

2024-08-01

Robust Multivariate Estimation And Inference With The Minimum Density Power Divergence Estimator

Ebenezer Nkum
University of Texas at El Paso

Follow this and additional works at: https://scholarworks.utep.edu/open_etd



Part of the [Mathematics Commons](#), and the [Statistics and Probability Commons](#)

Recommended Citation

Nkum, Ebenezer, "Robust Multivariate Estimation And Inference With The Minimum Density Power Divergence Estimator" (2024). *Open Access Theses & Dissertations*. 4195.
https://scholarworks.utep.edu/open_etd/4195

This is brought to you for free and open access by ScholarWorks@UTEP. It has been accepted for inclusion in Open Access Theses & Dissertations by an authorized administrator of ScholarWorks@UTEP. For more information, please contact lweber@utep.edu.

ROBUST MULTIVARIATE ESTIMATION AND INFERENCE WITH THE
MINIMUM DENSITY POWER DIVERGENCE ESTIMATOR

EBENEZER NKUM

Doctoral Program in Data Science

APPROVED:

Abhijit Mandal, Ph.D., Chair

Michael Pokojovy, Ph.D., Co-Chair

Suneel Babu Chatla, Ph.D.

Nilotpal Sanyal, Ph.D.

Thomas M. Fullerton Jr., Ph.D.

Stephen L. Crites, Ph.D.
Dean of the Graduate School

©Copyright

by

EBENEZER NKUM

2024

To my
MOTHER and FATHER
with love

ROBUST MULTIVARIATE ESTIMATION AND INFERENCE WITH THE
MINIMUM DENSITY POWER DIVERGENCE ESTIMATOR

by

EBENEZER NKUM

DISSERTATION

Presented to the Faculty of the Graduate School of

The University of Texas at El Paso

in Partial Fulfillment

of the Requirements

for the Degree of

DOCTOR OF PHILOSOPHY

Department of Mathematical Sciences

THE UNIVERSITY OF TEXAS AT EL PASO

AUGUST 2024

Acknowledgements

First and foremost, I give glory to Almighty God, whose divine guidance and grace have been with me every step of the way. His strength, wisdom, and perseverance have carried me through this journey, and I am eternally grateful for His blessings.

I would like to express my deepest gratitude to my research advisor (Chair), Dr. Abhijit Mandal, from the Department of Mathematical Sciences at The University of Texas at El Paso, and my Co-Chair, Dr. Michael Pokojovy, from the Data Science, Mathematics & Statistics at Old Dominion University. Their advice, encouragement, and unwavering support have been invaluable throughout this journey. Dr. Mandal's insightful guidance and Dr. Pokojovy's constant belief in my abilities have been the bedrock of my success. Their dedication to my work and their willingness to invest their time and expertise have been truly remarkable. I wish all students the honor and opportunity to experience their extraordinary mentorship.

I would also like to express sincere gratitude to my committee members: Dr. Suneel Chatla, Dr. Nilotpal Sanyal, and Dr. Thomas M. Fullerton Jr., all from The University of Texas at El Paso. Their suggestions, comments, and additional guidance were invaluable to the completion of this work. Their expertise and thoughtful feedback have significantly shaped this dissertation.

I am also very grateful to my professors, Dr. Su Xiaogang, Dr. Asim Kumer Dey, and Dr. Art Duval. Each of these individuals has played a critical role in my development at The University of Texas at El Paso. Their advice and encouragement have prepared me for the task ahead in the world of Mathematics.

I want to extend my appreciation to Dr. Ori Rosen, the new Program Director, for his support and guidance, and to the former Program Director, Dr. Amy Wagler, for her

invaluable contributions. Additionally, I would like to thank our Program Coordinator, Dr. Azucena Zamora, for her invaluable support and assistance throughout my studies.

I would also like to thank Dr. Andrews T. Anum and Dr. Michael Pokojovy for providing me with the initial resources for my dissertation.

Finally, I am deeply grateful to my family and colleagues. Their unwavering support, encouragement, and moral support have made this journey a memorable one. I do not have the words to express all my feelings here, only that I love you all and could not have done this without you.

Abstract

The estimation of the location vector and scatter matrix plays a crucial role in many multivariate statistical methods. However, the classical likelihood-based estimation is greatly influenced by outliers, potentially leading to unreliable decisions. Hence, a fundamental challenge in multivariate statistics is to develop robust alternatives that can maintain performance in the presence of outliers and deviations from the assumed data distribution. Unfortunately, methods with good global robustness often substantially sacrifice efficiency. To address this, we propose the adoption of Minimum Density Power Divergence (MDPD) estimation, a well-established robust technique known for its efficiency and statistical robustness to outliers and model violations. Focusing on multivariate contaminated Gaussian models, we present the first-order optimality conditions associated with minimizing the loss function underlying the MDPD estimator. We also describe a computationally efficient iterative algorithm designed to converge to a local minimum of the loss function. Additionally, we develop a robust one-way Multivariate Analysis of Variance (MANOVA) test based on the MDPD estimator, which is particularly useful for analyzing multiple dependent variables simultaneously, especially when a significant correlation between dependent variables exists. The asymptotic properties of the MDPD estimator and proposed MANOVA test are derived under suitable regularity conditions. Extensive Monte Carlo simulations are further conducted to empirically evaluate the statistical efficiency and quantitative robustness of the proposed methods. Furthermore, we provide real-world dataset examples on robust principal component analysis (PCA), multivariate regression, and one-way MANOVA tests. Our proposed approach is observed to be competitive or superior when compared both to classical likelihood-based methods and robust techniques based on the Minimum Covariance Determinant (MCD) estimator.

Contents

	Page
Acknowledgements	v
Abstract	vii
Table of Contents	viii
List of Figures	xi
Chapter	
1 Introduction	1
2 Fundamentals of Robust Statistical Methods	7
2.1 Influence Function	7
2.2 Breakdown Point	8
2.3 Affine Equivariance	10
2.4 Efficiency	12
2.5 Computational Complexity	13
3 Review of Robust Statistical Estimators	17
3.1 Classical Estimators of Multivariate Location and Scatter	17
3.2 Robust Estimators of Multivariate Location and Scatter	19
3.2.1 Multivariate M-Estimators	19
3.2.2 Minimum Covariance Determinant (MCD)	21
4 Minimum Density Power Divergence (MDPD)	31
4.1 Multivariate Gaussian MDPD	33
4.2 Asymptotic Distribution of MDPD Estimators	36
4.2.1 Asymptotic Distribution of the Location Vector	37
4.2.2 Asymptotic Distribution of the Scatter Matrix	37
4.3 Gross Breakdown Point	37

5	Multivariate Gaussian MDPD Algorithm	43
6	Simulation Study for MDPD Estimators	47
6.1	Data generating setup	48
6.2	Evaluation Criterion	49
6.2.1	No Contamination	49
6.2.2	Point Contamination	52
6.2.3	Cluster Contamination	55
6.3	Empirical Breakdown Point	59
7	Real Data Analysis with MDPD Estimators	62
7.1	PCA Diagnostics	62
7.2	Multivariate Regression Diagnostics	69
8	One-Way Robust Multivariate Analysis of Variance (MANOVA)	74
8.1	Parameter Estimation	76
8.2	Hypothesis Testing	79
8.3	Choosing Optimal α	81
8.4	Small Sample Correction	82
8.5	Monte Carlo Simulations	82
8.5.1	Accuracy of the Distribution of the Test Statistic	84
8.5.2	Empirical Levels for Varying Sample Sizes and Dimensions	87
8.5.3	Empirical Levels for Different Contaminations and Error Distributions	90
8.5.4	Empirical Levels for Different Sample Sizes	91
8.5.5	Empirical Powers for Different Contaminations and Error Distributions	93
8.5.6	Empirical Powers for Different Sample Sizes	95
8.6	Real Data Analysis	97
9	Conclusions	104
	References	107
	Appendix	
A	118

Appendix A	118
A.1 Integrals for DPD Measure	118
A.2 Estimating Equations	119
A.3 Asymptotic Distribution of MDPD Estimator	121
A.4 Asymptotic Variance of MDPD Estimator of $\boldsymbol{\mu}$	122
A.4.1 Score Function of $\boldsymbol{\mu}$	122
A.4.2 Calculation of $\mathbf{J}_{\boldsymbol{\mu}\boldsymbol{\mu}}$	123
A.4.3 Calculation of $\boldsymbol{\xi}_{\boldsymbol{\mu}}$	123
A.4.4 Calculation of $\mathbf{K}_{\boldsymbol{\mu}\boldsymbol{\mu}}$	124
A.4.5 Calculation of $\mathbf{J}_{\boldsymbol{\mu}\boldsymbol{\mu}}^{-1}\mathbf{K}_{\boldsymbol{\mu}\boldsymbol{\mu}}\mathbf{J}_{\boldsymbol{\mu}\boldsymbol{\mu}}^{-1}$	125
A.5 Asymptotic Variance of MDPD Estimator of $\boldsymbol{\Sigma}$	125
A.5.1 Score Function of $\boldsymbol{\Sigma}$	125
A.5.2 Calculation of $\mathbf{J}_{\boldsymbol{\Sigma}\boldsymbol{\Sigma}}$	125
A.5.3 Calculation of $\boldsymbol{\xi}_{\boldsymbol{\Sigma}}$	127
A.5.4 Calculation of $\mathbf{K}_{\boldsymbol{\Sigma}\boldsymbol{\Sigma}}$	128
A.5.5 Calculation of $\mathbf{J}_{\boldsymbol{\Sigma}\boldsymbol{\Sigma}}^{-1}\mathbf{K}_{\boldsymbol{\Sigma}\boldsymbol{\Sigma}}\mathbf{J}_{\boldsymbol{\Sigma}\boldsymbol{\Sigma}}^{-1}$	130
A.6 Covariance Matrices	130
A.6.1 Calculation of $\mathbf{J}_{\boldsymbol{\Sigma}\boldsymbol{\mu}}$	130
A.6.2 Calculation of $\mathbf{K}_{\boldsymbol{\Sigma}\boldsymbol{\mu}}$	132
A.7 MANOVA Setup	132
A.7.1 \mathbf{J} and \mathbf{K} Matrices at the Model	132
A.8 Test Statistic	133
A.8.1 Covariance Matrix of $\hat{\boldsymbol{\mu}}$	134
A.9 Regularity Conditions	135
A.10 Derivation of Breakdown Point	136
B	138
Curriculum Vitae	145

List of Figures

4.1	Maximum nominal breakdown point for the MDPD estimator.	41
6.1	MSE and SD of errors of the mean vector and scatter matrix estimates for uncontaminated data: simulation scenario with $p = 2$, and $bdp = 0.25$	50
6.2	MSE and SD of errors of the mean vector and scatter matrix estimates for uncontaminated data: simulation scenario with $p = 5$, and $bdp = 0.0924$. . .	50
6.3	MSE and SD of errors of the mean vector and scatter matrix estimates for uncontaminated data: simulation scenario with $p = 10$, and $bdp = 0.0502$. . .	51
6.4	MSE and SD of errors of the mean vector and scatter matrix estimates for uncontaminated data: simulation scenario with $p = 15$, and $bdp = 0.0345$. . .	51
6.5	MSE and Kullback–Leibler divergence of the mean vector and scatter matrix estimates for point contamination in simulation scenarios: $p = 2$, $n = 100$, $\varepsilon = 12.5\%$, and $bdp = 0.25$	53
6.6	MSE and Kullback–Leibler divergence of the mean vector and scatter matrix estimates for point contamination in simulation scenarios: $p = 5$, $n = 200$, $\varepsilon = 6.2\%$, and $bdp = 0.12$	53
6.7	MSE and Kullback–Leibler divergence of the mean vector and scatter matrix estimates for point contamination in simulation scenarios: $p = 10$, $n = 500$, $\varepsilon = 3.4\%$, and $bdp = 0.067$	54
6.8	MSE and Kullback–Leibler divergence of the mean vector and scatter matrix estimates for point contamination in simulation scenarios: $p = 15$, $n = 1000$, $\varepsilon = 1.5\%$, and $bdp = 0.046$	54
6.9	MSE and Kullback–Leibler divergence of the mean vector and scatter matrix estimates for cluster contamination in simulation scenarios: $p = 2$, $n = 100$, $\varepsilon = 12.5\%$, and $bdp = 0.25$	56

6.10	MSE and Kullback–Leibler divergence of the mean vector and scatter matrix estimates for cluster contamination in simulation scenarios: $p = 5$, $n = 200$, $\varepsilon = 6.2\%$, and $bdp = 0.12$ (cluster contamination).	57
6.11	MSE and Kullback–Leibler divergence of the mean vector and scatter matrix estimates for cluster contamination in simulation scenarios: $p = 10$, $n = 500$, $\varepsilon = 3.4\%$, and $bdp = 0.067$ (cluster contamination).	57
6.12	MSE and Kullback–Leibler divergence of the mean vector and scatter matrix estimates for cluster contamination in simulation scenarios: $p = 15$, $n = 1000$, $\varepsilon = 1.5\%$, and $bdp = 0.046$ (cluster contamination).	58
7.1	Boxplots for the counterfeit part of the Swiss Banknote dataset.	63
7.2	Explained variance by the classical and robust PCA	65
7.3	Overall empirical influence for the eigenvectors versus the Mahalanobis distance based on classical scatter matrix for the counterfeit notes in the Swiss banknote dataset.	67
7.4	Diagnostic Plot: Overall empirical influence for the eigenvectors versus the robust distance based on MDPD (Panel a) and FastMCD (Panel b) for the counterfeit notes in the Swiss banknote dataset.	68
7.5	Density plot of the predictor variables.	69
7.6	Caption for	72
8.1	Significance Level (Panel a) and Power (Panel b) for Uncontaminated Data.	86
8.2	Significance Level (Panel a) and Power (Panel b) for 5% random contamination.	86
8.3	Significance Level (Panel a) and Power (Panel b) for 5% cluster contamination.	87
8.4	The empirical level of different tests in uncontaminated data (top left) and in the presence of 1% random outliers at random locations (top right), 5% random outliers (bottom left), and 5% clustered outliers (bottom right). In all cases, $k = 4$ and $p = 5$	92

8.5	The power of different tests in uncontaminated data (top left) and in the presence of 1% random outliers at random locations (top right), 5% random outliers (bottom left), and 5% clustered outliers (bottom right). In all cases, $k = 4$, $p = 5$ and $\boldsymbol{\mu}_1 = (-0.4, 0.02, -0.2, -0.1, -0.4)^T$	96
8.6	Boxplot of the log-transformed lithological data from four Oslo transects. .	99
8.7	MCD($\alpha = 0.75$) distance-distance plot of the log-transformed lithological data from four Oslo transects (most influential outliers are colored red). . .	100

Chapter 1

Introduction

Robustness in parameter estimation is a critical aspect of statistical analysis, especially when dealing with multivariate datasets prone to outliers. Robust estimation aims to develop methods that provide stable and unbiased parameter estimates even when the data contains outliers. Traditional statistical estimates employed to characterize such datasets are susceptible to the detrimental effects of outliers. A single outlier can significantly influence an optimal classical method under the model assumption ([Maronna et al., 2019](#)).

[Tukey \(1960\)](#) initiated a critical discussion regarding the robustness of statistical methods, raising the question: What happens if the true distribution deviates slightly from the assumed normal one? This inquiry revealed that traditional measures, such as the sample mean and variance, perform poorly under such conditions. This pivotal discussion has significantly contributed to the development of robust estimation methods, which seek to offer more reliable performance in the presence of deviations from model assumptions.

Building on this foundation, [Huber \(1992\)](#) provides an insightful reflection on the fundamental aspects of estimation theory, tracing its origins back to Gauss and his development of the least squares method. According to [Huber \(1992\)](#), Gauss primarily adopted the normal distribution and a quadratic loss function for their mathematical convenience, particularly in computational contexts. Over time, the critical rationale behind these choices faded from memory, often overshadowed by the central limit theorem. However, the central limit theorem can, at most, explain why most distributions occurring in practice are “approximately” normal. Therefore, statistical estimation has emerged, providing tools capable of

maintaining performance even in the presence of outliers or deviations from the assumed distribution. These methods have become essential in various fields, from economics and engineering to biology. The extent to which these estimators are robust against outliers has been extensively studied (see [Hampel, 1974](#); [Donoho and Huber, 1983](#)). We will discuss it extensively in Chapter 3.

In classical multivariate analysis, the sample mean vector and the sample covariance (or scatter) matrix are one of the foundational elements. They play a critical role in various analytical techniques, including multivariate analysis of variance, principal component analysis, factor analysis, canonical correlation analysis, discriminant analysis, classification, and clustering ([Zuo, 2006](#)). These estimators are optimal or efficient estimators of location and scatter parameters under multivariate normal distributions. However, they are notably vulnerable to outliers, often not apparent through visual inspection. For instance, while Mahalanobis distance based on these classical estimators can effectively detect single outliers, it typically faces challenges with multiple outliers, which may not exhibit large distances due to the masking effect. This effect occurs when multiple outliers collectively influence the mean and covariance estimates to the extent that it obscures their anomalous nature (see [Davies and Gather, 1993](#); [Muthukrishnan and Mahesh, 2014](#)). Also, the swamping effect can arise, where genuine data points are misidentified as outliers because the presence of true outliers has skewed the estimation metrics. For illustrative examples of the sensitivity of the sample mean vector and the sample covariance matrix to outliers, refer to [Maronna \(1976\)](#) and [Rousseeuw and Leroy \(2005\)](#) for detailed discussions and analyses.

Over the years, the field of robust estimation has significantly developed, incorporating a diverse range of resilient techniques against outliers while retaining the desirable properties of classical non-robust estimators, such as efficiency, affine equivariance, and consistency. In their work, [Ronchetti and Huber \(2009\)](#) and [Hampel et al. \(1986\)](#) and [Anum and Pokojovy \(2024\)](#) categorized robust estimators into classes such as M-, L-, R-, and S-estimators along with other types like A-, D-, W-, and P-estimators. Among the most notable robust

estimators methodologies for location and scatter are the minimum covariance determinant (MCD) estimator of [Rousseeuw \(1984\)](#) (see also [Rousseeuw and Driessen \(1999\)](#); [Hubert et al. \(2012\)](#); [Pokojoy and Jobe \(2022\)](#) for their computational algorithms), Stahel-Donoho estimator (see [Donoho, 1982](#); [Stahel, 1981](#)), M-estimators ([Maronna, 1976](#)) and much more. For further discussion and systematic review (see [Maronna et al., 2019](#)).

The quest for robust alternatives began with [Bickel \(1964\)](#), who introduced the coordinate-wise median and the coordinate-wise Hodges-Lehmann estimator as more robust options than the sample mean vector. Furthering this effort, [Bickel \(1965\)](#) proposed the metrically trimmed and winsorized means in the multivariate domain based on [Tukey \(1960\)](#) ideas. Although these estimators significantly enhance robustness against outliers and contaminated data, they do not maintain the desirable property of affine equivariance, a critical natural property for maintaining consistency under linear transformations of data. Beginning with [Huber \(1972\)](#) work on the “peeling” procedure, [Gnanadesikan and Kettenring \(1972\)](#) work on iterative trimming, location estimators were introduced that are affine-equivariant, though little is known about their properties. [Hampel \(1973\)](#) was the first to propose an affine equivariant iterative procedure for the estimation of the scatter matrix, equivalent to an M-estimator of the scatter matrix. Similarly, [Maronna \(1976\)](#) extended M-estimators to general M-estimators of multivariate location and scatter parameters, which are affine equivariant by design (see also [Huber, 1977](#)). Multivariate M-estimates demonstrate robust local properties; they are minimally influenced by small data deviations and maintain good efficiency with the normal model and a wide array of other population models. However, studies by [Stahel \(1981\)](#) and [Maronna \(1976\)](#) have indicated that multivariate M-estimates lack global robustness, evidenced by their relatively low breakdown points ($\leq 1/(p+1)$) in high dimensions. Following this, high breakdown point affine equivariant estimators for multivariate location and scatter were introduced by [Stahel \(1981\)](#), [Donoho \(1982\)](#), [Rousseeuw \(1985\)](#), [Davies \(1987\)](#), among others. Interestingly, many high-breakdown point estimators suffer from poor local robustness properties. For instance, MCD has low efficiency when applied to the normal model. To enhance its efficiency,

[Lopuhaä and Rousseeuw \(1991\)](#) proposed a one-step reweighting procedure to improve its efficiency. In contrast, [Kent and Tyler \(1996\)](#) proposed constrained M-estimators, which effectively combine good local and global robustness properties. These estimators can be tuned to achieve both high breakdown points and strong local robustness. As evidenced in literature, one objective in robust statistics is to develop affine equivariant alternatives to the sample mean and covariance matrix estimators. These alternatives should feature high breakdown points and also demonstrate high efficiency across a wide variety of population models.

The pursuit of robust estimation is not confined to traditional parametric models but extends to more general settings. Researchers have devised estimators inspired by the maximum likelihood approach, replacing it with minimum distance or divergence estimation to enhance robustness. The pioneering work of [Beran \(1977\)](#), followed by enhancements proposed by [Basu and Lindsay \(1994\)](#), exemplifies this trend. These estimators harness divergence measures like Hellinger distance and density power divergence to navigate the complexities of robust estimation in continuous parametric models. Under appropriate regularity conditions, [Beran \(1977\)](#) demonstrated that the proposed estimators achieve complete asymptotic efficiency within the model. The methodology hinges on the selection of bandwidth, a process that may introduce certain adverse effects in continuous models. This is because the estimators necessitate the use of nonparametric smoothing techniques, which are known to have a substantial impact on kernel-based procedures. In fact, one such non-parametric smoothing technique is kernel density estimation, which is relied upon to provide an estimate of the population density. A significant modification to the original procedure proposed by [Beran \(1977\)](#) was introduced by [Basu and Lindsay \(1994\)](#). To mitigate the reliance on bandwidth selection, they adopted a novel approach by applying the same kernel for smoothing both the model and the data.

In this same vein, [Basu et al. \(1998\)](#) pioneered a class of minimum density power divergence estimators, distinct yet related to the approach proposed by [Windham \(1995\)](#), designed

for robust estimation within the broader framework of general parametric models. These estimators provide an approach to parameter estimation, aiming to balance breakdown point and efficiency.

Therefore, this dissertation seeks to estimate robust mean vector and scatter matrix by minimizing the power density divergence measure by [Basu et al. \(1998\)](#) within the multivariate Gaussian model. The univariate scenerio has been extensively studied by [Anum and Pokojovy \(2024\)](#), where it is shown to balance the breakdown point and efficiency effectively. Focusing on multivariate contaminated Gaussian models that commonly arise in various practical situations, we present first-order optimality conditions associated with minimizing the loss function underlying the MDPD estimator. We further develop a computationally efficient iterative algorithm designed to converge to a local extremum of the MDPD objective. We also propose a robust one-way MANOVA Wald-type test based on the MDPD estimator, expanding on the robust ANOVA Wald-type test studied by [Das et al. \(2022\)](#) and derive the asymptotic properties of the proposed test. Robust test based on MDPD estimator has been shown to demonstrate significantly superior performance compared to likelihood-based tests in the presence of outliers (see [Basu et al., 1998, 2015](#)), while remaining highly competitive with pure data. Consequently, tests based on MDPD estimators are invaluable tools in robust statistics (see [Basu et al., 2011; Pardo, 2018](#)).

We organized the rest of the dissertation as follows. Chapter 2 discusses fundamentals of robust statistical methods including the influence function, breakdown point, affine equivariance, and efficiency. Chapter 3 provides a comprehensive review of statistical estimators, covering both classical and robust multivariate location and scatter estimators. Chapter 4 introduces the minimum density power divergence (MDPD) estimator, discussing its theoretical framework and properties, such as the asymptotic distribution. We derived and extended the breakdown point of the MDPD estimator in the multivariate setup. Chapter 5 presents the MDPD algorithm, detailing the iterative procedure for estimating the mean vector and covariance matrix under the multivariate Gaussian model and evaluating the

algorithm's computational efficiency. In Chapter 6, we conduct an extensive simulation study comparing the performance of the MDPD estimator against other robust estimators under various scenarios, highlighting its robustness and efficiency. Chapter 7 demonstrates the MDPD estimator's practical application through real-world data examples, including principal component analysis and multivariate regression. Chapter 8 introduces a new robust MANOVA testing procedure based on the MDPD estimator. We derive the asymptotic null distribution, conduct extensive simulations, and provide real-data examples to validate its robustness and efficiency. Finally, Chapter 9 concludes the dissertation with a summary of the findings, their implications, and potential directions for future research.

Chapter 2

Fundamentals of Robust Statistical Methods

Robustness in statistics generally refers to the resilience of statistical methods to deviations from assumptions, such as distributional assumptions, the presence of outliers, or other anomalies in the data. Consider a statistic T_n that is regarded as functional $T(\cdot)$ evaluated at the empirical distribution F_n . The empirical distribution assigns mass $1/n$ to each sample point $\mathbf{X}_i, i = 1, \dots, n$. In the following we describe three most popular properties or robustness measures of functional $T(F)$ or statistic $T(F_n)$ and efficiency of an estimator.

2.1 Influence Function

The influence function of robust estimators plays a crucial role in understanding the impact of individual data points on the estimation process. The influence function of an estimator at a given probability distribution, as introduced by [Hampel \(1968, 1974\)](#), is essentially the Gateaux derivative of the estimator, viewed as a functional, at some distribution. It quantifies how infinitesimally small changes in the data points impact the estimator, thus helping to assess the estimator's sensitivity to local perturbations in the data. It measures the effect of a small fraction of outliers placed at a given point. This measure is pivotal in studying the local robustness properties of statistical estimators, as it allows for evaluating how the estimator reacts to slight deviations in the distribution from which we sampled data.

Consider the estimator T at a distribution F in a point \mathbf{x} , then the influence function is given by

$$\text{IF}(\mathbf{x}; T, F) = \lim_{\varepsilon \rightarrow 0} \frac{T[(1 - \varepsilon)F + \varepsilon\delta_{\mathbf{x}}] - T(F)}{\varepsilon}. \quad (2.1)$$

Here, $\delta_{\mathbf{x}}$ represents the probability measure determined by the point mass 1 at a given point $\mathbf{x} \in \mathbb{R}^p$ and $T(F)$ is the parameter value, where T maps any distribution F to \mathbb{R}^p for multivariate location, and to $\mathbb{R}^{p \times p}$ for multivariate scatter, which represents the set of all positive semidefinite $p \times p$ matrices. Because it considers only point-mass contamination, the influence function assesses the local robustness of the functional $T(\cdot)$.

[Hampel \(1974\)](#) utilized the influence function to analyze various estimators, including trimmed and Winsorized means, Huber-estimators, and maximum likelihood and M-estimators. [Shoemaker \(1984\)](#) calculated the influence curve for a class of scale estimators, specifically the mid-variance, and used it to approximate the distribution of the sample mid-variance. [Peracchi \(1990\)](#) summarized the influence function approach to robust estimation of parametric models, extending Hampel’s optimality results for M-estimators. For detail analysis, refer to [Hampel et al. \(1986\)](#). These studies collectively highlight the importance of the influence curve in assessing the robustness of estimators.

2.2 Breakdown Point

Among the various metrics for robustness, the breakdown point has emerged as a pivotal global measure. The seminal works by [Hodges Jr \(1967\)](#), which introduced the concept of tolerance to extreme values, paved the way for the development of the breakdown point. This critical notion was further elaborated upon and introduced into the statistical literature by [Hampel \(1968, 1971\)](#) and [Ronchetti and Huber \(2009\)](#). The breakdown point, introduced by [Donoho and Huber \(1983\)](#), is a crucial measure of robustness in statistical estimation. Generally, it represents the maximum proportion of contaminated data points an estimator can withstand while providing meaningful results about the uncontaminated

data. Specifically, it quantifies the maximum proportion of outliers or contaminated data that an estimator can handle before it starts producing incorrect or infinitely biased results.

Consider the mixture distribution $K = (1 - \varepsilon)F + \varepsilon H$, where H represents any arbitrary distribution. The asymptotic breakdown point for the location is defined by the following equation:

$$\varepsilon^*(T, K) = \inf_{\varepsilon \in (0,1)} \left\{ \varepsilon > 0 : \sup_H \|T(K)\| = \infty \right\}. \quad (2.2)$$

This formulation captures the smallest proportion of contamination ε that leads the estimate of the mean to become unbounded, thereby assessing the robustness of the mean estimator against outliers.

For a positive definite semi-definite scatter matrix, whose eigenvalues are ordered as $\lambda_1 \geq \dots \geq \lambda_p \geq 0$, the implosion breakdown point of a scatter functional S at distribution F is specified by:

$$\varepsilon_{\text{imp}}^*(S, F) = \inf_{\varepsilon \in (0,1)} \left\{ \varepsilon > 0 : \inf_H \{\lambda_p(S(F))\} = 0 \right\}. \quad (2.3)$$

This value indicates the minimum proportion of contamination that causes the smallest eigenvalue of the scatter matrix to become zero, which can be interpreted as a collapse in the dimensionality of variability.

Furthermore, the explosion breakdown point of a scatter functional S is described as:

$$\varepsilon_{\text{exp}}^*(S, F) = \inf_{\varepsilon \in (0,1)} \left\{ \varepsilon > 0 : \sup_H \{\lambda_1(S(F))\} = \infty \right\}. \quad (2.4)$$

This parameter delineates the maximum proportion of contamination ε that results in the largest eigenvalue of the scatter matrix reaching infinity, signifying extreme sensitivity to outliers affecting the spread or variability of the data.

The asymptotic breakdown point then equals the smallest of the two ([Becker et al., 2014](#))

$$\varepsilon^*(S, F) = \min \left(\varepsilon_{\text{exp}}^*(S, F), \varepsilon_{\text{imp}}^*(S, F) \right). \quad (2.5)$$

A simple and appealing finite sample version of this concept was introduced by [Donoho and Huber \(1983\)](#). It is defined as the minimum proportion of observations that must be replaced with arbitrary values to render the estimate ineffective.

Consider a data set matrix $\mathbf{X}_{n \times p}$, the breakdown point of a location estimator T_n is defined as:

$$\varepsilon_n^*(T_n; \mathbf{X}_n) = \min \left\{ \frac{m}{n} : \sup ||T_n(\mathbf{X}_{n,m}) - T_n(\mathbf{X}_n)|| = \infty \right\}, \quad (2.6)$$

where $1 \leq m \leq n$ and the supremum is over all datasets $\mathbf{X}_{n,m}$ obtained by replacing any m data points of \mathbf{X}_n by arbitrary values. For a multivariate scatter estimator \mathbf{S}_n , we have

$$\varepsilon_n^*(\mathbf{S}_n; \mathbf{X}_n) = \min \left\{ \frac{m}{n} : \sup \max_i |\log(\lambda_i(\mathbf{S}_n(\mathbf{X}_{n,m}))) - \log(\lambda_i(\mathbf{S}_n(\mathbf{X}_n)))| = \infty \right\}, \quad (2.7)$$

where the eigenvalues λ_i are sorted $0 \leq \lambda_p(S_n) \leq \dots \leq \lambda_1(S_n)$. With this definition, we consider the scatter estimator to break down when λ_1 become arbitrary large (“explosion”) and/or λ_p become arbitrary close to 0 (“implosion”). Implosion is a problem since it makes the scatter matrix singular especially in many cases where its inverse is required.

The concept of breakdown point, a measure of robustness in statistical estimation, has been explored in various contexts. [Stromberg and Ruppert \(1992\)](#) introduced a new definition for the breakdown point in nonlinear regression, in time series (see [Ma and Genton, 2000](#)), and in more general situations (see [He and Simpson, 1993](#); [Genton, 2003](#)).

2.3 Affine Equivariance

Affine equivariance in robust estimation refers to the property where the estimation process remains consistent even when the data undergoes affine transformations. This concept is crucial in scenarios where data may be corrupted or altered. Affine equivariance implies that the estimator transforms appropriately under any non-singular reparametrization of the space of the \mathbf{X}_i . The data might, for instance, be rotated, translated, or rescaled (for example through a change of the measurement units). For any non-singular matrix

$\mathbf{A} \in \mathbb{R}^{p \times p}$ and any p -dimensional column vector $\mathbf{b} \in \mathbb{R}^p$, we refer to p -variate location and scatter matrix estimators $\mathbf{T}_n : \mathbb{R}^{n \times p} \rightarrow \mathbb{R}^p$ and $\mathbf{C}_n : \mathbb{R}^{n \times p} \rightarrow \mathbb{R}^{p \times p}$ as affine equivariant if

$$\mathbf{T}_n(\mathbf{X}\mathbf{A}^T + \mathbb{1}_n\mathbf{b}^T) = \mathbf{A}\mathbf{T}_n(\mathbf{X}) + \mathbf{b}, \quad (2.8)$$

$$\mathbf{C}_n(\mathbf{X}\mathbf{A}^T + \mathbb{1}_n\mathbf{b}^T) = \mathbf{A}\mathbf{C}_n(\mathbf{X})\mathbf{A}^T. \quad (2.9)$$

The concept of affine equivariance in robust estimation is explored in several studies. [Yu et al. \(2015\)](#) investigates the robustness of the affine equivariant scatter estimator, particularly in terms of breakdown point and influence function. [Ostrovskii and Rudi \(2019\)](#) propose an estimator for the covariance matrix of heavy-tailed distributions, demonstrating its affine equivariance properties. [Mendelson and Zhivotovskiy \(2020\)](#) present a robust covariance estimator under L_4 - L_2 norm equivalence, achieving optimal performance for Gaussian vectors. [Fan et al. \(2019\)](#) focus on robust covariance estimation in approximate factor models, introducing a framework for joint covariance matrix estimation and recovery. While this condition appears to be quite intuitive, it has been shown that several prominent multivariate location estimators do not meet this criterion, see [Lopuhaä and Rousseeuw \(1991\)](#). It might be feasible to identify estimators with a high breakdown point if one considers weakening the affine equivariance requirement to, for example, rigid-motion equivariance or merely location equivariance, see [Donoho and Gasko \(1992\)](#).

For multivariate estimators, being consistent under affine transformations (equivariance) is expected. However, it is challenging to ensure this consistency while also maintaining a high level of robustness against outliers (high breakdown point). The Stahel-Donoho estimator represents one of the earliest significant advancements in the field of multivariate statistics, combining affine equivariance with a high breakdown point. This robust estimator was independently proposed by [Stahel \(1981\)](#) and [Donoho \(1982\)](#), marking a pivotal moment in the recognition and development of robust statistical methods capable of maintaining reliability under data contamination and also versatile, maintaining consistency under various affine transformations. This dual capability ensures that the estimator is effective across different scales and orientations, thereby handling real-world data complexities with

resilience against anomalies and outliers.

Further contributions include the affine equivariant maximum depth estimator introduced by [Zuo and Lai \(2009\)](#), which exhibits both a high breakdown point and efficiency. Transformation retransformation medians, with an appropriate choice of the transformation matrix, can achieve a high breakdown point while maintaining affine equivariance ([Chakraborty and Chaudhuri, 1999](#)). Minimum volume ellipsoid estimator is another affine equivariant estimator established by [Davies \(1992\)](#). Similarly, minimum covariance determinant (MCD) of [Rousseeuw \(1984\)](#) combines these desirable properties. [Lopuhaä and Rousseeuw \(1991\)](#) provided a comprehensive overview of the breakdown points of various affine equivariant estimators for multivariate location and covariance matrices, illuminating the extensive research in this area. However, caution is necessary when using high-breakdown estimators as they can behave unpredictably in certain situations [Jurczyk, 2008](#). Another approach involves constructing projection-based affine equivariant multivariate estimators, which offer more resistance to contamination compared to existing estimators [Zuo, 2004](#). For recent discussions about the breakdown point and equivariance, see [Davies and Gather \(2005\)](#).

2.4 Efficiency

Efficiency is a fundamental concept in statistics that measures the quality of an estimator. It evaluates how well an estimator performs in terms of its covariance, relative to the best possible unbiased estimator.

The efficiency of an estimator T of a parameter vector $\boldsymbol{\theta}$ is defined by the ratio of the Cramér-Rao lower bound to the actual covariance matrix of the estimator:

$$e(T) = \det(\mathcal{I}^{-1}(\boldsymbol{\theta})) \cdot \det(\text{cov}(T)),$$

where $\mathcal{I}(\boldsymbol{\theta})$ represents the Fisher information matrix. This efficiency measure indicates how closely the estimator's covariance matrix approaches the theoretical lower bound provided by the Fisher information.

Relative efficiency compares the performance of two estimators, expressed in terms of the determinants of their covariance matrices, that is, their generalized variance:

$$e(T_1, T_2) = \frac{\det(\text{cov}(T_2))}{\det(\text{cov}(T_1))}.$$

If the relative efficiency is greater than 1, estimator T_1 is more efficient than T_2 . This comparison provides a direct way to evaluate which estimator has a smaller overall variability.

For robust statistics, the challenge is to design estimators that maintain high efficiency at the model distribution while improving resistance to outliers and model deviations. Robust estimators aim to balance robustness and efficiency, ensuring that they are not unduly influenced by atypical data points while still providing accurate parameter estimates. In their discussion on robust estimation, [Ronchetti and Huber \(2009\)](#) emphasize the importance of this balance, highlighting that robust estimators can be designed to have high efficiency at the model distribution while still providing protection against outliers.

2.5 Computational Complexity

Robust statistical methods are crucial for analyzing data with outliers or deviations from assumed distributional forms. However, these methods often entail increased computational complexity, which can affect their practical application. This section discusses the computational challenges inherent in robust statistical methods, particularly concerning the breakdown point, the complexity of the optimization problems they involve, and advances to mitigate these issues. By understanding the computational complexities associated with different robust methods and the complexity of the optimization problems they involve, researchers can effectively balance robustness and computational efficiency in their analyses. These strategies and advances enable the practical application of robust methods to large and complex datasets, ensuring that the benefits of robustness are not overshadowed by prohibitive computational costs.

Computational Complexity and Breakdown Point

The breakdown point is a critical measure of an estimator’s robustness, representing the smallest fraction of contamination that can cause the estimator to yield arbitrarily large errors. Robust methods often achieve high breakdown points, but this robustness comes at a computational cost.

For instance, the least trimmed squares (LTS) estimator is known for its high breakdown point. However, achieving this robustness involves solving a combinatorial optimization problem where the subset of data that minimizes the sum of squared residuals must be selected. This process becomes computationally infeasible for large datasets due to the exponential growth of possible subsets ([Rousseeuw and Leroy, 2005](#)).

Similarly, the minimum covariance determinant (MCD) estimator, designed to identify the subset of observations with the smallest determinant of the covariance matrix, involves solving an NP-hard problem. The computational intensity of finding the exact MCD grows exponentially with the dataset size, making it impractical for large datasets without approximation or heuristic methods. The development of the FASTMCD algorithm by [Rousseeuw and Driessen \(1999\)](#) and the more recent DetMCD have made MCD ([Hubert et al., 2012](#)) practical for larger datasets by significantly reducing computation time while preserving robust properties.

Complexity of Optimization Problems

Robust methods often involve solving highly nonlinear optimization problems. These problems are computationally intensive due to their iterative nature and the need for robust initial estimates. Many robust estimators, such as M-estimators, rely on iterative optimization procedures. These procedures can be computationally demanding, especially for large datasets, as they require recalculating weights and re-estimating parameters at each iteration. The iterative reweighted least squares (IRLS) algorithm is a prime example, where robustness is achieved through iterative adjustments ([Maronna et al., 2019](#)). However, the

iterative nature can significantly increase computational costs. Additionally, the choice of initial estimates can greatly influence the convergence and computational efficiency of robust methods. Starting from robust initial estimates, such as the sample median, can help ensure that the procedure inherits a high breakdown point and improves convergence ([Huber and Ronchetti, 2011](#)). However, identifying appropriate initial estimates itself can be a computational challenge.

Advances to Mitigate Computational Challenges

To address the computational challenges associated with robust statistical methods, several advances have been made:

First, approximation algorithms offer a balance between computational efficiency and robustness. Approximate versions of the LTS estimator have been developed to reduce computational costs while maintaining reasonable robustness properties ([Croux and Haesbroeck, 2003](#)). These algorithms often use heuristic or probabilistic approaches to find near-optimal solutions without exhaustive searches.

Moreover, the development of algorithms like FASTMCD and DetMCD has significantly reduced the computation time for MCD estimators. FASTMCD, introduced by [Rousseeuw and Driessen \(1999\)](#), uses concentration steps to approximate the MCD. These steps involve iteratively refining subsets of the data to find the one with the smallest determinant of the covariance matrix, making the algorithm feasible for larger datasets. DetMCD by [Hubert et al. \(2012\)](#) further enhances this by providing deterministic algorithms that improve both speed and accuracy.

Additionally, leveraging advanced computational techniques, such as parallel computing and specialized hardware (e.g., GPUs), can significantly enhance the performance of robust algorithms. Parallelizing the computations involved in iterative procedures or employing distributed computing frameworks can reduce the time required for robust estimation ([Buchanan and Fitzgibbon, 2005](#)).

Furthermore, careful consideration of the specific problem and dataset characteristics can guide the choice of robust methods. Selecting algorithms with lower computational complexity when feasible can help balance robustness and efficiency. For example, when dealing with moderately sized datasets, simpler robust methods such as Huber’s M-estimator may be preferred over more complex alternatives like the LTS or MCD estimators ([Huber and Ronchetti, 2011](#)).

Chapter 3

Review of Robust Statistical Estimators

In this chapter, we provide an in-depth examination of various statistical methods used to estimate multivariate location and scatter parameters. The focus is on both classical and robust estimators, detailing their mathematical foundations, properties, and practical implications.

3.1 Classical Estimators of Multivariate Location and Scatter

In multivariate analysis, one of the fundamental goal is to estimate the parameters $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ from a sample $\mathbf{X}_n = (\mathbf{x}_1, \dots, \mathbf{x}_n)^T$. The maximum likelihood estimates of $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ under the assumption of multivariate normality are the sample mean vector and the sample scatter/covariance matrix:

$$\bar{\mathbf{x}} = \frac{1}{n} \sum_{i=1}^n \mathbf{x}_i \quad \text{and} \quad \mathbf{S} = \frac{1}{n} \sum_{i=1}^n (\mathbf{x}_i - \bar{\mathbf{x}})(\mathbf{x}_i - \bar{\mathbf{x}})^T,$$

where \mathbf{x}_i represents the i -th observation of a n -sample dataset, with each observation being a p -dimensional vector, \mathbf{S} is a $p \times p$ symmetric matrix with diagonal elements representing the variances of individual variables. A crucial geometric property of these estimators is their affine equivariance, ensuring that they adapt appropriately under any affine transformation of the data. This property is essential for their applicability across different

measurement scales and units as well as correlations.

Classical estimators of multivariate location and scatter are unbiased and efficient under the assumption of multivariate normality. However, they lack robustness, making them highly sensitive to outliers. This sensitivity can lead to significant distortions in their estimation, rendering them unreliable when datasets contain significant anomalies.

- (1) The influence function of classical estimators of multivariate location and scatter is unbounded (Hampel et al., 1986). Let $T(F)$ denote a statistical functional that maps a distribution F to its location vector, denoted as $\boldsymbol{\mu}_F$. Similarly, let $S(F)$ be the statistical functional that maps F to its scatter matrix, denoted as $\boldsymbol{\Sigma}_F$. For any vector $\mathbf{x} \in \mathbb{R}^p$, the influence functions of the location vector and the scatter matrix can be expressed as follows:

$$\begin{aligned}\text{IF}(\mathbf{x}; T, F) &= \mathbf{x} - \boldsymbol{\mu}_F, \\ \text{IF}(\mathbf{x}; S, F) &= (\mathbf{x} - \boldsymbol{\mu}_F)(\mathbf{x} - \boldsymbol{\mu}_F)^T - \boldsymbol{\Sigma}_F.\end{aligned}$$

The influence function of the location vector represents the deviation of the point \mathbf{x} from the mean $\boldsymbol{\mu}_F$. The influence function of the scatter matrix calculates the outer product of the deviation $(\mathbf{x} - \boldsymbol{\mu}_F)$ with itself, adjusted by subtracting the covariance matrix $\boldsymbol{\Sigma}_F$. This provides a measure of how much \mathbf{x} influences the estimation of the covariance matrix relative to its current estimate. Clearly, both influence functions are unbounded in \mathbf{x} .

- (2) The asymptotic breakdown point of the classical location vector and scatter matrix is zero (Huber and Ronchetti (2011)). Refer to Section 2.2 for extensive discussion.
- (3) For finite-sample breakdown point estimators of location and scatter, Lopuhaä and Rousseeuw (1991) proved that any affine equivariance location estimator $\hat{\boldsymbol{\mu}}$ satisfies

$$\varepsilon_n^*(\hat{\boldsymbol{\mu}}, \mathbf{X}_n) \leq \frac{1}{n} \left\lfloor \frac{n+1}{2} \right\rfloor,$$

and any affine equivariant scatter estimator $\hat{\Sigma}$ satisfies the sharp bound (Davies, 1987)

$$\varepsilon_n^*(\hat{\Sigma}, \mathbf{X}_n) \leq \frac{1}{n} \left\lceil \frac{n - p + 1}{2} \right\rceil,$$

if the original sample before contamination is in general position, that is no hyperplane contains more than p points. For samples that are not in general position, the upper bound is lower and depends on the maximal number of observations on a single hyperplane. However, the asymptotic breakdown point for affine equivariant for location and scatter estimators is always at most 0.5.

3.2 Robust Estimators of Multivariate Location and Scatter

When it comes to statistical analysis, the validity and reliability of analytical results are primarily dependent on the robustness of estimators, especially in the presence of outliers or deviations from assumed distributions. These estimators are crucial in environmental science, economics, and biological research, where anomalies can compromise data integrity. Robust estimators are particularly interesting because they yield reliable and accurate findings despite outliers. The following discussion explores important robust estimators of multivariate location and scatter, highlighting their value and adaptability in handling complex and contaminated data. It lays the groundwork for a thorough comparison with the minimum density power divergence (MDPD) approaches covered later in the dissertation.

3.2.1 Multivariate M-Estimators

The general definition of multivariate M-estimators for location and scatter parameters, denoted as $\hat{\mu}$ and $\hat{\Sigma}$, is given by the following equations, as defined by Maronna (1976):

$$\sum_{i=1}^n W_1(d_i^2)(\mathbf{x}_i - \hat{\boldsymbol{\mu}}) = \mathbf{0}, \quad (3.1)$$

$$\frac{1}{n} \sum_{i=1}^n W_2(d_i^2)(\mathbf{x}_i - \hat{\boldsymbol{\mu}})(\mathbf{x}_i - \hat{\boldsymbol{\mu}})^T = \hat{\boldsymbol{\Sigma}}, \quad (3.2)$$

in which the statistical distance $d(\mathbf{x}_i, \hat{\boldsymbol{\mu}}, \hat{\boldsymbol{\Sigma}})$ (Mahalanobis distance) is given by

$$d_i = d(\mathbf{x}_i, \hat{\boldsymbol{\mu}}, \hat{\boldsymbol{\Sigma}}) = \sqrt{(\mathbf{x}_i - \hat{\boldsymbol{\mu}})^T \hat{\boldsymbol{\Sigma}}^{-1} (\mathbf{x}_i - \hat{\boldsymbol{\mu}})}.$$

It is essential that $\hat{\boldsymbol{\mu}}$ is a real vector and $\hat{\boldsymbol{\Sigma}}$ a symmetric positive definite matrix. The functions $W_1(t)$ and $W_2(t)$ are real-valued and well-defined for all $t \geq 0$.

For $\mathbf{X} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, $W_1(d^2) = W_2(d^2) = 1$, which corresponds to the sample mean and the sample covariance matrix.

[Maronna \(1976\)](#) stipulates necessary conditions for functions $W_1(t)$ and $W_2(t)$ to ensure the existence, uniqueness, and consistency of these estimators. The sufficient conditions are that $\sqrt{t}W_1(t)$ and $tW_2(t)$ are bounded and that $tW_2(t)$ is nondecreasing. If a multivariate M-estimator satisfies the condition that $tW_2(t)$ is nondecreasing, it is called monotone, otherwise, it is called redescending.

M-estimators are also characterized by several key properties:

- (1) Multivariate M-estimators are affine equivariant.
- (2) Multivariate M-estimators are asymptotically normal, assuming regularity conditions on W_1 and W_2 (see [Ronchetti and Huber, 2009](#)).
- (3) Monotone M-estimators for Equations (3.1) and (3.2) always have a unique solution, established through a reweighting algorithm. This algorithm includes:
 - (a) Begin with initial estimators $\boldsymbol{\mu}_0$ and $\boldsymbol{\Sigma}_0$, for example the vector coordinate-wise median and a diagonal matrix where the diagonal elements are the squared normalized Median Absolute Deviations (MADs) of the variables.

