DEVELOPING A COMPOSITE INDEX TO CATEGORIZE MANUFACTURING SECTOR ENTERPRISES IN SRI LANKA BY USING PRINCIPAL COMPONENT ANALYSIS

Pinidiya Pathirage Nuwan Aruna Shantha Perera

(168810D)

Degree of Master of Science

Department of Mathematics

University of Moratuwa
Sri Lanka

August 2020

DEVELOPING A COMPOSITE INDEX TO CATEGORIZE MANUFACTURING SECTOR ENTERPRISES IN SRI LANKA BY USING PRINCIPAL COMPONENT ANALYSIS

Pinidiya Pathirage Nuwan Aruna Shantha Perera

(168810D)

Thesis submitted in partial fulfilment of the requirements for the degree of Master of Science in Operational Research

Department of Mathematics

University of Moratuwa
Sri Lanka

August 2020

Declaration of the candidate & supervisor

I declare that this is my own work and this dissertation does not incorporate without acknowledgment 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 acknowledgment is made in the text.

Also, I hereby grant to the University of Moratuwa the non-exclusive right to reproduce and distribute my thesis/dissertation, in whole or in part in print, electronic or any other medium. I retain the right to use this content in whole or part in future works (such as articles or books).

…………………………..   ……………………….

Signature: Date:

P. P. N. A. S Perera

The above candidate has carried out research for the Masters Dissertation under my supervision.

……………………………   ………………………

Signature of the supervisor: Date :

Mr. A.R. Dissanayake

Senior Lecturer,

Department of Mathematics,

University of Moratuwa,

Sri Lanka.

Abstract

The industry sector, manufacturing industries play a prominent role in accomplishing economic growth in countries all over the world. Presently, Sri Lanka does not have a commonly accepted standard to categorize manufacturing enterprises. Different organisations use different definitions and there is no consistency between them. The most common criterion is the number of persons employed in the company. Though this is simple, it disregards important characteristics such as annual turnover, assets, energy consumption, etc. Hence, an establishment with fewer employees and large turnover categorized into small scale establishment and the number of employees significantly large but turnover not sufficient to large scale categories also mark as a large scale enterprise. Therefore policy-making stage on small-medium enterprises (SME) very difficult to identify enterprises categories exactly. So, Identifying manufacturing sector enterprises on a generally accepted criterion is a long-felt necessity to the country.

The main focus of this study is to develop a statistical method, to categorize manufacturing enterprises (5 or more persons engaged) in Sri Lanka. Developing a composite index and define the index boundaries to identify small, medium, and large manufacturing industries by considering the composite index mean value. One of the variable reduction methods called the principal component analysis (PCA) technique is used to define the index. Five reliable and significant variables were considered for the study. Data were collected from the Annual Survey of Industries 2017 (ASI) which is conducted by the Industries, Trade, Construction, and Services Division of the Department of Census and Statistics of Sri Lanka.

398 establishments out of the 1792 size sample were misclassified referred to two criteria (Turnover and Number of employees) as per the Ministry of Industry and Commerce (MOI) definition. Treating this misclassification is one of the main objectives of this study to come up with a solution.The analysis was addressed correctly to misclassified establishments in an accepted manner. Composite Index value less than or equal to zero (negative values) grouped as small scale and composite index value zero to 0.9983 categorized as meadium scale. Index values more than 0.9983 grouped as large scale establishments.

Eventually, by introducing cut-off index value, a newly entered establishment could also be categorized. Further cut-off point can be re-valued by changing base year when an Economic Census being done. The introduction of a consistent methodology to categorize which led to granting aid for the right establishment and paying taxes from the right establishment, which is very important for the development of the country.

Keywords: Composite Index, Principal Component Analysis, Dimensional Reduction, Categorization

Acknowledgments

It would not have been effectively achieved without several committed individuals who gave me great support and assistance, both professional and personal.

First of all, I would like to express my heartfelt gratitude to my supervisor Mr. Rohana Dissanayake, Senior Lecturer, Department of Mathematics, Faculty of Engineering, University of Moratuwa in Sri Lanka. Throughout the process, he has provided me with an abundance of assistance. I am indebted to him for his valuable support, guidance, comment, and advice. I would like to express my sincere gratitude to Dr. Pasan Edirweera, Senior Lecturer, Department of Mathematics, University of Moratuwa, and Course coordinator of MSc. in Operational Research, for his great support and guidance.

I am also thankful to all the lecturers who shared their valuable knowledge with us during the degree program and all the non-academic staff of the Department of Mathematics for their support.

I would also like to convey my special thanks to Mr. J. S. N. P Dharmawardena, Senior Statistician who is attached to the Department of Census and Statistics Sri Lanka for his valuable guidance, advice, and encouragement during research work and data dissemination unit staff of the Department of Census and Statistics Sri Lanka.

I am also grateful to my wife who has always supported me throughout my academic and career life. Her love and unfailing support has always inspired me in any overcoming obstacles of my research, especially during the final stage of my study.

Last but not least, I would like to thank my parents, brother, and sister for their endless motivation and support given me throughout the Masters degree and every person who has contributed directly or indirectly to the success of this thesis.

Table of Contents

List of Figures

List of Tables

List of Abbreviations

Abbreviations	Description
CI	Composite Index
CORM	Correlation Matrix
COVM	Covariance Matrix
CV	Coefficient of Variance
DCS	Department of Census and Statistics
EU	European Union
FA	Factor Analysis
GVA	Gross Value Added
ISIC	International Standard of Industrial Classification
KMO	Kaiser Meyer Olkin
MOI	Ministry of Industry and Commerce
MSME	Micro, Small & Medium Enterprises
PC	Principal Component
PCA	Principal Component Analysis
SD	Standard Deviation
SLSIC	Sri Lanka Standard of Industrial Classification
SLSIC 2D	Sri Lanka Standard of Industrial Classification 2 Digit
SME	Small Medium Enterprise
WB	World Bank

List of Appendices

CHAPTER 1

INTRODUCTION

1.1 Overview

Sri Lankan economy mainly depends on agricultural and non – agricultural sectors. According to the existing environment non-agricultural sector plays a vital role in between the two sectors. Under the non-agricultural sector, there are three main sub-sectors called Industry, Trade, and Services.

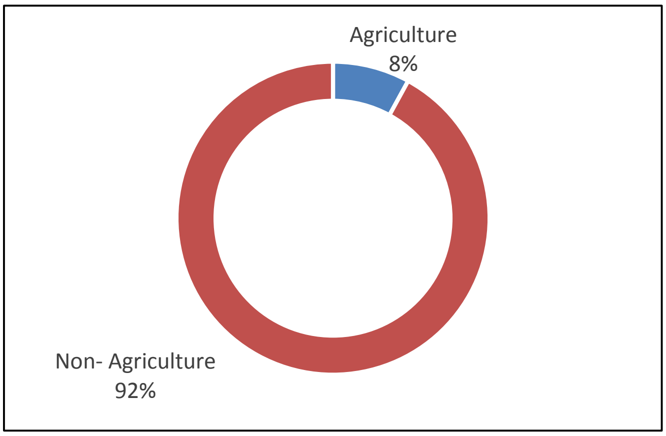

Figure 1.1 Gross Value Added (GVA) share by main sectors, current prices, 2019

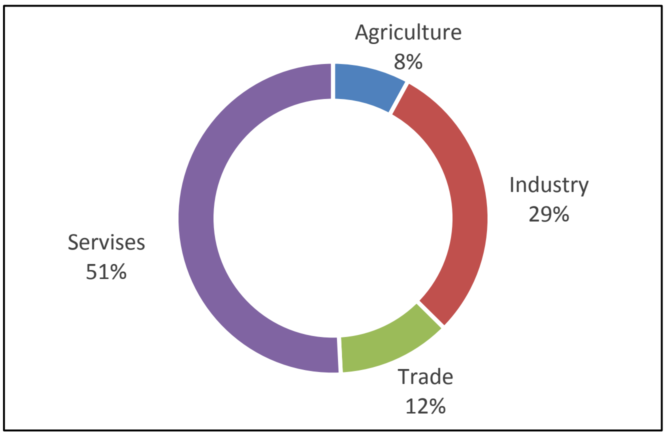

Figure 1.2 Gross Value Added (GVA) share by sub-sectors, current prices, 2019

(Source: National Accounts Estimates of Sri Lanka Gross Domestic Product (GDP) and other Macroeconomic indicators Base year: 2010)

Figure 1.1 shows that non – agricultural sector covered 92% of the gross value added in the whole economy. Meanwhile, the agriculture sector covered by 8%. Figure 1.2 shows the further division of sub-sectors of the non-agricultural sector.

Out of this non-agriculture sector share, industry sub-sector contribution was 29% to the gross value added.

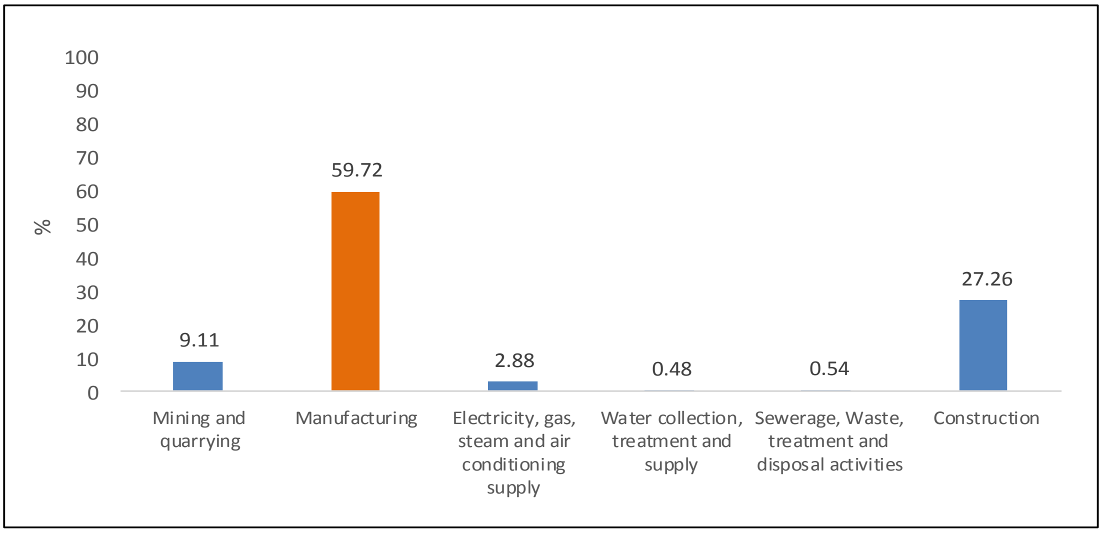

Figure 1.3 Economic sections wise contribution to the Gross Value Added (GVA) in industry sub-sector, current prices, 2019

The manufacturing section contribution to the industry sub-sector is 59.7% while all other sections show a relatively low percentage value. Therefore the manufacturing section is the key player in the industry sub-sector.

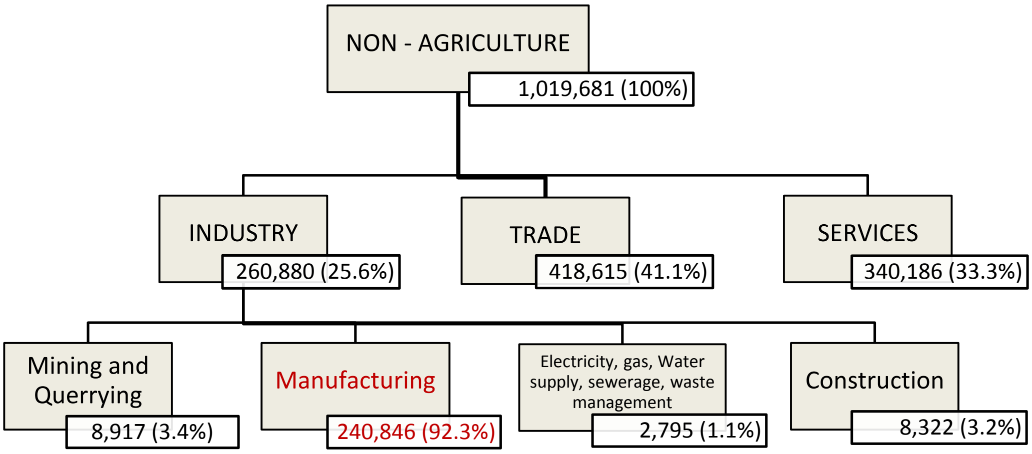

Figure 1.4 Distribution of the number of establishments

(Source: Listing Stage, Economic Census 2013/14)

The above flow chart clearly shows that the manufacturing section under the industry sub-sector has been dominated by covering 240,846 (92.3%) number of establishments of the total industry sub-sector. This is also trustworthy evidence to prove the importance of the manufacturing section in the country's economy. Because of these reasons studying the manufacturing section is opportune.

In this study, five or more persons engaged manufacturing establishments are considered as the study population, and relevant secondary data collected from the Department of Census and Statistics (DCS) in Sri Lanka.

1.2 Industry Sub-Sector

According to the International Standard Industrial Classification of All Economic Activities (ISIC), Rev.4 industry sub-sector mainly consists of five sections as follows.

- Mining and Quarrying
- Manufacturing
- Electricity, gas, steam and air conditioning supply
- Water supply, sewerage, waste management, and remediation activities
- Construction

Economic Census 2013/2014 listing stage report which was published by the DCS Sri Lanka reveals that 260,880 (25.6%) establishments are engaged in an industrial activity covered by the industry sub-sector. Those activities mainly consist of the above five sections.

1.2.1 Manufacturing Section

The manufacturing section refers the industries which involve the manufacturing and processing of items and accommodate in either the creation of new commodities or in value addition. The final products can either serve as a finished good for final consumption or as intermediate goods used in the production process.

The industry sub-sector main contributor is the manufacturing section and out of the total industry sub-sector establishments, 240,846 (92%) establishments are engaged in manufacturing activity.

According to the Annual Survey of Industries 2017 report total manufacturing establishments with which five or more persons engaged reported as 18,210, it's covered 86% out of the total industry sub-sector (five or more persons engaged establishments) except for the construction sector.

1.3 Importance of the Establishment Categorization

Every country has a responsibility to work on stabilizing its economy. In this process, the government has to support all institutions on different levels. Not only the government but are non-government agencies also engaging in this economy stabilizing process.

Ensuring and maintaining a definite definition of the categorization of establishments that contributes to the economy is significant to any country. But in Sri Lanka, different agencies use various definitions for industry categorization and there is no steadiness between them. The most common criterion to select the category of the establishment is the number of persons engaged in the establishment.

Though this is simple, it disregards important characteristics such as annual turnover, assets, energy consumption, etc. Hence, a company with fewer employees but which has a large turnover, categorizes into small scale enterprises by considering only the number of persons engaged. Likewise, a company with the number of employees significantly large but turnover is not sufficient to the level of a large scale category also mark as a large scale enterprise. Therefore, one can identify a conflict between these two cases. This study caters to the above problem using a statistical approach.

