ROBUST LINEAR DISCRIMINANT ANALYSIS USING MOM-*Qn* AND WMOM-*Qn* ESTIMATORS: COORDINATE-WISE APPROACH

HAMEEDAH NAEEM MELIK

MASTER OF SCIENCE (STATISTICS)
UNIVERSITI UTARA MALAYSIA
2017

Permission to Use

In presenting this thesis in fulfilment of the requirements for a postgraduate degree from Universiti Utara Malaysia, I agree that the Universiti Library may make it freely available for inspection. I further agree that permission for the copying of this thesis in any manner, in whole or in part, for scholarly purpose may be granted by my supervisor(s) or, in their absence, by the Dean of Awang Had Salleh Graduate School of Arts and Sciences. It is understood that any copying or publication or use of this thesis or parts thereof for financial gain shall not be allowed without my written permission. It is also understood that due recognition shall be given to me and to Universiti Utara Malaysia for any scholarly use which may be made of any material from my thesis.

Requests for permission to copy or to make other use of materials in this thesis, in whole or in part, should be addressed to:

Dean of Awang Had Salleh Graduate School of Arts and Sciences
UUM College of Arts and Sciences
Universiti Utara Malaysia
06010 UUM Sintok

Abstrak

Kaedah analisis diskriminan linear (RLDA) teguh menjadi pilihan yang lebih baik untuk masalah pengklasifikasi berbanding dengan analisis diskriminan linear (LDA) klasik disebabkan kemampuan kaedah tersebut dalam mengatasi isu titik terpencil. LDA klasik bergantung kepada penganggar lokasi dan skala yang biasa iaitu min sampel dan kovarians matriks. Sensitiviti penganggar ini ke arah data terpencil akan menjejaskan proses pengelasan. Untuk mengurangkan isu ini, penganggar teguh lokasi dan kovarians dicadangkan. Sehubungan itu, dalam kajian ini, dua RLDA untuk pengelasan dua kumpulan telah diubah suai menggunakan dua penganggar lokasi yang amat teguh yang dinamakan Penganggar-M satu langkah terubahsuai (MOM) dan Penganggar-M satu langkah terubahsuai terwinsor (WMOM). Satu penganggar skala yang amat teguh, *Qn*, disepadukan dalam kriteria pemangkasan MOM dan WMOM, menghasilkan dua RLDA yang baharu yang masing-masing dikenali sebagai RLDA_{MQ} dan RLDA_{WMQ}. Dalam pengiraan RLDA yang baharu, min biasa digantikan dengan MOM-*Qn* dan WMOM-*Qn*. Prestasi kaedah RLDA baharu diuji ke atas data simulasi begitu juga data sebenar, dan seterusnya dibandingkan dengan LDA klasik. Bagi data simulasi, beberapa pemboleh ubah telah dimanipulasi untuk mewujudkan pelbagai keadaan yang sering berlaku dalam kehidupan sebenar. Pembolehubah tersebut ialah kehomogenan kovarians (sama dan tidak sama), saiz sampel (seimbang dan tidak seimbang), dimensi pembolehubah, dan peratus pencemaran. Secara umumnya, keputusan menunjukkan bahawa prestasi RLDA baharu adalah lebih baik daripada LDA klasik dari segi purata ralat kesilapan pengelasan, walaupun RLDA yang baharu mempunyai kelemahan iaitu memerlukan lebih banyak masa pengiraan. RLDA_{MQ} memberi hasil yang terbaik pada saiz sampel seimbang manakala RLDA_{WMQ} lebih baik dari yang lainnya pada keadaan saiz sampel tidak seimbang. Apabila data kewangan yang sebenar dipertimbangkan, RLDA_{MQ} menunjukkan keupayaan dalam menangani data terpencil dengan ralat kesilapan pengelasan yang paling kecil. Sebagai penutup, kajian ini telah mencapai objektif utama iaitu untuk memperkenalkan RLDA baharu untuk mengklasifikasi data multi pembolehubah dua kumpulan dengan kehadiran titik terpencil.

Kata kunci: Ralat kesilapan pengelasan, Penganggar-M satu langkah terubahsuai, Data terpencil, Analisis diskriminan linear teguh, Terwinsor.

Abstract

Robust linear discriminant analysis (RLDA) methods are becoming the better choice for classification problems as compared to the classical linear discriminant analysis (LDA) due to their ability in circumventing outliers issue. Classical LDA relies on the usual location and scale estimators which are the sample mean and covariance matrix. The sensitivity of these estimators towards outliers will jeopardize the classification process. To alleviate the issue, robust estimators of location and covariance are proposed. Thus, in this study, two RLDA for two groups classification were modified using two highly robust location estimators namely Modified One-Step M-estimator (MOM) and Winsorized Modified One-Step M-estimator (WMOM). Integrated with a highly robust scale estimator, *Qn*, in the trimming criteria of MOM and WMOM, two new RLDA were developed known as $RLDA_{MQ}$ and $RLDA_{WMQ}$ respectively. In the computation of the new RLDA, the usual mean is replaced by MOM-*Qn* and WMOM-*Qn* accordingly. The performance of the new RLDA were tested on simulated as well as real data and then compared against the classical LDA. For simulated data, several variables were manipulated to create various conditions that always occur in real life. The variables were homogeneity of covariance (equal and unequal), samples (balanced and unbalanced), dimension of variables, and the percentage of contamination. In general, the results show that the performance of the new RLDA are more favorable than the classical LDA in terms of average misclassification error for contaminated data, although the new RLDA have the shortcoming of requiring more computational time. $RLDA_{MQ}$ works best under balanced sample sizes while $RLDA_{WMQ}$ surpasses the others under unbalanced sample sizes. When real financial data were considered, $RLDA_{MQ}$ shows capability in handling outliers with lowest misclassification error. As a conclusion, this research has achieved its primary objective which is to develop new RLDA for two groups classification of multivariate data in the presence of outliers.

Keywords: Misclassification Error, Modified One-Step M-Estimator, Outliers, Robust linear discriminant analysis, Winsorized.

Acknowledgement

I am grateful to the Almighty Allah for giving me the opportunity to complete my Master's thesis in Universiti Utara Malaysia. This achievement would not have been possible without the guidance and help of several individuals who contributed their assistance in the preparation of this thesis towards the completion of my study. It gives me great pleasure to acknowledge their support.

First and foremost, I would like to express my deepest appreciation and gratitude to my supervisor, Dr. Nor Aishah Ahad for her valuable support and guidance throughout this study. I could not have imagined being under such a great tutelage. Your constructive advice and constant availability all through my study is well appreciated. I would like to also thank my co-supervisor Prof. Dr. Sharipah Soaad Syed Yahaya who supported me and assisted me through all stages of my research and the preparation of the thesis. I am highly honored to have had the pleasure of working with you. My sincere gratitude is extended to all academic and administrative staff in the Department of Quantitative Sciences and College of Arts and Sciences Universiti Utara Malaysia.

My special appreciation also goes to my father who has been a great and wise teacher in my life and my lovely mother for her infinite patience especially during my absence. Your sincere flow of love has accompanied me all the way in my long struggle and has pushed me to pursue my dreams. My heartfelt gratitude also goes to my two sisters and brother for their patience, prayers and moral support all through this wonderful journey.

Finally, I would like to thank everyone who has directly or indirectly helped me during this research. Your support is greatly appreciated. Allah blesses you.

Table of Contents

Permission to Use ii

Abstrak iii

Abstract iv

Acknowledgement v

Table of Contents vi

List of Tables ix

List of Figures xi

List of Abbreviations xii

CHAPTER ONE INTRODUCTION 1

1.1 Overview 1

1.2 Linear Discriminant Analysis (LDA) Method 4

1.3 Problem Statement 9

1.4 Objectives of the Study 11

1.5 Significance of the Study 12

1.6 Scope of the Study 12

CHAPTER TWO LITERATURE REVIEW 14

2.1 Discriminant Analysis 14

2.1.1 Discriminant Function 15

2.2 Linear Discriminant Analysis (LDA) 18

2.2.1 Fisher LDA 18

2.2.2 Limitations of LDA 20

2.2.2.1 Small Sample Size Problem (SSS) 20

2.2.2.2 Overfitting or Underfitting 22

2.2.2.3 Distribution Assumption 24

2.3 Multivariate Outliers 26

2.4 Misclassification Error 28

2.5 Trimming 30

2.6 Robust LDA 32

2.6.1 Robust Estimators 34

2.6.2 Properties of Robust Estimators 35

2.6.3 Types of Robust Estimators 37

2.6.3.1 Modified One-Step M-Estimator (MOM) 37

2.6.3.2 Winsorized Modified One-Step M-Estimator (WMOM) 38

2.7 Scale Estimators 40

2.7.1 Q_n 41

2.8 Variance Estimators 42

2.8.1 The Traditional Approach 43

2.8.2 Cross-Validation (CV) 45

2.9 Summary 47

CHAPTER THREE RESEARCH METHODOLOGY 48

3.1 Research Design 48

3.2 Research Framework 49

3.2.1 Generation of Data 50

3.2.2 Properties of Data 50

3.2.3 Assumptions of the Discriminant Model 51

3.3 Linear Discriminant Analysis (LDA) 53

3.4 Modified One-Step M-Estimator with Q_n (MOM-Q_n) 56

3.5 Winsorized Modified One-Step M-Estimator with Q_n (WMOM-Q_n) 57

3.6 Cross Validation (CV) 59

3.7 Variables Manipulated 59

3.7.1 Dimension of Variable (p) and Sample Size (n) 60

3.7.2 Percentage of Contamination (ε), Shifts in Location (μ) and Population (κ) 61

CHAPTER FOUR RESULT AND ANALYSIS 63

4.1 Introduction 63

4.2 Misclassification Error Analysis with Simulation Study 63
4.2.1 Equal Covariance Matrices 64
4.2.1.1 Balanced Sample Sizes 64
4.2.1.2 Unbalanced Sample Sizes 73
4.2.2 Unequal Covariance Matrices 79
4.2.2.1 Balanced Sample Sizes 79
4.2.2.2 Unbalanced Sample Sizes 84
4.3 Computational Time Analysis with Simulation Study 89
4.3.1 Equal Covariance Matrices with Balanced Sample Sizes 89
4.3.2 Equal Covariance Matrices with Unbalanced Sample Sizes 95
4.3.3 Unequal Covariance Matrices with Balanced Sample Sizes 100
4.3.4 Unequal Covariance Matrices with Unbalanced Sample Sizes 104
4.4 Misclassification Error Analysis with Real Data 108
CHAPTER FIVE CONCLUSION AND FUTURE WORK 110
5.1 Conclusion 110
5.2 Comparison between the Linear Models 113
5.3 Implication of Study 116
5.4 Limitation of Study and Future Work 117
REFERENCES 118

List of Tables

Table 3.1 Simulation Conditions 60

Table 4.1 Mean Misclassification Error for Linear Discriminant Models with Balanced Sample Sizes, Equal Covariance Matrices and $p = 2$ 66

Table 4.2 Mean of Misclassification Error for Linear Discriminant Models with Balanced Sample Sizes, Equal Covariance Matrices and $p = 6$ 70

Table 4.3 Mean of Misclassification Error for Linear Discriminant Models with Balanced Sample Sizes, Equal Covariance Matrices and $p = 10$ 71

Table 4.4 Mean Misclassification Error for Linear Discriminant Models with Unbalanced Sample Sizes, Equal Covariance Matrices and $p = 2$ 74

Table 4.5 Mean of Misclassification Error for Linear Discriminant Models with Unbalanced Sample Sizes, Equal Covariance Matrices and $p = 6$ 76

Table 4.6 Mean of Misclassification Error for Linear Discriminant Models with Unbalanced Sample Sizes, Equal Covariance Matrices and $p = 10$ 77

Table 4.7 Mean Misclassification Error for Linear Discriminant Models with Balanced Sample Sizes, Unequal Covariance Matrices and $p = 2$ 80

Table 4.8 Mean of Misclassification Error for Linear Discriminant Models with Balanced Sample Sizes, Unequal Covariance Matrices and $p = 6$ 82

Table 4.9 Mean of Misclassification Error for Linear Discriminant Models with Balanced Sample Sizes, Unequal Covariance Matrices and $p = 10$ 83

Table4.10 Mean Misclassification Error for Linear Discriminant Models with Unbalanced Sample Sizes, Unequal Covariance Matrices and $p = 2$ 85

Table 4.11 Mean of Misclassification Error for Linear Discriminant Models with Unbalanced Sample Sizes, Unequal Covariance Matrices and $p = 6$ 86

Table 4.12 Mean of Misclassification Error for Linear Discriminant Models with Unbalanced Sample Sizes, Unequal Covariance Matrices and $p = 10$ ... 87

Table 4.13 Computational Time (in seconds) for Linear Discriminant Models with Balanced Sample Sizes, Equal Covariance Matrices and $p = 2$ 90

Table 4.14 Computational Time (in seconds) for Linear Discriminant Models with Balanced Sample Sizes, Equal Covariance Matrices and $p = 6$ 91

Table 4.15 Computational Time (in seconds) for Linear Discriminant Models with Balanced Sample Sizes, Equal Covariance Matrices and $p = 10$ 92

Table 4.16 Average Computational Time (in seconds) for Linear Discriminant Models with Balanced Sample Sizes, Equal Covariance Matrices 93

Table 4.17 Computational Time (in seconds) for Linear Discriminant Models with Unbalanced Sample Sizes, Equal Covariance Matrices and $p = 2$ 95

Table 4.18 Computational Time (in seconds) for Linear Discriminant Models with Unbalanced Sample Sizes, Equal Covariance Matrices and $p = 6$ 96

Table 4.19 Computational Time (in seconds) for Linear Discriminant Models with Unbalanced Sample Sizes, Equal Covariance Matrices and p=10 97

Table 4.20 Average Computational Time (in seconds) for Linear Discriminant Models with Unbalanced Sample Sizes, Equal Covariance Matrices 98

Table 4.21 Computational Time (in seconds) for Linear Discriminant Models with Balanced Sample Sizes, Unequal Covariance Matrices and p=2 101

Table 4.22 Computational Time (in seconds) for Linear Discriminant Models with Balanced Sample Sizes, Unequal Covariance Matrices and p=6 102

Table 4.23 Computational Time (in seconds) for Linear Discriminant Models with Balanced Sample Sizes, Unequal Covariance Matrices and p=10 103

Table 4.24 Average Computational Time (in seconds) for Linear Discriminant Models with Balanced Sample Sizes, Unequal Covariance Matrices ... 104

Table 4.25 Computational Time (in seconds) for Linear Discriminant Models with Unbalanced Sample Sizes, Unequal Covariance Matrices and p=2 105

Table 4.26 Computational Time (in seconds) for Linear Discriminant Models with Unbalanced Sample Sizes, Unequal Covariance Matrices and p=6 106

Table 4.27 Computational Time (in seconds) for Linear Discriminant Models with Unbalanced Sample Sizes, Unequal Covariance Matrices and p=10 107

Table 4.28 Average Computational Time (in seconds) for Linear Discriminant Models with Unbalanced Sample Sizes, Unequal Covariance Matrices 108

Table 4.29 Error Rates for Linear Models using Real Data................................... 109

Table 5.1 Summary of Results for Equal Covariance Matrices and Balanced Sample Size Analysis 113

Table 5.2 Summary of Results for Equal Covariance Matrices and Unbalanced Sample Size Analysis...114

Table 5.3 Summary of Results for Unequal Covariance Matrices and Balanced Sample Size Analysis...115

Table 5.4 Summary of Results for Unequal Covariance Matrices and Unbalanced Sample Size Analysis...115

Table 5.5 Summary of Results for Performance of Models with Respect to Presence of Contaminations...116

List of Figures

Figure 2.1: Masking and Swamping Effects on Outliers... 27

Figure 3.1: The Research Flowchart 49

Figure 4.1: Average Computational Time (in seconds) for Linear Discriminant Models with Balanced Sample Sizes, Equal Covariance Matrices and $p = 2$ 94

Figure 4.2: Average Computational Time (in seconds) for Linear Discriminant Models with Balanced Sample Sizes, Equal Covariance Matrices and p=6 94

Figure 4.3: Average Computational Time (in seconds) for Linear Discriminant Models with Balanced Sample Sizes, Equal Covariance Matrices and p=10 94

Figure 4.4: Average Computational Time (in seconds) for Linear Discriminant Models with Unbalanced Sample Sizes, Equal Covariance Matrices and p=2 99

Figure 4.5: Average Computational Time (in seconds) for Linear Discriminant Models with Unbalanced Sample Sizes, Equal Covariance Matrices and p=6 ... 99

Figure 4.6: Average Computational Time (in seconds) for Linear Discriminant Models with Unbalanced Sample Sizes, Equal Covariance Matrices and p=10 .. 99

List of Abbreviations

MOM	Modified One-step M-estimator
WMOM	Winosrized Modified One-step M-estimator
CA	Classical Approach
Q_n	A scale estimator
CV	Cross- Validation
LDA	Linear Discriminant Analysis
MOM-Q_n	Modified One-Step M-Estimator with Q_n
WMOM-Q_n	Winsorized Modified One-Step M-Estimator with Q_n
$RLDA_{MQ}$	RLDA with MOM-Q_n
$RLDA_{WMQ}$	RLDA with WMOM-Q_n
QDA	Quadratic Discriminant Analysis
LR	Logistic Regression
RDA	Regularized Discriminant Analysis
MVE	Minimum Volume Ellipsoid
MCD	Minimum Covariant Determinant
MAD	Mean Absolute Deviation
PCA	Principal Component Analysis
RLDA	Robust Linear Discriminant Analysis
KPCA	Kernel Principal Component Analysis
CKFD	Complete Kernel Fisher Discriminant
KFD	Kernel Fisher Discriminant
LLDA	Locally Linear Discriminant Analysis

MODA	Multimodal Oriented Discriminant Analysis
$\mathrm{MAD_n}$	Median Absolute Deviation
S_n	A scale estimator
T_n	A scale estimator
LSE	Least-Squares Estimation
MSE	Mean Squared Error
AER	Apparent Error Rates

CHAPTER ONE
INTRODUCTION

1.1 Overview

Statistical classification techniques are basically of two types; cluster analysis and discriminant analysis. In cluster analysis, the rule to classify and the independent variables that describe the classification of the object are known but the category of the object is not known. Whereas, in discriminant analysis the object groups and several training examples of objects that have been grouped are known and the model of classification is also given. Discriminant analysis is one of the methods that give more information to the structure of multivariate data; which are data arising from variables greater than one (Fidler & Leonardis, 2003). The construction of a discriminant procedure comes from a training sample used for classifying every member of the sample. One of the primary objectives of discriminant analysis is to make inference about the unknown class membership of a new observation.

As stated in Chen and Muirhead (1994), distributional assumptions on the observation which involves the measurement of groups separately and the examination of the properties of the intended algorithms are the major root of statistical considerations in discriminant analysis. These rationales form the two stages of separation and allocation of the discriminant analysis. The separation stage is aimed to obtain functions known as discriminant functions which can conveniently make a separation of the groups, while the allocation stage involves assigning an unclassified object to one of the given groups using discriminant functions. On the other hand, the most crucial stage is the separation stage where the outcomes on the discriminant analysis are determined (Yan & Dai, 2011).

The main application of discriminant analysis is for the classification of objects, which may include but not limited to classification of things, people, customers, amongst others, into certain subsets of a given population. These subsets may be groups of two or more based on certain features of the population, such as, gender, income, height, etc. Generally, an object is assigned into subgroups based on the observed properties of the object. Hence, discriminant analysis can be said to predict the membership in a set or population based on the values observed on a number of continuous variables. This analysis specifically predicts a classification of dependent variables, with respect to given continuous independent responses. Therefore, it can be said that the discriminant analysis data consists of sample observations with given group membership alongside their continuous variables values.

For instance, considering an attempt to make a classification of a set of graduated students into two groups: students that graduated in five years or less and students who did not. The continuous variables could be chosen as undergraduate grade point average and examining the prediction model can give insights on each variable individually and in combination predicted the non-completion or completion of the program. Likewise, the classification can be made of loan applicants into risk categories: bad, moderate and good. Also, the continuous variables here can be income, age, years in current job and debt burden for the prediction of the credit risk category of each individual. Similarly, a predictive model can be developed for the classification of an individual into a certain risk category using discriminant analysis.

There are certain assumptions associated with discriminant analysis, one of these assumptions includes being highly sensitive to outliers and the number of predictor variables must be less than the size of the smallest group. Other major assumptions include multivariate normality, constant variance and homogeneous covariance matrices, independent variables should be correlated to the dependent variable only and not itself, linear relation between independent variables and all cases must be independent (Sajtos & Mitev, 2007). Interesting application areas of discriminant analysis is seen to be found in almost all scientific and traditional fields, ranging from health to social sciences, economy to industry, amongst many others (Cacoullos, 2014). It likewise has been introduced in recent interdisciplinary areas like data mining and pattern recognition (McLachlan, 2004). Certain features of a discriminant analysis platform includes the choice of fitting methods which ranges from linear, quadratic, regularized and wide linear. This births the following common discriminant analysis methods; Linear Discriminant Analysis (LDA), Quadratic Discriminant Analysis (QDA) and Logistic Regression (LR).

Under LDA we assume that the density for a specific variable X, given every class k is following a Gaussian distribution. In LDA, you simply assume for different k that the covariance matrix is identical. By making this assumption, the classifier becomes linear. QDA is not really that much different from LDA except that you assume that the covariance matrix can be different for each class and so, we will estimate the covariance matrix separately for each class. The only difference of LDA from QDA is that we do not assume that the covariance matrix is identical for different classes. On the other hand, logistic regression for classification is a discriminative modeling approach, where we estimate the posterior probabilities of classes given X directly without assuming the marginal distribution on X. Recall that the model of LDA satisfies the assumption of the linear logistic model. Therefore, the difference between logistic regression and LDA is that the linear logistic model only specifies the conditional distribution of the variable while the LDA model specifies its joint distribution.

The focus of this study however is on the implementation of LDA. Certain advantages attuned to LDA for data classification includes the ability to easily handle cases such as class frequencies within the dataset of an unequal where their performances are based on randomly generated data (Balakrishnama & Ganapathiraju, 1998). Similarly, when using LDA the limitation of all classes of the dataset having equal variance can be easily bypassed in a case where the confidence interval for the classes are defined differently (Khan Mohammadi, Garmarudi & De la Guardia, 2013).

Also, in a situation where the distribution of the samples are not known for certain, it is more advisable to use the easy to implement, straightforward LDA approach, since the adoption of the other discriminant analysis methods (QDA, RDA, LR) does not guarantee the best results from the dataset (Martínez & Kak, 2001). This is evident in the wide adoption of LDA in fields ranging from engineering (Jin, Zhao, Chow & Pecht, 2014), action and image recognition (Yan, Ricci, Subramanian, Liu & Sebe, 2014), applied statistics (Estoup, Lombaert, Marin, Guillemaud, Pudlo, Robert & Cornuet, 2012) amongst many others.

1.2 Linear Discriminant Analysis (LDA) Method

The LDA has been the mostly adopted technique of dimensionality reduction with the main aim of projecting a dataset on a lower-dimensionality space with a reasonable class separation to avoid overfitting and also reduce the cost of computation (Raschka, 2014). The LDA as introduced by Fisher (1936) is a very imperative and archetypal technique in discriminant analysis as it has good use in practical applications. Other discriminant analysis methods which can be used to generate discriminant functions include support vector, flexible discriminant, support vector machine (Santos, Guyomarch, & Bruzek, 2014; Arjmandi & Pooyan, 2012).

Originally, LDA was described for a two-class problem but was later generalized by Rao (1948) to multi-class linear discriminant analysis suitable for adoption in a multi-group setting or for the analysis of multivariate data. As aforementioned, LDA performs dimensionality reduction whereas time preservation on the class information is likewise possible. Hence, for the initial description of LDA for two classes, it is assumed that the N-dimensional set samples is given as $\left\{x^{(1)},\ x^{(2)},\ ...,\ x^{(N_i)}\right\};\ i=1,2$, whereas N_1 belong to class say ω_1 and N_2 to ω_2 which can obtain a scalar y by introducing x samples in the line. However, in its multi-class modification, Fisher (1936) generalizes for problems within the multi-class which are the type of problems under consideration in this study. Now, instead of a single projection y, we now aim to obtain say $(k-1)$ projections $[y_1,\ y_2,\ ...,\ y_{k-1}]$ for k classes using $(k-1)$ projection vectors denoted w_i (where $i=1,2,\cdots,k-1$) that can be rewritten by columns into a projection matrix $W=[w_1\,|\,w_2\,|...|\,w_{k-1}]$ such that $y_1 = w_1^T x$ which implies that $y=W^T x$. Details on the derivation for the Fisher's LDA will be discussed in the following chapter.

Since, the major goal of LDA is for the detection of the group of a test object within the class or the group with its closest mean. Thus, if population is of multivariate normal and the different groups have same covariance then LDA will perform better (Pohar, Blas & Turk, 2004). The fundamental concept of LDA is for the classification of an object or population to one or more classes or subgroups with respect to a measurement vector (discriminator) (Alrawasdeh, Sabri & Ismail, 2012). A combination of the available data sets, which are samples of objects, may be made for the enhancement of the understanding of the groups' differences.

In a practical environment, LDA parameters are unknown and there is a need to estimate these parameters from the sample data. Conventionally, the rules of discriminant analysis are most times based on the covariance matrix and the empirical mean of the sample data. However, discriminant analysis has a high level of vulnerability to outliers which is seen to be present in many real world multivariate data sets (Sajtos & Mitev, 2007). This is likewise the case when considering LDA, which due to the constraint that the LDA parameters are highly affected by outlying observations gives room for misclassification of new observations (Kim, Magnami & Boyd, 2006; Pires & Branco, 2010; Jin & An, 2011).

Other limitations of LDA includes not being able to produce more than $k-1$ projections, that is, if the classification error estimates state that additional characterizes are required, there is a need to adopt another method to provide these characterizes (Li & Yuan, 2005). Also, if the distributions are significantly non-Gaussian, none of the complex structures of the data which may be needed for classification will be preserved by the LDA projections (Yu, 2011). LDA will also not perform well if the discriminatory information is within the variance of the data and not the mean. These setbacks make part of the reasons why researchers ventured into introduction of robust estimators which will most importantly handle the presence of outliers (Cheng, Li, Lai, Song & Yu, 2016).

The major advantage for the introduction of the robust methods is due to the fact that most of the practical real world application data sets are made up a large number of variables and observations. In the process of treating these real-life data, certain irregularities termed as outliers are observed. The concept for robustifying LDA initially involved the replacement of the classical mean vectors and covariance matrices by its robust counterparts. This approach was called the plug-in method (Zollanvari & Dougherty, 2015).

There are numerous alternative estimators available to replace the classical sample mean vector and covariance matrix estimators. Some are more computationally extensive then others, and differences with regard to various theoretical properties may be exhibited as well. However, the simplest way to create a resistant multivariate estimator is to address each coordinate individually. Following the lead of univariate location estimation, the sample mean is replaced by the more resistant median for each of the variables. The dispersion matrix estimator becomes a covariance calculation that is centered by this coordinate-wise median. This coordinate-wise approach is adopted with a number of location statistics, such as the M-estimator, the Modified one-step M-estimator (MOM) and its winsorized counterpart as adopted in this research for example.

A wide range of robust estimators exist and have been adopted for handling the outliers in the data that the conventional LDA approach cannot handle (Filzmoser & Todorov, 2013; Todorov & Pires, 2007), ranging from the M-estimators, minimum volume ellipsoid (MVE), minimum covariant determinant (MCD) and S-estimators as introduced by Campbell (1980), Rousseeuw (1984, 1985) and Davies (1987) respectively. However, some other important robust location estimators were highlighted in the work of Haddad (2013) which effectively take into consideration the possible shapes of the data distribution. These includes the trimmed mean suitable for adoption to datasets with symmetrical distribution (Pei, 2002) and the Hodges-Lehmann location estimator (Brown & Kildea, 1978), a median based estimator that does not implement trimming to deal with outliers.

Past literatures have displayed the effective adoption of these robust estimators; the trimmed mean and Hodges-Lehmann (Abu-Shawiesh & Abdullah, 2001; Alfaro & Ortega, 2008). Although the latter (Hodges-Lehmann) is quite effective but only in cases where the distribution is just slightly heavier tailed because in situations where the distribution is heavily tailed, the performance is not as impressive. In same vein, the trimmed mean has high dependence on the percentage choice of omitted objects in the computation (Alfaro & Ortega, 2008) whereas it is important to ensure that trimming is performed in such a way that there is no loss of information (Othman, Keselman, Padmanabhan, Wilcox & Fradette, 2004) and this led to the introduction of the Modified One-Step M-Estimator (MOM) by Wilcox and Keselman (2003). The MOM estimator is an estimator that is suitable for adoption with datasets having asymmetrical distributions. The MOM estimator is a robust estimator with high breakdown point which performs trimming (deleting proportion of extreme values) with respect to the distribution as it empirically investigates the need to trim an observation or not, as well as the amount of trimming required.

These characteristics inform the use of MOM. However, since the focus of this work is to obtain more accurate analysis, the MOM estimator is winsorized hence also utilizing the Winsorized Modified One-Step M-Estimator (WMOM) (Ali, Yahaya and Omar, 2013; Haddad, Yahaya and Alfaro, 2012). Certain modifications will be made when adopting both the MOM and WMOM, which includes introducing the robust scale estimator (Q_n) and the winsorized covariance matrix.

When performing any statistical analysis, it is very insightful to adopt diagnostic procedures in assessing how efficient the analysis is and in this work the verification procedure will involve analyzing the misclassification error using cross-validation approach (Klaus, 2013). A certain rule will be stated to define the acceptable misclassification error and this analysis will be the yardstick in determining the improvement ratio of these newly adopted concepts in comparison to other existing results in literature.

However, this research intends to introduce new approaches for robustifying linear discriminant analysis with real life application data considered from previously existing literature. Comparison between the robust approaches and the classical LDA are made.

1.3 Problem Statement

LDA is one of the conventional approaches adopted in discriminant analysis and the procedures in these traditional approaches have been seen to suffer from a major lack of robustness, as the presence of outliers gives room for less accurate computations (Ayanendranath, Smarajit & Sumitra, 2004; Cheng et al., 2016). Due to the constraint that LDA parameters are highly affected by outliers which gives room for misclassification of new observations (Kim, Magnami & Boyd, 2006; Jin & An, 2011; Okwonu & Othman, 2013), obtaining robust alternatives to the discriminant rules has been put in place. This involves replacing usual estimates of parameters by robust ones (Ali & Yahaya, 2013; Yahaya, Ali & Omar, 2011).

The distribution mean which involves measuring the location and the variance-covariance which is about the shape measurement are the two statistics commonly implemented for data analysis when outliers are present (Ben-Gal, 2005; Rousseeuw and Leory, 1987). Also, the introduction of the multidimensional distribution parameters of the robust estimators will improve detection procedures in presence of outliers' performances. Earlier studies (Tiku & Balakrishnan, 1984; Campbell, 1980; Randles, Brofitt, Ramberg & Hogg, 1978; Ahmed & Lachenbruch, 1977) considered the use of robust estimator, but of low breakdown point estimators. The results of these estimators however were not encouraging and within a close interval of time, research started springing up to handle this drawback by adopting equivariant estimators with high breakdown point and this approach gave more encouraging results (Ali, Yahaya & Omar, 2015; Nkiruka, Onyeagu & Okeke, 2015; Lim, Yahaya, Idris, Ali & Omar, 2014; Croux & Dehon, 2001). Among the introduced concepts in robust LDA is the MOM and its winsorized counterpart known as WMOM (Yahaya, 2005). These two estimators are going to be the focus of this study with comparison being made to the classical form of analyzing linear discriminant models.

A well-adapted approach for handling outliers is the robust concept known as trimming. There are two type of trimming approaches namely symmetrical and asymmetrical trimming. The former approach is to trim symmetrically both left and right tail of the data based on the predetermined amount. On the other hand, the latter approach trimmed data based on the distribution of the data (Xao, Yahaya, Abdullah & Yusof, 2014). The amount of trimming is based on either predetermined or empirically determined. However, the predetermined trimming will be unnecessary when the data is normal, trimming based on the former approach will trim the data regardless of the shape, not in the case of the latter. One estimator using the latter approach is MOM. MOM is a location estimator which is winsorized for obtaining better results and hence the newly introduced WMOM (Haddad, Yahaya & Alfaro, 2012) which goes a step further from the MOM by performing a replacement of the largest and smallest values of the continuity of the consistent data after trimming has taken place using the MOM criteria (Haddad, 2013). Yahaya, Lim, Ali and Omar

(2016a, 2016b) adopted these concepts of trimming and winsorizing to deal with outliers where simulated and real financial data were used to test the performance of their proposed methods. In the separate woks, the authors estimated the covariance matrix for the robust estimators using a product of the conventional spearman correlation coefficient and the rescaled median absolute deviation to form a new robust discriminant rule. These new discriminant rules (Yahaya et al., 2016a; 2016b) performed better when the performance (in terms of misclassification error) was compared with the conventional LDA and existing modified robust LDA models in literature. However, it is important to also consider other approaches of developing new robust discriminant rules with better accuracy than previously existing robust models in literature. Therefore, in a nutshell this study will be implementing the coordinate-wise robust estimators in linear discriminant analysis by introducing high accuracy scale estimators with efficiency measured by analyzing the misclassification error using cross-validation approach.

1.4 Objectives of the Study

1. To develop modified LDA models using two estimators, that is, the Modified one-step M-estimator (MOM) and Winsorized Modified One-Step M-Estimator (WMOM) as the robust location and Q_n as the scale measure respectively.
2. To compare the performance of the proposed robust LDA with the classical LDA based on the average misclassification error and the computational time via simulation.
3. To apply the proposed robust LDA on real financial data.

1.5 Significance of the Study

This study is aimed at adopting two newly introduced location estimators for robustifying LDA referred to as the MOM and its winsorized counterpart WMOM. These estimators were chosen due to ability to perform trimming so well and thereby reducing the volume of information or data that goes to waste. Therefore, more encouraging results are expected to be obtained in comparison to the classical approaches and also other estimators adopted for robust LDA in previous literature.

The study contribution focuses on knowledge development in statistical classification techniques especially within LDA. LDA is extensively used in engineering, action and image recognition, applied statistics amongst many other vast research areas. However, when working with large number of quality characteristics the presence of outliers is unavoidable, and this situation will cause the result to be misrepresented. This problem can be rectified when robust LDA based on two robust estimators are considered, that is, MOM and WMOM. The advantage of using the proposed robust LDA is because there would be no restrictions by the normality assumption which is the conventional requirement of the traditional LDA. This implies that the original data can be worked with without bothering about the shape of the distributions.

1.6 Scope of the Study

This study considers robustifying LDA only. Likewise, just two robust estimators are considered, which are, MOM and WMOM. Two robust scale estimators are integrated in the modifications which are the robust Q_n estimator and the winsorized covariance matrix. The application of the LDA approach uses Fisher techniques in comparison to the newly constructed robust estimators. The application problem is limited to real financial data from just one selected field of study.

CHAPTER TWO
LITERATURE REVIEW

2.1 Discriminant Analysis

Discriminant analysis is used in multivariate analysis as a statistical tool for separating unique sets of observations or objects with respect to multivariate data (Härdle & Simar, 2012). This analysis is mainly used when assigning a data unit into a specific category or population group as a result of specific features or measurements. One of the traditional techniques proposed for discriminant analysis was done by Fisher (1936) which can be used to analyze group differences in multivariate study (Harlow 2014; Press, 2012; Betz, 1987). Basically, the analysis is used for explanatory purposes where there are need to identify, describe and comprehend unique differences among group members. It can also identify the set of continuous function that will be best characterized or capture group differences. Likewise, it is used to define group differences dimensionality which is similar to the definition of continuous variables dimensionality in factor analysis. Most importantly, it provides a comprehensive description of group differences of results obtained from variance multivariate analysis (Stevens 2012; Borgen & Seling, 1978). These aforementioned benefits of discriminate analysis proved it as an imperative tool for predictive analysis.