(b) At iteration k , let $d_{i,k} = d(\mathbf{x}_i, \widehat{\boldsymbol{\mu}}_k, \widehat{\boldsymbol{\Sigma}}_k)$ and compute

$$\begin{aligned}\widehat{\boldsymbol{\mu}}_{k+1} &= \frac{\sum_{i=1}^n W_1(d_{i,k}^2) \mathbf{x}_i}{\sum_{i=1}^n W_1(d_{i,k}^2)}, \\ \widehat{\boldsymbol{\Sigma}}_{k+1} &= \frac{1}{n} \sum_{i=1}^n W_2(d_{i,k}^2) (\mathbf{x}_i - \boldsymbol{\mu}_{k+1})(\mathbf{x}_i - \boldsymbol{\mu}_{k+1})^T.\end{aligned}$$

The algorithm assures convergence to the unique solution for monotone M-estimators, independent of initial values. The choice of initial values may only influence the number of iterations and not the end value (Becker et al., 2014). However, for redescending M-estimators, initial values significantly influence the quality of the solution.

- (4) Their influence functions are bounded if $\sqrt{t}W_1(t)$ and $tW_2(t)$ are bounded.
- (5) The asymptotic breakdown point of monotone M-estimators is given by $\varepsilon^* \leq \frac{1}{p+1}$, where p is the number of variables.
- (6) Redescending M-estimators can offer higher breakdown points but are challenging to compute efficiently.

However, while monotone M-estimators achieve optimal breakdown points in univariate cases, they exhibit lower breakdown points in higher dimensions often not exceeding $1/(p+1)$ where p is the number of variables (see Clement, 2020), highlighting a key limitation in their robustness for multivariate applications. This issue arises due to contamination restricted to specific planes, leading to low breakdown points. However, in the absence of such contamination restricted to specific planes, M-estimators can achieve breakdown points close to $1/2$, indicating higher robustness (Sunanta, 2018). For further review, refer to Maronna et al. (2019)

3.2.2 Minimum Covariance Determinant (MCD)

The Minimum Covariance Determinant (MCD) method of Rousseeuw (1984, 1985) is highly robust and affine equivariant estimator of location and scatter. The MCD objective is to

find h observations (out of n) whose classical covariance matrix has the lowest determinant. The MCD estimate of location is then the average of these h points, and the MCD estimate of scatter is their covariance matrix multiplied with a correction/scale factor to obtain consistency in the case of multivariate normality.

That is, given an $n \times p$ data matrix $\mathbf{X} = (\mathbf{x}_1, \dots, \mathbf{x}_n)^T$ with n observations and p -variate vectors $\mathbf{x}_i = (x_{i1}, \dots, x_{ip})^T$, and a fixed tuning constant h such that $[(n+p+1)/2] \leq h \leq n$, the raw MCD estimators for location and scatter respectively are defined as follows:

- (1) the mean of the h observations for which covariance matrix has the lowest determinant

$$\hat{\boldsymbol{\mu}}_{\text{MCD}} = \frac{1}{h} \sum_{i \in \mathcal{I}_{\text{MCD}}} \mathbf{x}_i. \quad (3.3)$$

- (2) the corresponding covariance matrix multiply by a correction factor k_{MCD}

$$\hat{\boldsymbol{\Sigma}}_{\text{MCD}} = \frac{k_{\text{MCD}}}{h-1} \sum_{i \in \mathcal{I}_{\text{MCD}}} (\mathbf{x}_i - \hat{\boldsymbol{\mu}}_{\text{MCD}})(\mathbf{x}_i - \hat{\boldsymbol{\mu}}_{\text{MCD}})^T, \quad (3.4)$$

where the “best set” is given by

$$\mathcal{I}_{\text{MCD}} = \arg \min_{\substack{\mathcal{I} \in \{1, \dots, n\} \\ |\mathcal{I}|=h}} \left\{ \det \left(\frac{1}{h-1} \sum_{i \in \mathcal{I}} \left(\mathbf{x}_i - \frac{1}{h} \sum_{j \in \mathcal{I}} \mathbf{x}_j \right) \left(\mathbf{x}_i - \frac{1}{h} \sum_{j \in \mathcal{I}} \mathbf{x}_j \right)^T \right) \right\}. \quad (3.5)$$

The correction factor k_{MCD} that makes the MCD scatter estimate consistent at the normal model is given by $k_{\text{MCD}} = \alpha / F_{\chi_{p+2}^2}(q_\alpha)$ with $q_\alpha = \chi_{p,\alpha}^2$ and $\alpha = \lim_{n \rightarrow \infty} h(n)/n$ (Croux and Haesbroeck, 1999). A finite-sample correction or consistency factor was also proposed by Pison et al. (2002) for small sample sizes.

The MCD estimators are characterized by bounded influence functions, which exhibit discontinuities, as detailed by (Croux and Haesbroeck, 1999). These estimators achieve \sqrt{n} -consistency and while asymptotic normality has been established for the location estimates, it remains unconfirmed for the scatter estimates (Butler et al., 1993). The MCD estimator is most robust when $h = [(n+p+1)/2]$, asymptotically attaining the highest breakdown

point of 50% under appropriate assumptions but then MCD unfortunately suffers from low efficiency at the normal model. This is especially true at the h selected in order for the estimators to achieve the highest breakdown point (see [Lopuhaä and Rousseeuw, 1991](#); [Roelant et al., 2009](#); [Croux and Haesbroeck, 1999](#)). To increase the efficiency while maintaining high robustness, [Lopuhaä and Rousseeuw \(1991\)](#) proposed a reweighting procedure which was analyzed by [Lopuhaä \(1999\)](#). The reweighted RMCD estimators are given by

$$\hat{\boldsymbol{\mu}}_{\text{RMCD}} = \frac{\sum_{i=1}^n W(d_i^2) \mathbf{x}_i}{\sum_{i=1}^n W(d_i^2)}, \quad (3.6)$$

$$\hat{\boldsymbol{\Sigma}}_{\text{RMCD}} = \frac{k_{\text{RMCD}}}{n-1} \sum_{i=1}^n W(d_i^2) (\mathbf{x}_i - \hat{\boldsymbol{\mu}}_{\text{RMCD}}) (\mathbf{x}_i - \hat{\boldsymbol{\mu}}_{\text{RMCD}})^T, \quad (3.7)$$

where $d_i^2 = D^2(\mathbf{x}_i, \hat{\boldsymbol{\mu}}_{\text{MCD}}, \hat{\boldsymbol{\Sigma}}_{\text{MCD}}) = (\mathbf{x}_i - \hat{\boldsymbol{\mu}}_{\text{MCD}})^T \hat{\boldsymbol{\Sigma}}_{\text{MCD}}^{-1} (\mathbf{x}_i - \hat{\boldsymbol{\mu}}_{\text{MCD}})$ is the raw squared Mahalanobis distance base on the MCD estimates and W an appropriate weight function.

While the method is highly effective for detecting outliers, its primary limitation is inapplicability when the number of dimensions exceeds the size of the subset. To mitigate this issue, the minimum regularized covariance determinant (MRCD) approach was developed, incorporating a regularization term to effectively manage high-dimensional data ([Boudt et al., 2020](#)). The MRCD method introduces a regularization technique where the scatter matrix is a convex combination of a target matrix and the sample covariance matrix of the subset, allowing for estimation in any dimension and enhancing robustness. The method's robustness is further enhanced by the minimum weighted covariance determinant estimator, which uses implicit weights and has been shown to outperform other estimators ([Kalina, 2022](#)). The MCD's efficiency and robustness are underscored by its influence function and asymptotic variances ([Croux and Haesbroeck, 1999](#)). The existence and non-singularity of the derivative for the MCD estimator was proven, and its asymptotic normality and limiting covariance structure were described in [Cator and Lopuhaä \(2010\)](#).

The exact MCD estimator is very challenging to compute because it necessitates evaluation of all possible combinations of n observations taken h at a time. This computation grows

exponentially with the size of the dataset, making it computationally intensive, especially for large datasets (see [Hubert and Debruyne, 2010](#); [Pokojoy and Jobe, 2022](#)). To address this challenge, several fast computing algorithms have been developed. Notable among these are FastMCD of [Rousseeuw and Driessen \(1999\)](#), DetMCD of [Hubert et al. \(2012\)](#), and Projection Pursuit MCD of [Pokojoy and Jobe \(2022\)](#). These algorithms enable more accurate approximations of the MCD estimator in high-dimensional spaces, contributing to its growing popularity in academic literature.

FastMCD

FastMCD is an efficient affine equivariance algorithm proposed by [Rousseeuw and Driessen \(1999\)](#). It uses the basic idea of sorting by squared Mahalanobis distances, determinants and selective iteration or nested extensions call the concentration steps to find h observations (out of n) whose covariance matrix has the lowest determinant. For small datasets, FastMCD can typically finds the exact MCD, whereas for larger datasets it gives more accurate approximate results.

The FastMCD algorithm is essential for robust multivariate data analysis, particularly in its execution of the concentration step (C-step). This step operates on the premise of refining estimates for the location and scatter matrix of a data set. Initially, these estimates are provided as $\hat{\boldsymbol{\mu}}_{\text{old}}$ for the location and $\hat{\boldsymbol{\Sigma}}_{\text{old}}$ for the scatter matrix. The process follows from the C-step algorithm below:

- (1) **Distance Computation:** Calculate the Mahalanobis distance for each data point i from $\hat{\boldsymbol{\mu}}_{\text{old}}$ and $\hat{\boldsymbol{\Sigma}}_{\text{old}}$ using the equation

$$d_{\text{old}}(i) = D(\mathbf{x}_i, \hat{\boldsymbol{\mu}}_{\text{old}}, \hat{\boldsymbol{\Sigma}}_{\text{old}})$$

for $i = 1, 2, \dots, n$ where n is the numbers of observations.

- (2) **Sorting and Subset Selection:** Sort the distances, $d_{\text{old}}(i)$, and permute them in ascending order to identify a subset H of h observations with the smallest distances,

that is, for the permutation π , where $d_{\text{old}}(\pi(1)) \leq d_{\text{old}}(\pi(2)) \leq \dots d_{\text{old}}(\pi(n))$, set $H = \{\pi(1), \pi(2), \dots, \pi(h)\}$.

- (3) **Estimation Update:** Update the estimates for the mean and covariance by computing:

$$\begin{aligned}\hat{\boldsymbol{\mu}}_{\text{new}} &= \sum_{i \in H} \frac{\mathbf{x}_i}{h}, \\ \hat{\boldsymbol{\Sigma}}_{\text{new}} &= \sum_{i \in H} \frac{(\mathbf{x}_i - \hat{\boldsymbol{\mu}}_{\text{new}})(\mathbf{x}_i - \hat{\boldsymbol{\mu}}_{\text{new}})^T}{h-1}.\end{aligned}$$

Rousseeuw and Driessen (1999) established that $\det(\hat{\boldsymbol{\Sigma}}_{\text{new}}) \leq \det(\hat{\boldsymbol{\Sigma}}_{\text{old}})$, with equality only if $\hat{\boldsymbol{\Sigma}}_{\text{new}} = \hat{\boldsymbol{\Sigma}}_{\text{old}}$, that is determinant of the updated scatter matrix $\hat{\boldsymbol{\Sigma}}_{\text{new}}$ does not exceed that of $\hat{\boldsymbol{\Sigma}}_{\text{old}}$, indicating convergence towards a limit as the process iterates. To ensure convergence, C-steps is applied iteratively, resulting in a sequence that converges within a finite number of steps due to the finite number of h -subsets. However, this method does not guarantee that the final iteration yields the global minimum of the MCD objective function. Therefore, to obtain an approximate MCD solution, they start with numerous (typically 500) initial h -subsets $H_1 \subset \{1, 2, \dots, n\}$, apply C-steps to each, and select the solution with the lowest overall determinant.

To construct an initial subset H_1 , a random $(p+1)$ -subset J is drawn and $\hat{\boldsymbol{\mu}} = \sum_{i \in J} \mathbf{x}_i / (p+1)$ and $\hat{\boldsymbol{\Sigma}} = \sum_{i \in J} (\mathbf{x}_i - \hat{\boldsymbol{\mu}})(\mathbf{x}_i - \hat{\boldsymbol{\mu}})^T / p$ are computed. If $\hat{\boldsymbol{\Sigma}}$ is singular, random points are added to J until it becomes nonsingular. Next C-step is applied to $(\hat{\boldsymbol{\mu}}, \hat{\boldsymbol{\Sigma}})$ yielding $(\hat{\boldsymbol{\mu}}_1, \hat{\boldsymbol{\Sigma}}_1)$, and so on. Each iteration involves recalculating the covariance matrix and its inverse, which can be computationally intensive. To manage this, the FASTMCD algorithm restricts the full application of C-steps to the 10 subsets that exhibit the lowest determinant values in their corresponding covariance matrices, thus balancing computational efficiency with statistical robustness. Only for the 10 subsets with lowest determinant further C-steps are taken until convergence. Upon convergence, the estimates from the FASTMCD algorithm, denoted as $\hat{\boldsymbol{\mu}}_{\text{Raw}}$ and $\hat{\boldsymbol{\Sigma}}_{\text{Raw}}$, align closely with the empirical mean and covariance

matrix of the h -subset yielding the lowest determinant. Further refinement uses these as base estimates to calculate $\hat{\boldsymbol{\mu}}_{\text{FastMCD}}$ and $\hat{\boldsymbol{\Sigma}}_{\text{FastMCD}}$, incorporating a weighted average approach where weights are determined based on the χ^2 distribution's quantile, enhancing the statistical efficiency of the final estimates while maintaining high robustness. This procedure is very efficient for small sample sizes n , but as n increases, the computation time escalates significantly. For large n , FastMCD employs partitioning strategy, which mitigates the computational burden by avoiding the need to perform all calculations on the entire dataset.

The FastMCD algorithm has been incorporated into several statistical software packages, enhancing its accessibility and utility in robust statistical analysis. This wide-ranging implementation underscores the algorithm's importance and versatility in handling diverse data analysis scenarios.

Deterministic MCD (DetMCD) Algorithm

The DetMCD algorithm, proposed by [Hubert et al. \(2012\)](#), offers a deterministic approach to estimate multivariate location and scatter, designed to produce consistent results by eliminating the randomness typically associated with subset selection in classical MCD algorithms. Unlike FastMCD it is permutation invariant, less sensitive to point contamination and runs even faster than FastMCD.

The DetMCD computes a small number of deterministic initial estimates, followed by concentration steps. The procedure begins by standardizing each variable \mathbf{X}_i by subtracting its median and dividing by the Q_n -scale estimator of [Rousseeuw and Croux \(1993\)](#), ensuring location and scale equivariance. i.e.,

$$\mathbf{Z}_j = \frac{\mathbf{X}_j - \text{median}(\mathbf{X}_j)}{Q_n(\mathbf{X}_j)}. \quad (3.8)$$

The resulting $n \times p$ matrix is denoted \mathbf{Z}_n and six initial estimates \mathbf{S}_k are developed for the correlation or covariance of \mathbf{Z} .

Computation of Initial Scatter Estimates

Several methods are employed to estimate the initial scatter:

- (1) \mathbf{S}_1 is based on Spearman's rank correlation of the transformed variables (the hyperbolic tangent of each column of \mathbf{Z} , a bounded function that reduces the effect of large coordinatewise outliers):

$$\mathbf{Y}_j = \tanh(\mathbf{Z}_j), \quad \mathbf{S}_1 = \text{corr}(\mathbf{Y}), \quad j = 1, \dots, p.$$

- (2) Let \mathbf{R}_j be the ranks of the column \mathbf{Z}_j and computes a rank covariance matrix: $\mathbf{S}_2 = \text{corr}(\mathbf{R})$.

- (3) \mathbf{S}_3 is computed from correlation of the normal scores of the ranks \mathbf{R}_j , that is

$$\mathbf{T}_j = \Phi^{-1} \left(\frac{\mathbf{R}_j - 1/3}{n + 1/3} \right), \quad \mathbf{S}_3 = \text{corr}(\mathbf{T}).$$

- (4) \mathbf{S}_4 is based on the spatial sign covariance matrix ([Visuri et al., 2000](#)). It is constructed as

$$\mathbf{K}_i = \frac{\mathbf{Z}_i}{\|\mathbf{Z}_i\|}, \quad \mathbf{S}_4 = \frac{1}{n} \sum_{i=1}^n \mathbf{K}_i \mathbf{K}_i'.$$

- (5) \mathbf{S}_5 represents the covariance matrix of the $[n/2]$ standardized observations \mathbf{z}_i with smallest norm, which corresponds to the first step of the BACON algorithm ([Billor et al., 2000](#)).

- (6) \mathbf{S}_6 estimate is the raw OGK estimator, where median and Q_n is used for location ($m(\cdot)$) and scale ($s(\cdot)$).

The \mathbf{S}_k of the covariance or correlation matrix of \mathbf{Z} may have inaccurate eigenvalues, so the following iterative procedure is used to refine these estimates:

- (1) Compute the eigenvectors \mathbf{E} of \mathbf{S}_k and transform the data:

$$\mathbf{V} = \mathbf{Z}\mathbf{E}.$$

(2) Estimate the scatter matrix by:

$$\hat{\Sigma}_k = \mathbf{E} \mathbf{L} \mathbf{E}^T, \quad \mathbf{L} = \text{diag}(\mathbf{Q}_2(\mathbf{V}_1), \dots, \mathbf{Q}_2(\mathbf{V}_p)).$$

(3) Update the center using the median of the transformed data:

$$\hat{\mu}_k = \hat{\Sigma}_k^{1/2} \times \text{median}(\mathbf{Z} \hat{\Sigma}_k^{-1/2}).$$

For each of the six estimates $(\hat{\mu}_k(\mathbf{Z}), \hat{\Sigma}_k(\mathbf{Z}))$, compute the statistical distances $d_{ik} = D(\mathbf{z}_i, \hat{\mu}_k(\mathbf{Z}), \hat{\Sigma}_k(\mathbf{Z}))$. For each initial estimate k , the $h_0 = \lfloor n/2 \rfloor$ observations with smallest d_{ik} are retained and the statistical distances (denoted as d_{ik}^*) based on these h_0 observations are computed. Following this, for each of the six estimates, the h observations \mathbf{x}_i that correspond to the smallest recalculated distances d_{ik}^* are selected. The algorithm then proceeds with iterative C-steps on these selected observations until convergence is achieved. The solution with smallest determinant is called the (raw) DetMCD. Then a weighting step can be applied as in the FAST-MCD algorithm, yielding the final DetMCD. Despite not achieving perfect affine equivariance, the deviation from affine equivariance in the DetMCD is minimal, making it an effective and practical tool for robust statistical analysis in multivariate contexts.

The Projection Pursuit Minimum Covariance Determinant

The projection pursuit minimum covariance determinant (PPMCD) algorithm introduced by Pokojovy and Jobe (2022), which was motivated by Peña and Prieto (2001), is the first deterministic, affine equivariant and permutation invariant algorithm of the MCD estimator of location and scatter matrix which in contrast to the earlier branch-and-bound type heuristics (Hawkins and Olive, 1999) is not restricted to small (n, p) values. It incorporates a projection pursuit approach and the concentration step (C-step) of Rousseeuw and Driessen (1999). Similar to DetMCD, PPMCD construct a deterministic initial estimate $(\hat{\mu}_0, \hat{\Sigma}_0)$, that assures permutation invariance and improve it through the C-step iteration. But unlike

DetMCD, PPMCD initial estimate is chosen so that it can preserve the affine equivariance of the MCD. The initial estimate $(\hat{\boldsymbol{\mu}}_0, \hat{\boldsymbol{\Sigma}}_0)$ is constructed through a projection pursuit procedure, that is, p orthogonal directions that are suitable for filtering out outliers are determined and univariate robust estimation is used to recover the projected locations and variances along the principal axes.

Computation of Initial Scatter Estimates

The initial estimator pair for the PPMCD algorithm is obtained with a type of non-linear principal component analysis (PCA) based on projection indices with outlier filtering:

1. The first projection index compares the ratio of sample variance to the raw minimum covariance determinant (MCD) scatter estimator, defined as:

$$Q_{\text{var}}(Z) = \log \left(\frac{s^2}{\hat{\sigma}_{\text{MCD}}^2} \right),$$

where $s^2 = \frac{1}{n-1} \sum_{i=1}^n (z_i - \bar{z})^2$ is the sample variance, and $\hat{\sigma}_{\text{MCD}}^2$ is the raw MCD scatter estimator in a univariate setting. Larger values of $Q_{\text{var}}(Z)$ indicate a significant difference between the sample variance and the MCD scatter estimator, highlighting directions where the projected data deviate more from the Gaussian distribution. This distinction helps in identifying “good” observations and outliers. However, the measure’s effectiveness diminishes when the variable Z contains a significant proportion of closely clustered outliers.

2. The second projection index is derived from a Gaussian bimodality test. That is

$$H_0 : z_i \sim \mathcal{N}(\mu_0, \sigma_0^2), \quad \text{vs.} \quad H_1 : z_i \sim (1 - \varepsilon)\mathcal{N}(\mu_1, \sigma_1^2) + \varepsilon\mathcal{N}(\mu_2, \sigma_2^2),$$

with $\varepsilon \in (0, 1)$ and $(\mu_1, \sigma_1^2) \neq (\mu_2, \sigma_2^2)$. The test statistic is:

$$Q_{\text{bimod}}(Z) = \frac{s^2}{(1 - \hat{\varepsilon}_{\text{ML}})\hat{\sigma}_{1,\text{ML}}^2 + \hat{\varepsilon}_{\text{ML}}\hat{\sigma}_{2,\text{ML}}^2 + \hat{\varepsilon}_{\text{ML}}(1 - \hat{\varepsilon}_{\text{ML}})(\hat{\mu}_{1,\text{ML}} - \hat{\mu}_{2,\text{ML}})^2},$$

where $\hat{\varepsilon}_{\text{ML}}, \hat{\mu}_{1,\text{ML}}, \hat{\mu}_{2,\text{ML}}, \hat{\sigma}_{1,\text{ML}}, \hat{\sigma}_{2,\text{ML}}$ are maximum likelihood estimators.

Larger $Q_{\text{bimod}}(Z)$ values favor the alternative hypothesis H_1 , indicating bimodality,

while smaller values support the null hypothesis H_0 , suggesting unimodality. In their observation, Pokojovy and Jobe (2022) suggested to use $Q_{\text{var}}(\cdot)$ to locally improve the initial guess of a “potential” direction and $Q_{\text{bimod}}(\cdot)$ to verify if the local improvement turns out to be adequate.

Using the PP step as part of the PP MCD algorithm, as detailed in Algorithm 3.1 and 3.2 by Pokojovy and Jobe (2022), a single warmstart is developed. The subsequent steps of the algorithm, including the C-step, rescaling, and reweighting, follow the DetMCD procedure. Unlike methods that involve randomization, the PP MCD algorithm ensures that both raw and reweighted estimators are deterministic and reproducible. This consistency is due to the fact that all statistical quantities in this method are computed through summation, making them independent of the sample indices of the observations. As a result, the estimators exhibit permutation invariance. Moreover, the PP MCD estimators are affine equivariant.

Chapter 4

Minimum Density Power Divergence (MDPD)

Consider a parametric family of models $\{F_{\boldsymbol{\theta}}, \boldsymbol{\theta} \in \boldsymbol{\Theta}\}$ with densities $f_{\boldsymbol{\theta}}$ defined with respect to Lebesgue measure and let \mathcal{G} denote the class of all (multivariate) distributions having densities with respect to the Lebesgue measure. Suppose $G \in \mathcal{G}$ is the true distribution with density g . The class of power divergence measure between the model density $f_{\boldsymbol{\theta}}$ and the true density g introduced by [Basu et al. \(1998\)](#) for robust estimation in general parametric models is given as:

$$d_{\alpha}(f_{\boldsymbol{\theta}}, g) = \begin{cases} \int \left\{ f_{\boldsymbol{\theta}}^{1+\alpha}(\mathbf{x}) - \left(1 + \frac{1}{\alpha}\right) f_{\boldsymbol{\theta}}^{\alpha}(\mathbf{x})g(\mathbf{x}) + \frac{1}{\alpha}g^{1+\alpha}(\mathbf{x}) \right\} d\mathbf{x}, & \alpha > 0, \\ \int g(\mathbf{x}) \log \left(\frac{g(\mathbf{x})}{f_{\boldsymbol{\theta}}(\mathbf{x})} \right) d\mathbf{x}, & \alpha = 0, \end{cases} \quad (4.1)$$

where $f_{\boldsymbol{\theta}}$ is the density with respect to the Lebesgue measure belonging to a family of parametric models $\{F_{\boldsymbol{\theta}}, \boldsymbol{\theta} \in \boldsymbol{\Theta}\}$, g is the density of true distribution $G \in \mathcal{G}$, and $\alpha \geq 0$ is the tuning parameter that balances efficiency with robustness.

When $\alpha = 0$, the $d_{\alpha}(f_{\boldsymbol{\theta}}, g)$ is defined as

$$\lim_{\alpha \rightarrow 0} d_{\alpha}(f_{\boldsymbol{\theta}}, g) = \int g(\mathbf{x}) \log \left(\frac{g(\mathbf{x})}{f_{\boldsymbol{\theta}}(\mathbf{x})} \right) d\mathbf{x},$$

and the minimum density power divergence (MDPD) function is the same the Kullback-Leibler (KL) divergence, and its minimization is equivalent to the maximization of the log-likelihood function, a method that lack robustness in parameter estimation. Conversely,

when $\alpha = 1$, the measure corresponds to the L_2 distance between the model density $f_{\boldsymbol{\theta}}$ and the true density g . However, the robustness of the minimum L_2 distance estimator comes at a significant cost in terms of asymptotic efficiency. Thus, the optimal range for α is between 0 and 1, providing a smooth bridge between the L_2 distance and the Kullback-Leibler divergence. While values of $\alpha > 1$ can be considered, they render the method highly inefficient. Given a family of parametric models, the parameter $\boldsymbol{\theta}$ is estimated by minimizing the density power divergence measure over the parameter space $\boldsymbol{\Theta}$. The resulting estimator is known as the minimum density power divergence (MDPD) estimator.

To estimate the parameter $\boldsymbol{\theta}$ using the MDPD, we express the estimator as $\widehat{\boldsymbol{\theta}}_{\alpha} = T_{\alpha}(G_n)$, where G_n represents the empirical distribution function derived from the observed sample. For $\alpha > 0$, the estimator $\widehat{\boldsymbol{\theta}}_{\alpha}$ is obtained by minimizing the following function $H_n(\boldsymbol{\theta})$ (observe that the last term in Equation (4.1) does not depend on $\boldsymbol{\theta}$):

$$H_n(\boldsymbol{\theta}) = \int_{\mathbb{R}^p} f_{\boldsymbol{\theta}}^{1+\alpha}(\mathbf{x}) d\mathbf{x} - \left(1 + \frac{1}{\alpha}\right) n^{-1} \sum_{i=1}^n f_{\boldsymbol{\theta}}^{\alpha}(\mathbf{x}_i). \quad (4.2)$$

Thus, the MDPD estimator $\widehat{\boldsymbol{\theta}}_{\alpha}$ is given by:

$$\widehat{\boldsymbol{\theta}}_{\alpha} = \arg \min_{\boldsymbol{\theta} \in \boldsymbol{\Theta}} H_n(\boldsymbol{\theta}). \quad (4.3)$$

Let $\mathbf{u}_{\boldsymbol{\theta}}(\mathbf{x}) = \frac{\partial}{\partial \boldsymbol{\theta}} \log f_{\boldsymbol{\theta}}(\mathbf{x})$ denote the score function of the model. Under the assumption of differentiability, the minimization of $H_n(\boldsymbol{\theta})$ leads to the following estimating equation:

$$\mathbf{U}_{\alpha,n}(\boldsymbol{\theta}) = \frac{1}{n} \sum_{i=1}^n \mathbf{u}_{\boldsymbol{\theta}}(\mathbf{x}_i) f_{\boldsymbol{\theta}}^{\alpha}(\mathbf{x}_i) - \int_{\mathbb{R}^p} \mathbf{u}_{\boldsymbol{\theta}}(\mathbf{x}) f_{\boldsymbol{\theta}}^{1+\alpha}(\mathbf{x}) d\mathbf{x} = \mathbf{0}. \quad (4.4)$$

Here, $\mathbf{0}$ represents a zero vector of dimension p . The robustness of this estimator against outliers arises from the weighting of the score function $\mathbf{u}_{\boldsymbol{\theta}}(\mathbf{x})$ by the power of the density $f_{\boldsymbol{\theta}}^{\alpha}(\mathbf{x}_i)$. This characteristic indicates that the MDPD estimator function as an M-estimator, which solve the equation of the form $\sum_{i=1}^n \psi(\mathbf{x}_i|\boldsymbol{\theta}) = \mathbf{0}$, with

$$\psi(\mathbf{x}|\boldsymbol{\theta}) = \mathbf{u}_{\boldsymbol{\theta}}(\mathbf{x}_i) f_{\boldsymbol{\theta}}^{\alpha}(\mathbf{x}_i) - \int_{\mathbb{R}^p} \mathbf{u}_{\boldsymbol{\theta}}(\mathbf{x}) f_{\boldsymbol{\theta}}^{1+\alpha}(\mathbf{x}) d\mathbf{x}. \quad (4.5)$$

The asymptotic consistency and normality results were consequently developed. Under suitable regularity conditions, [Basu et al. \(1998\)](#) showed that the MDPD estimator has compelling statistical characteristics. These encompass its asymptotic normality, its capacity for affine equivariance (particularly in the context of location and scatter estimation), its resilience in terms of breakdown point and efficiency, among other noteworthy qualities.

4.1 Multivariate Gaussian MDPD

Consider the p -variate MDPD objective function which follows from Equation (4.2) (the detailed results are given by [Pokojoy et al. \(2024a\)](#)):

$$H_n(\boldsymbol{\theta}) = \int_{\mathbb{R}^p} f_{\boldsymbol{\theta}}^{1+\alpha}(\mathbf{x}) d\mathbf{x} - \left(1 + \frac{1}{\alpha}\right) n^{-1} \sum_{i=1}^n f_{\boldsymbol{\theta}}^{\alpha}(\mathbf{x}_i). \quad (4.6)$$

Restricting our attention to the Gaussian distribution $\mathcal{N}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, we obtain the parameter $\boldsymbol{\theta} = (\boldsymbol{\mu}^T, \text{vec}(\boldsymbol{\Sigma})^T)^T$ and the density

$$\begin{aligned} f_{\boldsymbol{\theta}}(\mathbf{x}) &= (2\pi)^{-\frac{p}{2}} |\boldsymbol{\Sigma}|^{-\frac{1}{2}} \exp\left(-\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1}(\mathbf{x} - \boldsymbol{\mu})\right) \\ &= (2\pi)^{-\frac{p}{2}} |\boldsymbol{\Sigma}|^{-\frac{1}{2}} \exp\left(-\frac{1}{2}\|\boldsymbol{\Sigma}^{-1/2}(\mathbf{x} - \boldsymbol{\mu})\|^2\right) \end{aligned}$$

for $\boldsymbol{\mu} \in \mathbb{R}^p$ and $\boldsymbol{\Sigma} \in \text{SPD}(p \times p)$ (SPD denotes semi-positive definite), where

$$\text{SPD}(p \times p) := \{\mathbf{A} \in \mathbb{R}^{p \times p} \mid \mathbf{A}^T = \mathbf{A} \text{ and } \boldsymbol{\xi}^T \mathbf{A} \boldsymbol{\xi} > 0 \text{ for all } \boldsymbol{\xi} \in \mathbb{R}^p\}.$$

Refer to Appendix A for detailed calculations.

To evaluate the first integral on the right-hand side of Equation (4.6), we compute

$$\begin{aligned} \int_{\mathbb{R}^p} f_{\boldsymbol{\theta}}^{1+\alpha}(\mathbf{x}) d\mathbf{x} &= \int_{\mathbb{R}^p} (2\pi)^{-\frac{p}{2}(1+\alpha)} |\boldsymbol{\Sigma}|^{-\frac{1}{2}(1+\alpha)} \exp\left(-\frac{1}{2}(1+\alpha)\|\boldsymbol{\Sigma}^{-\frac{1}{2}}(\mathbf{x} - \boldsymbol{\mu})\|^2\right) d\mathbf{x} \\ &= (2\pi)^{-\frac{p}{2}(1+\alpha)} |\boldsymbol{\Sigma}|^{-\frac{1}{2}(1+\alpha)} \int_{\mathbb{R}^p} \exp\left(-\frac{1}{2}(1+\alpha)\|\boldsymbol{\Sigma}^{-\frac{1}{2}}(\mathbf{x} - \boldsymbol{\mu})\|^2\right) d\mathbf{x} \\ &= (2\pi)^{-\frac{p}{2}(1+\alpha)} |\boldsymbol{\Sigma}|^{-\frac{1}{2}(1+\alpha)} |\boldsymbol{\Sigma}|^{\frac{1}{2}} \int_{\mathbb{R}^p} \exp\left(-\frac{1}{2}(1+\alpha)\|\mathbf{u}\|^2\right) d\mathbf{u} \\ &= (2\pi)^{-\frac{p}{2}(1+\alpha)} |\boldsymbol{\Sigma}|^{-\frac{\alpha}{2}} \int_{\mathbb{R}^p} \exp\left(-\frac{1}{2}(1+\alpha)\|\mathbf{u}\|^2\right) d\mathbf{u}, \end{aligned}$$

where we invoked the transformation theorem by substituting $\mathbf{u} = \Sigma^{-\frac{1}{2}}(\mathbf{x} - \boldsymbol{\mu})$ and, thus, $\mathbf{x} = |\Sigma|^{\frac{1}{2}}\mathbf{u} + \boldsymbol{\mu}$ and $d\mathbf{x} = |\Sigma|^{\frac{1}{2}}d\mathbf{u}$. Similarly, substituting $\sqrt{\frac{1+\alpha}{2}}\mathbf{u} = \frac{1}{\sqrt{2}}\mathbf{v}$ and, thus,

$$\sqrt{1+\alpha}\mathbf{u} = \mathbf{v}, \quad \mathbf{u} = \frac{1}{\sqrt{1+\alpha}}\mathbf{v} \quad \text{and} \quad d\mathbf{u} = \left(\frac{1}{\sqrt{1+\alpha}}\right)^p d\mathbf{v},$$

we obtain

$$\begin{aligned} \int_{\mathbb{R}^p} f_{\boldsymbol{\theta}}^{1+\alpha}(\mathbf{x})d\mathbf{x} &= (2\pi)^{-\frac{p}{2}(1+\alpha)}|\Sigma|^{-\frac{\alpha}{2}}\left(\frac{1}{\sqrt{1+\alpha}}\right)^p \int_{\mathbb{R}^p} \exp\left(-\frac{1}{2}\|\mathbf{v}\|^2\right)d\mathbf{v} \\ &= \left((2\pi)^{-\frac{p}{2}} \int_{\mathbb{R}^p} \exp\left(-\frac{1}{2}\|\mathbf{v}\|^2\right)d\mathbf{v}\right) \left((2\pi)^{-\frac{\alpha p}{2}}(1+\alpha)^{-\frac{p}{2}}|\Sigma|^{-\frac{\alpha}{2}}\right) \\ &= (2\pi)^{-\frac{\alpha p}{2}}(1+\alpha)^{-\frac{p}{2}}|\Sigma|^{-\frac{\alpha}{2}} =: \psi(\Sigma). \end{aligned}$$

With the newly introduced function $\psi(\cdot)$, $H_n(\boldsymbol{\mu}, \Sigma)$ can then be expressed as

$$H_n(\boldsymbol{\mu}, \Sigma) = \psi(\Sigma) - \left(1 + \frac{1}{\alpha}\right) n^{-1} \sum_{i=1}^n f_{\boldsymbol{\theta}}^{\alpha}(\mathbf{x}_i). \quad (4.7)$$

We derive the estimates below. First, we compute the “gradient” of $H_n(\cdot)$.

For any $(\boldsymbol{\mu}, \Sigma)$ from the open subset $\mathbb{R}^p \times \text{SPD}(p \times p)$ of $\mathbb{R}^p \times \mathbb{R}^{p \times p}$, we can express the following:

$$\begin{aligned} \frac{\partial H_n}{\partial \boldsymbol{\mu}} &= \frac{\partial \psi}{\partial \boldsymbol{\mu}}(\Sigma) - \left(1 + \frac{1}{\alpha}\right) n^{-1} \sum_{i=1}^n \frac{\partial f_{\boldsymbol{\theta}}^{\alpha}}{\partial \boldsymbol{\mu}}(\mathbf{x}_i), \\ \frac{\partial H_n}{\partial \Sigma} &= \frac{\partial \psi}{\partial \Sigma}(\Sigma) - \left(1 + \frac{1}{\alpha}\right) n^{-1} \sum_{i=1}^n \frac{\partial f_{\boldsymbol{\theta}}^{\alpha}}{\partial \Sigma}(\mathbf{x}_i). \end{aligned}$$

Recalling $\mathbb{R}^p \times \text{SPD}(p \times p)$ is open, the Lagrange optimality conditions read as

$$\frac{\partial H_n}{\partial \boldsymbol{\mu}}(\boldsymbol{\mu}, \Sigma) = \mathbf{0}_p, \quad \frac{\partial H_n}{\partial \Sigma}(\boldsymbol{\mu}, \Sigma) = \mathbf{0}_{p \times p}. \quad (4.8)$$

The various parts are derived below:

$$\begin{aligned} \frac{\partial f_{\boldsymbol{\theta}}^{\alpha}}{\partial \boldsymbol{\mu}}(\mathbf{x}_i) &= f_{\boldsymbol{\theta}}^{\alpha}(\mathbf{x}_i)(-\alpha \Sigma^{-1}(\mathbf{x}_i - \boldsymbol{\mu})), \\ \frac{\partial f_{\boldsymbol{\theta}}^{\alpha}}{\partial \Sigma}(\mathbf{x}_i) &= -\frac{\alpha}{2} f_{\boldsymbol{\theta}}^{\alpha}(\mathbf{x}_i)(\Sigma^{-1} - \Sigma^{-1}(\mathbf{x}_i - \boldsymbol{\mu})(\mathbf{x}_i - \boldsymbol{\mu})' \Sigma^{-1}) \\ \frac{\partial \psi}{\partial \boldsymbol{\mu}}(\Sigma) &= \mathbf{0}_{p \times 1}, \\ \frac{\partial \psi}{\partial \Sigma}(\Sigma) &= -\frac{\alpha}{2} (2\pi)^{-\frac{\alpha p}{2}} (1+\alpha)^{-\frac{p}{2}} |\Sigma|^{-\frac{\alpha}{2}} \Sigma^{-1}, \end{aligned}$$

The first condition in Equation (4.8) can be expressed as

$$\begin{aligned}
\mathbf{0} &= \frac{\partial H_n}{\partial \boldsymbol{\mu}}(\boldsymbol{\mu}, \boldsymbol{\Sigma}) \\
&= -\left(1 + \frac{1}{\alpha}\right) n^{-1} \sum_{i=1}^n \frac{\partial f_{\boldsymbol{\theta}}^{\alpha}}{\partial \boldsymbol{\mu}}(\mathbf{x}_i) \\
&= \left(1 + \frac{1}{\alpha}\right) n^{-1} \alpha \boldsymbol{\Sigma}^{-1} \sum_{i=1}^n f_{\boldsymbol{\theta}}^{\alpha}(\mathbf{x}_i)(\mathbf{x}_i - \boldsymbol{\mu}) \\
&= \sum_{i=1}^n \mathbf{x}_i f_{\boldsymbol{\theta}}^{\alpha}(\mathbf{x}_i) - \boldsymbol{\mu} \sum_{i=1}^n f_{\boldsymbol{\theta}}^{\alpha}(\mathbf{x}_i).
\end{aligned}$$

Dividing the equation by $(2\pi)^{-p/2} |\boldsymbol{\Sigma}|^{-1/2}$ and introducing the weights

$$w_i(\boldsymbol{\mu}, \boldsymbol{\Sigma}) = \exp\left(-\frac{1}{2}(\mathbf{x}_i - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1}(\mathbf{x}_i - \boldsymbol{\mu})\right), \quad (4.9)$$

we obtain the first optimality condition

$$\begin{aligned}
\boldsymbol{\mu} \sum_{i=1}^n w_i^{\alpha}(\boldsymbol{\mu}, \boldsymbol{\Sigma}) &= \sum_{i=1}^n \mathbf{x}_i w_i^{\alpha}(\boldsymbol{\mu}, \boldsymbol{\Sigma}), \\
\boldsymbol{\mu} &= \frac{\sum_{i=1}^n \mathbf{x}_i w_i^{\alpha}(\boldsymbol{\mu}, \boldsymbol{\Sigma})}{\sum_{i=1}^n w_i^{\alpha}(\boldsymbol{\mu}, \boldsymbol{\Sigma})}.
\end{aligned}$$

Proceeding to the second condition in Equation (4.8), we get

$$\begin{aligned}
&\frac{\partial H_n}{\partial \boldsymbol{\Sigma}}(\boldsymbol{\mu}, \boldsymbol{\Sigma}) = \mathbf{0} \\
&-\frac{\alpha}{2}(2\pi)^{-\frac{\alpha p}{2}}(1 + \alpha)^{-\frac{p}{2}} |\boldsymbol{\Sigma}|^{-\frac{\alpha}{2}} \boldsymbol{\Sigma}^{-1} - \left(1 + \frac{1}{\alpha}\right) n^{-1} \sum_{i=1}^n \frac{\partial f_{\boldsymbol{\theta}}^{\alpha}}{\partial \boldsymbol{\Sigma}}(\mathbf{x}_i) = \mathbf{0} \\
&-\frac{\alpha}{2}(2\pi)^{-\frac{\alpha p}{2}}(1 + \alpha)^{-\frac{p}{2}} |\boldsymbol{\Sigma}|^{-\frac{\alpha}{2}} \boldsymbol{\Sigma}^{-1} + \\
&\frac{\alpha}{2} \left(1 + \frac{1}{\alpha}\right) n^{-1} \sum_{i=1}^n f_{\boldsymbol{\theta}}^{\alpha}(\mathbf{x}_i) (\boldsymbol{\Sigma}^{-1} - \boldsymbol{\Sigma}^{-1}(\mathbf{x}_i - \boldsymbol{\mu})(\mathbf{x}_i - \boldsymbol{\mu})' \boldsymbol{\Sigma}^{-1}) = \mathbf{0}.
\end{aligned}$$

Multiplying both sides of the equation with $2\alpha^{-1}(2\pi)^{\alpha p/2} |\boldsymbol{\Sigma}|^{\alpha/2} \boldsymbol{\Sigma}$ and recalling Equation (4.9), we get

$$\begin{aligned}
-(1 + \alpha)^{-\frac{p}{2}} \boldsymbol{\Sigma} + \frac{1}{n} \left(1 + \frac{1}{\alpha}\right) \sum_{i=1}^n w_i^{\alpha}(\boldsymbol{\mu}, \boldsymbol{\Sigma}) \boldsymbol{\Sigma} &= \frac{1}{n} \left(1 + \frac{1}{\alpha}\right) \sum_{i=1}^n w_i^{\alpha}(\boldsymbol{\mu}, \boldsymbol{\Sigma}) (\mathbf{x}_i - \boldsymbol{\mu})(\mathbf{x}_i - \boldsymbol{\mu})^T \\
\boldsymbol{\Sigma} &= \frac{1}{D_{\alpha}} \sum_{i=1}^n w_i^{\alpha}(\boldsymbol{\mu}, \boldsymbol{\Sigma}) (\mathbf{x}_i - \boldsymbol{\mu})(\mathbf{x}_i - \boldsymbol{\mu})^T,
\end{aligned}$$

where $D_\alpha(\boldsymbol{\mu}, \boldsymbol{\Sigma}) = \sum_{i=1}^n w_i^\alpha(\boldsymbol{\mu}, \boldsymbol{\Sigma}) - n\alpha(1 + \alpha)^{-\frac{p}{2}-1}$. Therefore, the MDPD estimates for the location vector and scatter matrix are given by:

$$\boldsymbol{\mu} = \frac{\sum_{i=1}^n \mathbf{x}_i w_i^\alpha(\boldsymbol{\mu}, \boldsymbol{\Sigma})}{\sum_{i=1}^n w_i^\alpha(\boldsymbol{\mu}, \boldsymbol{\Sigma})}, \quad (4.10)$$

$$\boldsymbol{\Sigma} = \frac{\sum_{i=1}^n w_i^\alpha(\boldsymbol{\mu}, \boldsymbol{\Sigma}) (\mathbf{x}_i - \boldsymbol{\mu})(\mathbf{x}_i - \boldsymbol{\mu})^T}{\sum_{i=1}^n w_i^\alpha(\boldsymbol{\mu}, \boldsymbol{\Sigma}) - n\alpha(1 + \alpha)^{-\frac{p}{2}-1}} \quad (4.11)$$

The existence and uniqueness of these estimating equations are given by [Maronna \(1976\)](#) under general assumptions about the functions w 's.

4.2 Asymptotic Distribution of MDPD Estimators

Theorem 4.2.1. *Under the appropriate regularity conditions (refer to [Basu et al., 1998](#)), with probability tending to 1 as $n \rightarrow \infty$, there exist $\widehat{\boldsymbol{\theta}}$ such that*

1. $\widehat{\boldsymbol{\theta}}$ is consistent estimator of $\boldsymbol{\theta}$
2. $n^{1/2}(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}) \sim \mathcal{N}(\mathbf{0}, \mathbf{J}^{-1} \mathbf{K} \mathbf{J}^{-1})$, that is asymptotically distribution multivariate normal with mean vector zero and scatter matrix $\mathbf{J}^{-1} \mathbf{K} \mathbf{J}^{-1}$, where

$$\begin{aligned} \mathbf{J} &= \int_{\mathbb{R}^p} (\mathbf{u}_\theta(\mathbf{x}) \otimes \mathbf{u}_\theta(\mathbf{x})) f_\theta^{1+\alpha}(\mathbf{x}) d\mathbf{x} \\ &\quad + \int_{\mathbb{R}^p} (i_\theta(\mathbf{x}) - \alpha \mathbf{u}_\theta(\mathbf{x}) \otimes \mathbf{u}_\theta(\mathbf{x})) (g(\mathbf{x}) - f_\theta(\mathbf{x}) f_\theta^\alpha) d\mathbf{x}, \\ \mathbf{K} &= \int_{\mathbb{R}^p} (\mathbf{u}_\theta(\mathbf{x}) \otimes \mathbf{u}_\theta(\mathbf{x})) f_\theta^{1+2\alpha}(\mathbf{x}) d\mathbf{x} - \boldsymbol{\xi}_\theta \otimes \boldsymbol{\xi}_\theta, \\ \boldsymbol{\xi}_\theta &= \int_{\mathbb{R}^p} \mathbf{u}_\theta(\mathbf{x}) f_\theta^{1+\alpha}(\mathbf{x}) d\mathbf{x}, \end{aligned}$$

with $\boldsymbol{\xi}_\theta = \int_{\mathbb{R}^p} \mathbf{u}_\theta(\mathbf{x}) f_\theta^\alpha(\mathbf{x}) g(\mathbf{x}) d\mathbf{x}$ and the information function $i_\theta(\mathbf{x}) := -\frac{\partial \mathbf{u}_\theta(\mathbf{x})}{\partial \boldsymbol{\theta}}$.

For the proof of Theorem 4.2.1, refer to (Theorem 6.4.1 of [Lehmann and Casella, 1998](#); [Das et al., 2022](#)). We however evaluate (see [Nkum et al., 2024](#)) the integral in the Theorem (4.2.1) as follows:

4.2.1 Asymptotic Distribution of the Location Vector

The asymptotic normality results for the multivariate Gaussian location vector MDPD estimator are given as:

$$n^{1/2} \left(\hat{\boldsymbol{\mu}}_{\text{MDPD}}^n - \hat{\boldsymbol{\mu}} \right) \xrightarrow{\mathcal{D}} \mathcal{N} \left(\mathbf{0}, \left(1 + \frac{\alpha^2}{1 + 2\alpha} \right)^{p/2+1} \boldsymbol{\Sigma} \right), \quad \text{as } n \rightarrow \infty. \quad (4.12)$$

The detailed calculations are provided in Appendix A.3.

4.2.2 Asymptotic Distribution of the Scatter Matrix

The asymptotic normality results for the multivariate Gaussian scatter matrix MDPD estimator are given as:

$$n^{1/2} \left(\hat{\boldsymbol{\Sigma}}_{\text{MDPD}}^n - \hat{\boldsymbol{\Sigma}} \right) \xrightarrow{\mathcal{D}} \mathcal{N} \left(\mathbf{0}, \frac{4(1 + \alpha)^{p+4}}{(1 + 2\alpha)^{(p+4)/2}} \left(\mathbf{K}(\boldsymbol{\alpha}) - \alpha^2 \right) (\boldsymbol{\Sigma} \otimes \boldsymbol{\Sigma}) \right) \quad \text{as } n \rightarrow \infty, \quad (4.13)$$

where

$$\begin{aligned} \mathbf{K}(\boldsymbol{\alpha}) &= \left(\alpha^2 \mathbf{I}_{pp} + \mathbf{K}_{pp} + \text{vec}(\boldsymbol{\Sigma}^{-1}) \text{vec}(\boldsymbol{\Sigma})^T \right)^{-1} \times \left(4\alpha^2 \mathbf{I}_{pp} + \mathbf{K}_{pp} + \text{vec}(\boldsymbol{\Sigma}^{-1}) \text{vec}(\boldsymbol{\Sigma})^T \right) \\ &\quad \times \left(\alpha^2 \mathbf{I}_{pp} + \mathbf{K}_{pp} + \text{vec}(\boldsymbol{\Sigma}^{-1}) \text{vec}(\boldsymbol{\Sigma})^T \right)^{-1}. \end{aligned}$$

The detailed calculations are provided in Appendix A.5.