Such effective categorization helps a country to develop policies and implement them in target sectors in a correct manner. Similarly, it helps for the smooth distribution of subsidies and earning taxes as well as granting loans throughout the economy.

1.4 Current Practice and Issues

When considering locally and globally, most of the countries do not have proper manufacturing establishments categorizing methods that are based on a statistical approach.

Sri Lanka does not have generally accepted criteria for establishment categorization and,different agencies use different criteria based on their objectives and there is no consistency between them. Identifying SMEs on commonly acceptable criteria as a long-felt need of the country. (Economic Census 2013/14 Press Release DCS)

International and Local institutions use their definitions for their purposes. Because of this issue the government national level industry sector policymaking stage, they have faced a lot of difficulties when categorizing a particular establishment.

1.4.1 Current definitions for manufacturing sector categorization

1.4.1.1 Local Institutions

(a) <u>Ministry of Industry and Commerce (MOI)</u>

Table 1.1 Defining Establishment in Sri Lanka (Ministry of Industry and Commerce)

Sector	**Criteria**	**Large**	**Medium**	**Small**
Manufacturing Sector	Annual Turnover	More than Rs. Mn 750	Rs.Mn 251-750	Rs.Mn 16-250
	No. of Employees	More than 300	51-300	11-50

(Source: National Policy Framework for Small Medium Enterprise (SME) Development 2016)

In terms of their definition, the total number of employees and annual turnover are considered in defining the establishment category. In the event of an establishment falling under more than one category at the same time, the level of employment should be the deciding factor.

(b) Department of Census and Statistics (DCS), Sri Lanka

Table 1.2 Defining Establishment in Sri Lanka (Department of Census and Statistics)

Major Economic Sector	**Groups**	**Criteria (Number of Persons Engaged)**
Industry and Construction	Small	5 to 24
	Medium	25 to 199
	Large	200 and above

(Source: Report on Listing Stage, Economic Census 2013/14)

Initially, they identified three significant variables namely;

- Number of Persons Engaged
- Annual Turnover
- Assets

Out of these three variables, the number of persons engaged is shown to be the most trustworthy and consistent variable for defining establishment, from the data, collected at the listing stage of the Economic Census.
These two main organizations clearly use total employment as a deciding factor of the categorization of the establishments. And they never try to consider a systematic and scientific method to categorize establishments.

1.4.1.2 International Institutions

(i) European Union (EU)

Table 1.3 Defining Enterprises in the European Union

Company category	**Staff headcount**	**Financial ceilings**		
		Turnover	**or**	**Balance sheet total**
Large	> 250	> € 50 m		> € 43 m
Medium	< 250	≤ € 50 m		≤ € 43 m
Small	< 50	≤ € 10 m		≤ € 10 m

(Source: Evaluation of the SME definition final report, September 2012)

European Commission (EU) determines the criteria for defining establishment: number of employees, annual turnover, and annual balance sheet total. It is determined that meeting the criteria of the number of employees is compulsory while satisfying another from the two financial criteria is a choice of the enterprise.

(ii) <u>World Bank (WB)</u>

The World Bank uses three quantitative standards for defining enterprises: number of employees, total assets in U.S. dollars, and annual sales in U.S. dollars (IEG, 2008). An enterprise must encounter the quantitative criteria of the number of employees and at least one financial criterion to be categorized as a micro, small, or medium business.

Table 1.4 Defining SMEs in World Bank

Enterprise indicators	**Number of employees**	**Total assets**	**or**	**Total annual sales**
Large	> 300	> $15,000,000		> $15,000,000
Medium	> 50; ≤ 300	> $3,000,000; ≤ $15,000,000		> $3,000,000; ≤ $15,000,000
Small	> 10; ≤ 50	> $100,000; ≤ $3,000,000		> $100,000; ≤ $3,000,000
Micro	< 10	≤ $100,000		≤ $100,000

(Source: Independent Evaluation Group (IEG) 2008)

Above four definitions, their main threshold selection factor is the number of employees that belong to a particular establishment but excluding DCS Sri Lanka other three institutions are considering another optional factor to decide their company category.

If more than one factor gets into the company grouping process there is a chance to particular company figures are falling into two or more factors at the same time.

1.4.1.3 Other Countries

(i) India

Table 1.5 Definition of MSMEs in India

Manufacturing Enterprises - investment in Plant and Machinery	
Description	**USD($)**
Micro	up to $ 62,500
Small	above $ 62,500 & up to $ 1.25 million
Medium	above $ 1.25 million & up to $ 2.5 million

(As per Micro, Small & Medium Enterprises Development (MSMED) Act, 2006)

(Source: SME Chamber of India)

(ii) Malaysia

Table 1.6 Definition of MSMEs in Malaysia

Size	**Micro**		**Small**		**Medium**	
	Sales Turnover	**Employees**	**Sales Turnover**	**Employees**	**Sales Turnover**	**Employees**
Manufacturing	< RM 300,000	< 5 Employees	RM 300,000 to < 15 Million	5 to < 75 Employees	RM 15 million to ≤ 50 Million	75 to ≤ 200

(Source: SME Corp. Malaysia)

According to the Malasiyan definition, If an establishment satisfies either one criterion across the different sizes of operation, then the smaller size will be applicable. For example, if an establishment's sales turnover falls under micro but employment falls under small, the establishment will be deemed as a micro-enterprise.

(iii) Bangladesh

Table 1.7 Definition of SMEs in Bangladesh

Sector	**Small**		**Medium**	
	Fixed Asset other than Land and Building (Tk)	**Employed Manpower (not above)**	**Fixed Asset other than Land and Building (Tk)**	**Employed Manpower (not above)**
Industrial	50,000 - 1,50,00,000	50	1,50,00,000-20,00,00,000	150

(Source: SME & Special Programmes Department Bangladesh Bank Head Office, Dhaka)

1.5 Research Question

The considered sample was categorized into small, medium, and large categories concerning MOI currently use definition (Table 1.1). They use two criteria (turnover and number of employees) to define the scale of a particular company. This study proposes to cover only five or more persons engaged manufacturing establishments so that data do not represent establishments that have less than five persons engaged. Due to this limitation, micro sector establishments are not going to be categorized in this study. Table 1.8 shows how to distribute study sample establishments concerning the MOI definition.

Table 1.8 No. of establishment satisfying both criteria according to MOI definition

Category	**MOI Selection criteria**		**No. of Establishment**
	No. of Employee	**Annual Turnover**	
	Cutoff	**Cutoff**	
Small	Less than 50	Less than Rs. 250 Mn	1089
Medium	51-300	Rs. 251 - 750 Mn	166
Large	More than 300	More Than Rs 750 Mn	139
Total			**1394**

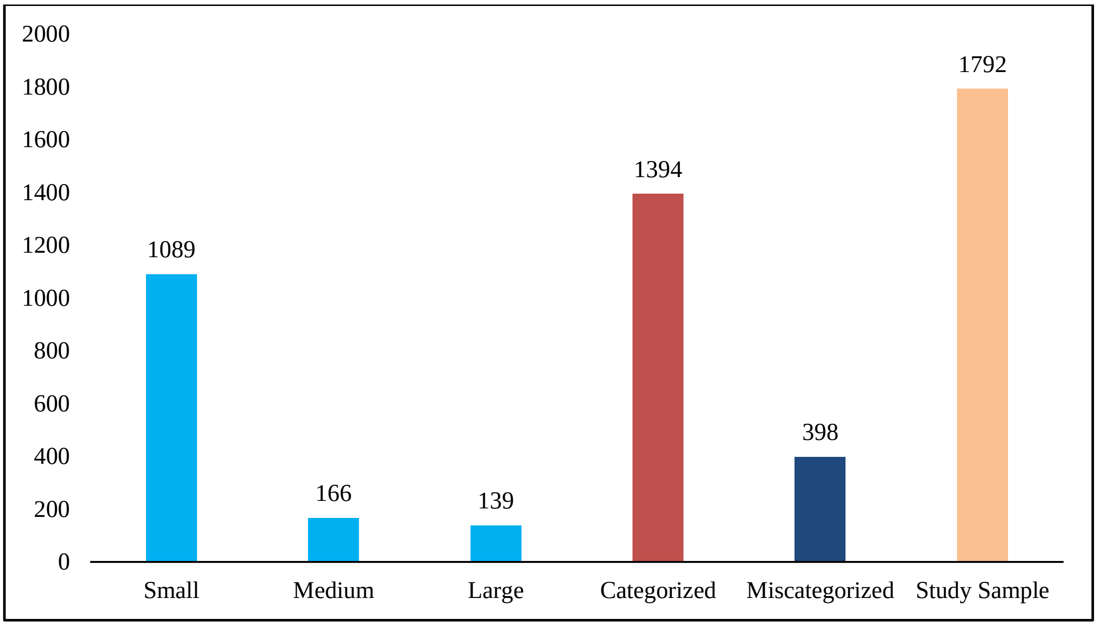

Figure 1.5 Distribution of No. of establishment category by MOI – definition

According to table 1.8, only 1,394 establishments are satisfied with both criteria to define by MOI out of 1792 establishments in the study sample and the rest of the 398 establishments can not be categorized by considering both criteria together. Because of that MOI has to use their second option (decide category by only considering the number of employees size) to categorize establishments that are falling into different categories according to their employee size and annual turnover. Table 1.9 clearly shows that how misclassified establishments are distributed into different categories according to their value of selection criteria.

Table 1.9 No. of establishment categorized and miscategorized by MOI definition

		Annual Turnover			Total	Miscategorized	%
		Small	**Medium**	**Large**			
No. of Employees	**Small**	1089	21	3	**1113**	24	*6.0*
	Medium	223	166	81	**470**	304	*76.4*
	Large	29	41	139	**209**	70	*17.6*
Total		**1341**	**228**	**223**	**1792**		
Miscategorized		252	62	84		398	
%		*63.3*	*15.6*	*21.1*			*22.2*

Table 1.9 describe miscategorized establishments by using two criteria (Turnover and Number of employees) as per the MOI definition. For example, column number one and raw number two, cell represent the number of establishments that are falling under small scale according to turnover and medium scale according to No. of employees.

In this table shaded diagonal cells are indicated No. of establishments correctly classified according to both criteria.

According to Table 1.9, 398 establishments are misclassified and they represent 22 percent of the total sample size. 304 establishments are classified into medium-scale according to no. of employees but that companies are classified into the small scale (223) and large scale (81) establishments when considering annual turnover as a selection criterion. In this kind of situation, MOI has selected all 304 establishments as a medium scale without concerning their annual turnover. Small and large scale establishments that are classified according to their no. of employees also have the same issue.

When policy developers are going to develop new policies for the manufacturing sector establishments they have to identify what are the companies they capture and their scale. But sometimes MOI classification misleads the selection of the scale.

For example, if going to implement financial aid packages to small scale companies one can see that according to above table 1.9 only 24 companies are eligible for that as they are on a small scale concerning no. of employee criteria. But it is not a fair decision because, when considering annual turnover criteria there are 252 establishments are on a small scale due to low turnover (though their no. of employees are higher). The said establishments need to have an opportunity to represent this eligible group. They missed the chance to be eligible for this tremendous opportunity, due to misclassification.

This study focuses on, treating this issue in the categorization of manufacturing establishments, via a statistical approach.

1.6 Research Objectives

The main objective of this study is to analyze these variables statistically and develop a composite index for each manufacturing establishment (5 or more persons engaged) in the selected sample. The calculated composite index base year is considered as 2016 and using a particular composite index value identifies a new methodology to categorize manufacturing establishments into small, medium, and large groups.

1.7 Limitation of the Study

This study only considers manufacturing establishments that have five or more persons engaged therefore no chance to cover the non-manufacturing industries, trade, and services sectors. Not only that, but there is also no opportunity to cover less than five persons engaged in manufacturing companies because of the unavailability of data. But most of the time persons engaged less than five, which meant that they have engaged in some household self-employment activities and are also categorized under the micro sector.

When going to analyze establishment-wise data some establishments show significantly high values of the considered variables. Further analyzing particular manufacturing activities one can identify that high values companies are the main manufacturer and they have maintained the monopoly of the particular manufacturing activity. This type of establishment can be identifying as an outlier.

Outliers are significantly affected to the analyzing process so that need to treat them before beginning the data analysis.

1.8 Thesis Outline

The rest of the thesis is organized as follows:

Chapter 2:

In this chapter; literature review, illustrate papers have published by other researchers on SME classification, papers base on Composite Index (CI), and the use of Principal Component Analysis (PCA) which helps for this research.

Chapter 3:

This chapter; methodology, defined the study variables that were used in this research. Further explanation of the data obtained, and the methodology used in data analysis chapter. Finally describe how to develop a composite index by using PCA. Moreover, the proposed establishment categorization also mentions in this chapter.

Chapter 4:

Hereby illustrate the data analysis part by using Statistical Package for the Social Sciences (SPSS version 25) in PCA. Furthermore, the findings will be discussed.

Chapter 5:

Eventually, the conclusion will be discussed in this chapter. Further, the opportunities opened up for further studies will be discussed. This comes out with results on categorizing manufacturing sector establishments in Sri Lanka.

CHAPTER 2

LITERATURE REVIEW

2.1 Literature related to Area of Study

“The classification of SMEs varies from country to country, region to region, sometimes within one nation. SMEs can be defined in terms of many factors such as the number of persons employed, capital invested, turnover, or a combination of the two or more. It is a fact that there is no single or unique definition in regard to SMEs.”(S. Vijayakumar, 2013)

Perera et al., (2017) has identified that Sri Lanka doesn’t have a generally established criterion to categorize SMEs and also different organizations use several criteria and there is no consistency among them. The most common criterion is the number of employees in the company. Though this is simple, it disregards an important characteristic such as annual turnover. Hence, a company with fewer employees and large turnover categories to small scale establishment. And also he mentions a commonly accepted criterion for SMEs is a long-felt requirement of the country (Perera et al., 2017).

Moreover, Perera et al. (2017) try to cluster establishments based on the Model-Based Clustering (EM) Algorithm. The Decision Tree algorithm is used to define the cluster margins for small, medium, and large industries, and introduced nine rules to classify industries.

Government or private sector institutions that are engaging in the policy development process need clear and deep knowledge of the sector and their boundaries that they are going to implement new policies. If there is any confusion in the categorization process, sometimes their attention goes to the wrong sectors that no need such kinds of policies or development strategy. Then valuable resources of the country have distributed in the wrong path. So that, effective establishment categorization methodology is very significant for any country for its economic policymaking.

2.2 Theories Related to Area of Study

2.2.1 Introduction

This study mainly concentrated on developing a statistical methodology to categorize the manufacturing sector establishments that five or more persons engaged in Sri Lanka based on their company's input and output performance. Since several variables are reflecting the establishment performance, the classification is proposed to be done by using Principal Component Analysis (PCA) through constructing a Composite Index (CI).