The predictive nature of discriminate analysis has made it as a useful tool for objects classification and categorization. This permits classification and categorization of group membership in form of predictor scores combination or best linearity outcome (Lee & Choi, 2013). There is similarity between multiple regression and discriminant analysis where both are used for continuous variable prediction in order to comprehend better differences in group members. However, multiple regressions differs to discriminant analysis in terms that multiple regression function on continuous criterion of variables while discriminant analysis function on categorized criterion with respect to group membership. In other words, multiple-regression is suitable when there is a continuous dependent variable whereas discriminant analysis is suitable for categorized dependent variable with more than two levels (Cohen, Cohen, West & Aiken, 2013).

In summary, it can be inferred that discriminant analysis has lots of applications to both applied and theoretical cases. The analysis can be used for behavioural dynamism and comprehension of changes dues to their group differences and nature. This analysis is based on an equation known as discriminant function for membership prediction as shown in studies such as Morrison (1976), Ender (2014) and Guh, Shiue and Yu (2014). These studies were all based on Fisher (1936) discriminant function that can distinguish two multivariate normal features with the aid of variance matrix. It can also be expended to new samples in order to identify high-risk members with a category or class (Damico, Nettleton, Damico & Nelson, 2014). An example of this is the prediction of dropout within an education intervention of those that will relapse during smoking cessation intervention (Kao, Lee & Tai, 2015).

2.1.1 Discriminant Function

From the above section, it has been established that discriminant analysis is used for predictive features and it produces linear functions which enable researchers to identify and comprehend members differences within their groups (Cohen et al., 2013). The generalized expression for linear equation is given as $Y = Xb + \varepsilon$ where ε is the error term, while the linear multiple regression is:

$$\hat{Y} = b_1 X_1 + b_2 X_2 + \cdots + b_p X_p + a \tag{2.1}$$

In the above equation, b_1 through b_p are known as the weights value of variable X, while a is the constant (as expressed in the Y – intercept within the regression line) where Y is the prediction continuous variable. For the prediction of Y, the least squared errors are used which is based on the selections of weights chosen to minimize the squared errors in the prediction for the variable of Y and the other variable of Y. In the same manner, the linear equation which is the fundamental part of discriminant analysis known as the discriminant function is given as:

$$D = b_1 X_1 + b_2 X_2 + \cdots + b_p X_p + a \tag{2.2}$$

This is the derivation of linear equation which is based on discriminant analysis where D the predictable is categorized variable and it defines other group membership. Ds can also be known as the group centroids which are formed based on each group maximize weight means using the discriminant analysis. In other word, this is the weights selected to handle the maximize ratio of the sum of squares within groups to the sum of the squares (Uray, 2008). In practice, it is found that similar variable have smaller weights whereas different variables have greater weights. It is important to observe that this technique focus more on group differences than group similarities.

Thus, it can be seen that a discriminant analysis produces a discriminant function which is applied to variables and the resultant weights pinpoint the important of each variable to its group differences and its contribution within the group. Likewise, the method depicts the significant function of a variable in a group and to each member's in the group (Vapnik, 2013). For explanatory purpose, associated error term and significant function are used which enhance group differences nature. On the other hand, calculation and comparison on each individual group centroids are done to determine the group member probabilities for predictive purposes (McGarigal, Cushman & Stafford, 2013).

As multiple regression equation is used for the calculation of prediction of score based on criterion on each subject, discriminant function is also used to compute discriminant score for individual members. Furthermore, the general express of linear equation ($Y = Xb + \varepsilon$) and it can be seen that each variable are multiplied with resultant discriminant weight on each scores. In addition, it can be seen that raw score multiplication with unstandardized weights will generate discriminant score as of the original variable. Similarly, a standardized weights multiplication with standard score will generate standard score unit discriminant (Harlow, 2014). Based on this fact, the calculation of discriminant score standardization can be obtained for the i^{th} individual as

$$D_i = b_1 X_{1i} + b_2 X_{2i} + \cdots + b_p X_{pi} + a \tag{2.3}$$

For better comprehension of the fundamental of centroid, assume that the dissemination of members' discriminant scores of a group is taken to be the set of continuous scores then there will be certain distribution with mean of standard deviation. This mean is referred as the group centroid and it is determined by introducing the weights' discriminant of the group means on each variable. Similarly, the discriminant function produces the maximize difference between the group centroids. Thus, minimize intersection between the scores groups' distributions. This pictures the important of the need to implement discriminant scores and centroids in the prediction of group membership and the major issue here is the enhancement of classification and categorization.

2.2 Linear Discriminant Analysis (LDA)

The previous section has introduced the fundamental concept, benefits and application of discriminant analysis. It has pointed to Fisher LDA as one of the classical approach used in discriminant analysis (Fisher, 1936). Thus, these sections deeper present the method of LDA with theoretical and application implications.

2.2.1 Fisher LDA

Basically in discriminant analysis the criteria for class separation are expressed in respect to within-class and between-class scatter matrices which is in the form of a covariance matrix and are invariant under coordinate shifts.

For a defined dataset $x \in \mathrm{R}^{m \times n}$ describing n samples $x_1, \cdots, x_n$ of c classes $C_1, \cdots, C_c$. Then, each sample $x_i \in R^m$ has an assigned class label $\omega \in \{\omega_1, \cdots, \omega_c\}$ describing to which class it belongs. All those expressions are combined in the class label list $1 \in R^n$. If the classes are well separated clusters in the original $m-$space having the same prior probabilities. Further assume that the classes are normally distributed $N\left(\mu_i, \sum_i\right)$, such that each class can be represented by its mean vector and its var-covariance (Uray, 2008). Let n_i represent the samples number in class C_i, that is $n_i = |C_i|$ such that $n = \sum_{i=1}^{c} n_i$. Then, the class means are given as

$$\mu_i = E[x \mid \omega_i] \tag{2.4}$$

and the expected the mixture distribution of the vector (the grand mean) is given as

$$\mu = E[x] = \sum_{i=1}^{c} P(\omega_i) \mu_i \tag{2.5}$$

The between-class scatter matrix is the expected vectors' scatter which within the mixture mean

$$S_b = \sum_{i=1}^{c} P(\omega_i)(\mu_i - \mu)(\mu_i - \mu)^T \tag{2.6}$$

The within-class scatter matrix shows the scatter of samples around their respective class expected vectors:

$$S_w = \sum_{i=1}^{c} P(\omega_i) E\left[(x-\mu_i)(x-\mu_i)^T \mid \omega_i\right] = \sum_{i=1}^{c} P(\omega_i) \sum{}_i \tag{2.7}$$

Thus, Fisher Discriminant Analysis is meant to minimize the distance within the classes and also maximize the distance between the classes simultaneously (Uray, 2008). That is, the Fisher Criterion describes the class separated by

$$J = tr\left(S_w^{-1} S_b\right) \tag{2.8}$$

where J is invariant under any non-singular transformation. For this reason it is possible to simplify the task by optimizing J in a much lower-dimensional space. The linear transformation from a $m-$dimensional space X to a $n-$dimensional space Y with $n < m$ is generally described as

$$Y = W^T X \tag{2.9}$$

where the projection W is a $m \times n$ matrix with linearly independent columns. The scatter matrices in the projected space can be calculated as

$$\tilde{S}_w = W^T S_w W \tag{2.10}$$

and

$$\tilde{S}_b = W^T S_b W . \tag{2.11}$$

2.2.2 Limitations of LDA

Previous section has given better understanding and comprehension into LDA concepts and application. This section will examine issues and drawbacks associated with LDA. The section will further describe problem and efforts made by previous studies on it.

2.2.2.1 Small Sample Size Problem (SSS)

In the usage of LDA, singularity is usually defines scatter matrix within-class S_w and the needs for this is due to the measurement of the sample space which is typically higher than the sample values of the training set. To resolve this concern within SSS, Belhumeur, Hespanha and Kriegman (1997) and Swets and Weng (1996) employed the use of PCA as dimension reduction for the calculation of LDA. This was done to reduce the dimensional measurement of the space feature to $n-c$ that S_w will not be singular afterward. Subsequently, LDA is implemented in the reduction of the space to obtain the intended discriminative information resulting in total dimensionality reduction to $c-1$. This can be seen in Belhumeur et al. (1997) study where the resultant face images and Eigen images are known as the Fisher faces.

Furthermore, Yang and Yang (2003) proposed that the derivation of null space of S_w without reduction or attenuation in optimization of information discrimination can be achieved by optimizing the discriminant vectors which will resolve the concern. This study gave the idea of LDA in the null space of S_w where the first computation of the null space of S_w is achieved and the second S_b is estimated. The eigenvectors analogous to the largest eigenvalues of the estimated between-class scatter matrix are chosen as final transformation. The introduction of Direct LDA (D-LDA) approach by Yu and Yang (2001) is meant to resolve this concern. This concept made use of discarding the null space of S_b whereas the important null space of S_w is retained. This is done by inversely replacing the diagonalized orders of S_w and S_b. The calculation of the Fisher criterion can be achieve directly. In addition, the introduced D-LDA is equivalent to subspace-based LDA and aid the resolution of the issue in SSS. Nevertheless, Gao and Davis (2006) argued that D-LDA is just a distinctive case of LDA and subspace-based LDA equivalence cannot be assured most especially for classes that are not well separated.

Additionally, Cevikalp, Neamtu and Barkana (2005) proposed Discriminative Common Vectors to resolve this issue of SSS for LDA. It made use of optimal projection vectors in the null space of S_w to overcome computational difficulties issues. This is achieved by removing the difference between the samples in each class in order to extracting common properties of classes within the training set. However, this approach demands the need for dimension of the sample space must be larger than the rank of S_w which is usually a big issue in SSS. This implies that for a good result the null space must be big which means the bigger the null space the better the results. Based on this, it can be concludes that success of this approach largely depends on the size of the null space of S_w. Similarly Uray (2008) developed Regularized LDA (RLDA) approach to resolve this issue. The approach made use of a diagonal matrix kI (where I is the identity matrix and $k > 0$) as an addition to the class of scatter matrix. It defines S_w as a positive which is expressed as $S_w + kI$ and is always non-singular. Therefore, any small perturbation to the scatter matrix within-class will not change or affect the resultant matrix projection. Moreover, the unique advantage of this approach is based on its simplicity and ability to still keep full discriminative information.

2.2.2.2 Overfitting or Underfitting

Another important issue in LDA has to do with its fixed values of freedom degrees (fixed number of free parameters). The main issue is that model parameters do not always generalize accurately while new data cannot be used to describe it properly. This issue can be viewed in two dimensions namely overfitting and underfitting.

The issue of overfitting usually occurs in three different cases. First, when the statistical model has too many parameters which imply that the given degrees of freedom within the parameter selection exceed the information given by the available data. Secondly, when there is not enough training data and the model is based on specific features which are not represented within test data. Thirdly, training data is highly correlated and multicollinearity in nature. In those three cases, the eigenvalues are closely together for the eigenvectors to have a small eigenvalues which can be unstable in nature. On the other hand, underfitting is usually occur when statistical model has few parameters. This implies that the information given by the data cannot be described based on the degree of freedom of parameter selection. Thus, the linearity is not achievable with this class boundaries cases and it calls for a robust classifier to make it more achievable in nature.

There are many studies in the vast literature that implement similar approach in their problem solving applications such as Yang et al., (2004). They implemented Kernel LDA approach based on Kernel Principal Component Analysis (KPCA). This approach utilizes kernel functions to derive high-order correlation randomly among the input variants (Lu et al, 2004; Lu et al., 2003) to overcome the issues of linearity within LDA in multi-class cases. This was achieved by inventing the KFD (KPCA plus LDA) algorithm within LDA in multi-class cases. Furthermore, the concept of Complete Kernel Fisher Discriminant (CKFD) was introduced based on the fusion of regular and irregular discriminative information where regular discriminant information is denoted with range space of S_w. According to Yang et al (2005) this is known as the irregular discriminant information which is S_w of the null space.

Likewise, Kim and Kittler (2005) developed a Locally Linear Discriminant Analysis (LLDA) which can handle linearity issues within LDA. This is to ensure that global non-linear structures are localized linearly and local structures can also be linearly aligned. It implies that locally linear functions can be transformed to a cluster with small possibility within a maximal class object differences in a single class objects that is multimodally distributed. LLDA maximizes the separation of classes locally for the promotion of consistency between the multiple local representations of single class objects. LLDA has a high computational efficiency and overfitting reduction which is the major advantage it has over KFD whereas both can achieve similar results.

2.2.2.3 Distribution Assumption

The LDA solution is only optimized statistically if the distribution of the various class samples is Gaussian with same covariance matrices but different means. However, the introduction of distribution assumption can make a huge difference. For instance assumption based on small approximation of training samples number can cause inaccuracy in the result. This approximation can be in three different forms which include the true covariance, the sample mean and the true mean by the sample covariance. This will introduce outlying data with negative influence error into the model.

Based on the data structure, it is possible to achieve mixed model using only one transformation matrix over the whole data. For the creation of mixture model, set of classes will be partitioned into several clusters. Calculation is based on each cluster to get an appropriate transformation matrix. In LDA, the mixture model is introduced with the PCA mixture model with each class with K mixture components (Kim et al., 2001; 2003). The concept is to express the n-dimensional data with the combination of densities partitioned cluster components whereby Gaussian function modelled the density conditional function. The result produces mixture of each component cluster mean with the transformation matrix and diagonal matrix eigenvalues components of covariance matrix. The outcomes give individual scatter matrices for each component which made it possible to determine label classes for each component whereas the Fisher criterion will be maximized for each cluster independently. This will make it possible to project onto each LDA space and assigned these components to the nearest classes (Kim et al., 2003).

In another study by Torre and Black (2001), Multimodal Oriented Discriminant Analysis (MODA) was proposed as an extension of LDA to handle multimodal

Gaussian distributions of classes of different covariances. Probabilistic interpretation was used for this case of optimal discriminant analysis which has different multimodal and covariances distributions. This is obtained by maximizing the classes of the linear transformation which defines the difference between two Gaussian distributions the Kullback-Leiber (KL) divergence employed (Fukunaga, 2013). Afterward, the training data is clustered in each class with aids of multi-way normalized cuts (Yu and Shi, 2003). Thus, MODA can be projected to maximize the KL divergence between different classes' clusters and not between same classes' clusters.

However, Loog, Duin and Haeb-Umbach (2001) pointed the issue of distribution assumption as a limitation of LDA. This established the unsuitability of Fisher criterion within multi-class cases and can be expressed based on the scatter matrix between-class in terms of class mean differences in decomposition as given by $c-$ class Fisher criterion into $\frac{1}{2}c(c-1)$ two-class Fisher criteria. From this expression, it can be seen that classes lying far from others are overemphasized which create overlapping issues with other classes. Therefore, a weighting scheme based on the Mahalanobis distance among different classes is required whereby the distribution of the classes depends on the Bayes error rate among the different classes. It implies that the Bayes error among different two classes largely depend on the Mahalanobis distance. Although, the outlier classes can be considered accurate based on the weights however, due to various estimation involved the outcome is not guaranteed to be optimal. Many studies have proposed verification method to establish optimal results by adopting weighting function from dissimilar classes (Tang et al., 2005; Li et al., 2006) adopted the idea of weighting class pairs, at which the weighting function relates to the dissimilarity between two classes.

2.3 Multivariate Outliers

To give a specific definition to outliers is dependent on certain hidden assumptions with regard to the feature of the data and the method applied for detection, although some definitions have been generalized to encompass a number of methods with variations in data (Ben-Gal, 2005). Certain definitions given in literature includes defining an outlier as an observation with large deviation from other observations in the data making it suspicious that such observation was generated by a different mechanism (Hawkins, 1980). Another definition is by Barnett and Lewis (1994) indicating that an outlier which is also referred as outlying observation is the member that appears to be differ significantly among other sample members within a group which is similar to the definition in Johnson (1992) defining an outlier as a form of observation within a data is appearing inconsistent in comparison to the rest of that data set. However, in recent literature Ali (2013) described an outlier generally as a point(s) that differs surprisingly from the remaining data set. One of the problems caused by the presence of outliers in data sets is non-normality and the number of outliers in a sample data will simultaneously increase with an increase in the dimensions of the data.

The detection of outliers is obtainable when multivariate analysis is conducted and the interactions within the different variables are compared in the data class because outliers cannot be detected easily when the data dimensions is multivariate (Ali, 2013; Ben-Gal, 2005). Datasets having clustered or multiple outliers are subjected to the effects of swamping and masking. Acuna and Rodriguez, (2004) gives an insightful, non-rigorous, easy to comprehend mathematical definitions of these effects as described in the figure below.

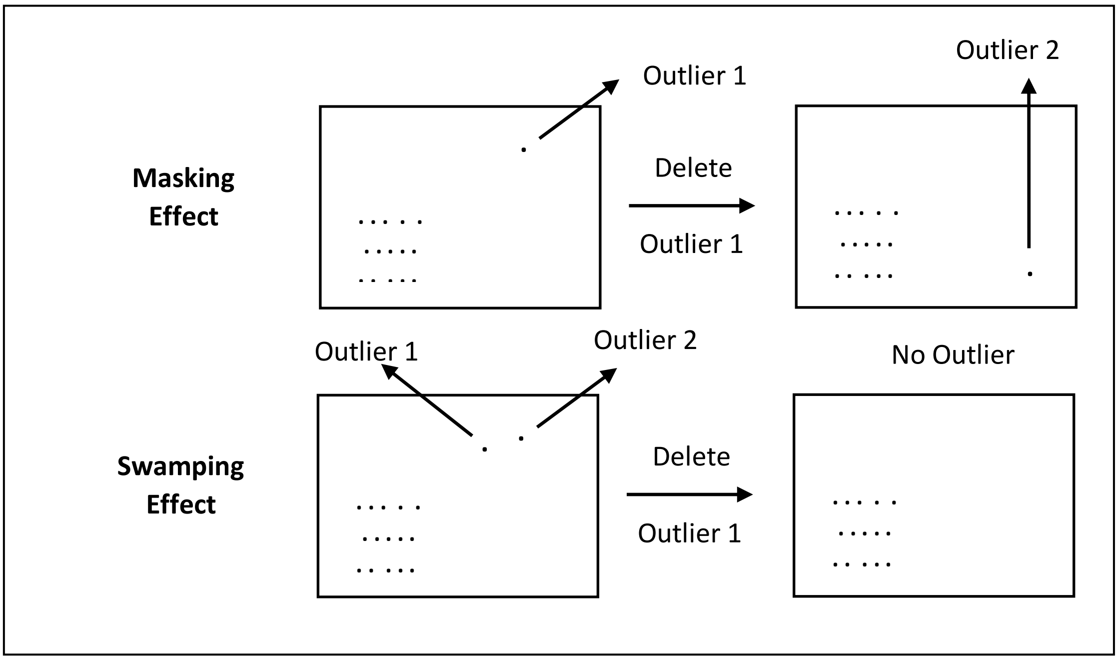

Figure 2.1: Masking and Swamping Effects on Outliers (Acuna & Rodriguez, 2004)

When considering masking effect, what happens is that an outlier masks another outlier, if the later outlier is being considered as an outlier on its own, but in the absence of the earlier outlier. Therefore, after the earlier outlier is deleted, then there is emergence another outlier which is the later one instance. Masking is said to have occurred when the mean and covariance estimates are skewed by a group of outliers towards itself. This make the distance from the mean to the outlying position so small it becomes almost impossible to find the outliers.

For swamping effect on the other hand, the concept follows that an outlier swamps another outlier, in the case where the later outlier can only be regarded as an outlier when the earlier outlier is present. This implies that, upon deletion of the earlier outlier, the later one become a non-outlying observation. The case of swamping comes into existence when a group or cluster of outliers skew the mean and covariance estimates to itself while simultaneously skewing away from other non-outlying observations, where the distance between these occurrences to the mean is huge, creating seemingly outliers which gives an incorrect declaration to the outliers (Ali, 2013; Ben-Gal, 2005; Iglewics and Martinez, 1982).

There can be a resolution of the problems associated to masking and swamping by the use of robust estimates of location and scatter which are less prone to outliers. Some of these resolutions were proposed far back in literature by Beckman and Cook (1983) and Rousseeuw and Leroy (1987) on the use of robust estimates. It was mentioned in their individual works that robust estimation is one of the best approaches in handling outliers as the use of these robust estimates most of the times triggers an improvement in the performance of the process of detecting the outliers in the system (Ali, 2013). Thus, adopting robust estimation is encouraged and hence it is necessary to develop robust estimates to avoid these errors form having negative impact of the model or analysis.

2.4 Misclassification Error

In discussing unconditional misclassification error, which is independent of data, on a population sample commonly referred to as misclassification frequency in literature (Shao, Wang, Deng & Wang, 2011; Dabney & Storey, 2007). Assuming a *d*-dimensional data *x* with class *k*, group-specific centroids μ_k and a common covariance matrix Σ. Consider $x^{(k)}$ as vector sample taken from the distribution of the multivariate normal $N\left(\mu_k, \Sigma\right)$ associated with class *k*. Thus, the conventional algorithm which is assigned to the highest score yielding class as defined by (Klaus, 2013):

$$d_k\left(x\right) = \mu_k^T \Sigma^{-1} x - \tfrac{1}{2} \mu_k^T \Sigma^{-1} \mu_k + \log\left(\pi_k\right) \tag{2.12}$$

Using the scalar product of

$$\Delta_k\left(x\right) = \left(\omega^{(k,pool)}\right)^T \delta_k\left(x\right) + \log\left(\pi_k\right) \tag{2.13}$$

where $\delta\left(x\right)$ is a Mahalanobis transformed variable and $\omega^{(k,pool)}$ is the corresponding weight variable defined in Klaus (2013). A misclassification on the population level of $x^{(k)}$ in (2.13) occurs if between any two classes *k* and *l*,

$$\left(\omega^{(k,pool)}\right)^T \delta_k\left(x\right) + \log\left(\pi_k\right) < \max_l \left[\omega^{(l,pool)}\right]^T \delta_l\left(x^{(k)}\right) + \log\left(\pi_l\right) \tag{2.14}$$

which is equivalent to the condition

$$\min_{l \neq k} \frac{\left[\omega^{(k,l)}\right]^T \left[P^{-1/2} V^{-1/2} \left(x^{(k)} - \frac{\mu_k + \mu_l}{2}\right)\right] + \log\left(\dfrac{\pi_k}{\pi_l}\right)}{\sqrt{\left[\omega^{(k,l)}\right]^T \left[\omega^{(k,l)}\right]}} < 0 \tag{2.15}$$

P and *V* are resultants from the decomposition of the covariance matrix Σ, where *V* contains the variances $\left(V = diag\left\{\sigma_1^2, \cdots, \sigma_d^2\right\}\right)$ and *P* is the correlation matrix $P = \left(\rho_{ij}\right)$. As $x^{(k)} \sim N(\mu_k, \Sigma)$ holds for all $k \in \left\{1, \cdots, K\right\}$, then unconditional (which is expected) probability of misclassifying a sample taken as class *k* into a wrong class $j \neq k$ can be deduced from the above formula as:

$$P\left(j \neq k \mid k\right) = \Phi\left(-\min_{l \neq k} \frac{\left[\omega^{(k,l)}\right]^T \left[\omega^{(k,l)}\right] + 2\log\left(\frac{\pi_k}{\pi_l}\right)}{2\sqrt{\left[\omega^{(k,l)}\right]^T \left[\omega^{(k,l)}\right]}}\right) \tag{2.16}$$

This translates into a misclassification error (total probability error) given to be

$$P\left(error\right) = \sum_{k=1}^{K} P\left(j \neq k \mid k\right) \times P\left(k\right) \tag{2.17}$$

$$= \sum_{k=1}^{K} \Phi\left(-\min_{l \neq k} \frac{\left[\omega^{(k,l)}\right]^T \left[\omega^{(k,l)}\right] + 2\log\left(\frac{\pi_k}{\pi_l}\right)}{2\sqrt{\left[\omega^{(k,l)}\right]^T \left[\omega^{(k,l)}\right]}}\right) \times \pi_k \tag{2.18}$$

Observe that the final equation obtained is the outcome of adopting an expectation operator two times. The first application is with respect to the model parameters $\omega^{(k,l)}$ and the other application is with respect to the transformed data $\delta_k\left(x^{(k)}\right) - \delta_l\left(x^{(k)}\right) = P^{-1/2} V^{-1/2}\left(x^{(k)} - \frac{\mu_k + \mu_l}{2}\right)$. The first application obtained the population version of the intended statistical model and the second outcome will be an unconditional (notwithstanding the dependent on the data) error rate (Klaus, 2013).

2.5 Trimming

Consider a set of observations $\left[x_1, x_2, \cdots, x_n\right]$ arranged from the smallest to the biggest where the trimmed mean is represented as $k\%$.Let $\overline{x_t}$ denote the average of the remaining values after removing the smallest and largest $k\%$ observations from the original set. The variance of the trimmed mean is defined by $\frac{s_w^2}{n \cdot \left(1 - 2\gamma\right)^2}$ where s_w^2 is the winsorized variance, n is referred as the untrimmed sample size while γ is given as trimming proportion whereas the trimmed standard deviation is given as $\frac{s_w}{\sqrt{n}\cdot(1-2\gamma)}$ (Haddad, 2013).

To obtain the best amount of trimming is dependent on specific conditions relative to the data although Wilcox (2003) recommended an amount of 20%. Bearing in mind that when the sample size is smaller, the trimming would be less (say, 10%) (Bennett, 2009).

When adopting the usage of the trimmed means, removal of the outlying observations on the predetermined amount simultaneously reduces that effect of the tails of the distribution where the predetermined method for trimming is used by the common trimmed mean. Adopting this approach, about 10% to 20% can be trim on both observation tail sides. When considering a normal distribution or a distributed light-tailed however, it may be more suitable to not trim at all and if there is need to trim, just a few observations should be trimmed (Yahaya, Lim, Ali & Omar, 2016a).

A large number of researchers have discussed about the trimming method that adopted the concept of just the predetermined amount of symmetric trimming as seen in the works of Keselman et al., (2002) and Wilcox (2003). In a case where there exist skewed distributions, then there should be a difference on the amount of trimming per tails, which implies that there should be more trimming on the skewed tail. Although, in a situation where predetermined symmetric trimming is adopted, then irrespective of the shape of the tails, symmetrical trimming is conducted.

Keselman et al., (2008) adopted asymmetric trimming with particular case of application to hinge estimators as initially introduced by Reed and Stark (1996). This asymmetric trimming was adopted for determining the right amount to be trimmed from each tail of a distribution. Bearing in mind that the method in Keselman et al., (2008) still used predetermined trimming percentages. From the results obtained, the breakdown point of trimmed mean is same to the percentage of trimming and this implies that the trimmed mean is not suitable for handling large data of extreme value. Hence the trimmed mean is not so robust. In comparison of trimmed means with the means of the actual data, it was observed that the power of the trimmed mean procedure increased drastically (Wilcox et al., 2002).

2.6 Robust LDA

As discussed in previous sections, the major aim of LDA is the prediction of variables within group members. Rousseeuw (1985) and Hastie et al. (1995) mentioned that LDA is based on assigning an unidentified prediction within a group with minimum classification error rate obtained. This implies that LDA will only perform optimally if the equality of variance covariance matrices and normality assumptions are not violated (Kim, Magnani & Boyd, 2006). Additionally, studies have proved that classical LDA relies on sample covariance matrix and mean that are sensitive to outliers (Hubert & Van Driessen, 2004; Hubert, Rousseeuw & Van, 2008; Croux, Filzmoser, & Joossens, 2008; Ella, Van & Williem, 2009). It also shown that misclassification is gotten due to susceptibility of the classical sample covariance matrix and sample mean to outlying observations (Jin & An, 2011).

Robustness of an approach or method is basically determined by its property of sensitive to outliers. This is defined by the common measure for robustness which the breakdown point is defining the amount of noise an estimator can handle (Joossens, 2006). The estimator breakdown point is the portion of the dataset which is totally polluted without disturbing the estimator. The measurement of robustness which lies between 0% and 50% is given as the breakdown point. Therefore, it can be concluded that the higher the breakdown point the more robust is the method.

There are many studies in the vast literature that examined the issue of robustness of LDA. These approaches have been implemented to use estimators such as studies by Campbell (1980), Maronna (1976), Davies (1987), Lopuhaa (1989), and Rousseeuw and Van Drissen (1999). Robust multivariate estimators have been used by inserting it into classical estimators to produce robust multivariate approach that is more useful for outlier diagnostics rather than high breakdown robust approaches. Thus, robustness classical method involves the detection of outliers and deletion of breakout from the datasets. Past literatures like Fung (1995, 1996) have proposed approaches to analyze outliers when considering discriminant analysis. The concept follows a removal of the outliers from the dataset and then adopting the conventional discriminant analysis with respect to the observations left making the classical method a robust one. It is worth taking note of that in the detection of these observations, the use of robust estimators is important in order to bypass the masking effect. This affirms the well known fact that analysis based on non-robust estimates not in all cases recognizes the presence of all outliers (Joossens, 2006).

One key importance in adopting robust estimators is for the identification data deviation or detection of the presence of outliers. Outliers are seen to influence or affect the results obtained from classical discriminant analysis due to the fact that the discriminant rules rely solely on estimates of parameters of the population. Generally, the estimated means is shifted by the outliers which simultaneously and blow up the dispersion matrices. In comparison to conventional classical statistical estimation approaches, robust estimation gives more vivid distinctions between a good data and an outlier. This is because conventional statistical approaches faintly display the presence of outliers in the dataset (Hampel, 2001). As aforementioned, robust estimators can be applied in two major ways, where the first involves removing the outliers while the second approach directly substitutes the robust estimators (Beckman & Cook, 1983). The former approach is used in this research. In the review in the following subsection, a number of robust estimators are discussed from the wide range of robust estimators of multivariate data in literature available (Maronna & Zamar, 2012; Maronna et al., 2006).

2.6.1 Robust Estimators

Robust estimators were suggested in early literatures by Huber (1964) because the normality assumption needed to be satisfied in conventional estimation methods, most time does not hold. Later on, Hampel (1985) highlighted that the main objective of applying robust estimation is to observe the divergence in the data or outliers. This is because, in comparison to conventional approaches, robust estimates distinguish more vividly between a useable data and one affected by outliers unlike classical approaches where the difference cannot be spotted most times (Yahaya, Lim, Ali & Omar, 2016b). There are two approaches to adopting robust estimation methods. The initial concept follows identifying and removing the outliers, then classical estimation approaches is used on the useable data left, while in the second concept, the robust estimators are applied directly instead of classical estimation methods (Beckman & Cook, 1983). It is worth noting that in finding and selecting robust estimators, there are certain properties these estimators are expected to possess and this is discussed in the following subsection.

2.6.2 Properties of Robust Estimators

To determine how good or useful a robust estimator for multivariate data is, four main properties can be implemented to assess the estimator robustness which includes the affine equivariance, statistical and computational efficiency of the estimator and the breakdown point (BP) (Jensen et al., 2007). The BP as presented way back in literature (Hampel, 1974; 1971) is the quantity for analyzing how robust an estimator is in the presence of outliers. The BP of an estimator is the smallest fraction of the observations that have to be replaced to make the estimator unbounded (Rousseeuw, 1991). In this definition, one can choose which observations are replaced, as well as the magnitude of the outliers in the least favourable way. Another definition in Ali (2013) described the BP as the least ratio part of outliers which can result an estimator to take on arbitrary higher ration values. A higher BP implies a more robust estimator which simultaneously has less chances of being affected by the masking effect. A number of authors have stated that in practical and real life applications, if the BP is higher than or equals 20%, then it is not acceptable (Zuo, 2006).

The second measure that can be used to assess a robust estimator is the affine equivariance. This property is a very desirable and important property when considering statistical estimations. This is due to the fact that when an estimator, be it robust or not, is affine equivariant, a change in the scale of measurement or its affine transformations is not expected to have an effect on the properties of the estimator.

The third measure is in relation with the statistical efficiency of the estimator. The focus of this property has to do with how efficient the estimator adopts the usable data made available, as efficiency is a very significant quality to be considered for any mathematical or statistical procedure (Zuo, 2006). In the work by Huber (1964), the issues surrounding robustness and efficiency of statistical approach was taken into consideration with focus on the widely known minimax approach which follows the concept of minimizing worst-case asymptotic variance. Generally, robust estimators are not the most efficient estimators and a good example is the univariate median. Although this estimator has the best BP and is also the most robust affine equivariant location estimator with the least maximum bias at symmetric distributions (Huber 1964). It is weakly efficient with respect to its mean for normal or other light-tailed models irrespective of its outstanding robust properties.

The last measure or property a robust estimator is expected to possess is in terms of its computational efficiency as this translates to easier and faster computations. Just like the measurement of data in either megabytes or gigabytes is a normal norm, certain real life or practical applications expected the detection of outliers to occur within few seconds of minutes. This becomes a major issue for consideration when robust estimators are considered for large multivariate data. Data of this class are seen to arise in areas including quality control, healthcare, information, machinery, financial and agriculture because all these industries mentioned deal with products translating into multi-dimensional data. Therefore, when robust estimators are applied, the computational efficiency becomes low with large datasets as the computational burden (in terms of time and cost) is very high. Pena and Prieto (2001) suggested the development of special methods to deal with these special cases is suitable. Therefore, for large datasets, achieving computational efficiency may be a bit difficult. It still remains very germane in effectiveness to the detection of outliers (Angiulli and Pizzuti, 2005).

2.6.3 Types of Robust Estimators

There are in existence a large number of location and scale robust estimators for multivariate data. A number of these estimators follow the concept of the robust scale of Mahalanobis distance, and such estimators include the M-estimator (Campbell, 1980), MCD (minimum covariance determinant), MVE (minimum volume ellipsoid) estimators (Rousseeuw 1984, 1985), *S* and *T* estimators (Davies, 1987; Lopuhaä & Rousseeuw, 1991). Some other robust estimators follow the concept of projections, such as the Stahel-Donoho estimate (SDE), Kurtosis1 (Pena & Prieto, 2001) and *P* estimates (Maronna, Stahel, & Yohai, 1992). Large attention and focus has been placed on the MVE and MCD estimators introduced by Rousseeuw (1984; 1985).

2.6.3.1 Modified One-Step M-Estimator (MOM)

One of the latest additions to the family of robust statistics is a measure of central tendency modified from the one-step M-estimator referred to as modified one-step M-estimator (MOM). The MOM estimators like trimmed means can be used for equality of measurement of the typical scores across treatment groups. Besides the drawback of lower BP for trimmed means, another concern is the amount of trimming is usually fixed prior to data analysis. One approach to this problem is to ponder on the degree of the trimming as a function of the observations, and one-step M-estimators represent this approach. On the other hand, the one-step M-estimator empirically determines whether an observation should be trimmed, or the possibility of no trimming as well as different amount of trimming in the left versus the right (Wilcox & Keselman, 2003).

MOM is just the average of the values left after all extreme values (if any) are discarded. This estimator is derived from the one-step M-estimator (Staudte & Sheather, 1990) after some modification. Mathematically, Wilcox and Keselman (2003) defined the MOM estimator as

$$\hat{\theta}_j = \frac{\sum_{i=i_1+1}^{n_j-i_2} Y_{(i)j}}{n_j - i_1 - i_2} \tag{2.21}$$

where $Y_{(i)j}$ is the i^{th} order observations in j^{th} characteristic variable,

i_1 is the number of $Y_{(i)j}$ that satisfies the criteria $\left(Y_{(i)j} - \hat{M}_j\right) < -K^* \left(\text{scale estimator}\right)$,

i_2 is the number of $Y_{(i)j}$ that satisfies the criteria $\left(Y_{(i)j} - \hat{M}_j\right) > K^* \left(\text{scale estimator}\right)$,

n_j is the size of the data set for each variable,

$\hat{M}_j$ is the median of the data in each j^{th} variable,

and the scale estimator is the median absolute deviation (MAD_n). The constant $K = 2.24$ is motivated to give a good efficiency for the robust scale estimators (MAD_n) when the sample is taken from a normal distribution (Othman et al.,2004).

2.6.3.2 Winsorized Modified One-Step M-Estimator (WMOM)

In WMOM, the trimmed observations are replaced by the highest and the lowest values of the remaining data. The WMOM is one of the central tendency measurements. It is the arithmetic mean, which is resulted by replacing outliers from each end of the data with the next largest and smallest values of the continuity of the consistent data after performing the trimming by using the MOM criteria (Ahmad Mahir & Al-Khazaleh, 2009). According to Wilcox (2003), the mean is one of the most popular location estimators. Nevertheless, there is a problem concerns with this measurement in the tail of the distribution that may affects its value. This problem becomes clearer through the unbounded influence function of the BP of zero, where the influence function measures how the estimator reacts to the small proportion of outliers and the lack of robustness against outliers Hampel (1974).