4.3 Gross Breakdown Point

The breakdown point of an estimator, which essentially measures the fraction of erroneous data that can be handled before the estimator becomes unreliable, serves as a global indicator of the estimator's robustness. In this context, we assess the gross-error breakdown point for the minimum density power divergence estimator of normal distribution parameters under specific contamination scenarios ([Ronchetti and Huber, 2009](#), p. 97).

Consider a p -variate normal distribution of a random variable $\mathbf{x} = (x_1, x_2, \dots, x_p) \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, where $\boldsymbol{\mu}$ is the mean vector and $\boldsymbol{\Sigma}$ is the covariance matrix. Let $\alpha > 0$ and let g

be the multivariate normal density $\mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, expressed as $\phi_{\boldsymbol{\mu}, \boldsymbol{\Sigma}}(\cdot) = |\boldsymbol{\Sigma}|^{-1/2} \phi(\boldsymbol{\Sigma}^{-1/2}(\cdot - \boldsymbol{\mu}))$ and $f_{\boldsymbol{\theta}}$ be the $\mathcal{N}(\mathbf{m}, \mathbf{S})$, we define the contaminated model with point contamination

$$q(\mathbf{x}) = (1 - \varepsilon)g(\mathbf{x}) + \varepsilon\delta_{\mathbf{z}}(\mathbf{x}), \quad (4.14)$$

where $\delta_{\mathbf{z}}(\mathbf{x})$ is the Dirac delta function and $\mathbf{z} \rightarrow \infty$. The data are a random sample drawn from q , and the target parameters are $\boldsymbol{\theta} = (\boldsymbol{\mu}, \boldsymbol{\Sigma})$.

From the class of minimum density divergence in Equation (4.1), consider the maximizer of as presented by [Basu et al. \(1998\)](#):

$$\begin{aligned} \psi(\mathbf{m}, \mathbf{S}) &\equiv (1 + \alpha) \int q(\mathbf{x}) f_t^\alpha(\mathbf{x}) d\mathbf{x} - \alpha \int f_t^{1+\alpha}(\mathbf{x}) d\mathbf{x} \\ &= (1 + \alpha)(1 - \varepsilon) \left\{ \int \phi_{\boldsymbol{\mu}, \boldsymbol{\Sigma}}(\mathbf{x}) \phi_{\mathbf{m}, \mathbf{S}}^\alpha(\mathbf{x}) d\mathbf{x} + \varepsilon \int \delta_{\mathbf{z}}(\mathbf{x}) \phi_{\mathbf{m}, \mathbf{S}}^\alpha(\mathbf{x}) d\mathbf{x} \right\} - \alpha \int \phi_{\mathbf{m}, \mathbf{S}}^{1+\alpha}(\mathbf{x}) d\mathbf{x}. \end{aligned}$$

Let $\mathbf{x}, \mathbf{c}, \mathbf{m} \in \mathbb{R}^p$ and \mathbf{D}, \mathbf{S} be positive definite matrices. The multivariate normal density is given by:

$$\phi_{\boldsymbol{\mu}, \boldsymbol{\Sigma}}(\mathbf{x}) = \frac{1}{(2\pi)^{p/2} |\boldsymbol{\Sigma}|^{1/2}} \exp \left(-\frac{1}{2} (\mathbf{x} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1} (\mathbf{x} - \boldsymbol{\mu}) \right).$$

Now, let's consider the integral (detail derivation is given in Appendix (A.10)):

$$\begin{aligned} &\int \phi_{\mathbf{c}, \mathbf{D}}(\mathbf{x}) \phi_{\mathbf{m}, \mathbf{S}}^\alpha(\mathbf{x}) d\mathbf{z} \\ &= \frac{1}{(2\pi)^{\alpha p/2} |\mathbf{D}|^{1/2} |\mathbf{D}^{-1} + \alpha \mathbf{S}^{-1}|^{1/2} |\mathbf{S}|^{\alpha/2}} \\ &\times \exp \left(-\frac{1}{2} \mathbf{c}' \mathbf{D}^{-1} \mathbf{c} - \frac{\alpha}{2} \mathbf{m}' \mathbf{S}^{-1} \mathbf{m} + \frac{1}{2} (\mathbf{D}^{-1} \mathbf{c} + \alpha \mathbf{S}^{-1} \mathbf{m})^T (\mathbf{D}^{-1} + \alpha \mathbf{S}^{-1})^{-1} (\mathbf{D}^{-1} \mathbf{c} + \alpha \mathbf{S}^{-1} \mathbf{m}) \right). \end{aligned}$$

Also consider the integral:

$$\begin{aligned} \int \delta(\mathbf{x} - \mathbf{z}) \phi_{\mathbf{m}, \mathbf{S}}^\alpha(\mathbf{x}) d\mathbf{x} &= \phi_{\mathbf{m}, \mathbf{S}}^\alpha(\mathbf{z}) \\ &= \frac{1}{|\mathbf{S}|^{\alpha/2} (2\pi)^{\alpha p/2}} \exp \left(-\frac{\alpha}{2} (\mathbf{x} - \mathbf{m})^T \mathbf{S}^{-1} (\mathbf{x} - \mathbf{m}) \right), \end{aligned}$$

where: $\int \delta(\mathbf{x} - \mathbf{z}) \phi(\mathbf{x}) d\mathbf{x} = \phi(\mathbf{z})$ by the property of the Dirac delta function.

Finally, we have the integral:

$$\begin{aligned} \int \phi_{\mathbf{m}, \mathbf{S}}^{1+\alpha}(\mathbf{x}) d\mathbf{x} &= \frac{1}{(2\pi)^{(1+\alpha)p/2} |\mathbf{S}|^{(1+\alpha)/2}} \int \exp \left(-\frac{1+\alpha}{2} (\mathbf{x} - \mathbf{m})^T \mathbf{S}^{-1} (\mathbf{x} - \mathbf{m}) \right) (1 + \alpha) d\mathbf{x} \\ &= \frac{1}{(2\pi)^{\alpha p/2} |\mathbf{S}|^{\alpha/2}} \cdot \frac{1}{(1 + \alpha)^{p/2}}. \end{aligned}$$

Let $\mathbf{B} = \mathbf{\Sigma}\mathbf{S}^{-1}$. It follows that $\psi_1(\mathbf{m}, A) \equiv (2\pi)^{\alpha p/2} |\mathbf{S}|^{\alpha/2} \psi(\mathbf{m}, A)$ is given by

$$\begin{aligned} \psi_1(\mathbf{m}, \mathbf{B}) = & \\ \mathbf{B}^{\alpha/2} \left(\frac{(1+\alpha)(\varepsilon-1) \exp\{-\frac{1}{2}\boldsymbol{\mu}'\mathbf{\Sigma}^{-1}\boldsymbol{\mu} + \frac{1}{2}(\mathbf{I}\boldsymbol{\mu} + \alpha\mathbf{B}\mathbf{m})^T\mathbf{\Sigma}^{-1}(\mathbf{I} + \alpha\mathbf{B})^{-1}(\mathbf{I}\boldsymbol{\mu} + \alpha\mathbf{B}\mathbf{m}) - \frac{\alpha}{2}\mathbf{m}'\mathbf{\Sigma}^{-1}\mathbf{B}\mathbf{m}\}}{|\mathbf{I} + \alpha\mathbf{B}|^{1/2}} \right. & \\ \left. + \varepsilon(1+\alpha) \exp\left(-\frac{\alpha}{2}(\mathbf{x} - \mathbf{m})^T\mathbf{\Sigma}^{-1}\mathbf{B}(\mathbf{x} - \mathbf{m})\right) - \frac{\alpha}{(1+\alpha)^{p/2}} \right). & \end{aligned}$$

We now wish to maximise this quantity over \mathbf{B} (rather than \mathbf{S} and \mathbf{m} . First, $\psi_1(\mathbf{m}, 0) = 0$.

For $\mathbf{B} > 0$, $\psi_1(\mathbf{m}, \mathbf{B})$ consists mainly of two ridges which have heights

$$\mathbf{B}^{\alpha/2} \left[\varepsilon(1+\alpha) - \left(\frac{\alpha}{(1+\alpha)^{p/2}} \right) \right] \text{ at } \mathbf{m} = \mathbf{x},$$

and

$$\mathbf{B}^{\alpha/2} \left[\frac{(1+\alpha)(1-\varepsilon)}{|\mathbf{I} + \alpha\mathbf{B}|^{1/2}} - \left(\frac{\alpha}{(1+\alpha)^{p/2}} \right) \right] \text{ at } \mathbf{m} = \boldsymbol{\mu}.$$

If the height at $\mathbf{m} = \mathbf{x}$ is negative, which occurs when $\varepsilon < K \equiv \frac{\alpha}{(1+\alpha)^{p/2+1}}$, then $\mathbf{B} = 0$ would be optimal only if the ridge height at $\mathbf{m} = \boldsymbol{\mu}$ is also negative for all $\mathbf{B} > 0$. This happens if $\varepsilon > 1 - K$. However, since $K < 1/2$, the condition $1 - K < \varepsilon < K$ is impossible. Therefore, $\mathbf{B} = 0$ cannot maximise $\psi_1(\mathbf{m}, \mathbf{B})$.

Conversely, if the ridge height at $\mathbf{m} = \mathbf{x}$ is positive, the value along this ridge will increase to ∞ as $\mathbf{B} \rightarrow \infty$. In contrast, the values along the $\mathbf{m} = \boldsymbol{\mu}$ ridge remain finite: They might be positive at some points, but they will have a finite maximum at a finite \mathbf{B} and approach a negative value as $\mathbf{B} \rightarrow \infty$. Therefore, if the ridge height at $\mathbf{m} = \mathbf{x}$ is positive, the maximum occurs at $\mathbf{m} = \mathbf{x}$ with $\mathbf{S} = \mathbf{0}$, meaning \mathbf{B} will increase indefinitely, leading to the highest possible values for $\psi_1(\mathbf{m}, \mathbf{B})$. This implies that, with enough outliers, the normal fit tries to match the extreme data points, leading to a mean vector $\mathbf{x} \rightarrow \infty$ and and scatter matrix $\mathbf{0}$. This is simultaneous location vector and scatter matrix breakdown in the sense that location “explodes” and scatter “implodes” ([Hampel et al., 1986](#)).

Therefore, the nominal breakdown occurs if

$$\varepsilon > \varepsilon_*(\alpha) = \frac{\alpha}{(1+\alpha)^{p/2+1}}. \quad (4.15)$$

To maximize ε , observe that

$$\begin{aligned}\varepsilon_*(\alpha) &= \frac{\alpha}{(1+\alpha)^{p/2+1}}, \\ \varepsilon'_*(\alpha) &= \frac{d}{d\alpha} \left(\frac{\alpha}{(1+\alpha)^{p/2+1}} \right).\end{aligned}$$

Using the quotient rule:

$$\begin{aligned}\varepsilon'_*(\alpha) &= \frac{(1+\alpha)^{p/2+1} - \alpha \left(\frac{p}{2} + 1\right) (1+\alpha)^{p/2}}{(1+\alpha)^{2(p/2+1)}} \\ &= \frac{(1+\alpha) - \alpha \left(\frac{p}{2} + 1\right)}{(1+\alpha)^{p/2+2}} \\ &= \frac{1 - \alpha \frac{p}{2}}{(1+\alpha)^{p/2+2}}.\end{aligned}$$

We set $\varepsilon'_*(\alpha) = 0$ to find the critical points. Solving $\varepsilon'_*(\alpha) = 0$ gives:

$$1 - \alpha \frac{p}{2} = 0 \implies \alpha = \frac{2}{p}.$$

Thus, maximum nominal breakdown point α occurs at $\frac{2}{p}$.

Let's now examine it's asymptotic behaviour:

Consider:

$$\varepsilon_*\left(\frac{2}{p}\right) = \frac{\frac{2}{p}}{\left(1 + \frac{2}{p}\right)^{p/2+1}} = \left(\frac{p}{p+2}\right)^{p/2} \cdot \frac{2}{p+2}.$$

Let $x = \frac{p}{2}$. We know from Euler identity that, as $x \rightarrow \infty$, we obtain:

$$\lim_{x \rightarrow \infty} \left(1 - \frac{1}{x+1}\right)^n = e^{-1},$$

thus:

$$\begin{aligned}\left(\frac{p}{p+2}\right)^{p/2} &= \left(1 - \frac{2}{p+2}\right)^{p/2} = \left(1 - \frac{1}{\frac{p}{2}+1}\right)^{p/2} \\ &= e^{-1}.\end{aligned}$$

Therefore, the asymptotic nominal breakdown point, independent of the tuning parameter α , is given by:

$$\varepsilon > \frac{2e^{-1}}{p+2}.$$

We observe that, ε is decreasing function of p . From Figure 4.1 we observe that as p increases the maximum nominal breakdown point decreases. This reduction in the breakdown point with increasing p implies that the robustness of the estimator diminishes in higher-dimensional spaces. Consequently, the estimator becomes more susceptible to outliers and deviations from the model assumptions as the dimensionality of the data increases.

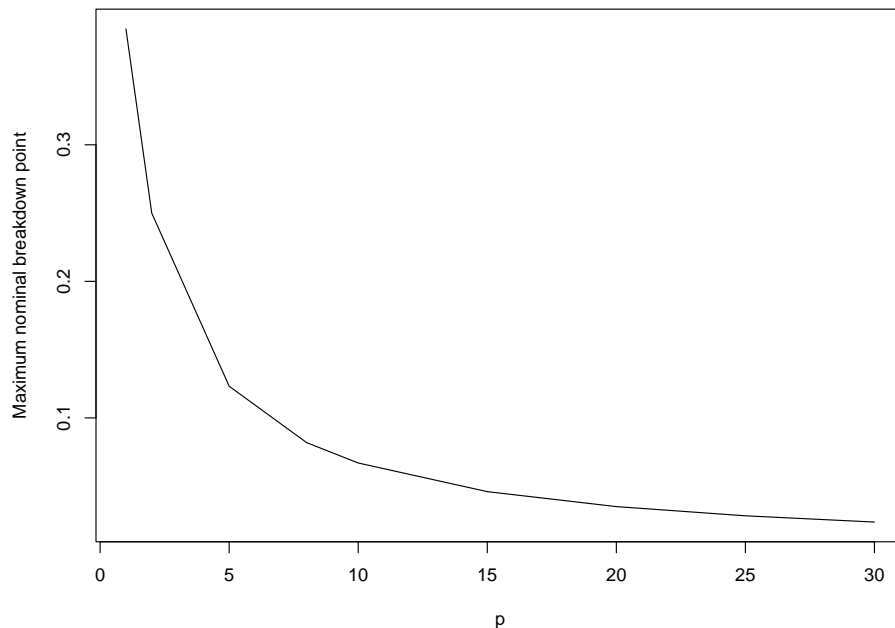


Figure 4.1: Maximum nominal breakdown point for the MDPD estimator.

A notable limitation of the MDPD estimator is that its breakdown point (bdp) reduces to 0 as the dimensionality p increases, making it less effective in very high-dimensional contexts. This reduction in the breakdown point indicates that the estimator's robustness significantly diminishes as the number of variables increases, rendering it more vulnerable

to the presence of outliers and deviations from model assumptions. Consequently, applying this method to datasets with dimensions not exceeding 15 is ideal. In higher dimensions, alternative robust estimation methods or dimensionality reduction techniques should be considered to maintain robustness and effectiveness.

Chapter 5

Multivariate Gaussian MDPD

Algorithm

The minimum density power divergence (MDPD) method is a robust M-estimator-based statistical technique that balances efficiency and robustness in estimating multivariate parameters. We introduce the multivariate Gaussian MDPD algorithm, providing a detailed computational framework for implementing this robust estimation method. The algorithm is precious for handling data with outliers or deviations from the assumed distribution, ensuring reliable estimates of the mean vector and covariance matrix. We will explore the algorithm's iterative fixed-point approach, initialization, convergence criteria, and the robustness parameter α , which controls the influence of outliers.

M-estimators are characterized by their estimating equations, leading to an iterative algorithm. The convergence of this procedure is well-established, as detailed in [Maronna \(1976\)](#). The solution for monotone M-estimators is unique but they lack robust in high dimension. Thus, while the initial point may influence the number of iterations required, they do not affect the final results, guaranteeing the robustness and reliability of these estimators. On the hand, redescending M-estimators are very efficient, have a high breakdown point and unlike many other methods, they do not suffer from masking effect. However reascending estimators may not have a unique solution. So the initial point for the iterative solution should be carefully chosen possibly by another estimator.

[Basu et al. \(1998\)](#) observed that the influence functions for the estimation of the standard

normal location have a rescinding nature. However, [Pokojoy et al. \(2024a\)](#) outlined a custom fixed-point iterative algorithm to estimate the mean vector and scatter matrix using the minimum density power divergence (MDPD) method, which always converges to a local minimum of the MDPD loss function.

The algorithm starts with initial estimators $\hat{\boldsymbol{\mu}}_0$ and $\hat{\boldsymbol{\Sigma}}_0$. The convergence of the algorithm is determined based on either an error threshold or a maximum number of iterations, whichever is reached first. It is designed to robustly estimate the mean vector and covariance matrix of a multivariate data set using the minimum density power divergence (MDPD) method. Robust estimation is crucial when dealing with data that may contain outliers or deviate from the assumed distribution. The MDPD approach provides a way to balance efficiency and robustness by introducing a parameter α that controls the trade-off between these two aspects.

The process begins with the initialization of data and parameters: The algorithm set $n \times p$ data matrix as an input, where n is the number of observations and p is the number of variables. A small positive value ϵ (e.g., 10^{-8}) is chosen to set the convergence tolerance for the iterative updates. The robustness parameter α is selected to control the influence of outliers, which can be predefined or computed based on a given breakdown point (bdp). The breakdown represents the proportion of contaminated data the estimator can handle before breaking down.

The algorithm starts with an initial mean vector $\boldsymbol{\mu}^{(0)}$ and covariance matrix $\boldsymbol{\Sigma}^{(0)}$. Here, the classical mean vector and scatter matrix was used. The iterative process updates these estimates using a fixed-point approach until convergence is reached base on the error threshold or the maximum iteration. This procedure converges to a local minimum (see [Maronna, 1976](#)). Similar to the MCD procedure, a minimum subset size is required to achieve a specified breakdown point. A novel addition to the fixed-point iteration is a weight adjustment mechanism. If the covariance is too small, the algorithm identifies the smallest inflation factor for the current iteration to ensure that the resulting weights sum

up to at least the minimum subset size (h_{\min}). This mechanism ensures that the effective sample size remains robust by maintaining the minimum subset size, leading to the desired breakdown point and convergence to the local minimum.

This algorithm is particularly significant in its robustness, efficiency, and flexibility. By incorporating the robustness parameter α , it can handle outliers and deviations from the assumed distribution, providing more reliable estimates. The iterative fixed-point approach ensures efficient computation, balancing robustness with computational feasibility. This makes the algorithm especially useful in applications where data quality is a concern, and robust statistical methods are required to obtain accurate estimates of central tendency and dispersion. Furthermore, the fixed-point iteration algorithm with MDPD preserves affine equivariance, an important property in robust statistics. Affine equivariance means that if the data undergoes an affine transformation, the estimates produced by the algorithm will undergo the same transformation. The preservation of affine equivariance further enhances its applicability and reliability in various practical scenarios (as already discussed in Section 2.3).

Algorithm 1: Fixed-Point Iteration

Data: The n by p matrix, where $n > p$

- 1 Select $\varepsilon > 0$ small, e.g., $\varepsilon := 10^{-9}$, K (maximum iteration, default is 500) and choose $\alpha \in [0, 2/p]$ ($\alpha = 0$ is the classical estimates and $\alpha = 2/p$ gives the asymptotically maximum bdp)
- 2 Initialize the mean $\boldsymbol{\mu}^{(0)}$ and covariance matrix $\boldsymbol{\Sigma}^{(0)}$ using:

$$\boldsymbol{\mu}^{(0)} := \bar{\mathbf{x}} \equiv \frac{1}{n} \sum_{i=1}^n \mathbf{x}_i, \quad \boldsymbol{\Sigma}^{(0)} := \mathbf{S} \equiv \frac{1}{n} \sum_{i=1}^n (\mathbf{x}_i - \bar{\mathbf{x}})(\mathbf{x}_i - \bar{\mathbf{x}})^T$$

for $k = 0, 1, 2, \dots, K$ **do**

- 3 Compute the updated mean $\boldsymbol{\mu}^{(k+1)}$ and covariance matrix $\boldsymbol{\Sigma}^{(k+1)}$ using weighted averages:

$$\boldsymbol{\mu}^{(k+1)} = \frac{\sum_{i=1}^n (w_i^{(k)})^\alpha \mathbf{x}_i}{\sum_{i=1}^n (w_i^{(k)})^\alpha}, \quad \boldsymbol{\Sigma}^{(k+1)} = \frac{\sum_{i=1}^n (w_i^{(k)})^\alpha (\mathbf{x}_i - \boldsymbol{\mu}^{(k)})(\mathbf{x}_i - \boldsymbol{\mu}^{(k)})^T}{\sum_{i=1}^n w_i^\alpha - n\alpha(1 + \alpha)^{-\frac{p}{2}-1}}$$

where the weights $w_i^{(k)}$ are given by:

$$w_i(\boldsymbol{\mu}^{(k)}, \boldsymbol{\Sigma}^{(k)}) = \exp\left(-\frac{1}{2}(\mathbf{x}_i - \boldsymbol{\mu}^{(k)})^T \boldsymbol{\Sigma}^{-1}(\mathbf{x}_i - \boldsymbol{\mu}^{(k)})\right)$$

- 4 Adjust weights if necessary (h_{\min} is the minimum subset size):

- 5 **if** $\sum_i w_i^{(k+1)} < h_{\min}$ **then**

- 6
$$r = \arg \min_{r \in [0, 2]} \left| \sum_{i=1}^n \exp\left(-\frac{1}{2}\alpha(\mathbf{x}_i - \boldsymbol{\mu}^{(k)})^T \boldsymbol{\Sigma}^{-1}(\mathbf{x}_i - \boldsymbol{\mu}^{(k)})r\right) - h_{\min} \right|$$

$$w_i^{(k+1)} = \exp\left(-\frac{r}{2}\alpha(\mathbf{x}_i - \boldsymbol{\mu}^{(k)})^T \boldsymbol{\Sigma}^{-1}(\mathbf{x}_i - \boldsymbol{\mu}^{(k)})\right)$$

- 7 **end**

- 8 Compute the squared error:

- 9
$$\epsilon \leftarrow \frac{1}{n} \sum_{i=1}^n (w_i^{(k+1)} - w_i^{(k)})^2$$

- 10 **if** $\epsilon < \varepsilon$ **then**

- 10 | **break**

- 11 **end**

- 12 **end**

- 13 **return** *Best* $\boldsymbol{\mu}$, $\boldsymbol{\Sigma}$ and *iter*
-

Chapter 6

Simulation Study for MDPD Estimators

In this Chapter, we present a comparative analysis of the minimum density power divergence (MDPD) estimator against the FASTMCD and DetMCD algorithms based on simulated data to assess the empirical performance of the MDPD estimator. The simulation study is designed to evaluate the robustness and accuracy of various estimators for the mean vector and scatter matrix of a multivariate data set. The simulation process involves several key steps: generating multivariate data, introducing contamination to simulate outliers, applying different robust estimation methods, and computing error measures to evaluate each estimator's performance. These steps are essential for providing a comprehensive assessment of the estimators' robustness to outliers and their overall accuracy in parameter estimation.

We conducted all simulations using the R programming language. For the MDPD estimator, we implemented the fixed-point iteration algorithm to estimate the mean vector and scatter matrix, as provided by the R package from [Pokojoy et al. \(2024b\)](#). This custom implementation leverages the robustness parameter α to control the influence of outliers and efficiency. FASTMCD and DetMCD estimators were computed using the `covMcd` function from the `robustbase` package in R, which implements well-established algorithms for robust covariance estimation. By comparing the MDPD estimator with the FASTMCD and DetMCD algorithms, this simulation study aims to provide valuable insights into the relative strengths and weaknesses of each method. The results will highlight the scenarios

in which the MDPD estimator excels, particularly in terms of handling data contamination and achieving reliable parameter estimates.

6.1 Data generating setup

We generated data matrices of various sizes from a multivariate normal distribution. Specifically, we considered sample sizes of $n = 30, 50, 100, 200, 500$ and dimensions $p = 2, 5, 10, 15$. Each data set consists of n observations of p variables, drawn from a standard multivariate Gaussian distribution

$$\mathbf{x}_i \stackrel{\text{i.i.d}}{\sim} (1 - \varepsilon)\mathcal{N}_p(\mathbf{0}, \mathbf{I}) + \varepsilon\mathcal{N}_p\left(\mathbf{1}\sqrt{ncp/p}, \mathbf{I}\right), \quad \text{for } i = 1, 2, \dots, n,$$

where \mathbf{I} is the $p \times p$ identity matrix. Since MDPD estimators are affine equivariant, we can assume, without loss of generality, that the location vector is $\mathbf{0}$ and the scatter matrix is \mathbf{I} .

To simulate contaminated data, we introduced outliers by modifying a fraction of the observations. The contamination fraction ε was varied to evaluate its impact on the robustness of the estimators. Outliers were generated by adding a constant shift to the contaminated observations. This shift is determined by the non-centrality parameter (ncp) which ensures that the shift is significant relative to the data's dimensions, and is then scaled by a factor δ . This process effectively moves the contaminated observations away from the central distribution, creating realistic outliers. The contamination was applied as follows:

$$\mathbf{x}_i = \begin{cases} \mathbf{1}\sqrt{\frac{ncp}{p}} + \delta\mathbf{z}_i, & \text{for } i = 1, \dots, [(\varepsilon n)] \\ \mathbf{z}_i, & \text{for } i = [(\varepsilon n)] + 1, \dots, n \end{cases}, \quad (6.1)$$

where $\mathbf{z}_i \stackrel{\text{i.i.d}}{\sim} \mathcal{N}_p(\mathbf{0}, \mathbf{I})$ and $ncp = 10, 25, 100, 250, 500$ and $\delta = 0.001, 1.0, 3.0$. The contamination described gives rise two types of contamination, point contamination when $\delta = 1.0$ and cluster contamination when $\delta = 1.0$ and $\delta = 3$, where the value of δ influences the

dispersion of the contamination cluster. This simulation setup is similar to [Hubert et al. \(2012\)](#) and [Pokojoy and Jobe \(2022\)](#).

6.2 Evaluation Criterion

To assess the performance of the estimators, several error measures were adopted:

(1) **Squared Error of the Mean Vector:**

$$e_{\boldsymbol{\mu}} = \|\widehat{\boldsymbol{\mu}}(\mathbf{X})\|^2 = \sum_{i=1}^p \widehat{\boldsymbol{\mu}}_i^2.$$

(2) **Normalized Frobenius Norm Error of the Covariance Matrix:**

$$e_{\boldsymbol{\Sigma}} = \frac{1}{p} \left(\text{tr}(\widehat{\boldsymbol{\Sigma}}\widehat{\boldsymbol{\Sigma}}) - 2\text{tr}(\widehat{\boldsymbol{\Sigma}}) + p \right).$$

(3) **Kullback-Leibler Divergence:**

$$e_{KL} = \frac{1}{2} \left(\text{tr}(\widehat{\boldsymbol{\Sigma}}) - p + \sum_{i=1}^p (\widehat{\boldsymbol{\mu}}_i)^2 - \log \det(\widehat{\boldsymbol{\Sigma}}) \right).$$

For each of the error measures, 1,000 replications were performed, and the average error and the standard deviation of the error were reported for comparison. Each of these performance measures should be as close to zero as possible.

6.2.1 No Contamination

Figure 6.1 to Figure 6.4 below display the average error of the mean vector and scatter estimates for different simulation scenarios, characterized by varying dimensions (p) and nominal breakdown points (bdp). Each bar in the figures represents the standard deviations of the error measures, providing an indication of the variability in the estimates.

The results across all scenarios consistently show that MDPD performs comparably to FastMCD and DetMCD in terms of estimating the mean vector and scatter. A significant

observation is that while the average error decreases with increasing sample sizes for all methods, the MDPD method tends to exhibit less variability in its estimates. This is evidenced by the smaller standard deviation bars in the figures, indicating more stable and reliable performance across different simulations. In contrast, the DetMCD method appears to exhibit greater variability in its estimates, as indicated by the larger error bars.

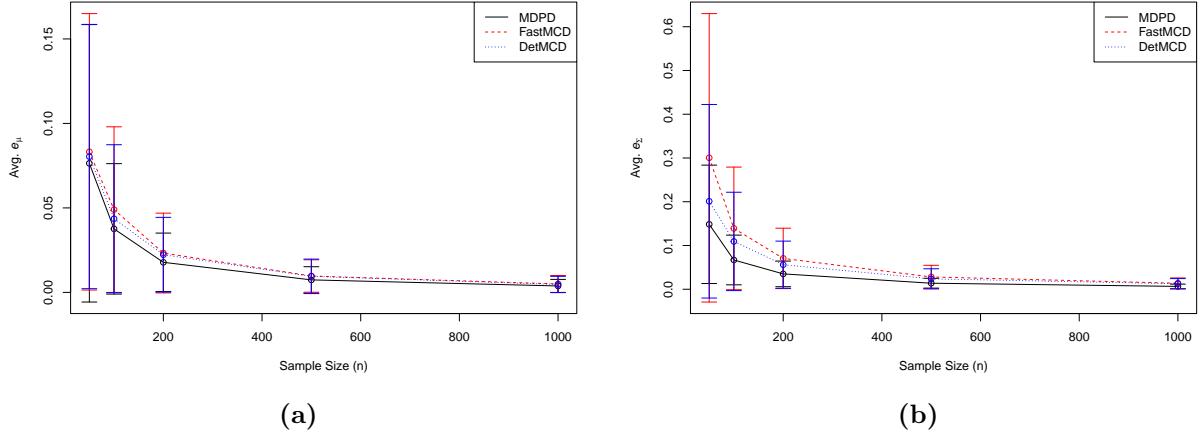


Figure 6.1: MSE and SD of errors of the mean vector and scatter matrix estimates for uncontaminated data: simulation scenario with $p = 2$, and $bdp = 0.25$.

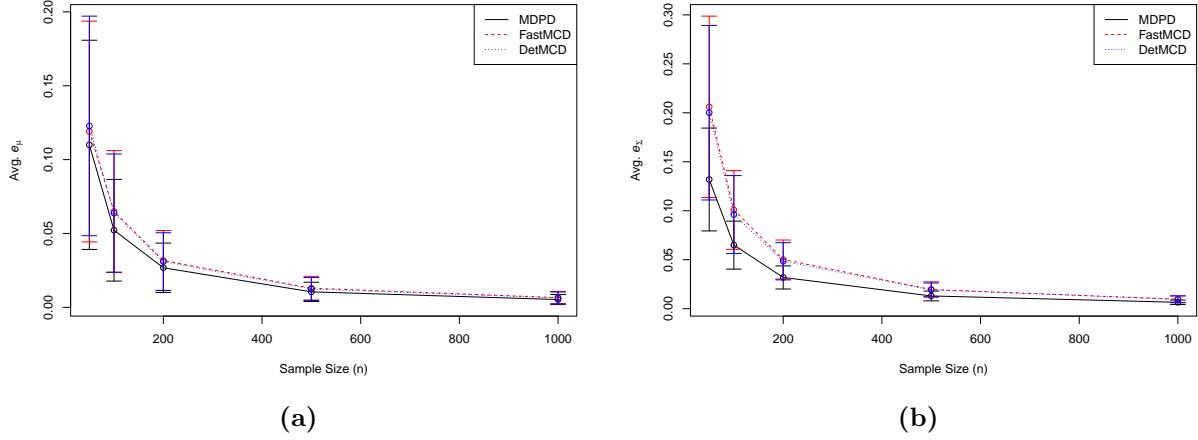
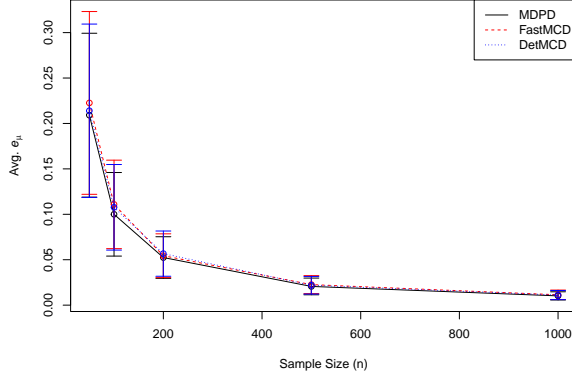
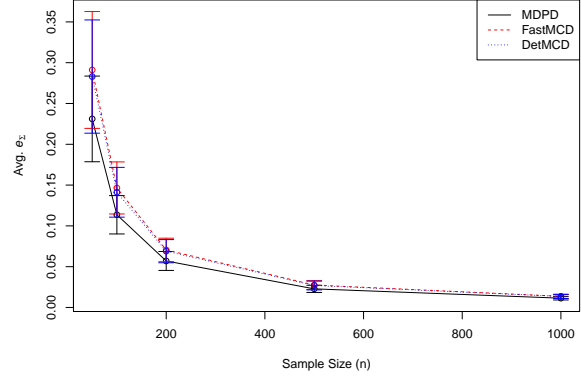


Figure 6.2: MSE and SD of errors of the mean vector and scatter matrix estimates for uncontaminated data: simulation scenario with $p = 5$, and $bdp = 0.0924$.

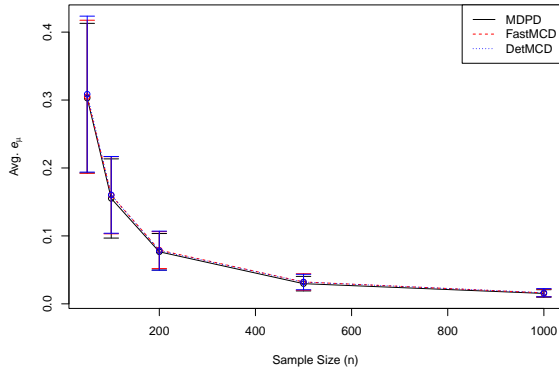


(a)

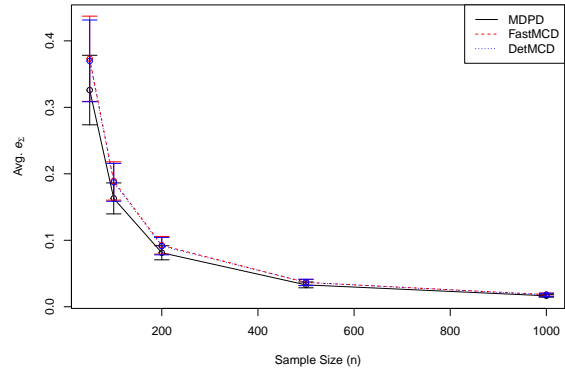


(b)

Figure 6.3: MSE and SD of errors of the mean vector and scatter matrix estimates for uncontaminated data: simulation scenario with $p = 10$, and $bdp = 0.0502$.



(a)



(b)

Figure 6.4: MSE and SD of errors of the mean vector and scatter matrix estimates for uncontaminated data: simulation scenario with $p = 15$, and $bdp = 0.0345$.

As the dimensionality increases, see Figures 6.3 and 6.4, where $p = 10$ and $p = 15$ respectively, the average error for all methods slightly increases. This increase is expected as the methods lose efficiency with higher dimensions. Despite this, the MDPD method continues to perform on par with the other robust methods. Importantly, it maintains the advantage of lower variability, which is crucial for high-dimensional data analysis where stability is a

significant concern. On the other hand, the DetMCD method shows increased variability, which might affect its reliability in higher dimensions.

6.2.2 Point Contamination

The performance of the minimum divergence power divergence (MDPD) method generally exhibits superiority in scatter estimation (e_{Σ}) when compared to FastMCD and DetMCD, particularly in scenarios with varying levels of point contamination. As observed in Figures 6.5 to 6.8, MDPD maintains lower Mean Squared Error (MSE) and Kullback-Leibler (KL) divergence for scatter estimates across different dimensions and sample sizes. This trend is supported by the data in Table 6.1, which shows that MDPD consistently demonstrates lower standard deviations for scatter estimation (e_{Σ}), indicating higher robustness and stability against contamination.

For instance, in Figure 6.5 ($p = 2, n = 100$) and Figure 6.6 ($p = 5, n = 200$), MDPD shows lower variability and more stable performance in scatter estimation as the number of contamination points (ncp) increases. This trend continues in Figure 6.7 ($p = 10, n = 500$) and Figure 6.8 ($p = 15, n = 1000$), where MDPD retains its efficiency even in larger sample sizes and higher dimensions, maintaining its lead over FastMCD and DetMCD in terms of scatter estimation accuracy and consistency.

However, for mean vector estimation (e_{μ}) MDPD's performance is more variable. FastMCD and DetMCD often show better results, particularly in scenarios with higher contamination levels. As evidenced in Figures 6.6 and 6.7, FastMCD and DetMCD exhibit lower MSE and KL divergence for mean vector estimates in these cases. Table 6.1 confirms these findings, showing that FastMCD and DetMCD have lower standard deviations for mean vector estimation showing that FastMCD and DetMCD have lower standard deviations for mean vector estimation. However, as the n increases in each scenario, performance is at par.

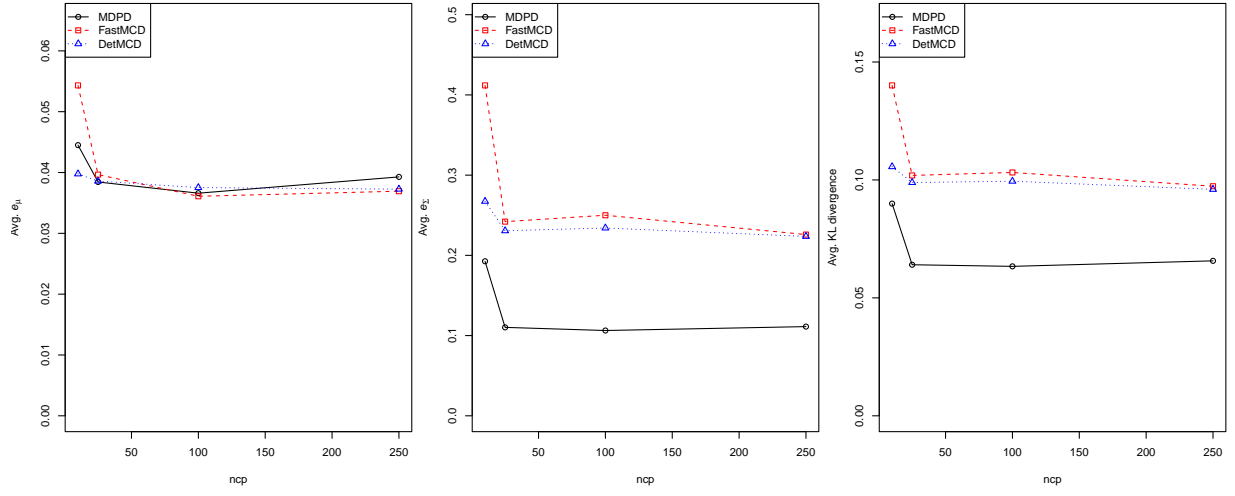


Figure 6.5: MSE and Kullback–Leibler divergence of the mean vector and scatter matrix estimates for point contamination in simulation scenarios: $p = 2$, $n = 100$, $\varepsilon = 12.5\%$, and $bdp = 0.25$.

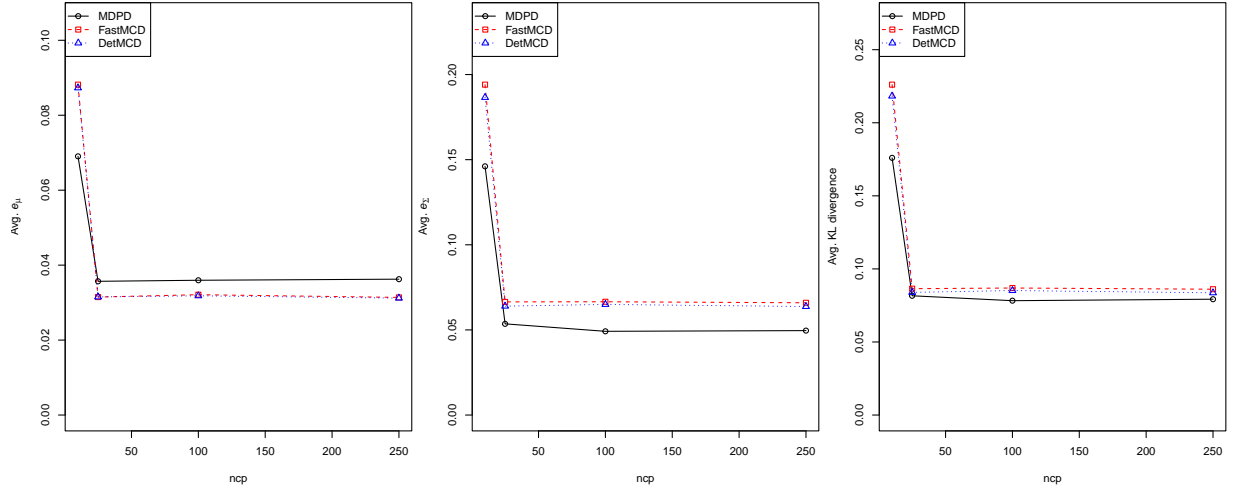


Figure 6.6: MSE and Kullback–Leibler divergence of the mean vector and scatter matrix estimates for point contamination in simulation scenarios: $p = 5$, $n = 200$, $\varepsilon = 6.2\%$, and $bdp = 0.12$.

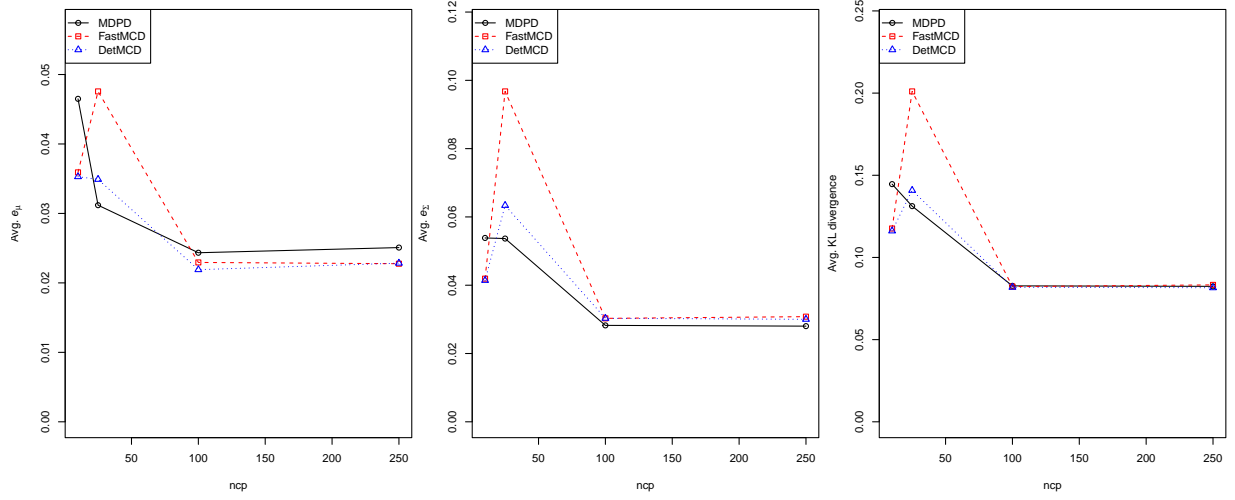


Figure 6.7: MSE and Kullback–Leibler divergence of the mean vector and scatter matrix estimates for point contamination in simulation scenarios: $p = 10$, $n = 500$, $\varepsilon = 3.4\%$, and $bdp = 0.067$.

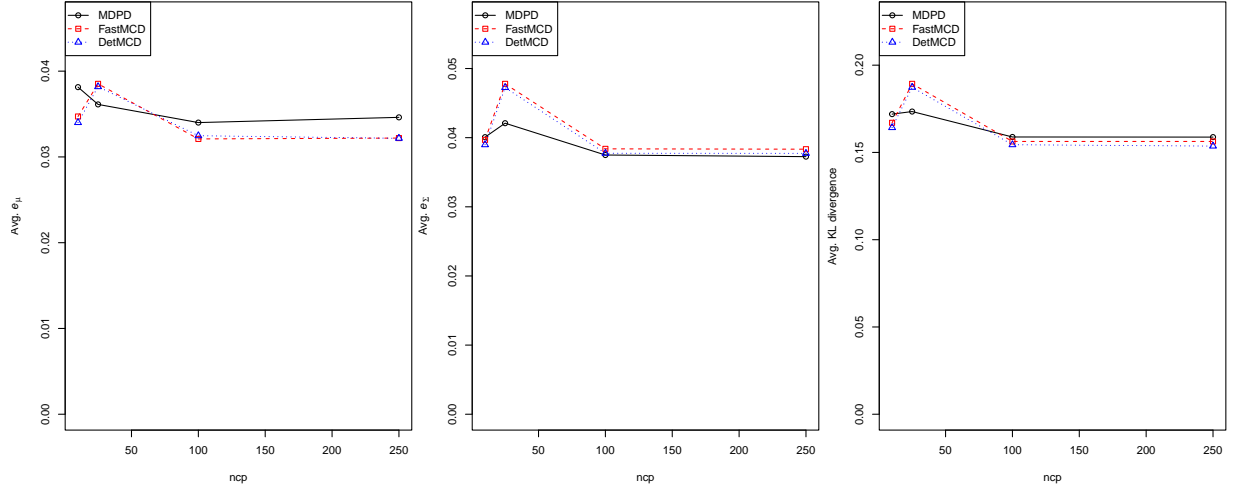


Figure 6.8: MSE and Kullback–Leibler divergence of the mean vector and scatter matrix estimates for point contamination in simulation scenarios: $p = 15$, $n = 1000$, $\varepsilon = 1.5\%$, and $bdp = 0.046$.

In summary, when matched for bdp , MDPD demonstrates clear superiority in scatter estimation, consistently outperforming FastMCD and DetMCD in terms of lower MSE, KL

divergence, and variability. This makes MDPD the preferred method for robust scatter estimation under point contamination. Conversely, for mean vector estimation, FastMCD and DetMCD exhibit better performance and less variability, especially at higher contamination levels, indicating their robustness and reliability in these scenarios. Thus, while MDPD excels in scatter estimation, FastMCD and DetMCD are more suitable for mean vector estimation in contaminated datasets.

Table 6.1: Standard Deviation of Errors for Point Contamination.

n	p	bdp	ε	ncp	e_μ			e_Σ			e_{KL}		
					MDPD	FastM	DetM	MDPD	FastM	DetM	MDPD	FastM	DetM
100	2	0.25	0.125	10	0.0472	0.094	0.0547	0.3013	0.976	0.5142	0.0876	0.2203	0.1201
				25	0.037	0.039	0.0377	0.1171	0.2092	0.1838	0.0439	0.0621	0.0575
				100	0.035	0.0357	0.0359	0.1011	0.2029	0.1831	0.0427	0.0609	0.0555
				250	0.039	0.0367	0.0382	0.1044	0.1751	0.1823	0.0435	0.0542	0.0568
200	5	0.1232	0.0616	10	0.0402	0.0403	0.0409	0.0631	0.0494	0.0489	0.0582	0.0384	0.0397
				25	0.0221	0.0198	0.0205	0.0233	0.0277	0.0262	0.027	0.028	0.0266
				100	0.0212	0.0199	0.0189	0.0193	0.0265	0.0261	0.0243	0.0267	0.0267
				250	0.0232	0.0204	0.02	0.0194	0.0272	0.0252	0.0252	0.0276	0.0257
500	10	0.067	0.0335	10	0.0178	0.0147	0.0155	0.0101	0.0076	0.0075	0.0226	0.0178	0.018
				25	0.0138	0.0226	0.0225	0.0163	0.0386	0.045	0.0298	0.0681	0.0808
				100	0.0106	0.01	0.0096	0.0056	0.0061	0.0063	0.0147	0.0139	0.0146
				250	0.0109	0.0096	0.0099	0.0053	0.0064	0.0059	0.0137	0.0146	0.0139
1000	15	0.046	0.0153	10	0.0079	0.0069	0.0066	0.0028	0.0026	0.0025	0.0107	0.0099	0.0098
				25	0.0073	0.0081	0.0078	0.0038	0.0035	0.0035	0.0127	0.0114	0.0112
				100	0.0064	0.0059	0.006	0.0024	0.0025	0.0024	0.0093	0.0093	0.009
				250	0.0061	0.0061	0.0059	0.0024	0.0025	0.0025	0.0094	0.0093	0.0093

6.2.3 Cluster Contamination

The performance of the Mean Divergence Power Divergence (MDPD) method under cluster contamination scenarios generally mirrors its performance under point contamination. In both scenarios, MDPD demonstrates superiority in scatter estimation (e_Σ), consistently

maintaining lower Mean Squared Error (MSE) and Kullback-Leibler (KL) divergence compared to FastMCD and DetMCD. This is evidenced in Figures 6.9 to 6.12 for cluster contamination and Figures 6.5 to 6.8 for point contamination, as well as in Tables 6.1 and 6.2, which show lower standard deviations for scatter estimation for (e_{Σ}) MDPD.