2.2.2 Principal Component Analysis (PCA)

The principal component analysis is concerned with explaining the variance-covariance structure of a set of variables through a few linear combinations of these variables.

There are two objectives on PCA,

1) Data reduction
2) Interpretation

"Principal component analysis (PCA) is a multivariate technique that analyses a data table in which observations are described by several inter-correlated quantitative dependent variables. Its goal is to extract the important information from the statistical data to represent it as a set of new orthogonal variables called principal components, and to display the pattern of similarity between the observations and of the variables as points in spot maps." (Abdi and Williams, 2010).

PCA can be achieved by decomposing a data covariance matrix by its value or decomposing a singular value, usually after standardizing the attribute data. Component scores and loadings are typically discussed in the results of a PCA (Shaw, 2003).

2.2.3 Composite Index (CI)

In recent times, composite indicators have gained astounding popularity in a variety of research areas. CI examines multi-dimensional attributes that a single variable does not adequately capture. CI should be based on the theoretical concept or definition that enables the collection, combination, and weighting of individual measures or variables in a way that represents the dimensions or structure of the phenomenon being evaluated.[1]

Composite indicators also seem easier for the general public to interpret than to identify similar patterns across several different indicators, and they have also been proven useful in benchmarking the performance of countries. (Saltelli, 2007)

According to Dharmawardena et al.(2015), in constructing composite indices, variables need to be minimized and the principal components analysis (PCA) is typically used for this objective. While PCA is carried out to resolve unit and value dependence issues after standardizing variables, variables drop their inherent variability. Two alternatives were tested to fix the issue.

- Converting variables by dividing their means.
- Meaningful alteration to original variables to convert them as unit less (Dharmawardena et al., 2015)

Laurent et al.(2010) argued that composite indicators can also produce deceptive policy messages if they are incorrectly designed or misinterpreted. Results could allow users (especially policymakers) to draw clear analyzes or politic conclusions. Ideally, the composite indicator should calculate multidimensional concepts that cannot be captured by a single indicator (Laurent et al., 2010).

[1] The Organization for Economic Co-operation and Development, (2004) The OECD-JRC Handbook on Practices for Developing Composite Indicators.

2.2.4 Other Related Papers

Principal Component Analysis (PCA) is a popular multivariate technique used to minimize dimensionality. With this methodology, the actual number of dimensions is supposed to be determined when accounting for almost all explained variance.

When conducting a PCA, the key reason for using CORM rather than COVM is that the outcomes of analyzes for various sets of random variables are more explicitly comparable, since the COVM-based PCA is sensitive to the units of measurement used for each variable. PCA, therefore, operates on standardized data, scaled by their standard deviation, in the CORM method. Then all variables will be scaled down with zero means and unit variances (Jolliffe, 2002).

On the other hand, Jolliffe (2002) suggests that if there are significant variations of the variances between the variables, such variables whose variances are greatest appear to occupy the first few PCs. In this case, the inherent uncertainty of PCAs with uniform data cannot be captured. It appears to be misleading to draw conclusions about the dominance of variance for real, non-standardized data (Jolliffe, 2002).

Jolliffe (2002) suggested that the COVM method could be completely suitable for the collection of variables with different variances, but measured on the same scale. Another drawback of PCs derived from CORM is that they have coefficients for standardized variables and are thus less straightforward to interpret directly. (Joliffe, 2002). There is also a need to create scale-independent composite indexes while retaining the inherent variability of the variables.

Standard PCA is based either on correlation matrix (CORM) or covariance matrix (COVM). When dependent on CORM, the scale dependence may be eliminated but the inherent variability cannot be maintained. On the other hand, when PCA is based on COVM, inherent variability can be preserved but it is not feasible to eliminate scale dependence. A solution to this issue suggests scaling each indicator by its mean. This leads to PCs, which are scale-independent while retaining the observed variability (Dharmawardena et al., 2017).

CHAPTER 3

METHODOLOGY

3.1 Introduction

This study mainly focuses on developing a statistical-based methodology to categorize five or more persons engaged, manufacturing establishments in Sri Lanka into small, medium, and large sectors. Individual establishment wise annual quantitative data that was collected by DCS, Annual Survey of Industries 2017 used for this classification and proposed to do by using principal component analysis and constructing a composite index (CI).

3.2 Data and Variables

3.2.1 Study Sample

DCS has used a stratified random sampling technique to select the Annual Survey of Industries sample from the population. This study has covered 1792 individual manufacturing establishments that were covered all manufacturing activities and all districts.

3.2.2 Sources of Data

The Department of Census and Statistics of Sri Lanka (DCS) continuously conduct establishment surveys annually, quarterly, and monthly. The Annual Survey of Industries (ASI) is the main survey that provides principal indicators of the industrial sector in Sri Lanka conducted by the division of Industry, Construction, Trade, and Services.

Annual Survey of Industries 2017[2] microdata was considering as the secondary data source of this study.

3.2.3 Variable selection

Analyzing the ASI 2017 questionnaire (see appendix A) identifies five significant variables that are direct can measure company performance. Table 4.1 shows selected variables and their measurement units and abbreviations.

---

[2] Reference year 01.01.2016 to 31.12.2016

Table 3.1 Variables considered

Sn. No.	Variable Name	Measured Unit	Variable Name Abbreviations
1	Number of employees Engaged	Numbers (No.)	Emp_Eng
2	Annual Output	Rupees (LKR)	Output
3	Opening Fixed Capital Assets	Rupees (LKR)	Open_FCA
4	Total Energy Consumption	Rupees (LKR)	Energy_Cons
5	Salary and Wages	Rupees (LKR)	Sal_Wages

3.2.3.1 Number of Employees Engaged

The manufacturing establishment has engaged a different kind of people. According to ASI 2017 questionnaire, DCS has categorized persons engaged into four groups named operatives, other employees, working proprietors, and active partners, and unpaid family workers.

Out of these groups, only operatives and other employees have been considered as employees who engaged in the establishment. These two groups entitle to received salaries and wages.

3.2.3.2 Annual Output

"The value of output was obtained from the value of shipments, the value of own account capital formations, and receipts of industrial and non-industrial services adjusted for changes in inventories of finished goods and change in work-in-progress during the reference period.[3] "

DCS has been used in the following equation to calculate the annual output value of the particular establishment.

3 Annual Survey of Industries – 2017 Final Report (Reference Period: 01.01.2016 to 31.12.2016)

Value of products moved out	+
changes in inventories of finished goods	+
Change in work-in-progress	+
Receipts from Industrial and non-industrial Services	+
Own account capital formation	=
Value of output	

3.2.3.3 Opening Fixed Capital Assets

An establishment has a different kind of fixed assets to use their day-to-day operations. As an example, land, buildings, plant and machinery, motor vehicles, etc. This variable consists of the total value of the assets at the beginning of the reference period.

3.2.3.4 Total Energy Consumption

The manufacturing company uses various sources to fulfill its energy requirement. Electricity and fuel are the main energy sources that can be captured from the DCS questionnaire. The total energy consumption value represents these two sources together.

3.2.3.5 Salary and Wages

According to International Recommendations for Industrial Statistics 2008, salary and wages have defined as follows.
"Wages and salaries are defined as all payments, whether in cash or in-kind, made by the employer during the reference period in connection with work done by all persons included in the count of employees, regardless of whether they are paid based on working time, output or piecework, or whether payments are made regularly or not."[4]

[4] International Recommendations for Industrial Statistics 2008 (Page No.: 65)

3.3 Methods of Analysis

3.3.1 Data preparation

Before start, the analyzing process needs to prepare the data, mainly concern about missing values and how to treat them. This study used a mean substitution technique for handling the missing data.

In a mean substitution, the mean value of the variable is used instead of the incomplete data value for the same variable. This helps researchers to use the data obtained in an incomplete dataset. The theoretical context of the mean substitution is that the mean is a rational approximation for a randomly chosen observation of the normal distribution. In addition, this method does not add new information, but only increases the sample size. (Malhotra, 1987)

When calculating the mean value to impute missing values consider manufacturing activity and different size classes of persons engaged.

Ex: if missing the value of energy consumption of the food manufacturing company that has 50 persons engaged, calculate the imputation value as follow,

Compute energy consumption mean value for different size classes of persons engaged (ex: 5-24, 25-99, 100-299 etc.) in the food manufacturing sector and identify relevant class matched to the particular company and get the mean value of that class to impute the missing value.

Plot the boxplot for every variable and identify outliers that are misleading the analysis and remove them from the data set and finally transformed the data into a new data set, dividing the original data point by its mean value as follows.

$$Y_{ij} = X_{ij} / \bar{X}_i$$

Yij = Transformed value of the i^{th} variable j^{th} value

X_{ij} Original value of the i^{th} variable j^{th} value

$\bar{X}_i$ is the mean of the i^{th} variable

The data were analyzed using statistical software SPSS version 25 and MS Excel has been used to graph generate.

3.3.2 Principal Component Analysis

3.3.2.1 Selection of Approach (CROM or COVM)

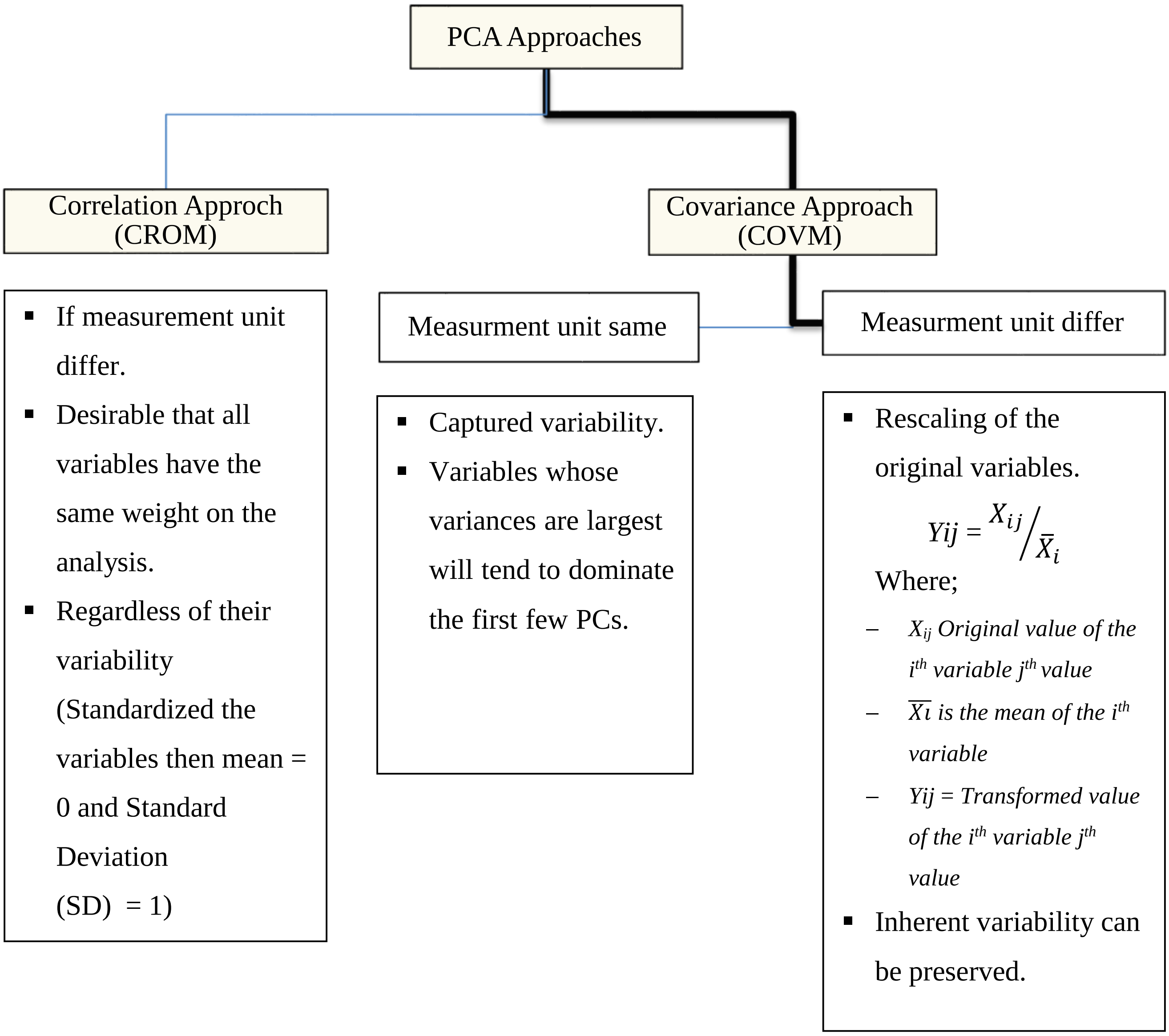

Figure 3.1 PCA Approach selection procedure

3.3.2.2 KMO and Bartlett's Test

Sample adequacy and the Sphericity assumptions are the prerequisites to perform PCA. The sample adequacy of PCA is measured by using Kaiser Meyer Olkin test (KMO test). In order to satisfy the sample adequacy, Kaiser Meyer Olkin test value should have minimum of 0.5.

The assumption of sphericity is measured using the Bartlett Sphericity Test. The null hypothesis for Bartlett’s Test of sphericity is the original correlation matrix is an identity matrix. In order to continue PCA, Bartlett’s Test null hypothesis should be rejected. Therefore, the significance value for Bartlett’s Test of Sphericity, should be less than significance level 0.05 which implies data do not produce an identity matrix and thus approximately multivariate normal.

3.3.3 Develop a Composite Index (CI)

The following equation can be used to construct a composite index for each establishment. Component scores were generated on the result of PCA analysis and variance proportion was calculated by using eigenvalues of the selected component after relevant rotation.

$$CI_j = \sum_{i=1}^{k} v_i C_{ij} \qquad 3.1$$

Where, i = 1, 2,…,k; k is the number of common factors,

j = 1, 2, …………..,n; n is the number of establishments.