To solve this problem, more attention should be given to the value of means whenever they are nearer to the center. Therefore, the advantage of using the robust measures of the center, winsorized MOM is that it can be used instead of the usual mean, which is thought to be another easier solution for tackling the sensitivity of the mean. As winsorized mean is known to be less sensitive than the mean but still give a reasonable estimate of central measure (Wilcox & Keselman, 2003), thus, it is assumed that winsorized MOM should also perform better than the usual mean.

The construction of the Winsorized sample is proceeding as follow (Wilcox, 1997). For each random variable $X_j = \left\{x_{1j}, \cdots, x_{mj}\right\}$, $j = 1, \cdots, p$, the winsorized sample is obtained from

$$w_{ij} = \left\{ \begin{array}{ccc} x_{(i_1+1)j} & & if \ x_{ij} \le x_{(i_1+1)j} \\ x_{ij} & if \ x_{(i_1+1)j} < x_{ij} < x_{(m-i_2)j} & \\ x_{(m_j-i_2)j} & & if \ x_{ij} \ge x_{(m-i_2)j} \end{array} \right\} \tag{2.22}$$

where

i_1: number of the smallest outliers data.

i_2: number of the largest outliers data.

Therefore, the estimated winsorized MOM for *l-th* variable as follows (Haddad, 2013):

$$\overline{w_{ij}} = \frac{1}{m_j}\left[\sum_{k=1}^{m_j} w_{ij}\right] \qquad (2.23)$$

2.7 Scale Estimators

There are fewer studies in literature that considers robust scale estimators in comparison to studies that have considered robust location estimators. Irrespective of this, scale estimators are still very relevant in statistical applications as these scale estimators can be used in descriptive statistics for data analysis, in comparison for variations in datasets**,** construction of confidence intervals**,** standardization of observations, formulation of concepts for outlier detection**,** as objective functions in regression, as auxiliary estimates for location M-estimators amongst many others (Rousseeuw & Croux, 1992). It is very essential that in adopting the scale estimator to these application areas, there is no breakdown of the estimator either in tending to zero (imploding) or in growing so large (exploding). Normally, the expected BP is about 50%, which means that there may be replacement of almost half of the data before the estimate becomes non-useable. This is not like the sample standard deviation where the presence of just one outlier can be very problematic.

Similarly, Yahaya et al., (2004) study suggested four scale estimators which include Q_n, S_n, T_n and the most implemented robust scale estimator in the vast literature, MAD_n. The rationale for this is because of its high BP. Likewise, it has high bounded stimulus function which is capable of sustaining the robustness. For the following sections reviewing these notable scale estimators, let $X = (x_1, x_2, \cdots, x_n)$ this can be a random sample while the sample median can be represented with $med_i x_i$. The following subsection will further discuss about the Q_n scale estimator which is the chosen estimator integrated with the robust estimator.

2.7.1 Q_n

Rousseeuw and Croux (1993) and Croux and Rousseeuw (1992) proposed two measures of scale as which is a form of alternatives for median absolute deviation and Gini's mean difference known as the S_n and Q_n estimators. However, the focus of interest is the Q_n estimator. This estimator has been identified to have a BP of about 50% and a smooth bounded influence function. For a model distribution F which has a density f which is the influence function for Q_n and represented with

$$F\left(x;Q,F\right)=d\,\frac{\frac{1}{4}-F\left(x+d^{-1}\right)+F\left(x-d^{-1}\right)}{\int f\left(y+d^{-1}\right)f\left(y\right)dy} \tag{2.27}$$

The Q_n-estimator also possesses large gross-error sensitivity in comparison to other estimators. It is not dependent upon a measure of location and calculates the distance of each shot length from every other shot length, hence being appropriate for asymmetric distributions. Properties of the Q_n estimator include it having an explicit and simple formula, which is equally appropriate for distributions of asymmetric. Considering the case of a normal distribution, the Q_n-estimator asymptotic variance efficiency is seen to be higher than other estimators, therefore making it more efficient. This is seen in the simulation study by Rousseeuw (1991). The square of the Q_n-estimator, that is, $(Q_n)^2$ can be used as an estimate of σ^2. Even though both Q_n

and $(Q_n)^2$ are biased estimators of σ and σ^2 respectively, they are efficient estimators of their respective targets.

The estimator Q_n for a random sample $X_1, X_2, \cdots, X_n$ with model distribution F is defined as:

$$Q_n = 2.2219\left\{\left|X_i - X_j\right|;\ i < j;\ i = 1, 2, 3, \cdots, n;\ j = 1, 2, 3, \cdots, n\right\}_{\{k\}} \tag{2.28}$$

where $k = \binom{h}{2}\binom{n}{2} \Big/ 4$ and $h = \left[\frac{n}{2}\right] + 1$ which is roughly half the number of observations. Here the symbol $(.)$ represents the combination and the symbol $[.]$ is used to take only the integer part of a fraction. The Q_n estimator is the *g-th* order statistic of the $\binom{n}{2}$ inter point distances and the value 2.2219 is chosen to make Q_n a consistent estimator of scale for normal data (Abu-Shawiesh, Banik & Kibria, 2011).

2.8 Variance Estimators

Statistical analysis main purpose is to extract information from the data and identity various data properties within the data. Generally statistical analysis is based on information which is a function of data and it is analyzed using statistical principle and tool. Examples of such principles are the likelihood principle, robustness, sufficiency and the substitution principle. One of the key concepts in statistical analysis is sample which depicts an estimate of a defined population. In a data collection, sampling distribution is used as a random quantity with probability distribution.

There is need for sampling distribution knowledge which is linked with the nature of analysis. These are needed to prevent again estimation problem which arises from inaccurate estimator and creates huge estimation errors. The basic knowledge of accuracy understanding and measures such as bias, variance, and mean squared error of the estimator is needed. These accuracy understanding and measures are based on estimator's sampling distribution. An accuracy measure is normally used for selection of best estimator from a class of suitable estimators. The sampling distribution of a statistic and its features usually depend on the fundamental population. Most times, these estimation or approximation from given data have huge problems. The selection of estimator is based on the nature of the data and not on other factors. Popularly, there is a traditional approach used for estimating or approximating the sampling distribution of a statistic and its characteristics. Nevertheless, there are many identified limitations with this traditional approach. Thus the following sub-sections will discuss in details these approaches.

2.8.1 The Traditional Approach

Statistical analysis is used for data approximation or estimation to identify facts and information. The accuracy of these facts and information is measured by the given statistic (estimator), such as the variance, the bias and the mean squared error. For traditional approach, theoretical formula is derived based on a postulated model and this is used to run an empirical analysis where the accuracy depends on the established formula. This can be best picture in an example such as a variance. If $X_1, X_2, \cdots, X_n$ denote the data set of n distributed identically and independently (i.i.d.) observations from an unknown distribution F and let $T_n = T_n\left(X_1, \cdots, X_n\right)$ be a given statistic.

Then the variance of T_n is

$$\operatorname{var}\left(T_n\right) = \int \left[T_n\left(x\right) - \int T_n\left(y\right) d \prod_{i=1}^{n} F\left(y_i\right) \right]^2 d \prod_{i=1}^{n} F\left(x_i\right), \quad i = 1, \ldots, n \tag{2.29}$$

where $x_i = \left(x_1, \cdots, x_n\right)$ and $y_i = \left(y_1, \cdots, y_n\right)$. When T_n is simple, this can be obtain as an explicit expression of $\operatorname{var}\left(T_n\right)$ as unknown function of some quantities and then estimate $\operatorname{var}\left(T_n\right)$ by substituting the unknown quantities with their estimates. For example, if $T_n = \bar{X}_n^2$, then

$$\operatorname{var}\left(\bar{X}_n^2\right) = \frac{4\mu^2\alpha_2}{n} + \frac{4\mu\alpha_3}{n^2} + \frac{\alpha_4}{n^3} \tag{2.30}$$

where $\mu = E\left(X_1\right)$ and $\alpha_k = E\left(X_1 - \mu\right)^k$ is the k th central moment of X_1. We can then estimate $\operatorname{var}\left(\bar{X}_n^2\right)$ by substituting μ, and α_k with their estimators $\bar{X}_n$ and

$$\hat{\alpha}_k = \frac{1}{n-1} \sum_{i=1}^{n} \left(X_i - \bar{X}_n\right)^k, \; k = 2, 3, 4 \tag{2.31}$$

However, there are not many statistics as simple as the sample mean. For most statistics, the expression $\operatorname{var}\left(T_n\right) = \int \left[T_n\left(x\right) - \int T_n\left(y\right) d \prod_{i=1}^{n} F\left(y_i\right) \right]^2 d \prod_{i=1}^{n} F\left(x_i\right)$ is too complicated to be useful in obtaining estimators of $\operatorname{var}\left(T_n\right)$, and it is very hard or impossible to obtain an exact and explicit formula for $\operatorname{var}\left(T_n\right)$. Thus, in the traditional approach, we try to simplify the problem by considering approximations or asymptotic expansions of $\operatorname{var}(T_n)$.

However, there were some limitations identified with this traditional approach which include (Shao & Tu, 2012):

1. There is need for large sample size n in order to achieve a higher level of accurate variance (or accuracy measure) estimators. This fact is because of approximated formula usage.
2. Wrong in postulated model and theoretical formula can result in wrong in obtained estimator or results.
3. Wrong in derivations of suitable theoretical formula which might be due to unknowledgeable experiences in mathematical symbol and theoretical statistics.
4. Difficult in empirical processes expression which usually affect the result of the approach.
5. Wrong expression of model function parameters.
6. Complicated and misused of model parameters.

2.8.2 Cross-Validation (CV)

The cross-validation (CV) approach is very close with the jackknife for the case of selection of explanatory variables. These can be used to make futuristic predictions based on Cox's regression, linear, generalized linear and nonlinear models. Allen (1974) and Stone (1974) were one of the first researcher to introduced this approach which was a modified version of Jackknife and uses the concept of deleting one line within the approach to reach a better outcome. The approach is basically used in selection of bandwidth in nonparametric density estimation and nonparametric regression. For a better definition of this approach, Let $\hat{\beta}_{\alpha,i}$ be the Least Square Estimation (LSE) of β under model α after removing the pair $\left(y_i, x_i^{'}\right)$, that is,

$$\hat{\beta}_{\alpha,i} = \left(\sum_{j \neq i} x_{j\alpha} x_{j\alpha}^{'}\right)^{-1} \sum_{j \neq i} x_{j\alpha} y_j,\ i = 1, \cdots, n \tag{2.34}$$

Since y_j and $\hat{\beta}_{\alpha,i}$ are independent, $\overline{mse}(\alpha)$ can be estimated by

$$\overline{mse}_{CV}(\alpha) = \frac{1}{n} \sum_{i=1}^{n} \left(y_j - x_i^{'} \hat{\beta}_{\alpha,i}\right)^2 \tag{2.35}$$

where MSE denotes the mean squared error.

CV is advantageous over other approaches of estimating misclassification error such as the apparent error rate, data-splitting or bootstrap methods. Although, the apparent error rate (AER) will be adopted in comparison to CV when considering real data analysis. AER provides estimates of misclassification error which can be significantly biased towards zero although the bias decreases as the training sample increases. CV method on the other hand selects a model by minimizing $\overline{mse}_{CV}(\alpha)$. An essential asymptotic requirement for any given model selection procedure is its consistency in the sense that $\lim_{n \to \infty} P\left\{\hat{\alpha} = \alpha_0\right\} = 1$ (where $\hat{\alpha}$ is the model selected by using the given procedure), and $\overline{mse}_{CV}(\alpha)$ has been found to be an almost unbiased estimator of $\overline{mse}(\alpha)$. This provides a good justification for the application of this method.

2.9 Summary

This chapter has given an introduction to Robust LDA by taking a foundational review of Discriminant Analysis and the introduction of LDA. The reason why it is necessary to robustify LDA was justified as certain major drawbacks and limitations of the LDA approach was discussed. Furthermore, a review of common Robust Estimators and Variance Estimators were given and their advantages and drawbacks were also highlighted. The next chapter will introduce the methodology of this study by considering the specific robust and the variance estimators adopted in this study, which is the CV approach implemented on the MOM and WMOM.

CHAPTER THREE
RESEARCH METHODOLOGY

In this chapter, the steps involved in achieving the research objectives as aforementioned in Chapter One are discussed. The segmenting of this chapter is in the following sections; the first section gives the general view of the research design followed by the research framework. The framework gives the procedure for actualization of the research objectives and certain important preliminaries are also stated. The following sections now give detailed description of the robust estimators together with the evaluation approach for the robust estimator which in this case is the cross-validation technique. Finally, a summary of this chapter is given at the end.

3.1 Research Design

The approach for the evaluation of these robust estimators adopted involves computing the misclassification error of each technique using CV procedure. As a result of the shortcomings of LDA, where less accurate computations are obtained because of the presence of outliers detected, which simultaneously results in misclassification of new observations. There is expedient need to implement robust alternatives in discriminant analysis. Hence, the new Robust LDA techniques introduced in this research, that is, RLDA_{MQ} (MOM robust estimator with Q_n) and RLDA_{WMQ} (WMOM robust estimator with Q_n), will be adopted to certain simulation scenarios and evaluated in comparison to each other. In the utilization of each technique, focus is placed on a multivariate discriminant analysis problem, where $X = \left(X_1, X_2, \cdots, X_p\right)$ is a vector for independent variables of p dimensions.

It is expected that at the end of this study, researchers will have a better understanding on which robust estimator is suitable in certain real life scenarios modeled via simulation and consideration of real financial data.

3.2 Research Framework

In this section, a description on the research flow is followed. This is shown graphically by designing a flowchart for easier understanding and comprehension. The initial part of this research involves considering the modification of the robust estimators and then simulation analysis is considered for certain variable conditions. In addition, real financial data is also considered with comparison made amongst the linear models. Finally, comparison is made based on the average misclassification error using CV. Figure 3.1 presents the research flowchart.

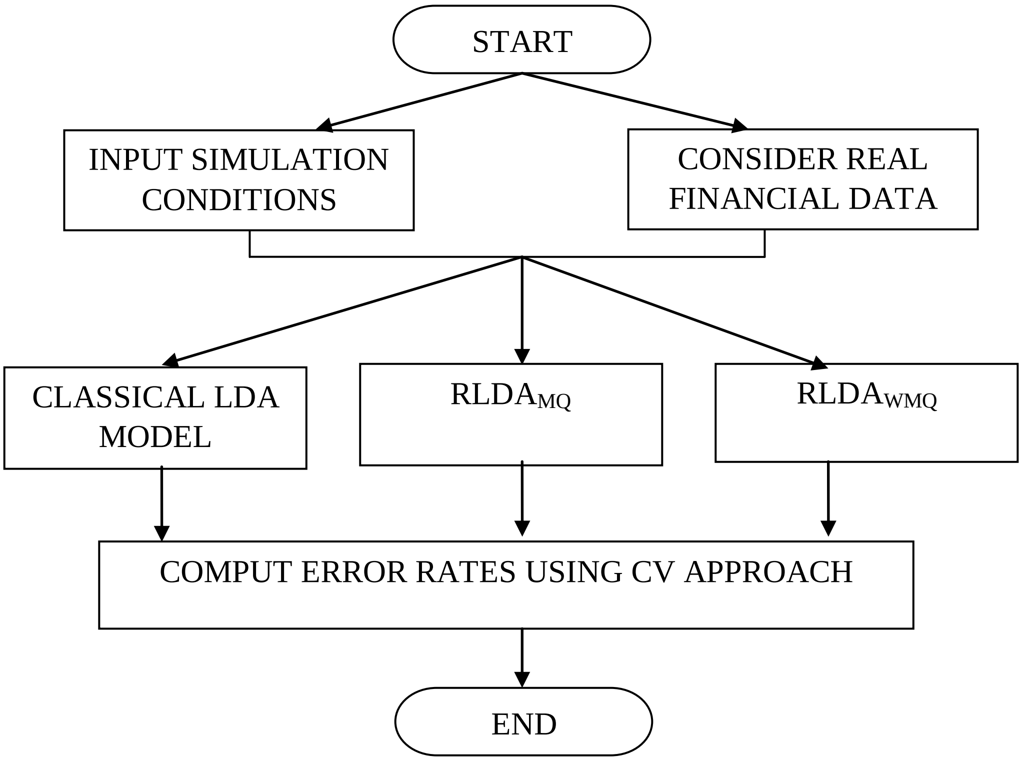

Figure 3.1 The Research Flowchart

From the flowchart above, it can be clearly observed how the research objectives of this study will be achieved. The modified linear discriminant models is adopted in comparison to the classical LDA model to know which approach is best suitable in detection of outliers. The misclassification error is computed and then comparison is made with simulations on financial data adapted from real life with MATLAB (2017a, 64-bits) which runs on a 2.00GHz laptop with 3.00GB RAM.

Before delving into a detailed description and process of implementation of the robust estimators and their corresponding evaluation estimators, it is worthwhile to understand certain preliminaries associated with the area of study.

3.2.1 Generation of Data

This study will be adopting simulation with real life scenarios considered in previously existing studies. This is for justification on a standard basis for comparison of the robust techniques considered in this study in comparison to works by other researchers. However, the data considered is expected to have certain properties as listed below.

3.2.2 Properties of Data

The following properties are essential to be noted for the analysis data:

1. Sample size: It is a grounded rule that the sample size of the smallest group should be larger than the number of variables under consideration (Austin & Steyerberg, 2015).

2. Sample division: The usual norm is to consider a random division of the observation set into smaller subsets, where the initial set is considered the training set, which is to approximate the model, while the other set is for estimation of test set which is implemented to estimate the results' reliability (Feng, Xu & Mannor, 2015).
3. Outliers: Conventionally, in discriminant analysis it is a requirement that the analysis data should follow an independent random sample pattern, that is, outliers should not be present. However, since the focus in this research is on robust discriminant analysis techniques, the presence of outliers is suitably handled.

3.2.3 Assumptions of the Discriminant Model

It is worth taking note that the discriminant model follows certain unique assumptions as discussed below (Poulsen & French, 2003):

1. Normal distribution: It is presumed that the analysis data considered is from a multivariate normal distribution. However, if the normality assumption is violated, the outcome is not considered fatal for the data at all times; rather it is assumed that the resultant significance tests may still have a level of reliability.
2. Correlation between mean and variance: The consistency in the validity of significance tests is not guaranteed when there is a large variability in the data group having extraordinary means on certain variables. Hence, the need for a correlation between the variance/standard deviations with the mean is of paramount importance.

3. Homogeneity of variances or covariances: The assumption in this case follows the case that when the variables are written in matrix form, the resultant matrices are homogeneous across groups through variance or covariance of the variables. Although too much focus is not needed to be placed on minor deviations, it is still a better approach to evaluate the within groups variances and correlation matrices before the end conclusions of any study is accepted.
4. Matrices ill-conditioning: The process of computation involved in discriminant analysis requires matrix inversion of the variance or covariance matrix in the model. A situation where a variable cannot be inverted or where there is redundancy of a variable with other variables, the case is described as ill-conditioned. Hence, this assumption is very important to assure the existence of solution.
5. Low multicollinearity of the variables: This assumption is important because in a case where the contrary exists, that is when there is high multicollinearity between two variables or more, the resulting coefficients will be predict group membership properly.

Certain verification approaches can be used to confirm if these assumptions hold, amongst which includes normality test and equality test of covariance matrices.

To investigate the misclassification error in the analysis data, certain techniques were adopted. The sections discussed subsequently gives the detailed steps involved in each technique, starting from the conventional approach to the utilization of the robust estimators as shown in Figure 3.1.

3.3 Linear Discriminant Analysis (LDA)

The derivation of the LDA formula as described by Teknomo (2015) is given before the iterative algorithm is discussed. Now, considering g groups, the conventional Bayes' rule minimizes the total error of classification by assigning the object x to group i which has the largest conditional probability where $P(i|x) > P(j|x) \;\forall j \neq i$.

Since $P(i|x)$ (that is, the probability of class given the measurement) cannot be computed directly from the measurement, $P(x|i)$ (the probability of measurement given the class) is initially computed. Then Bayes theorem which states

$$P(i|x) = \frac{P(x|i)P(i)}{\sum_{\forall j} P(x|j)P(j)} \tag{3.1}$$

is adopted. The object x is assigned to group i if

$$\frac{P(x|i)P(i)}{\sum_{\forall k} P(x|k)P(k)} > \frac{P(x|j)P(j)}{\sum_{\forall k} P(x|j)P(j)} \;\forall j \neq i \tag{3.2}$$

Since both denominators are same, the expression can be rewritten to state that the object x is assigned to group i if $P(x|i)P(i) > P(x|j)P(j) \;\forall j \neq i$.

In a case where there are many groups and many dimensions of measurement with each dimension having many values, calculating the conditional probability $P(x|i)$ requires a lot of data. Hence, it is reasonable to assume that the data follows certain theoretical distribution such as the data following from a multivariate normal distribution with formula given as

$$P(x|i) = \left(\frac{1}{(2\pi)^{\frac{1}{2}} |C_i|^{\frac{1}{2}}} \right) \exp\left(-\frac{1}{2}(x-\mu_i)^T C_i^{-1} (x-\mu_i) \right) \tag{3.3}$$

where μ_i and C_i are the vector mean and covariance matrix of group i respectively. Substituting (3.3) into (3.1), the object x can be assigned to group i if

$$\begin{aligned}&\left(\frac{P(i)}{(2\pi)^{\frac{1}{2}}|C_i|^{\frac{1}{2}}}\right)\exp\left(-\frac{1}{2}(x-\mu_i)^T C_i^{-1}(x-\mu_i)\right)\\&>\left(\frac{P(j)}{(2\pi)^{\frac{1}{2}}|C_j|^{\frac{1}{2}}}\right)\exp\left(-\frac{1}{2}(x-\mu_j)^T C_j^{-1}(x-\mu_j)\right),\ i\neq j.\end{aligned} \tag{3.4}$$

Since the factor of $(2\pi)^{\frac{1}{2}}$ is same for both sides, it can be cancelled out to give

$$\left(\frac{P(i)}{|C_i|^{\frac{1}{2}}}\right)\exp\left(-\frac{1}{2}(x-\mu_i)^T C_i^{-1}(x-\mu_i)\right)>\left(\frac{P(j)}{|C_j|^{\frac{1}{2}}}\right)\exp\left(-\frac{1}{2}(x-\mu_j)^T C_j^{-1}(x-\mu_j)\right),\ i\neq j \tag{3.5}$$

Taking the logarithm of both sides,

$$\begin{aligned}&-\frac{1}{2}\ln(|C_i|)+\ln(P(i))-\frac{1}{2}(x-\mu_i)^T C_i^{-1}(x-\mu_i)\\&>-\frac{1}{2}\ln(|C_j|)+\ln(P(j))-\frac{1}{2}(x-\mu_j)^T C_j^{-1}(x-\mu_j),\ i\neq j\end{aligned} \tag{3.6}$$

which is same as

$$\begin{aligned}&\ln(|C_i|)-2\ln(P(i))+(x-\mu_i)^T C_i^{-1}(x-\mu_i)\\&<\ln(|C_j|)-2\ln(P(j))+(x-\mu_j)^T C_j^{-1}(x-\mu_j),\ i\neq j\end{aligned} \tag{3.7}$$

Now if all the covariance matrices are equal; $C=C_i=C_j$, then further simplification can be made to give

$$\begin{aligned}&\ln(|C|)-2\ln(P(i))+(x-\mu_i)^T C^{-1}(x-\mu_i)\\&<\ln(|C|)-2\ln(P(j))+(x-\mu_j)^T C^{-1}(x-\mu_j),\ i\neq j\end{aligned}\ . \tag{3.8}$$

Rewriting $(x-\mu_i)^T C^{-1}(x-\mu_i)$ as $xC^{-1}x^T - 2\mu_i C^{-1}x^T + \mu_i C^{-1}\mu_i^T$. Therefore the inequality in (3.8) becomes

$$\begin{aligned}&\ln(|C|) - 2\ln(P(i)) + xC^{-1}x^T - 2\mu_i C^{-1}x^T + \mu_i C^{-1}\mu_i^T \\ &< \ln(|C|) - 2\ln(P(j)) + xC^{-1}x^T - 2\mu_j C^{-1}x^T + \mu_j C^{-1}\mu_j^T,\ i \neq j\end{aligned} \quad (3.9)$$

Eliminating the first and third terms of both sides in (3.9) as they do not have an effect on the group decision gives,

$$-2\ln(P(i)) - 2\mu_i C^{-1}x^T + \mu_i C^{-1}\mu_i^T < -2\ln(P(j)) - 2\mu_j C^{-1}x^T + \mu_j C^{-1}\mu_j^T,\ i \neq j \quad (3.10)$$

Thus, $\ln(P(i)) + \mu_i C^{-1}x^T - \frac{1}{2}\mu_i C^{-1}\mu_i^T > \ln(P(j)) + \mu_j C^{-1}x^T - \frac{1}{2}\mu_j C^{-1}\mu_j^T,\ i \neq j$ and suppose $f_i = \mu_i C^{-1}x_k^T - \frac{1}{2}\mu_i C^{-1}\mu_i^T + \ln(P(i))$, where k is the number of groups. Object with measurement x can be assigned to group i if $f_i > f_j \ \forall\ i \neq j$. This follows from the assumptions that the discriminant analysis follows a multivariate normal distribution with same covariance matrix. Given below is the iterative algorithm of LDA.

Consider x_i independent variables features of a multivariate data of g groups having corresponding dependent variables y.

Step 1: Compute μ_i and μ; the characteristic mean in the group i is given as the average of x_i and the global mean vector respectively.

Step 2: Obtain the covariance matrix C_i of group i defined as $C_i = \frac{\left(x_i^0\right)^T x_i^0}{n_i}$, where x_i^0 is the corrected mean data which is the characteristic for i, x_i group which can be the global minus mean vector μ.

Step 3: Compute the pooled within the group covariance matrix $C(r,s) = \frac{1}{n}\sum_{i=1}^{g} n_i C_i(r,s)$ for every matrix in entry (r,s).

Step 4: Determine the inverse of the matrix pooled covariance C^{-1} and the prior probability vector P assumed to be equal to the total sample of each group divided by the total samples.

Step 5: Assign object x to group i that has maximum f_i, where $f_i = \mu_i C^{-1} x_i^T - \frac{1}{2}\mu_i C^{-1} \mu_i^T + \ln\left(p_i\right)$.

3.4 Modified One-Step M-Estimator with Q_n (MOM-Q_n)

This approach involves combining the MOM statistic with the highly robust Q_n scale estimator. MOM as obtained from the conventional one-step M-estimator (Haddad, 2013; Staudte and Sheather, 1990) but with certain modifications is simply the average of the values remaining after the removal of all extreme values (if there is existence of any). The robust Q_n scale estimator on the other hand as proposed by Rousseeuw and Croux (1993) is a well suitable estimator with the advantage of high efficiency. The algorithm to combine these two techniques is described iteratively below.

Step 1: Trim the data to be analyzed using the default scale estimator MAD_n for determining the extreme values in MOM criterions. Let $\hat{M}_j$ be the median for group j,

$$MAD_{nj} = \frac{MAD_j}{0.6745};\ MAD_j = Median\left|Y_{1j} - \hat{M}_j\right|, \left|Y_{2j} - \hat{M}_j\right|, \cdots, \left|Y_{nj} - \hat{M}_j\right|.$$

Step 2: Compute $\hat{\theta}_j$ following equation (2.21).

Step 3: Calculate Q_n from equation (2.28).

Step 4: Replace the default scale estimator MAD_n in Step 2 with the Q_n estimator to obtain i_1 as the number of observations Y_{ij} such that $\left(Y_{ij} - \hat{M}_j\right) < -2.24\left(Q_{nj}\right)$ and i_2 is the number of observations $\left(Y_{ij} - \hat{M}_j\right) > 2.24\left(Q_{nj}\right)$.

Step 5: Compute $\hat{\theta}_j$ based on the Q_n estimator in Step 4.

3.5 Winsorized Modified One-Step M-Estimator with Q_n (WMOM-Q_n)

This approach involves the WMOM statistic. To improve performance of the conventional MOM estimator, the estimator is winsorized and this new estimator is merged with the advantageous robust Q_n scale estimator and its counterpart winsorized covariance matrix. The conventional algorithm of WMOM automatically adopts using its winsorized covariance matrix, however a step further is taken to investigate the behaviour when merged with highly advantageous Q_n scale estimator. Given below is the general approach of the WMOM algorithm and also insight is given on how to merge this estimator with the Q_n scale estimator.

Step 1: Eliminate the outliers from the analysis data and after that, the data is winsorized

Step 2: Compute the winsorized MOM for *j*-th variable and the winsorized covariance matrix between w_i and w_j variables respectively and represented with:

$$\overline{w_j} = \frac{1}{m_j}\sum_{i=1}^{m_j} w_{ij}$$

and the vector of winsorized MOM estimator is obtained from

$$\overline{w} = \begin{bmatrix} \overline{w_1} \\ \vdots \\ \overline{w_p} \end{bmatrix} \text{ and } S_{WMADn}\left(w_i, w_j\right) = \frac{1}{m-1}\left[\sum_{k=1}^{m} w_{ki} w_{kj} - m^{*}\overline{w_i}\,\overline{w_j}\right]$$

Step 3: Replace the standard mean vector in Step 2 then inverse the matrix covariance with the winsorized $MOM\left(\overline{X}_{WMAD_n}\right)$. Thus, the inverse of the matrix winsorized covariance $S^{-1}_{WMAD_n}$ from Step 2 which is given as MAD_n is the default scale estimator for the criterion on trimming.

However, Yahaya et al. (2006) pinpointed that highly robust scale estimators such as the Q_n scale estimator could improve the results and for this reason, there is a replacement of the conventional MAD_n with this robust scale estimators in the trimming criterion. This approach follows the same algorithm for the first three steps but with an extra Step 4 as described below.

Step 4: Replace the default scale estimator MAD_n in Step 3 with the Q_n estimator in equation (2.28).

3.6 Cross Validation (CV)

The CV is a computer intensive technique and the main objective of the CV approach is the selection of the more suitable model or technique by minimizing the MSE (Bengio & Grandvalet, 2004). Considering the $k-$fold CV which is the most common approach, the concept is simplified and explained iteratively in the steps below.

Step 1: Compute μ_i; the mean of characteristics in the group i

Step 2: Obtain the covariance matrix C_i of group i defined as $C_i = \frac{\left(x_i^0\right)^2}{n_i}$, where x_i^0 is the corrected data mean which is the characteristics for group i, x_i the global minus mean vector μ_i

Step 3: Obtain the MSE for each group i from $MSE_i = \frac{\left(n_i - 1\right)C_i^2}{n_i}$

Step 4: Evaluate the error from the cross validation formula; $CV = \frac{\sum MSE_i}{k}$

3.7 Variables Manipulated

The performance of the robust estimators computed in terms of misclassification error was investigated on a number of simulation settings. To obtain these simulation conditions, five variables were manipulated which include: dimension of variable (p), percentage of contamination(ε), sample size of the training data(n_1, n_2), shift in location of the population (μ) and shift in shape of the population(κ). These manipulations are important to show the advantages and disadvantages of the robust estimators. The choice of variables to be manipulated follows from prior adoption in previous studies, such as Haddad (2013) and Lim et al., (2016) amongst others. Table 3.1 gives a summary of the simulation scenarios.

Table 3.1
Simulation Conditions

Variable	**Descriptions**
Dimension of variable (p)	2, 6, 10
Percentage of contamination (ε)	10, 20, 40
Sample size of the training data (n_1, n_2)	Balanced: (20,20), (50,50), (100,100) Unbalanced: (50,20), (100,50), (100,20)
Shift in location of the population (μ)	0, 3, 5
Shift in shape of the population (κ)	9, 25, 100

3.7.1 Dimension of Variable (*p*) and Sample Size (*n*)

Variation was introduced in the dimension of variable to show that the research can handle multivariate data. As highlighted in Table 3.1, *p* takes the values 2, 6 and 10. Likewise, manipulations were also made in the sample sizes to show its relationship to the variable dimension. The group sample sizes were set to be balanced and unbalanced to check the effect on the misclassification error per variable dimension. Alfaro and Ortega (2008; 2009), Rousseeuw and Zomeren (1990) and Jensen et al., (2007) reported that if *n* is fixed and *p* is increased, then the effect will cause the misclassification error to decrease. Misclassification errors are more serious in unbalanced sample sizes than in balanced. Hence, the motivation to examine the robustness of the estimators for both balanced and unbalanced sample sizes to investigate the variation in results obtained.

This concept of considering different dimension of variables with respect to varying sample sizes is discussed in works of Jensen et al., (2007), Alfaro and Ortega (2009), Chenouri et al., (2009), Yusof, Othman and Yahaya (2010), Yusof, Abdullah, Yahaya and Othman (2011), Li, Shao and Deng (2015) amongst others. These authors chose *n* values ranging from 25 to 1000. To check the performance of the robust estimators in this research, the values of *p* are set as 2, 6, and 10 with $\left(n_1, n_2\right)$ taking values as highlighted in Table 3.1.

3.7.2 Percentage of Contamination (ε), Shifts in Location (μ) and Population (κ)

To further check whether the robust estimators can adequately handle the presence of outliers, we shifted the mean (centrality) to a certain values for both location and population denoted by μ and κ respectively. The larger the shift, the more extreme is the values of the outliers. In this study, we used 3 levels of shifts in location ($\boldsymbol{\mu}$), that is 0 representing no outliers, 3 for moderate outlier values and 5 for extreme outlier values. Likewise, the shifts in population take values (9, 25, 100) with the percentage of contamination chosen at 10%, 20% and 40%. These variables were manipulated in line with works of Lim et al., (2016), Alfaro and Ortega (2009), Mohammadi et al., (2011). Although a number of their works chose the percentage to be between 10% and 20%.

Note that this research considers both balanced and unbalanced sample sizes which give rise to two data structures, which are the cases of equal and unequal covariance matrices. Therefore for both balanced and unbalanced data, analysis will be made based on the nature of the covariance matrices.

Considering the first case of equal covariance matrices, each group $\pi_j,\ j=1,2$ has a separate mean μ_j but the same covariance matrix I_p. Therefore, the data was contaminated for the equal covariance matrices as follows

$$\pi_1:(1-\varepsilon)N_p\left(\mu_j, I_p\right)+\varepsilon N_p\left(\mu_j+\mu, \kappa I_p\right) \tag{3.9a}$$

and

$$\pi_2:(1-\varepsilon)N_p\left(\mu_j, I_p\right)+\varepsilon N_p\left(\mu_j-\mu, \kappa I_p\right). \tag{3.9b}$$

On the other hand, for the unequal covariance matrices, each group $\pi_j,\ j=1,2$ also has a separate mean μ_j but in this case, each group has their covariance matrix computed as jI_p. This leads to the expression of data contamination for the unequal covariance matrix condition given as

$$\pi_1:(1-\varepsilon)n_1N_p\left(0, I_p\right)+\varepsilon n_1N_p\left(0+\mu, \kappa I_p\right) \tag{3.10a}$$

$$\pi_2:(1-\varepsilon)n_2N_p\left(1, 2I_p\right)+\varepsilon n_2N_p\left(1-\mu, \kappa I_p\right). \tag{3.10b}$$

CHAPTER FOUR
RESULT AND ANALYSIS

4.1 Introduction

This chapter investigates misclassification error for each of the considered approaches: Classical Approach (CA), RLDA with MOM-Q_n ($RLDA_{MQ}$) and RLDA with WMOM-Q_n ($RLDA_{WMQ}$). This analysis starts by first considering the simulation data by comparing which linear discriminant model has better outlier detection ability when the sample sizes are balanced or unbalanced for both cases of equal and unequal covariance matrices respectively. For this cause, the mean misclassification is considered for multivariate variable dimensions (2, 6, 10) with the sample size of the training data (n_1, n_2) taking values as highlighted in Table 3.1. A step further is taken in the analysis phase to investigate the computational rigor involved in adopting these linear discriminant models by calculating the computational time. In addition, analysis is also made on certain real financial data and the results from the analysis conducted are also reported.