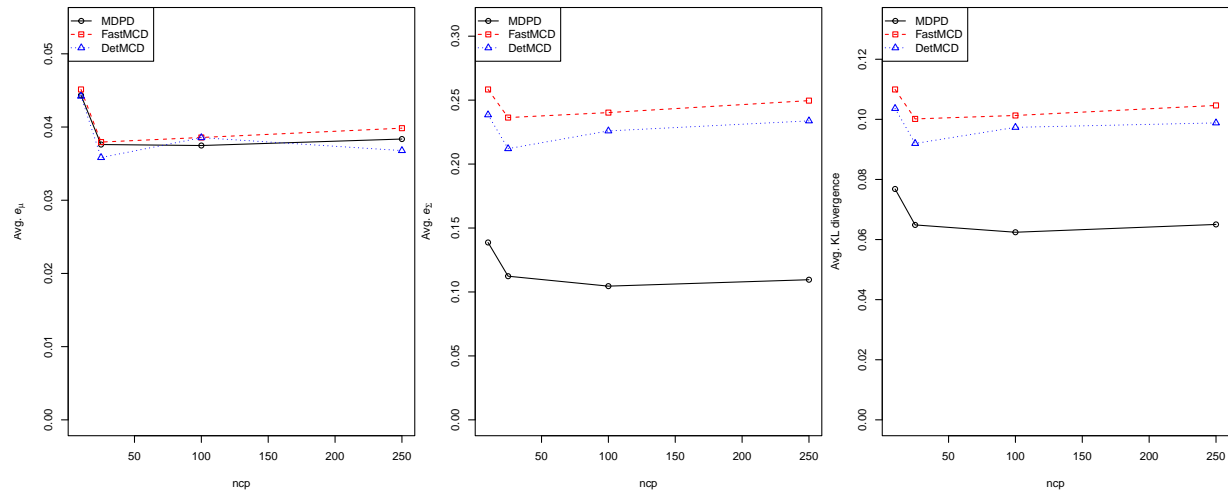


Figure 6.9: MSE and Kullback–Leibler divergence of the mean vector and scatter matrix estimates for cluster contamination in simulation scenarios: $p = 2$, $n = 100$, $\varepsilon = 12.5\%$, and $bdp = 0.25$.

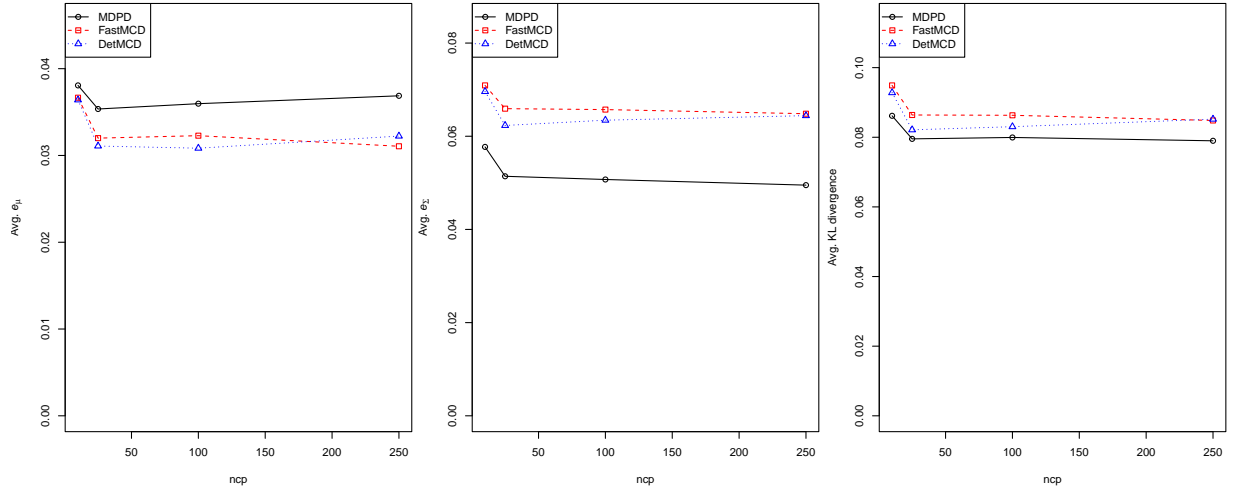


Figure 6.10: MSE and Kullback–Leibler divergence of the mean vector and scatter matrix estimates for cluster contamination in simulation scenarios: $p = 5$, $n = 200$, $\varepsilon = 6.2\%$, and $bdp = 0.12$ (cluster contamination).

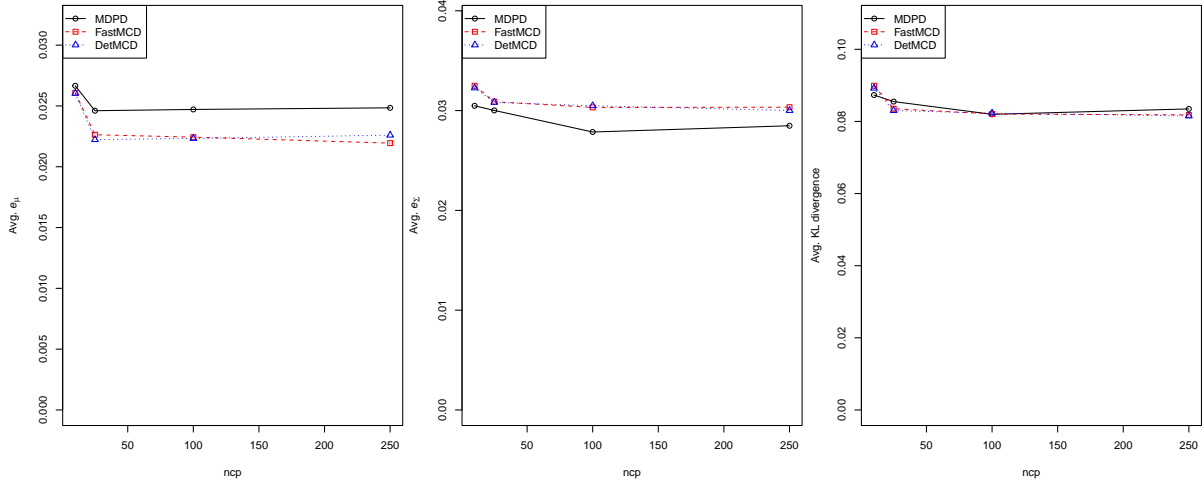


Figure 6.11: MSE and Kullback–Leibler divergence of the mean vector and scatter matrix estimates for cluster contamination in simulation scenarios: $p = 10$, $n = 500$, $\varepsilon = 3.4\%$, and $bdp = 0.067$ (cluster contamination).

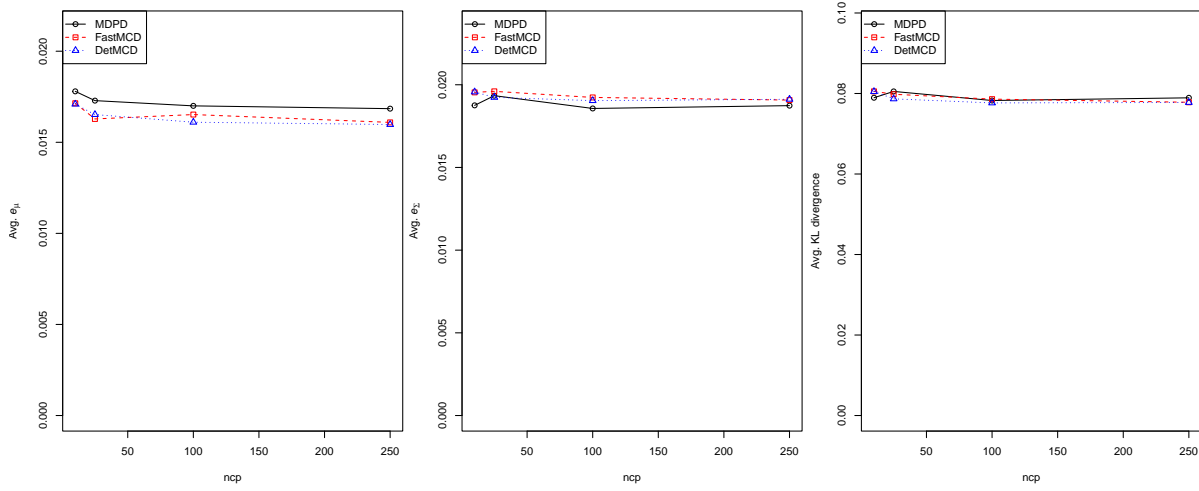


Figure 6.12: MSE and Kullback–Leibler divergence of the mean vector and scatter matrix estimates for cluster contamination in simulation scenarios: $p = 15$, $n = 1000$, $\varepsilon = 1.5\%$, and $bdp = 0.046$ (cluster contamination).

Similarly, for mean vector estimation (e_{μ}), FastMCD and DetMCD generally perform better than MDPD both point and cluster contamination scenarios. FastMCD and DetMCD consistently exhibit lower MSE and KL divergence for mean vector estimates, particularly at higher contamination levels, as shown in both sets of figures and tables.

However, a notable difference under cluster contamination is the slightly increased variability in scatter estimation (e_{Σ}) across all methods. This increased variability is reflected in the higher standard deviations observed in Table 6.2 compared to Table 6.1. Despite this increase, MDPD maintains its relative advantage over FastMCD and DetMCD in scatter estimation.

Table 6.2: Standard Deviation of Errors for Cluster Contamination.

n	p	bdp	eps	ncp	e_μ			e_Σ			e_{KL}		
					MDPD	FastM	DetM	MDPD	FastM	DetM	MDPD	FastM	DetM
100	2	0.25	0.125	10	0.0451	0.0441	0.0477	0.1363	0.2271	0.2222	0.0528	0.0702	0.0684
				25	0.0362	0.035	0.0354	0.1075	0.1983	0.168	0.0424	0.058	0.0532
				100	0.0373	0.0381	0.0374	0.1005	0.21	0.1811	0.0408	0.0614	0.0555
				250	0.0391	0.0401	0.0368	0.0971	0.2018	0.1867	0.0414	0.0602	0.0578
200	5	0.1232	0.0616	10	0.0242	0.0236	0.0238	0.0246	0.0302	0.03	0.0281	0.0314	0.0312
				25	0.0218	0.0208	0.0195	0.021	0.0259	0.0255	0.0249	0.0266	0.0257
				100	0.0233	0.0204	0.0197	0.0199	0.0263	0.025	0.0253	0.0267	0.0255
				250	0.0229	0.0188	0.0195	0.0196	0.0264	0.0249	0.0252	0.0262	0.0256
500	10	0.067	0.0335	10	0.0117	0.0111	0.0118	0.0063	0.0066	0.0068	0.0152	0.0156	0.0159
				25	0.0107	0.0101	0.0099	0.0063	0.0064	0.0064	0.0152	0.0146	0.0145
				100	0.0109	0.0106	0.0102	0.0054	0.0064	0.0062	0.0141	0.015	0.0142
				250	0.0113	0.0096	0.0102	0.0057	0.0062	0.006	0.0146	0.0138	0.014
1000	15	0.046	0.0153	10	0.0069	0.0065	0.0062	0.0024	0.0026	0.0026	0.0095	0.0096	0.0094
				25	0.0062	0.0061	0.0059	0.0024	0.0026	0.0024	0.0091	0.0095	0.0089
				100	0.0062	0.006	0.0058	0.0024	0.0026	0.0025	0.0091	0.0096	0.0091
				250	0.0061	0.0058	0.0059	0.0024	0.0025	0.0025	0.0094	0.0093	0.0092

The rest of the results under various scenarios are presented in Appendix B across several tables. Similar observations can be made, and it is also noted that as the sample size increases, the MSE of the (e_Σ) becomes comparable to that of FastMCD and DetMCD.

6.3 Empirical Breakdown Point

The breakdown point (bdp) is an important measure in robust statistics, representing the proportion of contamination an estimator can withstand before becoming unreliable. Given that algorithmic implementations of robust estimators often display lower than expected breakdown points, we empirically verify the nominal breakdown point as outlined in Section 4.2.

We conducted an extensive simulation study to assess the MDPD estimator under point contamination. Sample sizes considered were $n = 30, 50, 100, 200, 500, 1000$ and dimensions were $p = 2, 5, 10, 15$. Each combination of n and p involved 10,000 replications. Point contamination was introduced at a level $\delta = 0.001$, with a non-central parameter $ncp = 10,000$. The breakdown point threshold was set at $\varepsilon = 0.001$.

To evaluate if the MDPD estimator breaks down under point contamination, we employed the following condition. Let \mathbf{X} be the data matrix with n rows and p columns, where the first n_{out} rows are contaminated. The MDPD estimator is applied to this contaminated data.

Define:

- $\mathbf{X}_{\text{clean}} = \mathbf{X}[-(1 : n_{\text{out}}),]$, the uncontaminated portion of the data.
- $\bar{\mathbf{x}}$, the mean vector of $\mathbf{X}_{\text{clean}}$.
- \mathbf{S} , the covariance matrix of $\mathbf{X}_{\text{clean}}$.
- d_i , the Mahalanobis distance for the uncontaminated data points.

The Mahalanobis distance for each point in $\mathbf{X}_{\text{clean}}$ is given by:

$$d_i = (\mathbf{x}_i - \bar{\mathbf{x}})^T \mathbf{S}^{-1} (\mathbf{x}_i - \bar{\mathbf{x}}).$$

The average Mahalanobis distance for the uncontaminated data is:

$$\bar{d} = \frac{1}{n - n_{\text{out}}} \sum_{i=1}^{n - n_{\text{out}}} d_i.$$

The breakdown condition is assessed by comparing the minimum Mahalanobis distance of the contaminated data $\mathbf{X}_{\text{contaminated}}$ to 90% of the average Mahalanobis distance of the uncontaminated data:

$$\min (d_j^{\text{contaminated}}) \leq 0.90 \times \bar{d},$$

where $d_j^{\text{contaminated}}$ are the Mahalanobis distances of the contaminated points.

This condition is effective because the Mahalanobis distance is highly sensitive to outliers, making it a powerful tool for detecting the influence of contaminated data points. By comparing the minimum Mahalanobis distance of the contaminated points to 90% of the mean distance of the uncontaminated points, we establish a robust threshold that adapts to the overall scale and variability of the uncontaminated data. If the smallest Mahalanobis distance among the contaminated points is less than this threshold, it indicates that the contaminated points are not being flagged as outliers, suggesting that the estimator has broken down. The factor of 0.90 balances robustness and sensitivity, ensuring the estimator's robustness is adequately tested without being overly sensitive to minor deviations. Repeating the simulation 10,000 times for each combination of n and p ensures stable and reliable breakdown point estimation, averaging out the variability due to random noise.

The proportion of times the breakdown condition is satisfied over the total number of replications (10,000) provides the breakdown percentage. The empirical bdp is determined as the smallest proportion of contamination $\frac{n_{\text{out}}}{n}$ at which the breakdown percentage exceeds the threshold. Table 6.3 shows the results of the empirical breakdown point based on point contamination.

p / n	30	50	100	200	500	1000
2	0.3000	0.2800	0.2800	0.2850	0.2840	0.2850
5	0.1667	0.1600	0.1500	0.1500	0.1480	0.1490
10	0.1000	0.1000	0.0900	0.0850	0.0840	0.0840
15	0.0667	0.0600	0.0600	0.0600	0.0600	0.0590

Table 6.3: Empirical Breakdown Points for Different Sample Sizes and Dimensions.

Chapter 7

Real Data Analysis with MDPD Estimators

In this chapter, we showcase practical applicability and efficacy of the proposed minimum density power divergence (MDPD) estimator method through analysis of real-world datasets. We compare our method against the classical unbiased MLE and the minimum covariance determinant (MCD) estimator. We consider two distinct examples: Principal Component Analysis (PCA) with the Swiss Banknote dataset (see [Flury and Riedwyl, 1988](#)) and Multivariate Regression with the Pulp Fiber and Paper dataset (see [Whiting et al., 2018](#)).

7.1 PCA Diagnostics

The first dataset in this research is the Swiss Banknote dataset which has been a subject of interest in previous studies by [Flury and Riedwyl \(1988\)](#) and [Riani et al. \(2009\)](#) and was further used by [Pokojoy and Jobe \(2022\)](#) to benchmark their procedures against the state-of-art procedures of FastMCD and DetMCD. Specifically, 100 banknotes are genuine, and the remaining 100 are counterfeit. Six different variables were measured for each banknote, furnishing a multi-dimensional dataset. These variables capture various characteristics or attributes of the banknotes, and they serve as the basis for analysis in this research. Studies by [Willems et al. \(2009\)](#), [Pison and Van Aelst \(2004\)](#), [Salibián-Barrera et al. \(2006\)](#) and [Hubert et al. \(2012\)](#) reveal that the 100 counterfeit Swiss bank notes contains several

outlying observations and are highly correlated. Figure 7.1 shows the boxplots of counterfeit Swiss banknotes, illustrating their distribution. The boxplot analysis reveals that the banknotes' length, left, right, and diagonal dimensions are tightly clustered, indicating uniformity in these measurements. However, the bottom and top dimensions exhibit a broader distribution with potential outliers, suggesting more variability and anomalies. This variability in the bottom and top values is the most notable observation, hinting at possible distinctions or inconsistencies in those dimensions of the banknotes.

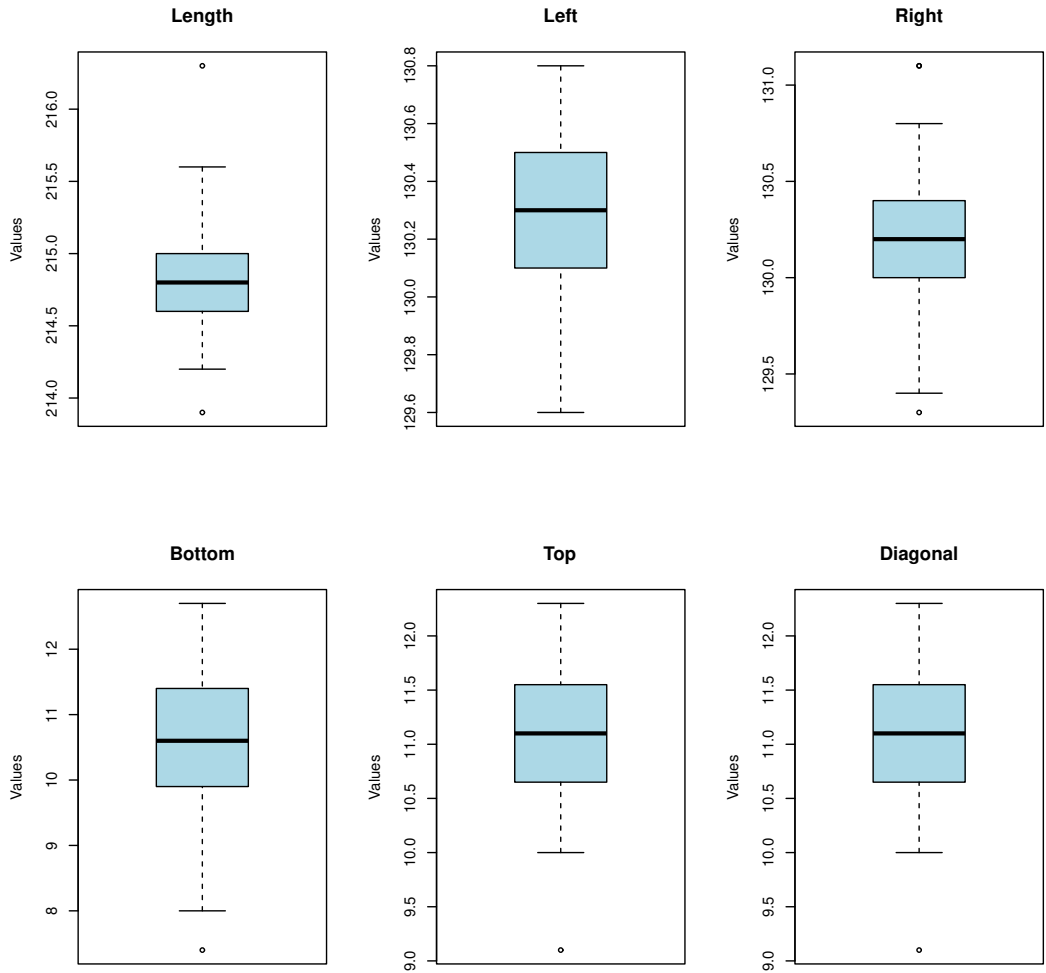


Figure 7.1: Boxplots for the counterfeit part of the Swiss Banknote dataset.

The boxplot analysis of the six variables for the counterfeit part of the Swiss banknote dataset reveals that outliers may be present. These outliers indicate that there are a few data points significantly different from the majority, which could suggest a potential presence measurement errors, data entry errors, or inherent variability in these specific features of the banknotes.

Principal component analysis (PCA) is a widely used statistical method that reduces high-dimensional data to a few components, which are linear combinations of the original variables. This reduction helps interpret and understand the main sources of variation in the data. PCA is commonly applied in fields like computer vision, chemometrics, and genetics, often serving as a preliminary step before further multivariate analysis techniques. The first principal component corresponds to the direction in which the projected observations have the largest variance, while the second component is orthogonal to the first and also maximizes the variance of the projected data points. Classical PCA based on the empirical scatter matrix is known to be highly sensitive to outlying observations. The first components are often attracted to outlying points, which may prevent them from capturing the variation of regular observations (Hubert et al., 2005). Therefore, following Croux and Haesbroeck (2000) and Hubert et al. (2005), we perform robust PCA on the dataset using the eigenvalues and eigenvectors of the robust estimator of the covariance matrix. The goal is to find linear combinations of the original variables that contain most of the information, even in the presence of outliers, and to flag outliers and determine their type using diagnostic plots proposed by Pison and Van Aelst (2004).

In the PCA analysis, we set the tuning parameter to $\alpha = 0.25$ (corresponding to a nominal breakdown point of 10.24%) for the MDPD and matched the FastMCD(bdp = 10.24%) to a breakdown point of 10.24%. The first three principal components are retained in our analysis. The robust MDPD PCA's first three principal components account for 92% of the total variation, the robust MCD PCA accounts for 90%, and the classical PCA method accounts for 90%. Figure 7.2 is a scree plot the cumulative explained variance obtained

from the eigenvalues of respective scatter estimates.

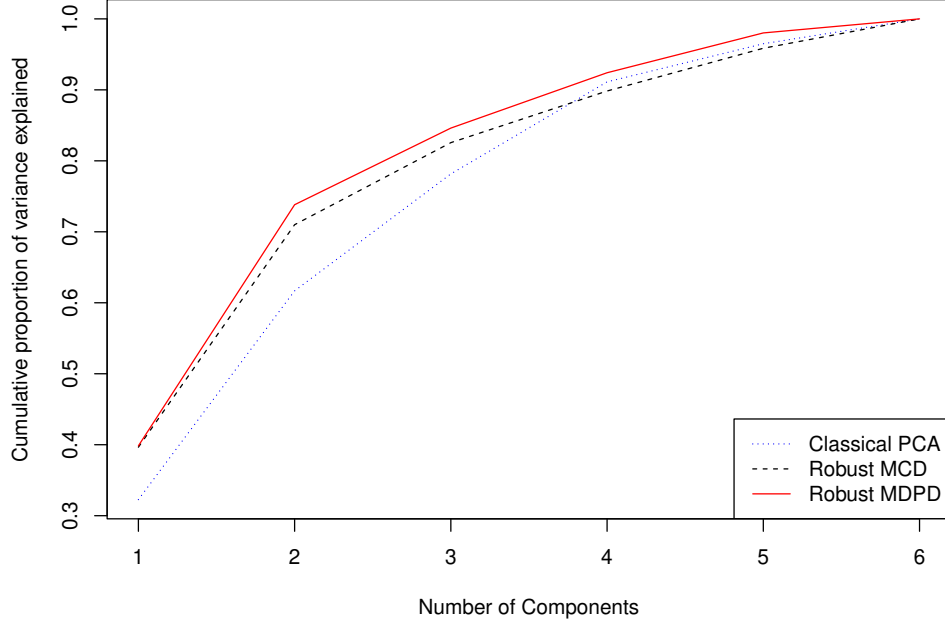


Figure 7.2: Explained variance by the classical and robust PCA

To identify which outliers significantly influence the PCA analysis, we employ the diagnostic plot introduced by [Pison and Van Aelst \(2004\)](#). This plot effectively visualizes the empirical influence of each observation on the first k eigenvectors retained for analysis against their robust distance, calculated using robust estimates of location and scatter. In Figures 7.3 and 7.4, the Mahalanobis distance or robust distance of each observation is displayed on the x -axis, while the overall empirical influence on the first $k = 3$ eigenvectors is plotted on the y -axis. Formally, the overall empirical influence of observation \mathbf{x}_i are given by:

$$EIF_k(\mathbf{x}_i, l) = \sqrt{\frac{1}{k} \sum_{j=1}^k (\tilde{z}_j^2 - \lambda_j \mathbf{e}_{\tilde{\mathbf{x}}}^t \mathbf{D}_{\tilde{\mathbf{x}}} \mathbf{e}_j)^2} \quad (7.1)$$

$$EIF_k(\mathbf{x}_i, v) = \sqrt{\frac{1}{kp} \sum_{j=1}^k \sum_{l=1}^p \left(\sum_{k=1, k \neq j}^p \left(\tilde{z}_k \tilde{z}_j - \frac{\lambda_k + \lambda_j}{2} \mathbf{e}_j^t \mathbf{D}_{\tilde{\mathbf{x}}} \mathbf{e}_k \right) \frac{\mathbf{e}_k}{\lambda_j - \lambda_k} \right)^2} \quad (7.2)$$

where:

- $EIF_k(\mathbf{x}_i, l)$ is the empirical influence function for the eigenvalues.
- $EIF_k(\mathbf{x}_i, v)$ is the empirical influence function for the eigenvectors.
- $\tilde{z}_j = \mathbf{e}_j^T \tilde{\mathbf{x}}$, where $\tilde{\mathbf{x}} = \Sigma_D^{-1/2}(\mathbf{x} - \boldsymbol{\mu})$ is the standardized observation.
- λ_j and λ_k are the eigenvalues of the covariance matrix Σ .
- \mathbf{e}_j and \mathbf{e}_k are the eigenvectors of the covariance matrix Σ .
- $\mathbf{D}_{\tilde{\mathbf{x}}} = \text{diag}(\tilde{\mathbf{x}}\tilde{\mathbf{x}}^T)$ is a diagonal matrix with elements derived from the standardized observation $\tilde{\mathbf{x}}$.
- $\sum_{k=1, k \neq j}^p$ indicates a summation over all p components except j .

Additionally, the Mahalanobis distance, which measures the distance between a point and a distribution, is given by:

$$D(\mathbf{x}_i, \boldsymbol{\mu}, \Sigma) = \sqrt{(\mathbf{x}_i - \boldsymbol{\mu})^T \Sigma^{-1} (\mathbf{x}_i - \boldsymbol{\mu})}$$

This approach follows the methodology detailed by [Critchley \(1985\)](#) in the context of Principal Component Analysis.

The graph is segmented into four quadrants by horizontal and vertical lines that represent cutoff values. The empirical influence cutoff is determined through Monte Carlo simulation, as outlined by [Pison and Van Aelst \(2004\)](#), corresponding to the 97.5% quantile of the overall empirical influences. The robust distance cutoff is the usual $\sqrt{\chi_{6,0.975}^2} = 3.80$. These quadrants classify observations into regular points, non-outlying influential points, non-influential outliers, and influential outliers. This classification is essential for diagnosing the data, as it helps identify which observations have a substantial impact on the PCA results. The influence on the eigenvectors is particularly critical, as it indicates whether outliers are altering the directions of the principal components, thereby affecting the interpretation and accuracy of the PCA.

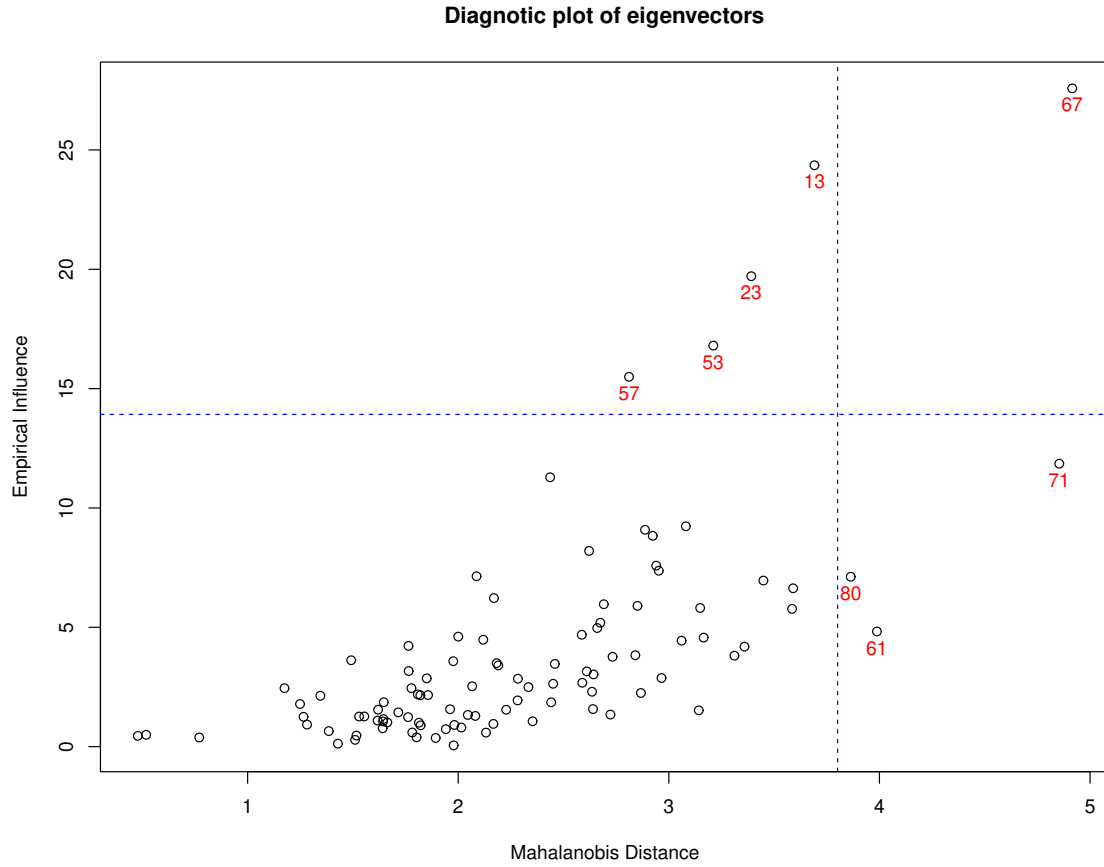


Figure 7.3: Overall empirical influence for the eigenvectors versus the Mahalanobis distance based on classical scatter matrix for the counterfeit notes in the Swiss banknote dataset.

In Figure 7.3, we observe that 4 data points are flagged as non-outlying influential, 3 are flagged as non-influential outliers, and only 1 data point is flagged as an influential outlier for the eigenvectors. This comparison to the robust counterparts reveals that potential influential outliers are masked in the classical PCA method, making it less reliable for PCA analysis when outliers are present. The classical PCA method's inability to effectively identify and manage influential outliers underscores the importance of using robust PCA methods to ensure accurate and resilient multivariate data analysis in the presence of outliers.

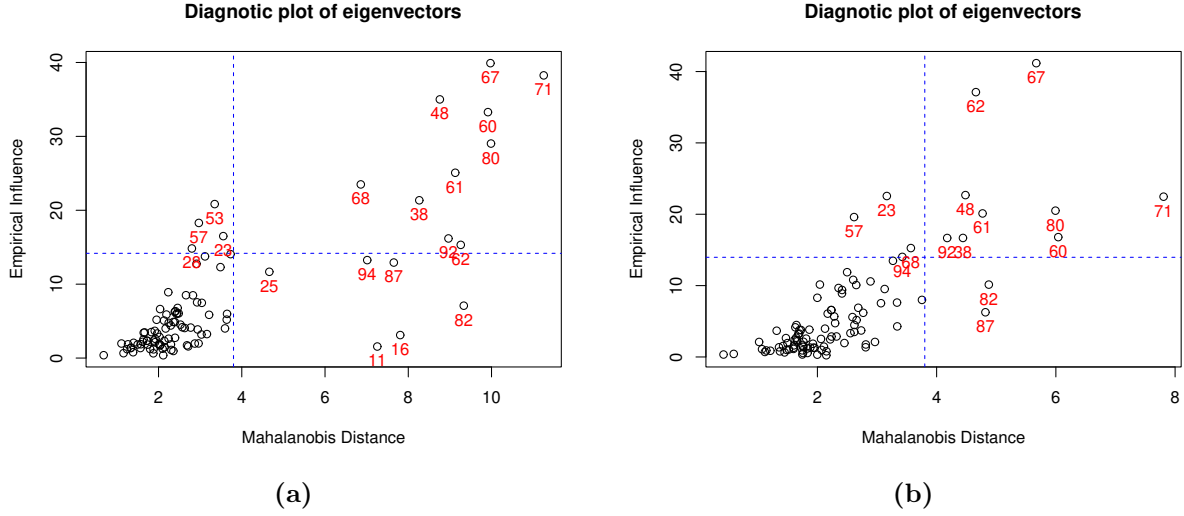


Figure 7.4: Diagnostic Plot: Overall empirical influence for the eigenvectors versus the robust distance based on MDPD (Panel a) and FastMCD (Panel b) for the counterfeit notes in the Swiss banknote dataset.

The robust PCA diagnostic in Figures 7.4(a) and 7.4(b), based on the empirical influence of the eigenvector, presents a completely distinct results compared to the classical PCA diagnostic. Figure 7.4(a) shows that using the robust MDPD method, 10 observations are flagged as influential outliers for the eigenvectors¹, with observations 71, 67, 60, and 48 having very large influence. Figure 7.4(b) is based on the robust FastMCD method, identifying 9 observations as influential outliers for the eigenvectors, with observation 67 having very large influence. These results suggest that, when matched for their breakdown point (bdp), the robust MDPD method and the robust FastMCD method perform similarly in identifying influential outliers. However, the MDPD method is less computationally intensive, making it a more efficient choice.

¹The outlier map of the Swiss bank notes studied in [Hubert et al. \(2012\)](#) reveals the same observations as influential outliers, comparing the score distance to the orthogonal distance.

7.2 Multivariate Regression Diagnostics

The second dataset we consider is the Pulp Fiber and Paper dataset, provided by [Whiting et al. \(2018\)](#). It offers a practical example of applying robust multivariate regression techniques. This dataset includes measurements of properties of pulp fibers and the paper made from them. The aim is to investigate the relationships between pulp fiber properties and the resulting paper properties. The dataset consists of 62 measurements of the following four pulp fiber characteristics: arithmetic fiber length, long fiber fraction, fine fiber fraction, and zero span tensile. Additionally, four paper properties were measured: breaking length, elastic modulus, stress at failure, and burst strength. This analysis can be performed with multivariate linear regression. In Figure 7.5 the density plot of the predictor variables are presented.

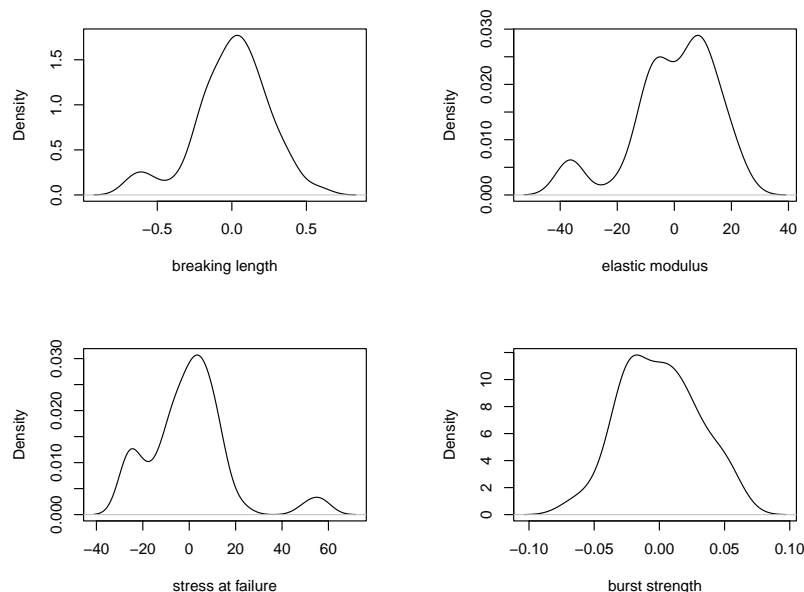


Figure 7.5: Density plot of the predictor variables.

Multivariate linear regression involves modeling the relationship between multiple predictors $\mathbf{x} = (x_1, \dots, x_p)^T$ and multiple responses $\mathbf{y} = (y_1, \dots, y_q)^T$. The model can be ex-

pressed as:

$$\mathbf{y} = \mathbf{B}^T \mathbf{x} + \boldsymbol{\alpha} + \boldsymbol{\varepsilon}, \quad (7.3)$$

where:

- \mathbf{B} is the $(p \times q)$ slope matrix, $\boldsymbol{\alpha}$ is the q -dimensional intercept vector,
- $\boldsymbol{\varepsilon} = (\varepsilon_1, \dots, \varepsilon_q)^T$ are the errors, assumed to be independently and identically distributed with mean $\mathbf{0}$ and covariance matrix $\boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}}$.

The joint distribution of (\mathbf{x}, \mathbf{y}) is assumed normal and is characterized by the mean vector $\boldsymbol{\mu}$ and scatter matrix $\boldsymbol{\Sigma}$, which can be partitioned as:

$$\boldsymbol{\mu} = \begin{pmatrix} \boldsymbol{\mu}_x \\ \boldsymbol{\mu}_y \end{pmatrix}, \quad \boldsymbol{\Sigma} = \begin{pmatrix} \boldsymbol{\Sigma}_{xx} & \boldsymbol{\Sigma}_{xy} \\ \boldsymbol{\Sigma}_{yx} & \boldsymbol{\Sigma}_{yy} \end{pmatrix}.$$

Traditionally, the mean vector $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$ are estimated by their empirical counterparts $\hat{\boldsymbol{\mu}}$ and $\hat{\boldsymbol{\Sigma}}$. The least squares estimators of the regression parameters can be derived from these estimates (see [Johnson and Wichern, 1998](#) and [Rousseeuw et al., 2004](#)):

$$\begin{aligned} \hat{\boldsymbol{\beta}} &= \hat{\boldsymbol{\Sigma}}_{xx}^{-1} \hat{\boldsymbol{\Sigma}}_{xy}, & \hat{\boldsymbol{\alpha}} &= \hat{\boldsymbol{\mu}}_y - \hat{\boldsymbol{\beta}}^T \hat{\boldsymbol{\mu}}_x, \\ \hat{\boldsymbol{\Sigma}}_{\boldsymbol{\varepsilon}} &= \hat{\boldsymbol{\Sigma}}_{yy} - \hat{\boldsymbol{\beta}}^T \hat{\boldsymbol{\Sigma}}_{xy} \hat{\boldsymbol{\beta}}, & \mathbf{r}_i &= \mathbf{y}_i - \hat{\boldsymbol{\beta}}^T \mathbf{x}_i - \hat{\boldsymbol{\alpha}}, \end{aligned} \quad (7.4)$$

The Mahalanobis distances for the residuals and the individual observation for the diagnostic plot (combines information on regression outliers and leverage points) are given by

$$d(\mathbf{r}_i) := \sqrt{\mathbf{r}_i^T (\hat{\boldsymbol{\Sigma}}_{\boldsymbol{\varepsilon}})^{-1} \mathbf{r}_i}, \quad d(\mathbf{x}_i) := \sqrt{(\mathbf{x}_i - \boldsymbol{\mu}_x)^T (\hat{\boldsymbol{\Sigma}}_{xx}^{-1})^{-1} (\mathbf{x}_i - \boldsymbol{\mu}_x)}. \quad (7.5)$$

Multivariate regression is widely used across various fields such as chemometrics, engineering, econometrics, psychometrics, and more. It allows for simultaneous modeling of multiple outcomes, providing insights into the relationships among predictors and responses. Recent advancements in multivariate regression have been explored in various studies, contributing to the understanding and application of this methodology in complex data analysis scenarios (see [Johnson and Wichern, 1998](#), [Barrett and Ling, 1992](#), [Breiman and Friedman,](#)

1997, Cook and Setodji, 2003 and Gleser, 1992). It is however well known that classical multiple regression is heavily sensitive to outliers in the data. This also applies to multivariate regression. Therefore, Rousseeuw et al. (2004) proposed using robust estimators for the center $\boldsymbol{\mu}$ and scatter matrix $\boldsymbol{\Sigma}$ in Equation (7.4) to construct a robust multivariate regression method that has the equivariance properties required for a multivariate regression estimator.

For robust multivariate regression method, using the MDPD estimators and the MCD estimators, the estimation of the coefficients, begins by combining the predictor variables \mathbf{X} (an $n \times p$ matrix) and response variables \mathbf{Y} (an $n \times q$ matrix) into a single dataset \mathbf{Z} , where \mathbf{Z} is an $n \times (p + q)$ matrix.

$$\mathbf{Z} = \begin{pmatrix} \mathbf{X} & \mathbf{Y} \end{pmatrix},$$

where \mathbf{Z} is an $n \times (p + q)$ matrix.

The robust covariance matrix $\hat{\boldsymbol{\Sigma}}_{\text{rob}}$ is then partitioned into the submatrices:

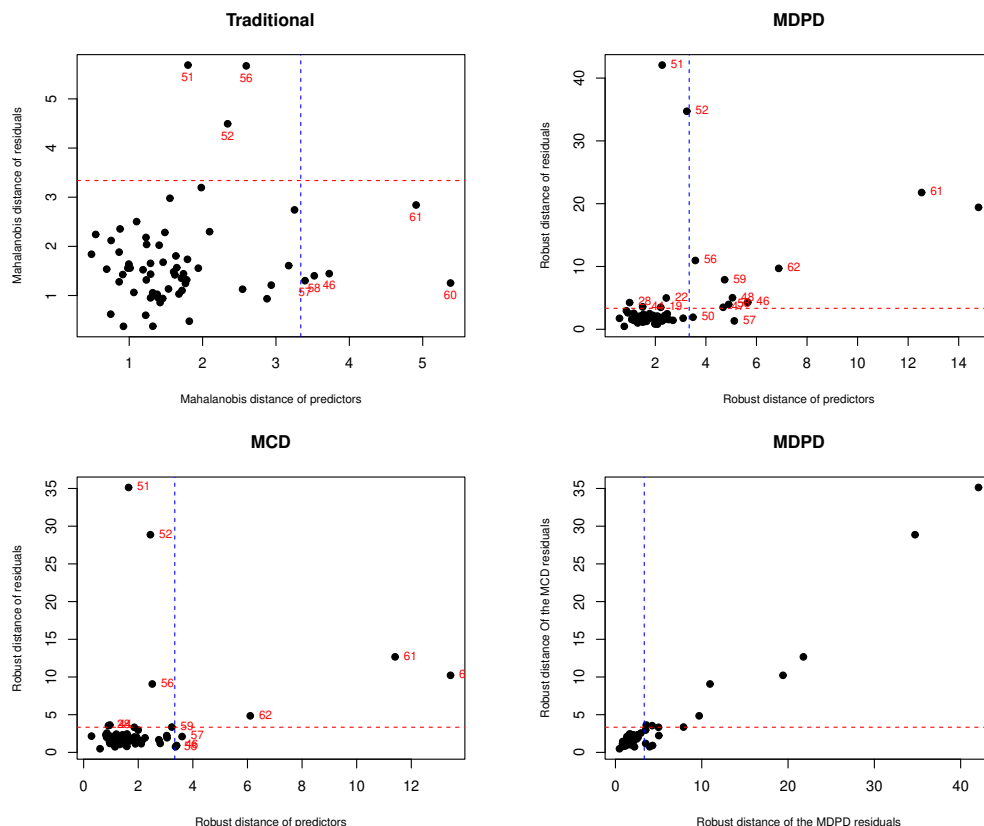
$$\hat{\boldsymbol{\Sigma}}_{\mathbf{z}} = \begin{pmatrix} \hat{\boldsymbol{\Sigma}}_{xx} & \hat{\boldsymbol{\Sigma}}_{xy} \\ \hat{\boldsymbol{\Sigma}}_{yx} & \hat{\boldsymbol{\Sigma}}_{yy} \end{pmatrix},$$

where:

- $\hat{\boldsymbol{\Sigma}}_{xx}$ is the robust estimate of the covariance matrix of \mathbf{X} ,
- $\hat{\boldsymbol{\Sigma}}_{yy}$ is the robust estimate of the covariance matrix of \mathbf{Y} ,
- $\hat{\boldsymbol{\Sigma}}_{xy}$ is the robust estimate of the covariance matrix between \mathbf{X} and \mathbf{Y} ,
- $\hat{\boldsymbol{\Sigma}}_{yx}$ is the transpose of $\hat{\boldsymbol{\Sigma}}_{xy}$.

The robust mean vectors are also estimated accordingly. Here, our focus is on the robust location vectors and scatter matrices estimated using the MCD and MDPD methods. For MCD, as an MLE, replacing classical mean vectors and covariance matrices with robust estimators suffices. However, MDPD, being an M-estimator, ideally requires minimizing

a specific criterion and iterative solutions for accurate β estimation. Despite this, our primary goal here is to further analyze the robust mean vectors and scatter matrices to better understand their practical applications and robustness against outliers.



Using the cutoff values, $\sqrt{\chi_{p,0.975}^2} = 3.34$ for both the horizontal and vertical lines, we observe that MDPD($\alpha = 0.16$, corresponding to $bdp = 0.1$) distance-distance plot flagged 9 observations as bad leverage points and MCD($bdp = 0.10$) flagged 3 as bad leverage points. However, the classical distance-distance plot did not flag any of the observation as

bad leverage. It is noteworthy that classical multivariate regression detected only three of these outliers (51, 52, and, 56) and considered four of the outliers (46, 58, 60, and, 61) to be good leverage. For the MSE of the robust methods, MDPD and MCD are head-to-head.

The results correspond to the results that were obtained in [Rousseeuw et al. \(2004\)](#). As it was investigated by [Rousseeuw et al. \(2004\)](#), the origin of the collected data reveals that all but the last four pulp samples (observations 59-62) were produced from fir wood. Moreover, it was found that many of the outlying samples were obtained using different pulping processes. For example, observation 62 is unique as it is the only sample from a chemi-thermomechanical pulping process. Observations 60 and 61 are the only samples from a solvent pulping process, while observations 51, 52, and 56 were obtained from a kraft pulping process.

Chapter 8

One-Way Robust Multivariate Analysis of Variance (MANOVA)

Multivariate analysis of variance (MANOVA) is a powerful statistical technique that extends univariate ANOVA to analyze multiple dependent variables simultaneously. It evaluates whether mean vectors of these dependent variables differ significantly across groups defined by one or more independent variables. It is particularly useful in marketing, educational research, genetics, economics, and medicine for examining group differences across multiple outcomes (see [Xu and Cui, 2008](#)). Developed by [Hotelling \(1992\)](#), MANOVA evaluates the equality of group means by examining the multivariate data, thereby accounting for the correlations between dependent variables which might be overlooked in separate univariate analyses.