C_{ij} is the Component Score of the i^{th} PC Component for j^{th} individual establishment

v_i is the variance proportion explained by the i^{th} component in the model.

i.e. $v_i = \frac{\lambda_i}{\sum_i^k \lambda_i}$, Where λ_i is the eigenvalue of the i^{th} component

3.4 Proposed Method to Categorization

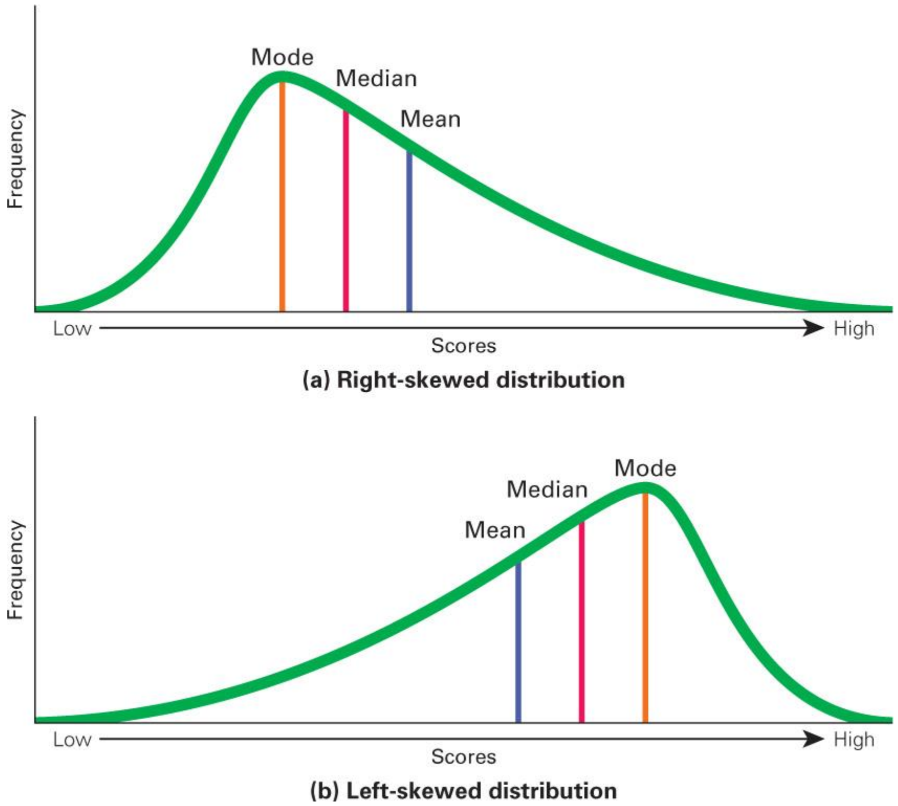

Figure 3.2 Skewness graphs with mean and median

Figure 3.2(a) and Figure 3.2(b), The X-axis represents scores low to high and the Y-axis represents frequencies. The mean value of the scores represents a blue line and the median value of the scores represents the red line.

Figure 3.2(a) shows that if the scores values are right-skewed distribute (positive skewed) then the scores mean value always greater than the scores median value because that if consider mean value as a cut-off point definitely, an area covered by less than the mean value acquired more than 50 percent of the sample units. If scores are left-skewed (negative skewed) then the scores mean value smaller than the median value as per figure 3.2(b) and also an area covered by less than the mean value cannot acquire at least half of the sample units.

Table 3.2 Manufacturing Establishment Categorization by DCS - 2013

Category (Scale)	Number of Establishments	%
Small	14,185	79.8
Medium	2,863	16.1
Large	720	4.1
Total	**17,768**	**100.0**

(Source: Report on Listing Stage, Economic Census 2013/14)

DCS is the National Statistical body that collects and disseminates actual data. DCS visits each organization to collect data in the listing stage in the economic census 2013. Thus the reliability of data collected by DCS is higher compared to other institutions. Hence their composition of the manufacturing establishment into respective categories can be used as a reliable estimate for this study.

Table 3.2 illustrate how DCS has grouped manufacturing establishments into different scales with their percentage by using economic census 2013/2014 listing stage data. Almost 80 percent has been covered by small scale establishments.

If composite indices show positively skewed distribution then the mean value of the CI can be considered as the best cut-off point and then the study sample can be divide into two different groups. Because of that number of establishments that are below the particular cut-off point represent more than 50 percent of the sample units and it can be identified as group 1. The other part let's define as group 2 that represents lesser than 50 percent of sample units.

After the division of two groups by using CI mean value, separately analyze CI values of group 2 again and calculate skewness and sketch the histogram. If it is also showing positive skewed features the previous scenario (method apply for identifying the first cut-off point) can be applied to recognize the second cut-off point that is separate medium and large-scale establishments.

According to table 3.2 out of the total medium and large categories, the medium category represents 79.9 percent. It indicates that more than 50 percent covered by medium-scale establishments in group 2. Due to this reason, the calculated mean value of group 2 can be selected as a second cut-off point. Then it opens up the way, clearly to identify medium-scale establishments and large scale establishments.

CHAPTER 4

DATA ANALYSIS AND RESULTS

4.1 Introduction

Manufacturing establishment categorization carried out by considering an establishment as a point of the data gathered. There are five significant variables were selected as the study variables by studying 1792 manufacturing establishment units covered by all districts in Sri Lanka.

4.2 Identification of Outliers

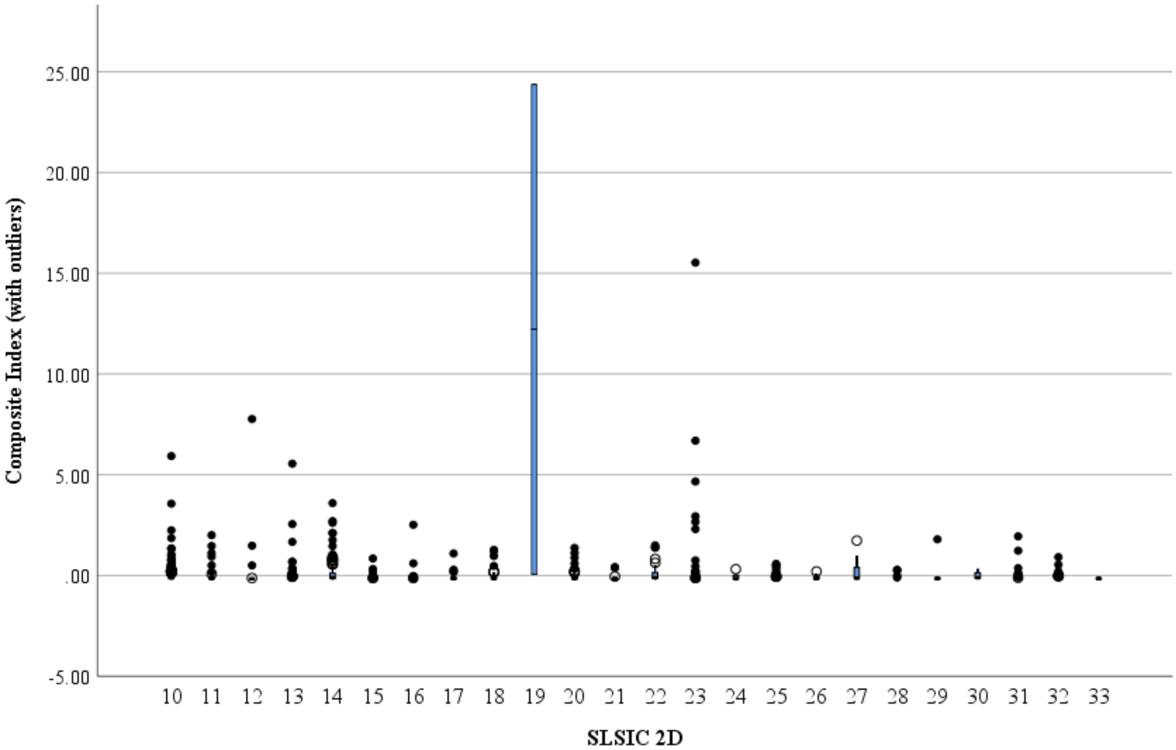

Figure 4.1 Boxplot of composite index value by SLSIC 2D

Figure 4.1 shows how individual establishment composite index values plot against their manufacturing activity. Here an establishment with regard to SLSIC 2D –"19-Manufacture of coke and refined petroleum products" activity and an establishment with regard to SLSIC 2D – "23-Manufacture of other non-metallic mineral products" activity have considerable deviation with other establishment's composite indices according to figure 4.1. It could predict that this deviation could be affected by extreme values of variables considered to build composite index. To identify this abnormality; analysis has to be done, by considering variable values separately with their respective manufacturing activity. Variables used for this study to build a composite index are Annual Output, Total Energy Consumption, Open Fixed Capital Assets, Employees Engaged, Salary and Wages.

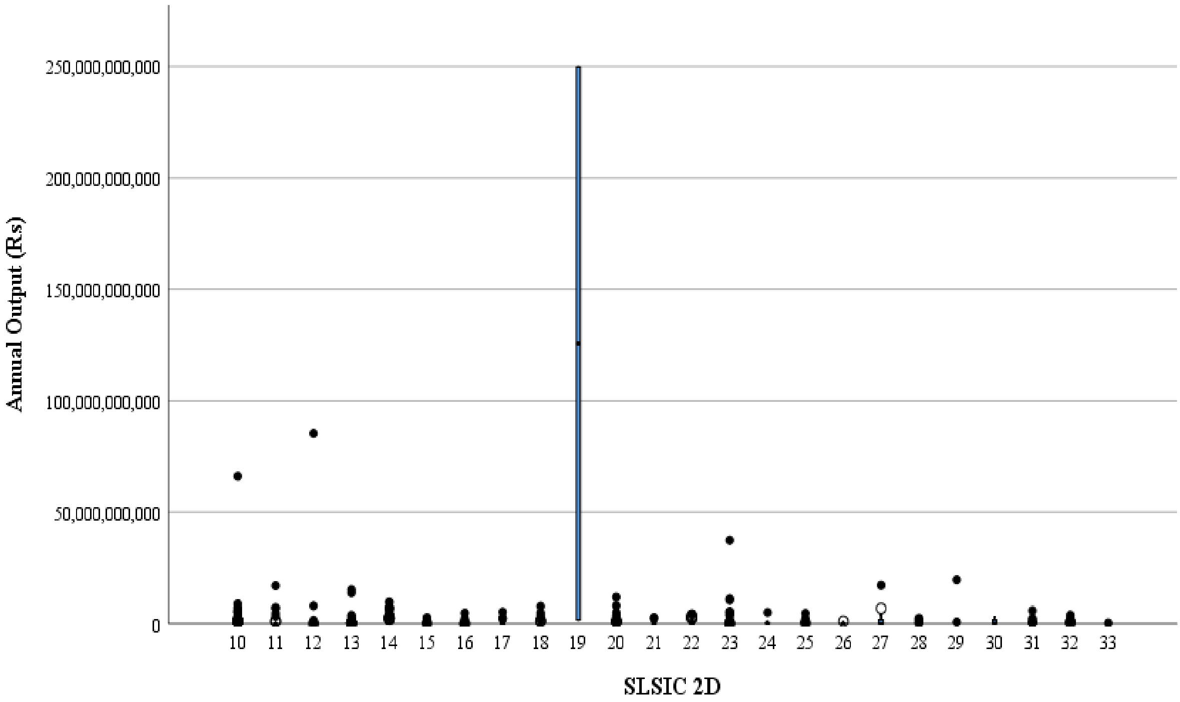

Figure 4.2 Boxplot of annual output by SLSIC 2D

Annual Output has been dominated by a single establishment with regard to SLSIC 2D – “19-Manufacture of coke and refined petroleum products” activity as shown in Figure 4.2. Deep analysis conducted by establishment level it can be identified that Similar establishment is dominating the composite index and annual output in Figure 4.1 and Figure 4.2 respectively.

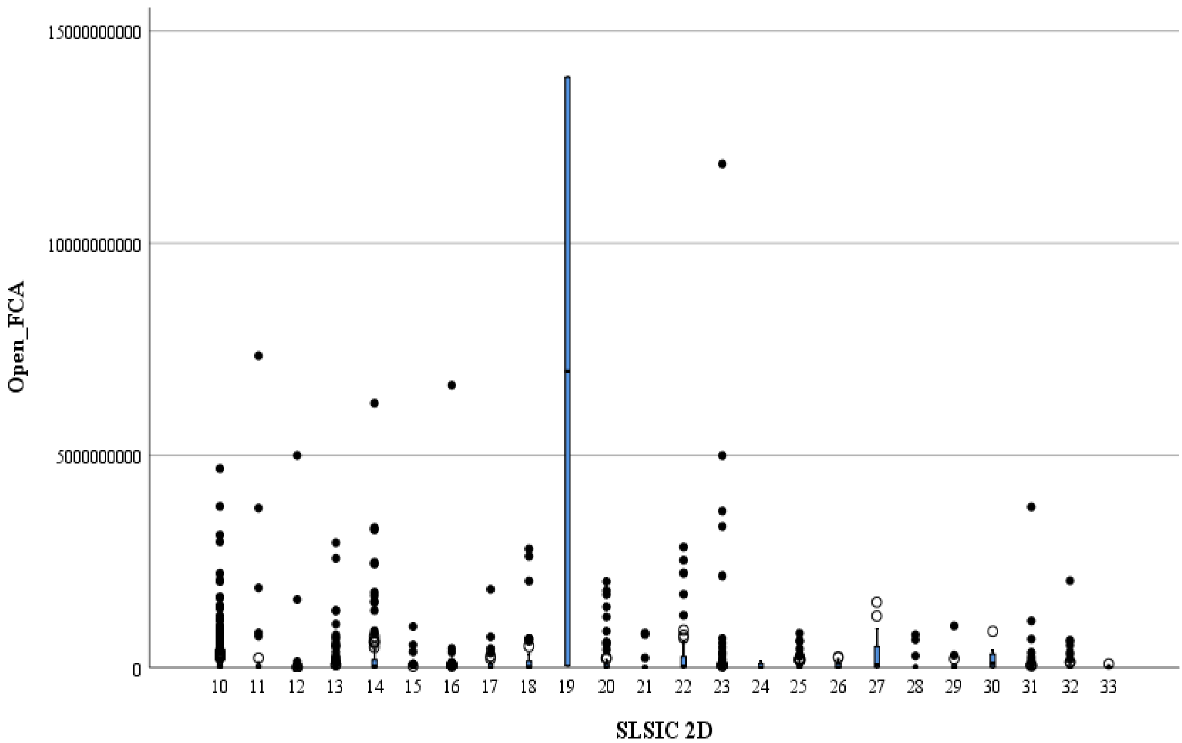

Figure 4.3 Boxplot of opening fixed capital assets by SLSIC 2D

Manufacturing activity-wise, the open fixed capital assets variable has been dominated by two establishments with their extreme values as shown in figure 4.3. The establishments show high composite index values concerning SLSIC 2D – "19 - Manufacture of coke and refined petroleum products" activity and SLSIC 2D – "23 - Manufacture of other non-metallic mineral products" activity in figure 4.1 are similar establishments show extreme values in open fixed capital assets variable.

Not only the composite index value of the particular establishment which has extreme value also other establishments indices values are significantly affected by these high values.