4.2 Misclassification Error Analysis with Simulation Study

This section shows the misclassification error of the linear discriminant model using simulation analysis when considering variable dimensions 2, 6 and 10 for balanced and unbalanced sample sizes. This section is divided into two parts; the first part shows the average misclassification error when the covariance matrices are equal for all variable dimensions while the second part gives the results for the unequal covariance matrices.

A testing sample of size 2000 from each population was generated and the misclassification error was computed by obtaining the proportion of misclassified testing sample observations in each population. It is known that as the size of the testing sample size increases, the accuracy and the running time will also be increased. However, a size of 2000 was chosen because if the testing sample size is increased beyond 2000, there is only a small amount of additional accuracy achieved (Hintze, 2008). Therefore, to avoid unrequired increase in running time, the size is chosen as 2000 which has also been adopted in similar studies as Lim et al., (2016) and Yahaya et al. (2016a). Thus, after choosing the testing sample size, the simulation process was repeated 2000 times and the mean misclassification error was recorded. Furthermore, the computation of each linear discriminant model was documented as same number of simulation trials.

4.2.1 Equal Covariance Matrices

The combination of the various variable settings as highlighted in Table 3.1 gave rise to 54 uncontaminated, 324 location contamination, 486 shape contamination and 972 location and shape contamination. This produces a total of 1836 different data conditions. The following subsections will highlight the results obtained for both the balanced and unbalanced sample sizes per variable dimension.

4.2.1.1 Balanced Sample Sizes

The first set of results discussed is the balanced sample sizes; $(n_1, n_2) = [(20, 20), (50, 50), (100, 100)]$. The results are considered for variable dimension $p = 2, 6, 10$ with focus on the average misclassification error of each linear discriminant model.

For the case where the variable dimension is two, the mean misclassification error analysis for CA, $\mathrm{RLDA}_{\mathrm{MQ}}$ and $\mathrm{RLDA}_{\mathrm{WMQ}}$ are shown for the balanced sample sizes (n_1, n_2).

Table 4.1

Mean Misclassification Error for Linear Discriminant Models with Balanced Sample Sizes, Equal Covariance Matrices and $p = 2$

			$(n_1,n_2)=(20,20)$			$(n_1,n_2)=(50,50)$			$(n_1,n_2)=(100,100)$		
ε	μ	κ	**CA**	**$RLDA_{MQ}$**	**$RLDA_{WMQ}$**	**CA**	**$RLDA_{MQ}$**	**$RLDA_{WMQ}$**	**CA**	**$RLDA_{MQ}$**	**$RLDA_{WMQ}$**
-	-	-	0.2511	0.2530	0.2518	0.2442	0.2449	0.2445	0.2420	0.2424	0.2421
10	**3**	-	0.3389	0.2867	0.3057	0.2960	0.2583	0.2746	0.2741	0.2496	0.2603
10	**5**	-	0.4987	0.2723	0.2997	0.4986	0.2519	0.2757	0.5010	0.2462	0.2668
10	**0**	**9**	0.3178	0.2549	0.2566	0.2759	0.2455	0.2465	0.2587	0.2427	0.2433
10	**0**	**25**	0.4205	0.2542	0.2564	0.3863	0.2452	0.2466	0.3447	0.2426	0.2434
10	**0**	**100**	0.4903	0.2538	0.2561	0.4842	0.2451	0.2466	0.4800	0.2425	0.2435
10	**3**	**9**	0.3884	0.2556	0.2597	0.3610	0.2456	0.2482	0.3270	0.2428	0.2445
10	**3**	**25**	0.4527	0.2544	0.2573	0.4441	0.2453	0.2471	0.4234	0.2426	0.2437
10	**3**	**100**	0.4937	0.2538	0.2564	0.4916	0.2451	0.2467	0.4929	0.2425	0.2435
10	**5**	**9**	0.4548	0.2570	0.2632	0.4732	0.2461	0.2500	0.4804	0.2430	0.2457
10	**5**	**25**	0.4755	0.2545	0.2581	0.4870	0.2452	0.2475	0.4917	0.2426	0.2440
10	**5**	**100**	0.4961	0.2537	0.2565	0.4963	0.2452	0.2468	0.5012	0.2425	0.2436
20	**3**	-	0.5770	0.4745	0.5379	0.6202	0.4009	0.5733	0.6542	0.3480	0.6045
20	**5**	-	0.6530	0.3925	0.5396	0.6911	0.2998	0.5827	0.7124	0.2710	0.6376
20	**0**	**9**	0.3624	0.2608	0.2687	0.3055	0.2470	0.2516	0.2745	0.2433	0.2463
20	**0**	**25**	0.4637	0.2576	0.2654	0.4277	0.2461	0.2511	0.3929	0.2429	0.2468
20	**0**	**100**	0.4995	0.2556	0.2619	0.4911	0.2457	0.2502	0.4896	0.2427	0.2462
20	**3**	**9**	0.5083	0.2624	0.2849	0.5334	0.2479	0.2613	0.5678	0.2437	0.2529
20	**3**	**25**	0.5041	0.2574	0.2690	0.5062	0.2463	0.2536	0.5237	0.2430	0.2480
20	**3**	**100**	0.5027	0.2557	0.2624	0.4993	0.2456	0.2503	0.5042	0.2428	0.2463
20	**5**	**9**	0.6039	0.2662	0.3039	0.6795	0.2492	0.2807	0.7158	0.2445	0.2671
20	**5**	**25**	0.5310	0.2578	0.2719	0.5590	0.2465	0.2556	0.6061	0.2431	0.2494
20	**5**	**100**	0.5053	0.2552	0.2625	0.5048	0.2456	0.2504	0.5144	0.2428	0.2464
40	**3**	-	0.7061	0.7080	0.7050	0.7328	0.7319	0.7319	0.7442	0.7430	0.7436
40	**5**	-	0.6955	0.6894	0.6857	0.7252	0.7013	0.7177	0.7389	0.7134	0.7332
40	**0**	**9**	0.4100	0.3059	0.3432	0.3491	0.2613	0.2883	0.3063	0.2496	0.2639
40	**0**	**25**	0.4804	0.3075	0.3687	0.4571	0.2593	0.3272	0.4346	0.2487	0.2959
40	**0**	**100**	0.4975	0.2801	0.3227	0.4965	0.2514	0.3132	0.4940	0.2452	0.3092
40	**3**	**9**	0.6106	0.3547	0.4844	0.6767	0.2804	0.5185	0.7162	0.2579	0.5479
40	**3**	**25**	0.5174	0.3101	0.3969	0.5499	0.2606	0.3757	0.5867	0.2490	0.3566
40	**3**	**100**	0.5005	0.2792	0.3247	0.5035	0.2518	0.3178	0.5076	0.2454	0.3130
40	**5**	**9**	0.6693	0.3967	0.5791	0.7172	0.3079	0.6661	0.7372	0.2727	0.7192
40	**5**	**25**	0.5446	0.3129	0.4203	0.5992	0.2620	0.4185	0.6453	0.2497	0.4264
40	**5**	**100**	0.5023	0.2811	0.3289	0.5080	0.2523	0.3235	0.5159	0.2454	0.3258
Performance			**2.9%**	**91.2%**	**5.9%**	**2.9%**	**97.1%**	**2.9%**	**2.9%**	**97.1%**	**0%**

Recall that there are five simulated variables (see Table 3.1) which includes dimension of variable (p), percentage of contamination (ε), sample size of the training data (n_1,n_2), shift in location of the population (μ) and shift in shape of the population (κ). Therefore, for every p (such as $p=2$ in Table 4.1), discussions will be made with respect to the other four variables.

Firstly, considering the percentage of contamination (ε), Table 3.1 states that the percentage will vary from 10% to 20% and then 40%. It is observed that as ε increases, the mean misclassification error also increases at constant μ and κ. Although certain exemptions were noticed for CA at $(n_1, n_2) = (20, 20)$ where the mean misclassification error of $(\varepsilon, \mu, \kappa) = (20, 0, 100)$ which is 0.4995 is greater than the mean misclassification error of $(\varepsilon, \mu, \kappa) = (40, 0, 100)$ which is 0.4975. Similarly, when ε increases from 20% to 40% for $(\mu, \kappa) = (3, 100)$ and $(\mu, \kappa) = (5, 100)$, likewise anomality is observed with CA as $0.5027 > 0.5005$ and $0.5053 > 0.5023$ respectively. However, as (n_1, n_2) increased to $(50, 50)$ and $(100, 100)$, stability is observed as all mean misclassification error increased with increasing ε. This implies that CA's ability to correctly detect outliers is weak for small sample sizes.

Secondly, considering (n_1, n_2) it cannot be generally concluded that as the sample size increases, the mean misclassification error reduces or increases. Therefore, in this case, the discussion considers the linear discriminant model with the highest performance percentage as the sample size increases. From the performance row, CA maintains a performance percentage of 2.9% at all sample sizes and this is the point where CA obtains the least mean misclassification error when the data is clean (no contamination) with $\varepsilon = 0,\ \mu = 0,\ \kappa = 0$. For the RLDA$_{WMQ}$ on the other hand, a depreciating performance percentage is observed from 5.9% to 2.9% and then 0%. This implies that for small sample sizes of simulation data with equal covariance matrix and balanced sample sizes, RLDA$_{WMQ}$ shows minimal high performance but as the sample size increases, its performance tends to zero. This leaves the RLDA$_{MQ}$ as the estimator with highest percentage performance as the sample size increases.

Thirdly, considering the shift in location of the population (μ), one notable behavior as μ increased from 3 to 5 at $\kappa = 0$ is the decrease in the mean misclassification error. Although, this observation was not consistent at 10% and 20% contamination, it was consistent as 40% contamination. This implies that when there is no shift in the shape of the population at increasing percentage of contamination, linear discriminant models have better accuracy in detecting of outliers. In addition, the behavior of the $\mathrm{RLDA_{MQ}}$ as an accurate model at this level of variable manipulations is observed in the minimal and convergent results obtained even as μ increased. Some of these convergence is observed at $(\varepsilon,\mu,\kappa) = (10,0,100)$, $(\varepsilon,\mu,\kappa) = (10,3,100)$ of $(n_1,n_2) = (20,20)$ and $(\varepsilon,\mu,\kappa) = (10,\mu,25)$ of $(n_1,n_2) = (100,100)$.

Finally, considering the shift in shape of the population (κ), the general behavior for all the linear discriminant models is sharp reduction as soon as κ goes from 0 to 9 before its gradual descent to convergence at $\kappa = 25,100$. This pattern is seen for the $\mathrm{RLDA_{MQ}}$ at $(n_1,n_2) = (100,100)$ and 40% of contamination where there is a sharp reduction from 0.7430 and 0.7134 to 0.2579 and 0.2727 for μ equals 3 and 5 respectively.

A general overview of the results obtained in Table 4.1 shows CA obtaining the least mean misclassification error at no contamination. This is in line with the theory that the classical LDA approach will perform optimally when the assumptions of the LDA are fulfilled. Although, $\mathrm{RLDA_{MQ}}$ and $\mathrm{RLDA_{WMQ}}$ also gave favourable results as the difference between the mean misclassification error of the robust estimators and the classical approach is within the range $\left[3\times10^{-4}, 1.9\times10^{-3}\right]$ which shows convergence in results. However, as soon as there is a little contamination in the data, say 10%

contamination $\varepsilon = 10$, the mean misclassification error for CA is seen to increase from a prior mean of 0.2511 to 0.3389, 0.2442 to 0.2960, and 0.2420 to 0.2741 for $(n_1, n_2) = (20, 20)$, $(50, 50)$ and $(100,100)$ respectively. This shows CA cannot handle the presence of outliers in data as a more increase is observed as the percentage of contamination increased to 20% and then 40% (Lim et al., 2016). Likewise, the RLDA_{WMQ} robust model had an increased mean misclassification error value but the RLDA_{MQ} approach maintained a controlled increase, thus being the best approach at the 10% and 20% levels of contamination.

The best mean misclassification error is seen to fluctuate between RLDA_{MQ} and RLDA_{WMQ} when there was only a shift in location of the population such as the tie in value of 0.7319 at 40% percent contamination with moderate outlier value of $\mu = 3$ when $(n_1, n_2) = (50, 50)$. However, as the sample size of the training sample increased to (100, 100), RLDA_{MQ} showed a consistent control of a least mean misclassification error. Hence making this robust model the best at p=2. Although, higher variable dimensions are also considered to investigate the general behavior of the linear discriminant models.

Therefore, for the case where the variable dimension is p=6 and p=10, the mean misclassification error analysis for CA, RLDA_{MQ} and RLDA_{WMQ} are shown for the balanced sample sizes (n_1, n_2) in Tables 4.2 and 4.3 respectively.

Table 4.2

Mean of Misclassification Error for Linear Discriminant Models with Balanced Sample Sizes, Equal Covariance Matrices and $p = 6$

			$(n_1,n_2)=(20,20)$			$(n_1,n_2)=(50,50)$			$(n_1,n_2)=(100,100)$		
ε	μ	κ	**CA**	**RLDA$_{MQ}$**	**RLDA$_{WMQ}$**	**CA**	**RLDA$_{MQ}$**	**RLDA$_{WMQ}$**	**CA**	**RLDA$_{MQ}$**	**RLDA$_{WMQ}$**
-	-	-	0.1409	0.1442	0.1417	0.1214	0.1226	0.1218	0.1157	0.1163	0.1159
10	**3**	-	0.3915	0.2728	0.3249	0.3286	0.1937	0.2654	0.2740	0.1574	0.2195
10	**5**	-	0.4998	0.2438	0.2970	0.5004	0.1697	0.2616	0.4991	0.1418	0.2395
10	**0**	**9**	0.2108	0.1484	0.1490	0.1812	0.1247	0.1269	0.1505	0.1172	0.1190
10	**0**	**25**	0.2543	0.1481	0.1491	0.2696	0.1246	0.1273	0.2252	0.1172	0.1193
10	**0**	**100**	0.2725	0.1476	0.1486	0.4413	0.1245	0.1273	0.4310	0.1171	0.1194
10	**3**	**9**	0.2679	0.1541	0.1611	0.2757	0.1271	0.1354	0.2414	0.1184	0.1239
10	**3**	**25**	0.2655	0.1485	0.1517	0.3288	0.1250	0.1295	0.3142	0.1173	0.1205
10	**3**	**100**	0.2733	0.1476	0.1493	0.4572	0.1245	0.1278	0.4562	0.1172	0.1197
10	**5**	**9**	0.3253	0.1625	0.1743	0.3809	0.1306	0.1453	0.4000	0.1202	0.1312
10	**5**	**25**	0.2783	0.1497	0.1534	0.3812	0.1255	0.1314	0.4072	0.1175	0.1217
10	**5**	**100**	0.2742	0.1477	0.1497	0.4675	0.1246	0.1281	0.4736	0.1172	0.1198
20	**3**	-	0.5365	0.4698	0.5085	0.5611	0.4313	0.5322	0.5866	0.3913	0.5504
20	**5**	-	0.5668	0.4141	0.4491	0.6101	0.3300	0.4906	0.6526	0.2670	0.5397
20	**0**	**9**	0.2514	0.1567	0.1638	0.1980	0.1277	0.1354	0.1587	0.1185	0.1233
20	**0**	**25**	0.3613	0.1541	0.1605	0.3534	0.1270	0.1351	0.2921	0.1181	0.1240
20	**0**	**100**	0.4694	0.1524	0.1586	0.4871	0.1262	0.1327	0.4684	0.1179	0.1231
20	**3**	**9**	0.3933	0.1693	0.1934	0.4948	0.1338	0.1637	0.5381	0.1214	0.1435
20	**3**	**25**	0.4204	0.1553	0.1660	0.4977	0.1277	0.1403	0.5044	0.1185	0.1279
20	**3**	**100**	0.4780	0.1526	0.1581	0.5036	0.1265	0.1340	0.4960	0.1180	0.1237
20	**5**	**9**	0.4956	0.1882	0.2279	0.6776	0.1433	0.2067	0.7669	0.1265	0.1838
20	**5**	**25**	0.4625	0.1572	0.1712	0.5911	0.1287	0.1453	0.6490	0.1190	0.1315
20	**5**	**100**	0.4846	0.1524	0.1586	0.5146	0.1266	0.1346	0.5147	0.1180	0.1242
40	**3**	-	0.6433	0.6476	0.6404	0.7165	0.7153	0.7129	0.7677	0.7635	0.7644
40	**5**	-	0.6137	0.5988	0.5940	0.6793	0.6243	0.6553	0.7300	0.6573	0.7035
40	**0**	**9**	0.3240	0.2062	0.2315	0.2487	0.1475	0.1781	0.1893	0.1275	0.1443
40	**0**	**25**	0.4563	0.2087	0.2424	0.4247	0.1458	0.2142	0.3682	0.1266	0.1762
40	**0**	**100**	0.4991	0.1827	0.2059	0.4949	0.1364	0.1922	0.4853	0.1223	0.1871
40	**3**	**9**	0.6382	0.2749	0.3990	0.7623	0.1898	0.4842	0.8194	0.1497	0.5321
40	**3**	**25**	0.5355	0.2175	0.2762	0.5995	0.1501	0.2726	0.6495	0.1288	0.2485
40	**3**	**100**	0.5035	0.1834	0.2090	0.5101	0.1364	0.1969	0.5128	0.1224	0.1943
40	**5**	**9**	0.7232	0.3428	0.5134	0.8173	0.2541	0.6878	0.8526	0.1953	0.7845
40	**5**	**25**	0.5805	0.2289	0.3000	0.6701	0.1556	0.3324	0.7379	0.1315	0.3441
40	**5**	**100**	0.5070	0.1844	0.2112	0.5195	0.1363	0.2007	0.5306	0.1226	0.1995
Performance			**2.9%**	**91.2%**	**5.9%**	**2.9%**	**94.2%**	**2.9%**	**2.9%**	**97.1%**	**0%**

Table 4.3

Mean of Misclassification Error for Linear Discriminant Models with Balanced Sample Sizes, Equal Covariance Matrices and $p = 10$

			$(n_1,n_2)=(20,20)$			$(n_1,n_2)=(50,50)$			$(n_1,n_2)=(100,100)$		
ε	μ	κ	**CA**	**RLDA$_{MQ}$**	**RLDA$_{WMQ}$**	**CA**	**RLDA$_{MQ}$**	**RLDA$_{WMQ}$**	**CA**	**RLDA$_{MQ}$**	**RLDA$_{WMQ}$**
-	-	-	0.0980	0.1013	0.0992	0.0707	0.0718	0.0711	0.0635	0.0640	0.0636
10	**3**	-	0.4202	0.3058	0.3534	0.3629	0.1984	0.2946	0.3102	0.1386	0.2391
10	**5**	-	0.4996	0.2734	0.3133	0.5003	0.1628	0.2513	0.4995	0.1142	0.2595
10	**0**	**9**	0.1421	0.1056	0.1053	0.1426	0.0742	0.0765	0.1078	0.0650	0.0671
10	**0**	**25**	0.1521	0.1053	0.1051	0.2256	0.0742	0.0770	0.1745	0.0651	0.0676
10	**0**	**100**	0.1540	0.1048	0.1047	0.3263	0.0742	0.0770	0.3968	0.0651	0.0678
10	**3**	**9**	0.1979	0.1165	0.1229	0.2392	0.0786	0.0889	0.2223	0.0674	0.0756
10	**3**	**25**	0.1616	0.1067	0.1090	0.2563	0.0747	0.0794	0.2549	0.0654	0.0693
10	**3**	**100**	0.1547	0.1049	0.1054	0.3348	0.0742	0.0775	0.4292	0.0651	0.0681
10	**5**	**9**	0.2581	0.1320	0.1444	0.3294	0.0860	0.1065	0.3637	0.0711	0.0887
10	**5**	**25**	0.1747	0.1091	0.1125	0.2869	0.0756	0.0823	0.3404	0.0658	0.0712
10	**5**	**100**	0.1558	0.1050	0.1060	0.3412	0.0742	0.0779	0.4520	0.0651	0.0683
20	**3**	-	0.5237	0.4689	0.4979	0.5436	0.4436	0.5206	0.5616	0.4121	0.5340
20	**5**	-	0.5432	0.4235	0.4128	0.5787	0.3610	0.2822	0.6115	0.2960	0.5044
20	**0**	**9**	0.1977	0.1157	0.1205	0.1470	0.0771	0.0844	0.1083	0.0666	0.0720
20	**0**	**25**	0.2575	0.1136	0.1171	0.2858	0.0767	0.0840	0.2469	0.0665	0.0726
20	**0**	**100**	0.2864	0.1123	0.1139	0.4678	0.0763	0.0822	0.4671	0.0664	0.0716
20	**3**	**9**	0.3049	0.1396	0.1574	0.4063	0.0882	0.1214	0.4972	0.0726	0.1016
20	**3**	**25**	0.2798	0.1167	0.1236	0.4314	0.0781	0.0905	0.4937	0.0671	0.0776
20	**3**	**100**	0.2879	0.1124	0.1150	0.4884	0.0764	0.0834	0.5017	0.0664	0.0722
20	**5**	**9**	0.3826	0.1738	0.2039	0.5863	0.1069	0.1789	0.7423	0.0828	0.1620
20	**5**	**25**	0.3030	0.1213	0.1309	0.5366	0.0801	0.0975	0.6630	0.0685	0.0838
20	**5**	**100**	0.2897	0.1122	0.1157	0.5027	0.0766	0.0841	0.5242	0.0665	0.0728
40	**3**	-	0.6018	0.6046	0.5995	0.6742	0.6705	0.6703	0.7323	0.7270	0.7281
40	**5**	-	0.5769	0.5482	0.5571	0.6354	0.5765	0.6118	0.6864	0.6070	0.6575
40	**0**	**9**	0.2639	0.1611	0.1746	0.1886	0.0940	0.1197	0.1346	0.0741	0.0900
40	**0**	**25**	0.4187	0.1674	0.1795	0.3927	0.0927	0.1432	0.3367	0.0739	0.1180
40	**0**	**100**	0.4956	0.1434	0.1509	0.4915	0.0851	0.1257	0.4865	0.0703	0.1205
40	**3**	**9**	0.5762	0.2454	0.3202	0.7777	0.1509	0.4159	0.8609	0.1069	0.5053
40	**3**	**25**	0.5214	0.1781	0.2038	0.6076	0.0994	0.1961	0.6915	0.0768	0.1876
40	**3**	**100**	0.5050	0.1439	0.1549	0.5087	0.0855	0.1308	0.5222	0.0704	0.1282
40	**5**	**9**	0.6744	0.3242	0.4236	0.8497	0.2393	0.6307	0.8995	0.1776	0.7829
40	**5**	**25**	0.5792	0.1923	0.2251	0.6973	0.1078	0.2513	0.7897	0.0809	0.2890
40	**5**	**100**	0.5116	0.1457	0.1562	0.5208	0.0859	0.1350	0.5448	0.0703	0.1320
Performance			**2.9%**	**82.4%**	**14.7%**	**2.9%**	**91.2%**	**5.9%**	**2.9%**	**97.1%**	**0%**

From Tables 4.2 and 4.3, CA still maintains the position of the linear model with the least mean values for the misclassification error when $\varepsilon=0$, $\mu=0$, $\kappa=0$ with the robust models following closely. Likewise, just as in the case of p=2, as soon as the percentage of contamination was introduced with moderate and extreme outlier value $(\mu=3,5)$, the best mean values moved to the robust models.

However, considering the other variables, a consistent behavior is now observed with respect to the percentage of contamination (ε). There is a consistent increase in the mean misclassification error as ε increases at constant μ and κ as no exemptions were noticed. This implies that the accurate behavior of linear discriminant models with respect to contamination is better observed at high dimension of variables. In addition, highlighting the model with the highest performance percentage as the sample size increases, RLDA_{MQ} remains the estimator with highest percentage performance as the sample size increases. Furthermore, the notable behavior with the shift in location of the population (μ) increasing from 3 to 5 at $\kappa = 0$ causing a decrease in the mean misclassification error is still consistent with robust estimators. Whereas, CA is observed to display fluctuations in following the other robust linear models. This inconsistent behavior still follows CA even when considering the shift in shape of the population (κ). The robust models are consistent with the pattern of sharp reduction as κ goes from 0 to 9 before its gradual descent to convergence at $\kappa = 25{,}100$ while CA is inconsistent.

Generally, it is observed that the misclassification error is inversely proportional to the dimension of the variables, that is, as p increases, mean misclassification error reduces, except when there is no shift in shape of the population $(\kappa = 0)$. For instance, when considering the increase from p=2 to p=6, the mean misclassification error reduces to about half of its initial value and further decrease was also observed when p increased to 10.This same pattern is not observed for the classical model which did not display such convergence with respect to the increase in the dimension of the variables.

As a result of these values observed in Tables 4.2 and 4.3, RLDA_{MQ} remains the best model even when the dimension of the variable is increased. Although, CA and RLDA_{WMQ} also performed well in certain conditions such as $\varepsilon = 40$, $\mu = 3$ and 5, $\kappa = 0$ for p=6, however the values are not significant. In addition, from the performance percentages, RLDA_{WMQ} has an increased performance rate of 5.9% to 14.7% when the dimension of variables was increased to p=10 although still lagging way behind RLDA_{MQ}. Therefore, the RLDA_{MQ} model performed most favorably especially in the case when the sample sizes of the training data were increased which is a difficult scenario for detecting outliers. Therefore, the RLDA_{MQ} performs better than the classical model and its winsorized counterpart when adopted to data with balanced sample sizes and equal covariance matrix.

4.2.1.2 Unbalanced Sample Sizes

The next set of results discussed are the unbalanced sample sizes; $(n_1, n_2) = \left[(50,20),(100,50),(100,20)\right]$. The results are also considered for variable dimensions $p = 2,6,10$.

Table 4.4

Mean Misclassification Error for Linear Discriminant Models with Unbalanced Sample Sizes, Equal Covariance Matrices and $p = 2$

			$(n_1,n_2)=(50,20)$			$(n_1,n_2)=(100,50)$			$(n_1,n_2)=(100,20)$		
ε	μ	κ	**CA**	**RLDA$_{MQ}$**	**RLDA$_{WMQ}$**	**CA**	**RLDA$_{MQ}$**	**RLDA$_{WMQ}$**	**CA**	**RLDA$_{MQ}$**	**RLDA$_{WMQ}$**
-	-	-	0.2897	0.3106	0.2855	0.2684	0.2746	0.2658	0.3552	0.3771	0.3463
10	**3**	-	0.4885	0.4368	0.4509	0.4836	0.3723	0.4308	0.4997	0.4877	0.4952
10	**5**	-	0.5000	0.4116	0.4327	0.5000	0.3441	0.4213	0.5000	0.4783	0.4914
10	**0**	**9**	0.4881	0.3592	0.3168	0.4909	0.3033	0.2892	0.4998	0.4385	0.4048
10	**0**	**25**	0.5000	0.3664	0.3182	0.5000	0.3083	0.2914	0.5000	0.4458	0.4085
10	**0**	**100**	0.5000	0.3697	0.3181	0.5000	0.3104	0.2920	0.5000	0.4486	0.4095
10	**3**	**9**	0.4949	0.3641	0.3355	0.4988	0.3069	0.3054	0.5000	0.4430	0.4284
10	**3**	**25**	0.5000	0.3667	0.3238	0.5000	0.3089	0.2964	0.5000	0.4463	0.4161
10	**3**	**100**	0.5000	0.3695	0.3194	0.5000	0.3105	0.2933	0.5000	0.4487	0.4114
10	**5**	**9**	0.4983	0.3702	0.3508	0.4999	0.3114	0.3211	0.5000	0.4492	0.4456
10	**5**	**25**	0.5000	0.3675	0.3279	0.5000	0.3096	0.3003	0.5000	0.4473	0.4219
10	**5**	**100**	0.5000	0.3695	0.3203	0.5000	0.3106	0.2940	0.5000	0.4489	0.4127
20	**3**	-	0.5017	0.4969	0.5007	0.5015	0.4946	0.5012	0.5001	0.4997	0.5000
20	**5**	-	0.5040	0.4843	0.4962	0.5050	0.4569	0.5024	0.5002	0.4985	0.4994
20	**0**	**9**	0.4995	0.4234	0.3695	0.4998	0.3579	0.3392	0.5000	0.4843	0.4649
20	**0**	**25**	0.5000	0.4405	0.3633	0.5000	0.3771	0.3397	0.5000	0.4913	0.4644
20	**0**	**100**	0.5000	0.4476	0.3541	0.5000	0.3853	0.3312	0.5000	0.4933	0.4584
20	**3**	**9**	0.4999	0.4330	0.4138	0.5000	0.3691	0.3988	0.5000	0.4879	0.4864
20	**3**	**25**	0.5000	0.4410	0.3764	0.5000	0.3773	0.3555	0.5000	0.4911	0.4716
20	**3**	**100**	0.5000	0.4470	0.3556	0.5000	0.3855	0.3347	0.5000	0.4934	0.4605
20	**5**	**9**	0.5000	0.4437	0.4418	0.5000	0.3836	0.4435	0.5000	0.4917	0.4943
20	**5**	**25**	0.5000	0.4422	0.3861	0.5000	0.3791	0.3686	0.5000	0.4915	0.4764
20	**5**	**100**	0.5000	0.4471	0.3579	0.5000	0.3856	0.3372	0.5000	0.4934	0.4620
40	**3**	-	0.5333	0.5132	0.5342	0.5667	0.5448	0.5679	0.5015	0.5004	0.5017
40	**5**	-	0.5226	0.5244	0.5264	0.5485	0.5611	0.5547	0.5009	0.5020	0.5021
40	**0**	**9**	0.5000	0.4961	0.4888	0.5000	0.4859	0.4884	0.5000	0.4999	0.4999
40	**0**	**25**	0.5000	0.4996	0.4860	0.5000	0.4983	0.4917	0.5000	0.5000	0.4999
40	**0**	**100**	0.5000	0.5000	0.4474	0.5000	0.4998	0.4577	0.5000	0.5000	0.4972
40	**3**	**9**	0.5000	0.4986	0.4985	0.5000	0.4959	0.4999	0.5000	0.5000	0.5000
40	**3**	**25**	0.5000	0.4997	0.4911	0.5000	0.4986	0.4958	0.5000	0.5000	0.4999
40	**3**	**100**	0.5000	0.5000	0.4491	0.5000	0.4998	0.4617	0.5000	0.5000	0.4974
40	**5**	**9**	0.5000	0.4995	0.5000	0.5000	0.4990	0.5008	0.5000	0.5000	0.5000
40	**5**	**25**	0.5000	0.4997	0.4929	0.5000	0.4989	0.4976	0.5000	0.5000	0.5000
40	**5**	**100**	0.5000	0.5000	0.4503	0.5000	0.4998	0.4639	0.5000	0.5000	0.4975
Performance			**2.9%**	**17.7%**	**79.4%**	**2.9%**	**32.4%**	**64.7%**	**11.8%**	**29.4%**	**79.4%**

From Table 4.4, certain observations were noted with respect to the manipulated variables. Firstly, it is observed that as ε increases, the mean misclassification error also increases at constant μ and κ with convergence towards an approximate mean misclassification error of 0.5. Secondly, considering (n_1,n_2), a discussion could not be presented as to what happens when the sample sizes increases in reduces. This is because such a pattern does not exist with respect to the sample size values, thus, the observation drawn is that the maximum mean misclassification also tends to an approximate misclassification error of 0.5 as the sample size increases. From the performance row, due to the convergence of all the linear discriminant models towards an 0.5 mean misclassification, a higher performance percentage is observed as the sample size increased. Although, the model with the highest percentage was $RLDA_{WMQ}$ having the least misclassification error even when there was no contamination. Thirdly, as the shift in location of the population (μ) increased, the mean misclassification error decreased while also converging to an approximate 0.5. Finally, a sharp reduction was not observed as the shift in the shape of the population varied since the prior values were already close or tending to 0.5.

Generally, unlike the case of the balanced sample sizes, at the point where the data has no contamination, $RLDA_{WMQ}$ has the least mean misclassification error. The $RLDA_{WMQ}$ performs better than the CA at this point because the combination of simulation scenario does not satisfy all assumptions of the LDA. Observing the results at 10% contamination, a maximal mean misclassification error value of 0.5000 is observed and the $RLDA_{WMQ}$ model still performs better since the CA and $RLDA_{MQ}$ models attain this high error value faster. This is also similar to the results computed at the 20% and 40% contamination levels.

Therefore, for the case where the variable dimension is p=6 and p=10, the mean misclassification error analysis for CA, $RLDA_{MQ}$ and $RLDA_{WMQ}$ are shown for the unbalanced sample sizes (n_1, n_2) in Tables 4.5 and 4.6 respectively.