Random samples are collected from each of the k populations with common covariance matrix and can be arranged as the following table:

	Sample 1 from $N_p(\boldsymbol{\mu}_1, \boldsymbol{\Sigma})$	Sample 2 from $N_p(\boldsymbol{\mu}_2, \boldsymbol{\Sigma})$...	Sample k from $N_p(\boldsymbol{\mu}_k, \boldsymbol{\Sigma})$
	\mathbf{X}_{11}	\mathbf{X}_{21}	...	\mathbf{X}_{k1}
	\mathbf{X}_{12}	\mathbf{X}_{22}	...	\mathbf{X}_{k2}
	\vdots	\vdots		\vdots
	\mathbf{X}_{1n_1}	\mathbf{X}_{2n_2}	...	\mathbf{X}_{kn_k}
Total	$\mathbf{X}_{1\cdot}$	$\mathbf{X}_{2\cdot}$...	$\mathbf{X}_{k\cdot}$
Mean	$\bar{\mathbf{X}}_{1\cdot}$	$\bar{\mathbf{X}}_{2\cdot}$...	$\bar{\mathbf{X}}_{k\cdot}$

The model for each of observation vector is given by

$$\begin{aligned}\mathbf{X}_{ij} &= \boldsymbol{\mu} + \boldsymbol{\alpha}_i + \boldsymbol{\varepsilon}_{ij} \\ &= \boldsymbol{\mu}_i + \boldsymbol{\varepsilon}_{ij}, \quad i = 1, 2, \dots, k; \quad j = 1, 2, \dots, n_i,\end{aligned}\tag{8.1}$$

\mathbf{X}_{ij} represents the j -th observation in the i -th group, where the total number of observations is $N = \sum_{i=1}^k n_i$. The term $\boldsymbol{\mu}_i$ denotes the unobserved fixed effect for the i -th group, $\boldsymbol{\mu}$ is the overall mean, and $\sum_{i=1}^k \boldsymbol{\alpha}_i = \mathbf{0}$. We assume the random errors $\boldsymbol{\varepsilon}_{ij}$ are independent $N_p(\mathbf{0}, \boldsymbol{\Sigma})$ variables.

Classical one-way MANOVA assumes multivariate normality, homogeneity of covariance matrices, and absence of outliers. Under the classical assumptions that all groups follow multivariate normal distributions, numerous test statistics have been explored in the literature, with one of the most frequently utilized being the likelihood-ratio test. The likelihood-ratio test is a powerful statistical method widely used for hypothesis testing in various fields. It compares the likelihood of the data under one hypothesis against the likelihood under an alternative hypothesis, aiding in model selection and parameter estimation (see [Morris et al., 1997](#), [Johnson and Petkau, 1995](#)). In the context of MANOVA, this test statistic is more commonly referred to as Wilks' Lambda.

In practice, traditional statistical assumptions are often violated, leading to misleading conclusions. For instance, the presence of outliers can severely affect test statistics, resulting in inflated Type I error rates or reduced power, as noted by [Huber and Ronchetti \(2011\)](#) and [Hampel et al. \(1986\)](#). To address these issues, robust statistical methods that can handle deviations from these assumptions are essential.

[Todorov and Filzmoser \(2010\)](#) proposed a robust Wilks' Lambda statistic using minimum covariance determinant estimators, demonstrating improved performance under various distributions. [Van Aelst and Willems \(2011\)](#) developed robust tests using S-estimators and MM-estimators, introducing a fast robust bootstrap method for estimating null distributions. [Finch and French \(2013\)](#) compared 16 test statistics across multiple conditions, finding that alternatives outperform standard MANOVA when assumptions are violated.

Xu and Cui (2008) presented a robustified MANOVA method for detecting differentially expressed genes in microarray data, utilizing probe-level information and permutation tests, showing improved power, especially with small group numbers.

These studies highlight the importance of robust MANOVA techniques when dealing with non-normal data, unequal covariance matrices, or outliers, offering practitioners various options for more reliable multivariate analyses. Recently, Das et al. (2022) proposed an alternative robust ANOVA Wald-test based on the Minimum Density Power Divergence (MDPD) estimator. They proved that this test follows a χ^2 distribution and demonstrated that it has substantially superior performance compared to the likelihood-based test in the presence of outliers. The Wald test is a versatile statistical tool applicable to various models, including ANOVA and MANOVA (Randall et al., 1997). Its simplicity and generality make the Wald test appealing practically. This is part of our broader effort to assess the performance of the asymptotic distribution of the estimated parameters in this study.

In this section, we propose to extend the robust ANOVA Wald-test based on MDPD by Das et al. (2022) to MANOVA. This robust approach aims to mitigate the influence of outliers and heavy-tailed distributions, providing more reliable inference in the presence of data contamination. We derive the asymptotic properties of the proposed test and examine its finite-sample performance through extensive Monte Carlo simulations. Additionally, we apply the robust MANOVA to real-world datasets and compare the results with those from classical MANOVA and other robust methods.

8.1 Parameter Estimation

Consider the density power divergence measure between the true density g and the model density $f_{\boldsymbol{\theta}}$ in Equation (4.1) and let $\boldsymbol{\theta} = (\boldsymbol{\mu}_1^T, \boldsymbol{\mu}_2^T, \dots, \boldsymbol{\mu}_k^T, \text{vec}(\boldsymbol{\Sigma})^T)^T$ denote the parameter

of the MANOVA Model 8.1. For $\alpha > 0$, the DPD measure is empirically written as

$$\hat{d}_\alpha(f_\theta, g) = \frac{1}{N} \sum_{i=1}^k \sum_{j=1}^{n_i} \int f_\theta^{1+\alpha}(\mathbf{y}_{ij}) d\mathbf{y}_{ij} - \frac{1+\alpha}{N\alpha} \sum_{i=1}^k \sum_{j=1}^{n_i} f_\theta^\alpha(\mathbf{y}_{ij}) + c(\alpha), \quad (8.2)$$

where $c(\alpha) = \frac{1}{N\alpha} \sum_{i=1}^k \sum_{j=1}^{n_i} \int g^{1+\alpha}(\mathbf{y}_{ij}) d\mathbf{y}_{ij}$ is independent of θ . Based on the results in Appendix A.2, we rewrite Equation (8.2) as

$$\hat{d}_\alpha(f_\theta, g) = (2\pi)^{-\alpha p/2} |\Sigma|^{-\alpha/2} (1+\alpha)^{-p/2} \left[1 - \frac{1+\alpha}{N\alpha} \sum_{i=1}^k \sum_{j=1}^{n_i} \exp \left\{ -\frac{\alpha}{2} (\mathbf{y}_{ij} - \boldsymbol{\mu}_i)^T \Sigma^{-1} (\mathbf{y}_{ij} - \boldsymbol{\mu}_i) \right\} \right] + c(\alpha). \quad (8.3)$$

The MDPD estimator of the θ is obtained by directly minimizing the DPD measure in Equation (8.3) or by solving the fixed point iteration (detail derivation is given in Appendix A.2) as follow:

$$\begin{aligned} \boldsymbol{\mu}_i &= \frac{\sum_{j=1}^{n_i} \mathbf{y}_{ij} \exp \left\{ -\frac{\alpha}{2} (\mathbf{y}_{ij} - \boldsymbol{\mu}_i)^T \Sigma^{-1} (\mathbf{y}_{ij} - \boldsymbol{\mu}_i) \right\}}{\sum_{j=1}^{n_i} \exp \left\{ -\frac{\alpha}{2} (\mathbf{y}_{ij} - \boldsymbol{\mu}_i)^T \Sigma^{-1} (\mathbf{y}_{ij} - \boldsymbol{\mu}_i) \right\}} \quad \text{for } i = 1, 2, \dots, k, \\ \Sigma &= \frac{\sum_{i=1}^k \sum_{j=1}^{n_i} (\mathbf{y}_{ij} - \boldsymbol{\mu}_i)(\mathbf{y}_{ij} - \boldsymbol{\mu}_i)^T \exp \left\{ -\frac{\alpha}{2} (\mathbf{y}_{ij} - \boldsymbol{\mu}_i)^T \Sigma^{-1} (\mathbf{y}_{ij} - \boldsymbol{\mu}_i) \right\}}{\sum_{i=1}^k \sum_{j=1}^{n_i} \exp \left\{ -\frac{\alpha}{2} (\mathbf{y}_{ij} - \boldsymbol{\mu}_i)^T \Sigma^{-1} (\mathbf{y}_{ij} - \boldsymbol{\mu}_i) \right\} - \frac{N\alpha}{(1+\alpha)^{\frac{p}{2}+1}}}. \end{aligned} \quad (8.4)$$

The estimation of $\boldsymbol{\mu}_i$ and Σ is given in the iterative Algorithm 2

Algorithm 2: MDPD estimators for the MANOVA model

Data: \mathbf{Y} : $N \times p$ matrix of the response variable, X : factor block indicator, α : tuning parameter

- 1 Select $\varepsilon > 0$ small, e.g., $\varepsilon := 10^{-8}$, R (maximum iteration) and choose $\alpha \in [0, 2/p]$ ($\alpha = 0$ is the classical Wald test and $\alpha = 2/p$ gives the maximum bdp)
- 2 Initialize the mean $\boldsymbol{\mu}_i^{(0)}, i = 1, 2, \dots, k$, and covariance matrix $\boldsymbol{\Sigma}^{(0)}$ using:

$$\begin{aligned}\boldsymbol{\mu}_i^{(0)} &:= \mathbf{m}_i \equiv \text{coordinate-wise mean or median of } \mathbf{Y}_i, \\ \boldsymbol{\Sigma}^{(0)} &:= \mathbf{S} \equiv \text{sample covariance matrix or MAD of } \mathbf{Y}\end{aligned}$$

for $k = 0, 1, 2, \dots, R$ **do**

- 3 Compute the updated mean $\boldsymbol{\mu}_i^{(k+1)}, i = 1, 2, \dots, k$, and covariance matrix $\boldsymbol{\Sigma}^{(k+1)}$:

$$\boldsymbol{\mu}_i^{(k+1)} = \frac{\sum_{j=1}^{n_i} \mathbf{y}_{ij} \exp\left(-\frac{\alpha}{2}(\mathbf{y}_{ij} - \boldsymbol{\mu}_i^{(k)})^T (\boldsymbol{\Sigma}^{(k)})^{-1} (\mathbf{y}_{ij} - \boldsymbol{\mu}_i^{(k)})\right)}{\sum_{j=1}^{n_i} \exp\left(-\frac{\alpha}{2}(\mathbf{y}_{ij} - \boldsymbol{\mu}_i^{(k)})^T (\boldsymbol{\Sigma}^{(k)})^{-1} (\mathbf{y}_{ij} - \boldsymbol{\mu}_i^{(k)})\right)},$$

$$\boldsymbol{\Sigma}^{(k+1)} = \frac{\sum_{i=1}^k \sum_{j=1}^{n_i} \exp\left(-\frac{\alpha}{2}(\mathbf{y}_{ij} - \boldsymbol{\mu}_i^{(k)})^T (\boldsymbol{\Sigma}^{(k)})^{-1} (\mathbf{y}_{ij} - \boldsymbol{\mu}_i^{(k)})\right) (\mathbf{y}_{ij} - \boldsymbol{\mu}_i^{(k)})(\mathbf{y}_{ij} - \boldsymbol{\mu}_i^{(k)})^T}{\sum_{i=1}^k \sum_{j=1}^{n_i} \exp\left(-\frac{\alpha}{2}(\mathbf{y}_{ij} - \boldsymbol{\mu}_i^{(k)})^T (\boldsymbol{\Sigma}^{(k)})^{-1} (\mathbf{y}_{ij} - \boldsymbol{\mu}_i^{(k)})\right) - \frac{N\alpha}{(1+\alpha)^{(p+2)/2}}},$$

- 4 Compute the relative change in $\hat{\boldsymbol{\mu}}$ and $\hat{\boldsymbol{\Sigma}}$:

$$\text{rel}_{\boldsymbol{\mu}} = \frac{\|\boldsymbol{\mu}_i^{(k+1)} - \boldsymbol{\mu}_i^{(k)}\|}{\|\boldsymbol{\mu}^{(k)}\|} \quad \text{rel}_{\boldsymbol{\Sigma}} = \frac{\|\boldsymbol{\Sigma}^{(k+1)} - \boldsymbol{\Sigma}^{(k)}\|_F}{\|\boldsymbol{\Sigma}^{(k)}\|_F}$$

- 5 Check for convergence:

- 6 **if** $\text{rel}_{\boldsymbol{\mu}} < \delta$ **and** $\text{rel}_{\boldsymbol{\Sigma}} < \epsilon$ **then**

- 7 | break

- 8 **end**

- 9 **end**

- 10 **return** $\hat{\boldsymbol{\mu}}$ and $\hat{\boldsymbol{\Sigma}}$
-

The algorithm for MDPD estimators in the MANOVA model starts by selecting a small value for ε , setting a maximum number of iterations, and choosing a tuning parameter α . Initial values for the mean and covariance matrix are calculated using either coordinate-wise means or medians and the sample covariance matrix or mean absolute deviation (MAD). For each iteration, it computes new mean and covariance matrix estimates using exponential weighting, which dampens the influence of outliers. It checks for convergence by comparing

the relative changes in the mean and covariance matrix to δ . If these changes are smaller than ε , the algorithm stops; otherwise, it continues updating the estimates until convergence or the maximum number of iterations is reached. The final output consists of the robust estimates for the mean and covariance matrix.

8.2 Hypothesis Testing

Consider the MANOVA test, where the null hypothesis asserts that the mean vectors across different groups are identical, formulated as:

$$H_0 : \boldsymbol{\mu}_1 = \boldsymbol{\mu}_2 = \cdots = \boldsymbol{\mu}_k \quad \text{vs.} \quad H_1 : \boldsymbol{\mu}_i \neq \boldsymbol{\mu}_j \quad \text{for at least one of } 1 \leq i \neq j \leq k. \quad (8.5)$$

The null hypothesis H_0 is equivalent to $\mathbf{m}(\boldsymbol{\theta}) = \mathbf{0}_{p(k-1)}$, where

$$\mathbf{m}(\boldsymbol{\theta}) = \begin{pmatrix} \boldsymbol{\mu}_1 - \boldsymbol{\mu}_2 \\ \boldsymbol{\mu}_2 - \boldsymbol{\mu}_3 \\ \vdots \\ \boldsymbol{\mu}_{k-1} - \boldsymbol{\mu}_k \end{pmatrix}.$$

Observe here that, $\mathbf{m}(\cdot)$ imposes $p(k-1)$ constraints for the null hypothesis. With slight abuse of notation, let $\mathbf{m}(\boldsymbol{\theta}) = \mathbf{m}(\boldsymbol{\mu})$, since $\mathbf{m}(\boldsymbol{\theta})$ function does not depend on $\boldsymbol{\Sigma}$.

Definition 8.2.1. Let $\hat{\boldsymbol{\theta}}$ be the MDPD estimator of $\boldsymbol{\theta}$. The Wald-type test statistic for testing the null hypothesis in (8.5) is

$$\mathbf{W}_N = N \mathbf{m}^T(\hat{\boldsymbol{\theta}}) \left[\mathbf{M}^T(\hat{\boldsymbol{\theta}}) \mathbf{J}^{-1}(\hat{\boldsymbol{\theta}}) \mathbf{K}(\hat{\boldsymbol{\theta}}) \mathbf{J}^{-1}(\hat{\boldsymbol{\theta}}) \mathbf{M}(\hat{\boldsymbol{\theta}}) \right]^{-1} \mathbf{m}(\hat{\boldsymbol{\theta}}), \quad (8.6)$$

where

- $\mathbf{M}(\hat{\boldsymbol{\theta}})$ is the gradient matrix of the constraint function, where the gradient is given by:

$$\mathbf{M}(\boldsymbol{\theta}) = \frac{\partial \mathbf{m}^T(\hat{\boldsymbol{\theta}})}{\partial \boldsymbol{\theta}}.$$

- $\mathbf{J}(\hat{\boldsymbol{\theta}})$ and $\mathbf{K}(\hat{\boldsymbol{\theta}})$ are defined in Appendix A.
- $N = \sum_{i=1}^k n_i$.

For the gradient matrix, we define \mathbf{M}_μ as:

$$\mathbf{M}_\mu = \mathbf{H}_\mu \otimes \mathbf{I}_p,$$

where \mathbf{I}_p is the $p \times p$ identity matrix and \mathbf{H}_μ is defined as:

$$\mathbf{H}_\mu = \begin{pmatrix} 1 & 0 & 0 & 0 & \cdots & 0 & 0 & 0 \\ -1 & 1 & 0 & 0 & \cdots & 0 & 0 & 0 \\ 0 & -1 & 1 & 0 & \cdots & 0 & 0 & 0 \\ 0 & 0 & -1 & 1 & \cdots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \cdots & -1 & 1 & 0 \\ 0 & 0 & 0 & 0 & \cdots & 0 & -1 & 1 \\ 0 & 0 & 0 & 0 & \cdots & 0 & 0 & -1 \end{pmatrix}_{k \times (k-1)}.$$

Then, it can be shown that $\mathbf{M}(\boldsymbol{\theta}) = (\mathbf{M}_\mu^T, \mathbf{0}_{(k-1)p \times p^2})^T$. From Appendix A.8, the test statistic \mathbf{W}_N in Equation (8.6) is simplified as

$$\mathbf{W}_N = N \left(\frac{(1+2\alpha)}{(1+\alpha)^2} \right)^{(p+2)/2} \mathbf{m}^T(\hat{\boldsymbol{\mu}}) \left[\mathbf{M}_\mu^T \mathbf{S}^{-1} \mathbf{M}_\mu \right]^{-1} \mathbf{m}(\hat{\boldsymbol{\mu}}), \quad (8.7)$$

where \mathbf{S} is $k \times k$ block diagonal matrix. The i -th diagonal blocks are the $\frac{n_i}{N} \widehat{\boldsymbol{\Sigma}}_{p \times p}^{-1}$ matrix and the off-diagonal blocks are $p \times p$ zero matrices.

Theorem 8.2.1. *Under the regularity conditions in Appendix A.9, the asymptotic null distribution of the proposed Wald-type test statistic given in (8.7) is $\chi_{p(k-1)}^2$.*

The proof of the theorem is straight-forward from [Basu et al. \(2017\)](#).

8.3 Choosing Optimal α

The asymptotic distribution of the MDPD estimator plays a crucial role in selecting the optimal value of the DPD parameter α . As the performance of the corresponding estimator directly impacts the MANOVA test, it is essential to choose α that balances robustness and efficiency of $\hat{\boldsymbol{\mu}}$ effectively, as discussed in [Das et al. \(2022\)](#). In practical applications, one might either choose a fixed α to achieve a specific robustness level, trading off some efficiency, or determine an optimal α based data-driven approach. There are two primary methods we consider choosing the optimum α in the context of DPD MANOVA: one focusing on optimal efficiency and the other on optimal breakdown point.

Optimum Efficiency

The first method involves selecting α that minimizes the mean squared error (MSE) of the estimator $\hat{\boldsymbol{\mu}}$. The empirical estimate of the MSE as a function of a pilot estimator $\boldsymbol{\mu}^p$ is given by:

$$\widehat{\text{MSE}}(\alpha) = (\hat{\boldsymbol{\mu}} - \boldsymbol{\mu}^p)^T (\hat{\boldsymbol{\mu}} - \boldsymbol{\mu}^p) + \text{tr}(\hat{\boldsymbol{\Sigma}}_{\boldsymbol{\mu}}), \quad (8.8)$$

where $\hat{\boldsymbol{\Sigma}}_{\boldsymbol{\mu}}$ is the asymptotic scatter matrix of $\hat{\boldsymbol{\mu}}$, given by (detailed calculations is provided in Appendix A.8.1)

$$\hat{\boldsymbol{\Sigma}}_{\boldsymbol{\mu}} = \left(1 + \frac{\alpha^2}{1 + 2\alpha}\right)^{p/2+1} \boldsymbol{S}^{-1}.$$

To implement this technique, a pilot value α_p is initially selected to guide the optimization. The process is iterated by using the optimum α from the previous stage as the pilot estimator for the current stage, continuing until convergence is achieved. Specifically, the iterations proceed until the change in the α value between iterations falls below a predefined tolerance level. This method reduces sensitivity to the initial value of $\boldsymbol{\mu}^p$, as long as the initial estimate is robust.

Optimum Breakdown Point

The second method focuses on selecting the α that maximizes the breakdown point. The alpha value is predetermined and added as a fixed parameter, specifically calculated as $\alpha = \frac{2}{p}$, where p is the number of variables. This value of alpha is chosen because it maximizes the breakdown point, ensuring the estimator's robustness to the highest possible proportion of contamination in the data. The breakdown point is a critical robustness measure indicating the estimator's ability to handle outliers without being unduly influenced, thus ensuring reliable results even in the presence of significant data contamination.

8.4 Small Sample Correction

Small sample correction is employed to improve the accuracy of the p -value for small sample sizes. This involves generating $R = 1000$ additional samples under the null hypothesis and computing the test statistic for each sample. The mean and variance of these generated test statistics are then used to adjust the location and scale of the original test statistic distribution. Specifically, the test statistic is adjusted using the formula $T_{\text{adj}} = \frac{T - \hat{\mu}_T}{\hat{\sigma}_T}$, where $\hat{\mu}_T$ and $\hat{\sigma}_T^2$ are the sample mean and variance of the generated test statistics. The adjusted p -value is then calculated using the chi-squared distribution with $p(k - 1)$ degrees of freedom. This correction mitigates the biases present in small samples, enhancing the robustness and reliability of the statistical inference in the MANOVA test based on MDPD.

8.5 Monte Carlo Simulations

In this section, we conduct a comprehensive Monte Carlo study to evaluate the performance of the proposed Wald-type MDPD test for MANOVA, following a similar setup as described in [Das et al. \(2022\)](#) and [Todorov and Filzmoser \(2010\)](#). This study investigates the empirical performance of the proposed test statistic under a variety of scenarios, including different sample sizes, dimensionalities, error distributions, and types of outliers.

We compare the performance of our proposed method with four MANOVA tests: the classical Wald test, the classical Wilks' Lambda test, the rank-based Wilks' Lambda test, and the robust MCD Wilks' Lambda test. These tests are implemented using the `rrcov` package in R. For our proposed Wald-type MDPD test, we considered four fixed values of the tuning parameter α for different p 's. Additionally, we determined an optimal value of α through a data-driven adaptive approach referred to as optimum efficiency ("DPD(OE)") and an optimal value based on the maximum breakdown point ("DPD(OB)") as discussed in Section 8.3.

Data Generation

The data generation process involved generating data from multiple populations with different distributional properties and contamination scenarios. Two types of data distributions were considered:

1. Multivariate Normal Distribution:

$$\mathbf{Y}_i \sim N_p(\mathbf{0}, \mathbf{I}), \quad i = 1, 2, \dots, n,$$

where $N_p(\mathbf{0}, \mathbf{I})$ denotes a p -dimensional multivariate normal distribution with location vector $\mathbf{0}$ and identity scatter matrix \mathbf{I}_p .

2. Multivariate t -Distribution:

$$\mathbf{Y}_i \sim t_p(q),$$

where $t_p(q)$ denotes a p -dimensional multivariate t -distribution with q degrees of freedom.

Contamination Scenarios

To assess the robustness of the MANOVA tests, different contamination scenarios were introduced:

- **Random Contamination:** A specified proportion (e.g., 5%) of the data points were replaced with outliers generated from a different mean vector $\boldsymbol{\mu}_c$ and covariance matrix \mathbf{I} . That is, $\epsilon\%$ original standard Gaussian errors is replaced with $\boldsymbol{\varepsilon}_{ij} \sim \mathcal{N}(5\mathbf{1}^T, \mathbf{I})$ in the generalized MANOVA Model 8.5.
- **Clustered Contamination:** Outliers were introduced in specific populations to simulate clusters of contaminated data. This involved replacing a subset of the data points in certain population groups with outliers. That is, concentrated vertical outliers are generated by substituting $\epsilon\%$ of errors in the first block by $\boldsymbol{\varepsilon}_{ij} \sim \mathcal{N}(5\mathbf{1}^T, \mathbf{I})$.

Evaluation Metrics

The performance of each method was evaluated using the following metrics:

- **Significance Level:** The proportion of replications where the test rejects the null hypothesis when it is true. This assesses the test's ability to maintain the nominal significance level.
- **Power:** The proportion of replications where the test correctly rejects the null hypothesis when it is false. This evaluates the test's sensitivity to detect true effects.
- **Mean Squared Error (MSE):** Measures the accuracy of the estimated parameters by comparing the estimated values to the true values.

8.5.1 Accuracy of the Distribution of the Test Statistic

In Theorem 8.2.1, the Wald test statistic for the DPD MANOVA is shown to follow the $\chi^2_{p(k-1)}$ distribution. To investigate the accuracy of this approximation, we generate $N = 1000$ samples from a standard multivariate Gaussian distribution, considering various values for the dimension p and the number of groups k , as well as both equal and unequal sample sizes n_i for $i = 1, 2, \dots, k$. For the level simulation, all mean vectors are set to $(0, 0, 0, 0, 0)^T$ and are generated from a multivariate normal distribution with a common

spherical scatter matrix \mathbf{I} . For the power simulation, we employ a specific alternative hypothesis configuration: each group $n_i, i = 1, \dots, k$ is generated from a multivariate normal distribution with a common spherical scatter matrix \mathbf{I} . The mean of the first group is $(-0.4, 0.02, -0.2, -0.1, -0.4)^T$, while the means of all other groups are $(0, 0, 0, 0, 0)^T$.

For each method, we calculate 1000 p -values and plot them using histograms. Under the null hypothesis, we expect the p -values to be uniformly distributed, indicating proper calibration of the test. The uniform distribution of p -values means that each p -value is equally likely. If the test is well-calibrated, approximately 5% of the p -values should be less than 0.05 (assuming a 5% significance level), 10% should be less than 0.10, and so on. This indicates that the test maintains the expected rate of Type I errors (false positives). For power analysis, where the null hypothesis is false, we anticipate the p -values to be right-skewed, demonstrating the test's ability to detect deviations from the null hypothesis.

We compare the distribution of the proposed Wald-type MDPD test with the the classical Wald test equivalent to $\text{MDPD}(\alpha = 0)$, the classical Wilks' Lambda test, the rank-based Wilks' Lambda test ([Nath and Pavur, 1985](#)), and the robust MCD Wilks' Lambda ([Todorov and Filzmoser, 2010](#)) test concurrently under each scenario. This comparison is conducted under both pure and contaminated data conditions to assess the robustness and accuracy of the tests. In particular, we can expect that some methods may tend to be more conservative or more liberal in specific scenarios, impacting their Type I error rates and power. This comprehensive comparison allows us to identify which tests are likely to maintain appropriate levels of significance and power under various conditions. Figures 8.1, 8.2, and 8.3 illustrate the significance level (Panel a) and power (Panel b) for different contamination scenarios with $p = 5$, $k = 4$, $n_1 = n_2 = n_3 = n_4 = 100$. In the case of pure data (Figure 8.1), the p -values for all methods appear uniformly distributed, indicating good calibration under the null hypothesis. The power analysis shows right-skewed distributions, demonstrating the tests' ability to detect true effects. For 5% random contamination (Figure 8.2), the MDPD with larger α and $\text{MCD}(bdp = 0.5)$ tests maintain a more uniform

distribution compared to classical tests, indicating better robustness, while still showing right-skewed p -values for power. Under 5% cluster contamination (Figure 8.3), the MDPD and $MCD(bdp = 0.5)$ tests again exhibit more uniform p -value distributions and effective power, highlighting their robustness. Overall, these plots confirm that the MDPD test and $MCD(bdp = 0.5)$, especially with an appropriate tuning parameter, tends to be more robust to contamination while maintaining reasonable level and power, making it suitable for practical applications involving potentially contaminated data.

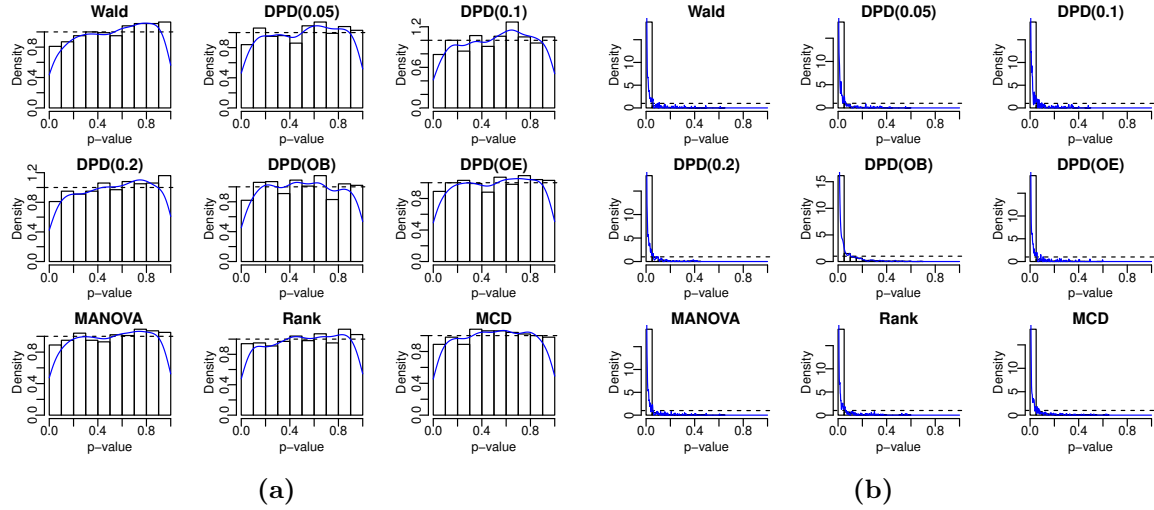


Figure 8.1: Significance Level (Panel a) and Power (Panel b) for Uncontaminated Data.

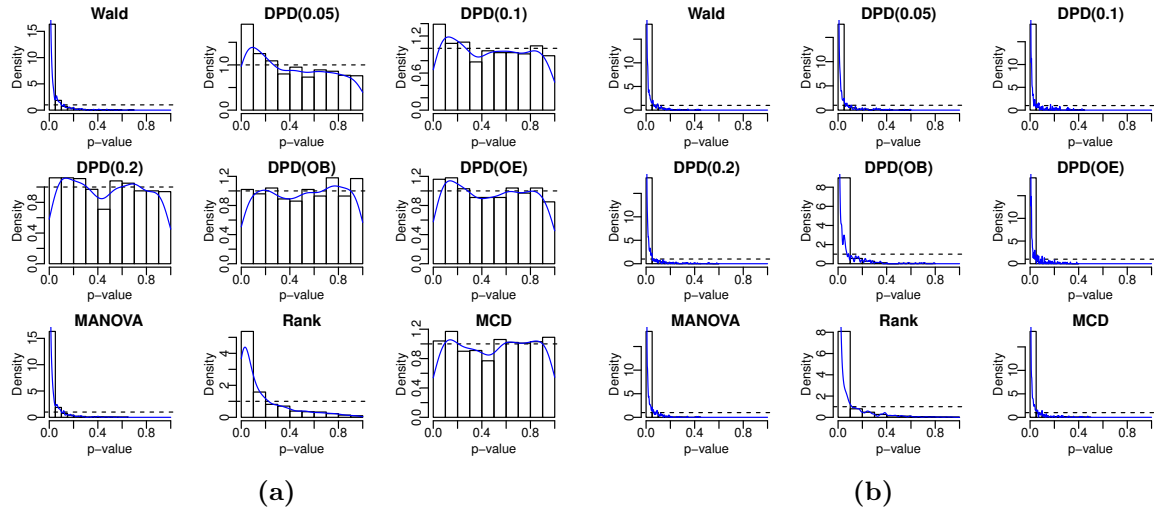


Figure 8.2: Significance Level (Panel a) and Power (Panel b) for 5% random contamination.

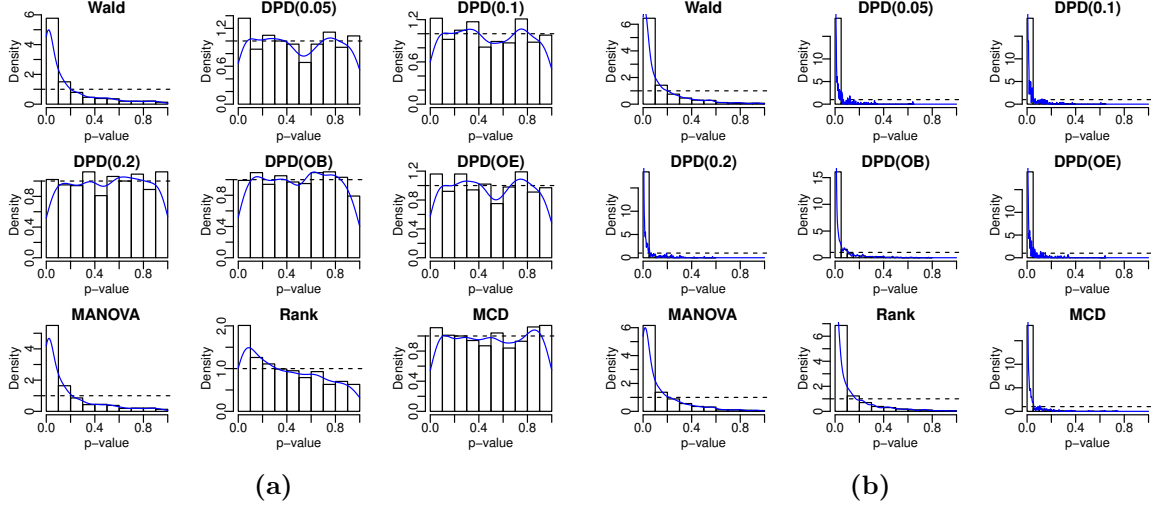


Figure 8.3: Significance Level (Panel a) and Power (Panel b) for 5% cluster contamination.

8.5.2 Empirical Levels for Varying Sample Sizes and Dimensions

To compare the empirical level (that is, observe measure of the type I error rate or size) of the test statistics, consider a one-way MANOVA model under the null hypothesis

$$H_0 : \boldsymbol{\mu}_1 = \boldsymbol{\mu}_2 = \cdots = \boldsymbol{\mu}_k.$$

We assume that the observations come from identical multivariate distribution, where each $\boldsymbol{\mu}_i$ is the zero vector ($\boldsymbol{\mu}_i = (\mathbf{0}, \dots, \mathbf{0})^T$) and the scatter matrix is identity matrix (\mathbf{I}). This assumption is valid as the test statistics are affine equivariant. The observations are thus generated from a multivariate Gaussian distribution $\mathbf{y}_{ij} \sim N(\mathbf{0}, \mathbf{I})$. The empirical level is computed as the percentage of test statistics in 1000 replications that exceed the nominal χ^2 critical value. This percentage serves as an estimate of the actual significance level, offering a practical measure of the true type I error rate. The type I error rates (nominal significance levels) α are set to 0.10, 0.05 and 0.01. With 1000 replications, the two standard deviation intervals around the nominal levels are calculated as follows: (0.0905, 0.1095), (0.0431, 0.0569), and (0.0069, 0.0135). We will consider several dimensions $p = 2, 5, 8, 10$ and groups $k = 2, 3, 4$ and sample sizes $n_i, i = 1, 2, 3, 4$. The specific sample sizes for the different groups in the simulations are detailed in the following Table 8.1

Table 8.1: Group sizes for the simulation study

Two groups	Three groups	Four groups
(n_1, n_2)	(n_1, n_2, n_3)	(n_1, n_2, n_3, n_4)
(30, 30)	(30, 30, 30)	(30, 30, 30, 30)
(50, 50)	(50, 50, 50)	(50, 50, 50, 50)
(100, 100)	(100, 100, 100)	(100, 100, 100, 100)
(50, 30)	(50, 40, 40)	(50, 40, 35, 30)
(50, 100)	(100, 80, 50)	(100, 80, 50, 30)

In Table 8.2, we present the significance levels of various MANOVA tests for uncontaminated multivariate normal distributions. The simulation setup involves comparing the MDPD Wald-type based test with Optimum Efficiency (MDPD(OE)), the robust MCD Wilk's Lambda, the classical Wilk's Lambda (CWL), the classical Rank-based test (RANK). These tests are evaluated under three-group ($k = 3$) scenarios with different dimensions (p) and total sample sizes ($n = n_1 + n_2 + n_3$). The analysis is based on $N = 1000$ Monte Carlo replications, examining the empirical significance levels $\alpha = 0.05$ and $\alpha = 0.1$.

The primary objective of this simulation is to assess how well each test maintains its nominal significance level across various dimensional settings and sample sizes. By systematically varying the parameters p , n_1 , n_2 , and n_3 we aim to provide a comprehensive comparison of the reliability of each test under ideal conditions. This evaluation is crucial for understanding the performance of these tests in practical applications and guiding researchers in selecting the most appropriate test for their specific multivariate data analysis needs. Additionally, we examine how the empirical significance level converges to the nominal level as the sample size increases for each dimension p , highlighting the improvement in test accuracy with larger sample sizes.

In Table 8.2, we observe that all the methods perform well, falling within the two standard

deviation intervals around the nominal levels. The CWL and Rank tests consistently align with the nominal level across different dimensions (p) and sample sizes, with minor deviations. The MDPD(OE) test shows some variability, with significance levels fluctuating around the nominal value, sometimes falling below and sometimes within the interval. However, the MDPD(OE) test can be tuned with different α to improve its performance on uncontaminated data. The MCD($bdp = 0.5$) test mostly falls within this interval, even in higher-dimensional settings, indicating good performance. As the sample size increases, all the methods perform similarly, maintaining significance levels close to the nominal values.

Table 8.2: Empirical Levels for MDPD(OE) test, robust MCD Wilk's Lambda test, classical Wilk's Lambda (CWL) test, and classical Rank-based test for different dimensions p and the sample sizes $n = n_1 + n_2 + n_3$ (Based on $N = 1000$ Monte Carlo replications with nominal significance levels $\alpha = 0.1$ and 0.05).

p	n_1	n_2	n_3	$\alpha = 0.05$				$\alpha = 0.1$			
				MDPD(OE)	MCD	CWL	Rank	MDPD(OE)	MCD	CWL	Rank
2	30	30	30	0.0550	0.0560	0.0640	0.0630	0.1000	0.1040	0.1140	0.1140
2	50	30	45	0.0600	0.0510	0.0570	0.0550	0.1190	0.1010	0.1160	0.1150
2	50	50	50	0.0500	0.0520	0.0440	0.0400	0.1090	0.1010	0.1000	0.1000
2	50	100	85	0.0610	0.0590	0.0620	0.0610	0.1000	0.1180	0.1050	0.0960
2	100	100	100	0.0500	0.0460	0.0470	0.0440	0.0950	0.0820	0.0840	0.0830
5	30	30	30	0.0580	0.0550	0.0550	0.0570	0.1130	0.0960	0.1080	0.1170
5	50	30	45	0.0410	0.0400	0.0450	0.0480	0.0900	0.1030	0.0930	0.0970
5	50	50	50	0.0440	0.0410	0.0450	0.0420	0.0860	0.0910	0.0900	0.0930
5	50	100	85	0.0680	0.0580	0.0630	0.0570	0.1110	0.0910	0.0980	0.1040
5	100	100	100	0.0500	0.0540	0.0520	0.0580	0.0950	0.1080	0.1060	0.1190
8	50	100	85	0.0570	0.0570	0.0520	0.0500	0.1150	0.0990	0.0960	0.1080
8	100	100	100	0.0520	0.0580	0.0430	0.0510	0.1040	0.0970	0.0890	0.1020
8	200	100	135	0.0490	0.0550	0.0530	0.0510	0.0950	0.0920	0.0990	0.1040
8	150	150	150	0.0530	0.0510	0.0500	0.0480	0.0990	0.0980	0.0950	0.0970
8	200	200	200	0.0600	0.0540	0.0530	0.0460	0.1120	0.1010	0.1090	0.0970
8	200	200	200	0.0600	0.0540	0.0530	0.0460	0.1120	0.1010	0.1090	0.0970
10	100	100	100	0.0430	0.0520	0.0440	0.0520	0.0960	0.1000	0.0970	0.0900
10	200	100	135	0.0450	0.0510	0.0460	0.0490	0.0910	0.1080	0.0930	0.0940
10	150	150	150	0.0700	0.0590	0.0600	0.0690	0.1220	0.1180	0.1130	0.1200
10	200	200	200	0.0520	0.0510	0.0490	0.0510	0.1050	0.1000	0.0990	0.1020

8.5.3 Empirical Levels for Different Contaminations and Error Distributions

The empirical significance level is computed as the proportion of test statistics in 1000 replications that exceed the nominal critical value at a 5% level of significance. The results are reported in Table (8.3). The setup involves data with dimensionality $p = 5$, number of groups $k = 4$ and group sizes $n_1 = n_2 = n_3 = n_4 = 100$ under four scenarios: no contamination (normal distribution), 5% random contamination, 5% clustered contamination, and t_3 distribution. For each scenario, we calculate the empirical significance level (Type I error rate) and the Mean Squared Error (MSE) of the estimated mean vector ($\hat{\mu}$), scaled by the sample size N .

The tests analyzed include MCD($bdp = 0.5$), MDPD tests with varying tuning parameters (0.05, 0.1, 0.2), MDPD(OB), MDPD(OE), MANOVA, and Rank-based test.

The results indicate that the MCD($bdp = 0.5$) test provides moderate robustness, with empirical levels close to the nominal 5% across all scenarios. For example, under 5% random contamination, the MCD($bdp = 0.5$) test has an empirical level of 0.0530 and an MSE of 88.27. However, its MSE values are relatively higher compared to some MDPD-based tests, indicating less accuracy in mean estimation under contamination.

Under the t_3 distribution scenario, which introduces heavy-tailed data, the empirical significance levels increase notably for most tests. For instance, the MCD test reaches an empirical level of 0.1700 and an MSE of 125.38, reflecting the challenge posed by heavy-tailed distributions. The MDPD tests, particularly with tuning parameters of 0.1 and 0.2, again show better robustness with lower empirical levels and MSE values, such as the MDPD(0.2) test, which has an empirical level of 0.0480 and an MSE of 102.85.

The CWL test, however, performs poorly under contamination scenarios, showing high empirical levels and MSE values. For instance, under 5% clustered contamination, it reaches an empirical level of 0.4350 and an MSE of 365.30.

These findings suggest that MDPD-based tests offer a balanced performance, providing robustness against contamination while maintaining reasonable estimation accuracy, making them preferable to classical tests in contaminated data scenarios. The $\text{MCD}(bdp = 0.5)$ test offers some robustness but is outperformed by the MDPD tests in terms of MSE. The t_3 distribution scenario highlights the challenges for all tests, but MDPD-based methods still show relatively better performance.

Table 8.3: The Empirical level of different tests and the MSE of $\hat{\mu}$ (times N) for the corresponding estimators for different error distributions.

Test	c = 0, Normal		c=5%, Random		c = 5%, Clustered		t_3	
	Level	MSE	Level	MSE	Level	MSE	Level	MSE
MCD	0.0480	86.2293	0.0530	88.2672	0.0480	86.4968	0.1700	125.3849
DPD(0.05)	0.0520	80.2406	0.1170	162.0318	0.0740	80.7253	0.0200	126.5639
DPD(0.1)	0.0430	82.0029	0.0630	85.3388	0.0670	82.3185	0.0340	110.0233
DPD(0.2)	0.0450	88.2504	0.0550	91.6373	0.0560	88.1470	0.0480	102.8518
DPD(OB)	0.0460	109.5123	0.0510	112.9685	0.0580	108.5080	0.0400	113.4719
DPD(OE)	0.0570	79.6490	0.0550	84.2180	0.0630	80.6465	0.0450	104.5226
MANOVA	0.0550	79.6091	0.8160	1305.9504	0.4350	365.3034	0.0330	241.4604
Rank	0.0510	—	0.3770	—	0.1090	—	0.0400	—

8.5.4 Empirical Levels for Different Sample Sizes

Figure 8.4 illustrates the performance of different MANOVA tests under various conditions: uncontaminated data (top left), 1% random outliers (top right), 5% random outliers (bottom left), and 5% clustered outliers (bottom right). Overall, the Wald test and classical Wilk's Lambda test (CWL) maintain their nominal significance levels well in uncontaminated data. The MDPD(OE) and $\text{MCD}(bdp = 0.5)$ tests demonstrate superior robustness across all scenarios, maintaining nominal significance levels effectively, even in the presence

of random and clustered outliers. The classical Wald test shows more variability and sensitivity to outliers, particularly with higher percentages of random outliers. As n increases, it has tendency to reject every test. These trends highlight the MDPD(OE) and MCD tests' robustness to data contamination compared to the Wald, classical Wilk's Lambda test (CWL), and rank-based tests, which exhibit varying degrees of sensitivity, especially with increasing outlier presence.

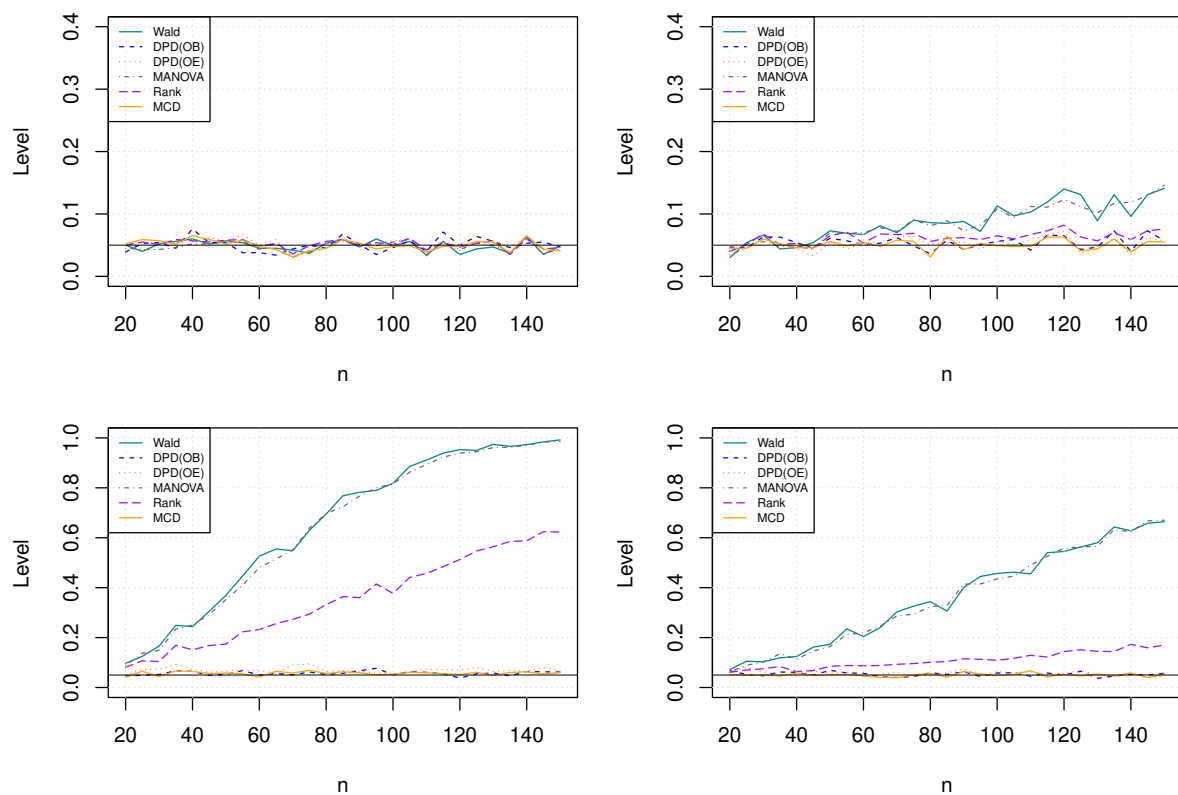


Figure 8.4: The empirical level of different tests in uncontaminated data (top left) and in the presence of 1% random outliers at random locations (top right), 5% random outliers (bottom left), and 5% clustered outliers (bottom right). In all cases, $k = 4$ and $p = 5$.

8.5.5 Empirical Powers for Different Contaminations and Error Distributions

To evaluate the power of the robust MDPD Wald-type test compared to the classical Wald test, the classical Wilks' Lambda test (CWL), the classical rank-based test, and the MCD Wilks' Lambda test, we will generate data under the alternative hypothesis H_a , where not all group means $\boldsymbol{\mu}_k$ are equal ($i = 1, \dots, k$). We will analyze the likelihood of failing to reject the null hypothesis H_0 (the frequency of type II errors) using the same combinations of dimension p , number of groups k , and sample sizes $n_i, i = 1, \dots, k$ as those used in previous experiments for studying significance levels. This comparison will help in understanding the relative effectiveness of these tests under the given conditions.

For this simulation, we will employ a specific alternative hypothesis configuration: each group $n_i, i = 1, \dots, k$ is generated from a multivariate normal distribution with a common spherical scatter matrix \mathbf{I} . The mean of the first group is $(-0.4, 0.02, -0.2, -0.1, -0.4)^T$, while the means of all other groups are $(0, 0, 0, 0, 0)^T$.