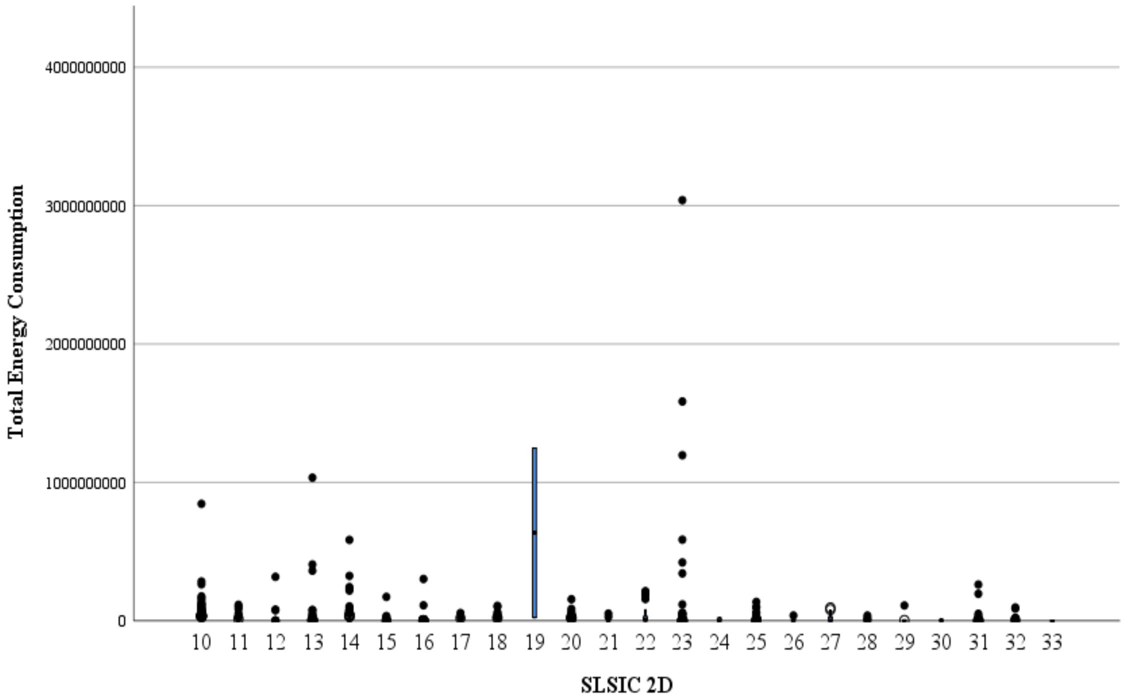

Figure 4.4 Boxplot of total energy consumption by SLSIC 2D

As described in figure 4.3 the same establishment in SLSIC 2D – “23 - Manufacture of other non-metallic mineral products” activity shows a deviation in figure 4.4 with regard to the Total Energy Consumption variable which significantly affecting to CI.

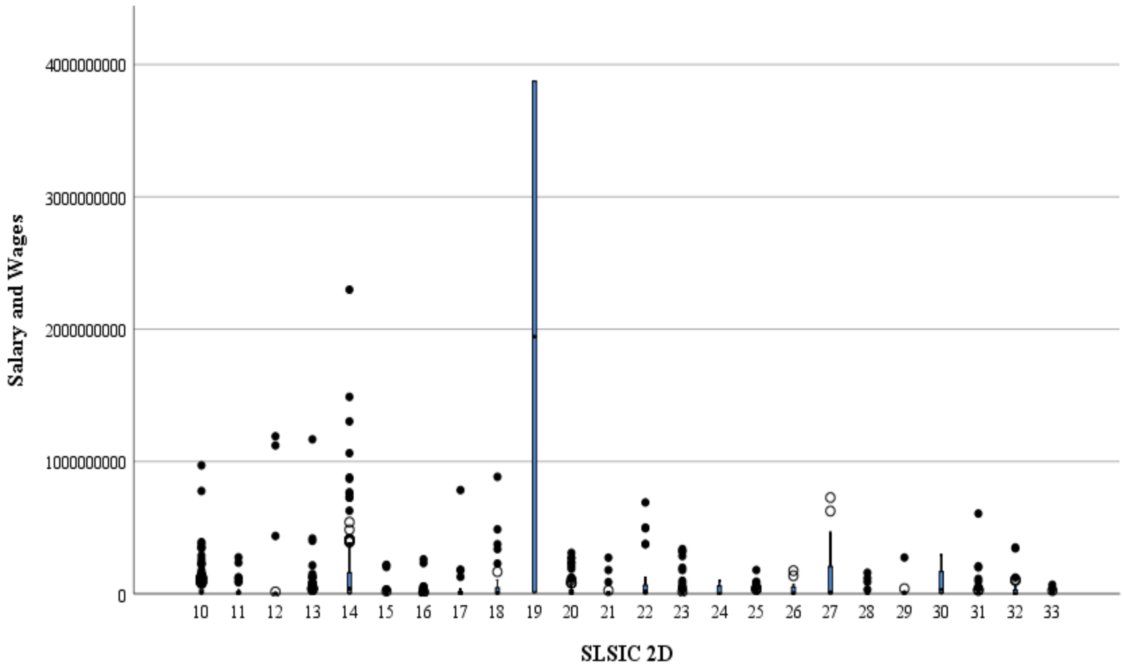

Figure 4.5 Boxplot of salary and wages by SLSIC 2D

Salary and Wages variable also indicate an extreme value concerning SLSIC 2D – "19 - Manufacture of coke and refined petroleum products" activity as shown in figure 4.5. The same establishment comes under SLSIC 2D – "19 - Manufacture of coke and refined petroleum products" activity in figure 4.1, 4.2, 4.3, and 4.5 have shown deviation here with regard to salary and wages.

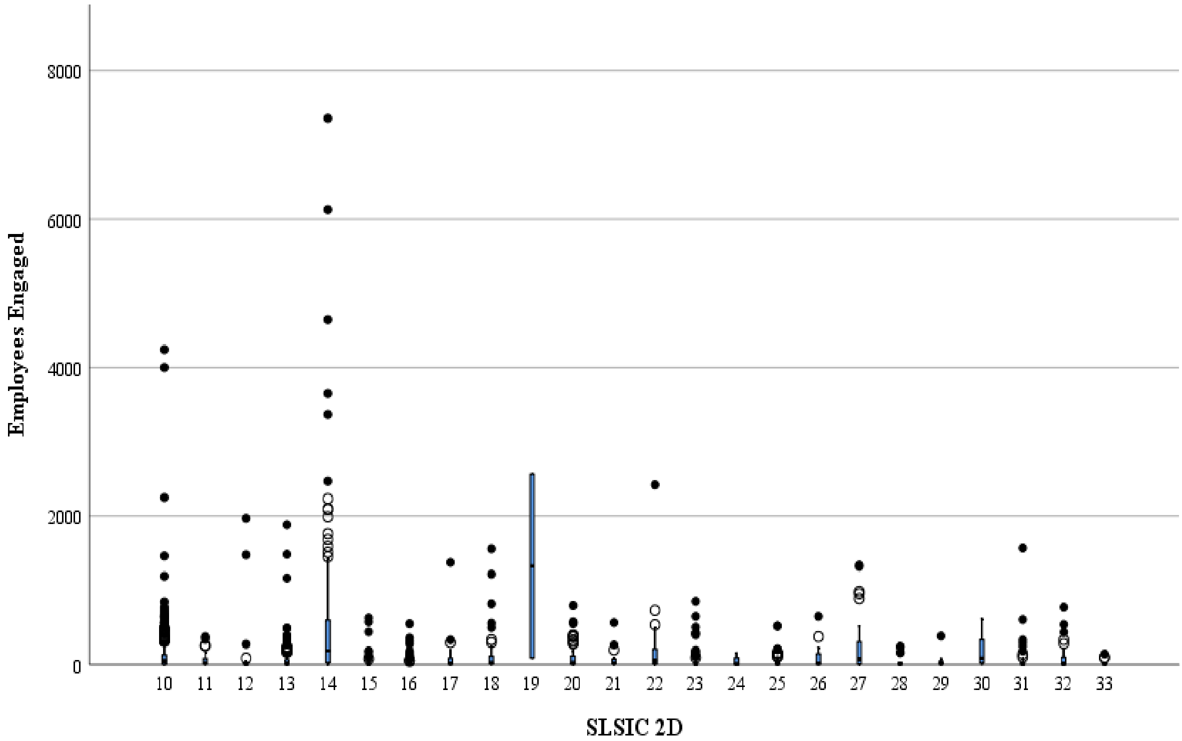

Figure 4.6 Boxplot of employees engaged by SLSIC 2D

Compare to the other variables Employees Engaged has not shown any extreme values with respect to their manufacturing activities.

When if the analysis was carried out with these extreme values the composite index values will be misled. This was recognized with some knowing large companies tend to degrade their scale. Hence analysis will be conducted by removing these outliers.

4.3 Descriptive analysis

4.3.1 Descriptive Analysis of Original Variables

Table 4.1 Descriptive statistics of the original variables

Variable	Mean	Standard Deviation (SD)	Coefficient of Variation (CV)
Output	445,372,921.54	1,396,529,888.61	3.136
Open_FCA	157,833,100.52	503,943,318.17	3.193
Sal_Wages	44,319,826.45	127,274,245.50	2.872
Energy_Cons	13,556,220.55	67,398,931.74	4.972
Emp_Eng	139.61	400.39	2.868

It is evident from the methodology chapter of Table 3.1 that the range of variables considered was in a different unit. And table 4.1 reveals that theyiwere with hugely scattered variability. That was not only due to the magnitude of the numbers but also due to the inherent property of the variables. One can see the output variable has the highest standard deviation (1,396,529,888.61) among those variables but its CV value (3.136) is the third-place among the other CV values. However, Open_FCA keeps in second-highest standard deviation (503,943,318.17), and its CV value (3.193) also in second place. Energy_Cons shows the highest CV value (4.972) but its standard deviation (67,398,931.74) in fourth place. Therefore, the inherent property of the variables could be captured by using the CV value in table 4.1, even though CVs are independent of the scales.

Table 4.2 Correlation matrix of the original variables

Original Variable	**Output**	**Energy_Cons**	**Open_FCA**	**Emp_Eng**	**Sal_Wages**
Output	**1**	0.512	0.628	0.459	0.584
Energy_Cons	0.512	**1**	0.513	0.358	0.43
Open_FCA	0.628	0.513	**1**	0.444	0.547
Emp_Eng	0.459	0.358	0.444	**1**	0.905
Sal_Wages	0.584	0.430	0.547	0.905	**1**

Correlations estimate the strength of the linear relationship between two (and only two) variables. Correlation coefficients range from minus 1.0 (a perfect negative correlation) to positive 1.0 (a perfect positive correlation). Table 4.2 shows the correlation between the variables. “Number of employees engaged” and “Salary and Wages” has a significantly high positive correlation (0.905). Other couples of variables show positive correlations and “Number of employees engaged” and “Total Energy Consumption” has a minimum correlation value of 0.358.

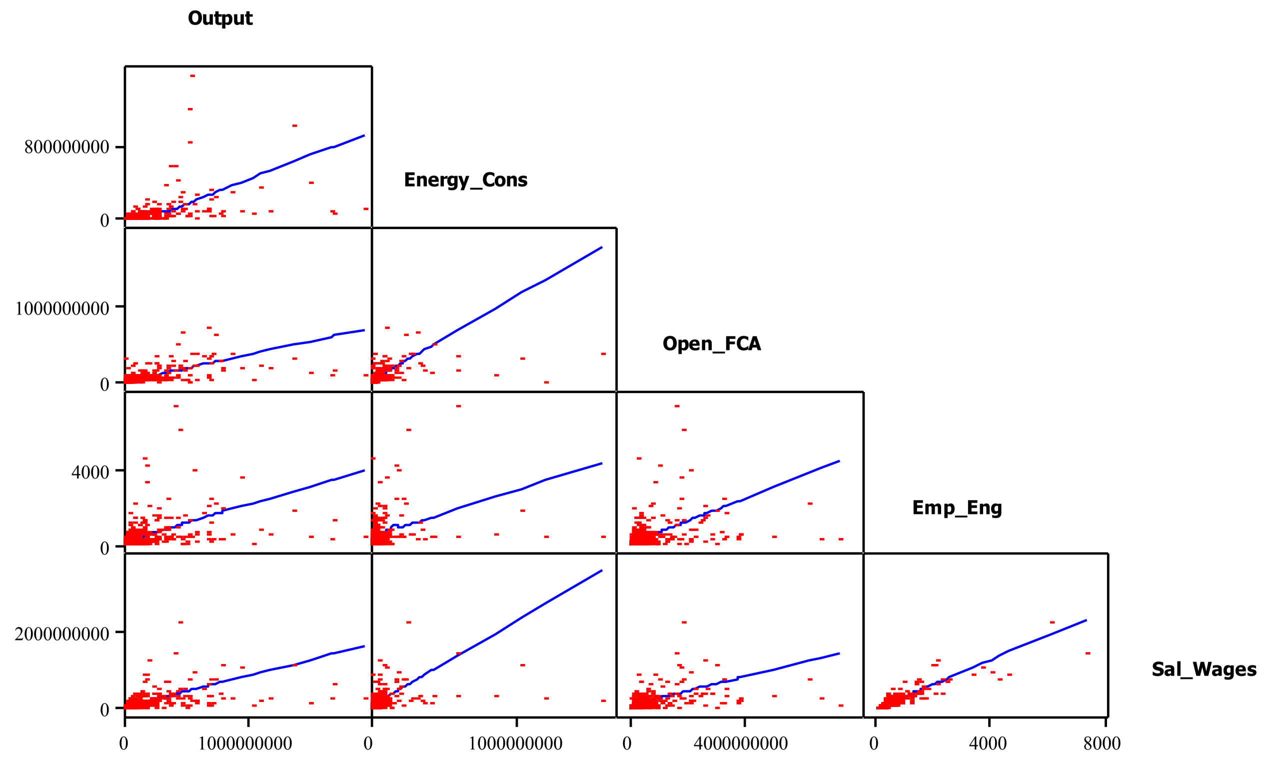

Figure 4.7 Multiple Scatter diagram for study variables

Table 4.3 Variance-covariance matrix

Original Variable	Output	Energy_Cons	Open_FCA	Emp_Eng	Sal_Wages
Output	1,950,295,729,789,820,000	48,177,045,353,635,000	442,139,343,353,316,000	256,474,340,505	103,739,646,649,739,000
Energy_Cons	48,177,045,353,635,000	4,542,616,000,178,480	17,415,394,779,893,900	9,661,273,264	3,691,850,218,355,780
Open_FCA	442,139,343,353,316,000	17,415,394,779,893,900	253,958,867,924,909,000	89,620,562,276	35,059,214,023,968,200
Emp_Eng	256,474,340,505	9,661,273,264	89,620,562,276	160,309	46,132,447,750
Sal_Wages	103,739,646,649,739,000	3,691,850,218,355,780	35,059,214,023,968,200	46,132,447,750	16,198,733,568,374,000

Table 4.3 variance-covariance matrix indicates the natural variability of the variables. But they are indifferent scales and different units of measurements. Because of that variability between each variable can never be compared. Therefore the variables, considered for the study were not suitable for applying PCA as they are with inherent variability, they are not scale-independent and they have different units of measurements.