Table 4.5

Mean of Misclassification Error for Linear Discriminant Models with Unbalanced Sample Sizes, Equal Covariance Matrices and $p = 6$

			$(n_1,n_2)=(50,20)$			$(n_1,n_2)=(100,50)$			$(n_1,n_2)=(100,20)$		
ε	μ	κ	**CA**	**RLDA$_{MQ}$**	**RLDA$_{WMQ}$**	**CA**	**RLDA$_{MQ}$**	**RLDA$_{WMQ}$**	**CA**	**RLDA$_{MQ}$**	**RLDA$_{WMQ}$**
-	-	-	0.1428	0.1520	0.1417	0.1268	0.1296	0.1261	0.1681	0.1813	0.1638
10	**3**	-	0.4609	0.3673	0.4025	0.4511	0.2841	0.3781	0.4955	0.4580	0.4787
10	**5**	-	0.5000	0.3375	0.3721	0.4998	0.2531	0.3621	0.5000	0.4413	0.4661
10	**0**	**9**	0.3213	0.1758	0.1582	0.3474	0.1416	0.1373	0.4668	0.2334	0.2027
10	**0**	**25**	0.4592	0.1786	0.1587	0.4989	0.1436	0.1384	0.4998	0.2410	0.2055
10	**0**	**100**	0.4972	0.1800	0.1588	0.5000	0.1445	0.1387	0.5000	0.2446	0.2064
10	**3**	**9**	0.4043	0.1862	0.1772	0.4634	0.1477	0.1524	0.4946	0.2529	0.2429
10	**3**	**25**	0.4637	0.1800	0.1630	0.4996	0.1445	0.1419	0.4999	0.2441	0.2151
10	**3**	**100**	0.4969	0.1801	0.1596	0.5000	0.1446	0.1395	0.5000	0.2449	0.2083
10	**5**	**9**	0.4547	0.2032	0.2019	0.4934	0.1576	0.1742	0.4992	0.2826	0.2925
10	**5**	**25**	0.4696	0.1827	0.1675	0.4999	0.1457	0.1454	0.5000	0.2482	0.2247
10	**5**	**100**	0.4968	0.1803	0.1603	0.5000	0.1447	0.1401	0.5000	0.2454	0.2100
20	**3**	-	0.5104	0.4868	0.5044	0.5110	0.4779	0.5079	0.5010	0.4980	0.5007
20	**5**	-	0.5212	0.4559	0.4781	0.5238	0.4171	0.5003	0.5022	0.4897	0.4960
20	**0**	**9**	0.4684	0.2246	0.1914	0.4793	0.1673	0.1606	0.4996	0.3224	0.2795
20	**0**	**25**	0.4999	0.2391	0.1862	0.5000	0.1774	0.1609	0.5000	0.3494	0.2772
20	**0**	**100**	0.5000	0.2457	0.1789	0.5000	0.1824	0.1566	0.5000	0.3615	0.2656
20	**3**	**9**	0.4972	0.2553	0.2529	0.4999	0.1886	0.2262	0.5000	0.3683	0.3841
20	**3**	**25**	0.5000	0.2430	0.1991	0.5000	0.1803	0.1735	0.5000	0.3554	0.3032
20	**3**	**100**	0.5000	0.2458	0.1812	0.5000	0.1826	0.1586	0.5000	0.3619	0.2701
20	**5**	**9**	0.4996	0.3018	0.3175	0.5000	0.2232	0.3143	0.5000	0.4199	0.4496
20	**5**	**25**	0.5000	0.2501	0.2110	0.5000	0.1849	0.1859	0.5000	0.3650	0.3258
20	**5**	**100**	0.5000	0.2465	0.1833	0.5000	0.1830	0.1604	0.5000	0.3630	0.2743
40	**3**	-	0.5637	0.5379	0.5646	0.5861	0.5644	0.5876	0.5088	0.5041	0.5096
40	**5**	-	0.5507	0.5412	0.5497	0.5659	0.5680	0.5722	0.5070	0.5100	0.5118
40	**0**	**9**	0.4995	0.4279	0.4052	0.4997	0.3397	0.3743	0.5000	0.4924	0.4929
40	**0**	**25**	0.5000	0.4784	0.4158	0.5000	0.4303	0.4200	0.5000	0.4995	0.4968
40	**0**	**100**	0.5000	0.4923	0.3075	0.5000	0.4685	0.3175	0.5000	0.5000	0.4600
40	**3**	**9**	0.5000	0.4753	0.4892	0.5000	0.4415	0.4979	0.5000	0.4991	0.5000
40	**3**	**25**	0.5000	0.4813	0.4433	0.5000	0.4430	0.4638	0.5000	0.4997	0.4986
40	**3**	**100**	0.5000	0.4925	0.3135	0.5000	0.4686	0.3272	0.5000	0.5000	0.4639
40	**5**	**9**	0.5000	0.4925	0.4993	0.5000	0.4857	0.5036	0.5000	0.4999	0.5000
40	**5**	**25**	0.5000	0.4855	0.4567	0.5000	0.4544	0.4812	0.5000	0.4998	0.4993
40	**5**	**100**	0.5000	0.4923	0.3168	0.5000	0.4699	0.3350	0.5000	0.5000	0.4670
Performance			**0%**	**26.5%**	**73.5%**	**2.9%**	**44.1%**	**53%**	**0%**	**35.3%**	**64.7%**

Table 4.6

Mean of Misclassification Error for Linear Discriminant Models with Unbalanced Sample Sizes, Equal Covariance Matrices and $p = 10$

			$(n_1,n_2)=(50,20)$			$(n_1,n_2)=(100,50)$			$(n_1,n_2)=(100,20)$		
ε	μ	κ	**CA**	**RLDA$_{MQ}$**	**RLDA$_{WMQ}$**	**CA**	**RLDA$_{MQ}$**	**RLDA$_{WMQ}$**	**CA**	**RLDA$_{MQ}$**	**RLDA$_{WMQ}$**
-	-	-	0.0862	0.0913	0.0860	0.0707	0.0725	0.0704	0.0958	0.1035	0.0936
10	**3**	-	0.4478	0.3452	0.3886	0.4323	0.2609	0.3608	0.4880	0.4367	0.4627
10	**5**	-	0.5003	0.3193	0.3565	0.4998	0.2285	0.3455	0.4999	0.4215	0.4457
10	**0**	**9**	0.1823	0.1039	0.0970	0.2064	0.0791	0.0783	0.3366	0.1323	0.1167
10	**0**	**25**	0.2412	0.1052	0.0975	0.4644	0.0800	0.0791	0.4811	0.1361	0.1187
10	**0**	**100**	0.2609	0.1057	0.0973	0.5000	0.0803	0.0794	0.4997	0.1378	0.1192
10	**3**	**9**	0.2785	0.1169	0.1180	0.3803	0.0862	0.0950	0.4511	0.1547	0.1561
10	**3**	**25**	0.2584	0.1068	0.1016	0.4775	0.0809	0.0824	0.4853	0.1391	0.1262
10	**3**	**100**	0.2622	0.1057	0.0981	0.5000	0.0805	0.0800	0.4994	0.1381	0.1206
10	**5**	**9**	0.3597	0.1387	0.1477	0.4621	0.0986	0.1211	0.4872	0.1931	0.2134
10	**5**	**25**	0.2799	0.1096	0.1062	0.4870	0.0826	0.0862	0.4902	0.1444	0.1351
10	**5**	**100**	0.2638	0.1059	0.0987	0.5000	0.0806	0.0805	0.4993	0.1384	0.1217
20	**3**	-	0.5149	0.4801	0.5046	0.5182	0.4677	0.5111	0.5030	0.4949	0.5018
20	**5**	-	0.5283	0.4452	0.4548	0.5358	0.4047	0.4858	0.5060	0.4813	0.4902
20	**0**	**9**	0.3432	0.1314	0.1180	0.4000	0.0932	0.0937	0.4913	0.1942	0.1671
20	**0**	**25**	0.4933	0.1376	0.1156	0.5000	0.0977	0.0937	0.5000	0.2127	0.1653
20	**0**	**100**	0.5000	0.1399	0.1109	0.5000	0.1000	0.0909	0.5000	0.2208	0.1558
20	**3**	**9**	0.4666	0.1701	0.1798	0.4984	0.1162	0.1561	0.4999	0.2596	0.2870
20	**3**	**25**	0.4955	0.1427	0.1261	0.5000	0.1010	0.1043	0.5000	0.2217	0.1878
20	**3**	**100**	0.5000	0.1406	0.1131	0.5000	0.1003	0.0926	0.5000	0.2218	0.1600
20	**5**	**9**	0.4921	0.2285	0.2523	0.5000	0.1584	0.2533	0.5000	0.3492	0.3929
20	**5**	**25**	0.4974	0.1523	0.1385	0.5000	0.1064	0.1162	0.5000	0.2375	0.2129
20	**5**	**100**	0.5000	0.1411	0.1142	0.5000	0.1006	0.0938	0.5000	0.2226	0.1626
40	**3**	-	0.5749	0.5545	0.5749	0.5998	0.5812	0.6006	0.5192	0.5105	0.5204
40	**5**	-	0.5590	0.5366	0.5497	0.5788	0.5574	0.5773	0.5158	0.5150	0.5196
40	**0**	**9**	0.4924	0.3139	0.2928	0.4970	0.2108	0.2562	0.5000	0.4503	0.4598
40	**0**	**25**	0.5000	0.4016	0.3171	0.5000	0.3035	0.3220	0.5000	0.4912	0.4804
40	**0**	**100**	0.5000	0.4397	0.2100	0.5000	0.3628	0.2154	0.5000	0.4978	0.3897
40	**3**	**9**	0.5000	0.4278	0.4669	0.5000	0.3700	0.4937	0.5000	0.4947	0.4996
40	**3**	**25**	0.5000	0.4153	0.3635	0.5000	0.3285	0.4034	0.5000	0.4936	0.4911
40	**3**	**100**	0.5000	0.4396	0.2147	0.5000	0.3642	0.2258	0.5000	0.4979	0.3961
40	**5**	**9**	0.5000	0.4724	0.4948	0.5000	0.4623	0.5068	0.5000	0.4994	0.5000
40	**5**	**25**	0.5000	0.4325	0.3926	0.5000	0.3621	0.4505	0.5000	0.4963	0.4962
40	**5**	**100**	0.5000	0.4407	0.2200	0.5000	0.3666	0.2331	0.5000	0.4980	0.4023
Performance			**0%**	**35.3%**	**64.7%**	**0%**	**61.8%**	**38.2%**	**0%**	**38.2%**	**61.8%**

Considering all variables in Tables 4.5 and 4.6, the following conclusions were drawn. An increase in the percentage of contamination, brought about a simultaneous increase in the mean misclassification error. Although, an in-depth insight was not possible for the CA approach as there was a limitation of all misclassification error values being equal or tending towards 0.5. This setback was not observed with the robust models as the increase in the dimension of variables had brought about a convergent reduction in the mean misclassification error. Therefore, the pattern of the errors with respect to the variables was better understood with the robust LDAs than the conventional LDA for unbalanced sample sizes. Other observations includes reduction in mean misclassification error as the shift in location of the population (μ) increased, a sharp reduction observed in the robust models only as the shift in the shape of the population increased, and a higher performance percentage observed from the $RLDA_{MQ}$ as p increased although on a general performance scale, $RLDA_{WMQ}$ had greater frequency.

The classical approach falls short in competing with the robust models and the $RLDA_{WMQ}$ still maintains the position of the linear model with the least mean values for the misclassification error when $\varepsilon = 0,\ \mu = 0,\ \kappa = 0$. When the contamination was introduced, at the point where there is little (say $\kappa = 9$) or no shift in $(\kappa = 0)$, $RLDA_{MQ}$ was seen to perform better. This is observed for both cases of balanced and unbalanced sample sizes. This implies that the $RLDA_{MQ}$ cannot handle large concentration of contamination. Therefore, considering the general performance percentage results, $RLDA_{WMQ}$ competes best when adapted to data with unbalanced sample sizes and equal covariance matrix except for the case when $(n_1, n_2) = (100, 50)$ where $RLDA_{MQ}$ had a higher performance percentage. However, this singular performance value is not consistent enough to earn the $RLDA_{MQ}$ model to be better than $RLDA_{WMQ}$.

This leads to the discussion on the analysis for datasets with unequal covariance matrix.

4.2.2 Unequal Covariance Matrices

Similar to the case for the equal covariance matrices, the combination of the various variable settings as highlighted in Table 3.1 also gave rise to 54 uncontaminated, 324 location contamination, 486 shape contamination and 972 location and shape contamination. This produces a total of 1836 different data conditions. The following subsections will highlight the results obtained for both the balanced and unbalanced sample sizes per variable dimension.

4.2.2.1 Balanced Sample Sizes

This subsection considers the balanced sample sizes; $(n_1, n_2) = [(20,20),(50,50),(100,100)]$ for variable dimension $p = 2,6,10$. The mean misclassification error is computed for each linear discriminant model.

Table 4.7

Mean Misclassification Error for Linear Discriminant Models with Balanced Sample Sizes, Unequal Covariance Matrices and $p = 2$

			$(n_1, n_2) = (20, 20)$			$(n_1, n_2) = (50, 50)$			$(n_1, n_2) = (100, 100)$		
ε	μ	κ	**CA**	**RLDA$_{MQ}$**	**RLDA$_{WMQ}$**	**CA**	**RLDA$_{MQ}$**	**RLDA$_{WMQ}$**	**CA**	**RLDA$_{MQ}$**	**RLDA$_{WMQ}$**
-	-	-	0.3169	0.3192	0.3177	0.3069	0.3078	0.3071	0.3038	0.3043	0.3039
10	**3**	-	0.3863	0.3652	0.3740	0.3512	0.3316	0.3407	0.3302	0.3192	0.3236
10	**5**	-	0.4850	0.3694	0.4032	0.4896	0.3382	0.3788	0.4931	0.3267	0.3616
10	**0**	**9**	0.3620	0.3225	0.3242	0.3294	0.3086	0.3096	0.3152	0.3047	0.3052
10	**0**	**25**	0.4366	0.3213	0.3252	0.4106	0.3084	0.3103	0.3781	0.3045	0.3056
10	**0**	**100**	0.4903	0.3203	0.3249	0.4865	0.3079	0.3105	0.4805	0.3045	0.3057
10	**3**	**9**	0.4189	0.3246	0.3347	0.3969	0.3102	0.3164	0.3713	0.3058	0.3101
10	**3**	**25**	0.4623	0.3213	0.3285	0.4558	0.3084	0.3124	0.4395	0.3047	0.3072
10	**3**	**100**	0.4943	0.3202	0.3251	0.4931	0.3079	0.3110	0.4925	0.3045	0.3061
10	**5**	**9**	0.4693	0.3268	0.3435	0.4784	0.3114	0.3234	0.4846	0.3068	0.3149
10	**5**	**25**	0.4805	0.3216	0.3305	0.4890	0.3084	0.3137	0.4921	0.3047	0.3084
10	**5**	**100**	0.4964	0.3204	0.3257	0.4977	0.3079	0.3113	0.5008	0.3045	0.3064
20	**3**	-	0.5366	0.4941	0.5197	0.5718	0.4739	0.5478	0.6024	0.4581	0.5770
20	**5**	-	0.6182	0.5141	0.5759	0.6546	0.4753	0.6170	0.6702	0.4483	0.6427
20	**0**	**9**	0.3917	0.3294	0.3367	0.3511	0.3107	0.3151	0.3278	0.3057	0.3083
20	**0**	**25**	0.4691	0.3261	0.3375	0.4411	0.3099	0.3179	0.4146	0.3053	0.3101
20	**0**	**100**	0.4987	0.3232	0.3326	0.4911	0.3086	0.3158	0.4891	0.3049	0.3102
20	**3**	**9**	0.5036	0.3389	0.3700	0.5231	0.3163	0.3464	0.5486	0.3094	0.3321
20	**3**	**25**	0.5016	0.3273	0.3452	0.5051	0.3101	0.3254	0.5180	0.3058	0.3167
20	**3**	**100**	0.5025	0.3231	0.3340	0.4987	0.3087	0.3172	0.5014	0.3050	0.3110
20	**5**	**9**	0.5772	0.3467	0.4028	0.6342	0.3208	0.3945	0.6638	0.3125	0.3886
20	**5**	**25**	0.5242	0.3281	0.3515	0.5486	0.3104	0.3313	0.5845	0.3060	0.3234
20	**5**	**100**	0.5052	0.3230	0.3348	0.5037	0.3087	0.3182	0.5115	0.3050	0.3122
40	**3**	-	0.6568	0.6541	0.6555	0.6798	0.6766	0.6790	0.6886	0.6858	0.6880
40	**5**	-	0.6566	0.6485	0.6521	0.6793	0.6570	0.6751	0.6879	0.6596	0.6840
40	**0**	**9**	0.4270	0.3648	0.3862	0.3820	0.3269	0.3457	0.3495	0.3130	0.3244
40	**0**	**25**	0.4813	0.3734	0.4069	0.4662	0.3319	0.3791	0.4459	0.3141	0.3542
40	**0**	**100**	0.4972	0.3546	0.3886	0.4975	0.3209	0.3807	0.4935	0.3099	0.3755
40	**3**	**9**	0.5860	0.4280	0.5057	0.6370	0.3846	0.5458	0.6674	0.3539	0.5837
40	**3**	**25**	0.5160	0.3798	0.4316	0.5453	0.3398	0.4308	0.5727	0.3189	0.4274
40	**3**	**100**	0.5005	0.3577	0.3918	0.5039	0.3205	0.3831	0.5073	0.3106	0.3863
40	**5**	**9**	0.6355	0.4747	0.5703	0.6701	0.4465	0.6249	0.6842	0.4232	0.6516
40	**5**	**25**	0.5420	0.3850	0.4505	0.5858	0.3459	0.4735	0.6217	0.3232	0.4928
40	**5**	**100**	0.5023	0.3570	0.3959	0.5085	0.3218	0.3903	0.5162	0.3107	0.3921
Performance			**2.9%**	**97.1%**	**0%**	**2.9%**	**97.1%**	**0%**	**2.9%**	**97.1%**	**0%**

Table 4.7 shows the mean misclassification error just like Table 4.1 but in this case, the covariance matrices are unequal. In discussing each of the variables, similar behavior is exhibited between Table 4.1 and Table 4.7. For instance, considering the percentage of contamination (ε), it is similarly observed that as ε increases, the mean misclassification error also increases at constant μ and κ. Certain exemptions

were again noticed for CA at $(n_1, n_2) = (20, 20)$ where the mean misclassification error of $(\varepsilon, \mu, \kappa) = (20, 0, 100)$ which is 0.4987 is greater than the mean misclassification error of $(\varepsilon, \mu, \kappa) = (40, 0, 100)$ which is 0.4972. Similarly, when ε increases from 20% to 40% for $(\mu, \kappa) = (3, 100)$ and $(\mu, \kappa) = (5, 100)$, same anomality is observed with CA as $0.5025 > 0.5005$ and $0.5052 > 0.5023$ respectively. Although, as (n_1, n_2) increased, there is stability as all mean misclassification errors increased with increasing ε.

Considering (n_1, n_2), the discussion focuses on the linear discriminant model with the highest performance percentage as the sample size increases. CA maintains a performance percentage of 2.9% at all sample sizes and this performance is obtained when the data has no contamination. RLDA_{WMQ} on the other hand shows no performance as its percentage takes value of 0% all through. This leaves the RLDA_{MQ} as the estimator with highest percentage performance as the sample size increases.

The change in shift in location of the population (μ) helped draw the conclusion that when $\kappa = 0$ with increasing ε, linear discriminant models have better accuracy in detecting of outliers. This is because of the behavior of μ as it increased from 3 to 5 at $\kappa = 0$, a decrease was recorded in the mean misclassification error. Although, this observation was not consistent at 10% and 20% contamination, it was consistent as 40% contamination. On the other hand, increase in the shift in shape of the population (κ) showed decrease in the mean misclassification error.

Generally, the introduction of contamination makes the mean misclassification error is seen to increase for CA making the RLDA_{MQ} the best approach in the presence of contamination. The mean misclassification error results for $p = 6,10$ using CA, RLDA$_{MQ}$ and RLDA$_{WMQ}$ are shown for the balanced sample sizes (n_1, n_2) in Tables 4.8 and 4.9 respectively.

Table 4.8

Mean of Misclassification Error for Linear Discriminant Models with Balanced Sample Sizes, Unequal Covariance Matrices and $p = 6$

			$(n_1,n_2)=(20,20)$			$(n_1,n_2)=(50,50)$			$(n_1,n_2)=(100,100)$		
ε	μ	κ	**CA**	**RLDA$_{MQ}$**	**RLDA$_{WMQ}$**	**CA**	**RLDA$_{MQ}$**	**RLDA$_{WMQ}$**	**CA**	**RLDA$_{MQ}$**	**RLDA$_{WMQ}$**
-	-	-	0.2342	0.2376	0.2354	0.2069	0.2092	0.2075	0.1986	0.1997	0.1989
10	**3**	-	0.3842	0.3395	0.3602	0.3400	0.2824	0.3132	0.2980	0.2468	0.2752
10	**5**	-	0.4715	0.3605	0.4021	0.4817	0.3047	0.3855	0.4843	0.2674	0.3637
10	**0**	**9**	0.2722	0.2428	0.2431	0.2439	0.2119	0.2141	0.2215	0.2010	0.2026
10	**0**	**25**	0.3090	0.2418	0.2432	0.3190	0.2119	0.2154	0.2829	0.2010	0.2036
10	**0**	**100**	0.3301	0.2411	0.2432	0.4511	0.2116	0.2157	0.4375	0.2008	0.2040
10	**3**	**9**	0.3246	0.2506	0.2593	0.3256	0.2170	0.2289	0.2979	0.2034	0.2126
10	**3**	**25**	0.3217	0.2425	0.2471	0.3678	0.2127	0.2196	0.3542	0.2012	0.2063
10	**3**	**100**	0.3310	0.2410	0.2439	0.4650	0.2116	0.2166	0.4598	0.2009	0.2045
10	**5**	**9**	0.3731	0.2610	0.2774	0.4150	0.2234	0.2478	0.4266	0.2069	0.2268
10	**5**	**25**	0.3345	0.2443	0.2508	0.4110	0.2135	0.2232	0.4268	0.2016	0.2088
10	**5**	**100**	0.3320	0.2415	0.2446	0.4738	0.2118	0.2171	0.4751	0.2009	0.2049
20	**3**	-	0.5067	0.4776	0.4942	0.5413	0.4712	0.5247	0.5696	0.4548	0.5476
20	**5**	-	0.5429	0.4878	0.5174	0.5986	0.4707	0.5624	0.6438	0.4552	0.6019
20	**0**	**9**	0.3053	0.2508	0.2556	0.2639	0.2165	0.2234	0.2311	0.2029	0.2076
20	**0**	**25**	0.3911	0.2494	0.2556	0.3828	0.2156	0.2272	0.3323	0.2026	0.2111
20	**0**	**100**	0.4764	0.2463	0.2521	0.4878	0.2140	0.2250	0.4699	0.2020	0.2105
20	**3**	**9**	0.4258	0.2699	0.2973	0.4984	0.2297	0.2731	0.5267	0.2100	0.2482
20	**3**	**25**	0.4402	0.2515	0.2638	0.4985	0.2173	0.2382	0.4995	0.2034	0.2197
20	**3**	**100**	0.4841	0.2466	0.2542	0.5034	0.2142	0.2267	0.4957	0.2020	0.2119
20	**5**	**9**	0.5057	0.2922	0.3400	0.6349	0.2466	0.3430	0.7026	0.2210	0.3308
20	**5**	**25**	0.4736	0.2541	0.2719	0.5733	0.2190	0.2490	0.6157	0.2043	0.2291
20	**5**	**100**	0.4899	0.2467	0.2551	0.5133	0.2143	0.2284	0.5129	0.2020	0.2129
40	**3**	-	0.6162	0.6123	0.6133	0.6900	0.6781	0.6867	0.7341	0.7221	0.7312
40	**5**	-	0.5958	0.6079	0.5913	0.6664	0.6466	0.6585	0.7129	0.6721	0.7042
40	**0**	**9**	0.3590	0.2917	0.3072	0.3021	0.2390	0.2584	0.2547	0.2143	0.2263
40	**0**	**25**	0.4607	0.3081	0.3290	0.4357	0.2457	0.3044	0.3893	0.2178	0.2623
40	**0**	**100**	0.4984	0.2930	0.3050	0.4948	0.2341	0.2985	0.4854	0.2122	0.2926
40	**3**	**9**	0.6071	0.3711	0.4643	0.7038	0.3192	0.5485	0.7508	0.2742	0.6079
40	**3**	**25**	0.5315	0.3176	0.3625	0.5844	0.2568	0.3731	0.6238	0.2248	0.3610
40	**3**	**100**	0.5035	0.2922	0.3099	0.5101	0.2346	0.3049	0.5127	0.2125	0.3033
40	**5**	**9**	0.6737	0.4273	0.5438	0.7499	0.4079	0.6526	0.7798	0.3777	0.7121
40	**5**	**25**	0.5713	0.3281	0.3885	0.6449	0.2669	0.4339	0.6985	0.2322	0.4687
40	**5**	**100**	0.5067	0.2922	0.3121	0.5185	0.2354	0.3108	0.5294	0.2128	0.3110
Performance			**2.9%**	**94.2%**	**2.9%**	**2.9%**	**97.1%**	**0%**	**2.9%**	**97.1%**	**0%**

Table 4.9

Mean of Misclassification Error for Linear Discriminant Models with Balanced Sample Sizes, Unequal Covariance Matrices and $p = 10$

			$(n_1,n_2)=(20,20)$			$(n_1,n_2)=(50,50)$			$(n_1,n_2)=(100,100)$		
ε	μ	κ	**CA**	**RLDA$_{MQ}$**	**RLDA$_{WMQ}$**	**CA**	**RLDA$_{MQ}$**	**RLDA$_{WMQ}$**	**CA**	**RLDA$_{MQ}$**	**RLDA$_{WMQ}$**
-	-	-	0.2005	0.2041	0.2019	0.1607	0.1630	0.1612	0.1483	0.1496	0.1486
10	**3**	-	0.3985	0.3538	0.3738	0.3527	0.2802	0.3220	0.3107	0.2311	0.2781
10	**5**	-	0.4647	0.3791	0.4090	0.4755	0.3137	0.3977	0.4803	0.2642	0.3808
10	**0**	**9**	0.2282	0.2089	0.2077	0.2019	0.1653	0.1672	0.1776	0.1510	0.1530
10	**0**	**25**	0.2411	0.2081	0.2077	0.2697	0.1653	0.1686	0.2409	0.1512	0.1548
10	**0**	**100**	0.2442	0.2073	0.2079	0.3618	0.1651	0.1689	0.4123	0.1511	0.1554
10	**3**	**9**	0.2791	0.2224	0.2292	0.2883	0.1734	0.1872	0.2767	0.1556	0.1682
10	**3**	**25**	0.2505	0.2099	0.2125	0.3012	0.1663	0.1730	0.3086	0.1517	0.1581
10	**3**	**100**	0.2449	0.2073	0.2088	0.3694	0.1651	0.1697	0.4395	0.1511	0.1560
10	**5**	**9**	0.3286	0.2406	0.2540	0.3705	0.1858	0.2152	0.3992	0.1632	0.1927
10	**5**	**25**	0.2623	0.2124	0.2173	0.3301	0.1680	0.1775	0.3774	0.1527	0.1617
10	**5**	**100**	0.2458	0.2076	0.2093	0.3750	0.1651	0.1704	0.4585	0.1511	0.1565
20	**3**	-	0.4880	0.4689	0.4789	0.5200	0.4672	0.5074	0.5461	0.4557	0.5284
20	**5**	-	0.5096	0.4757	0.4890	0.5615	0.4667	0.5335	0.6046	0.4563	0.5683
20	**0**	**9**	0.2593	0.2199	0.2211	0.2167	0.1703	0.1766	0.1848	0.1538	0.1595
20	**0**	**25**	0.3083	0.2202	0.2212	0.3274	0.1702	0.1802	0.2961	0.1537	0.1640
20	**0**	**100**	0.3397	0.2162	0.2188	0.4694	0.1688	0.1786	0.4690	0.1533	0.1634
20	**3**	**9**	0.3594	0.2484	0.2674	0.4358	0.1897	0.2355	0.5008	0.1657	0.2147
20	**3**	**25**	0.3330	0.2238	0.2307	0.4448	0.1725	0.1916	0.4941	0.1552	0.1746
20	**3**	**100**	0.3417	0.2164	0.2203	0.4884	0.1691	0.1805	0.5019	0.1532	0.1648
20	**5**	**9**	0.4235	0.2816	0.3133	0.5723	0.2194	0.3125	0.6857	0.1868	0.3151
20	**5**	**25**	0.3552	0.2282	0.2389	0.5288	0.1761	0.2033	0.6285	0.1575	0.1867
20	**5**	**100**	0.3435	0.2167	0.2216	0.5015	0.1693	0.1820	0.5230	0.1533	0.1660
40	**3**	-	0.5684	0.5657	0.5656	0.6572	0.6435	0.6528	0.7183	0.7003	0.7138
40	**5**	-	0.5484	0.5730	0.5457	0.6252	0.6183	0.6172	0.6833	0.6468	0.6729
40	**0**	**9**	0.3120	0.2580	0.2628	0.2521	0.1929	0.2092	0.2075	0.1660	0.1789
40	**0**	**25**	0.4299	0.2812	0.2782	0.4074	0.2035	0.2515	0.3635	0.1718	0.2205
40	**0**	**100**	0.4950	0.2670	0.2581	0.4899	0.1928	0.2463	0.4858	0.1654	0.2490
40	**3**	**9**	0.5587	0.3436	0.4084	0.7176	0.2877	0.5123	0.7867	0.2431	0.6005
40	**3**	**25**	0.5173	0.2905	0.3071	0.5916	0.2152	0.3208	0.6572	0.1793	0.3244
40	**3**	**100**	0.5038	0.2672	0.2611	0.5078	0.1938	0.2544	0.5204	0.1655	0.2586
40	**5**	**9**	0.6284	0.4009	0.4834	0.7771	0.3848	0.6334	0.8220	0.3654	0.7232
40	**5**	**25**	0.5662	0.3023	0.3320	0.6692	0.2294	0.3853	0.7421	0.1893	0.4454
40	**5**	**100**	0.5100	0.2682	0.2647	0.5199	0.1944	0.2601	0.5428	0.1659	0.2673
Performance			**2.9%**	**73.5%**	**23.6%**	**2.9%**	**94.2%**	**2.9%**	**2.9%**	**97.1%**	**0%**

From Tables 4.8 and 4.9, CA still maintains the position of the linear model with the least mean values for the misclassification error when $\varepsilon = 0$, $\mu = 0$, $\kappa = 0$. As the percentage of contamination was introduced with moderate and extreme outlier values, the best mean values moved to the robust models. A consistent behavior is observed in the percentage of contamination (ε) because as ε increases at constant

μ and κ the mean misclassification error increases. In terms of performance, RLDA_{MQ} stayed as the estimator with highest percentage performance as the sample size increases. A consistent pattern was not observed with the shift in location of the population (μ) and the shift in shape of the population (κ). However, from observation, the conclusion that can be drawn that the mean misclassification error is inversely proportional to the dimension of the variables, that is, as p increases, mean misclassification error reduces.

4.2.2.2 Unbalanced Sample Sizes

The next set of results discussed are the unbalanced sample sizes; $(n_1, n_2) = [(50, 20), (100, 50), (100, 20)]$. The results will also be considered for variable dimensions $p = 2, 6, 10$.

Table 4.10

Mean Misclassification Error for Linear Discriminant Models with Unbalanced Sample Sizes, Unequal Covariance Matrices and $p = 2$

			$(n_1,n_2)=(50,20)$			$(n_1,n_2)=(100,50)$			$(n_1,n_2)=(100,20)$		
ε	μ	κ	**CA**	**RLDA$_{MQ}$**	**RLDA$_{WMQ}$**	**CA**	**RLDA$_{MQ}$**	**RLDA$_{WMQ}$**	**CA**	**RLDA$_{MQ}$**	**RLDA$_{WMQ}$**
-	-	-	0.3267	0.3533	0.3231	0.3059	0.3145	0.3031	0.3608	0.3873	0.3545
10	**3**	-	0.4571	0.4398	0.4326	0.4557	0.4035	0.4254	0.4856	0.4715	0.4718
10	**5**	-	0.4826	0.4442	0.4481	0.4902	0.4065	0.4573	0.4940	0.4739	0.4787
10	**0**	**9**	0.4675	0.3913	0.3499	0.4678	0.3432	0.3259	0.4965	0.4303	0.3950
10	**0**	**25**	0.4995	0.3998	0.3507	0.4999	0.3513	0.3306	0.5000	0.4389	0.3977
10	**0**	**100**	0.5000	0.4042	0.3509	0.5000	0.3551	0.3318	0.5000	0.4424	0.3981
10	**3**	**9**	0.4856	0.3970	0.3689	0.4929	0.3486	0.3479	0.4993	0.4359	0.4166
10	**3**	**25**	0.4996	0.4005	0.3572	0.5000	0.3521	0.3374	0.5000	0.4393	0.4051
10	**3**	**100**	0.5000	0.4040	0.3522	0.5000	0.3552	0.3333	0.5000	0.4424	0.3996
10	**5**	**9**	0.4934	0.4021	0.3831	0.4985	0.3547	0.3692	0.4998	0.4413	0.4323
10	**5**	**25**	0.4997	0.4012	0.3617	0.5000	0.3532	0.3430	0.5000	0.4400	0.4099
10	**5**	**100**	0.5000	0.4039	0.3532	0.5000	0.3554	0.3346	0.5000	0.4425	0.4009
20	**3**	-	0.4840	0.4860	0.4804	0.4905	0.4871	0.4881	0.4936	0.4948	0.4913
20	**5**	-	0.4851	0.4853	0.4798	0.4887	0.4842	0.4858	0.4939	0.4940	0.4895
20	**0**	**9**	0.4946	0.4327	0.3893	0.4958	0.3881	0.3688	0.4999	0.4679	0.4428
20	**0**	**25**	0.5000	0.4509	0.3888	0.5000	0.4119	0.3781	0.5000	0.4796	0.4437
20	**0**	**100**	0.5000	0.4594	0.3802	0.5000	0.4235	0.3715	0.5000	0.4843	0.4355
20	**3**	**9**	0.4991	0.4431	0.4276	0.4999	0.4035	0.4306	0.5000	0.4750	0.4717
20	**3**	**25**	0.5000	0.4524	0.4008	0.5000	0.4133	0.3970	0.5000	0.4799	0.4528
20	**3**	**100**	0.5000	0.4591	0.3824	0.5000	0.4233	0.3755	0.5000	0.4843	0.4377
20	**5**	**9**	0.4997	0.4520	0.4498	0.5000	0.4193	0.4652	0.5000	0.4810	0.4842
20	**5**	**25**	0.5000	0.4526	0.4093	0.5000	0.4153	0.4103	0.5000	0.4804	0.4585
20	**5**	**100**	0.5000	0.4591	0.3839	0.5000	0.4237	0.3782	0.5000	0.4845	0.4398
40	**3**	-	0.4855	0.4872	0.4856	0.4964	0.4891	0.4974	0.4906	0.4946	0.4900
40	**5**	-	0.4826	0.4853	0.4823	0.4932	0.4885	0.4936	0.4913	0.4938	0.4894
40	**0**	**9**	0.4997	0.4881	0.4766	0.4999	0.4762	0.4766	0.5000	0.4979	0.4973
40	**0**	**25**	0.5000	0.4981	0.4845	0.5000	0.4963	0.4915	0.5000	0.4998	0.4986
40	**0**	**100**	0.5000	0.4997	0.4577	0.5000	0.4996	0.4735	0.5000	0.5000	0.4917
40	**3**	**9**	0.5000	0.4955	0.4955	0.5000	0.4937	0.4988	0.5000	0.4995	0.4998
40	**3**	**25**	0.5000	0.4983	0.4897	0.5000	0.4968	0.4953	0.5000	0.4999	0.4993
40	**3**	**100**	0.5000	0.4997	0.4610	0.5000	0.4996	0.4756	0.5000	0.5000	0.4921
40	**5**	**9**	0.5000	0.4977	0.4966	0.4999	0.4979	0.4973	0.5000	0.4998	0.4998
40	**5**	**25**	0.5000	0.4986	0.4917	0.5000	0.4974	0.4978	0.5000	0.4999	0.4995
40	**5**	**100**	0.5000	0.4997	0.4618	0.5000	0.4995	0.4772	0.5000	0.5000	0.4920
Performance			**2.9%**	**5.9%**	**94.2%**	**0%**	**35.3%**	**64.7%**	**0%**	**14.7%**	**88.2%**

Table 4.11

Mean of Misclassification Error for Linear Discriminant Models with Unbalanced Sample Sizes, Unequal Covariance Matrices and $p = 6$

			$(n_1,n_2)=(50,20)$			$(n_1,n_2)=(100,50)$			$(n_1,n_2)=(100,20)$		
ε	μ	κ	**CA**	**$\mathrm{RLDA_{MQ}}$**	**$\mathrm{RLDA_{WMQ}}$**	**CA**	**$\mathrm{RLDA_{MQ}}$**	**$\mathrm{RLDA_{WMQ}}$**	**CA**	**$\mathrm{RLDA_{MQ}}$**	**$\mathrm{RLDA_{WMQ}}$**
-	-	-	0.2362	0.2500	0.2351	0.2149	0.2203	0.2138	0.2512	0.2684	0.2481
10	**3**	**-**	0.4037	0.3761	0.3792	0.4013	0.3355	0.3676	0.4440	0.4224	0.4236
10	**5**	**-**	0.4351	0.3870	0.3940	0.4477	0.3492	0.4000	0.4596	0.4274	0.4294
10	**0**	**9**	0.3320	0.2737	0.2509	0.3360	0.2360	0.2273	0.4343	0.3068	0.2774
10	**0**	**25**	0.4439	0.2778	0.2522	0.4931	0.2400	0.2302	0.4989	0.3132	0.2800
10	**0**	**100**	0.4949	0.2794	0.2520	0.5000	0.2418	0.2308	0.5000	0.3157	0.2800
10	**3**	**9**	0.3919	0.2844	0.2710	0.4349	0.2439	0.2466	0.4791	0.3209	0.3067
10	**3**	**25**	0.4499	0.2792	0.2570	0.4965	0.2411	0.2344	0.4995	0.3154	0.2873
10	**3**	**100**	0.4946	0.2794	0.2529	0.5000	0.2419	0.2316	0.5000	0.3161	0.2817
10	**5**	**9**	0.4310	0.2993	0.2927	0.4755	0.2563	0.2719	0.4922	0.3399	0.3375
10	**5**	**25**	0.4567	0.2816	0.2615	0.4983	0.2426	0.2390	0.4997	0.3184	0.2940
10	**5**	**100**	0.4946	0.2797	0.2537	0.5000	0.2420	0.2323	0.5000	0.3165	0.2831
20	**3**	**-**	0.4436	0.4427	0.4382	0.4577	0.4455	0.4513	0.4625	0.4683	0.4559
20	**5**	**-**	0.4519	0.4400	0.4434	0.4671	0.4389	0.4593	0.4652	0.4659	0.4545
20	**0**	**9**	0.4352	0.3139	0.2790	0.4395	0.2639	0.2520	0.4938	0.3613	0.3263
20	**0**	**25**	0.4993	0.3311	0.2816	0.5000	0.2809	0.2596	0.5000	0.3822	0.3304
20	**0**	**100**	0.5000	0.3383	0.2738	0.5000	0.2891	0.2548	0.5000	0.3910	0.3206
20	**3**	**9**	0.4853	0.3384	0.3305	0.4975	0.2873	0.3215	0.4998	0.3902	0.3942
20	**3**	**25**	0.4997	0.3343	0.2933	0.5000	0.2839	0.2750	0.5000	0.3853	0.3470
20	**3**	**100**	0.5000	0.3385	0.2762	0.5000	0.2893	0.2573	0.5000	0.3915	0.3241
20	**5**	**9**	0.4940	0.3676	0.3714	0.4995	0.3203	0.3900	0.5000	0.4193	0.4339
20	**5**	**25**	0.4999	0.3390	0.3029	0.5000	0.2887	0.2897	0.5000	0.3910	0.3615
20	**5**	**100**	0.5000	0.3391	0.2780	0.5000	0.2898	0.2594	0.5000	0.3921	0.3270
40	**3**	**-**	0.4819	0.4753	0.4819	0.5088	0.4960	0.5098	0.4730	0.4786	0.4716
40	**5**	**-**	0.4722	0.4771	0.4719	0.4941	0.4937	0.4941	0.4697	0.4782	0.4653
40	**0**	**9**	0.4932	0.4290	0.4025	0.4938	0.3790	0.3849	0.5000	0.4743	0.4724
40	**0**	**25**	0.5000	0.4765	0.4404	0.5000	0.4537	0.4480	0.5000	0.4955	0.4904
40	**0**	**100**	0.5000	0.4913	0.3916	0.5000	0.4836	0.4161	0.5000	0.4991	0.4628
40	**3**	**9**	0.4997	0.4690	0.4771	0.4999	0.4570	0.4917	0.5000	0.4935	0.4982
40	**3**	**25**	0.5000	0.4787	0.4572	0.5000	0.4623	0.4771	0.5000	0.4965	0.4952
40	**3**	**100**	0.5000	0.4910	0.3938	0.5000	0.4836	0.4206	0.5000	0.4992	0.4651
40	**5**	**9**	0.4998	0.4824	0.4876	0.4998	0.4836	0.4956	0.5000	0.4973	0.4991
40	**5**	**25**	0.5000	0.4817	0.4674	0.5000	0.4698	0.4877	0.5000	0.4971	0.4973
40	**5**	**100**	0.5000	0.4908	0.3972	0.5000	0.4841	0.4243	0.5000	0.4992	0.4675
Performance			**0%**	**23.6%**	**76.4%**	**0%**	**47.1%**	**52.9%**	**0%**	**20.6%**	**79.4%**