This setup allows for a straightforward assessment where the number of groups k can be at most $p + 1$. However, for simplicity, we will restrict our analysis to scenarios with three or four groups. Specifically, the data sets are derived from the following p -variate Gaussian distributions, where each group $n_i, i = 1, \dots, k$ has a distinct mean $\boldsymbol{\mu}_k$ but shares the same covariance matrix \mathbf{I} :

$$n_k \sim \mathcal{N}_p(\boldsymbol{\mu}_k, \mathbf{I}), \quad i = 1, \dots, k,$$

Table 8.4: The empirical power of different tests and the MSE of $\hat{\mu}$ (times N) for the corresponding estimators for different error distributions.

Test	c = 0, Normal		c=5%, Random		c = 5%, Clustered		t_3	
	Power	MSE	Power	MSE	Power	MSE	Power	MSE
MCD	0.9170	87.5590	0.9150	90.3165	0.9120	87.7463	0.8990	128.8784
DPD(0.05)	0.9240	80.9935	0.8860	163.1828	0.9500	81.9158	0.6340	129.7988
DPD(0.1)	0.9450	82.7934	0.9420	86.9558	0.9480	83.7001	0.7810	113.5364
DPD(0.2)	0.9140	89.0459	0.9160	93.5743	0.9230	90.0149	0.8480	106.1143
DPD(OB)	0.8070	110.2515	0.8360	115.5322	0.8040	111.4666	0.8190	116.2280
DPD(OE)	0.9430	80.3067	0.9470	85.7154	0.9460	81.8225	0.8500	107.7026
MANOVA	0.9410	80.2901	0.9100	1334.1531	0.4890	372.2143	0.4590	240.3558
Rank	0.9270	—	0.7080	—	0.5630	—	0.7480	—

The results in Table 8.4 indicate that the MCD test provides high power across all scenarios, with power values close to or above 0.9. For instance, under the normal distribution, the MCD test has a power of 0.9170 and an MSE of 87.56 and under the t_3 distribution scenario, which introduces heavy-tailed data, the MCD test has a power of 0.899 and an MSE of 128.88.

Similarly, MDPD tests with tuning parameters of 0.05, 0.1, and 0.2 also exhibit high power, particularly under the normal distribution and contamination scenarios. For example, DPD(0.1) achieves the highest power of 0.9450 under the normal distribution with an MSE of 82.79. Under 5% random contamination, DPD(0.1) maintains a high power of 0.9420 with an MSE of 86.9558 and for the t_3 distribution, DPD(0.2) shows a power of 0.8480 with an MSE of 106.11, indicating robust performance even with heavy-tailed data.

The MDPD(OE) test demonstrates high power and low MSE across all scenarios. For example, under the normal distribution, DPD(OE) has a power of 0.9430 with an MSE of

80.3067. Under 5% random contamination, the power is 0.9470 with an MSE of 85.7154, and under 5% clustered contamination, the power is 0.9460 with an MSE of 81.82. In the t_3 distribution scenario, the power is 0.8500 with an MSE of 107.7.

The MANOVA test exhibits high power under the normal distribution (0.9410) with low MSE (80.29). However, its performance deteriorates significantly under contamination scenarios. For instance, under 5% random contamination, the power is 0.9100, but the MSE skyrockets to 1334.1531. Under 5% clustered contamination, the power drops to 0.4890 with an MSE of 372.21, and under the t_3 distribution, the power is 0.4590 with an MSE of 240.34, indicating poor robustness.

8.5.6 Empirical Powers for Different Sample Sizes

Figure 8.5 illustrates the power of the robust MDPD Wald-type test, classical Wald test, classical Wilk's Lambda (CWL), and classical rank-based test under the following scenarios: uncontaminated data, 1% random outliers, 5% random outliers, and 5% clustered outliers. Across all scenarios, the tests' power increases with sample size n . In uncontaminated data, all tests perform well, with the classical Wald test showing high power, closely followed by the classical Wilk's Lambda and rank-based test. The presence of outliers reveals the robustness of the MDPD and MCD test, maintaining higher power compared to the classical tests, which experience notable power reduction, especially with clustered outliers. Overall, the robust MDPD(OE) Wald-type test consistently demonstrates superior performance in contaminated data scenarios, highlighting its value in practical applications where data contamination is a concern.

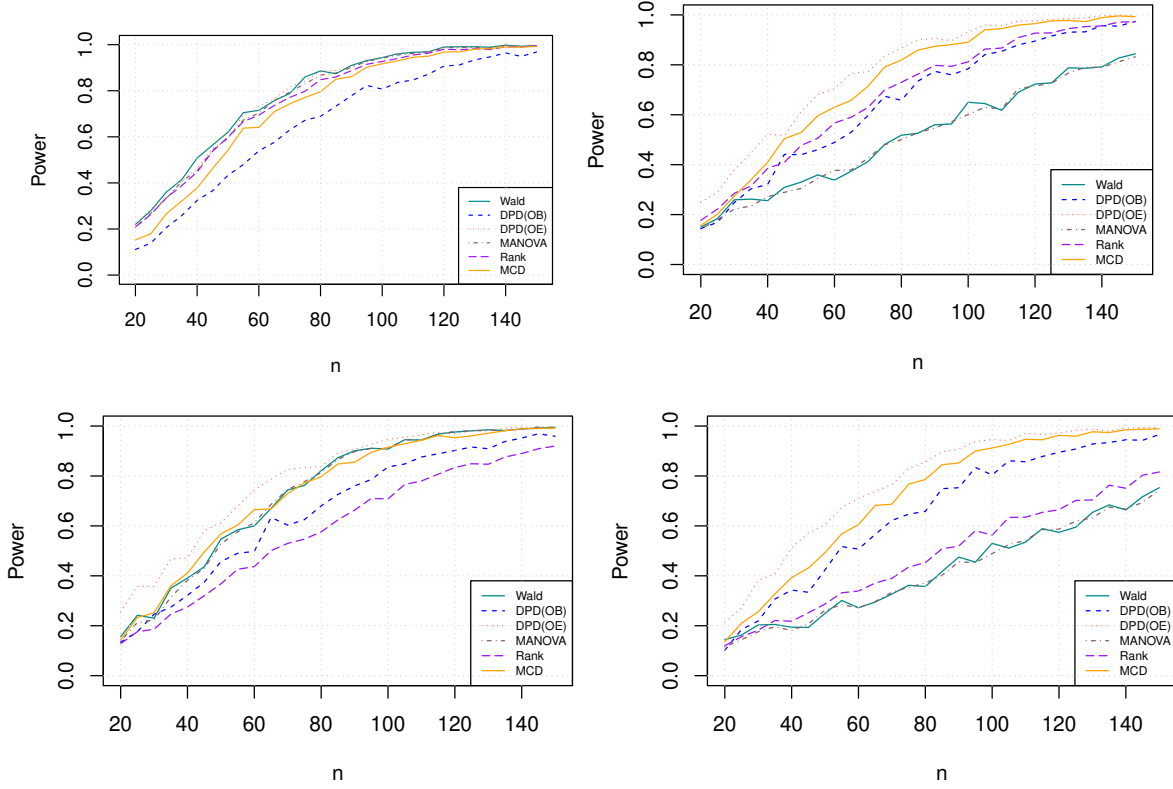


Figure 8.5: The power of different tests in uncontaminated data (top left) and in the presence of 1% random outliers at random locations (top right), 5% random outliers (bottom left), and 5% clustered outliers (bottom right). In all cases, $k = 4$, $p = 5$ and $\mu_1 = (-0.4, 0.02, -0.2, -0.1, -0.4)^T$.

When comparing the robust MDPD and robust MCD tests, the MDPD test demonstrates higher power due to its robustness against outliers and model misspecifications. This robustness allows the MDPD test to better detect true differences, even in the presence of anomalies, thereby increasing its probability of correctly rejecting the null hypothesis when it is false. On the other hand, the MCD test excels in maintaining a better significance level performance, effectively controlling the probability of Type I errors. This is achieved by focusing on minimizing the determinant of the covariance matrix for a subset of the data,

making it more conservative and less likely to falsely reject the null hypothesis. Therefore, while the MDPD test is more sensitive to detecting true effects, the MCD test prioritizes reducing false positives, illustrating the trade-off between robustness and significance level control.

8.6 Real Data Analysis

In this study, we analyze a subset of the dataset derived from the Oslo transect, initially discussed by [Reimann et al. \(2007\)](#). The dataset includes samples collected from various plant species along a 120 km transect passing through Oslo, Norway’s largest city. The samples encompass 40 individual specimens from each of the following species: birch (*Betula pubescens* Ehrh.), European mountain ash (*Sorbus aucuparia* (L.)), bracken fern (*Pteridium aquilinum* (L.) Kuhn), and spruce needles (*Picea abies* (L.) Karsten). These samples were analyzed to determine the concentrations of 25 chemical elements along with the loss on ignition for each type of plant material.

The transect was chosen to cross a variety of lithologies, including Precambrian gneisses, Cambro-Silurian sediments, and magmatic rocks of the Oslo Rift. This geological diversity is expected to influence the uptake of chemical elements by plants. Previous studies have shown that factors such as geology, soil pH, and sea spray significantly affect elemental concentrations in plant leaves. For example, Ca, Ba, and Ni concentrations were closely linked to the geological substrate, while Mn concentrations were strongly influenced by soil pH. Additionally, sea spray played a critical role in the distribution of elements like B and Na, with concentrations decreasing systematically with distance from the coast.

For our specific analysis, we will use the proposed robust test statistic to determine whether there are significant differences in the mean concentrations of selected nutrients—phosphorus (P), potassium (K), zinc (Zn), and copper (Cu)—across different lithological groups. Similar to the analyses conducted by [Todorov and Filzmoser \(2010\)](#) and [Pison et al. \(2002\)](#), we

focus on the influence of lithology. Our working dataset, after excluding observations with missing values, consists of 332 observations across these four chemical elements. To account for the skewness typical in geochemical data, we applied a logarithmic transformation to the variables before performing statistical analyses. Table 8.5 shows the group sizes of the four lithological groups in the Oslo transect data, and Figure 8.6 presents the boxplots for the different groups.

Table 8.5: Oslo transect data: Names of the lithology groups.

Lithological group	Description	Group Sizes
CAMSED	Cambro-Silurian sedimentary rocks	98
GNEISS_O	Precambrian gneisses–Oslo	89
GNEISS_R	Precambrian gneisses–Randsfjord	32
MAGM	Magmatic rocks of the Oslo Rift	113

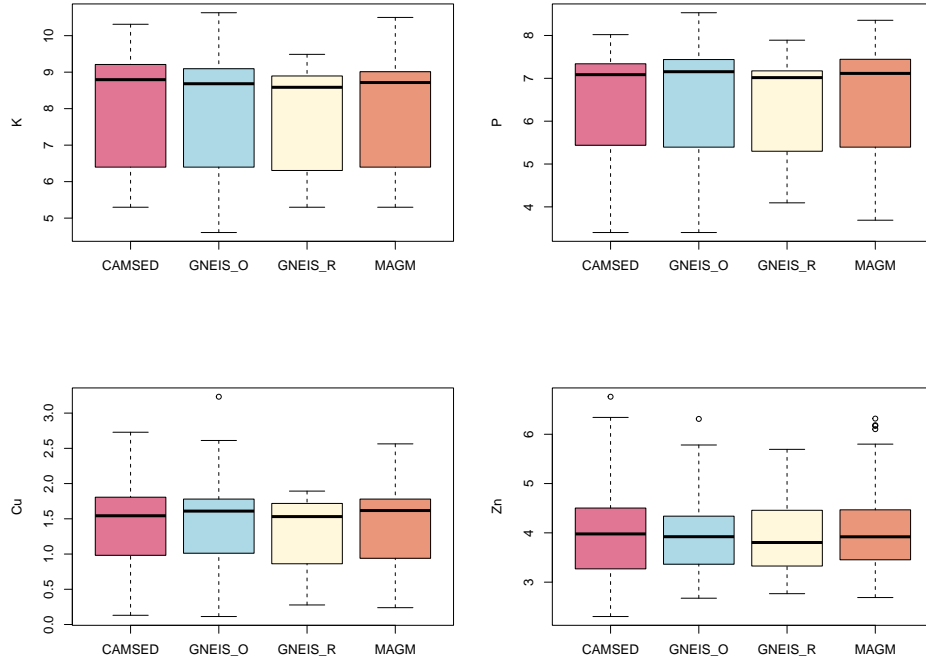


Figure 8.6: Boxplot of the log-transformed lithological data from four Oslo transects.

The box plots and distance-distance plots offer a detailed analysis of the log-transformed concentrations of elements (K, P, Cu, Zn) across four lithological groups: CAMSED, GNEIS_O, GNEIS_R, and MAGM. The box plots (Figure 8.6) show that the median values for Potassium (K) and Phosphorus (P) are similar across all groups, with consistent interquartile ranges (IQR), indicating uniform central tendencies and variability. Copper (Cu) also displays consistent median values, though it has a notable outlier in the GNEIS_O group. Zinc (Zn), while having similar IQRs, presents several outliers, particularly in the MAGM group.

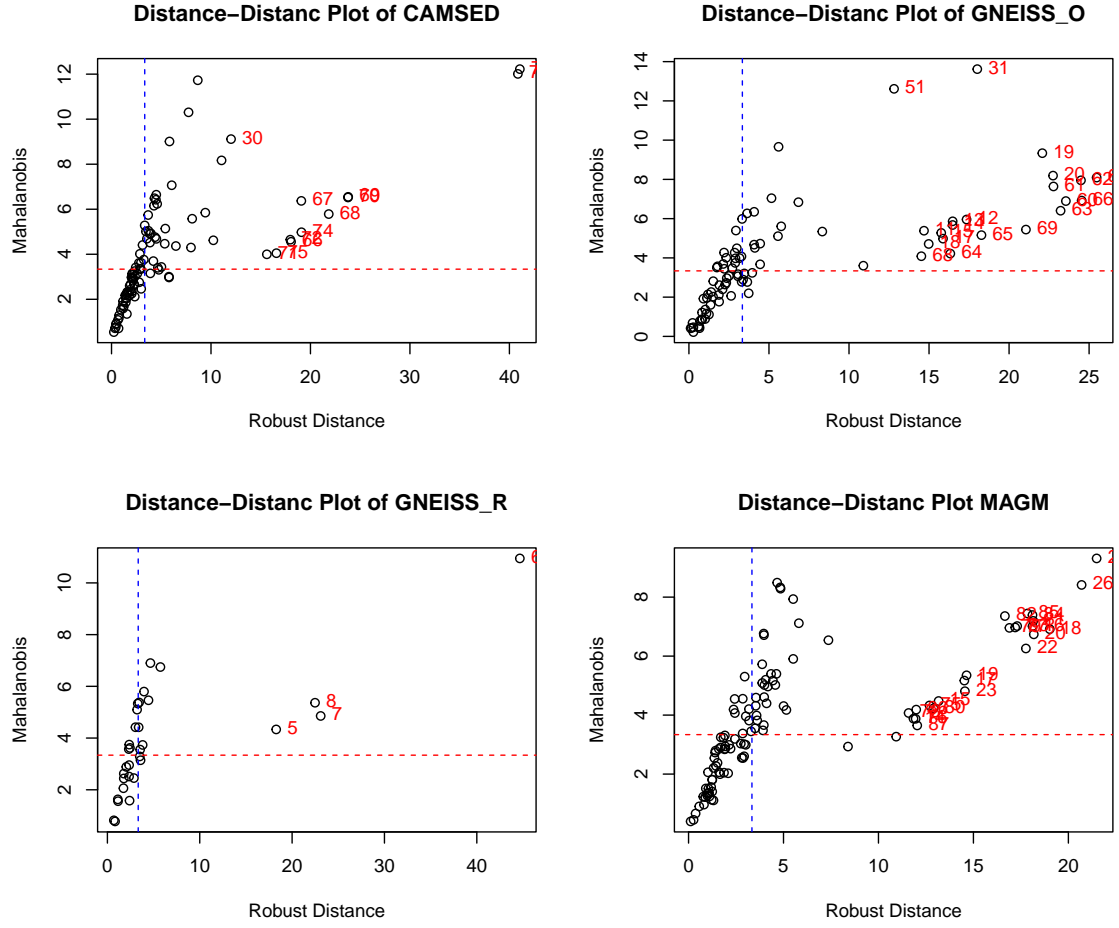


Figure 8.7: MCD($\alpha = 0.75$) distance-distance plot of the log-transformed lithological data from four Oslo transects (most influential outliers are colored red).

The distance-distance plots (Figure 8.7) provide a detailed examination of the data by comparing Mahalanobis distances and robust distances using MCD (with $\alpha = 0.75$) for each observation. Here, $\alpha = 0.25$ corresponds to a breakdown point of 25%. For identifying the most influential outliers, we flagged all observations with robust distances greater than $\chi^2_{p,0.975} = 11.16$ and Mahalanobis distances greater than $\sqrt{\chi^2_{p,0.975}} = 3.34$. These outliers are highlighted in red in Figure 8.7.

The distance-distance plots reveal several outliers which can significantly affect the results of

classical statistical tests. These outliers, particularly those identified using robust distance metrics, underline the importance of robust statistical methods. The presence of significant outliers in the dataset necessitates the use of robust MANOVA to accurately compare the means across the lithological groups. Traditional MANOVA might be unduly influenced by these outliers, potentially leading to misleading conclusions. Robust MANOVA, by accounting for these outliers, provides a more reliable and accurate comparison of group means, ensuring that the conclusions drawn are robust and reflective of the true underlying data structure.

To analyze the means of four groups, denoted as μ_1 , μ_2 , μ_3 , and μ_4 , a one-way MANOVA was conducted to test the hypothesis $H_0 : \mu = \mu_2 = \mu_3 = \mu_4$. The classical Wilks' Lambda statistic (MANOVA) for this dataset yielded a p -value of 0.7757. Similarly, the classical Wald test produced a p -value of 0.7609. On the other hand, the classical Rank test, which is a robust method, resulted in a lower p -value of 0.1133. Despite this reduction, the p -value for the Rank test is still not significant. These results from the classical tests indicate that the hypothesis of equal means cannot be rejected even at the 10% significance level.

Conversely, using the robust Wald-type minimum density power divergence test (MDPD Wald-type test) and the robust MCD Lambda test, we obtained p -values of 0.0036 and 0.0025, respectively, allowing us to reject the null hypothesis even at the 1% significance level. This robust approach highlights the necessity of considering outliers to draw accurate conclusions in the presence of influential data points.

Furthermore, the influential outliers highlighted in red in Figure 8.7 were removed to understand their impact on the outcomes of the various tests. The group sizes of the data with the influential outliers removed are summarized in Table 8.6, and the corresponding p -values are summarized in Table 8.7.

Table 8.6: Oslo Transect Data: Names of the lithology groups.

Lithological group	Description	Group Sizes
CAMSED	Cambro-Silurian sedimentary rocks	86
GNEISS_O	Precambrian gneisses–Oslo	68
GNEISS_R	Precambrian gneisses–Randsfjord	28
MAGM	Magmatic rocks of the Oslo Rift	89

After removing the very influential outliers and performing the tests again, all methods rejected the null hypothesis at the 1% significance level.

Table 8.7: p -values for the various method with very influential outliers removed.

Method	p -values (full data)	p -values (influential outliers removed)
Classical Wald	0.7609	0.0061
Classical Wilks' Lambda	0.7757	0.0096
Classical Rank	0.1133	0.0030
DPD(bdp=14.5%)	0.0035	0.0031
MCD(bdp=14.5%)	0.0021	0.0098

In conclusion, the analysis of the Oslo transect dataset underscores the critical role of robust statistical methods in the presence of influential outliers. Traditional MANOVA tests, including Wilks' Lambda and the classical Wald test, failed to detect significant differences in mean concentrations of selected chemical elements across lithological groups. However, using robust methods like the robust Wald-type MDPD and the robust MCD Lambda test showed significant differences, even at very high significance levels. Removing influential outliers confirmed these findings, leading to all methods rejecting the null hypothesis. The study demonstrates that accounting for outliers is essential for accurate geochemical data

analysis. This approach ensures that conclusions drawn reflect the underlying data structure, facilitating a better understanding and interpretation of environmental and geological influences on plant chemistry.

Chapter 9

Conclusions

In this dissertation, we have embarked on an in-depth exploration of robust statistical methodologies aimed at enhancing the reliability and accuracy of multivariate data analysis, particularly in the face of outliers. Our focus on the minimum density power divergence (MDPD) estimator has unveiled its significant potential in providing a flexible framework to mitigate the adverse impacts of data contamination. The MDPD approach, with its adjustable robustness parameter α , has demonstrated consistent superiority over classical methods through various simulation studies, particularly under conditions of data contamination. This adaptability, coupled with its efficiency, underscores the practical utility of MDPD in real-world applications where outliers are prevalent.

Comparative analysis with other robust methods, such as the minimum covariance determinant (MCD), revealed that the MDPD estimator excels in reducing Mean Squared Error (MSE) and Kullback-Leibler (KL) divergence when matched for the same breakdown point. These findings highlight the robust nature of the MDPD estimator and validate its effectiveness in statistical estimation, making it a preferable choice in contaminated data environments.

Extending the MDPD framework to Multivariate Analysis of Variance (MANOVA) has proven particularly fruitful. The development of a robust MANOVA based on the MDPD estimator has addressed the complexities of multivariate data, providing more reliable hypothesis testing outcomes than classical MANOVA tests. Monte Carlo simulations confirmed that this robust MANOVA test maintains appropriate significance levels and power

across various contamination scenarios, thus offering a valuable tool for robust hypothesis testing in multivariate contexts.

Real-world applications further underscore the necessity and efficacy of these robust methodologies. Analyzing datasets such as the Swiss Banknote Dataset for Principal Component Analysis (PCA), the Pulp Fiber and Paper dataset for Multivariate Regression, and lithological data from the Oslo transect for Robust MANOVA demonstrated the limitations of traditional statistical assumptions in practical scenarios. For the Swiss Banknote dataset, robust MDPD-based PCA identified influential outliers that classical PCA missed, significantly enhancing the analysis’s reliability and sensitivity. In the Pulp Fiber and Paper dataset, robust MDPD-based multivariate regression techniques effectively flagged outliers and bad leverage points that classical methods overlooked, leading to more accurate regression models and a better understanding of the relationships between pulp fiber properties and paper properties. For the lithological data from the Oslo transect, robust MANOVA tests highlighted significant differences in mean concentrations of selected nutrients across different lithological groups, which classical MANOVA methods failed to detect. This empirical evidence underscores the practical importance of adopting robust methods in diverse fields, as the robust MDPD-based tests identified significant differences between groups that classical methods overlooked, showcasing their enhanced sensitivity and reliability.

A notable limitation of the MDPD estimator is that its breakdown point (bdp) reduces to zero as the dimensionality p increases, making it less effective in very high-dimensional contexts. Therefore, applying this method to datasets with dimensions not exceeding 15 is ideal.

Looking ahead, this research paves the way for numerous future directions. Extending these robust methodologies to other multivariate techniques, such as cluster analysis and discriminant analysis, could broaden their applicability and utility. Additionally, exploring the integration of MDPD with other robust statistical methods may enhance its performance and robustness in practical applications.

In conclusion, the robust statistical methods developed and validated in this dissertation offer substantial improvements over classical approaches, particularly in the presence of data contamination. The MDPD estimator and its applications in PCA, multivariate regression, and MANOVA provide powerful tools for accurate and reliable multivariate data analysis. These contributions represent significant advancements in the field of robust multivariate statistics, with wide-ranging implications for both theoretical developments and practical applications. By embracing these advanced robust methods, researchers and practitioners can achieve more dependable and precise results, even in the most challenging data environments.

References

- Anderson, T. W. (2018). *An Introduction to Multivariate Statistical Analysis*, volume 121 of A. Wiley, New York, NY.
- Anum, A. T. and Pokojovy, M. (2024). A hybrid method for density power divergence minimization with application to robust univariate location and scale estimation. *Communications in Statistics–Theory and Methods*, 53(14):5186–5209.
- Barrett, B. E. and Ling, R. F. (1992). General classes of influence measures for multivariate regression. *Journal of the American Statistical Association*, 87(417):184–191.
- Basu, A., Ghosh, A., Mandal, A., Martín, N., and Pardo, L. (2017). A Wald-type test statistic for testing linear hypothesis in logistic regression models based on minimum density power divergence estimator.
- Basu, A., Harris, I. R., Hjort, N. L., and Jones, M. (1998). Robust and efficient estimation by minimising a density power divergence. *Biometrika*, 85(3):549–559.
- Basu, A. and Lindsay, B. G. (1994). Minimum disparity estimation for continuous models: efficiency, distributions and robustness. *Annals of the Institute of Statistical Mathematics*, 46:683–705.
- Basu, A., Mandal, A., Martin, N., and Pardo, L. (2015). Robust tests for the equality of two normal means based on the density power divergence. *Metrika*, 78(5):611–634.
- Basu, A., Shioya, H., and Park, C. (2011). *Statistical Inference: The Minimum Distance Approach*. CRC Press, Boca Raton, Florida.
- Becker, C., Fried, R., and Kuhnt, S. (2014). *Robustness and Complex Data Structures: Festschrift in Honour of Ursula Gather*. Springer Berlin Heidelberg, New York, NY.

- Beran, R. (1977). Minimum Hellinger distance estimates for parametric models. *The Annals of Statistics*, 5(3):445–463.
- Bickel, P. J. (1964). On some alternative estimates for shift in the p -variate one sample problem. *The Annals of Mathematical Statistics*, 35(3):1079–1090.
- Bickel, P. J. (1965). On some robust estimates of location. *The Annals of Mathematical Statistics*, 36(3):847–858.
- Billor, N., Hadi, A. S., and Velleman, P. F. (2000). BACON: blocked adaptive computationally efficient outlier nominators. *Computational Statistics & Data Analysis*, 34(3):279–298.
- Boudt, K., Rousseeuw, P. J., Vanduffel, S., and Verdonck, T. (2020). The minimum regularized covariance determinant estimator. *Statistics and Computing*, 30(1):113–128.
- Breiman, L. and Friedman, J. H. (1997). Predicting multivariate responses in multiple linear regression. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 59(1):3–54.
- Buchanan, A. M. and Fitzgibbon, A. W. (2005). Damped Newton algorithms for matrix factorization with missing data. In *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, volume 2, pages 316–322. IEEE.
- Butler, R., Davies, P., and Jhun, M. (1993). Asymptotics for the minimum covariance determinant estimator. *The Annals of Statistics*, pages 1385–1400.
- Cator, E. A. and Lopuhaä, H. P. (2010). Asymptotic expansion of the minimum covariance determinant estimators. *Journal of Multivariate Analysis*, 101(10):2372–2388.
- Chakraborty, B. and Chaudhuri, P. (1999). A note on the robustness of multivariate medians. *Statistics & Probability Letters*, 45(3):269–276.
- Clement, E. P. (2020). On the improvement of multivariate ratio method of estimation in

- sample surveys by calibration weightings. *Asian Journal of Probability and Statistics*, 10(1):1–12.
- Cook, R. D. and Setodji, C. M. (2003). A model-free test for reduced rank in multivariate regression. *Journal of the American Statistical Association*, 98(462):340–351.
- Critchley, F. (1985). Influence in principal components analysis. *Biometrika*, 72(3):627–636.
- Croux, C. and Haesbroeck, G. (1999). Influence function and efficiency of the minimum covariance determinant scatter matrix estimator. *Journal of Multivariate Analysis*, 71(2):161–190.
- Croux, C. and Haesbroeck, G. (2000). Principal component analysis based on robust estimators of the covariance or correlation matrix: influence functions and efficiencies. *Biometrika*, 87(3):603–618.
- Croux, C. and Haesbroeck, G. (2003). Implementing the bianco and yohai estimator for logistic regression. *Computational Statistics & Data Analysis*, 44(1-2):273–295.
- Das, J., Beyaztas, B. H., Mac-Ocloo, M. K., Majumdar, A., and Mandal, A. (2022). Testing equality of multiple population means under contaminated normal model using the density power divergence. *Entropy*, 24(9):1189.
- Davies, L. (1992). The asymptotics of Rousseeuw’s minimum volume ellipsoid estimator. *The Annals of Statistics*, 20(4):1828–1843.
- Davies, L. and Gather, U. (1993). The identification of multiple outliers. *Journal of the American Statistical Association*, 88(423):782–792.
- Davies, P. L. (1987). Asymptotic behaviour of S-estimates of multivariate location parameters and dispersion matrices. *The Annals of Statistics*, pages 1269–1292.
- Davies, P. L. and Gather, U. (2005). Breakdown and groups. *The Annals of Statistics*, 33(3):977–1035.

- Donoho, D. L. (1982). Breakdown properties of multivariate location estimators. Technical report, Technical report, Harvard University, Boston, MA.
- Donoho, D. L. and Gasko, M. (1992). Breakdown properties of location estimates based on halfspace depth and projected outlyingness. *The Annals of Statistics*, 20(4):1803–1827.
- Donoho, D. L. and Huber, P. J. (1983). *The notion of breakdown point. In A Festschrift for Erich L. Lehmann.* Wadsworth, Belmont, California.
- Fan, J., Wang, W., and Zhong, Y. (2019). Robust covariance estimation for approximate factor models. *Journal of Econometrics*, 208(1):5–22.
- Finch, H. and French, B. (2013). A Monte Carlo comparison of robust MANOVA test statistics. *Journal of Modern Applied Statistical Methods*, 12:35–81.
- Flury, B. and Riedwyl, H. (1988). *Multivariate Statistics: A Practical Approach.* Chapman & Hall, London, UK.
- Genton, M. (2003). Breakdown-point for spatially and temporally correlated observations. In *Developments in Robust Statistics: International Conference on Robust Statistics 2001*, pages 148–159, Heidelberg. Physica-Verlag HD.
- Gleser, L. J. (1992). The importance of assessing measurement reliability in multivariate regression. *Journal of the American Statistical Association*, 87(419):696–707.
- Gnanadesikan, R. and Kettenring, J. R. (1972). Robust estimates, residuals, and outlier detection with multiresponse data. *Biometrics*, 28(1):81–124.
- Hampel, F. R. (1968). *Contributions to the Theory of Robust Estimation.* Ph.D. dissertation, University of California, Berkeley.
- Hampel, F. R. (1971). A general qualitative definition of robustness. *The Annals of Mathematical Statistics*, 42(6):1887–1896.

- Hampel, F. R. (1973). Robust estimation: A condensed partial survey. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*, 27(2):87–104.
- 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., Ronchetti, E. M., Rousseeuw, P. J., and Stahel, W. (1986). *Robust Statistics: The Approach Based on Influence Functions*. John Wiley & Sons, New York.
- Hawkins, D. M. and Olive, D. J. (1999). Improved feasible solution algorithms for high breakdown estimation. *Computational Statistics & Data Analysis*, 30(1):1–11.
- He, X. and Simpson, D. G. (1993). Lower bounds for contamination bias: Globally minimax versus locally linear estimation. *The Annals of Statistics*, 21(1):314–337.
- Hodges Jr, J. L. (1967). Efficiency in normal samples and tolerance of extreme values for some estimates of location. In *Proceedings of the fifth Berkeley Symposium on Mathematical Statistics and Probability*, volume 1, pages 163–186, Oakland, California. Univ of California Press.
- Hotelling, H. (1992). *The Generalization of Student's Ratio*, pages 54–65. Springer, New York.
- Huber, P. J. (1972). The 1972 Wald Lecture Robust Statistics: A Review. *The Annals of Mathematical Statistics*, 43(4):1041–1067.
- Huber, P. J. (1977). Robust covariances. In *Statistical Decision Theory and Related Topics*, pages 165–191. Academic Press, Purdue University.
- Huber, P. J. (1992). Robust estimation of a location parameter. In *Breakthroughs in Statistics: Methodology and Distribution*, pages 492–518. Springer, New York, NY.
- Huber, P. J. and Ronchetti, E. M. (2011). *Robust Statistics*. John Wiley & Sons, Germany.

- Hubert, M. and Debruyne, M. (2010). Minimum covariance determinant. *Wiley Interdisciplinary Reviews: Computational Statistics*, 2(1):36–43.
- Hubert, M., Rousseeuw, P. J., and Vanden Branden, K. (2005). ROBPCA: A new approach to robust principal component analysis. *Technometrics*, 47(1):64–79.
- Hubert, M., Rousseeuw, P. J., and Verdonck, T. (2012). A deterministic algorithm for robust location and scatter. *Journal of Computational and Graphical Statistics*, 21(3):618–637.
- Johnson, D. and Petkau, A. (1995). A likelihood ratio test for equality of deviations from randomness. *Researches on Population Ecology*, 37(2):203–209.
- Johnson, R. A. and Wichern, D. W. (1998). *Applied Multivariate Statistical Analysis*. Prentice-Hall, Upper Saddle River, NJ, 4th edition.
- Jurczyk, T. (2008). High breakdown point estimation in regression. In *WDS*, volume 8, pages 94–99. Citeseer.
- Kalina, J. (2022). The minimum weighted covariance determinant estimator revisited. *Communications in Statistics-Simulation and Computation*, 51(7):3888–3900.
- Kent, J. T. and Tyler, D. E. (1996). Constrained M-estimation for multivariate location and scatter. *The Annals of Statistics*, 24(3):1346–1370.
- Lehmann, E. L. and Casella, G. (1998). *Theory of Point Estimation*. Springer-Verlag, New York, 2nd edition.
- Lopuhaä, H. P. (1999). Asymptotics of reweighted estimators of multivariate location and scatter. *Annals of Statistics*, 27(5):1638–1665.
- Lopuhaä, H. P. and Rousseeuw, P. J. (1991). Breakdown points of affine equivariant estimators of multivariate location and covariance matrices. *The Annals of Statistics*, 19(1):229–248.

- Ma, Y. and Genton, M. G. (2000). Highly robust estimation of the autocovariance function. *Journal of Time Series Analysis*, 21(6):663–684.
- Maronna, R. A. (1976). Robust M-estimators of multivariate location and scatter. *The Annals of Statistics*, 4(1):51–67.
- Maronna, R. A., Martin, R. D., Yohai, V. J., and Salibián-Barrera, M. (2019). *Robust Statistics: Theory and Methods (with R)*. John Wiley & Sons, Montrea, Canada.
- Mendelson, S. and Zhivotovskiy, N. (2020). Robust covariance estimation under L_4 – L_2 norm equivalence. *The Annals of Statistics*, 48(3):1648–1664.
- Morris, A., Whittaker, J., and Curnow, R. (1997). A likelihood ratio test for detecting patterns of disease-marker association. *Annals of Human Genetics*, 61(4):335–350.
- Muthukrishnan, R. and Mahesh, K. (2014). Robust procedure for estimating multivariate location and scatter. *American Inter-national Journal of Research in Science, Technology. Eng. Math*, 6:189–195.
- Nath, R. and Pavur, R. (1985). A new statistic in the one-way multivariate analysis of variance. *Computational Statistics & Data Analysis*, 2(4):297–315.
- Nkum, E., Pokojovy, M., and Mandal, A. (2024). Robust manova with minimum density power divergence estimator. Unpublished working paper.
- Ostrovskii, D. M. and Rudi, A. (2019). Affine invariant covariance estimation for heavy-tailed distributions. In *Conference on Learning Theory*, pages 2531–2550. PMLR.
- Pardo, L. (2018). *Statistical Inference Based on Divergence Measures*. Chapman and Hall/CRC, New York, 1st edition.
- Peña, D. and Prieto, F. J. (2001). Multivariate outlier detection and robust covariance matrix estimation. *Technometrics*, 43(3):286–310.
- Peracchi, F. (1990). Robust M-estimators. *Econometric Reviews*, 9(1):1–30.

- Petersen, K. B., Pedersen, M. S., et al. (2008). The Matrix Cookbook. *Technical University of Denmark, Kongens Lyngby, Denmark*, 7(15):510.
- Pison, G. and Van Aelst, S. (2004). Diagnostic plots for robust multivariate methods. *Journal of Computational and Graphical Statistics*, 13(2):310–329.
- Pison, G., Van Aelst, S., and Willems, G. (2002). Small sample corrections for LTS and MCD. *Metrika*, 55:111–123.
- Pokojovy, M., Anum, A. T., Nkum, E., and Mandal, A. (2024a). An new computational algorithm for Gaussian minimum density divergence estimation. Unpublished manuscript.
- Pokojovy, M., Anum, A. T., Nkum, E., and Mandal, A. (2024b). MDPD R package. <https://github.com/mpokojovy/mdpd>.
- Pokojovy, M. and Jobe, J. M. (2022). A robust deterministic affine-equivariant algorithm for multivariate location and scatter. *Computational Statistics & Data Analysis*, 172:107475.
- Randall, R. L., Arthur Woodward, J., and Bonett, D. G. (1997). A Wald test for the multivariate analysis of variance: Small sample critical values. *Communications in Statistics-Simulation and Computation*, 26(4):1275–1299.
- Reimann, C., Arnoldussen, A., Boyd, R., Finne, T. E., Koller, F., Nordgulen, Ø., and Englmaier, P. (2007). Element contents in leaves of four plant species (birch, mountain ash, fern and spruce) along anthropogenic and geogenic concentration gradients. *Science of the Total Environment*, 377(2-3):416–433.
- Riani, M., Atkinson, A. C., and Cerioli, A. (2009). Finding an unknown number of multivariate outliers. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 71(2):447–466.
- Roelant, E., Van Aelst, S., and Willems, G. (2009). The minimum weighted covariance determinant estimator. *Metrika*, 70(2):177–204.

- Ronchetti, E. M. and Huber, P. J. (2009). *Robust Statistics*. John Wiley & Sons, Hoboken, NJ, USA.
- Rousseeuw, P. J. (1984). Least median of squares regression. *Journal of the American Statistical Association*, 79(388):871–880.
- Rousseeuw, P. J. (1985). Multivariate estimation with high breakdown point. *Mathematical Statistics and Applications*, 8(283-297):37.
- Rousseeuw, P. J. and Croux, C. (1993). Alternatives to the median absolute deviation. *Journal of the American Statistical Association*, 88(424):1273–1283.
- Rousseeuw, P. J. and Driessen, K. V. (1999). A fast algorithm for the minimum covariance determinant estimator. *Technometrics*, 41(3):212–223.
- Rousseeuw, P. J. and Leroy, A. M. (2005). *Robust Regression and Outlier Detection*. John Wiley & Sons, Hoboken, New Jersey.
- Rousseeuw, P. J., Van Aelst, S., Van Driessen, K., and Gulló, J. A. (2004). Robust multivariate regression. *Technometrics*, 46(3):293–305.
- Salibián-Barrera, M., Van Aelst, S., and Willems, G. (2006). Principal components analysis based on multivariate MM-estimators with fast and robust bootstrap. *Journal of the American Statistical Association*, 101(475):1198–1211.
- Shoemaker, L. H. (1984). Robustness properties for a class of scale estimators. *Communications in Statistics-Theory and Methods*, 13(1):15–28.
- Stahel, W. (1981). *Robuste Schätzungen: Infinitesimale Optimalität und Schätzungen von Kovarianzmatrizen*. PhD thesis, 1981, ETH, Zurich, Switzerland.
- Stromberg, A. J. and Ruppert, D. (1992). Breakdown in nonlinear regression. *Journal of the American Statistical Association*, 87(420):991–997.

- Sunanta, O. (2018). Generalized point estimators for fuzzy multivariate data. *Austrian Journal of Statistics*, 47(1):33–44.
- Todorov, V. and Filzmoser, P. (2010). Robust statistic for the one-way MANOVA. *Computational Statistics & Data Analysis*, 54(1):37–48.
- Tukey, J. W. (1960). A survey of sampling from contaminated distributions. In *Contributions to Probability and Statistics*, pages 448–485, Redwood City, California. Stanford University Press. <https://api.semanticscholar.org/CorpusID:238386417>.
- Van Aelst, S. and Willems, G. (2011). Robust and efficient one-way MANOVA tests. *Journal of the American Statistical Association*, 106(494):706–718.
- Visuri, S., Koivunen, V., and Oja, H. (2000). Sign and rank covariance matrices. *Journal of Statistical Planning and Inference*, 91(2):557–575.
- Whiting, P., Lee, J., Roy, D. N., and Hong, M. (2018). Relationships between properties of pulp-fibre and paper. In Baker, C. F., editor, *Products of Papermaking, Transactions of the Xth Fundamental Research Symposium*, pages 159–182. FRC.
- Willems, G., Joe, H., and Zamar, R. (2009). Diagnosing multivariate outliers detected by robust estimators. *Journal of Computational and Graphical Statistics*, 18(1):73–91.
- Windham, M. P. (1995). Robustifying model fitting. *Journal of the Royal Statistical Society. Series B (Methodological)*, 57(3):599–609.
- Xu, J. and Cui, X. (2008). Robustified manova with applications in detecting differentially expressed genes from oligonucleotide arrays. *Bioinformatics*, 24(8):1056–1062.
- Yu, K., Dang, X., and Chen, Y. (2015). Robustness of the affine equivariant scatter estimator based on the spatial rank covariance matrix. *Communications in Statistics-Theory and Methods*, 44(5):914–932.
- Zuo, Y. (2004). Projection-based affine equivariant multivariate location estimators with the best possible finite sample breakdown point. *Statistica Sinica*, 14(4):1199–1208.

- Zuo, Y. (2006). Robust location and scatter estimators in multivariate analysis. In *Frontiers in Statistics*, pages 467–490, Lausanne, Switzerland. World Scientific.
- Zuo, Y. and Lai, S. (2009). On a robust and efficient maximum depth estimator. *Science in China Series A: Mathematics*, 52(6):1212–1232.

Appendix A

A.1 Integrals for DPD Measure

Suppose $\mathbf{X} \sim N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. The first term of the DPD measure in Equation (4.1) under the multivariate Gaussian model is given by:

$$\begin{aligned}
 \int f_{\boldsymbol{\theta}}^{1+\alpha}(\mathbf{x}) d\mathbf{x} &= \int \left(\left(\frac{1}{(2\pi)^{p/2} |\boldsymbol{\Sigma}|^{1/2}} \right)^{1+\alpha} \exp \left(-\frac{1+\alpha}{2} (\mathbf{x} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1} (\mathbf{x} - \boldsymbol{\mu}) \right) \right) d\mathbf{x} \\
 &= \frac{1}{(2\pi)^{p(1+\alpha)/2} |\boldsymbol{\Sigma}|^{(1+\alpha)/2}} \int \exp \left(-\frac{1+\alpha}{2} (\mathbf{x} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1} (\mathbf{x} - \boldsymbol{\mu}) \right) d\mathbf{x} \\
 &= \frac{1}{(2\pi)^{p(1+\alpha)/2} |\boldsymbol{\Sigma}|^{(1+\alpha)/2}} (2\pi)^{p/2} \left(\frac{1}{1+\alpha} \right)^{p/2} |\boldsymbol{\Sigma}|^{1/2} \\
 &= (2\pi)^{-p\alpha/2} (1+\alpha)^{-p/2} |\boldsymbol{\Sigma}|^{-\alpha/2}.
 \end{aligned} \tag{A.1}$$

We will compute a few more integrals required for the asymptotic distribution of the MDPD estimator.

$$\begin{aligned}
 &\int (\mathbf{x} - \boldsymbol{\mu}) f_{\boldsymbol{\theta}}^{1+\alpha}(\mathbf{x}) d\mathbf{x} \\
 &= \int (\mathbf{x} - \boldsymbol{\mu}) \left(\frac{1}{(2\pi)^{p/2} |\boldsymbol{\Sigma}|^{1/2}} \right)^{1+\alpha} \exp \left(-\frac{1+\alpha}{2} (\mathbf{x} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1} (\mathbf{x} - \boldsymbol{\mu}) \right) d\mathbf{x} \tag{A.2} \\
 &= \mathbf{0}.
 \end{aligned}$$

Since the first moment is an odd function over the symmetric range, the integral evaluates to zero.

Consider the integral of the second moment for a multivariate Gaussian distribution:

$$\int (\mathbf{x} - \boldsymbol{\mu})(\mathbf{x} - \boldsymbol{\mu})^T f_{\boldsymbol{\theta}}^{1+\alpha}(\mathbf{x}) d\mathbf{x}. \tag{A.3}$$

By substituting the density function f_{θ} of the multivariate Gaussian distribution, we get:

$$\begin{aligned}
& \int (\mathbf{x} - \boldsymbol{\mu})(\mathbf{x} - \boldsymbol{\mu})^T \left(\frac{1}{(2\pi)^{p/2} |\boldsymbol{\Sigma}|^{1/2}} \right)^{1+\alpha} \exp \left(-\frac{1+\alpha}{2} (\mathbf{x} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1} (\mathbf{x} - \boldsymbol{\mu}) \right) d\mathbf{x} \\
&= \left(\frac{1}{(2\pi)^{p/2} |\boldsymbol{\Sigma}|^{1/2}} \right)^{1+\alpha} \int (\mathbf{x} - \boldsymbol{\mu})(\mathbf{x} - \boldsymbol{\mu})^T \exp \left(-\frac{1+\alpha}{2} (\mathbf{x} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1} (\mathbf{x} - \boldsymbol{\mu}) \right) d\mathbf{x} \\
&= (2\pi)^{-p\alpha/2} |\boldsymbol{\Sigma}|^{-\alpha/2} \int \frac{1}{(2\pi)^{p/2} |\boldsymbol{\Sigma}|^{1/2}} (\mathbf{x} - \boldsymbol{\mu})(\mathbf{x} - \boldsymbol{\mu})^T \\
&\quad \times \exp \left(-\frac{1+\alpha}{2} (\mathbf{x} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1} (\mathbf{x} - \boldsymbol{\mu}) \right) d\mathbf{x} \\
&= (2\pi)^{-p\alpha/2} |\boldsymbol{\Sigma}|^{-\alpha/2} \int \frac{(1+\alpha)^{-p/2}}{(2\pi)^{p/2} |\boldsymbol{\Sigma}|^{1/2} (1+\alpha)^{-p/2}} (\mathbf{x} - \boldsymbol{\mu})(\mathbf{x} - \boldsymbol{\mu})^T \\
&\quad \times \exp \left(-\frac{1+\alpha}{2} (\mathbf{x} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1} (\mathbf{x} - \boldsymbol{\mu}) \right) d\mathbf{x}.
\end{aligned}$$

For a multivariate Gaussian distribution, we know from standard results (refer to [Petersen et al. \(2008\)](#)) that: $\mathbb{E}[(\mathbf{x} - \boldsymbol{\mu})(\mathbf{x} - \boldsymbol{\mu})^T] = \boldsymbol{\Sigma}$.

Therefore, we have

$$\begin{aligned}
& \int \frac{1}{(2\pi)^{p/2} |\boldsymbol{\Sigma}|^{1/2} (1+\alpha)^{-p/2}} (\mathbf{x} - \boldsymbol{\mu})(\mathbf{x} - \boldsymbol{\mu})^T \exp \left(-\frac{1+\alpha}{2} (\mathbf{x} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1} (\mathbf{x} - \boldsymbol{\mu}) \right) d\mathbf{x} \\
&= \frac{\boldsymbol{\Sigma}}{1+\alpha}.
\end{aligned}$$

Thus, the integral in (A.3) simplifies to:

$$\begin{aligned}
& \int (\mathbf{x} - \boldsymbol{\mu})(\mathbf{x} - \boldsymbol{\mu})^T f_{\theta}^{1+\alpha}(\mathbf{x}) d\mathbf{x} \\
&= (2\pi)^{-p\alpha/2} |\boldsymbol{\Sigma}|^{-\alpha/2} (1+\alpha)^{-p/2} \frac{\boldsymbol{\Sigma}}{1+\alpha} \\
&= (2\pi)^{-p\alpha/2} (1+\alpha)^{(-2-p)/2} |\boldsymbol{\Sigma}|^{-\alpha/2} \boldsymbol{\Sigma}.
\end{aligned} \tag{A.4}$$

A.2 Estimating Equations

Let's consider the probability density function of \mathbf{y}_{ij} in the MANOVA model:

$$f_{\theta}(\mathbf{y}_{ij}) = (2\pi)^{-p/2} |\boldsymbol{\Sigma}|^{-1/2} \exp \left\{ -\frac{1}{2} (\mathbf{y}_{ij} - \boldsymbol{\mu}_i)^T \boldsymbol{\Sigma}^{-1} (\mathbf{y}_{ij} - \boldsymbol{\mu}_i) \right\}. \tag{A.5}$$

For $\alpha > 0$, the DPD measure empirically can be written as

$$\widehat{d}_\alpha(f_\theta, g) = \frac{1}{N} \sum_{i=1}^k \sum_{j=1}^{n_i} \int f_\theta^{1+\alpha}(\mathbf{y}_{ij}) d\mathbf{y}_{ij} - \frac{1+\alpha}{N\alpha} \sum_{i=1}^k \sum_{j=1}^{n_i} f_\theta^\alpha(\mathbf{y}_{ij}) + c(\alpha), \quad (\text{A.6})$$

where $c(\alpha) = \frac{1}{N\alpha} \sum_{i=1}^k \sum_{j=1}^{n_i} \int g^{1+\alpha}(\mathbf{y}_{ij}) d\mathbf{y}_{ij}$ is independent of θ .