4.3.2 Descriptive Analysis of Transformed Variables

Table 4.4 Descriptive statistics of the transformed variables

Transformed Variable	Mean	Standard Deviation (SD) (CV of Original Variables)
Output_T	1.000	3.136
Energy_Cons_T	1.000	4.972
Open_FCA_T	1.000	3.193
Emp_Eng_T	1.000	2.868
Sal_Wages_T	1.000	2.872

Table 4.4 shows, after the transformation process all variables mean values equal to one and standard deviation (SD) equal to the CV value of the particular original variable. Due to transformation, all the variables were independent of the measurement units as well as scales. Also, the characteristics of the inherent variation of the variables were well protected.

The highest variability among variables was recorded from the transformed variable of "Total Energy Consumption" (Standard Deviation, 4.972) followed by the transformed variable of "Opening Fixed Capital Assets" (Standard Deviation, 3.193) this indicate the manufacturing establishments in Sri Lanka show more large scale characteristics because of their Energy Consumption and Opening Fixed Capital Assets values.

Table 4.5 Variance-Covariance matrix of the transformed variables

Variable	**Output**	**Energy_Cons**	**Open_FCA**	**Emp_Eng**	**Sal_Wages**
Output	9.83	7.98	6.29	4.12	5.26
Energy_Cons	7.98	24.72	8.14	5.10	6.14
Open_FCA	6.29	8.14	10.19	4.07	5.01
Emp_Eng	4.12	5.10	4.07	8.22	7.46
Sal_Wages	5.26	6.14	5.01	7.46	8.25

Variance-Covariance matrix of the new set of transformed variables was shown in table 4.5. Values of each variable are comparable as they are unitless and scale independent.

4.4 Application of PCA

Table 4.6 KMO and Bartlett's Test

Kaiser-Meyer-Olkin Measure of Sampling Adequacy (MSA)		0.729
Bartlett's Test of Sphericity	Approx. Chi-Square	5657.63
	df	10
	Sig.	**0.000**

The sample adequacy of this study was measured by using Kaiser Meyer Olkin test and the test results are shown in Table 4.6. Kaiser Meyer Olkin test value for this study was received as 0. 729 as can be seen in Table 4.6. Kaiseri(1974) recommends a bare minimum of 0.5 value, the value between 0.5 and 0.7 are mediocre, value between 0.7 and 0.8 are good, value between 0.8 and 0.9 are great and value between 0.9 and above are superb (Hutcheson & Sofroniou, 1999). As the Kaiser Meyer Olkin test value is 0. 729 for this case, it is evident that the data satisfies the sample adequacy as stated by Kaiser.

As can be seen in Table 4.6, Bartlett’s Test of Sphericity received a P-value of 0.000. By default, SPSS reports P-values as 0.000 if the P-value is less than 0.001. The null hypothesis of Bartlett’s Test is defined as the original correlation matrix is an identity matrix. A significant value of less than 0.05 suggests that these data do not generate an identity matrix and thus approximately multivariate normal and acceptable for further analysis.(Pallant,2013) Since the sample adequacy and Sphericity assumptions are satisfied, PCA can be continued.

Table 4.7 PCs extraction using total variance explained

Component	**Eigenvalues**	**% of Variance**	**Cumulative % of Variance**
1	**38.81**	**63.39**	**63.39**
2	**11.75**	**19.19**	**82.58**
3	6.28	10.27	92.84
4	3.71	6.06	98.91
5	0.67	1.09	100.00

The second column of Table 4.7 gives the eigenvalue, or amount of variance in the original variables accounted for by each component. The first PC eigenvalue covered almost 63 percent of the total variance and the second component covered 19 percent out of the total. These together describe over 82 percent of the data's overall variability. This brings us to the conclusion that a two-component solution is likely to be appropriate.

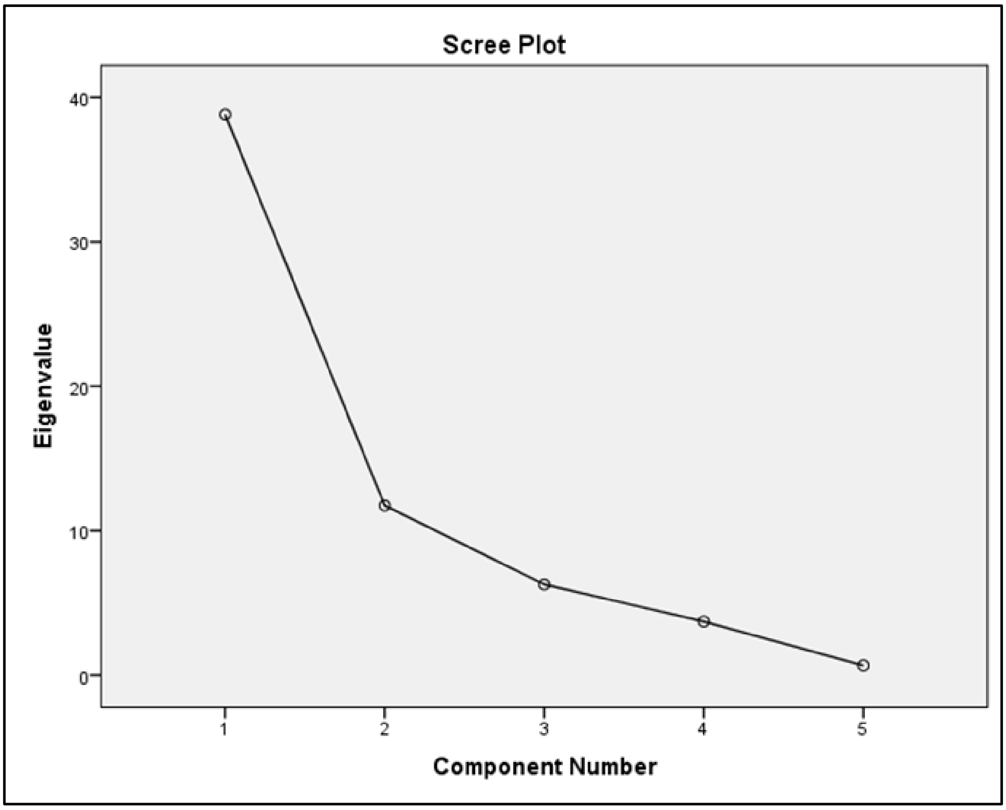

Figure 4.8 Scree plot

The scree plot in Figure 4.8 is a graph of the eigenvalues against all components. For deciding how many components to retain, the graph is useful. Where the curve begins to flatten is the point of concern. It can be seen that the curve, after component 2, starts to flatten.

Table 4.7 shows that the first two components have described a considerable amount of variance out of the total variance. Moreover above scree plot (Figure 4.8) also evidence that two components are enough to retain further analyses.

Table 4.8 Communalities

Variable	**Raw**		**Rescaled**	
	Initial	**Extraction**	**Initial**	**Extraction**
Output_T	9.832	6.213	1.000	0.632
Energy_Cons_T	24.719	24.331	1.000	0.984
Open_FCA_T	10.195	6.236	1.000	0.612
Emp_Eng_T	8.225	6.518	1.000	0.793
Sal_Wages_T	8.247	7.255	1.000	0.880

Extraction Method: Principal Component Analysis.

The communality is the variation in the variables observed, which are considered for by a common factor or common variance (Child, 2006). Initial communalities are measures of the variation in each variable taken into account by all components. Extraction communalities are estimates of the variance in each variable accounted for by the components.

Particular set of factors is said to explain a lot of the variance of a variable if it has a high communality (Kline,1994). Table 4.8 shows that all communalities related to considering variables are greater than 0.6 and it is reasonable validation of variance by individual variables.

Table 4.9 Component Matrix

Variable	PC 1	PC 2
Output_T	0.761	0.230
Energy_Cons_T	0.879	-0.461
Open_FCA_T	0.756	0.201
Emp_Eng_T	0.661	0.596
Sal_Wages_T	0.748	0.566

Extraction Method: Principal Component Analysis.

The component matrix indicates the correlations between the original variables and the components. Also, they represent the contribution of each component in estimating the original variables. If the absolute value of loading is greater than 0.5, that particular component is considered to have a significant contribution to the respective variable.

Table 4.10 Component Score Coefficient Matrix

Variable	Component	
	1	2
Output_T	0.193	0.192
Energy_Cons_T	0.560	-0.969
Open_FCA_T	0.199	0.175
Emp_Eng_T	0.140	0.418
Sal_Wages_T	0.159	0.398

Extraction Method: Principal Component Analysis.

Component scores for individual establishments can be calculated by using the below equation.

$$C_{ij} = \sum_{1}^{2} C_{ik} Z_{kj}$$

Where;

C_{ij} = Score for the *i*^th component on *j*^th establishment

C_{ik} = Component score coefficient for the i^{th} component on k^{th} variable

Z_{kj} = standardized value of the k^{th} variable on j^{th} establishment

Here, $i = 1,2$

$j = 1,2,\ldots.,1792$

$k = 1,2,\ldots,5$

The weights of PCs in Table 4.10 explain which variables are dominant in each PC. The first PC which accounts for 63.39 percent of the total variation in the data, is highly influenced by the variable "Energy_Cons_T".

$$\boldsymbol{PC1} = 0.193\ \text{Output\_T} + 0.560\ \text{Energy\_Cons\_T} + 0.199\ \text{Open\_FCA\_T} + 0.140\ \text{Emp\_Eng\_T} + 0.159\ \text{Sal\_Wages\_T}$$

The second PC which accounts for 19.19 percent of the total variation in the data is highly influenced by the variable "Energy_Cons_T".

$$\boldsymbol{PC2} = 0.192\ \text{Output\_T} - 0.969\ \text{Energy\_Cons\_T} + 0.175\ \text{Open\_FCA\_T} + 0.418\ \text{Emp\_Eng\_T} + 0.398\ \text{Sal\_Wages\_T}$$

4.5 Construction of the Composite index (CI)

Table 4.11 Eigenvalues and weights after rotation

PC	Eigenvalue	Variance Proportion (Weights) (Vi)
1	38.807	0.768
2	11.746	0.232

CI for the jth Establishment was calculated using equation 3.1. mention in chapter 3

$(CI)_j = 0.768 \times C1_j + 0.232 \times C2_j$

Where, $C1_j$, $C2_j$ are the component scores of first and second PC's for the jth establishment respectively.

Table 4.12 Descriptive statistics of the constructed CI

Statistics	**Value**
Mean	0.000
Std. Deviation	0.802
Minimum	-0.308
Maximum	8.869
Skewness	5.501

Table 4.12 indicates that the mean value of the calculated CI value is zero and the standard deviation equal to 0.802. Meanwhile Coefficient of Skewness of CI indicated a positive value, it is indicated that more than 50 percent of establishments have negative values for their respective CI.

4.6 Categorizing Small/ Medium/ Large Establishments by using CI

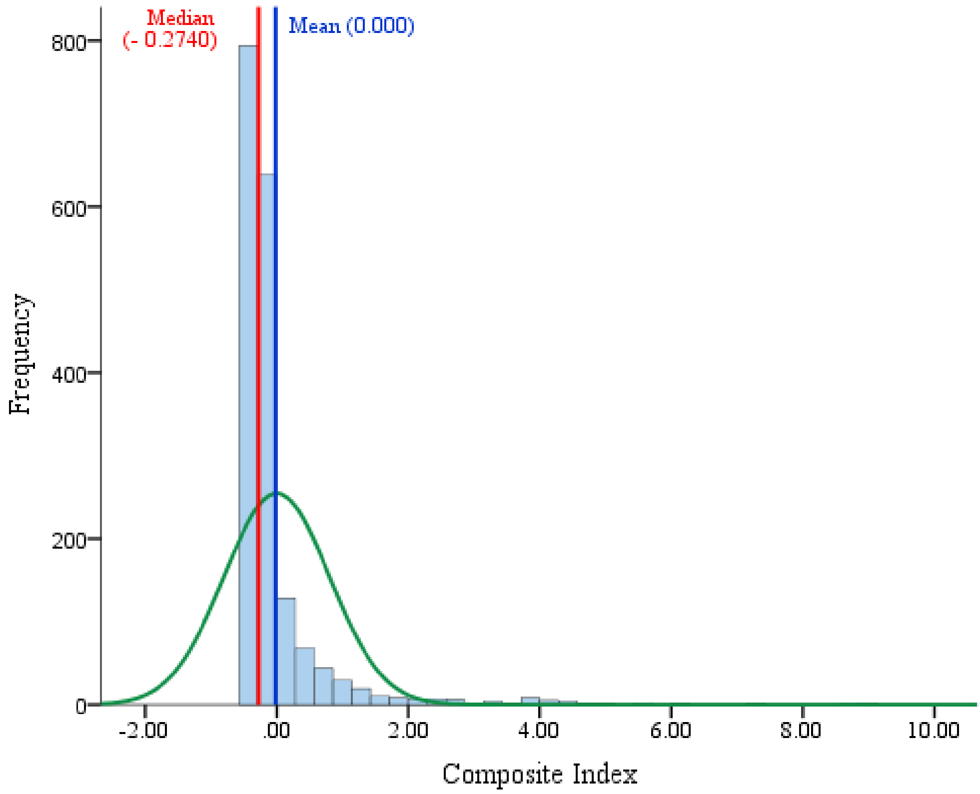

Figure 4.9 Histogram of the composite index value - All sample units

Figure 4.9 histogram shows that CI values of the study sample represent the right-skewed (positive skewed) graph. Table 4.12 clearly describes skewness of the CI value has a positive number (5.501) and that also evidence for proof of the right-skewed shape of the CI.One of the properties of right-skewed data is the mean value always greater than the median value. Figure 4.9 red line represents the median value (-0.2740) and the blue line represents the mean value (0.000) of the CI and it shows that the mean is the higher one compare to the median value. The median value represents the middle value of the CI and its divided CI values into two equal-size groups. Though mean value also divides CI into two groups but its sizes are different and the left side of the mean line represents a large number of establishments (more than 50 percent of establishments) that are CI value less than the mean value. Finally, conclude that if the CI value less than or equal to the mean value of the CI that particular establishment considers as a small-scale one.