Table 4.12

Mean of Misclassification Error for Linear Discriminant Models with Unbalanced Sample Sizes, Unequal Covariance Matrices and $p = 10$

			$(n_1, n_2) = (50, 20)$			$(n_1, n_2) = (100, 50)$			$(n_1, n_2) = (100, 20)$		
ε	μ	κ	**CA**	**RLDA$_{MQ}$**	**RLDA$_{WMQ}$**	**CA**	**RLDA$_{MQ}$**	**RLDA$_{WMQ}$**	**CA**	**RLDA$_{MQ}$**	**RLDA$_{WMQ}$**
-	-	-	0.1950	0.2063	0.1948	0.1703	0.1752	0.1694	0.2060	0.2195	0.2042
10	**3**	-	0.3914	0.3579	0.3712	0.3810	0.3180	0.3522	0.4210	0.3977	0.4029
10	**5**	-	0.4233	0.3715	0.3859	0.4280	0.3361	0.3831	0.4386	0.4049	0.4088
10	**0**	**9**	0.2519	0.2230	0.2073	0.2545	0.1875	0.1816	0.3368	0.2462	0.2252
10	**0**	**25**	0.2912	0.2250	0.2085	0.4377	0.1897	0.1838	0.4639	0.2497	0.2275
10	**0**	**100**	0.3061	0.2251	0.2086	0.4999	0.1907	0.1846	0.4991	0.2505	0.2279
10	**3**	**9**	0.3151	0.2382	0.2320	0.3671	0.1990	0.2055	0.4218	0.2654	0.2589
10	**3**	**25**	0.3033	0.2267	0.2136	0.4578	0.1915	0.1889	0.4712	0.2522	0.2345
10	**3**	**100**	0.3070	0.2250	0.2095	0.4999	0.1909	0.1854	0.4988	0.2507	0.2292
10	**5**	**9**	0.3631	0.2597	0.2605	0.4302	0.2177	0.2384	0.4596	0.2930	0.2977
10	**5**	**25**	0.3174	0.2301	0.2189	0.4721	0.1944	0.1944	0.4783	0.2565	0.2424
10	**5**	**100**	0.3080	0.2248	0.2100	0.4999	0.1911	0.1862	0.4988	0.2509	0.2303
20	**3**	-	0.4330	0.4236	0.4295	0.4420	0.4221	0.4366	0.4438	0.4460	0.4380
20	**5**	-	0.4425	0.4205	0.4343	0.4560	0.4179	0.4488	0.4478	0.4431	0.4382
20	**0**	**9**	0.3387	0.2543	0.2297	0.3621	0.2090	0.2025	0.4649	0.2940	0.2656
20	**0**	**25**	0.4812	0.2637	0.2312	0.4994	0.2198	0.2090	0.5000	0.3102	0.2700
20	**0**	**100**	0.5000	0.2641	0.2267	0.5000	0.2241	0.2050	0.5000	0.3143	0.2621
20	**3**	**9**	0.4377	0.2880	0.2869	0.4861	0.2393	0.2749	0.4973	0.3340	0.3424
20	**3**	**25**	0.4876	0.2678	0.2430	0.4999	0.2242	0.2233	0.5000	0.3147	0.2865
20	**3**	**100**	0.5000	0.2648	0.2291	0.5000	0.2246	0.2073	0.5000	0.3146	0.2647
20	**5**	**9**	0.4670	0.3262	0.3352	0.4965	0.2833	0.3484	0.4992	0.3791	0.3968
20	**5**	**25**	0.4917	0.2757	0.2552	0.5000	0.2317	0.2391	0.5000	0.3237	0.3032
20	**5**	**100**	0.5000	0.2650	0.2303	0.5000	0.2250	0.2091	0.5000	0.3150	0.2669
40	**3**	-	0.4762	0.4682	0.4760	0.5074	0.4941	0.5078	0.4610	0.4647	0.4598
40	**5**	-	0.4651	0.4741	0.4648	0.4900	0.4950	0.4895	0.4555	0.4643	0.4520
40	**0**	**9**	0.4657	0.3671	0.3379	0.4742	0.3098	0.3191	0.4994	0.4346	0.4334
40	**0**	**25**	0.5000	0.4366	0.3871	0.5000	0.3946	0.4003	0.5000	0.4811	0.4747
40	**0**	**100**	0.5000	0.4617	0.3343	0.5000	0.4389	0.3656	0.5000	0.4926	0.4337
40	**3**	**9**	0.4981	0.4355	0.4533	0.4998	0.4184	0.4824	0.5000	0.4815	0.4945
40	**3**	**25**	0.5000	0.4431	0.4147	0.5000	0.4106	0.4500	0.5000	0.4843	0.4864
40	**3**	**100**	0.5000	0.4608	0.3377	0.5000	0.4397	0.3722	0.5000	0.4928	0.4377
40	**5**	**9**	0.4992	0.4614	0.4749	0.4998	0.4632	0.4935	0.5000	0.4919	0.4974
40	**5**	**25**	0.5000	0.4510	0.4329	0.5000	0.4275	0.4740	0.5000	0.4879	0.4924
40	**5**	**100**	0.5000	0.4616	0.3403	0.5000	0.4405	0.3790	0.5000	0.4928	0.4400
Performance			**0%**	**29.4%**	**70.6%**	**0%**	**50%**	**52.9%**	**0%**	**26.5%**	**73.5%**

From Table 4.10, certain observations were noted with respect to the manipulated variables. Firstly, it is observed that as ε increases, the mean misclassification error also increases at constant μ and κ with convergence towards an approximate mean misclassification error of 0.5 for all the linear discriminant models. This is similar to the pattern observed in Table 4.4. This convergence towards an approximate mean misclassification is also observed with changes in (n_1, n_2). From the performance row, due to the convergence of all the linear discriminant models towards an 0.5 mean misclassification error, a higher performance percentage is observed as the sample size increased for all models. Although, the model with the highest percentage was $\mathrm{RLDA_{WMQ}}$ having the least misclassification error even when there was no contamination. As the shift in location of the population and shift in the shape of the population increased, the mean misclassification error decreased while also converging to an approximate 0.5.

As the dimensions of the variables increased, a concise analysis was not made for the CA approach as there was a limitation of all misclassification error values being equal or tending towards 0.5. This setback was not observed with the robust models as the increase in the dimension of variables had brought about a convergent reduction in the mean misclassification error. Therefore, other for the robust LDAs includes reduction in mean misclassification error as the shift in location of the population (μ) increased, a sharp reduction in mean misclassification error as the shift in the shape of the population increased, and a higher performance percentage observed from the $\mathrm{RLDA_{WMQ}}$ as p increased.

A series of conclusion can be drawn from the results documented in Tables 4.1- 4.12. The first observation is that when all assumptions of LDA are satisfied, the classical linear discriminant model performs best. However, in the presence of contamination, the robust models have better accuracy. Considering the general performance percentages, it is worth taking note of that the $\mathrm{RLDA_{WMQ}}$ model handles datasets with unbalanced sample sizes better than $\mathrm{RLDA_{MQ}}$ irrespective of the property of the covariance matrices. An insight has been displayed on which linear discriminant model performs better in terms of mean misclassification error at certain variable specifications, it is also important to investigate which approach requires more computation time.

4.3 Computational Time Analysis with Simulation Study

This section displays the computational time involved when adopting the linear discriminant models. Variable dimensions 2, 6 and 10 are considered for balanced and unbalanced sample sizes which are similar to the analysis for the misclassification error with equal and unequal covariance matrices. A testing sample of size 2000 from each population was also generated and the computational time was computed after repeating the process 2000 times.

4.3.1 Equal Covariance Matrices with Balanced Sample Sizes

The first set of results discussed is the computational time of adopting the models for balanced sample sizes; $(n_1, n_2) = [(20,20),(50,50),(100,100)]$ with equal covariance matrices. The results are considered for variable dimension $p = 2,6,10$ and discussion of the results will be made after the tables.

Table 4.13

Computational Time (in seconds) for Linear Discriminant Models with Balanced Sample Sizes, Equal Covariance Matrices and $p = 2$

			$(n_1, n_2) = (20, 20)$			$(n_1, n_2) = (50, 50)$			$(n_1, n_2) = (100, 100)$		
ε	μ	κ	**CA**	**RLDA$_{MQ}$**	**RLDA$_{WMQ}$**	**CA**	**RLDA$_{MQ}$**	**RLDA$_{WMQ}$**	**CA**	**RLDA$_{MQ}$**	**RLDA$_{WMQ}$**
-	-	-	2	8	4	2	25	13	2	223	121
10	**3**	-	2	7	4	2	25	13	3	230	124
20	**3**	-	2	7	4	2	25	14	2	232	124
40	**3**	-	2	8	4	2	25	13	2	230	120
10	**5**	-	2	8	4	2	25	14	3	228	125
20	**5**	-	2	8	4	2	26	14	2	226	123
40	**5**	-	2	8	4	2	24	14	2	225	124
10	**0**	**9**	2	7	4	2	25	13	2	224	124
20	**0**	**9**	2	8	4	2	25	13	2	220	122
40	**0**	**9**	2	8	4	2	25	13	2	221	121
10	**0**	**25**	2	7	4	2	25	14	2	225	123
20	**0**	**25**	2	8	4	2	25	13	2	222	121
40	**0**	**25**	2	8	4	2	26	13	2	221	120
10	**0**	**100**	2	8	4	2	26	13	2	227	120
20	**0**	**100**	2	8	4	2	26	13	2	223	121
40	**0**	**100**	2	8	4	2	25	13	2	226	118
10	**3**	**9**	2	7	4	2	25	13	2	231	125
20	**3**	**9**	2	7	4	2	26	14	2	228	125
40	**3**	**9**	2	7	4	2	25	13	2	229	123
10	**3**	**25**	2	7	4	2	26	13	2	226	126
20	**3**	**25**	2	7	4	2	26	13	2	224	120
40	**3**	**25**	2	7	4	2	26	13	2	235	123
10	**3**	**100**	2	7	4	2	25	13	2	229	127
20	**3**	**100**	2	8	4	2	25	13	2	231	122
40	**3**	**100**	2	8	4	2	25	13	2	231	125
10	**5**	**9**	2	7	4	2	26	14	3	229	126
20	**5**	**9**	2	8	4	2	25	13	2	232	125
40	**5**	**9**	2	7	4	2	25	13	2	229	125
10	**5**	**25**	2	8	4	2	25	14	3	229	126
20	**5**	**25**	2	8	4	2	25	14	2	229	124
40	**5**	**25**	2	8	4	2	25	13	2	229	127
10	**5**	**100**	2	7	4	2	25	13	2	230	127
20	**5**	**100**	2	7	4	2	25	14	2	227	124
40	**5**	**100**	2	7	4	2	25	13	2	221	125

Table 4.14

Computational Time (in seconds) for Linear Discriminant Models with Balanced Sample Sizes, Equal Covariance Matrices and $p = 6$

			$(n_1,n_2)=(20,20)$			$(n_1,n_2)=(50,50)$			$(n_1,n_2)=(100,100)$		
ε	μ	κ	**CA**	**RLDA$_{MQ}$**	**RLDA$_{WMQ}$**	**CA**	**RLDA$_{MQ}$**	**RLDA$_{WMQ}$**	**CA**	**RLDA$_{MQ}$**	**RLDA$_{WMQ}$**
-	-	-	5	25	11	5	74	38	6	693	349
10	**3**	-	5	29	11	5	78	38	5	651	362
20	**3**	-	5	29	10	5	78	38	6	683	373
40	**3**	-	5	28	10	5	77	38	5	704	368
10	**5**	-	5	29	11	6	79	39	6	702	368
20	**5**	-	5	29	11	5	79	39	5	702	377
40	**5**	-	5	28	10	5	79	40	5	704	366
10	**0**	**9**	5	26	10	6	78	40	6	704	351
20	**0**	**9**	5	28	11	5	77	40	6	705	353
40	**0**	**9**	5	27	10	5	77	41	5	700	354
10	**0**	**25**	5	27	11	6	74	39	6	700	354
20	**0**	**25**	5	28	11	6	74	40	5	708	353
40	**0**	**25**	5	27	11	5	74	41	5	709	352
10	**0**	**100**	5	27	11	5	76	40	5	705	355
20	**0**	**100**	5	27	11	5	76	40	5	708	355
40	**0**	**100**	5	27	11	5	77	40	5	706	351
10	**3**	**9**	5	28	11	5	79	39	6	703	358
20	**3**	**9**	5	27	11	5	78	39	5	700	347
40	**3**	**9**	5	27	10	5	77	39	5	701	341
10	**3**	**25**	5	26	11	6	72	39	6	694	341
20	**3**	**25**	5	24	11	5	79	39	6	705	349
40	**3**	**25**	5	25	11	5	73	39	5	703	342
10	**3**	**100**	5	26	11	5	80	38	5	701	333
20	**3**	**100**	5	27	11	5	79	39	5	709	335
40	**3**	**100**	5	27	11	5	79	38	5	704	345
10	**5**	**9**	5	29	11	6	78	38	5	708	391
20	**5**	**9**	5	29	11	5	80	38	6	707	396
40	**5**	**9**	5	29	10	5	78	37	5	704	376
10	**5**	**25**	5	28	11	5	78	38	6	706	366
20	**5**	**25**	5	29	11	5	79	37	6	701	368
40	**5**	**25**	5	28	10	5	79	38	6	705	367
10	**5**	**100**	5	28	11	5	77	39	5	702	361
20	**5**	**100**	5	28	11	5	77	38	5	704	370
40	**5**	**100**	5	29	10	5	79	38	5	701	378

Table 4.15

Computational Time (in seconds) for Linear Discriminant Models with Balanced Sample Sizes, Equal Covariance Matrices and $p = 10$

			$(n_1,n_2)=(20,20)$			$(n_1,n_2)=(50,50)$			$(n_1,n_2)=(100,100)$		
ε	μ	κ	**CA**	**RLDA$_{MQ}$**	**RLDA$_{WMQ}$**	**CA**	**RLDA$_{MQ}$**	**RLDA$_{WMQ}$**	**CA**	**RLDA$_{MQ}$**	**RLDA$_{WMQ}$**
-	-	-	8	46	17	9	133	63	9	1165	579
10	3	-	10	45	17	10	132	64	11	1041	596
20	3	-	10	46	18	11	133	65	10	1054	605
40	3	-	8	47	17	8	136	64	8	1039	599
10	5	-	10	45	17	10	133	65	11	1086	597
20	5	-	10	44	17	10	134	66	11	1153	597
40	5	-	9	45	17	8	132	66	8	1074	594
10	0	9	9	45	17	9	137	67	9	1107	602
20	0	9	9	45	17	9	136	66	9	1166	602
40	0	9	8	45	17	8	138	67	8	1194	600
10	0	25	11	45	17	11	137	66	11	1176	597
20	0	25	10	44	18	11	136	66	11	1179	608
40	0	25	8	46	17	9	135	66	8	1180	598
10	0	100	8	45	17	8	133	65	8	1063	597
20	0	100	8	44	17	8	134	64	8	1044	604
40	0	100	8	45	17	8	134	63	8	1075	596
10	3	9	10	46	17	10	135	65	11	1217	592
20	3	9	10	48	17	10	134	67	11	1174	592
40	3	9	8	47	17	8	134	65	8	1231	600
10	3	25	10	45	17	10	132	66	11	1152	600
20	3	25	10	46	17	10	135	65	11	1181	591
40	3	25	8	45	17	8	134	67	8	1178	595
10	3	100	8	45	17	8	132	66	8	1184	610
20	3	100	8	45	17	9	135	66	8	1165	568
40	3	100	8	45	18	8	136	67	8	1099	578
10	5	9	10	46	17	10	133	65	10	1186	596
20	5	9	10	45	17	10	132	65	11	1129	599
40	5	9	8	46	18	8	134	66	8	1119	600
10	5	25	10	45	17	10	133	65	11	1121	595
20	5	25	10	45	18	10	132	65	10	1129	595
40	5	25	8	45	18	8	132	65	9	1135	599
10	5	100	8	45	17	8	133	66	8	1038	597
20	5	100	8	45	17	8	134	67	8	1152	595
40	5	100	8	46	17	8	135	66	8	1173	600

From Tables 4.13 to 4.15, the model with the least computational time is the classical LDA followed by the RLDA$_{WMQ}$ and the highest being the RLDA$_{MQ}$ robust model. The RLDA$_{MQ}$ model requires the highest computational time in comparison to the other linear discriminant models, although the model gives a good level of accuracy when considering datasets with balanced sample sizes. An average of the computational time values is given in the following table.

Table 4.16

Average Computational Time (in seconds) for Linear Discriminant Models with Balanced Sample Sizes, Equal Covariance Matrices

	$(20,20)$			$(50,50)$			$(100,100)$		
	CA	**$\mathbf{RLDA_{MQ}}$**	**$\mathbf{RLDA_{WMQ}}$**	**CA**	**$\mathbf{RLDA_{MQ}}$**	**$\mathbf{RLDA_{WMQ}}$**	**CA**	**RLDAMQ**	**$\mathbf{RLDA_{WMQ}}$**
$p=2$	2	8	4	2	25	13	2	227	123
$p=6$	5	27	11	5	77	39	5	701	359
$p=10$	9	45	17	9	134	65	9	1134	596

Figures 4.1-4.3 give a better pictorial view of this increase in the average computational time as seen in Table 4.16 with simultaneous increase in variable dimensions.

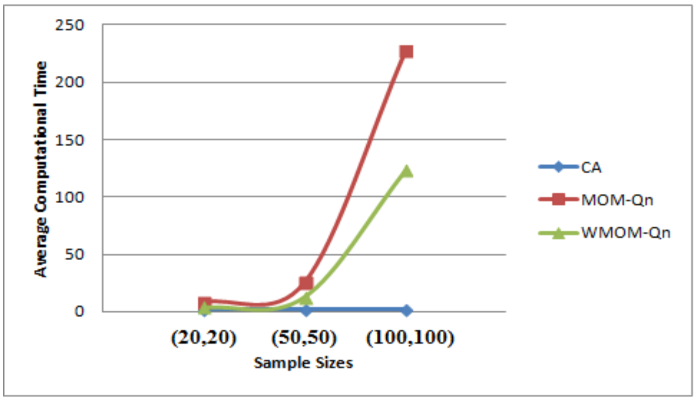

Figure 4.1 Average Computational Time (in seconds) for Linear Discriminant Models with Balanced Sample Sizes, Equal Covariance Matrices and $p = 2$

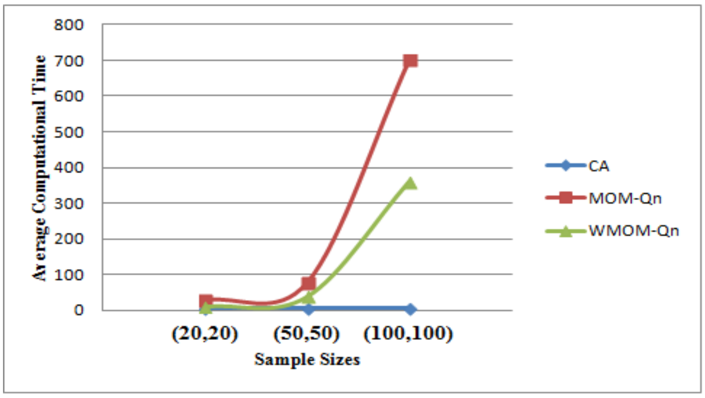

Figure 4.2 Average Computational Time (in seconds) for Linear Discriminant Models with Balanced Sample Sizes, Equal Covariance Matrices and $p = 6$

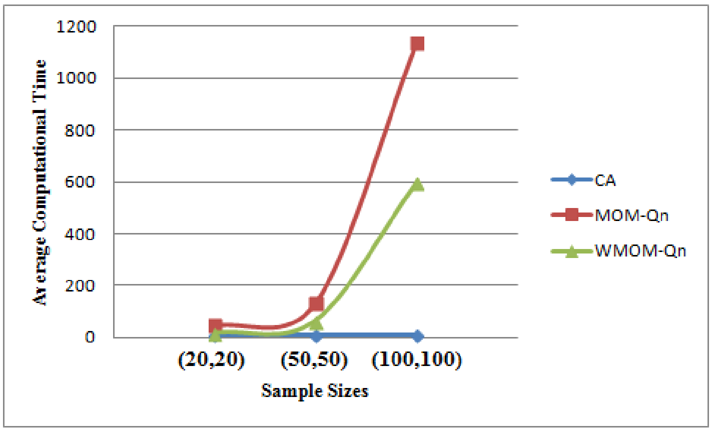

Figure 4.3 Average Computational Time (in seconds) for Linear Discriminant Models with Balanced Sample Sizes, Equal Covariance Matrices and $p = 10$

4.3.2 Equal Covariance Matrices with Unbalanced Sample Sizes

The next set of results discussed is the computational time of adopting the models for unbalanced sample sizes; $(n_1, n_2) = [(50,20),(100,50),(100,20)]$ still with the same equal covariance matrix property. The results are also considered for variable dimension $p = 2,6,10$ and discussion of the results are made after the tables.

Table 4.17

Computational Time (in seconds) for Linear Discriminant Models with Unbalanced Sample Sizes, Equal Covariance Matrices and $p = 2$

			$(n_1,n_2)=(50,20)$			$(n_1,n_2)=(100,50)$			$(n_1,n_2)=(100,20)$		
ε	μ	κ	**CA**	**RLDA$_{MQ}$**	**RLDA$_{WMQ}$**	**CA**	**RLDA$_{MQ}$**	**RLDA$_{WMQ}$**	**CA**	**RLDA$_{MQ}$**	**RLDA$_{WMQ}$**
-	-	-	2	17	7	2	126	61	2	118	62
10	**3**	-	2	17	8	2	126	62	2	128	60
20	**3**	-	2	18	9	2	126	62	2	121	61
40	**3**	-	2	19	8	2	127	61	2	117	63
10	**5**	-	2	17	9	2	127	70	2	126	61
20	**5**	-	2	17	9	2	128	71	2	128	63
40	**5**	-	2	17	9	2	129	63	2	128	59
10	**0**	**9**	2	17	7	2	132	63	2	119	63
20	**0**	**9**	2	17	7	2	129	64	2	125	60
40	**0**	**9**	2	17	7	2	129	68	2	126	62
10	**0**	**25**	2	17	9	2	134	63	2	122	63
20	**0**	**25**	2	17	9	2	132	71	2	112	63
40	**0**	**25**	2	18	8	2	123	69	2	128	63
10	**0**	**100**	2	16	7	2	123	59	2	120	64
20	**0**	**100**	2	17	7	2	128	66	2	124	64
40	**0**	**100**	2	17	8	2	128	62	2	126	61
10	**3**	**9**	2	18	8	2	123	66	2	121	64
20	**3**	**9**	2	19	8	2	125	69	2	126	58
40	**3**	**9**	2	18	9	2	121	60	2	126	61
10	**3**	**25**	2	18	8	2	119	66	2	128	61
20	**3**	**25**	2	19	8	2	126	64	2	129	62
40	**3**	**25**	2	19	8	2	126	70	2	122	65
10	**3**	**100**	2	19	9	2	122	61	2	129	63
20	**3**	**100**	2	19	9	2	122	64	2	126	63
40	**3**	**100**	2	18	9	2	124	62	2	129	61
10	**5**	**9**	2	18	9	2	127	64	2	126	58
20	**5**	**9**	2	19	9	2	128	69	2	131	57
40	**5**	**9**	2	19	9	2	124	70	2	129	58
10	**5**	**25**	2	19	9	2	125	68	2	130	60
20	**5**	**25**	2	17	9	2	124	68	2	125	64
40	**5**	**25**	2	19	9	2	123	64	2	116	60
10	**5**	**100**	2	18	8	2	125	68	2	128	63
20	**5**	**100**	2	19	8	2	126	69	2	127	63
40	**5**	**100**	2	18	8	2	126	64	2	129	64

Table 4.18

Computational Time (in seconds) for Linear Discriminant Models with Unbalanced Sample Sizes, Equal Covariance Matrices and $p = 6$

			$(n_1,n_2)=(50,20)$			$(n_1,n_2)=(100,50)$			$(n_1,n_2)=(100,20)$		
ε	μ	κ	**CA**	**$RLDA_{MQ}$**	**$RLDA_{WMQ}$**	**CA**	**$RLDA_{MQ}$**	**$RLDA_{WMQ}$**	**CA**	**$RLDA_{MQ}$**	**$RLDA_{WMQ}$**
-	**-**	**-**	5	55	23	5	402	197	5	366	199
10	**3**	**-**	5	52	24	5	401	208	5	367	194
20	**3**	**-**	5	53	25	5	395	209	6	365	198
40	**3**	**-**	5	51	23	5	394	206	5	366	193
10	**5**	**-**	5	57	26	5	397	210	5	382	194
20	**5**	**-**	5	54	26	5	396	193	5	382	195
40	**5**	**-**	5	52	24	5	399	211	5	370	192
10	**0**	**9**	5	54	23	5	395	213	5	385	197
20	**0**	**9**	5	56	25	5	400	195	5	381	193
40	**0**	**9**	5	56	24	5	404	196	5	381	194
10	**0**	**25**	5	56	26	5	410	213	5	386	194
20	**0**	**25**	5	56	23	5	410	214	5	387	192
40	**0**	**25**	5	57	26	5	416	195	5	373	199
10	**0**	**100**	5	49	26	5	401	195	5	380	198
20	**0**	**100**	5	49	24	5	407	216	5	382	199
40	**0**	**100**	5	53	26	5	413	211	5	374	197
10	**3**	**9**	5	58	24	5	403	213	5	366	196
20	**3**	**9**	5	57	24	5	408	214	5	367	198
40	**3**	**9**	5	56	25	5	410	196	5	379	191
10	**3**	**25**	5	56	22	5	416	210	5	373	199
20	**3**	**25**	5	57	24	5	419	214	5	378	195
40	**3**	**25**	5	56	25	5	404	200	5	377	194
10	**3**	**100**	5	57	24	5	396	192	5	378	200
20	**3**	**100**	5	57	26	5	400	213	5	381	194
40	**3**	**100**	5	57	25	5	410	216	5	368	194
10	**5**	**9**	5	56	26	5	414	195	5	369	183
20	**5**	**9**	5	56	27	5	418	194	5	369	192
40	**5**	**9**	5	58	24	5	400	191	5	362	194
10	**5**	**25**	5	57	26	5	399	204	5	368	199
20	**5**	**25**	5	57	26	5	411	206	5	377	200
40	**5**	**25**	5	58	24	5	415	212	5	377	189
10	**5**	**100**	5	56	24	5	411	208	5	372	194
20	**5**	**100**	5	55	23	5	403	218	5	350	199
40	**5**	**100**	5	57	26	5	407	216	5	354	195

Table 4.19

Computational Time (in seconds) for Linear Discriminant Models with Unbalanced Sample Sizes, Equal Covariance Matrices and p=10

			$(n_1,n_2)=(50,20)$			$(n_1,n_2)=(100,50)$			$(n_1,n_2)=(100,20)$		
ε	μ	κ	**CA**	**$RLDA_{MQ}$**	**$RLDA_{WMQ}$**	**CA**	**$RLDA_{MQ}$**	**$RLDA_{WMQ}$**	**CA**	**$RLDA_{MQ}$**	**$RLDA_{WMQ}$**
-	**-**	**-**	8	87	39	8	556	339	8	618	317
10	**3**	**-**	8	91	41	8	552	327	8	642	308
20	**3**	**-**	8	85	41	8	556	333	8	621	316
40	**3**	**-**	8	90	41	8	554	334	8	628	290
10	**5**	**-**	8	101	40	8	669	325	8	662	287
20	**5**	**-**	8	102	39	8	675	337	8	658	319
40	**5**	**-**	8	101	41	8	670	331	8	656	286
10	**0**	**9**	8	98	40	8	691	336	8	664	322
20	**0**	**9**	8	99	44	8	727	342	8	651	296
40	**0**	**9**	8	90	45	8	707	335	8	589	336
10	**0**	**25**	8	99	45	8	714	328	8	643	331
20	**0**	**25**	8	103	45	8	647	333	8	650	343
40	**0**	**25**	8	102	47	8	655	333	8	607	304
10	**0**	**100**	8	100	40	8	689	332	8	627	340
20	**0**	**100**	8	90	41	8	696	328	8	648	299
40	**0**	**100**	8	99	41	8	695	336	8	696	336
10	**3**	**9**	8	87	44	8	566	325	8	617	319
20	**3**	**9**	8	84	46	8	553	336	8	650	325
40	**3**	**9**	8	89	46	8	688	324	8	666	334
10	**3**	**25**	8	98	42	8	667	331	8	609	336
20	**3**	**25**	8	82	35	8	694	336	8	606	327
40	**3**	**25**	8	89	44	8	657	339	8	607	335
10	**3**	**100**	8	89	46	8	670	340	8	643	314
20	**3**	**100**	8	93	48	8	662	305	8	669	327
40	**3**	**100**	8	90	47	8	676	336	8	679	325
10	**5**	**9**	8	99	43	8	676	310	8	662	290
20	**5**	**9**	8	102	45	8	644	318	8	635	318
40	**5**	**9**	8	91	38	8	677	330	8	655	316
10	**5**	**25**	8	97	40	8	680	340	8	651	333
20	**5**	**25**	8	99	39	8	684	334	8	663	329
40	**5**	**25**	8	101	40	8	678	327	8	656	341
10	**5**	**100**	8	82	47	8	653	313	8	599	320
20	**5**	**100**	8	97	46	8	647	340	8	654	285
40	**5**	**100**	8	96	47	8	632	338	8	667	315

The model with the least computational time as seen in Tables 4.17- 4.19 still remains CA followed by the $RLDA_{WMQ}$ and then the highest being the $RLDA_{MQ}$ robust model. An average of the computational time values is given in the following Table 4.20 with the corresponding graphs in Figures 4.4-4.6.

Table 4.20

Average Computational Time (in seconds) for Linear Discriminant Models with Unbalanced Sample Sizes, Equal Covariance Matrices

	$(50,20)$			$(100,50)$			$(100,20)$		
	CA	**RLDA_{MQ}**	**RLDA_{WMQ}**	**CA**	**RLDA_{MQ}**	**RLDA_{WMQ}**	**CA**	**RLDA_{MQ}**	**RLDA_{WMQ}**
$p=2$	2	18	8	2	126	65	2	125	62
$p=6$	5	55	25	5	405	206	5	373	195
$p=10$	8	94	43	8	655	331	8	643	318

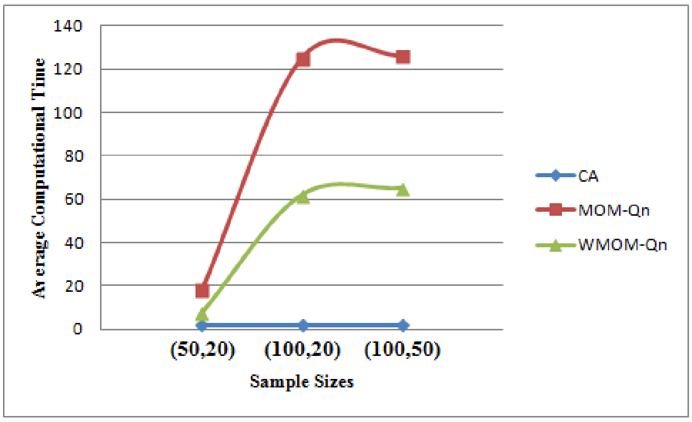

*Figure 4.4*Average Computational Time (in seconds) for Linear Discriminant Models with Unbalanced Sample Sizes, Equal Covariance Matrices and p=2

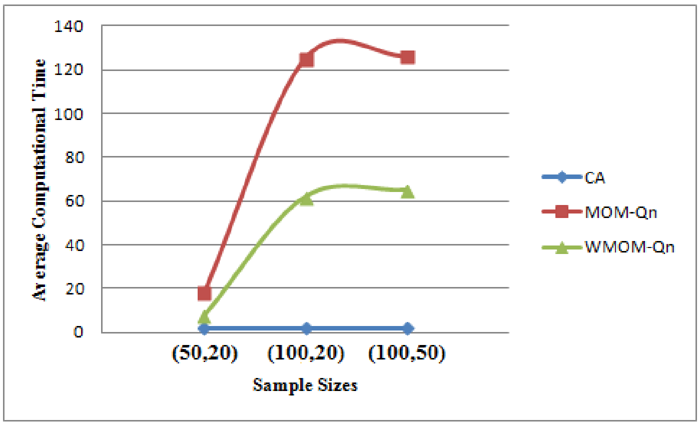

*Figure 4.5*Average Computational Time (in seconds) for Linear Discriminant Models with Unbalanced Sample Sizes, Equal Covariance Matrices and p=6

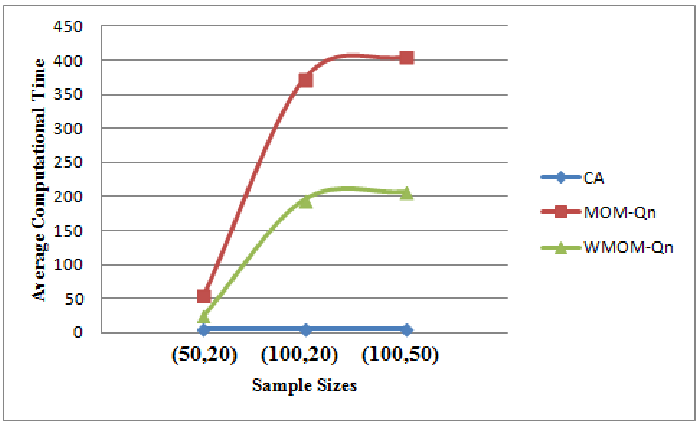

Figure 4.6 Average Computational Time (in seconds) for Linear Discriminant Models with Unbalanced Sample Sizes, Equal Covariance Matrices and p=10

Figures 4.4-4.6 do not show a growing spike in the average computational time as seen in Figures 4.1-4.3. This is as a result of the unbalanced sample sizes where n_2 is not greater than 50. Therefore, a close computation time is expected between the analysis for $(n_1, n_2) = (100, 20)$ and $(n_1, n_2) = (100, 50)$.

