 6.4	Output of NbClust Function	47
Figure 6.5	Output of k=2 clustering	48
Figure 6.6	Cluster plot	49
Figure 6.7	LDA model output period 2 k=2	50
Figure 6.8	Model accuracy and confusion matrices	51
Figure 6.9	Output of k=4 clustering	52
Figure 6.10	Cluster plot of Period 1	53
Figure 6.11	LDA model Period 1 k=4	55
Figure 6.12	Model accuracy and confusion matrices	55
Figure 6.13	Output of Elbow method	56
Figure 6.14	Output of Average Silhouette method	57
Figure 6.15	Output of NbClust function	58
Figure 6.16	Output of Period 2 k=2 clustering	59
Figure 6.17	Cluster plot of Period 2 k=2	60
Figure 6.18	Output of period 2 k=4 clustering	61
Figure 6.19	Cluster plot of period 2 k=4	62
Figure 6.20	LDA model period 2	64
Figure 6.21	Model accuracy and confusion matrix	64
Figure 6.22	Output of Elbow method	65
Figure 6.23	Output of Average Silhouette method	66
Figure 6.24	Output of NbClust function	67
Figure 6.25	Output of period 3 k=2 clustering	68
Figure 6.26	Cluster plot	69
Figure 6.27	Output of NbClust function	70
Figure 6.28	Output of period 3 k=4 clustering	71
Figure 6.29	Cluster plot	72
Figure 6.30	Announcement of Industrial ASPH comp	74
Figure 6.31	LDA model of period 3	75
Figure 6.32	Model accuracy and confusion matrix	75
Figure 6.33	Output of Average Silhouette method	76
Figure 6.34	Output of NbClust function	77
Figure 6.35	Output of period 4 k=3 clustering	78
Figure 6.36	Cluster plot	79
Figure 6.37	Output of period 4 k=6 clustering	80
Figure 6.38	Cluster plot	81
Figure 6.39	Output of Average Silhouette method	83
Figure 6.40	Output of NbClust function	84
Figure 6.41	Output of period 5 k=2 clustering	85

Figure 6.42	Output of period 5 k=3 clustering	87
Figure 6.43	Cluster plot	88
Figure 6.44	LDA model period 5	90
Figure 6.45	Model accuracy and confusion matrices	90

LIST OF TABLES

Table	Description	Page
Table 4.1	Excluded Companies from the data set	24
Table 6.1	Excluded Companies in Different Periods	43
Table 6.2	Clustering result of Period 1 k=2	50
Table 6.3	Clustering result of Period 1	54
Table 6.4	Clustering result of Period 2	63
Table 6.5	Clustering Result of Period 3	73
Table 6.6	CLustering Result of Period 4	82
Table 6.7	Clustering Result of Period 5	89

LIST OF ABBREVIATIONS

Abbreviation	Description
ASPI	All share price index
CSE	Colombo stock exchange
GICS	Global Industry Classification Standard
LDA	Linear Discriminant Analysis
MPT	Modern Portfolio Theory