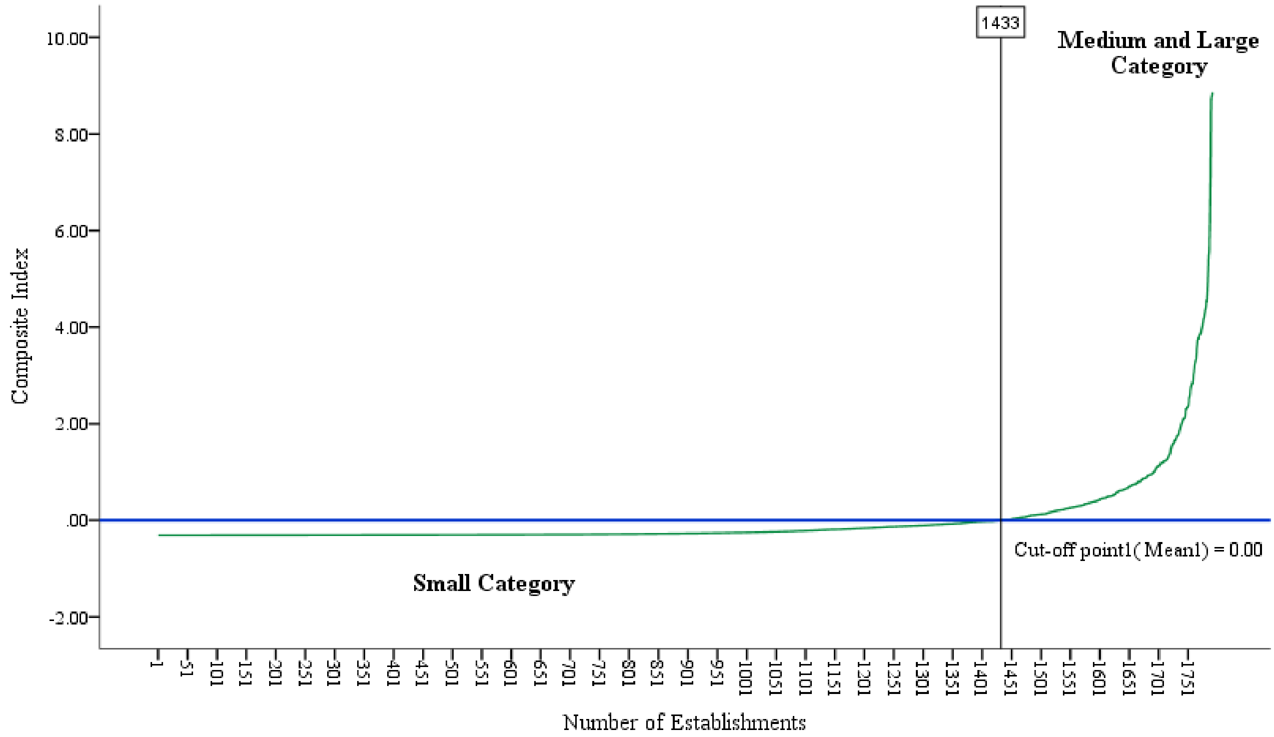

Figure 4.10 Line graph of the composite index value – All sample units

Figure 4.10 illustrates establishment-wise CI values in ascending order and the blue line represents the first cut-off point (mean value of CI). There are 1433 (80.0%) establishments are grouped as small category and other establishments that are greater than the CI mean value grouped as a medium and large category.

To decide the second cut-off point, medium and large category has to be separately analyzed in figure 4.11.

Table 4.13 Descriptive statistics of the medium and large category

Statistics	**Value**
Mean	0.998
Std. Deviation	1.394
Minimum	0.001
Maximum	8.869
Skewness	2.764

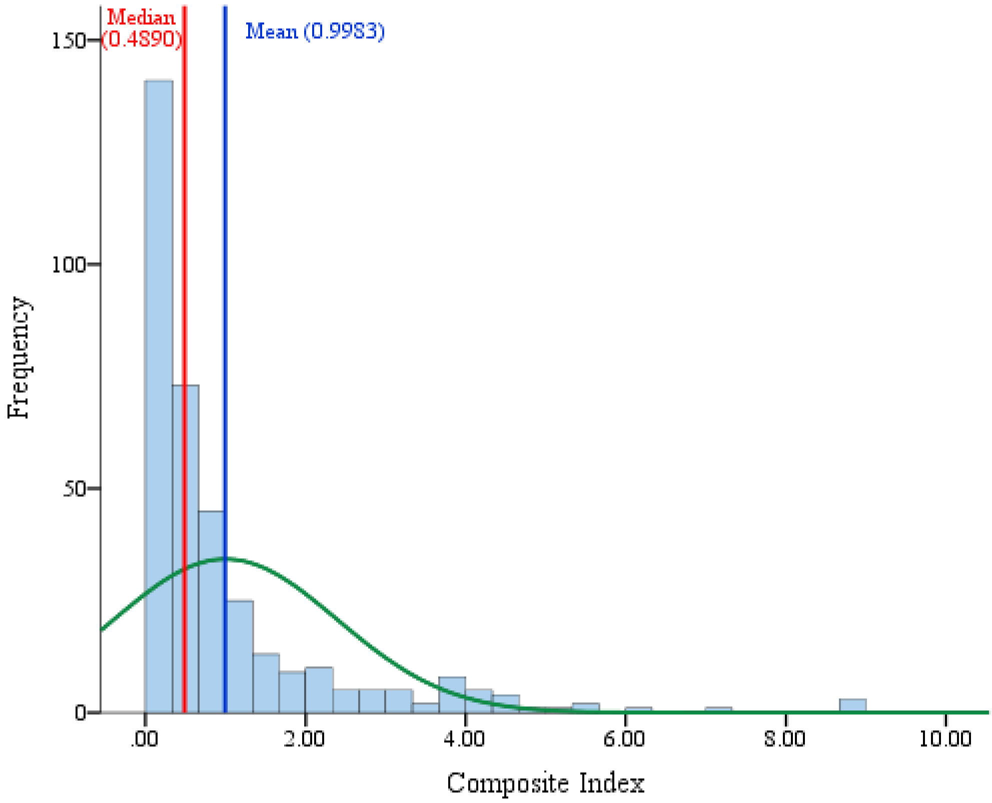

Figure 4.11 Histogram of the composite index value – medium and large category sample units

Figure 4.11 demonstrates the shape of the histogram that is developing by using frequencies of the CI values. It also shows a right-skewed pattern and its characteristics which are similar to figure 4.9 and it's also proof by figures in table 4.13. Because of this similarity and due to more than 50 percent of the establishments represent medium-scale features in the medium and large group; again CI mean value (0.9983) of this group can be considered as the second cut-off point.

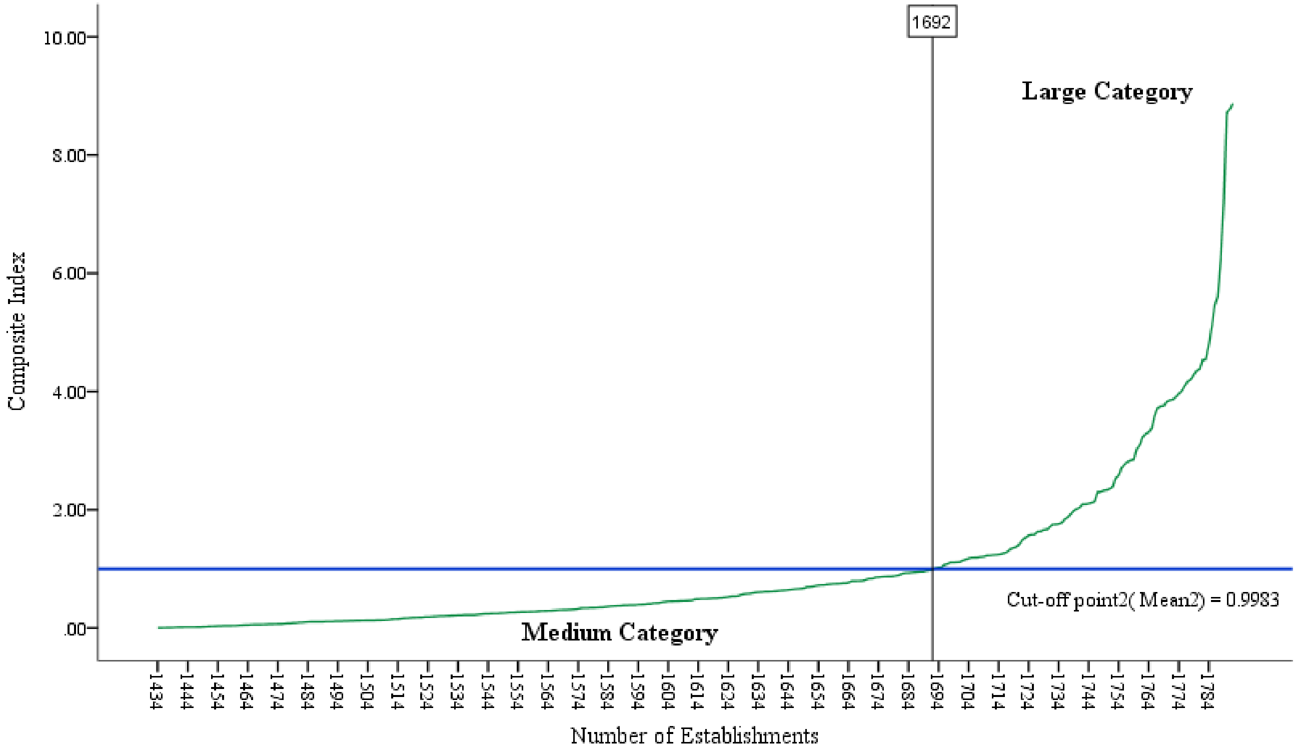

Figure 4.12 Line graph of the composite index value – medium and large category sample units

Figure 4.12 illuminates establishment-wise CI values of the medium and large category group in ascending order and the blue line represents the second cut-off point. There are 259 (14.4%) establishments grouped as the medium category that represents below the blue line and upper CI values grouped as large category and 100 (5.6%) establishments are categorized in that group.

Table 4.14 Cut-off values of proposed categorization

Cut-off Values of CI	**Category**	**No. of Establishments**	**%**
Less than or equal to zero (cutoff point[1])	Small	1,433	80.0
Greater than zero (cutoff point[1]) but less than or equal to 0.9983 (cutoff point[2])	Medium	259	14.4
Greater than 0.9983 (cutoff point[2])	Large	100	5.6
Total		**1,792**	**100.0**

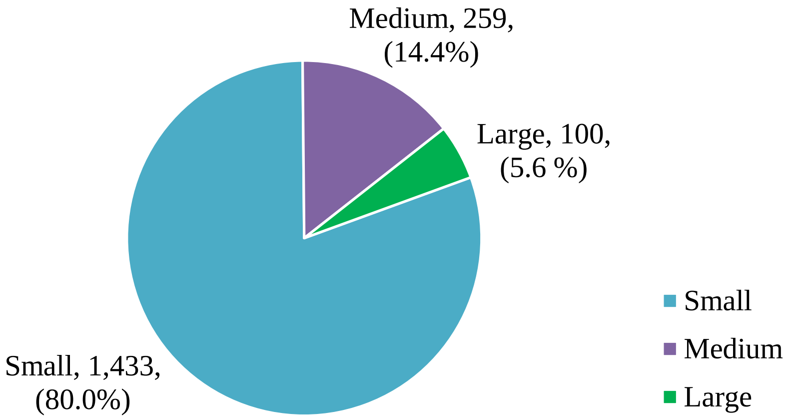

Figure 4.13 Distribution of establishments under the proposed method

According to table 4.14, 1433 (80.0%) number of establishments, out of 1792 establishments represent the small category and 259 (14.4%) establishments are categorized into medium scale. Meanwhile, the large scale represents 100 (5.6%) establishments out of the total.

Table 4.15 Crosstable of existing categorization and propose categorization

		Proposed Categorization			Total	%
		Small	Medium	Large		
MOI Categorization	Small	1,110	3	0	**1,113**	*62.1*
	Medium	317	145	8	**470**	*26.2*
	Large	6	111	92	**209**	*11.7*
Total		**1,433**	**259**	**100**	**1,792**	*100.0*
%		*80.0*	*14.4*	*5.6*	*100.00*	

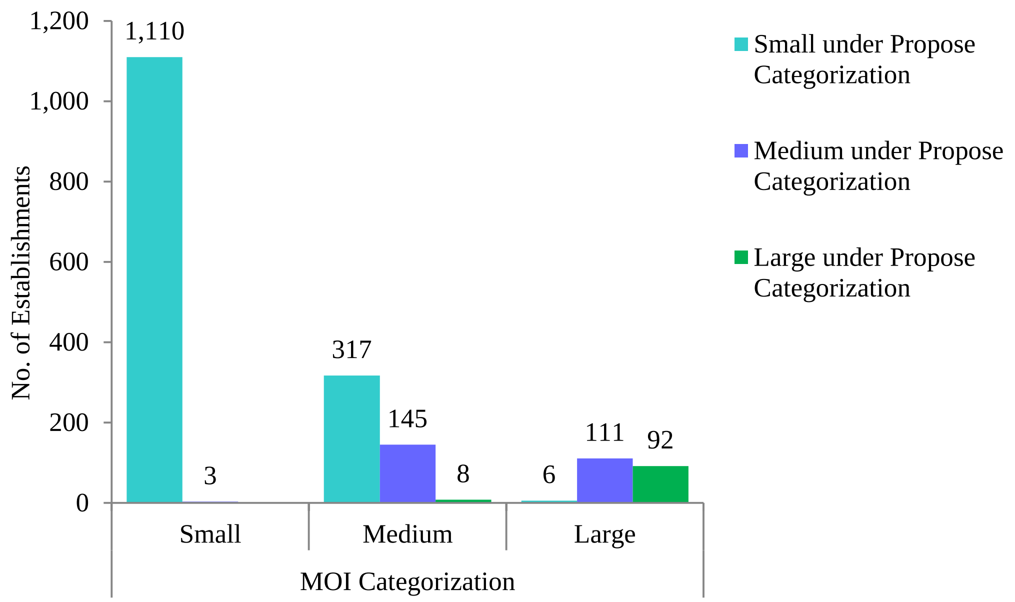

Figure 4.14 Comparision of MOI categorization with proposed categorization

Table 4.15 clearly shows the difference between currently using categorization according to MOI definition and proposed categorization. Under the proposed categorization 1,433 (80.0%) establishments are categorized into small scale while only 1,113 (62.1%) establishments were on small scale under MOI categorization.

That's mean some establishments show small scale characteristics in the previous categorization, but it is not categorized under small because of the limitation of the selection criteria.

Similarly, 259 (14.4%) establishments are categorized into medium-scale under the proposed categorization while 470 (26.2%) establishments are categorized into medium-scale under MOI categorization. That indicates previously grouped 209 medium-scale establishments did not have medium-scale characteristics and they have either small or large scale characteristics.

In large-scale establishments area also describes the same deviations concerning the previous categorization.

Considering the total establishments considered for the study, 1110 (61.9%), 145 (8.1%), and 92 (5.1%) of them remained unchanged as small, medium, and large establishments respectively. Hence comparing with the previous categorization, 445 (off-diagonal) (24.8%) establishments were found to deviate from their previous scale.

In table 4.15 diagonal values (1110, 145, 92) represent matching establishments in both categorization methods.

In chapter 1, table 1.9; it described 398 establishments are being misclassified concerning MOI definition within two criteria no. of employees and turnover. But according to the above explanation, it is 445 establishments were found that scales have been changed. Although the 47 (445 – 398) establishments were correctly classified in the previous method, it has been changed their scale in the proposed method. This may be happened due to extra variables usage in the proposed method. i.e. in the proposed method Energy consumption, Total assets, Annual salary were used in addition to the number of employees and annual turnover.