However, a general trend is still observed irrespective of the balanced or unbalanced sample sizes. This pattern follows that RLDA_{MQ} has the highest computational time followed by the RLDA_{WMQ} and the least is CA. This leads to the consideration of the computational time for the linear models when the covariance matrices are unequal.

4.3.3 Unequal Covariance Matrices with Balanced Sample Sizes

Now, we move over to considering the unequal covariance matrices for balanced sample sizes; $(n_1, n_2) = [(20, 20), (50, 50), (100, 100)]$. The computational time for each linear model per variable dimension $p = 2, 6, 10$ are presented.

Table 4.21

Computational Time (in seconds) for Linear Discriminant Models with Balanced Sample Sizes, Unequal Covariance Matrices and p=2

			$(n_1,n_2)=(20,20)$			$(n_1,n_2)=(50,50)$			$(n_1,n_2)=(100,100)$		
ε	μ	κ	**CA**	**RLDA$_{MQ}$**	**RLDA$_{WMQ}$**	**CA**	**RLDA$_{MQ}$**	**RLDA$_{WMQ}$**	**CA**	**RLDA$_{MQ}$**	**RLDA$_{WMQ}$**
-	-	-	2	8	4	2	25	13	2	219	113
10	**3**	-	2	8	4	2	26	12	2	221	121
20	**3**	-	2	8	4	2	26	13	2	222	116
40	**3**	-	2	7	5	2	25	13	2	221	118
10	**5**	-	2	8	4	2	26	12	2	231	118
20	**5**	-	2	8	4	2	26	13	2	228	119
40	**5**	-	2	7	4	2	25	13	2	224	110
10	**0**	**9**	2	8	4	2	26	14	2	219	119
20	**0**	**9**	2	8	4	2	26	14	2	223	118
40	**0**	**9**	2	8	4	2	25	13	2	211	108
10	**0**	**25**	2	7	4	2	24	14	2	228	118
20	**0**	**25**	2	8	4	2	24	13	2	211	119
40	**0**	**25**	2	7	4	2	25	13	2	221	120
10	**0**	**100**	2	7	4	2	24	13	2	223	115
20	**0**	**100**	2	7	4	2	24	13	2	234	119
40	**0**	**100**	2	7	4	2	25	13	2	220	118
10	**3**	**9**	2	8	4	2	25	13	2	221	118
20	**3**	**9**	2	8	4	2	26	13	2	238	122
40	**3**	**9**	2	7	4	2	23	14	2	237	119
10	**3**	**25**	2	8	4	2	24	14	2	242	121
20	**3**	**25**	2	7	4	2	25	13	2	241	117
40	**3**	**25**	2	7	4	2	25	13	2	224	118
10	**3**	**100**	2	8	4	2	25	13	2	224	119
20	**3**	**100**	2	8	4	2	25	14	2	222	118
40	**3**	**100**	2	7	4	2	23	14	2	221	120
10	**5**	**9**	2	7	4	2	25	12	2	234	114
20	**5**	**9**	2	8	4	2	26	13	2	234	117
40	**5**	**9**	2	7	4	2	26	14	2	232	119
10	**5**	**25**	2	7	4	2	25	12	2	223	117
20	**5**	**25**	2	8	4	2	25	14	2	234	118
40	**5**	**25**	2	8	4	2	25	13	2	234	118
10	**5**	**100**	2	8	4	2	24	14	2	232	115
20	**5**	**100**	2	8	4	2	25	13	2	247	113
40	**5**	**100**	2	7	4	2	26	14	2	242	116

Table 4.22

Computational Time (in seconds) for Linear Discriminant Models with Balanced Sample Sizes, Unequal Covariance Matrices and p=6

			$(n_1,n_2)=(20,20)$			$(n_1,n_2)=(50,50)$			$(n_1,n_2)=(100,100)$		
ε	μ	κ	**CA**	**$RLDA_{MQ}$**	**$RLDA_{WMQ}$**	**CA**	**$RLDA_{MQ}$**	**$RLDA_{WMQ}$**	**CA**	**$RLDA_{MQ}$**	**$RLDA_{WMQ}$**
-	-	-	5	23	10	5	75	36	5	800	408
10	**3**	-	5	23	10	5	78	37	5	800	409
20	**3**	-	5	23	9	5	79	38	5	794	405
40	**3**	-	5	23	9	5	76	37	5	758	401
10	**5**	-	5	23	11	5	77	37	5	729	394
20	**5**	-	5	24	11	5	70	37	5	723	397
40	**5**	-	5	24	10	5	78	37	5	709	401
10	**0**	**9**	5	24	10	5	76	37	5	788	338
20	**0**	**9**	5	23	10	5	75	37	5	788	351
40	**0**	**9**	5	23	10	5	76	37	5	766	353
10	**0**	**25**	5	23	11	5	75	36	5	715	352
20	**0**	**25**	5	24	10	5	75	35	5	766	351
40	**0**	**25**	5	23	10	5	75	36	5	760	354
10	**0**	**100**	5	23	11	5	75	35	5	770	357
20	**0**	**100**	5	24	10	5	75	35	5	806	354
40	**0**	**100**	5	23	10	5	75	35	5	811	355
10	**3**	**9**	5	24	11	5	75	36	5	705	368
20	**3**	**9**	5	24	10	5	75	35	5	711	369
40	**3**	**9**	5	24	10	5	76	37	5	710	357
10	**3**	**25**	5	23	10	5	76	37	5	709	357
20	**3**	**25**	5	25	11	5	75	37	5	698	358
40	**3**	**25**	5	23	10	5	76	34	5	715	357
10	**3**	**100**	5	22	10	5	76	37	5	706	384
20	**3**	**100**	5	22	11	5	77	35	5	701	385
40	**3**	**100**	5	24	10	5	73	38	5	709	359
10	**5**	**9**	5	22	11	5	75	37	5	701	394
20	**5**	**9**	5	22	10	5	74	36	5	707	396
40	**5**	**9**	5	24	11	5	76	34	5	702	394
10	**5**	**25**	5	23	10	5	76	34	5	697	393
20	**5**	**25**	5	22	10	5	75	37	5	687	389
40	**5**	**25**	5	22	10	5	76	37	5	691	391
10	**5**	**100**	5	23	10	5	75	34	5	694	388
20	**5**	**100**	5	23	11	5	77	38	5	705	391
40	**5**	**100**	5	23	10	5	75	35	5	704	388

Table 4.23

Computational Time (in seconds) for Linear Discriminant Models with Balanced Sample Sizes, Unequal Covariance Matrices and p=10

			$(n_1,n_2)=(20,20)$			$(n_1,n_2)=(50,50)$			$(n_1,n_2)=(100,100)$		
ε	μ	κ	**CA**	**$RLDA_{MQ}$**	**$RLDA_{WMQ}$**	**CA**	**$RLDA_{MQ}$**	**$RLDA_{WMQ}$**	**CA**	**$RLDA_{MQ}$**	**$RLDA_{WMQ}$**
-	-	-	8	46	17	8	142	68	8	1315	638
10	**3**	-	8	45	16	8	148	74	8	1297	639
20	**3**	-	8	45	17	8	144	73	8	1238	632
40	**3**	-	8	45	16	8	134	75	8	1335	628
10	**5**	-	8	46	16	8	142	73	8	1270	595
20	**5**	-	8	47	17	8	144	72	8	1315	629
40	**5**	-	8	45	17	8	146	75	8	1288	636
10	**0**	**9**	8	43	17	8	143	73	8	1284	633
20	**0**	**9**	8	44	17	8	141	71	8	1272	584
40	**0**	**9**	8	46	17	8	132	74	8	1280	624
10	**0**	**25**	8	45	17	8	132	72	8	1259	633
20	**0**	**25**	8	44	17	8	138	74	8	1104	662
40	**0**	**25**	8	45	17	8	137	72	8	1279	660
10	**0**	**100**	8	45	17	8	159	70	8	1314	645
20	**0**	**100**	8	44	17	8	160	74	8	1222	642
40	**0**	**100**	8	45	17	8	164	74	8	1229	645
10	**3**	**9**	8	45	17	8	148	71	8	1283	592
20	**3**	**9**	8	44	17	8	156	72	8	1275	630
40	**3**	**9**	8	44	17	8	156	74	8	1291	633
10	**3**	**25**	8	44	16	8	137	75	8	1287	625
20	**3**	**25**	8	43	17	8	138	74	9	1318	611
40	**3**	**25**	8	45	17	8	147	76	8	1206	623
10	**3**	**100**	8	44	16	8	149	71	8	1330	624
20	**3**	**100**	8	46	17	8	148	72	8	1362	625
40	**3**	**100**	8	46	17	8	144	72	8	1321	631
10	**5**	**9**	8	46	17	8	147	72	8	1227	657
20	**5**	**9**	8	46	17	8	147	72	8	1207	627
40	**5**	**9**	8	45	16	8	151	73	8	1202	642
10	**5**	**25**	8	45	17	8	150	71	8	1260	652
20	**5**	**25**	8	45	17	8	150	73	8	1242	659
40	**5**	**25**	8	45	17	8	150	73	8	1229	660
10	**5**	**100**	8	45	17	8	146	73	8	1111	638
20	**5**	**100**	8	45	17	8	146	72	8	1219	603
40	**5**	**100**	8	45	17	8	145	71	8	1314	624

In accordance with the results obtained from the equal covariance matrices property, Tables 4.21-4.23 shows the model with the least computational time still remains CA and the model with the highest computational time is the same $RLDA_{MQ}$ robust model. An average of the computational time values is also given for the analysis for unequal covariance for balanced sample sizes in the following table. Note that the graphs were not plotted as there is convergence in the average computational time values with Table 4.16, hence same shape of graph is expected.

Table 4.24

Average Computational Time (in seconds) for Linear Discriminant Models with Balanced Sample Sizes, Unequal Covariance Matrices

	$(20,20)$			$(50,50)$			$(100,100)$		
	CA	**$RLDA_{MQ}$**	**$RLDA_{WMQ}$**	**CA**	**$RLDA_{MQ}$**	**$RLDA_{WMQ}$**	**CA**	**$RLDA_{MQ}$**	**$RLDA_{WMQ}$**
$p=2$	2	8	4	2	25	13	2	228	117
$p=6$	5	23	10	5	75	36	5	733	376
$p=10$	8	45	17	8	146	73	8	1264	632

4.3.4 Unequal Covariance Matrices with Unbalanced Sample Sizes

The next set of results discussed are the computational time of adopting the models for unequal covariance matrix property having unbalanced sample sizes as defined previously $(n_1, n_2) = [(50,20),(100,50),(100,20)]$. The results are also considered for $p = 2,6,10$.

Table 4.25

Computational Time (in seconds) for Linear Discriminant Models with Unbalanced Sample Sizes, Unequal Covariance Matrices and p=2

			$(n_1,n_2)=(50,20)$			$(n_1,n_2)=(100,50)$			$(n_1,n_2)=(100,20)$		
ε	μ	κ	**CA**	**RLDA$_{MQ}$**	**RLDA$_{WMQ}$**	**CA**	**RLDA$_{MQ}$**	**RLDA$_{WMQ}$**	**CA**	**RLDA$_{MQ}$**	**RLDA$_{WMQ}$**
-	-	-	2	16	9	2	120	63	2	121	62
10	**3**	-	2	16	8	2	126	64	2	121	60
20	**3**	-	2	16	8	2	133	66	2	119	60
40	**3**	-	2	16	8	2	131	65	2	120	58
10	**5**	-	2	16	8	2	132	59	2	121	59
20	**5**	-	2	17	8	2	130	65	2	122	56
40	**5**	-	2	17	8	2	130	66	2	119	58
10	**0**	**9**	2	17	8	2	129	68	2	121	62
20	**0**	**9**	2	16	9	2	127	67	2	121	57
40	**0**	**9**	2	15	8	2	126	72	2	121	59
10	**0**	**25**	2	16	8	2	131	65	2	122	59
20	**0**	**25**	2	17	8	2	134	68	2	121	59
40	**0**	**25**	2	16	8	2	129	69	2	117	59
10	**0**	**100**	2	17	8	2	128	66	2	122	63
20	**0**	**100**	2	17	8	2	128	69	2	122	63
40	**0**	**100**	2	17	8	2	130	67	2	118	60
10	**3**	**9**	2	17	8	2	122	66	2	128	59
20	**3**	**9**	2	16	8	2	124	66	2	125	58
40	**3**	**9**	2	16	8	2	129	66	2	111	57
10	**3**	**25**	2	17	8	2	133	69	2	124	58
20	**3**	**25**	2	17	8	2	130	67	2	125	55
40	**3**	**25**	2	16	8	2	132	68	2	124	59
10	**3**	**100**	2	16	8	2	131	65	2	123	58
20	**3**	**100**	2	16	8	2	128	67	2	112	58
40	**3**	**100**	2	16	8	2	133	67	2	117	60
10	**5**	**9**	2	17	8	2	133	67	2	121	60
20	**5**	**9**	2	16	8	2	134	66	2	122	59
40	**5**	**9**	2	16	8	2	134	66	2	120	60
10	**5**	**25**	2	17	8	2	127	66	2	121	61
20	**5**	**25**	2	16	8	2	129	64	2	120	60
40	**5**	**25**	2	17	8	2	134	66	2	122	58
10	**5**	**100**	2	17	8	2	133	65	2	121	58
20	**5**	**100**	2	17	8	2	133	66	2	121	61
40	**5**	**100**	2	16	8	2	131	67	2	121	58

Table 4.26

Computational Time (in seconds) for Linear Discriminant Models with Unbalanced Sample Sizes, Unequal Covariance Matrices and p=6

			$(n_1,n_2)=(50,20)$			$(n_1,n_2)=(100,50)$			$(n_1,n_2)=(100,20)$		
ε	μ	κ	**CA**	**$RLDA_{MQ}$**	**$RLDA_{WMQ}$**	**CA**	**$RLDA_{MQ}$**	**$RLDA_{WMQ}$**	**CA**	**$RLDA_{MQ}$**	**$RLDA_{WMQ}$**
-	-	-	5	48	24	5	386	202	5	367	213
10	**3**	-	5	51	24	5	384	203	5	358	221
20	**3**	-	5	51	25	5	398	215	5	359	216
40	**3**	-	5	49	23	5	387	222	5	345	236
10	**5**	-	5	50	23	5	404	234	5	349	211
20	**5**	-	5	48	22	5	394	232	5	332	230
40	**5**	-	5	50	24	5	392	233	5	344	221
10	**0**	**9**	5	51	22	5	393	214	5	364	226
20	**0**	**9**	5	51	25	5	391	217	5	364	220
40	**0**	**9**	5	47	21	5	395	216	5	353	206
10	**0**	**25**	5	51	24	5	396	218	5	350	217
20	**0**	**25**	5	48	24	5	392	222	5	364	224
40	**0**	**25**	5	50	22	5	393	208	5	364	234
10	**0**	**100**	5	50	23	5	382	204	5	352	231
20	**0**	**100**	5	50	23	6	398	206	5	366	231
40	**0**	**100**	5	51	23	6	398	208	5	370	230
10	**3**	**9**	5	50	23	5	400	223	5	354	221
20	**3**	**9**	5	51	23	5	396	223	5	359	221
40	**3**	**9**	5	48	24	5	393	222	5	350	222
10	**3**	**25**	5	50	23	5	393	201	5	360	224
20	**3**	**25**	5	51	23	5	393	214	5	357	228
40	**3**	**25**	5	47	24	5	400	221	5	349	224
10	**3**	**100**	5	50	24	5	393	222	5	358	218
20	**3**	**100**	5	51	23	5	390	220	5	357	222
40	**3**	**100**	5	51	25	5	395	230	5	360	221
10	**5**	**9**	5	50	24	5	402	215	5	322	220
20	**5**	**9**	5	50	23	5	394	229	5	365	229
40	**5**	**9**	5	50	24	5	388	216	5	360	219
10	**5**	**25**	5	50	23	5	382	238	5	368	230
20	**5**	**25**	5	49	23	5	402	235	5	362	227
40	**5**	**25**	5	50	24	5	413	233	5	354	229
10	**5**	**100**	5	50	23	5	391	212	5	346	210
20	**5**	**100**	5	51	24	5	403	213	5	361	227
40	**5**	**100**	5	51	24	5	396	201	5	353	221

Table 4.27

Computational Time (in seconds) for Linear Discriminant Models with Unbalanced Sample Sizes, Unequal Covariance Matrices and p=10

			$(n_1,n_2)=(50,20)$			$(n_1,n_2)=(100,50)$			$(n_1,n_2)=(100,20)$		
ε	μ	κ	**CA**	**RLDA$_{MQ}$**	**RLDA$_{WMQ}$**	**CA**	**RLDA$_{MQ}$**	**RLDA$_{WMQ}$**	**CA**	**RLDA$_{MQ}$**	**RLDA$_{WMQ}$**
-	-	-	8	87	39	8	653	387	8	658	335
10	**3**	-	8	89	40	8	678	353	8	693	319
20	**3**	-	8	86	39	8	652	350	8	719	329
40	**3**	-	8	85	39	8	701	362	8	648	334
10	**5**	-	8	89	39	8	670	350	8	703	341
20	**5**	-	8	91	39	8	670	352	8	673	320
40	**5**	-	8	90	39	8	645	362	8	720	338
10	**0**	**9**	8	85	39	8	732	398	8	711	330
20	**0**	**9**	8	85	38	8	717	370	8	653	329
40	**0**	**9**	8	86	37	8	731	362	8	658	333
10	**0**	**25**	8	84	39	8	731	374	8	704	328
20	**0**	**25**	8	85	38	8	693	343	8	709	337
40	**0**	**25**	8	87	40	8	683	348	8	705	334
10	**0**	**100**	8	90	40	8	746	367	8	676	319
20	**0**	**100**	8	90	37	8	731	382	8	672	316
40	**0**	**100**	8	88	40	8	672	364	8	660	333
10	**3**	**9**	8	85	39	8	719	345	8	706	322
20	**3**	**9**	8	85	39	8	730	348	8	711	329
40	**3**	**9**	8	86	40	8	721	331	8	682	334
10	**3**	**25**	8	89	39	8	678	357	8	743	329
20	**3**	**25**	8	84	40	8	734	325	8	705	326
40	**3**	**25**	8	87	37	8	707	347	8	747	327
10	**3**	**100**	8	85	37	8	749	351	8	745	334
20	**3**	**100**	8	87	37	8	756	363	8	668	332
40	**3**	**100**	8	87	39	8	748	346	8	622	331
10	**5**	**9**	8	85	39	8	730	354	8	654	337
20	**5**	**9**	8	85	39	8	708	339	8	761	324
40	**5**	**9**	8	85	40	8	693	350	8	744	332
10	**5**	**25**	8	85	39	8	630	346	8	678	325
20	**5**	**25**	8	84	39	8	672	345	8	667	341
40	**5**	**25**	8	86	39	8	637	335	8	766	332
10	**5**	**100**	8	85	40	8	749	337	8	712	337
20	**5**	**100**	8	87	40	8	744	336	8	685	336
40	**5**	**100**	8	89	39	8	743	348	8	651	339

With reference to the tables displaying the computational time required to implement the models, a clear conclusion can be drawn to the computational rigor involved in using the RLDA$_{MQ}$ model. Although this model have its advantages when considering its misclassification error rate for balanced sample sizes. An average computational time calculation is made in the Table 4.28 for unequal sample sizes with unequal covariance matrices property. Also, just like the case unequal covariance for balanced. Sample sizes, Table 4.20 and Table 4.24 are in convergence. Therefore, the plots from Table 4.28 is expected to follow the shape of Figures 4.4-4.6.

Table 4.28

Average Computational Time (in seconds) for Linear Discriminant Models with Unbalanced Sample Sizes, Unequal Covariance Matrices

	$(50,20)$			$(100,50)$			$(100,20)$		
	CA	**RLDA$_{MQ}$**	**RLDA$_{WMQ}$**	**CA**	**RLDA$_{MQ}$**	**RLDA$_{WMQ}$**	**CA**	**RLDA$_{MQ}$**	**RLDA$_{WMQ}$**
$p=2$	2	16	8	2	130	66	2	121	59
$p=6$	5	50	23	5	394	218	5	356	223
$p=10$	8	86	39	8	705	354	8	694	331

4.4 Misclassification Error Analysis with Real Data

In addition to the simulation study, CA, RLDA$_{MQ}$ and RLDA$_{WMQ}$ were also tested on real data. The data considered was obtained from the financial sector in Malaysia and the analysis was based on classifying distressed and non-distressed banking institutions. The financial data were extracted from selected balance sheet in the annual report of 27 commercial banks from year1988 to 1999.

Out of the 27 banks considered, 17 observations were identified as non-distressed bank while the other 10 were identified to be distressed banks. The independent variables considered for the classification to show the variation in financial crisis were two, which are the ratio of total shareholder's fund to total assets and the ratio of total shareholder's fund to total equity.

The performance of the CA, RLDA$_{MQ}$ and RLDA$_{WMQ}$ models were based on two types of error rates: their apparent error rates (AER) and the estimate of their error rates using CV. These approaches evaluate the performance of the models by estimating the actual error rate.

Table 4.29

Error Rates for Linear Models using Real Data

Estimators	AER	CV
CA	0.1111	0.1111
MOM-Q_n	0.0741	0.1111
WMOM-Q_n	0.1111	0.1481

From Table 4.29, the real data results show that the robust models are quite suitable for detecting outliers. For in-depth comparison, the AER approach was implemented alongside CV. From AER, the robust $RLDA_{MQ}$ model has the least actual error rate of 0.0741 while CA and $RLDA_{WMQ}$ tie at value of 0.1111. Therefore, based on AER, $RLDA_{MQ}$ is more suitable to analyze the distress situation of the banks. In addition to AER, CV approach was likewise adopted with $RLDA_{MQ}$ and CA tying for the best approaches having actual error rate of 0.1111 while $RLDA_{WMQ}$ has error rate of 0.1481. Thus, it implies that following CV, CA and $RLDA_{WMQ}$ are more suitable to analyze the distressed banks. However, the reason for implementing these two approaches (AER and CV) to estimate the error rate, is to have a better conclusion of which LDA model performs better with real data. Therefore, the $RLDA_{MQ}$ is best overall, having AER error rate of 0.0741 and CV error rate of 0.1111.

This real data results indicate that the robust estimator $RLDA_{MQ}$ detects and eliminates outliers better because it has the smallest AER misclassification error rates. In addition, the robust models having equal error rates values at certain points to the classical approach are quite encouraging. This is because there is an expected increase in error rates for the robust models as a result of the summation involved in the different data conditions considered. Hence, the justification stands that the robust discriminant models are capable of handling outliers when classifying both simulation and real data. This is proven from the results obtained using the simulation study and the real data analysis results.

CHAPTER FIVE
CONCLUSION AND FUTURE WORK

This chapter concludes this research work by considering the research objectives highlighted in Chapter One of this research and discussing how each of the objectives were actualized. Furthermore, suggestions are made on what areas can be delved into for future research work.

5.1 Conclusion

This research has looked into a comparison of robust linear discriminant analysis based on coordinate-wise approach. MOM and its winsorized counterpart WMOM were adopted with the robust Q_n scale estimator. The reason for adopting these robust estimators was because of its novel property in being able to handle outliers in data.

The first objective of this research as highlighted in Chapter One was to modify the robust estimators. This objective was achieved in Chapter Three of this research where the RLDA_{MQ} and RLDA_{WMQ} algorithms were introduced (See Sections 3.4.1 and 3.5.1). Both algorithms begin with eliminating the extreme values in the data via trimming. Likewise, the modification done in both the MOM and WMOM algorithm involved replacing the default MAD_{n} scale estimator with the Q_n estimator.

The second objective considered conducting a simulation study on the modified robust estimators by creating scenarios that could be encountered in real life by manipulation of certain variables. The variables chosen to be manipulated were chosen from thorough investigation of past literature to know which variable can give concrete discussion basis when comparing the performance of the estimators. The chosen variables were highlighted in Section 3.6 and they include dimension of variable (p), percentage of contamination (ε), sample size of the training data (n_1, n_2), shift in location of the population (μ) and shift in shape of the population (κ). Take note that the sample of training data considered both the balanced and unbalanced type which introduced the cases of equal and unequal covariance matrices. Thus, the second objective of this research was achieved by using a testing sample of size 2000 from each population and the misclassification error was computed by obtaining the proportion of misclassified testing sample observations in each population. This process was repeated 2000 times and the mean misclassification error and computational time was recorded.

The third objective achieved in this research was showing the superiority of the robust linear models over the classical linear model. The actualization of this objective was discussed in detail in Chapter Four. Two basic criterions formed the basis for comparison which was the mean misclassification error and the computational time of each model to complete the simulation run. The results from the analysis show the classical linear discriminant model performing optimally when all assumptions were fulfilled with no contamination in the data. However, as soon as the percentage of contamination was increased, the $\mathrm{RLDA_{MQ}}$ model was consistently giving the least value for the mean misclassification error for balanced sample sizes while the $\mathrm{RLDA_{WMQ}}$ had better accuracy for unbalanced sample sizes. In addition, considering the each of the variables (dimension of variable (p), percentage of contamination (ε), sample size of the training data (n_1, n_2), shift in location of the population (μ) and shift in shape of the population (κ)), certain conclusions were drawn. Although some these conclusions do not suitably follow with the classical approach, on the basis of expectation, the conclusions hold. It follows that with increase in dimension of the variables, a more consistent behavior is observed with the linear models and thus better analysis can be made. It is expected that as the percentage of contamination increases, the misclassification error also increases. Therefore, the linear model with the least mean misclassification error is said to be the best due to its ability to retain low misclassification error despite the increase in contamination. However, there is an expected inverse behavior with the increase in the shift in location of the population and shift in shape of the population. An expected decrease in misclassification error was observed with increase in μ and κ, this is in line with the statistical norm that a change in the shape or location parameters alters the behavior of the data. However, more emphasis is placed on which linear model performs best with which covariance matrix property for either balanced or unbalanced sample sizes.

From Chapter Four, it was highlighted that the simulation conditions will be investigated for both equal and unequal covariance matrix property. It was observed that the linear models tend to same behavior irrespective of the covariance matrix property while variations were more evident with balanced or unbalanced sample sizes. When the sample sizes are balanced and all assumptions of the discriminant model is met, CA performs better when there is no contamination, although CA is overridden by the robust $RLDA_{WMQ}$ model when the sample sizes become unbalanced. Furthermore, in the presence of contamination, $RLDA_{MQ}$ has a higher performance for balanced sample sizes while $RLDA_{WMQ}$ takes the lead for unbalanced sample sizes. A clearer insight to the behavior of the linear models is displayed in the following section.

Hence the conclusion that the robust models are better than classical model, although one outweighs the other when certain variables are considered but the differences are negligible. It is worth taking note of that this accuracy comes with a cost of high computational time, however this can be handled by using high-performance server available nowadays. Therefore, this research has achieved its aim of comparing certain robust linear discriminant analysis models based on coordinate-wise approach. A general comparison among the linear models are shown in the following section.

5.2 Comparison between the Linear Models

Recall that the analysis considered certain variable settings, such as the property of the covariance matrices and the sample sizes. Based on these, we would first consider a general summary of the results when the covariance matrices and sample sizes are equal as shown in Table 5.1 below.

Table 5.1

Summary of Results for Equal Covariance Matrices and Balanced Sample Size Analysis

	$(n_1,n_2)=(20,20)$			$(n_1,n_2)=(50,50)$			$(n_1,n_2)=(100,100)$		
	CA	**$RLDA_{MQ}$**	**$RLDA_{WMQ}$**	**CA**	**$RLDA_{MQ}$**	**$RLDA_{WMQ}$**	**CA**	**$RLDA_{MQ}$**	**$RLDA_{WMQ}$**
P=2	1	31	2	1	33	1	1	33	0
	2.9%	91.2%	5.9%	2.9%	97.1%	2.9%	2.9%	97.1%	0%
P=6	1	31	2	1	32	1	1	33	0
	2.9%	91.2%	5.9%	2.9%	94.2%	2.9%	2.9%	97.1%	0%
P=10	1	28	5	1	31	2	1	33	0
	2.9%	82.4%	14.7%	2.9%	91.2%	5.9%	2.9%	97.1%	0%

Based on Tables 4.1 to 4.3, a percentage analysis of the total times the linear models had the least misclassification error was computed and displayed in Table 5.1. Note that per group calculation, the total of the percentage analysis does not equal 100% because there were cases of duplicate models holding the position of the least misclassification error.

Just as discussed in Chapter Four, the robust $RLDA_{MQ}$ model had a higher percentage of smallest mean error rates. Hence, being the best model for detecting outliers in cases of equal covariance matrices and balanced sample sizes. However, this is not the case for unbalanced sample sizes as the $RLDA_{WMQ}$ model performs better overall although the $RLDA_{MQ}$ competed favourably at high variable dimension. This is shown in the summary given in Table 5.2.

Table 5.2
Summary of Results for Equal Covariance Matrices and Unbalanced Sample Size Analysis

	$(n_1,n_2)=(50,20)$			$(n_1,n_2)=(100,50)$			$(n_1,n_2)=(100,20)$		
	CA	**$RLDA_{MQ}$**	**$RLDA_{WMQ}$**	**CA**	**$RLDA_{MQ}$**	**$RLDA_{WMQ}$**	**CA**	**$RLDA_{MQ}$**	**$RLDA_{WMQ}$**
P=2	1	6	27	1	11	22	4	10	27
	2.9%	17.7%	79.4%	2.9%	32.4%	64.7%	11.8%	29.4%	79.4%
P=6	0	9	25	1	15	18	0	12	22
	0%	26.5%	73.5%	2.9%	44.1%	53%	0%	35.3%	64.7%
P=10	0	12	22	0	21	13	0	13	21
	0%	35.3%	64.7%	0%	61.8%	38.2%	0%	38.2%	61.8%

In addition, the property of unequal covariance matrices was also considered for both balanced and unbalanced sample sizes, and a general summary of the results is given in Tables 5.3 and 5.4.

Table 5.3

Summary of Results for Unequal Covariance Matrices and Balanced Sample Size Analysis

	$(n_1,n_2)=(20,20)$			$(n_1,n_2)=(50,50)$			$(n_1,n_2)=(100,100)$		
	CA	**RLDA$_{MQ}$**	**RLDA$_{WMQ}$**	**CA**	**RLDA$_{MQ}$**	**RLDA$_{WMQ}$**	**CA**	**RLDA$_{MQ}$**	**RLDA$_{WMQ}$**
P=2	1	33	0	1	33	0	1	33	0
	2.9%	97.1%	0%	2.9%	97.1%	0%	2.9%	97.1%	0%
P=6	1	32	1	1	33	0	1	33	0
	2.9%	94.2%	2.9%	2.9%	97.1%	0%	2.9%	97.1%	0%
P=10	1	25	8	1	32	1	1	33	0
	2.9%	73.5%	23.6%	2.9%	94.2%	2.9%	2.9%	97.1%	0%

Table 5.4

Summary of Results for Unequal Covariance Matrices and Unbalanced Sample Size Analysis

	$(n_1,n_2)=(50,20)$			$(n_1,n_2)=(100,50)$			$(n_1,n_2)=(100,20)$		
	CA	**RLDA$_{MQ}$**	**RLDA$_{WMQ}$**	**CA**	**RLDA$_{MQ}$**	**RLDA$_{WMQ}$**	**CA**	**RLDA$_{MQ}$**	**RLDA$_{WMQ}$**
P=2	1	2	32	0	12	22	0	5	30
	2.9%	5.9%	94.2%	0%	35.3%	64.7%	0%	14.7%	88.2%
P=6	0	8	26	0	16	18	0	7	27
	0%	23.6%	76.4%	0%	47.1%	52.9%	0%	20.6%	79.4%
P=10	0	10	24	0	17	18	0	9	25
	0%	29.4%	70.6%	0%	50%	52.9%	0%	26.5%	73.5%

Tables 5.3 and 5.4 show similar results to Tables 5.1 and 5.2 respectively, where the robust RLDA$_{MQ}$ has upper hand accuracy when the sample sizes are balanced. In addition, the robust RLDA$_{MQ}$ model competes well for the unbalanced sample sizes too by displaying close results for $(n_1,n_2)=(100,50)$ when p=6 and p=10. In the case of the unbalanced sample sizes with unequal covariance matrices, the RLDA$_{WMQ}$ robust model performs better. Table 5.5 gives an overall summary of the results obtained.

Table 5.5

Summary of Results for Performance of Models with Respect to Presence of Contaminations

Conditions	Contamination	Best Method
Equal Covariance Matrices with Balanced Sample Sizes	NO	CA
Equal Covariance Matrices with Balanced Sample Sizes	YES	$RLDA_{MQ}$
Equal Covariance Matrices with Unbalanced Sample Sizes	NO	$RLDA_{WMQ}$
Equal Covariance Matrices with Unbalanced Sample Sizes	YES	$RLDA_{WMQ}$
Unequal Covariance Matrices with balanced Sample Sizes	NO	CA
Unequal Covariance Matrices with Balanced Sample Sizes	YES	$RLDA_{MQ}$
Unequal Covariance Matrices with Unbalanced Sample Sizes	NO	$RLDA_{WMQ}$
Unequal Covariance Matrices with Unbalanced Sample Sizes	YES	$RLDA_{WMQ}$

In addition to the real data analysis results (Table 4.29), both robust estimators are equally good in detection of outliers. Although one may outshine the other based on variable manipulations.

5.3 Implication of Study

The main aim of this research is to make a comparison between linear discriminant models for multivariate data analysis. It is generally known that as a reason of the presence of outliers, these linear discriminant models have been improved from the conventional classical form to robust models. Therefore, via this research, certain advances have been made regarding some robust models in comparison to the classical approach.

MOM and WMOM were modified by the introduction of the Q_n scale estimator on both robust models ($RLDA_{MQ}$ and $RLDA_{WMQ}$). To investigate the usability of both robust models, analysis was made on both simulation and real data. The results show the robust model performing favorably in detecting outliers. Therefore, $RLDA_{MQ}$ and

RLDA_{WMQ} are suitable for classification of multivariate date even in the presence of high contamination of outliers.

5.4 Limitation of Study and Future Work

Certain limitations were observed in this study and from these limitations there is room to explore more research. The first limitation is that only two robust estimators are considered for modification, the MOM and WMOM. Therefore, future research can consider modification of other coordinate-wise based robust models.

In addition, the robust scale estimator integrated in the modification is limited to Q_n estimator, hence creating room for exploring other scale estimators. Comparison could also be made between distance based robust estimators integrated with robust scale estimators and coordinate-wise based robust estimators integrated with robust scale estimators. Likewise, comparison was made between the robust models and the classical approach only, however a comparison between these modified robust models and previously existing robust models in literature can also be investigated.

Finally, in consideration of real data, the application problem was limited to real financial data which is just one selected field of study. Other areas of applications could also be considered. Similarly more recent database values can be culled for real time analysis to draw certain conclusions in varying fields of application.

REFERENCES

Abu-Shawiesh, M. O. A., Banik, S., & Golam Kibria, B. M. (2011). A simulation study on some confidence intervals for the population standard deviation. *Statistics and Operations Research Transactions*, 35(2), 83–102.

Abu-Shawiesh, M. O., & Abdullah, M. B. (2001). A new robust bivariate control chart for location. *Communications in Statistics-Simulation and Computation*, 30(3), 513-529.

Acuna, E., & Rodriguez, C. (2004). The treatment of missing values and its effect on classifier accuracy. *In Classification, Clustering, and Data Mining Applications*, 639-647.

Ahmed, S. W., & Lachenbruch, P. A. (1977). Discriminant analysis when scale contamination is present in the initial sample. *Classification and clustering*, 331-353.

Alfaro, J. L., & Ortega, J. F. (2008). A robust alternative to Hotelling's T^2 control chart using trimmed estimators. *Quality and Reliability Engineering International*, 24(5), 601-611.

Alfaro, J. L., & Ortega, J. F. (2009). A comparison of robust alternatives to Hotelling's T^2 control chart. *Journal of Applied Statistics*, 36(11–12), 1385–1396.

Ali H. (2013). *Efficient and highly robust Hotelling T^2 Control charts using reweighted mininum vector variance* . Unpublished Ph.D. thesis, Universiti Utara Malaysia.

Ali, H., & Yahaya, S. S. S. (2013). On robust Mahalanobis distance issued from minimum vector variance. *Far East Journal of Mathematical Sciences*, 74(2), 249.