From Equation (A.1), we have

$$\int f_\theta^{1+\alpha}(\mathbf{y}_{ij}) d\mathbf{y}_{ij} = (2\pi)^{-\alpha p/2} |\Sigma|^{-\alpha/2} (1+\alpha)^{-p/2}.$$

Therefore

$$\begin{aligned} \widehat{d}_\alpha(f_\theta, g) &= (2\pi)^{-\alpha p/2} |\Sigma|^{-\alpha/2} (1+\alpha)^{-p/2} \\ &\times \left[1 - \frac{1+\alpha}{N\alpha} \sum_{i=1}^k \sum_{j=1}^{n_i} \exp \left\{ -\frac{\alpha}{2} (\mathbf{y}_{ij} - \boldsymbol{\mu}_i)^T \Sigma^{-1} (\mathbf{y}_{ij} - \boldsymbol{\mu}_i) \right\} \right] + c(\alpha). \end{aligned} \quad (\text{A.7})$$

For the estimating equations, we have

$$\begin{aligned} \frac{\partial}{\partial \boldsymbol{\mu}_i} \widehat{d}_\alpha(f_\theta, g) &= 0 \text{ for } i = 1, 2, 3 \dots k, \\ \frac{\partial}{\partial \Sigma} \widehat{d}_\alpha(f_\theta, g) &= 0. \end{aligned}$$

Now

$$\begin{aligned} \frac{\partial}{\partial \boldsymbol{\mu}_i} \widehat{d}_\alpha(f_\theta, g) &= -(2\pi)^{-\alpha p/2} |\Sigma|^{-\alpha/2} \frac{\alpha(1+\alpha)}{2N\alpha} \\ &\times \sum_{j=1}^{n_i} \Sigma^{-1} (\mathbf{y}_{ij} - \boldsymbol{\mu}_i) \exp \left\{ -\frac{\alpha}{2} (\mathbf{y}_{ij} - \boldsymbol{\mu}_i)^T \Sigma^{-1} (\mathbf{y}_{ij} - \boldsymbol{\mu}_i) \right\} \\ &= \sum_{j=1}^{n_i} (\mathbf{y}_{ij} - \boldsymbol{\mu}_i) \exp \left\{ -\frac{\alpha}{2} (\mathbf{y}_{ij} - \boldsymbol{\mu}_i)^T \Sigma^{-1} (\mathbf{y}_{ij} - \boldsymbol{\mu}_i) \right\}, \end{aligned} \quad (\text{A.8})$$

$$\begin{aligned}
\frac{\partial}{\partial \Sigma} \widehat{d}_\alpha(f_{\boldsymbol{\theta}}, g) &= -\frac{\alpha}{2} (2\pi)^{-\alpha p/2} |\Sigma|^{-\alpha/2} \Sigma^{-1} (1+\alpha)^{-p/2} \\
&\quad - (2\pi)^{-\alpha p/2} |\Sigma|^{-\alpha/2} \frac{\alpha(1+\alpha)}{2N\alpha} \sum_{i=1}^k \sum_{j=1}^{n_i} \Sigma^{-1} \exp \left\{ -\frac{\alpha}{2} (\mathbf{y}_{ij} - \boldsymbol{\mu}_i)^T \Sigma^{-1} (\mathbf{y}_{ij} - \boldsymbol{\mu}_i) \right\} \\
&\quad - (2\pi)^{-\alpha p/2} |\Sigma|^{-\alpha/2} \frac{\alpha(1+\alpha)}{2N\alpha} \\
&\quad \times \sum_{i=1}^k \sum_{j=1}^{n_i} \Sigma^{-1} (\mathbf{y}_{ij} - \boldsymbol{\mu}_i) (\mathbf{y}_{ij} - \boldsymbol{\mu}_i)^T \Sigma^{-1} \exp \left\{ -\frac{\alpha}{2} (\mathbf{y}_{ij} - \boldsymbol{\mu}_i)^T \Sigma^{-1} (\mathbf{y}_{ij} - \boldsymbol{\mu}_i) \right\} \\
&= (1+\alpha)^{-p/2} \mathbf{I}_p - \frac{(1+\alpha)}{N\alpha} \sum_{i=1}^k \sum_{j=1}^{n_i} \exp \left\{ -\frac{\alpha}{2} (\mathbf{y}_{ij} - \boldsymbol{\mu}_i)^T \Sigma^{-1} (\mathbf{y}_{ij} - \boldsymbol{\mu}_i) \right\} \mathbf{I}_p \\
&\quad + \frac{(1+\alpha)}{N\alpha} \sum_{i=1}^k \sum_{j=1}^{n_i} \Sigma^{-1} (\mathbf{y}_{ij} - \boldsymbol{\mu}_i) (\mathbf{y}_{ij} - \boldsymbol{\mu}_i)^T \exp \left\{ -\frac{\alpha}{2} (\mathbf{y}_{ij} - \boldsymbol{\mu}_i)^T \Sigma^{-1} (\mathbf{y}_{ij} - \boldsymbol{\mu}_i) \right\}.
\end{aligned} \tag{A.9}$$

Thus, the estimating equations of $\boldsymbol{\theta}$ are simplified as:

$$\begin{aligned}
&\sum_{j=1}^{n_i} (\mathbf{y}_{ij} - \boldsymbol{\mu}_i) \exp \left\{ -\frac{\alpha}{2} (\mathbf{y}_{ij} - \boldsymbol{\mu}_i)^T \Sigma^{-1} (\mathbf{y}_{ij} - \boldsymbol{\mu}_i) \right\} = \mathbf{0}, \text{ for } i = 1, 2, 3 \dots k, \\
&\Sigma \left(\sum_{i=1}^k \sum_{j=1}^{n_i} \exp \left\{ -\frac{\alpha}{2} (\mathbf{y}_{ij} - \boldsymbol{\mu}_i)^T \Sigma^{-1} (\mathbf{y}_{ij} - \boldsymbol{\mu}_i) \right\} - \alpha N (1+\alpha)^{-\frac{p}{2}-1} \right) \\
&= \sum_{i=1}^k \sum_{j=1}^{n_i} (\mathbf{y}_{ij} - \boldsymbol{\mu}_i) (\mathbf{y}_{ij} - \boldsymbol{\mu}_i)^T \exp \left\{ -\frac{\alpha}{2} (\mathbf{y}_{ij} - \boldsymbol{\mu}_i)^T \Sigma^{-1} (\mathbf{y}_{ij} - \boldsymbol{\mu}_i) \right\}.
\end{aligned} \tag{A.10}$$

A.3 Asymptotic Distribution of MDPD Estimator

Let's consider the MANOVA model in (8.1), where $\boldsymbol{\theta} = (\boldsymbol{\mu}_1^T, \boldsymbol{\mu}_2^T, \dots, \boldsymbol{\mu}_k^T, \text{vec}(\Sigma)^T)^T$ is the parameter of the model. Suppose that the true distribution \mathbf{g} belongs to the parametric family $\{f_{\boldsymbol{\theta}}(\mathbf{x})\}$, where $\boldsymbol{\theta}$ represents the true parameter value. The matrices \mathbf{J} , \mathbf{K} , and the vector $\boldsymbol{\xi}_{\boldsymbol{\theta}}$ can be simplified as follows:

$$\begin{aligned}
\mathbf{J} &= \int_{\mathbb{R}^p} (\mathbf{u}_{\boldsymbol{\theta}}(\mathbf{x}) \otimes \mathbf{u}_{\boldsymbol{\theta}}(\mathbf{x})) f_{\boldsymbol{\theta}}^{1+\alpha}(\mathbf{x}) d\mathbf{x}, \\
\mathbf{K} &= \int_{\mathbb{R}^p} (\mathbf{u}_{\boldsymbol{\theta}}(\mathbf{x}) \otimes \mathbf{u}_{\boldsymbol{\theta}}(\mathbf{x})) f_{\boldsymbol{\theta}}^{1+2\alpha}(\mathbf{x}) d\mathbf{x} - \boldsymbol{\xi}_{\boldsymbol{\theta}} \otimes \boldsymbol{\xi}_{\boldsymbol{\theta}}, \\
\boldsymbol{\xi}_{\boldsymbol{\theta}} &= \int_{\mathbb{R}^p} \mathbf{u}_{\boldsymbol{\theta}}(\mathbf{x}) f_{\boldsymbol{\theta}}^{1+\alpha}(\mathbf{x}) d\mathbf{x}.
\end{aligned}$$

To derive the asymptotic distribution of $\widehat{\boldsymbol{\theta}}$, we need to compute \mathbf{J} and \mathbf{K} , and consequently, $\mathbf{J}^{-1}\mathbf{K}\mathbf{J}^{-1}$. These components are critical in determining the asymptotic variance of the minimum density power divergence (MDPD) estimator of $\widehat{\boldsymbol{\theta}}$.

The matrix \mathbf{J} can be written as:

$$\mathbf{J} = \begin{pmatrix} \mathbf{J}_{\mu_1\mu_1} & \mathbf{J}_{\mu_1\mu_2} & \cdots & \mathbf{J}_{\mu_1\mu_k} & \mathbf{J}_{\mu_1\Sigma} \\ \cdot & \mathbf{J}_{\mu_2\mu_2} & \cdots & \mathbf{J}_{\mu_2\mu_k} & \mathbf{J}_{\mu_2\Sigma} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \cdot & \cdot & \cdots & \mathbf{J}_{\mu_k\mu_k} & \mathbf{J}_{\mu_k\Sigma} \\ \cdot & \cdot & \cdots & \cdot & \mathbf{J}_{\Sigma\Sigma} \end{pmatrix},$$

Similarly, we decompose matrix \mathbf{K} based on μ_i and Σ . We will now calculate different components of \mathbf{J} and \mathbf{K} .

A.4 Asymptotic Variance of MDPD Estimator of μ

A.4.1 Score Function of μ

Let $f_{\boldsymbol{\theta}}(\mathbf{x})$ be the density function of a multivariate normal distribution:

$$f_{\boldsymbol{\theta}}(\mathbf{x}) = \frac{1}{(2\pi)^{p/2}|\boldsymbol{\Sigma}|^{1/2}} \exp\left(-\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1}(\mathbf{x} - \boldsymbol{\mu})\right), \quad (\text{A.11})$$

where $\boldsymbol{\mu}$ is the mean vector, $\boldsymbol{\Sigma}$ is the covariance matrix, and p is the dimensionality of \mathbf{x} .

The score function with respect to $\boldsymbol{\mu}$ is:

$$\begin{aligned} \mathbf{u}_{\boldsymbol{\mu}}(\mathbf{x}) &= \frac{\partial}{\partial \boldsymbol{\mu}} \log f_{\boldsymbol{\theta}}(\mathbf{x}) \\ &= \frac{\partial}{\partial \boldsymbol{\mu}} \left[-\frac{p}{2} \log(2\pi) - \frac{1}{2} \log |\boldsymbol{\Sigma}| - \frac{1}{2}(\mathbf{x} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1}(\mathbf{x} - \boldsymbol{\mu}) \right] \\ &= \boldsymbol{\Sigma}^{-1}(\mathbf{x} - \boldsymbol{\mu}). \end{aligned} \quad (\text{A.12})$$

A.4.2 Calculation of $\mathbf{J}_{\mu\mu}$

The calculation of the information matrix component $\mathbf{J}_{\mu\mu}$ is given by:

$$\begin{aligned}
\mathbf{J}_{\mu\mu} &= \int \mathbf{u}_\mu(\mathbf{x}) \otimes \mathbf{u}_\mu(\mathbf{x}) f_\theta^{1+\alpha}(\mathbf{x}) d\mathbf{x} \\
&= \int \mathbf{u}_\mu(\mathbf{x}) \mathbf{u}_\mu^T(\mathbf{x}) f_\theta^{1+\alpha}(\mathbf{x}) d\mathbf{x} \\
&= \int \Sigma^{-1}(\mathbf{x} - \boldsymbol{\mu}) (\Sigma^{-1}(\mathbf{x} - \boldsymbol{\mu}))^T f_\theta^{1+\alpha}(\mathbf{x}) d\mathbf{x} \\
&= \int \Sigma^{-1}(\mathbf{x} - \boldsymbol{\mu})(\mathbf{x} - \boldsymbol{\mu})^T \Sigma^{-1} f_\theta^{1+\alpha}(\mathbf{x}) d\mathbf{x} \\
&= \Sigma^{-1} \int (\mathbf{x} - \boldsymbol{\mu})(\mathbf{x} - \boldsymbol{\mu})^T f_\theta^{1+\alpha}(\mathbf{x}) d\mathbf{x} \Sigma^{-1}.
\end{aligned}$$

From Equation (A.4), we have:

$$\int (\mathbf{x} - \boldsymbol{\mu})(\mathbf{x} - \boldsymbol{\mu})^T f_\theta^{1+\alpha}(\mathbf{x}) d\mathbf{x} = (2\pi)^{-p\alpha/2} (1 + \alpha)^{(-2-p)/2} |\Sigma|^{-\alpha/2} \Sigma.$$

By substituting this result into the integral, we get:

$$\begin{aligned}
\mathbf{J}_{\mu\mu} &= \Sigma^{-1} (2\pi)^{-p\alpha/2} (1 + \alpha)^{(-2-p)/2} |\Sigma|^{-\alpha/2} \Sigma \Sigma^{-1} \\
&= (2\pi)^{-p\alpha/2} |\Sigma|^{-\alpha/2} (1 + \alpha)^{(-p-2)/2} \Sigma^{-1}.
\end{aligned} \tag{A.13}$$

A.4.3 Calculation of $\boldsymbol{\xi}_\mu$

The vector $\boldsymbol{\xi}_\mu$ is defined as:

$$\boldsymbol{\xi}_\mu = \int \mathbf{u}_\theta(\mathbf{x}) f_\theta^{1+\alpha}(\mathbf{x}) d\mathbf{x}. \tag{A.14}$$

From Equation (A.2), we have:

$$\begin{aligned}
\boldsymbol{\xi}_\mu &= \boldsymbol{\Sigma}^{-1} \int (\mathbf{x} - \boldsymbol{\mu}) \left(\frac{1}{(2\pi)^{p/2} |\boldsymbol{\Sigma}|^{1/2}} \right)^{1+\alpha} \exp \left(-\frac{(1+\alpha)}{2} (\mathbf{x} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1} (\mathbf{x} - \boldsymbol{\mu}) \right) d\mathbf{x} \\
&= \boldsymbol{\Sigma}^{-1} \left(\frac{1}{(2\pi)^{p/2} |\boldsymbol{\Sigma}|^{1/2}} \right)^{1+\alpha} \int (\mathbf{x} - \boldsymbol{\mu}) \exp \left(-\frac{(1+\alpha)}{2} (\mathbf{x} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1} (\mathbf{x} - \boldsymbol{\mu}) \right) d\mathbf{x}. \\
&= \mathbf{0}.
\end{aligned} \tag{A.15}$$

Because the integral of an odd function over the symmetric range of a Gaussian distribution is zero.

A.4.4 Calculation of $\mathbf{K}_{\mu\mu}$

The calculation of the matrix $\mathbf{K}_{\mu\mu}$ is given by:

$$\begin{aligned}
\mathbf{K}_{\mu\mu} &= \int \mathbf{u}_\mu(\mathbf{x}) \otimes \mathbf{u}_\mu(\mathbf{x}) f_\theta^{1+2\alpha}(\mathbf{x}) d\mathbf{x} - \boldsymbol{\xi}_\mu \otimes \boldsymbol{\xi}_\mu \\
&= \int \mathbf{u}_\mu(\mathbf{x}) \mathbf{u}_\mu^T(\mathbf{x}) f_\theta^{1+2\alpha}(\mathbf{x}) d\mathbf{x} - \boldsymbol{\xi}_\mu \boldsymbol{\xi}_\mu^T \\
&= \int \boldsymbol{\Sigma}^{-1} (\mathbf{x} - \boldsymbol{\mu}) (\boldsymbol{\Sigma}^{-1} (\mathbf{x} - \boldsymbol{\mu}))^T f_\theta^{1+2\alpha}(\mathbf{x}) d\mathbf{x}.
\end{aligned} \tag{A.16}$$

Here, we note that $\boldsymbol{\xi}_\mu = \mathbf{0}$. Additionally, from Equation (A.4), we have:

$$\int (\mathbf{x} - \boldsymbol{\mu}) (\mathbf{x} - \boldsymbol{\mu})^T f_\theta^{1+2\alpha}(\mathbf{x}) d\mathbf{x} = (2\pi)^{-p\alpha} (1 + 2\alpha)^{(-2-p)/2} |\boldsymbol{\Sigma}|^{-\alpha} \boldsymbol{\Sigma}.$$

Thus, we can express $\mathbf{K}_{\mu\mu}$ as:

$$\begin{aligned}
\mathbf{K}_{\mu\mu} &= \boldsymbol{\Sigma}^{-1} (2\pi)^{-p\alpha} |\boldsymbol{\Sigma}|^{-\alpha} (1 + 2\alpha)^{(-p-2)/2} \boldsymbol{\Sigma} \boldsymbol{\Sigma}^{-1} \\
&= (2\pi)^{-p\alpha} |\boldsymbol{\Sigma}|^{-\alpha} (1 + 2\alpha)^{(-p-2)/2} \boldsymbol{\Sigma}^{-1}.
\end{aligned} \tag{A.17}$$

A.4.5 Calculation of $\mathbf{J}_{\mu\mu}^{-1} \mathbf{K}_{\mu\mu} \mathbf{J}_{\mu\mu}^{-1}$

$$\begin{aligned}
\mathbf{J}_{\mu\mu}^{-1} \mathbf{K}_{\mu\mu} \mathbf{J}_{\mu\mu}^{-1} &= (2\pi)^{p\alpha/2} |\Sigma|^{\alpha/2} (1+\alpha)^{(p+2)/2} \Sigma \times (2\pi)^{-p\alpha} |\Sigma|^{-\alpha} (1+2\alpha)^{(-p-2)/2} \Sigma^{-1} \\
&\quad \times (2\pi)^{p\alpha/2} |\Sigma|^{\alpha/2} (1+\alpha)^{(p+2)/2} \Sigma. \\
&= (1+\alpha)^{(p+2)/2} \times (1+2\alpha)^{(-p-2)/2} \times (1+\alpha)^{(p+2)/2} \Sigma \\
&= \left(\frac{(1+\alpha)^2}{1+2\alpha} \right)^{(p+2)/2} \Sigma \\
&= \left(1 + \frac{\alpha^2}{1+2\alpha} \right)^{p/2+1} \Sigma.
\end{aligned} \tag{A.18}$$

A.5 Asymptotic Variance of MDPD Estimator of Σ

A.5.1 Score Function of Σ

The score function with respect to Σ is

$$\begin{aligned}
\mathbf{u}_{\Sigma}(\mathbf{x}) &= \frac{\partial}{\partial \Sigma} \log f_{\theta}(\mathbf{x}) \\
&= \frac{\partial}{\partial \Sigma} \left[-\frac{p}{2} \log(2\pi) - \frac{1}{2} \log |\Sigma| - \frac{1}{2} (\mathbf{x} - \boldsymbol{\mu})^T \Sigma^{-1} (\mathbf{x} - \boldsymbol{\mu}) \right] \\
&= -\frac{1}{2} (\Sigma^{-1} - \Sigma^{-1} (\mathbf{x} - \boldsymbol{\mu})(\mathbf{x} - \boldsymbol{\mu})^T \Sigma^{-1}).
\end{aligned} \tag{A.19}$$

Note that $\mathbf{u}_{\Sigma}(\mathbf{x})$ is defined as a matrix, whereas $\mathbf{u}_{\mu}(\mathbf{x})$ is a vector.

A.5.2 Calculation of $\mathbf{J}_{\Sigma\Sigma}$

Assume $\mathbf{x} \sim \mathcal{N}(\mathbf{0}, \Sigma)$ (refer to [Anderson, 2018](#), p. 67)

$$\mathbb{E}[\mathbf{x}\mathbf{x}^T \otimes \mathbf{x}\mathbf{x}^T] = \Sigma \otimes \Sigma + \mathbf{K}_{pp}(\Sigma \otimes \Sigma) + \text{vec}(\Sigma) \text{vec}(\Sigma)^T, \tag{A.20}$$

where vec is the vec-operator applied on matrix \mathbf{A} stacks the columns into a vector and \mathbf{K}_{pp} is the commutation matrix.

Also, for Kronecker product of $m \times n$ matrix \mathbf{A} and an $r \times q$ matrix \mathbf{B} (refer [Petersen et al., 2008](#))

$$\text{vec}(\mathbf{AXB}) = (\mathbf{B}^T \otimes \mathbf{A}) \text{vec}(\mathbf{X}). \quad (\text{A.21})$$

Substituting the score function in Equation (A.19), the calculation of the matrix $\mathbf{J}_{\Sigma\Sigma}$ is given by:

$$\begin{aligned} \mathbf{J}_{\Sigma\Sigma} &= \int \mathbf{u}_{\Sigma}(\mathbf{x}) \otimes \mathbf{u}_{\Sigma}(\mathbf{x}) f_{\theta}^{1+\alpha}(\mathbf{x}) d\mathbf{x} \\ &= \int \frac{1}{4} (\Sigma^{-1} - \Sigma^{-1}(\mathbf{x} - \mu)(\mathbf{x} - \mu)^T \Sigma^{-1}) \otimes (\Sigma^{-1} - \Sigma^{-1}(\mathbf{x} - \mu)(\mathbf{x} - \mu)^T \Sigma^{-1}) f_{\theta}^{1+\alpha}(\mathbf{x}) d\mathbf{x} \\ &= \frac{1}{4} \left\{ \int \Sigma^{-1} \otimes \Sigma^{-1} f_{\theta}^{1+\alpha}(\mathbf{x}) - 2\Sigma^{-1} \otimes \Sigma^{-1} \int (\mathbf{x} - \mu)(\mathbf{x} - \mu)^T f_{\theta}^{1+\alpha}(\mathbf{x}) d\mathbf{x} \Sigma^{-1} \right. \\ &\quad \left. + \Sigma^{-1} \int (\mathbf{x} - \mu)(\mathbf{x} - \mu)^T \Sigma^{-1} \otimes \Sigma^{-1}(\mathbf{x} - \mu)(\mathbf{x} - \mu)^T f_{\theta}^{1+\alpha}(\mathbf{x}) d\mathbf{x} \Sigma^{-1} \right\}, \end{aligned} \quad (\text{A.22})$$

From Equation (A.1), we have

$$\int \Sigma^{-1} \otimes \Sigma^{-1} f_{\theta}^{1+\alpha}(\mathbf{x}) = \Sigma^{-1} \otimes \Sigma^{-1} (2\pi)^{-p\alpha/2} (1 + \alpha)^{-p/2} |\Sigma|^{-\alpha/2}. \quad (\text{A.23})$$

From Equation (A.4), we have

$$\begin{aligned} &- 2\Sigma^{-1} \otimes \Sigma^{-1} \int (\mathbf{x} - \mu)(\mathbf{x} - \mu)^T f_{\theta}^{1+\alpha}(\mathbf{x}) d\mathbf{x} \Sigma^{-1} \\ &= -2\Sigma^{-1} \otimes \Sigma^{-1} (2\pi)^{-p\alpha/2} (1 + \alpha)^{(-2-p)/2} |\Sigma|^{-\alpha/2}. \end{aligned} \quad (\text{A.24})$$

And on the strength of Equation (A.20), let's evaluate

$$\begin{aligned} &\int \Sigma^{-1}(\mathbf{x} - \mu)(\mathbf{x} - \mu)^T \Sigma^{-1} \otimes \Sigma^{-1}(\mathbf{x} - \mu)(\mathbf{x} - \mu)^T \Sigma^{-1} f_{\theta}^{1+\alpha}(\mathbf{x}) d\mathbf{x} \\ &= \int (\Sigma^{-1} \otimes \Sigma^{-1}) ((\mathbf{x} - \mu)(\mathbf{x} - \mu)^T \otimes (\mathbf{x} - \mu)(\mathbf{x} - \mu)^T) (\Sigma^{-1} \otimes \Sigma^{-1}) f_{\theta}^{1+\alpha}(\mathbf{x}) d\mathbf{x} \\ &= (\Sigma^{-1} \otimes \Sigma^{-1}) \int ((\mathbf{x} - \mu)(\mathbf{x} - \mu)^T \otimes (\mathbf{x} - \mu)(\mathbf{x} - \mu)^T) f_{\theta}^{1+\alpha}(\mathbf{x}) d\mathbf{x} (\Sigma^{-1} \otimes \Sigma^{-1}). \end{aligned} \quad (\text{A.25})$$

Evaluating the middle integral using Equation (A.20) and Equation (A.21), we obtain:

$$\begin{aligned} &\int ((\mathbf{x} - \mu)(\mathbf{x} - \mu)^T \otimes (\mathbf{x} - \mu)(\mathbf{x} - \mu)^T) f_{\theta}^{1+\alpha}(\mathbf{x}) d\mathbf{x} \\ &= (2\pi)^{-p\alpha/2} |\Sigma|^{-\alpha/2} (1 + \alpha)^{-p/2} \mathbb{E}(\mathbf{y}\mathbf{y}^T \otimes \mathbf{y}\mathbf{y}^T), \quad \text{where } \mathbf{y} \sim N\left(\mathbf{0}, \frac{\Sigma}{1 + \alpha}\right) \\ &= (2\pi)^{-p\alpha/2} |\Sigma|^{-\alpha/2} (1 + \alpha)^{-p/2-2} (\Sigma \otimes \Sigma + \mathbf{K}_{pp}(\Sigma \otimes \Sigma) + \text{vec}(\Sigma) \text{vec}(\Sigma)^T). \end{aligned}$$

Therefore, we have

$$\begin{aligned}
& \int \boldsymbol{\Sigma}^{-1}(\mathbf{x} - \boldsymbol{\mu})(\mathbf{x} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{\Sigma}^{-1}(\mathbf{x} - \boldsymbol{\mu})(\mathbf{x} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1} f_{\boldsymbol{\theta}}^{1+\alpha}(\mathbf{x}) d\mathbf{x} \\
&= (2\pi)^{-p\alpha/2} |\boldsymbol{\Sigma}|^{-\alpha/2} (1 + \alpha)^{-p/2-2} (\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{\Sigma}^{-1}) \\
&\times (\boldsymbol{\Sigma} \otimes \boldsymbol{\Sigma} + \mathbf{K}_{pp}(\boldsymbol{\Sigma} \otimes \boldsymbol{\Sigma}) \text{vec}(\boldsymbol{\Sigma}) \text{vec}(\boldsymbol{\Sigma})^T) (\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{\Sigma}^{-1}) \\
&= (2\pi)^{-p\alpha/2} |\boldsymbol{\Sigma}|^{-\alpha/2} (1 + \alpha)^{-p/2-2} (\mathbf{I}_{pp} + \mathbf{K}_{pp} + \text{vec}(\boldsymbol{\Sigma}^{-1}) \text{vec}(\boldsymbol{\Sigma})^T) (\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{\Sigma}^{-1}).
\end{aligned} \tag{A.26}$$

Combining all the results, we have

$$\begin{aligned}
\mathbf{J}_{\boldsymbol{\Sigma}\boldsymbol{\Sigma}} &= \frac{1}{4} \left\{ \boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{\Sigma}^{-1} (2\pi)^{-p\alpha/2} (1 + \alpha)^{-p/2} |\boldsymbol{\Sigma}|^{-\alpha/2} \right. \\
&\quad - 2\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{\Sigma}^{-1} (2\pi)^{-p\alpha/2} (1 + \alpha)^{(-2-p)/2} |\boldsymbol{\Sigma}|^{-\alpha/2} \\
&\quad \left. + (2\pi)^{-p\alpha/2} |\boldsymbol{\Sigma}|^{-\alpha/2} (1 + \alpha)^{-p/2-2} (\mathbf{I}_{pp} + \mathbf{K}_{pp} + \text{vec}(\boldsymbol{\Sigma}^{-1}) \text{vec}(\boldsymbol{\Sigma})^T) (\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{\Sigma}^{-1}) \right\} \\
&= \frac{1}{4} \left\{ (2\pi)^{-p\alpha/2} |\boldsymbol{\Sigma}|^{-\alpha/2} (1 + \alpha)^{-p/2-2} \right. \\
&\quad \times ((1 + \alpha)^2 \mathbf{I}_p - 2(1 + \alpha) \mathbf{I}_{pp} + \mathbf{I}_{pp} + \mathbf{K}_{pp} + \text{vec}(\boldsymbol{\Sigma}^{-1}) \text{vec}(\boldsymbol{\Sigma})^T) \left. \right\} \\
&= \frac{1}{4} \left\{ (2\pi)^{-p\alpha/2} |\boldsymbol{\Sigma}|^{-\alpha/2} (1 + \alpha)^{-p/2-2} (\alpha^2 \mathbf{I}_{pp} + \mathbf{K}_{pp} + \text{vec}(\boldsymbol{\Sigma}^{-1}) \text{vec}(\boldsymbol{\Sigma})^T) (\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{\Sigma}^{-1}) \right\}.
\end{aligned} \tag{A.27}$$

A.5.3 Calculation of $\boldsymbol{\xi}_{\boldsymbol{\Sigma}}$

The vector $\boldsymbol{\xi}_{\boldsymbol{\Sigma}}$ is defined as:

$$\begin{aligned}
\boldsymbol{\xi}_{\boldsymbol{\Sigma}} &= \int \mathbf{u}_{\boldsymbol{\Sigma}}(\mathbf{x}) f_{\boldsymbol{\theta}}^{1+\alpha}(\mathbf{x}) d\mathbf{x}. \\
&= - \int \frac{1}{2} (\boldsymbol{\Sigma}^{-1} - \boldsymbol{\Sigma}^{-1}(\mathbf{x} - \boldsymbol{\mu})(\mathbf{x} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1}) f_{\boldsymbol{\theta}}^{1+\alpha}(\mathbf{x}) d\mathbf{x}. \\
&= - \frac{1}{2} \left((2\pi)^{-p\alpha/2} (1 + \alpha)^{-p/2} |\boldsymbol{\Sigma}|^{-\alpha/2} \boldsymbol{\Sigma}^{-1} - (2\pi)^{-p\alpha/2} (1 + \alpha)^{(-2-p)/2} |\boldsymbol{\Sigma}|^{-\alpha/2} \boldsymbol{\Sigma}^{-1} \right) \\
&= - \frac{1}{2} (2\pi)^{-p\alpha/2} (1 + \alpha)^{(-2-p)/2} |\boldsymbol{\Sigma}|^{-\alpha/2} \boldsymbol{\Sigma}^{-1} \alpha.
\end{aligned} \tag{A.28}$$

A.5.4 Calculation of $K_{\Sigma\Sigma}$

Substituting the score function in Equation A.19, the calculation of the matrix $K_{\Sigma\Sigma}$ is given by:

$$\begin{aligned}
K_{\Sigma\Sigma} &= \int \mathbf{u}_{\Sigma}(\mathbf{x}) \otimes \mathbf{u}_{\Sigma}(\mathbf{x}) f_{\theta}^{1+2\alpha}(\mathbf{x}) d\mathbf{x} - \xi_{\Sigma} \otimes \xi_{\Sigma}. \\
&= \int \frac{1}{4} (\Sigma^{-1} - \Sigma^{-1}(\mathbf{x} - \mu)(\mathbf{x} - \mu)\Sigma^{-1}) \otimes (\Sigma^{-1} - \Sigma^{-1}(\mathbf{x} - \mu)(\mathbf{x} - \mu)\Sigma^{-1}) f_{\theta}^{1+2\alpha}(\mathbf{x}) d\mathbf{x} \\
&\quad - \xi_{\Sigma} \otimes \xi_{\Sigma} \\
&= \frac{1}{4} \left\{ \int \Sigma^{-1} \otimes \Sigma^{-1} f_{\theta}^{1+2\alpha}(\mathbf{x}) - 2\Sigma^{-1} \otimes \Sigma^{-1} \int (\mathbf{x} - \mu)(\mathbf{x} - \mu)^T f_{\theta}^{1+2\alpha}(\mathbf{x}) d\mathbf{x} \Sigma^{-1} \right. \\
&\quad \left. + \int \Sigma^{-1}(\mathbf{x} - \mu)(\mathbf{x} - \mu)^T \Sigma^{-1} \otimes \Sigma^{-1}(\mathbf{x} - \mu)(\mathbf{x} - \mu)^T f_{\theta}^{1+2\alpha}(\mathbf{x}) d\mathbf{x} \Sigma^{-1} \right\} - \xi_{\Sigma} \otimes \xi_{\Sigma}.
\end{aligned} \tag{A.29}$$

From Equation (A.1), we have

$$\int \Sigma^{-1} \otimes \Sigma^{-1} f_{\theta}^{1+2\alpha}(\mathbf{x}) d\mathbf{x} = \Sigma^{-1} \otimes \Sigma^{-1} (2\pi)^{-p\alpha} (1 + 2\alpha)^{-p/2} |\Sigma|^{-\alpha}.$$

From Equation (A.4), we have

$$\begin{aligned}
&- 2\Sigma^{-1} \otimes \Sigma^{-1} \int (\mathbf{x} - \mu)(\mathbf{x} - \mu)^T f_{\theta}^{1+2\alpha}(\mathbf{x}) d\mathbf{x} \Sigma^{-1} \\
&= -2\Sigma^{-1} \otimes \Sigma^{-1} (2\pi)^{-p\alpha} (1 + 2\alpha)^{(-2-p)/2} |\Sigma|^{-\alpha}.
\end{aligned} \tag{A.30}$$

Similar to Equation (A.27)

$$\begin{aligned}
&\Sigma^{-1} \int (\mathbf{x} - \mu)(\mathbf{x} - \mu)^T \Sigma^{-1} \Sigma^{-1}(\mathbf{x} - \mu)(\mathbf{x} - \mu)^T \Sigma^{-1} f_{\theta}^{1+2\alpha}(\mathbf{x}) d\mathbf{x} \\
&= ((2\pi)^{-p\alpha} (1 + 2\alpha)^{-2-p/2} |\Sigma|^{-\alpha} (\mathbf{I}_{pp} + \mathbf{K}_{pp} + \text{vec}(\Sigma^{-1}) \text{vec}(\Sigma)^T)) \Sigma^{-1} \otimes \Sigma^{-1}.
\end{aligned} \tag{A.31}$$

Combining all the results, we obtain

$$\begin{aligned}
\mathbf{K}_{\Sigma\Sigma} &= \frac{1}{4} \left\{ \Sigma^{-1} \otimes \Sigma^{-1} (2\pi)^{-p\alpha} (1+2\alpha)^{-p/2} |\Sigma|^{-\alpha} \right. \\
&\quad - 2\Sigma^{-1} \otimes \Sigma^{-1} (2\pi)^{-p\alpha} (1+2\alpha)^{(-2-p)/2} |\Sigma|^{-\alpha} \\
&\quad + \left. \left((2\pi)^{-p\alpha} (1+2\alpha)^{-2-p/2} |\Sigma|^{-\alpha} (\mathbf{I}_{pp} + \mathbf{K}_{pp} + \text{vec}(\Sigma^{-1}) \text{vec}(\Sigma)^T) \right) \Sigma^{-1} \otimes \Sigma^{-1} \right\} \\
&\quad - \frac{1}{4} (2\pi)^{-p\alpha} (1+\alpha)^{-2-p/2} \alpha^2 |\Sigma|^{-\alpha} \Sigma^{-1} \otimes \Sigma^{-1} \\
&= (2\pi)^{-p\alpha} (1+2\alpha)^{-2-p/2} |\Sigma|^{-\alpha} \\
&\quad \times \left((1+2\alpha)^2 \mathbf{I}_I - 2(1+2\alpha) \mathbf{I}_{pp} + \mathbf{I}_{pp} + \mathbf{K}_{pp} + \text{vec}(\Sigma^{-1}) \text{vec}(\Sigma)^T \right) \Sigma^{-1} \otimes \Sigma^{-1} \\
&\quad - \frac{1}{4} (2\pi)^{-p\alpha} (1+\alpha)^{-2-p/2} \alpha^2 |\Sigma|^{-\alpha} \Sigma^{-1} \otimes \Sigma^{-1} \\
&= (2\pi)^{-p\alpha} (1+2\alpha)^{-2-p/2} |\Sigma|^{-\alpha} \left(4\alpha^2 \mathbf{I}_{pp} + \mathbf{K}_{pp} + \text{vec}(\Sigma^{-1}) \text{vec}(\Sigma)^T \right) \Sigma^{-1} \otimes \Sigma^{-1} \\
&\quad - \frac{1}{4} (2\pi)^{-p\alpha} (1+\alpha)^{-2-p} \alpha^2 |\Sigma|^{-\alpha} \Sigma^{-1} \otimes \Sigma^{-1}.
\end{aligned} \tag{A.32}$$

A.5.5 Calculation of $\mathbf{J}_{\Sigma\Sigma}^{-1}\mathbf{K}_{\Sigma\Sigma}\mathbf{J}_{\Sigma\Sigma}^{-1}$

Finally, $\mathbf{J}_{\Sigma\Sigma}^{-1}\mathbf{K}_{\Sigma\Sigma}\mathbf{J}_{\Sigma\Sigma}^{-1}$ can be simplified in the following steps:

$$\begin{aligned}
& \mathbf{J}_{\Sigma\Sigma}^{-1}\mathbf{K}_{\Sigma\Sigma}\mathbf{J}_{\Sigma\Sigma}^{-1} \\
&= 4 \left\{ (2\pi)^{p\alpha/2} |\Sigma|^{\alpha/2} (1+\alpha)^{p/2+2} \left(\alpha^2 \mathbf{I}_{pp} + \mathbf{K}_{pp} + \text{vec}(\Sigma^{-1}) \text{vec}(\Sigma)^T \right)^{-1} (\Sigma \otimes \Sigma) \right\} \\
&\times \left(\frac{1}{4} (2\pi)^{-p\alpha} (1+2\alpha)^{-2-p/2} |\Sigma|^{-\alpha} \left(4\alpha^2 \mathbf{I}_{pp} + \mathbf{K}_{pp} + \text{vec}(\Sigma^{-1}) \text{vec}(\Sigma)^T \right) \Sigma^{-1} \otimes \Sigma^{-1} \right. \\
&\quad \left. - \frac{1}{4} (2\pi)^{-p\alpha} (1+\alpha)^{-2-p} \alpha^2 |\Sigma|^{-\alpha} \Sigma^{-1} \otimes \Sigma^{-1} \right) \\
&\times 4 \left\{ (2\pi)^{p\alpha/2} |\Sigma|^{\alpha/2} (1+\alpha)^{p/2+2} \left(\alpha^2 \mathbf{I}_{pp} + \mathbf{K}_{pp} + \text{vec}(\Sigma^{-1}) \text{vec}(\Sigma)^T \right)^{-1} (\Sigma \otimes \Sigma) \right\} \\
&= \frac{4(1+\alpha)^{p+4}}{(1+2\alpha)^{(p+4)/2}} \left(\alpha^2 \mathbf{I}_{pp} + \mathbf{K}_{pp} + \text{vec}(\Sigma^{-1}) \text{vec}(\Sigma)^T \right)^{-1} (\Sigma \otimes \Sigma) \\
&\times \left(4\alpha^2 \mathbf{I}_{pp} + \mathbf{K}_{pp} + \text{vec}(\Sigma^{-1}) \text{vec}(\Sigma)^T \right) (\Sigma^{-1} \otimes \Sigma^{-1}) \\
&\times \left(\alpha^2 \mathbf{I}_{pp} + \mathbf{K}_{pp} + \text{vec}(\Sigma^{-1}) \text{vec}(\Sigma)^T \right)^{-1} (\Sigma \otimes \Sigma) - 4\alpha^2 (\Sigma \otimes \Sigma) \\
&= \frac{4(1+\alpha)^{p+4}}{(1+2\alpha)^{(p+4)/2}} \left\{ \left(\alpha^2 \mathbf{I}_{pp} + \mathbf{K}_{pp} + \text{vec}(\Sigma^{-1}) \text{vec}(\Sigma)^T \right)^{-1} \right. \\
&\times \left(4\alpha^2 \mathbf{I}_{pp} + \mathbf{K}_{pp} + \text{vec}(\Sigma^{-1}) \text{vec}(\Sigma)^T \right) \\
&\times \left. \left(\alpha^2 \mathbf{I}_{pp} + \mathbf{K}_{pp} + \text{vec}(\Sigma^{-1}) \text{vec}(\Sigma)^T \right)^{-1} - \alpha^2 \right\} (\Sigma \otimes \Sigma).
\end{aligned} \tag{A.33}$$

A.6 Covariance Matrices

A.6.1 Calculation of $\mathbf{J}_{\Sigma\mu}$

The matrix $\mathbf{J}_{\Sigma\mu}$ is given by:

$$\begin{aligned}
\mathbf{J}_{\Sigma\mu} &= \int \mathbf{u}_{\Sigma}(\mathbf{x}) \otimes \mathbf{u}_{\mu}(\mathbf{x}) f_{\theta}^{1+\alpha}(\mathbf{x}) d\mathbf{x} \\
&= - \int \frac{1}{2} (\Sigma^{-1} - \Sigma^{-1}(\mathbf{x} - \mu)(\mathbf{x} - \mu)^T \Sigma^{-1}) \otimes (\Sigma^{-1}(\mathbf{x} - \mu)) f_{\theta}^{1+\alpha}(\mathbf{x}) d\mathbf{x}.
\end{aligned} \tag{A.34}$$

To proceed, we decompose the integral into two parts, $\mathbf{I} = \mathbf{I}_1 + \mathbf{I}_2$, where:

$$\begin{aligned}\mathbf{I}_1 &= -\frac{1}{2}\Sigma^{-1} \otimes \Sigma^{-1} \int (\mathbf{x} - \boldsymbol{\mu}) f_{\boldsymbol{\theta}}^{1+\alpha}(\mathbf{x}) d\mathbf{x} \\ &= \mathbf{0}_{p^2 \times p} \quad \text{on the strength of Equation A.2.}\end{aligned}\tag{A.35}$$

Next, we consider \mathbf{I}_2 :

$$\begin{aligned}\mathbf{I}_2 &= \int (\Sigma^{-1}(\mathbf{x} - \boldsymbol{\mu})(\mathbf{x} - \boldsymbol{\mu})^T \Sigma^{-1}) \otimes (\Sigma^{-1}(\mathbf{x} - \boldsymbol{\mu})) f_{\boldsymbol{\theta}}^{1+\alpha}(\mathbf{x}) d\mathbf{x} \\ &= \Sigma^{-1}(\mathbf{x} - \boldsymbol{\mu})(\mathbf{x} - \boldsymbol{\mu})^T \Sigma^{-1} \otimes \Sigma^{-1}(\mathbf{x} - \boldsymbol{\mu}) f_{\boldsymbol{\theta}}^{1+\alpha}(\mathbf{x}) d\mathbf{x} \\ &= \Sigma^{-1}(\mathbf{x} - \boldsymbol{\mu})(\mathbf{x} - \boldsymbol{\mu})^T \Sigma^{-1} \otimes \Sigma^{-1}(\mathbf{x} - \boldsymbol{\mu}) d\mathbf{x} \\ &= \mathbf{0}_{p^2 \times p}.\end{aligned}\tag{A.36}$$

For each element of the Kronecker product in Equation (A.36), we recognize that the integral is over products of centered Gaussian variables. The terms involving $(z_i - \mu_i)$ where $i \neq j$ integrate to zero due to symmetry (mean is zero). Only terms where $p = q = k$ survive:

$$\int (x_p - \mu_p)^3 \exp\left(-\frac{1+\alpha}{2}(\mathbf{x} - \boldsymbol{\mu})^T \Sigma^{-1}(\mathbf{x} - \boldsymbol{\mu})\right) d\mathbf{x}.$$

Using the properties of the multivariate Gaussian distribution, we know that $\mathbb{E}[(x_p - \mu_p)^3] = 0$ for a Gaussian random variable (since odd moments of zero-mean Gaussian are zero). Thus, the integral evaluates to zero:

$$\sum_{p=1}^n a_{1p}^3 \int (x_p - \mu_p)^3 \exp\left(-\frac{1+\alpha}{2}(\mathbf{x} - \boldsymbol{\mu})^T \Sigma^{-1}(\mathbf{x} - \boldsymbol{\mu})\right) d\mathbf{x} = 0.$$

Thus

$$\Sigma^{-1}(\mathbf{x} - \boldsymbol{\mu})(\mathbf{x} - \boldsymbol{\mu})^T \Sigma^{-1} \otimes \Sigma^{-1}(\mathbf{x} - \boldsymbol{\mu}) d\mathbf{x} = \mathbf{0}_{p^2 \times p}.$$

Therefore

$$\mathbf{J}_{\Sigma\boldsymbol{\mu}} = \mathbf{0}_{p^2 \times p}.\tag{A.37}$$

A.6.2 Calculation of $K_{\Sigma\mu}$

Analogously, we have:

$$K_{\Sigma\mu} = \mathbf{0}_{p^2 \times p}. \quad (\text{A.38})$$

A.7 MANOVA Setup

A.7.1 J and K Matrices at the Model

We have

$$\mathbf{J}^{(ij)} = \int \mathbf{u}_\mu(\mathbf{y}_{ij}) \otimes \mathbf{u}_\mu(\mathbf{y}_{ij}) f_\theta^{1+\alpha}(\mathbf{y}) d\mathbf{y} = \int \mathbf{u}_\theta(\mathbf{y}_{ij}) \mathbf{u}_\theta(\mathbf{y}_{ij})^T f_\theta^{1+\alpha}(\mathbf{x}) d\mathbf{y}.$$

The matrix $\mathbf{J}^{(ij)}$ can be written as:

$$\mathbf{J}^{(ij)} = \begin{pmatrix} \mathbf{J}_{\mu_1\mu_1}^{(ij)} & \mathbf{J}_{\mu_1\mu_2}^{(ij)} & \cdots & \mathbf{J}_{\mu_1\mu_k}^{(ij)} & \mathbf{J}_{\mu_1\Sigma}^{(ij)} \\ \cdot & \mathbf{J}_{\mu_2\mu_2}^{(ij)} & \cdots & \mathbf{J}_{\mu_2\mu_k}^{(ij)} & \mathbf{J}_{\mu_2\Sigma}^{(ij)} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \cdot & \cdot & \cdots & \mathbf{J}_{\mu_k\mu_k}^{(ij)} & \mathbf{J}_{\mu_k\Sigma}^{(ij)} \\ \cdot & \cdot & \cdots & \cdot & \mathbf{J}_{\Sigma\Sigma}^{(ij)} \end{pmatrix},$$

where the individual blocks are:

$$\begin{aligned} \mathbf{J}_{\mu_i\mu_i}^{(ij)} &= (2\pi)^{-p\alpha/2} |\Sigma|^{-\alpha/2} (1+\alpha)^{(-2-p)/2} \Sigma^{-1}, \\ \mathbf{J}_{\mu_r\mu_s}^{(ij)} &= 0, \quad \text{for } r \neq s, \\ \mathbf{J}_{\Sigma\Sigma}^{(ij)} &= \frac{1}{4} \left\{ (2\pi)^{-p\alpha/2} |\Sigma|^{-\alpha/2} (1+\alpha)^{-p/2-2} \right. \\ &\quad \times \left(\alpha^2 \mathbf{I}_{pp} + \mathbf{K}_{pp} + \text{vec}(\Sigma^{-1}) \text{vec}(\Sigma)^T \right) \left(\Sigma^{-1} \otimes \Sigma^{-1} \right) \Big\}, \\ \mathbf{J}_{\Sigma\mu}^{(i)} &= \mathbf{0}. \end{aligned}$$

The $\mathbf{J}^{(ij)}$ matrix simplifies to

$$\mathbf{J}^{(ij)} = (2\pi)^{-p\alpha/2} |\Sigma|^{-\alpha/2} (1+\alpha)^{(-2-p)/2} \begin{bmatrix} \mathbf{D} & \mathbf{0}_{kp \times p^2} \\ \mathbf{0}_{p^2 \times kp}^T & \frac{(\mathbf{C}_\alpha)}{4(1+\alpha)} \Sigma^{-1} \otimes \Sigma^{-1} \end{bmatrix}, \quad (\text{A.39})$$

where \mathbf{D} is $k \times k$ block diagonal matrix. The i -th diagonal blocks are the $\Sigma_{p \times p}^{-1}$ matrix and the off-diagonal blocks are $p \times p$ zero matrices, with overall dimensions of $kp \times kp$ and $\mathbf{C}_\alpha = (\alpha^2 \mathbf{I}_{pp} + \mathbf{K}_{pp} + \text{vec}(\Sigma^{-1}) \text{vec}(\Sigma)^T)$

Therefore,

$$\begin{aligned} \mathbf{J} &= \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^k \sum_{j=1}^{n_i} \mathbf{J}^{(ij)} \\ &= (2\pi)^{-p\alpha/2} |\Sigma|^{-\alpha/2} (1 + \alpha)^{(-2-p)/2} \lim_{N \rightarrow \infty} \begin{bmatrix} \mathbf{S} & \mathbf{0}_{kp \times p^2} \\ \mathbf{0}_{p^2 \times kp}^T & \frac{(C_\alpha)}{4(1+\alpha)} \Sigma^{-1} \otimes \Sigma^{-1} \end{bmatrix}, \end{aligned} \quad (\text{A.40})$$

where \mathbf{S} is $\frac{n_i}{N} \mathbf{D}$ ($k \times k$ block diagonal matrix). The diagonal blocks are the $\Sigma_{p \times p}^{-1}$ matrix and the off-diagonal blocks are $p \times p$ zero matrices. Note that $\mathbf{S} = \mathbf{D}(\mathbf{n}/N) \otimes \Sigma_{p \times p}^{-1}$. Therefore, $\mathbf{S}^{-1} = \mathbf{D}(N/\mathbf{n}) \otimes \Sigma_{p \times p}$.