Table 4.16 Category wise DCS Percentages and Proposed method Percentages

Category (scale)	**DCS Categorization considering the population of Manufacturing establishments 2013**	**Categorization using Proposed method considering sample data 2016**
	%	**%**
Small	79.8	80.0
Medium	16.1	14.4
Large	4.1	5.6

2013/2014 listing stage data set is the most reliable data source to compare percentages, since DCS is the only institution that collects actual data for their census.

Table 4.16 describes the percentage distribution of categories according to DCS definitions (Refer to chapter 1 table 1.2), considering the population of manufacturing establishments 2013/2014 economic census listing report, and percentages of proposed method categorization by analyzing 2016 sample data. Hereby the results obtained through this analysis can be justified since the percentages are approximately having close relation referred to table 4.16.

CHAPTER 5

CONCLUSIONS AND RECOMMENDATIONS

5.1 Conclusion

Especially, scale of the manufacturing sector establishment cannot be decided only considering the number of employee engaged and the total turnover of a particular establishment.

Considering prevailing definitions of the country, not existing a consistent definition for categorizing establishments, is an issue; in the policymaking process. When the distribution of subsidiary and recovering taxes, desirable establishments may miss the opportunity in gaining aids, while some establishments may slip paying taxes to the country.

To overcome this issue this study considered five significant variables and developed a new threshold to identify the scale of the particular manufacturing establishment.
As a result, in this study by using a new statistical base methodology, treated in a precise way, for the mismatch scaled establishments.

According to chapter 1 table 1.9, it was identified 398 establishments have been mismatched in MOI categorization with respect to two criteria, while in the proposed method they were correctly categorized. Moreover, another 47 establishments have been changed their scale. This may occur with the introduction of more variables determining company scale.

Introducing a more accurate and reliable categorizing method, which affects the distribution of country resources in a precise manner. That led companies to grow while country development is in the best position.

5.2 Recommendations

This research covered only five or more persons engaged in manufacturing establishments so that there is a future opportunity to extend the proposed statistical methodology to cover less than five persons engaged in manufacturing establishments. Not only that, but this study can also be applied for construction, trade, and service sectors by carefully selecting appropriate variables.

DCS conducts an economic census every ten years. Before this census, they update their industry frame by conducting a listing. In the listing stage, they update newly entered and also closed establishments. Therefore every ten years change the population size. So that it is recommended to change the base year of this study and re-analysis the new sample and recalculate the cut-off points and to compare with existing cut-off points. If there are significant changes occur then can implement the needful changes to existing cut-off points for different scales.

If need to classify a newly entered manufacturing company that has five or more persons engaged it can be categorized using the proposed method concerning the base year 2016.

REFERENCES

Abdi, H., & Williams, L. J. (2010). Principal component analysis. *Wiley interdisciplinary reviews: computational statistics*, *2*(4), 433-459.

Child, D. (2006). *The essentials of factor analysis*. A&C Black.

Dharmawardena, P., Thattil, R. O., & Samita, S. (2017). Scale Independent Principal Component Analysis and Factor Analysis with Preserved Inherent Variability of the Indicators. *American Journal of Theoretical and Applied Statistics*, *6*(2), 90-94.

Dharmawardena, J. S. N. P., Thattil, R. O., & Samita, S. (2016). Adjusting variables in constructing composite indices by using principal component analysis: illustrated by Colombo district data. *Tropical Agricultural Research*, *27*(1).

Field, A. P., Miles, J., & Field, Z. (2012). Discovering statistics using R/Andy Field, Jeremy Miles, Zoë Field.

Hutcheson, G. D., & Sofroniou, N. (1999). The multivariate social scientist: Introductory statistics using generalized linear models. Sage.

Jolliffe, I. T. (2002). Principal components in regression analysis. *Principal component analysis*, 167-198.

Kline, P. (2014). *An easy guide to factor analysis*. Routledge.

Kaiser, H. F. (1974). An index of factorial simplicity. *Psychometrika*, *39*(1), 31-36.

Laurent, A., Olsen, S. I., & Hauschild, M. Z. (2010). Carbon footprint as environmental performance indicator for the manufacturing industry. *CIRP annals*, *59*(1), 37-40.

Liu, H., Shah, S., Jiang, W. (2004), "On-line outlier detection and data cleaning," *Computers and Chemical Engineering*, 28, 1635–1647.

Malhotra, N. K. (1987). Analyzing marketing research data with incomplete information on the dependent variable. *Journal of Marketing Research*, *24*(1), 74-84.

Perera, K. A. H. P. , Wickramarachchi, D. C. , Dharmawardena, J. S. N. P. , (2017). " Developing an Alternative Method to Categorize Small and Medium Enterprises in Sri Lanka",*Proceedings of the International Statistics Conference 2017 – IASSL, Sri Lanka*

PRESS RELEASE, Key indicators of Industry Trade and Services Sector Economic Census 2013/14 Department of Census and Statistics http://www.statistics.gov.lk/PressReleases/Files/en/EC_20150714E.pdf

Pallant, J. (2013). *SPSS survival manual*. McGraw-Hill Education (UK).
Shaw PJA (2003) Multivariate statistics for the Environmental Sciences. Hodder-Arnold, London

Saltelli, A. (2007). Composite indicators between analysis and advocacy. *Social indicators research*, *81*(1), 65-77.

The Organization for Economic Co-operation and Development, (2004) *The OECD-JRC Handbook on Practices for Developing Composite Indicators*. https://stats.oecd.org/glossary/detail.asp?ID=6278.

Vijayakumar, S. (2013). Small and Medium Enterprise Promotions for their Growth in Sri Lanka. *International Journal on Global Business Management & Research*, *1*(2), 46.

Williams, G. J., Baxter, R. A., He H. X., Hawkins S. and Gu L. (2002) "A Comparative Study of RNN for Outlier Detection in Data Mining" *IEEE International Conference on Data-mining (ICDM'02)*, Maebashi City, Japan, *CSIRO Technical Report CMIS-02/102*.

APPENDICES

Appendix A: Annual Survey of Industries 2017 - Questionnaire

ASI -Q

Reference No.

DEPARTMENT OF CENSUS AND STATISTICS

ANNUAL SURVEY OF INDUSTRIES - 2017

CONFIDENTIAL
This information is sought under the Statistics Ordinance as amended in 2000. Individual data will be treated as strictly confidential.

The data required is for the calender year 2016. If book figures are not available, carefully prepared estimates could be accepted. The 2016/2017 finance year data can be used for 2016 calender year, when first quarter of 2016/2017 are not significantly different.

01.Name of Establishment/ Propritor : ....

02.Address or location of Establishment : ....

03.Telephone Number :

04.Type of the Establishment (Please circle the relevant cage)

1. A Single Unit /Establishment .... 1 → go to Ques. No. 7
2. A Multi - Establishment .... → go to Ques. No. 5

05. Do you maintain seperate accounts for this establishment ? (Please circle the relevant cage)
1. Yes ..... 1 → go to Ques. No. 7
2. No ...... 2 → go to Ques. No. 6

06. Details of the parent company or Head Office :

(i) Name & Address: ....

(ii) Telephone Number : (iii) e- mail address:

(iv) No. of branches operated by your head office :

07. Province & District : .... 08. D.S. Division : ....

09. G.N. Division : .... 10. MC/UC/PS : ....

11. Ward/Village/Estate : ....

12. Type of legal organization : (Circle the apporpriate number)

1 Sole ownership	2 Partnership	3 Private Limited Liability Co.
4 Public Limited Liablitily Co.	5 Co-operative Society	6 Government
7 Semi Government	8 Non - profit institutes for serving household	9 Other (specify) ....

13. Description of Industry (Economic activity) : ....

14. Year of commencement of commercial operation of the establishment ....

15. Duration of the economic activity conducted (No. of months) :

16. Was your establishment engaged in research & development (R&D) activities in year 2016? (Pls. circle the relevent cage)
1. Yes 1
2. No 2

22. Employment and earnings

Persons engaged		Male No.	Female No.	Total No.	Salaries & Wages Rs.	Labour demand No.
		Employment and Earnings - 2016				
i		ii	iii	iv	v	vi
1. Operatives	(a) Skilled					
	(b) Unskilled					
2. Other employees	(a) Administrative/Managerial					
	(b) Technical/Supervisory					
	(c) Clerical and related					
	(d) Other					
3. Working Proprietors and Active Partners						
4. Unpaid family workers						
5. Total						

23. Any of the following changes of production noticed in 2016 compared to 2015 (Pls. circle the appropriate number)
1. Increase of the production
2. Decrease of the production
3. No noticeable change of the production

24. If there was any decline please circle appropriate reason/s

1 Lack of raw materials	6 Decline in the international demand
2 Lack of skilled Labours	7 High competitiveness in the market
3 Price increase of raw materials	8 Imports of finished goods
4 Decrease of the sale price in the market	9 Price increase of fuel/electricity
5 Decline in the domestic demand	10 Others (specify) ....

Remarks

....
....
....
....

I here by declare that the data given in this questionnaire is true and correct to the best of my knowledge and belief.

Signature : .... Name : ....

Designation : .... Date : .... Tel No: ....

Please despatch the completed questionnaire to the :

Director,
Industry, Construction, Trade and Services Division,
Department of Census and Statistics,
No 306/71, Polduwa Road,
Battaramulla.

Tel No: 011 - 2147401
Fax No: 011 -2147412
e-mail : industries@statistics.gov.lk

For office use only

Particulars	Field follow-up carried out by	Checked in the field by	Edited & coded in the office by	Data entry in the office by	Supervised in the office by
Name					
Designation					
Signature					
Date					

17. Capacity, Production moved out (shipments)and Stocks of finished goods

For office use only (CPC Code)	Description of Commodities produced	Unit of measurements	Production capacity	Total products moved out the establishment in 2016		Of which exports in 2016 Quantity	Stock of finished goods				For office use only
				Quantity	Value (at basic price*) Rs.		As at 01.01.2016		As at 31.12.2016		
							Quantity	Value (Rs.)	Quantity	Value (Rs.)	
i	ii	iii	iv	v	vi	vii	viii	ix	x	xi	xii
	1.										
	2.										
	3.										
	4.										
	5.										
	6										
	7.										
	8. Other products										
	9. Total										
	10. Semi finished goods										

* The basic price is the amount receivable by the producer from the purchaser for a unit of a good or services produced as output , minus any tax payable and plus any subsidy receivable by the producer as a consequence of its production or sale. It excludes any transport charges.

18.Receipts from Industrial & Non-industrial services rendered to others

Ser. No	Item	Value (Rs.)
i	ii	iii
1.	Receipts from industrial services rendered to others	
2.	Sales of all goods purchased for resale in the same condition as received.	
3.	Receipts for transport services rendered to others	
4.	Royality income	
5.	Revenue from sales of scrap	
6.	Other subsidies on production	
7.	Other revenue*	
8.	Total	

* Dividend receipts, Interest and discount receipts , Revenue from the outright sale of patent and licences, revenue from sales of land and used capital goods should not be included

19. Cost of Industrial & Non - Industrial Services purchased

Ser. No.	Item	Value (Rs.)
i	ii	iii
1.	Cost of Industrial services done by others for your establishments	
2.	Maintenance, Repair & Installation	
3.	Transport Services	
4.	Purchase of Communication Services	
5.	Advertising & Promotional Services	
6.	Financial Services (Except interest payments)	
7.	Non - life insurance premiums payable on establishment porperty.	
8.	Rental payments	
9.	Other taxes on production (eg: Property tax)	
10.	Electricity	
11.	Water	
12	Other non-Industrial Servics* (Legal services, consulting services, accounting and bookkeeping services, entertainment, meeting cost etc.)	
13.	Total	

*Following items should be excluded. Dividend and interest paid, fines, outright purchases of patent and licences, purchases of land and other capital goods, donations, bad debts and depreciations.

20. Value of Raw Materials, packing materials, goods purchased for resale and fuel

Ser. No.	For office use only (CPC Code)	Description of Items	Purchased in 2016 (Rs.)	Value of stocks as at 01.01.2016 (Rs.)	Value of stocks as at 31.12.2016 (Rs.)
i	ii	iii	iv	v	vi
1.		Raw materials			
		i.			
		ii.			
		iii.			
		iv.			
		v.			
2.		Packing materials			
		i.			
		ii.			
3.		Goods purchase for resale			
4.		Fuel			
		i.			
		ii.			
		iii.			
5.		Total			
For office use only					

21. Fixed Capital Assets - 2016

Ser. No.	Category	Book value at the beginning of the year* Rs.	Gross additions during the year** Rs.	Value of own account fixed assests Rs.	Depreciation during the year Rs.
i	ii	iii	iv	v	vi
01.	Tangible fixed assets				
	(i) Land				
	(ii) Building and other constructions				
	(iii) Plant & machinery				
	(iv) Transport equipments				
	(v.) Computer & accessories				
	(vi) Furniture & other office equipments				
	(vii) Work in progress				
	(viii) Other tangible fixed assets				
02.	Intangible fixed assets				
03.	Total				

*** Less depreciations and assets retired or sold out ** Less assets sold out during the year**

Appendix B: International Standard Industrial Classification of All Economic Activities (ISIC), Rev.4-Industry Sector

B - Mining and quarrying

05 - Mining of coal and lignite

06 - Extraction of crude petroleum and natural gas

07 - Mining of metal ores

08 - Other mining and quarrying

09 - Mining support service activities

C - Manufacturing

10 - Manufacture of food products

11 - Manufacture of beverages

12 - Manufacture of tobacco products

13 - Manufacture of textiles

14 - Manufacture of wearing apparel

15 - Manufacture of leather and related products

16 - Manufacture of wood and of products of wood and cork, except furniture; manufacture of articles of straw and plaiting materials

17 - Manufacture of paper and paper products

18 - Printing and reproduction of recorded media

19 - Manufacture of coke and refined petroleum products

20 - Manufacture of chemicals and chemical products

21 - Manufacture of basic pharmaceutical products and pharmaceutical preparations

22 - Manufacture of rubber and plastics products

23 - Manufacture of other non-metallic mineral products

24 - Manufacture of basic metals

25 - Manufacture of fabricated metal products, except machinery and equipment

26 - Manufacture of computer, electronic and optical products

27 - Manufacture of electrical equipment

28 - Manufacture of machinery and equipment n.e.c.

29 - Manufacture of motor vehicles, trailers and semi-trailers

30 - Manufacture of other transport equipment

31 - Manufacture of furniture

32 - Other manufacturing

33 - Repair and installation of machinery and equipment

D - Electricity, gas, steam and air conditioning supply

35 - Electricity, gas, steam and air conditioning supply

E - Water supply; sewerage, waste management and remediation activities

36 - Water collection, treatment and supply

37 - Sewerage

38 - Waste collection, treatment and disposal activities; materials recovery

39 - Remediation activities and other waste management services

F - Construction

41 - Construction of buildings

42 - Civil engineering

43 - Specialized construction activities