Ali, H., Yahaya, S. S. S., & Omar, Z.(2013). Robust hotelling T^2 control chart with consistent minimum vector variance. *Mathematical Problems in Engineering*, *2013*(Icoqsia), 695–702.

Ali, H., Yahaya, S., S. S., & Omar, Z. (2015). Enhancing minimum vector variance estimators using reweighted scheme. *Far East Journal of Mathematical Sciences*, 98(7), 819-830.

Allen, D. M. (1974). The relationship between variable selection and data augmentation and a method for prediction. *Technometrics*, 16, 125–127.

Alrawashdeh, M. J., Sabri, S. R. M., & Ismail, M. T. (2012). Robust linear discriminant analysis with financial ratios in special interval. *Applied Mathematical Sciences*, *6*(121), 6021–6034.

Angiulli, F., & Pizzuti, C. (2005). Outlier mining in large high-dimensional data sets. *IEEE Transactions on Knowledge and Data Engineering*, *17*(2), 203–215.

Arjmandi, M. K., & Pooyan, M. (2012). An optimum algorithm in pathological voice quality assessment using wavelet-packet-based features, linear discriminant analysis and support vector machine. *Biomedical Signal Processing and Control*, *7*(1), 3–19.

Austin, P. C., & Steyerberg, E. W. (2015). The number of subjects per variable required in linear regression analyses. *Journal of clinical epidemiology*, *68*(6), 627-636.

Ayanendranath, B., Smarajit B., & Sumitra, P. (2004). Robust discriminant analysis using weighted likelihood estimators. *Journal of Statistical Computation and Simulation*, 74(6), 445-460.

Balakrishnama, S., & Ganapathiraju, A. (1998). Linear discriminant analysis - a brief tutorial.*Institute for Signal and information Processing*, 18.

Barnett, V., & Lewis, T. (1994). *Outliers in Statistical Data*. New York: John Wiley.

Beckman, R. J., & Cook, R. D. (1983). Outliers. *Technometrics*,25(2), 119-149.

Belhumeur, P. N., Hespanha, J. P., & Kriegman, D. J. (1997). Eigenfaces vs. fisherfaces: Recognition using class specific linear projection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(7), 711–720.

Ben-Gal, I. (2005). Outliers detection. In: Maimon O. and Rockach, L. (Eds.), *Data Mining and Knowledge Discovery Handbook*: Heidelbery, Berlin: Springer, pp. 131-146.

Bengio, Y., & Grandvalet, Y. (2004). No unbiased estimator of the variance of k-fold cross-validation. *The Journal of Machine Learning Research*, 5, 1089-1105.

Bennett, P. J. (2009). *Introduction to the Bootstrap and Robust Statistics*, Bootstrapping PJ PSY711/712, 1-11.

Betz, N. E. (1987). Use of discriminant analysis in counseling psychology research. *Journal of Counseling Psychology*, 34(4), 393–403.

Borgen, F. H., & Seling, M. J. (1978). Uses of discriminant analysis following MANOVA: Multivariate statistics for multivariate purposes. *Journal of Applied Psychology*, 63(6), 689–697.

Brown, B. M. & Kildea, D. G. (1978). Reduced u-statistics and the Hodges-Lehmann estimator. *The Annals of Statistics*, 6, (4), 828-835.

Campbell, N. A. (1980). Robust procedures in multivariate analysis I robust covariance estimation. *Applied Statistics*,29(3), 231–237.

Cacoullos, T. (2014). *Discriminant analysis and applications*. Academic Press.

Cevikalp, H., Neamtu, M., Wilkes, M., & Barkana, A. (2005). Discriminative common vectors for face recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 27(1), 4–13.

Chen, Z. Y., & Muirhead, R. J. (1994). A comparison of robust linear discriminant procedures using projection pursuit methods. *Lecture Notes-Monograph Series*, 24, 163–176.

Chenouri, S., Steiner, S. H., & Mulayath, A. (2009). A multivariate robust control chart for individual observations, *Journal of Quality Technology*, 41(3), 259-271.

Cheng, G., Li, X., Lai, P., Song, F., & Yu, J.(2016). Robust rank screening for ultrahigh dimensional discriminant analysis. *Statistics and Computing*, 27(2), 535–545.

Cohen, J., Cohen, P., West, S. G., & Aiken, L. S. (2013). *Applied multiple regression/correlation analysis for the behavioral sciences*. Routledge.

Croux, C., & Dehon, C. (2001). Robust linear discriminant analysis using S-estimators. *The Canadian Journal of Statistics*, 29(3), 473–492.

Croux, C., & Rousseeuw, P. J. (1992). *Time-efficient algorithms for two highly robust estimators of scale*. In Computational Statistics, pp. 411-428, Physica-Verlag HD.

Dabney, A. R., & Storey, J. D. (2007). Normalization of two-channel microarrays accounting for experimental design and intensity-dependent relationships. *Genome Biology*, 8(3), R44.

Damico, J. S., Nettleton, S. K., Damico, H. L., & Nelson, R. L. (2014). Discriminant validity with a direct observational assessment system: Research with previously identified groups. *Clinical Linguistics & Phonetics*, 28(7–8), 617–626.

Davies, P.L. (1987). Asymptotic behaviour of S-estimators of multivariate location paramters and dispersion matrices. *The Annals of Statistics*, 15, 1269-1292.

Ella, R., Van Aelst, S. & Williem, G. (2009). The minimum weighted covariance determinant estimator. *Metrika*, 70, 177-204.

Ender, P. (2014, July). Profile Analysis. *In 2014 Stata Conference* (No. 1). Stata Users Group.

Estoup, A., Lombaert, E., Marin, J. M., Guillemaud, T., Pudlo, P., Robert, C. P., & Cornuet, J. M. (2012). Estimation of demo-genetic model probabilities with Approximate Bayesian Computation using linear discriminant analysis on summary statistics. *Molecular ecology resources*,12(5), 846-855.

Feng, J., Xu, H., & Mannor, S. (2014). Distributed Robust Learning. *arXiv preprint arXiv:*1409.5937.

Fidler, S., & Leonardis, A. (2003, June). *Robust LDA classification by subsampling. In Computer Vision and Pattern Recognition Workshop, 2003*. CVPRW'03. Conference on (Vol. 8, pp. 97-97). IEEE.

Filzmoser, P., & Todorov, V. (2013). Robust tools for the imperfect world. *Information Sciences*, 245, 4-20.

Fisher. R. A. (1936). The use of multiple measurements in taxonomic problems. *Annals of Eugenics*, 7(2),179–188.

Fukunaga, K. (2013). *Introduction to statistical pattern recognition*. Academic press.

Fung, W. K. (1995). Diagnostics in linear discriminant analysis. *Journal of the American Statistical Association*, 90(431), 952–956.

Fung, W. K. (1996). Diagnosing influential observations in quadratic discriminant analysis. *Biometrics*, 52(4), 1235–1241.

Gao, H., & Davis, J. W. (2006). Why direct LDA is not equivalent to LDA. *Pattern*

Recognition, *39*(5), 1002–1006.

Guh, R. S., Shiue, Y. R., & Yu, F. J. (2014). Real-time monitoring of the quality of multivariate processes with a SVM based classifier ensemble approach. *Journal of Quality*, 21(6), 427-454.

Haddad, F. S. (2013). *Statistical Process Control Using Modified Robust Hotelling's T2 Control Charts*. Unpublished Ph.D. thesis, Universiti Utara Malaysia.

Haddad, F. S., Syed-Yahaya, S. S., & Alfaro, J. L. (2012). Alternative Hotelling's T^2 charts using winsorized modified one-step M-estimator. *Quality and Reliability Engineering International*,29(4), 583–593.

Hampel, F. R. (2001). Robust statistics: A brief introduction and overview. *In First International Symposium on Robust Statistics and Fuzzy Techniques in Geodesy and GIS*., 295, 3-17.

Hampel, F. R. (1971). A general qualitative definition of robustness. *The Annals of Mathematical Statistics*, 1887-1896.

Hampel, F. R. (1974). The influence curve and its role in robust estimation. *Journal of the American Statistical Association*, 69(346), 383-393.

Hampel, F. R. (1985). The Breakdown Points of the Mean Combined with Some Rejection Rules. *Technometrics*, 27(2), 95.

Härdle, W. K., & Simar, L. (2012). Applied multivariate statistical analysis. *Springer Science & Business Media*.

Harlow, L. L. (2014). *The essence of multivariate thinking: Basic themes and methods*. Routledge.

Hastie, T., Buja, A., & Tibshirani, R. (1995). Penalized discriminant analysis. *The Annals of Statistics*, 73-102.

Hawkins, D. M. (1980). *Identification of Outliers.* London, Chapman and Hall.

Hintze, J. L. (2008). Quick start manual. *PASS Power Analysis and Sample Size System,* NCSS, Kaysville, Utah.

Huber, P. J. (1964). Robust Estimation of a Location Parameter. *The Annals of Mathematical Statistics*,35(1), 73–101.

Hubert, M., & Van Driesen, K. (2004). Fast and robust discriminant analysis. *Computational Statistics and Data Analysis*, 45, 301–320.

Hubert, M., Rousseeuw, P. J., & Van Aelst, S. (2008). High-Breakdown Robust Multivariate Methods. *Statistical Science*, 23(1), 92–119.

Iglewics, B., Martinez, J. (1982), Outlier Detection using robust measures of scale. *Journal of Statistical Computation and Simulation*, 15, 285-293.

Jensen, W. A., Birch, J. B. & Woodall, W. H. (2007). High breakdown estimation methods for phase I multivariate control charts. *Quality Reliability Engineering International*, 23, 615-629.

Jin, J. & An, J. (2011). Robust discriminant analysis and its application to identify protein coding regions of rice genes. *Mathematical Biosciences*, 232, 96-100.

Jin, X., Zhao, M., Chow, T. W., & Pecht, M. (2014). Motor bearing fault diagnosis using trace ratio linear discriminant analysis. *IEEE Transactions on Industrial Electronics*, 61(5), 2441-2451.

Johnson, R. (1992). *Applied Multivariate Statistical Analysis.* Prentice Hall.

Joossens, K. (2006). *Robust Discriminant Analysis.* Leuven, K. U. Leuven, Faculteit Economische en Toegepaste Economische Wetenschappen, 2006.

Kao, L. J., Lee, C. F., & Tai, T. (2015). Discriminant analysis and factor analysis: Theory and method. *Handbook of Financial Econometrics and Statistics,* 2461-2476.

Keselman, H. J., Wilcox, R. R., Algina, J., Othman, A. R., & Fradette, K. (2008). A comparative study of robust tests for spread: Asymmetric trimming strategies. *British Journal of Mathematical & Statistical Psychology*, 61, 235–253.

Keselman, H. J., Wilcox, R. R., Othman, A. R., Fradette, K., & Wilcox, R. R. (2002). Trimming , transforming statistics , and bootstrapping : circumventing the biasing effects of heterescedasticity and nonnormality. *Journal of Modern Applied Statistical Methods,* 1(2), 288-309.

Khan Mohammadi, M., Garmarudi, A. B., & De La Guardia, M. (2013). Feature selection strategies for quality screening of diesel samples by infrared spectrometry and linear discriminant analysis. *Talanta*, 104, 128–134.

Kim, H. C., Kim, D., & Bang, S. Y. (2001). A PCA mixture model with an efficient model selection method. *International Joint Conference on Neural Networks*, 430–435.

Kim, H. C., Kim, D., & Bang, S. Y. (2003). Face recognition using LDA mixture model. *Pattern Recognition Letters*, 24(15), 2815-2821.

Kim, S. J., Magnani, A., & Boyd, S. (2006). Optimal kernel selection in Kernel Fisher discriminant analysis. *In Proceedings of the 23rd International Conference on Machine Learning,* pp. 465-472. ACM.

Kim, T., & Kittler, J. (2005). Locally linear discriminant analysis for multi-modally distributed classes for face recognition with a single model image. *Pattern Analysis and Machine Intelligence, IEEE Transactions*, 27(3), 318–327.

Klaus, B. (2013). Effect size estimation and misclassification rate based variable selection in linear discriminant analysis. *Journal of Data Science*, 11(2013), 537-558.

Lee, S., & Choi, W. S. (2013). A multi-industry bankruptcy prediction model using back-propagation neural network and multivariate discriminant analysis. *Expert Systems with Applications,* 40(8), 2941–2946.

Li, C., Shao, Y., & Deng, N. (2015). Robust L1-norm two-dimensional linear discriminant analysis. *Neural Networks*, 65, 92–104.

Li, M., & Yuan, B. (2005). 2D-LDA: A statistical linear discriminant analysis for image matrix. *Pattern Recognition Letters*, 26(5), 527–532.

Li, T., Zhu, S., & Ogihara, M. (2006). Using discriminant analysis for multi-class classification: An experimental investigation. *Knowledge and Information Systems*, 10(4), 453–472.

Lim, Y. F., Yahaya, S. S. S., & Ali, H. (2016). Winsorization on linear discriminant analysis. *In Proceedings of the 4th International Conference on Quantitative Sciences and Its Applications*, 0500101-0500107.

Lim, Y. F., Yahaya, S. S. S., Idris, F., Ali, H., & Omar, Z. (2014). Robust linear discriminant models to solve financial crisis in banking sectors. *In Proceedings of the 3rd International Conference on Quantitative Sciences and Its Applications*, 798(Icoqsia), 794–798.

Loog, M., Duin R. P. W., & Haeb-Umbach R. (2001). Multiclass linear dimension reduction by weighted pairwise fisher criteria. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 23(7), 762 – 766.

Lopuhaa, H.P. (1989). On the relation between S- estimators and M-estimators of multivariate location and covariance. *The Annals of Statistics*, 17, 1662-1683.

Lu, C. D., Zhang, T. Y., Du, X. Z., & Li, C. P. (2004, August). A robust kernel PCA algorithm. In Machine Learning and Cybernetics. *Proceedings of 2004 International Conference*, 5, 3084-3087.

Lu, C., Zhang, T., Zhang, R., & Zhang, C. (2003). Adaptive robust kernel PCA algorithm. *Communication*, 621–624.

Mahir, R. A., & Al-Khazaleh, A. M. H. (2009). New method to estimate missing data by using the asymmetrical winsorized mean in a time series. *Applied Mathematical Sciences*, 3(35), 1715-1726

Maronna, R. A. R. D., Martin, D., & Yohai, V. (2006). *Robust statistics* (pp. 978-0). John Wiley & Sons, Chichester. ISBN.

Maronna, R. A., & Zamar, R. H. (2012). Robust estimates of location and dispersion for high-dimensional datasets. *Technometrics*.

Maronna, R. A., Stahel, W. A., & Yohai, V. J. (1992). Bias-robust multivariate scatter estimators of based on projections . *Journal of Multivariate Analysis*, 42(1), 141-161.

Maronna, R.A. (1976). Robust M-estimators of multivariate location and scatter. *The Annal of Statistics*, 4, 51-67.

Martínez, A. M., & Kak, A. C. (2001). Pca versus LDA. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 23(2), 228-233.

McGarigal, K., Cushman, S. A., & Stafford, S. (2013). Multivariate statistics for wildlife and ecology research. *Springer Science & Business Media*.

McLachlan, G. (2004). Discriminant analysis and statistical pattern recognition. *John Wiley & Sons*.

Mohammadi M., Midi, H., Arasan, J. & Al-Talib, B. (2011). High breakdown estimators to robustify phase II control charts. *Applied Sciences*, 11 (3), 503-511

Morrison, D. F. (1976). *Multivariate statistical methods.* New York: McGraw Hill.

Nkiruka, E., Onyeagu, S., & Okeke, J. U. (2015). Discriminant analysis by projection pursuit. *Global Journal of Science Frontier Research*, 15(6).

Okwonu, F. Z., & Othman, A. R. (2013). Heteroscedastic variance covariance matrices for unbiased two groups linear classification methods. *Applied Mathematical Sciences, 7*(138), 6855–6865.

Othman, A. R., Keselman, H. J., Padmanabhan, A. R., Wilcox, R. R., & Fradette, K. (2004). Comparing measures of the " typical " score across treatment groups. *British Journal of Mathematical and Statistical Psychology*. 57, 215 - 234.

Pei C. W. (2002). *The central limit theorem and comparing means, trimmed means, one step M-estimators and modified one step M-estimators under non normality*. Southern of California, Los Angeles, California.

Pena, D., & Prieto, J. F. (2001). Multivariate outlier detection and robust covariance matrix estimation. *Technometrics*, 3, 286-322.

Peña, D., & Prieto, J F. (2001). Multivariate outlier detection and robust covariance matrix estimation. *Technometrics August, 43*(3), 286–310.

Pires, A. M., & Branco, J. A. (2010). Projection-pursuit approach to robust linear discriminant analysis. *Journal of Multivariate Analysis*, 101(10), 2464–2485.

Pohar, M., Blas, M., & Turk, S. (2004). Comparison of Logistic regression and linear discriminant analysis : A Simulation study. *Metodoloski zvezki, 1*(1), 143–161.

Poulsen, J., & French, A. (2003). Discriminant function analysis (DA). Retrieved from *http://userwww.sfsu.edu/~efc/classes/biol710/discrim/discrim.pdf*.

Press, S. J. (2012). *Applied multivariate analysis: using Bayesian and frequentist methods of inference.* Courier Corporation.

Randles, R. H., Brofitt, J. D, Ramberg, J. S. & Hogg. R. V. (1978). Discriminant analysis based on ranks. *Journal of the American Statistical Association*, 73: 379-384.

Rao, C. R. (1948). The utilization of multiple measurements in problems of biological classification. *Journal of the Royal Statistical Society. Series B (Methodological)*, 10(2), 159-203.

Raschka, S. (2014) Linear Discriminant Analysis– bit by bit, retrieved from *http://sebastianraschka.com/Articles/2014_python_lda.html*

Reed, J. F., & Stark, D. B. (1996). Hinge estimators of location: Robust to asymmetry. *Computer methods and programs in biomedicine*, 49(1), 11-17.

Rousseeuw, P. J. (1984). Least median of squares regression. *Journal of the American statistical association*, 79(388), 871-880.

Rousseeuw, P.J. (1985). Multivariate estimators with high breakdown point. *Mathematical Statistics and its Applications*, B, 283-297.

Rousseeuw, P. J. (1991). Tutorial to robust statistics. *Journal of Chemometrics*, 5(1), 1-20.

Rousseeuw, P. J., & Croux, C. (1992). Explicit scale estimators with high breakdown point. *L1-Statistical Analysis and Related Methods,* 1, 77-92.

Rousseeuw, P. J., & Croux, C. (1993). Alternatives to the median absolute deviation. *Journal of the American Statistical association*, 88(424), 1273-1283.

Rousseeuw, P. J., & Leroy, A. M. (1987). *Robust regression and outlier detection. New York*: John Wiley.

Rousseeuw, P.J. and Van Driessen, K. (1999) A fast algorithm for the minimum covariance determinant estimator. *Technometrics*, 41: 212-223.

Rousseeuw, P.J., & Van Zomeren, B.C. (1990). Unmasking multivariate outliers and leverage points. *Journal of the American Statistical Association*. 85(411), 633-651.

Sajtos, L., & Mitev, A. (2007). SPSS research and data analysis handbook. *Alinea, Budapest*, 454-458.

Santos, F., Guyomarc'h, P., & Bruzek, J. (2014). Statistical sex determination from craniometrics: Comparison of linear discriminant analysis, logistic regression, and support vector machines. *Forensic science international*, 245, 204-e1.

Staudte, R. G., & Sheather, S. J. (1990). Robust estimation and testing: Wiley series in probability and mathematical statistics. *New York*.

Stevens, J. P. (2012). *Applied multivariate statistics for the social sciences*. Routledge.

Stone, M. (1974). Cross-validatory choice and assessment of statistical predictions. *Journal of the Royal Statistical Society*. Series B (Methodological), 111-147.

Swets, D. L., & Weng, J. J. (1996). Using discriminant eigenfeatures for image retrieval. *IEEE Transactions on pattern analysis and machine intelligence*, 18(8), 831-836.

Sajtos, L., & Mitev, A. (2007). SPSS research and data analysis handbook. *Alinea, Budapest*, 454-458.

Santos, F., Guyomarc'h, P., & Bruzek, J. (2014). Statistical sex determination from craniometrics: Comparison of linear discriminant analysis, logistic regression, and support vector machines. *Forensic Science International*, 245, 204-e1.

Shao, J., & Tu, D. (2012). *The jackknife and bootstrap*. Springer Science & Business Media.

Shao, J., Wang, Y., Deng, X., & Wang, S. (2011). Sparse linear discriminant analysis by thresholding for high dimensional data. *The Annals of Statistics*, 39(2), 1241-1265.

Staudte, R. G., & Sheather, S. J. (1990). Robust estimation and testing: Wiley series in probability and mathematical statistics. *New York.*

Stevens, J. P. (2012). *Applied multivariate statistics for the social sciences*. Routledge.

Stone, M. (1974). Cross-validatory choice and assessment of statistical predictions. J*ournal of the Royal Statistical Society. Series B (Methodological)*, 111-147.

Swets D. L. & Weng, J. (1996). Using Discriminant Eigenfeatures for Image Retrieval. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 18(8):831 – 837.

Tang, E. K., Suganthan, P. N., Yao, X., & Qin, A. K. (2005). Linear dimensionality reduction using relevance weighted LDA. *Pattern recognition*, 38, 485–493.

Teknomo, K. (2015). Discriminant Analysis Tutorial. *http://people.revoledu.com/kardi/tutorial/LDA/*

Tiku, M. L., & Balakrishnan, N. (1984). Robust multivariate classification procedures based on the MML estimators. *Communications in Statistics-Theory and Methods*, 13(8), 967-986.

Todorov, V., & Pires, A. M. (2007). Comparative performance of several robust linear discriminant analysis methods. R*EVSTAT Statistical Journal*, 5, 63-83.

Torre, F., & Black, M. J. (2001). Robust principal component analysis for computer vision. *In Computer Vision. ICCV 2001. Proceedings. Eighth IEEE International Conference*, 1, 362-369.

Uray, M. (2008). *Incremental, Robust, and Efficient Linear Discriminant Analysis Learning.* Published Ph.D. thesis, Graz University of Technology.

Vapnik, V. (2013). The nature of statistical learning theory. *Springer Science & Business Media.*

Werner, M. (2003). *Identification of multivariate outliers in large data sets*. PhD thesis, University of Colorado, Denver.

Wilcox, R. (1997). Introduction to robust estimation and hypothesis testing. *Statistical modeling and decision science.*

Wilcox, R. & Keselman, H. J. (2003). Repeated measures ANOVA based on a modified one-step M-estimator. *British Mathematical and Statistical Psychology,* 56(1):15–25.

Wilcox, R. (2002). Multiple comparisons among dependent groups based on a modified one -step M-estimator. *Biometrical*, 44, 466-477.

Wilcox, R. R., & Keselman, H. J. (2003). Modern robust data analysis methods: measures of central tendency. *Psychological methods*, 8(3), 254.

Xao, O. G., Yahaya, S. S. S., Abdullah, S., & Yusof, Z. M. (2014). H-statistic with winsorized modified one-step M-estimator for two independent groups design. *Germination of Mathematical Sciences Education and Research towards Global Sustainability*, 1605 (1), 928-931.

Yahaya, S. S. S. (2005). *Robust statistical procedures for testing the equality of central tendency parameters under skewed distributions*. Unpublished Ph. D. Thesis, Universiti Sains Malaysia.

Yahaya, S. S. S., Ali, H., & Omar, Z. (2011). An alternative hotelling T^2 control chart based on minimum vector variance (MVV). *Modern Applied Science*, 5(4), 132–

151.

Yahaya, S. S. S., Lim, Y. F., Ali, H., & Omar, Z. (2016a). Robust linear discriminant analysis. *Journal of Mathematics and Statistics,* 12(4), 312-316.

Yahaya, S. S. S., Lim, Y. F., Ali, H., & Omar, Z. (2016b). Robust linear discriminant analysis with automatic trimmed mean. *Electronic and Computer Engineering,* 8(10), 1-3.

Yahaya, S. S. S., Otluna, A. R., & Iiesehnan, H. J. (2004). Testing the equality of location parameters for skewed distributions using S1 with high breakdown robust scale estirnators, *In Theory and Applications of Recent Robust Methods,* (pp. 319-328). Birkhäuser Basel.

Yahaya, S. S. S., Othman, A. R., & Keselman, H. J. (2006). Comparing the typical score across independent groups based on different criteria for trimming. *Metodoloski zvezki*, 3(1), 49-62.

Yan, H., & Dai, Y. (2011). The comparison of five discriminant methods. *In Management and Service Science (MASS), 2011 International Conference, 1-4. IEEE.*

Yan, Y., Ricci, E., Subramanian, R., Liu, G., & Sebe, N. (2014). Multitask linear discriminant analysis for view invariant action recognition. *IEEE Transactions on Image Processing,* 23(12), 5599-5611.

Yang, J. & Yang, J. Y. (2003). Why can LDA be performed in PCA transformed space? *Pattern Recognition,* 36, 563 – 566.

Yang, J., Frangi, A. F., Yang, J. Y., Zhang, D., & Jin, Z. (2005). KPCA plus LDA: A complete kernel Fisher discriminant framework for feature extraction and recognition. I*EEE Transactions on pattern analysis and machine intelligence,* 27(2), 230-244.

Yang, J., Jin, Z., Yang, J. Y., Zhang, D., & Frangi, A. F. (2004). Essence of kernel Fisher discriminant: KPCA plus LDA. *Pattern Recognition,*37(10), 2097–2100.

Yu, H., & Yang, J. (2001). A direct LDA algorithm for high-dimensional data with application to face recognition. *Pattern Recognition,* 34(10), 2067–2070.

Yu, J. (2011). Localized Fisher discriminant analysis based complex chemical process monitoring. *AIChE Journal,* 57(7), 1817-1828.

Yu, S. X., & Shi, J. (2003). Multiclass spectral clustering. *Proceedings Ninth IEEE International Conference on Computer Vision,*1(1), 313–319.

Yusof, Z. M., Abdullah, S., Yahaya, S. S. S., & Othman, A. R. (2011). Type I error rates of Ft statistic with different trimming strategies for two groups case. *Modern Applied Science,*5(4), 236–242.

Yusof, Z. M., Othman, A. R., & Yahaya, S. S. S. (2010). Comparison of type I error rates between T1 and Ft statistics for unequal population variance using variable trimming. *Malaysian Journal of Mathematical Sciences,*4(2), 195–207.

Zollanvari, A., & Dougherty, E. R. (2015). Generalized consistent error estimator of linear discriminant analysis. *IEEETransactions on Signal Processing*, 63(11), 2804–2814.

Zuo, Y. (2006). The Frontiers in Statistics. In Peter Bickel on his 65th Birthday, *Robust location and scatter estimators in multivariate analysis.* Imperial College Press.

Appendix A

Program Calculates the Value of the Robust Scale Estimator *Qn*

```
function Result=Qn(X)
[s1 s2]=size(X);
dist=zeros(s1,s2);
count=0;
for i=1:s1
    for j=1:s1
        if i<j
            count=count+1;
            dist(count,s2)=abs(X(i,s2)-X(j,s2));
        end
    end
end
sortdist=sort(dist);
h=floor(s1/2)+1;
k=nchoosek(h,2);
Result=sortdist(k,s2)*2.2219;
```

Appendix B

Programs for Calculates Modified One-Step M-Estimator $RLDA_{MQ}$ and Winsorized Modified One-Step M-Estimator $RLDA_{WMQ}$ Sample with the scale estimator Qn

1- Program calculates the $RLDA_{MQ}$

```
function Result=MOM_Qn_sample(Y)
[S1 S2]=size(Y);
if S2>1
   disp('error Only vectors not coulumns or Matrices');
return;
end

Med=median(Y);
QN= Qn(Y);
const = 2.24;
Low=-const*QN;
High=const*QN;
k=0;
for i=1:S1,
   if ((Y(i) - Med) >= Low) && ((Y(i) - Med) <= High)
      k= k+1;
   end
end
X = zeros(k,S2);
k=1;
for i=1:S1
   if ((Y(i) - Med) >= Low) && ((Y(i) - Med) <= High)
      X(i) = Y(i);
      k= k+1;
   else
      X(i)=nan;
   end
end

Result=X;
```

2- Program calculates the $RLDA_{WMQ}$

```
function Result=WQn_sample(Y)
[S1 S2]=size(Y);
if S2>1
   disp('error Only vectors not coulumns or Matrices');
return;
end

Med=median(Y);
QN= Qn(Y);
const = 2.24;
Low=-const*QN;
```

```
High=const*QN;
k=0;
for i=1:S1,
   if ((Y(i) - Med) >= Low) && ((Y(i) - Med) <= High)
      k= k+1;
   end
end
X = zeros(k,S2);
k=1;
for i=1:S1
   if ((Y(i) - Med) >= Low) && ((Y(i) - Med) <= High)
      X(i) = Y(i);
      k= k+1;
   end
end
Max = max(X);
Min = min(X);
for i=1:S1
if ((Y(i) - Med) < Low)
   X(i) = Min;
elseif((Y(i) - Med) > High)
   X(i) = Max;
end
end
Result=X;
```

Appendix C

Programs for Simulation Study

1- Programs for Simulation $RLDA_{MQ}$

```
function result = simulation_MOM_Qn
clear all;
start_time = cputime;

N1=2000;
N2=2000;
n1=20;
n2=20;
p1=2;
err = 0.4;
R=2000;

miscl = zeros(R,1);

for r=1:R

   seed1 = 12954+r;
   randn('seed',seed1);

   G1=randn(N1,p1);
   G2=1+2*randn(N2,p1);

   V1 = repmat(1:1, [N1 1]);
   V2 = repmat(2:2, [N2 1]);

   test_data=[G1 V1
            G2 V2];

   [n,p] = size(test_data);

   seed = 3984+r;
   randn('seed',seed);

   X1=[randn((1-err)*n1,p1)
      3+randn(err*n1,p1)];
   X2=[1+2*randn((1-err)*n2,p1)
      -2+2*(randn(err*n2,p1))];

   MS_Qn1 = zeros(n1,p1);
   MS_Qn2 = zeros(n2,p1);
   Qn_X1=zeros(1,p1);
   Qn_X2=zeros(1,p1);

 for i=1:p1
  MS_Qn1(1:n1,i) = MOM_Qn_sample(X1(1:n1,i));
```

```
  MS_Qn2(1:n2,i) = MOM_Qn_sample(X2(1:n2,i));
 end

   dim = p-1;
   a = log (n2/n1);

  for i=1:p1
   Qn_X1(i) = Qn(X1(1:n1,i));
   Qn_X2(i) = Qn(X2(1:n2,i));
  end

   Product_Qn_X1=Qn_X1'*Qn_X1;
   Product_Qn_X2=Qn_X2'*Qn_X2;

   mu1  = nanmean(MS_Qn1); mu2  = nanmean(MS_Qn2);
   cov1 = corr(X1,'type','Spearman').*Product_Qn_X1;
   cov2 = corr(X2,'type','Spearman').*Product_Qn_X2;

   sigma  = ((n1-1)*cov1+(n2-1)*cov2)/(n1+n2-2);
   linear   = (mu1-mu2)/sigma;
   constant = 1/2*linear*(mu1+mu2)';
   scores   = linear*test_data(1:n,1:dim)' - constant ;

   group  = (scores < a) + 1;
   miscl(r) = mean(group ~= test_data(:,p)');
end

end_time = cputime;

result.average_MOM_Qn_miscl =mean(miscl);
result.std_dev_MOM_Qn_miscl =std(miscl);
result.exec_time = end_time-start_time;
```

2- Programs for Simulation $RLDA_{WMQ}$

```
function result = simulation_WMOM_Qn
clear all;
start_time = cputime;

N1=2000;
N2=2000;
n1=50;
n2=20;
p1=2;
err = 0.4;
R=2000;

miscl = zeros(R,1);

for r=1:R
```

```
    seed1 = 12954+r;
    randn('seed',seed1);

    G1=randn(N1,p1);
    G2=1+2*randn(N2,p1);

    V1 = repmat(1:1, [N1 1]);
    V2 = repmat(2:2, [N2 1]);

    test_data=[G1 V1
               G2 V2];

    [n,p] = size(test_data);

    seed = 3984+r;
    randn('seed',seed);

    X1=[randn((1-err)*n1,p1)
        3+randn(err*n1,p1)];
    X2=[1+2*randn((1-err)*n2,p1)
        -2+2*(randn(err*n2,p1))];

    WG1 = zeros(n1,p1);
    WG2 = zeros(n2,p1);

for i=1:p1
   WG1(1:n1,i) = WQn_sample(X1(1:n1,i));
   WG2(1:n2,i) = WQn_sample(X2(1:n2,i));
end

    dim = p-1;
    a = log (n2/n1);

    mu1  = mean(WG1); mu2  = mean(WG2);
    cov1 = cov(WG1); cov2 = cov(WG2);

    sigma  = ((n1-1)*cov1+(n2-1)*cov2)/(n1+n2-2);
    linear   = (mu1-mu2)/sigma;
    constant = 1/2*linear*(mu1+mu2)';
    scores   = linear*test_data(1:n,1:dim)' - constant ;

    group  = (scores < a) + 1;
    miscl(r) = mean(group ~= test_data(:,p)');
end

end_time = cputime;

result.average_WMOM_Qn_miscl =mean(miscl);
result.std_dev_WMOM_Qn_miscl =std(miscl);
result.exec_time = end_time-start_time;
```

Appendix D

Programs for Real Data

1- Programs for Real Data $RLDA_{MQ}$

```
[n,p] = size(datafull);
   [N,P] = size(datafull);

   dim = p-1;
   Dim = P-1;

   X1 = datafull(datafull(:,p)==1,1:dim);
   X2 = datafull(datafull(:,p)==2,1:dim);
   n1 = size(X1,1);
   n2 = size(X2,1);
   a = log (n2/n1);
   MS_Qn1 = zeros(n1,dim);
   MS_Qn2 = zeros(n2,dim);
   Qn_X1=zeros(1,dim);
   Qn_X2=zeros(1,dim);

 for i=1:dim
  MS_Qn1(1:n1,i) = MOM_Qn_sample(X1(1:n1,i));
  MS_Qn2(1:n2,i) = MOM_Qn_sample(X2(1:n2,i));
 end

  for i=1:dim
   Qn_X1(i) = Qn(X1(1:n1,i));
   Qn_X2(i) = Qn(X2(1:n2,i));
  end

   Product_Qn_X1=Qn_X1'*Qn_X1;
   Product_Qn_X2=Qn_X2'*Qn_X2;

   mu1  = nanmean(MS_Qn1); mu2  = nanmean(MS_Qn2);
   cov1 = corr(X1,'type','Spearman').*Product_Qn_X1;
   cov2 = corr(X2,'type','Spearman').*Product_Qn_X2;

   sigma  = ((n1-1)*cov1+(n2-1)*cov2)/(n1+n2-2);
   linear   = (mu1-mu2)/(sigma);
   constant = 0.5*linear*(mu1+mu2)';
   scores   = linear*datafull(1:N,1:Dim)' - constant ;
   group  = (scores < a) + 1;
   miscl = mean(group ~= datafull(:,P)');
```

2- Programs for Real Data $RLDA_{WMQ}$

```
[n,p] = size(datafull);
   [N,P] = size(datafull);

   dim = p-1;
   Dim = P-1;

   X1 = data27(data27(:,p)==1,1:dim);
   X2 = data27(data27(:,p)==2,1:dim);
   n1 = size(X1,1);
```

```
    n2 = size(X2,1);
    a = log (n2/n1);
    WG1 = zeros(n1,dim);
    WG2 = zeros(n2,dim);

for i=1:dim
   WG1(1:n1,i) = WQn_sample(X1(1:n1,i));
   WG2(1:n2,i) = WQn_sample(X2(1:n2,i));
end

    mu1  = mean(WG1); mu2  = mean(WG2);
    cov1 = cov(WG1);  cov2 = cov(WG2);

    sigma  = ((n1-1)*cov1+(n2-1)*cov2)/(n1+n2-2);
    linear   = (mu1-mu2)/(sigma);
    constant = 0.5*linear*(mu1+mu2)';
    scores   = linear*datafull(1:N,1:Dim)' - constant ;
    group  = (scores < a) + 1;
    miscl = mean(group ~= datafull(:,P)');
```