Similarly,

$$\begin{aligned} \mathbf{K} &= \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^k \sum_{j=1}^{n_i} \mathbf{J}^{(ij)} \\ &= (2\pi)^{-p\alpha} |\Sigma|^{-\alpha} (1 + 2\alpha)^{(-2-p)/2} \lim_{N \rightarrow \infty} \begin{bmatrix} \mathbf{S} & \mathbf{0}_{kp \times p^2} \\ \mathbf{0}_{p^2 \times kp}^T & \mathbf{V} \end{bmatrix}, \end{aligned} \quad (\text{A.41})$$

where

$$\mathbf{V} = \left\{ (4\alpha^2 \mathbf{I}_{pp} + \mathbf{K}_{pp} + \text{vec}(\Sigma^{-1}) \text{vec}(\Sigma)^T) - \frac{1}{4} \alpha^2 \right\} \Sigma^{-1} \otimes \Sigma^{-1}. \quad (\text{A.42})$$

A.8 Test Statistic

From Equation (A.40), we have

$$\mathbf{J}_N^{-1} = b \begin{bmatrix} \mathbf{S}^{-1} & \mathbf{0}_{kp \times p^2} \\ \mathbf{0}_{p^2 \times kp}^T & \mathbf{H} \end{bmatrix}, \quad (\text{A.43})$$

where $b = (2\pi)^{p\alpha/2} |\Sigma|^{\alpha/2} (1 + \alpha)^{(2+p)/2}$ and $\mathbf{H} = \left(\frac{(C_\alpha)}{4(1+\alpha)} \Sigma^{-1} \otimes \Sigma^{-1} \right)^{-1}$.

Also from Equation (8.2), we have

$$\mathbf{M}(\boldsymbol{\theta}) = \begin{bmatrix} \mathbf{M}_\mu \\ \mathbf{0}^T \end{bmatrix}. \quad (\text{A.44})$$

Thus,

$$\mathbf{M}^T \mathbf{J}_N^{-1} = b \begin{bmatrix} \mathbf{M}_\mu^T \mathbf{S}^{-1} & \mathbf{0} \end{bmatrix}. \quad (\text{A.45})$$

Also,

$$\mathbf{K}_N = c \begin{bmatrix} \mathbf{S} & \mathbf{0}_{kp \times p^2} \\ \mathbf{0}_{p^2 \times kp}^T & \mathbf{V} \end{bmatrix}, \quad (\text{A.46})$$

where $c = (2\pi)^{-p\alpha} |\Sigma|^{-\alpha} (1 + 2\alpha)^{(-2-p)/2}$.

Therefore,

$$\begin{aligned} \mathbf{M}^T \mathbf{J}_N^{-1} \mathbf{K}_N \mathbf{M} \mathbf{J}_N^{-1} &= b^2 c \begin{bmatrix} \mathbf{M}_\mu^T \mathbf{S}^{-1} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \mathbf{S} & \mathbf{0}_{kp \times p^2} \\ \mathbf{0}_{p^2 \times kp}^T & \mathbf{V} \end{bmatrix} \begin{bmatrix} \mathbf{M}_\mu^T \mathbf{S}^{-1} & \mathbf{0} \end{bmatrix}^T \\ &= b^2 c \begin{bmatrix} \mathbf{M}_\mu^T & \mathbf{0} \end{bmatrix} \begin{bmatrix} \mathbf{M}_\mu^T \mathbf{S}^{-1} & \mathbf{0} \end{bmatrix}^T \\ &= b^2 c \mathbf{M}_\mu^T \mathbf{S}^{-1} \mathbf{M}_\mu, \end{aligned} \quad (\text{A.47})$$

where $b^2 c = (1 + \alpha)^{2+p} (1 + 2\alpha)^{\frac{-2-p}{2}}$.

Therefore, we have

$$\begin{aligned} \mathbf{M}^T \mathbf{J}_N^{-1} \mathbf{K}_N \mathbf{M} \mathbf{J}^{-1} &= (1 + \alpha)^{2+p} (1 + 2\alpha)^{\frac{-2-p}{2}} \mathbf{M}_\mu^T \mathbf{S}^{-1} \mathbf{M}_\mu \\ &= \left(\frac{(1 + \alpha)^2}{(1 + 2\alpha)} \right)^{(p+2)/2} \mathbf{M}_\mu^T \mathbf{S}^{-1} \mathbf{M}_\mu. \end{aligned} \quad (\text{A.48})$$

A.8.1 Covariance Matrix of $\hat{\boldsymbol{\mu}}$

The asymptotic scatter matrix for the estimators is given by the following equations:

$$\mathbf{J}_{\mu\mu}^{-1} \mathbf{K}_{\mu\mu} \mathbf{J}_{\mu\mu}^{-1} = \left(1 + \frac{\alpha^2}{1 + 2\alpha} \right)^{p/2+1} \lim_{N \rightarrow \infty} \mathbf{S}^{-1}. \quad (\text{A.49})$$

For the scatter matrix involving the covariance estimator, we have:

$$\begin{aligned}
& \mathbf{J}_{\Sigma\Sigma}^{-1} \mathbf{K}_{\Sigma\Sigma} \mathbf{J}_{\Sigma\Sigma}^{-1} \\
&= \frac{4(1+\alpha)^{p+4}}{(1+2\alpha)^{(p+4)/2}} \left\{ \left(\alpha^2 \mathbf{I}_{pp} + \mathbf{K}_{pp} + \text{vec}(\Sigma^{-1}) \text{vec}(\Sigma)^T \right)^{-1} \right. \\
&\quad \times \left(4\alpha^2 \mathbf{I}_{pp} + \mathbf{K}_{pp} + \text{vec}(\Sigma^{-1}) \text{vec}(\Sigma)^T \right) \\
&\quad \left. \times \left(\alpha^2 \mathbf{I}_{pp} + \mathbf{K}_{pp} + \text{vec}(\Sigma^{-1}) \text{vec}(\Sigma)^T \right)^{-1} - \alpha^2 \right\} (\Sigma \otimes \Sigma).
\end{aligned} \tag{A.50}$$

In the first equation, the term $\left(1 + \frac{\alpha^2}{1+2\alpha}\right)^{p/2+1}$ modifies the asymptotic scatter matrix of the mean estimator, represented by the inverse of the scatter matrix \mathbf{S}^{-1} as the sample size N approaches infinity.

In the second equation, the expression involves the Kronecker product and the vectorization of the covariance matrix. The factors within the curly braces account for the dependency on α , the dimension p , and the interaction terms involving Σ^{-1} .

A.9 Regularity Conditions

To ensure the consistency and asymptotic properties of the Minimum Density Power Divergence (MDPD) estimator in a multivariate framework, we require the following assumptions::

- (A1) The true density $g(\mathbf{Y}_{ij})$ is supported over the entire real space \mathbf{R}^p .
- (A2) There is an open subset $\omega \subset \Theta$ containing the best fitting parameter θ such that \mathbf{J} is positive definite for all $\theta \in \omega$.
- (A3) Suppose

$$V_{\theta}(\mathbf{Y}_{ij}) = \exp \left\{ -\frac{\alpha}{2} (\mathbf{Y}_{ij} - \mu_i)^T \Sigma^{-1} (\mathbf{Y}_{ij} - \mu_i) \right\}.$$

There exist functions $M_{rst}(\mathbf{Y}_{ij})$ such that

$$\left| \frac{\partial^3 V_{\theta}(\mathbf{Y}_{ij})}{\partial \theta_r \partial \theta_s \partial \theta_t} \right| \leq M_{rst}(\mathbf{Y}_{ij}),$$

for all $\boldsymbol{\theta} \in \boldsymbol{\omega}$, where

$$\mathbb{E}_g(|M_{rst}(\mathbf{Y}_{ij})|) = \int_{\mathbf{y}_{ij}} |M_{rst}(\mathbf{y}_{ij})| g(\mathbf{y}_{ij}) d\mathbf{y}_{ij} < \infty,$$

for all r, s, t .

(A4) We denote $\boldsymbol{\delta}(\cdot)$ as the indicator function. Then, for all r and s , we have

$$\begin{aligned} & \lim_{\lambda \rightarrow \infty} \sup_{N > 1} \left\{ \frac{1}{N} \sum_{i=1}^k \sum_{j=1}^{n_i} \mathbf{E}_g \left[\left| \frac{\partial V_{\boldsymbol{\theta}}(\mathbf{Y}_{ij})}{\partial \boldsymbol{\theta}_r} \right| \boldsymbol{\delta} \left(\left| \frac{\partial V_{\boldsymbol{\theta}}(\mathbf{Y}_{ij})}{\partial \boldsymbol{\theta}_r} \right| > \lambda \right) \right] \right\} = \mathbf{0}, \\ & \lim_{\lambda \rightarrow \infty} \sup_{N > 1} \left\{ \frac{1}{N} \sum_{i=1}^k \sum_{j=1}^{n_i} \mathbf{E}_g \left[\left| \frac{\partial^2 V_{\boldsymbol{\theta}}(\mathbf{Y}_{ij})}{\partial \boldsymbol{\theta}_r \partial \boldsymbol{\theta}_s} - \mathbb{E}_g \left(\frac{\partial^2 V_{\boldsymbol{\theta}}(\mathbf{Y}_{ij})}{\partial \boldsymbol{\theta}_r \partial \boldsymbol{\theta}_s} \right) \right| \right. \right. \\ & \quad \left. \left. \times \boldsymbol{\delta} \left(\left| \frac{\partial^2 V_{\boldsymbol{\theta}}(\mathbf{Y}_{ij})}{\partial \boldsymbol{\theta}_r \partial \boldsymbol{\theta}_s} - \mathbf{E}_g \left(\frac{\partial^2 V_{\boldsymbol{\theta}}(\mathbf{Y}_{ij})}{\partial \boldsymbol{\theta}_r \partial \boldsymbol{\theta}_s} \right) \right| > \lambda \right) \right] \right\} = \mathbf{0}. \end{aligned}$$

(A5) Let

$$K_N = \frac{1}{N} \sum_{i=1}^k \sum_{j=1}^{n_i} K(\mathbf{Y}_{ij}). \quad (\text{A.51})$$

For all $\epsilon > 0$, we have

$$\left\{ \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^k \sum_{j=1}^{n_i} \mathbf{E}_g \left[\left\| K_N^{-1/2} \frac{\partial V_{\boldsymbol{\theta}}(\mathbf{Y}_{ij})}{\partial \boldsymbol{\theta}} \right\|^2 \boldsymbol{\delta} \left(\left\| K_N^{-1/2} \frac{\partial V_{\boldsymbol{\theta}}(\mathbf{Y}_{ij})}{\partial \boldsymbol{\theta}} \right\| > \epsilon \sqrt{N} \right) \right] \right\} = \mathbf{0}.$$

In the context of an independent heterogeneous setup, these conditions are crucial for stabilizing the matrices \mathbf{J} and \mathbf{K} which is necessary for the asymptotic distribution to exist.

A.10 Derivation of Breakdown Point

In this section, we derive the breakdown point integrals. The breakdown point is a measure of an estimator's robustness, indicating the proportion of incorrect observations the estimator can handle before giving incorrect (arbitrarily large) results. Here, we specifically

focus on the calculation of the integral $\int \phi_{\mathbf{c}, \mathbf{D}}(\mathbf{x}) \phi_{\mathbf{m}, \mathbf{S}}^\alpha(\mathbf{x}) d\mathbf{x}$, which plays a crucial role in this derivation. Now, we will calculate $\int \phi_{\mathbf{c}, \mathbf{D}}(\mathbf{x}) \phi_{\mathbf{m}, \mathbf{S}}^\alpha(\mathbf{x}) d\mathbf{x}$.

To calculate the integral $\int \phi_{\mathbf{c}, \mathbf{D}}(\mathbf{x}) \phi_{\mathbf{m}, \mathbf{S}}^\alpha(\mathbf{x}) d\mathbf{x}$, we start with the definitions of $\phi_{\mathbf{c}, \mathbf{D}}(\mathbf{x})$ and $\phi_{\mathbf{m}, \mathbf{S}}^\alpha(\mathbf{x})$. These are multivariate normal density functions with mean vectors \mathbf{c} and \mathbf{m} , and covariance matrices \mathbf{D} and \mathbf{S} , respectively.

$$\begin{aligned}
& \int \phi_{\mathbf{c}, \mathbf{D}}(\mathbf{x}) \phi_{\mathbf{m}, \mathbf{S}}^\alpha(\mathbf{x}) d\mathbf{x} \\
&= \frac{1}{(2\pi)^{p/2} |\mathbf{D}|^{1/2}} \cdot \frac{1}{(2\pi)^{p\alpha/2} |\mathbf{S}|^{\alpha/2}} \\
&\times \int \exp\left(-\frac{1}{2}(\mathbf{x} - \mathbf{c})^T \mathbf{D}^{-1}(\mathbf{x} - \mathbf{c})\right) \cdot \exp\left(-\frac{\alpha}{2}(\mathbf{x} - \mathbf{m})^T \mathbf{S}^{-1}(\mathbf{x} - \mathbf{m})\right) d\mathbf{x} \\
&= \frac{1}{(2\pi)^{p/2} |\mathbf{D}|^{1/2}} \cdot \frac{1}{(2\pi)^{p\alpha/2} |\mathbf{S}|^{\alpha/2}} \\
&\times \int \exp\left(-\frac{1}{2} \mathbf{x}' (\mathbf{D}^{-1} + \alpha \mathbf{S}^{-1}) \mathbf{x} + \mathbf{x}' (\mathbf{D}^{-1} \mathbf{c} + \alpha \mathbf{S}^{-1} \mathbf{m}) - \frac{1}{2} \mathbf{c}' \mathbf{D}^{-1} \mathbf{c} - \frac{\alpha}{2} \mathbf{m}' \mathbf{S}^{-1} \mathbf{m}\right) d\mathbf{x}.
\end{aligned} \tag{A.52}$$

To simplify the integral, we introduce the following variables:

$$\begin{aligned}
\mathbf{A} &= \mathbf{D}^{-1} + \alpha \mathbf{S}^{-1}, \\
\mathbf{B} &= \mathbf{D}^{-1} \mathbf{c} + \alpha \mathbf{S}^{-1} \mathbf{m}, \\
\mathbf{K} &= -\frac{1}{2} \mathbf{c}' \mathbf{D}^{-1} \mathbf{c} - \frac{\alpha}{2} \mathbf{m}' \mathbf{S}^{-1} \mathbf{m}.
\end{aligned}$$

Then, the integral becomes:

$$\begin{aligned}
\int \phi_{\mathbf{c}, \mathbf{D}}(\mathbf{x}) \phi_{\mathbf{m}, \mathbf{S}}^\alpha(\mathbf{x}) d\mathbf{x} &= \frac{1}{(2\pi)^{p/2} |\mathbf{D}|^{1/2}} \cdot \frac{1}{(2\pi)^{p\alpha/2} |\mathbf{S}|^{\alpha/2}} \cdot \exp(\mathbf{K}) \cdot \int \exp\left(-\frac{1}{2} \mathbf{x}' \mathbf{A} \mathbf{x} + \mathbf{x}' \mathbf{B}\right) d\mathbf{x} \\
&= \frac{1}{(2\pi)^{p/2} |\mathbf{D}|^{1/2}} \cdot \frac{1}{(2\pi)^{p\alpha/2} |\mathbf{S}|^{\alpha/2}} \exp(\mathbf{K}) \\
&\times \int \exp\left(-\frac{1}{2}(\mathbf{x} - \mathbf{A}^{-1} \mathbf{B})^T \mathbf{A}(\mathbf{x} - \mathbf{A}^{-1} \mathbf{B})\right) \exp\left(\frac{1}{2} \mathbf{B}' \mathbf{A}^{-1} \mathbf{B}\right) d\mathbf{x}
\end{aligned} \tag{A.53}$$

We can simplify the integral further by substituting $\mathbf{U} = \mathbf{x} - \mathbf{A}^{-1} \mathbf{B}$ and evaluate the integral using completing the square:

$$\begin{aligned}
& \int \phi_{\mathbf{c}, \mathbf{D}}(\mathbf{x}) \phi_{\mathbf{m}, \mathbf{S}}^\alpha(\mathbf{x}) d\mathbf{x} \\
&= \frac{1}{(2\pi)^{\alpha p/2} |\mathbf{D}|^{1/2} |\mathbf{D}^{-1} + \alpha \mathbf{S}^{-1}|^{1/2} |\mathbf{S}|^{\alpha/2}} \\
&\times \exp\left(-\frac{1}{2} \mathbf{c}' \mathbf{D}^{-1} \mathbf{c} - \frac{\alpha}{2} \mathbf{m}' \mathbf{S}^{-1} \mathbf{m} + \frac{1}{2} (\mathbf{D}^{-1} \mathbf{c} + \alpha \mathbf{S}^{-1} \mathbf{m})^T (\mathbf{D}^{-1} + \alpha \mathbf{S}^{-1})^{-1} (\mathbf{D}^{-1} \mathbf{c} + \alpha \mathbf{S}^{-1} \mathbf{m})\right).
\end{aligned}$$

Appendix B

Table B.1: MSE and Kullback–Leibler divergence of the mean vector and scatter matrices for simulation scenarios: $p = 2$, $ncp = 100$, $\delta = 0.001$

bdp	n	eps	e_μ			e_Σ			e_{KL}		
			MDPD	FastM	DetM	MDPD	FastM	DetM	MDPD	FastM	DetM
0.25	50	0.0833	0.0731	0.0756	0.0747	0.1737	0.3529	0.297	0.121	0.1571	0.1409
		0.1250	0.0765	0.0722	0.0733	0.1996	0.4293	0.3744	0.1281	0.1705	0.1569
		0.2500	0.08	0.0521	0.0504	0.3564	1.1817	1.1846	0.1608	0.343	0.3432
	100	0.0833	0.0389	0.0402	0.0387	0.0911	0.188	0.1688	0.0604	0.0886	0.0817
		0.1250	0.0366	0.0361	0.0375	0.1063	0.25	0.2341	0.0634	0.1032	0.0995
		0.2500	0.0432	0.0275	0.0269	0.2483	1.0185	0.9852	0.1069	0.307	0.2985
	200	0.0833	0.0186	0.0210	0.0218	0.0477	0.1081	0.0975	0.0302	0.052	0.0487
		0.1250	0.0180	0.0193	0.0188	0.0651	0.1613	0.1546	0.0358	0.0685	0.066
		0.2500	0.0191	0.0136	0.0128	0.196	0.8894	0.8798	0.0804	0.2768	0.2737
	500	0.0833	0.0075	0.0091	0.0089	0.0237	0.0587	0.0548	0.0142	0.0284	0.0269
		0.1250	0.0078	0.0077	0.0078	0.0399	0.1008	0.1006	0.0208	0.0434	0.0435
		0.2500	0.0084	0.0054	0.0055	0.1542	0.7974	0.8011	0.0632	0.2542	0.2551
0.1875	50	0.0625	0.0496	0.063	0.0629	0.08	0.2465	0.2075	0.0657	0.1227	0.1085
		0.0938	0.0521	0.067	0.0656	0.0818	0.2541	0.2455	0.0667	0.1234	0.1206
		0.1875	0.0548	0.0481	0.0481	0.1041	0.6717	0.6666	0.0744	0.2226	0.2212
	100	0.0625	0.0231	0.0353	0.0327	0.0429	0.1236	0.1146	0.0322	0.066	0.0614
		0.0938	0.0249	0.0312	0.0312	0.0426	0.1564	0.1523	0.0329	0.0731	0.0711
		0.1875	0.0283	0.0244	0.0243	0.0614	0.5001	0.5053	0.0408	0.1741	0.1749
	200	0.0625	0.0115	0.0181	0.0166	0.0218	0.0672	0.0653	0.0163	0.0364	0.0349
		0.0938	0.0124	0.0162	0.0165	0.0241	0.0938	0.0944	0.0174	0.0446	0.0449
		0.1875	0.0135	0.0116	0.0125	0.0334	0.4707	0.4527	0.0213	0.1635	0.1588
	500	0.0625	0.0051	0.0072	0.0063	0.009	0.038	0.0371	0.0068	0.0196	0.0189
		0.0938	0.0051	0.0062	0.0061	0.0108	0.064	0.0632	0.0076	0.0293	0.029
		0.1875	0.0054	0.0046	0.005	0.0198	0.4204	0.4181	0.0114	0.1486	0.148
0.125	50	0.0417	0.0411	0.0576	0.0563	0.0664	0.1552	0.142	0.0545	0.0907	0.0865
		0.0625	0.0428	0.0528	0.0513	0.0694	0.1806	0.1591	0.057	0.0939	0.0859
		0.1250	0.0513	0.046	0.0459	0.4735	0.3375	0.345	0.1103	0.1335	0.1359
	100	0.0417	0.0208	0.0283	0.0292	0.0341	0.0789	0.0732	0.0275	0.0472	0.0456
		0.0625	0.0232	0.0275	0.0277	0.0355	0.097	0.0996	0.0294	0.0519	0.0527
		0.1250	0.0246	0.0225	0.0235	0.1148	0.2474	0.2444	0.0405	0.0983	0.0982
	200	0.0417	0.0112	0.0148	0.0159	0.0166	0.0458	0.0415	0.0138	0.027	0.0257
		0.0625	0.0114	0.0143	0.014	0.0177	0.0592	0.0584	0.0145	0.0313	0.0307
		0.1250	0.0142	0.0118	0.011	0.1146	0.2202	0.2172	0.0272	0.0871	0.0856
	500	0.0417	0.0046	0.0058	0.0058	0.0068	0.0222	0.0229	0.0057	0.0126	0.0129
		0.0625	0.0044	0.0054	0.0054	0.0075	0.0368	0.0376	0.0059	0.0184	0.0187
		0.1250	0.0047	0.0047	0.0046	0.0086	0.1933	0.1925	0.0064	0.0762	0.0759

Table B.2: MSE and Kullback–Leibler divergence of the MDPD, FASTMCD, and DetMCD for the mean vector and scatter matrices for simulation scenarios: $p = 2$, $ncp = 100$, $\delta = 1$

bdp	n	eps	e_μ			e_Σ			e_{KL}		
			MDPD	FastM	DetM	MDPD	FastM	DetM	MDPD	FastM	DetM
0.25	50	0.0833	0.0777	0.0773	0.074	0.1676	0.3625	0.2867	0.1214	0.16	0.1378
		0.1250	0.0769	0.0713	0.0688	0.1965	0.4167	0.3637	0.1246	0.1675	0.1526
		0.2500	0.086	0.0538	0.054	0.3408	1.1667	1.1909	0.1598	0.3395	0.3467
	100	0.0833	0.037	0.0417	0.0388	0.0844	0.1848	0.1629	0.0567	0.0893	0.08
		0.1250	0.0375	0.0386	0.0385	0.1046	0.2402	0.2259	0.0624	0.1013	0.0973
		0.2500	0.039	0.0264	0.0256	0.2519	1.0057	1.0358	0.1059	0.3038	0.3108
	200	0.0833	0.0183	0.0213	0.0212	0.0434	0.1115	0.1019	0.0286	0.0534	0.05
		0.1250	0.0177	0.0194	0.0185	0.0624	0.161	0.1543	0.0345	0.0685	0.0657
		0.2500	0.0192	0.0134	0.0137	0.192	0.8836	0.878	0.079	0.2752	0.2738
	500	0.0833	0.0072	0.0086	0.0085	0.023	0.0573	0.0577	0.0137	0.0277	0.0278
		0.1250	0.0077	0.0078	0.008	0.0383	0.0999	0.1022	0.0201	0.0431	0.0441
		0.2500	0.008	0.0057	0.0054	0.1553	0.7986	0.7955	0.0635	0.2544	0.2536
0.1875	50	0.0625	0.0487	0.0698	0.0659	0.0821	0.2448	0.222	0.0649	0.1252	0.1134
		0.0938	0.0516	0.0613	0.0628	0.0849	0.2558	0.2447	0.0684	0.1214	0.1174
		0.1875	0.0547	0.0496	0.0498	0.1022	0.6487	0.6804	0.0737	0.2174	0.2249
	100	0.0625	0.0237	0.0363	0.0323	0.0425	0.1251	0.1129	0.0327	0.0669	0.0605
		0.0938	0.0241	0.033	0.0318	0.0446	0.1536	0.1516	0.0333	0.073	0.0712
		0.1875	0.027	0.024	0.0237	0.0578	0.5096	0.515	0.0387	0.1759	0.1772
	200	0.0625	0.0131	0.0172	0.0178	0.0225	0.0685	0.063	0.0172	0.0364	0.0346
		0.0938	0.0129	0.0165	0.017	0.0232	0.0958	0.0931	0.0172	0.0453	0.0447
		0.1875	0.014	0.0121	0.0126	0.0348	0.4572	0.456	0.0222	0.16	0.1597
	500	0.0625	0.0046	0.0067	0.0071	0.009	0.0382	0.0374	0.0065	0.0195	0.0193
		0.0938	0.0051	0.0065	0.0065	0.01	0.0643	0.0633	0.0072	0.0295	0.0292
		0.1875	0.0057	0.005	0.0048	0.0198	0.4249	0.4269	0.0116	0.1502	0.1507
0.125	50	0.0417	0.0448	0.0561	0.054	0.0687	0.1546	0.1473	0.057	0.0905	0.0864
		0.0625	0.042	0.0548	0.0529	0.0702	0.1681	0.1737	0.0563	0.0915	0.0919
		0.1250	0.0498	0.0457	0.0431	0.1742	0.3321	0.3299	0.0743	0.1319	0.1301
	100	0.0417	0.0211	0.0294	0.0281	0.0341	0.0786	0.0798	0.0277	0.0471	0.0474
		0.0625	0.0218	0.0282	0.0276	0.0364	0.1001	0.0889	0.0289	0.0533	0.049
		0.1250	0.0253	0.0223	0.023	0.0415	0.2567	0.2524	0.0324	0.1011	0.1003
	200	0.0417	0.0109	0.0151	0.0141	0.0174	0.044	0.0431	0.0141	0.0263	0.0255
		0.0625	0.0111	0.0136	0.0144	0.0189	0.0591	0.0568	0.0148	0.031	0.0304
		0.1250	0.0119	0.0118	0.0111	0.0226	0.2187	0.218	0.0165	0.0865	0.086
	500	0.0417	0.0044	0.0061	0.0059	0.0067	0.0221	0.0227	0.0055	0.0128	0.0129
		0.0625	0.0044	0.0055	0.0057	0.007	0.0378	0.0354	0.0056	0.0188	0.0179
		0.1250	0.0049	0.0044	0.0045	0.0091	0.1923	0.1883	0.0068	0.0757	0.0744

Table B.3: MSE and Kullback–Leibler divergence of the MDPD, FASTMCD, and DetMCD for the mean vector and scatter matrices for simulation scenarios: $p = 5$, $ncp = 100$, $\delta = 0.001$

bdp	n	eps	e_μ			e_Σ			e_{KL}		
			MDPD	FastM	DetM	MDPD	FastM	DetM	MDPD	FastM	DetM
0.1232	50	0.0411	0.1529	0.128	0.127	0.2049	0.2673	0.2558	0.3647	0.3415	0.3248
		0.0616	0.1501	0.1228	0.1201	0.2141	0.2804	0.2667	0.3671	0.3361	0.3204
		0.1232	0.158	0.1149	0.1143	0.2389	0.33	0.3359	0.3895	0.336	0.3404
	100	0.0411	0.0709	0.0636	0.0625	0.0973	0.1263	0.1196	0.1599	0.1681	0.1604
		0.0616	0.0725	0.0622	0.0622	0.1033	0.1301	0.1247	0.1653	0.1666	0.1606
		0.1232	0.0762	0.0587	0.0542	0.1159	0.1728	0.1733	0.1768	0.1863	0.1845
	200	0.0411	0.0353	0.0321	0.0326	0.0476	0.0612	0.0583	0.0771	0.0844	0.0811
		0.0616	0.036	0.0321	0.0318	0.0492	0.0665	0.065	0.0783	0.0869	0.0852
		0.1232	0.0377	0.0284	0.028	0.0561	0.1078	0.1071	0.0837	0.1175	0.1167
	500	0.0411	0.0135	0.0133	0.013	0.0189	0.0248	0.025	0.0299	0.0349	0.0349
		0.0616	0.0146	0.0129	0.0128	0.0199	0.0296	0.029	0.0313	0.0389	0.0382
		0.1232	0.0153	0.0114	0.0115	0.025	0.0697	0.0683	0.0361	0.0765	0.0754
0.0924	50	0.0308	0.1079	0.1188	0.117	0.1347	0.2076	0.2026	0.2337	0.2915	0.2837
		0.0462	0.1092	0.1169	0.1193	0.1372	0.2157	0.2053	0.2347	0.2861	0.2782
		0.0924	0.1135	0.1124	0.1103	0.1509	0.2273	0.2325	0.2478	0.2709	0.2738
	100	0.0308	0.0535	0.0589	0.0615	0.0685	0.1067	0.1036	0.1147	0.1488	0.1468
		0.0462	0.0548	0.0593	0.059	0.0679	0.1063	0.1038	0.1138	0.1461	0.1436
		0.0924	0.0594	0.0545	0.0545	0.142	0.1346	0.1321	0.1446	0.1555	0.1536
	200	0.0308	0.0266	0.0313	0.0308	0.0343	0.0534	0.0514	0.0563	0.0764	0.0739
		0.0462	0.0289	0.0303	0.0294	0.0343	0.0547	0.0531	0.0574	0.0751	0.0733
		0.0924	0.0301	0.0278	0.0281	0.0419	0.0767	0.0769	0.0617	0.0902	0.0907
	500	0.0308	0.0113	0.0123	0.0122	0.0134	0.0215	0.0211	0.0224	0.0311	0.0305
		0.0462	0.0112	0.012	0.0118	0.014	0.0244	0.0243	0.023	0.0332	0.0331
		0.0924	0.0117	0.011	0.0111	0.0155	0.0473	0.048	0.0247	0.0549	0.0556
0.0616	50	0.0205	0.1038	0.1122	0.1098	0.1263	0.1878	0.1846	0.2146	0.268	0.2631
		0.0308	0.1035	0.1193	0.1161	0.1283	0.1824	0.1848	0.219	0.2679	0.2685
		0.0616	0.1598	0.1067	0.1046	1.3863	0.2001	0.1996	0.7758	0.2517	0.2502
	100	0.0205	0.0517	0.0579	0.0567	0.0659	0.0927	0.0886	0.1095	0.1353	0.1305
		0.0308	0.0518	0.0573	0.0551	0.0709	0.094	0.0886	0.1119	0.1346	0.1273
		0.0616	0.1215	0.0535	0.0531	1.5361	0.099	0.1015	0.7834	0.1285	0.1299
	200	0.0205	0.026	0.0285	0.029	0.0334	0.0453	0.0441	0.0541	0.0672	0.066
		0.0308	0.0248	0.0277	0.0284	0.0367	0.0454	0.0455	0.0565	0.0655	0.0658
		0.0616	0.1016	0.0269	0.0271	1.7258	0.0556	0.0561	0.8425	0.0714	0.0719
	500	0.0205	0.0104	0.012	0.0114	0.0137	0.0181	0.0175	0.022	0.0272	0.0264
		0.0308	0.0111	0.0111	0.0112	0.017	0.0195	0.0192	0.0256	0.0279	0.0276
		0.0616	0.0938	0.0108	0.0108	1.8126	0.0304	0.0304	0.8636	0.0381	0.038

Table B.4: MSE and Kullback–Leibler divergence of the MDPD, FASTMCD, and DetMCD for the mean vector and scatter matrices for simulation scenarios: $p = 5$, $ncp = 100$, $\delta = 1$

bdp	n	eps	e_μ			e_Σ			e_{KL}		
			MDPD	FastM	DetM	MDPD	FastM	DetM	MDPD	FastM	DetM
0.1232	50	0.0411	0.1469	0.1305	0.1271	0.2049	0.2708	0.2547	0.3597	0.345	0.323
		0.0616	0.153	0.1245	0.1221	0.2116	0.2815	0.268	0.3684	0.3389	0.3221
		0.1232	0.1575	0.1136	0.1131	0.2403	0.3258	0.3264	0.387	0.3317	0.3337
	100	0.0411	0.072	0.0644	0.0632	0.0976	0.1265	0.1161	0.161	0.1683	0.1576
		0.0616	0.0713	0.0637	0.0641	0.0999	0.131	0.1252	0.1604	0.1681	0.1622
		0.1232	0.0773	0.0577	0.0546	0.1132	0.1739	0.173	0.1729	0.1866	0.1846
	200	0.0411	0.0357	0.0327	0.0321	0.0486	0.0621	0.0594	0.0783	0.0852	0.0818
		0.0616	0.036	0.0323	0.0308	0.0507	0.0657	0.0634	0.0799	0.0863	0.083
		0.1232	0.0384	0.0287	0.0285	0.0575	0.1063	0.1056	0.0856	0.1164	0.1157
	500	0.0411	0.014	0.0129	0.0128	0.0186	0.0253	0.0253	0.0298	0.0352	0.035
		0.0616	0.014	0.0122	0.0131	0.0202	0.0295	0.0293	0.0311	0.0383	0.0386
		0.1232	0.0152	0.0114	0.011	0.0249	0.0687	0.0689	0.036	0.0757	0.0756
0.0924	50	0.0308	0.1113	0.1159	0.1208	0.1359	0.2142	0.2003	0.2347	0.2963	0.2836
		0.0462	0.112	0.1202	0.1194	0.1342	0.2133	0.2078	0.2341	0.286	0.2787
		0.0924	0.1142	0.1107	0.1077	0.1489	0.2389	0.2403	0.2483	0.2781	0.2772
	100	0.0308	0.0554	0.0592	0.06	0.0686	0.1071	0.1037	0.1146	0.1483	0.1453
		0.0462	0.0559	0.0598	0.058	0.0707	0.1074	0.1049	0.1176	0.1466	0.1432
		0.0924	0.0569	0.0547	0.0553	0.0747	0.137	0.1343	0.1211	0.1575	0.1551
	200	0.0308	0.0272	0.0305	0.0291	0.0338	0.051	0.051	0.0558	0.074	0.0726
		0.0462	0.0286	0.03	0.0303	0.035	0.0538	0.0536	0.0581	0.0744	0.0741
		0.0924	0.0295	0.0283	0.0274	0.0368	0.077	0.0764	0.0598	0.0908	0.0898
	500	0.0308	0.011	0.0124	0.0118	0.0136	0.0213	0.0209	0.0224	0.0309	0.0301
		0.0462	0.0113	0.0122	0.0122	0.0141	0.0239	0.0237	0.023	0.0329	0.0326
		0.0924	0.0115	0.0107	0.0108	0.0159	0.0475	0.0478	0.0249	0.0549	0.0552
0.0616	50	0.0205	0.1045	0.1112	0.1114	0.1292	0.1871	0.182	0.2206	0.2648	0.2618
		0.0308	0.1061	0.1105	0.1108	0.1291	0.1854	0.186	0.2214	0.2685	0.2655
		0.0616	0.126	0.1046	0.1053	0.4948	0.2006	0.1975	0.4048	0.2501	0.2493
	100	0.0205	0.0508	0.0576	0.0588	0.0664	0.0888	0.089	0.1083	0.1323	0.1312
		0.0308	0.0555	0.0561	0.0559	0.0692	0.0913	0.0909	0.1122	0.1304	0.1298
		0.0616	0.0673	0.053	0.0537	0.3503	0.1018	0.0999	0.268	0.13	0.1291
	200	0.0205	0.0262	0.0291	0.0291	0.0339	0.045	0.0445	0.0549	0.0673	0.0668
		0.0308	0.0267	0.0283	0.0283	0.0362	0.0455	0.0449	0.0571	0.0661	0.0651
		0.0616	0.036	0.0269	0.0274	0.2561	0.0566	0.0572	0.1919	0.0721	0.0733
	500	0.0205	0.0104	0.0118	0.0115	0.0139	0.0182	0.0177	0.0221	0.0274	0.0267
		0.0308	0.0108	0.011	0.0108	0.0161	0.0199	0.0194	0.0245	0.0283	0.0276
		0.0616	0.019	0.0106	0.0104	0.199	0.0307	0.0306	0.1492	0.0383	0.0381

Table B.5: MSE and Kullback–Leibler divergence of the MDPD, FASTMCD, and DetMCD for the mean vector and scatter matrices for simulation scenarios: $p = 10$, $ncp = 250$, $\delta = 0.001$

bdp	n	eps	e_μ			e_Σ			e_{KL}		
			MDPD	FastM	DetM	MDPD	FastM	DetM	MDPD	FastM	DetM
0.067	50	0.0223	0.2664	0.2155	0.2196	0.3042	0.3326	0.3329	1.0252	0.8685	0.8634
		0.0335	0.2622	0.2215	0.2167	0.3028	0.3338	0.3285	1.0178	0.8771	0.8542
		0.0670	0.2712	0.2121	0.2123	0.318	0.3423	0.3405	1.0558	0.8104	0.8076
	100	0.0223	0.1216	0.1138	0.1096	0.1432	0.1547	0.1509	0.439	0.4174	0.4069
		0.0335	0.1252	0.1084	0.1106	0.1434	0.1546	0.154	0.4426	0.4085	0.4042
		0.0670	0.1268	0.1058	0.1065	0.1494	0.1594	0.1558	0.4549	0.3894	0.3862
	200	0.0223	0.0629	0.0574	0.0541	0.0695	0.0755	0.0748	0.2096	0.208	0.2034
		0.0335	0.062	0.0565	0.0552	0.0705	0.0754	0.0746	0.2098	0.2035	0.201
		0.0670	0.0651	0.0539	0.0547	0.0744	0.0813	0.082	0.2197	0.2006	0.2025
	500	0.0223	0.0244	0.0232	0.0232	0.0277	0.0298	0.0293	0.0816	0.0825	0.0813
		0.0335	0.0251	0.0228	0.0228	0.028	0.0308	0.03	0.0824	0.0833	0.0816
		0.0670	0.0253	0.0209	0.0214	0.0296	0.0356	0.0359	0.0857	0.0881	0.0889
0.0502	50	0.0167	0.2077	0.2174	0.2132	0.2327	0.2948	0.2914	0.7458	0.8065	0.7985
		0.0251	0.2157	0.213	0.208	0.2335	0.2942	0.2863	0.7472	0.7973	0.7858
		0.0502	0.2161	0.2075	0.2113	0.2414	0.2918	0.2894	0.7647	0.7625	0.7566
	100	0.01670	0.1042	0.1114	0.1069	0.1156	0.1444	0.1427	0.3514	0.4043	0.3956
		0.0251	0.1037	0.1117	0.1076	0.1167	0.1461	0.1413	0.3575	0.3996	0.3887
		0.0502	0.1093	0.1041	0.104	0.121	0.1497	0.1489	0.37	0.3782	0.3768
	200	0.01670	0.0533	0.0559	0.0557	0.0583	0.0708	0.0697	0.1752	0.1979	0.1957
		0.0251	0.0528	0.0551	0.0536	0.059	0.0714	0.07	0.1763	0.1956	0.192
		0.0502	0.0542	0.0525	0.0519	0.061	0.0738	0.0735	0.1814	0.1895	0.1882
	500	0.0167	0.0216	0.0218	0.0223	0.0233	0.028	0.0276	0.0696	0.0784	0.0777
		0.0251	0.0213	0.0218	0.0219	0.0236	0.0285	0.028	0.0701	0.0784	0.0774
		0.0502	0.0218	0.0208	0.0212	0.0243	0.0312	0.0312	0.0717	0.0801	0.0802
0.0335	50	0.0112	0.2092	0.2059	0.2079	0.2277	0.2592	0.2571	0.7334	0.7301	0.7292
		0.0167	0.2025	0.2006	0.2028	0.2301	0.2558	0.2546	0.7305	0.724	0.7271
		0.0335	0.1992	0.2015	0.2059	0.2263	0.2593	0.2594	0.7228	0.7311	0.7302
	100	0.0112	0.1012	0.1059	0.1051	0.1131	0.1322	0.1312	0.3465	0.3781	0.3746
		0.0167	0.0992	0.1075	0.106	0.1142	0.1312	0.1287	0.3476	0.3767	0.3689
		0.0335	0.1057	0.1047	0.1032	0.1182	0.132	0.1333	0.3584	0.3564	0.3607
	200	0.0112	0.0508	0.0529	0.0519	0.056	0.0652	0.0635	0.1684	0.186	0.1815
		0.0167	0.0516	0.0516	0.0525	0.0559	0.0648	0.0642	0.1682	0.182	0.1813
		0.0335	0.0518	0.0526	0.0509	0.0598	0.065	0.0664	0.1755	0.177	0.179
	500	0.0112	0.0201	0.0209	0.0209	0.0225	0.026	0.0256	0.0668	0.0741	0.073
		0.0167	0.0206	0.021	0.0214	0.0228	0.026	0.0257	0.0675	0.0731	0.0727
		0.0335	0.021	0.0207	0.0208	0.026	0.0273	0.0275	0.0721	0.0733	0.0738

Table B.6: MSE and Kullback–Leibler divergence of the MDPD, FASTMCD, and DetMCD for the mean vector and scatter matrices for simulation scenarios: $p = 10$, $ncp = 250$, $\delta = 1$

bdp	n	eps	e_μ			e_Σ			e_{KL}		
			MDPD	FastM	DetM	MDPD	FastM	DetM	MDPD	FastM	DetM
0.067	50	0.0223	0.2605	0.2241	0.2159	0.2967	0.3313	0.3271	1.0112	0.8692	0.8524
		0.0335	0.2670	0.2244	0.2221	0.3036	0.3374	0.3272	1.0207	0.8781	0.8539
		0.0670	0.2688	0.2122	0.2158	0.3249	0.3404	0.3386	1.0708	0.8038	0.8078
	100	0.0223	0.1252	0.1072	0.1108	0.1403	0.1534	0.1503	0.4351	0.413	0.406
		0.0335	0.1249	0.1105	0.108	0.1433	0.1534	0.1506	0.4432	0.4059	0.3998
		0.0670	0.1305	0.1081	0.1086	0.1496	0.1585	0.1575	0.4593	0.3895	0.39
	200	0.0223	0.0619	0.055	0.0569	0.0696	0.0747	0.0742	0.2087	0.205	0.2035
		0.0335	0.0606	0.0557	0.0542	0.0698	0.0762	0.0745	0.2085	0.2041	0.2003
		0.0670	0.0640	0.0533	0.0541	0.0746	0.0816	0.0801	0.2194	0.2011	0.1985
	500	0.0223	0.0246	0.023	0.0228	0.0277	0.0297	0.0293	0.0818	0.0821	0.0813
		0.0335	0.0248	0.0219	0.0226	0.0285	0.0303	0.03	0.0835	0.0818	0.0815
		0.0670	0.0250	0.0209	0.0217	0.0295	0.0352	0.0355	0.0853	0.0873	0.0881
0.0502	50	0.0167	0.2086	0.2086	0.213	0.234	0.2917	0.291	0.7466	0.7956	0.7944
		0.0251	0.2115	0.2109	0.2124	0.2323	0.2913	0.2939	0.7497	0.7976	0.7969
		0.0502	0.2120	0.2113	0.2087	0.2368	0.2968	0.2933	0.7581	0.7661	0.7644
	100	0.0167	0.1037	0.1105	0.1077	0.1146	0.1449	0.1419	0.3516	0.4048	0.3952
		0.0251	0.1054	0.1104	0.1089	0.116	0.1445	0.1435	0.3564	0.3974	0.3937
		0.0502	0.1091	0.1056	0.1073	0.1216	0.1505	0.1479	0.37	0.3794	0.3766
	200	0.01670	0.0524	0.0544	0.0568	0.0581	0.0705	0.0695	0.1750	0.1966	0.1956
		0.0251	0.0541	0.0536	0.0537	0.0591	0.0712	0.0698	0.1779	0.1945	0.1918
		0.0502	0.0554	0.0519	0.0523	0.0605	0.074	0.0744	0.1814	0.189	0.19
	500	0.0167	0.0207	0.0222	0.0215	0.0233	0.0279	0.0277	0.0689	0.0784	0.0776
		0.0251	0.0210	0.0215	0.022	0.0234	0.0281	0.0282	0.0693	0.0773	0.0778
		0.0502	0.0218	0.0214	0.0214	0.0245	0.0316	0.0312	0.0722	0.0811	0.0804
0.0335	50	0.0112	0.2047	0.2033	0.2029	0.2272	0.2541	0.254	0.7263	0.7237	0.722
		0.0167	0.2057	0.2007	0.2072	0.2237	0.2571	0.257	0.7202	0.7306	0.7311
		0.0335	0.2053	0.2046	0.204	0.2284	0.2562	0.2566	0.7306	0.7321	0.7235
	100	0.0112	0.1038	0.1054	0.1073	0.1129	0.1308	0.1325	0.3477	0.3754	0.3761
		0.0167	0.1016	0.1068	0.1068	0.1127	0.132	0.1302	0.3461	0.3751	0.3726
		0.0335	0.1077	0.1037	0.1035	0.114	0.1328	0.1338	0.352	0.3597	0.3591
	200	0.0112	0.0505	0.0533	0.0533	0.0568	0.0649	0.0635	0.1702	0.1856	0.1832
		0.0167	0.052	0.0517	0.0526	0.0564	0.0651	0.0648	0.1693	0.1838	0.1825
		0.0335	0.0534	0.0521	0.0511	0.0589	0.0659	0.0661	0.1752	0.1785	0.1781
	500	0.0112	0.0207	0.0217	0.0214	0.0224	0.0259	0.0259	0.0667	0.0741	0.074
		0.0167	0.0206	0.0211	0.0211	0.0228	0.0261	0.0261	0.0673	0.0735	0.0735
		0.0335	0.0209	0.0208	0.0208	0.0239	0.0272	0.0276	0.0698	0.0733	0.0739

Curriculum Vitae

Ebenezer Nkum was born on November 23, 1993, as the first male child of Margaret Nkum and Oswell Nkum. He entered the University of Cape Coast (UCC), Ghana after high school in 2013 for his undergraduate studies in BSc Actuarial Science. He graduated in 2018 as one of the best four graduating students from the Mathematics and Statistics Department. While pursuing his degree at UCC he served in various leadership positions, including Education and Career Development Secretary and the Vice President of Enyan Denkyira Students and Associates, UCC. In 2017, he rendered his national service as an Actuarial and Data Analyst at Vanguard Life Company Limited, Ghana. He was retained as a full employee and worked for one year.

In pursuit of becoming a researcher and professional in the field of Data Science, he entered the Graduate School at The University of Texas at El Paso in the Fall of 2019 for his Master's degree in Mathematical Sciences to lay a foundation for a future PhD in Data Science. He was a regular contributor and participant in a number of workshops and seminars in Mathematics, Computational Science, Data Science, and Statistics. While pursuing a Master's degree in Mathematics, he worked as a Teaching and Research Assistant. He was a member of Gamma Iota Sigma (GIS). He is an active member of the International Association of Black Actuaries and is working toward his designation as Fellow of the Society of Actuaries with five of their credential papers to his credit. He is also a member of the American Mathematical Society. He has participated in many conferences such as the Joint Statistical Meeting (JSM, Diversity and Mentoring - 2020), StatFest 2020, and the International Association of Black Actuaries (2020). After graduation, Ebenezer pursued a doctoral degree in Data Science at The University of Texas at El Paso.

Ebenezer Nkum is currently a PhD candidate in Data Science at The University of Texas at El Paso, expected to graduate in May 2024. His dissertation is titled "Robust Multivariate

Estimation and Inference with the Minimum Density Power Divergence Estimator.” During his PhD studies, he served as a Teaching Assistant and Research Associate, contributing to various research projects and teaching undergraduate courses. He was also actively involved in the American Mathematical Society UTEP Student Chapter, serving as President. Ebenezer’s research interests include robust statistical methods, machine learning, and data science applications in finance, engineering, and geology. He has published several research papers and presented his work at numerous conferences.

After completing his PhD, Ebenezer will join Cigna as an Actuarial Data Scientist. He has a long-term career plan of becoming a research scientist, focusing on the development and application of robust statistical techniques in high-dimensional data analysis.

Email: enkum@miners.utep.edu