SOME PROPERTIES OF ROBUST STATISTICS UNDER ASYMMETRIC MODELS

BY JUE WANG

A dissertation submitted to the

Graduate School—New Brunswick

Rutgers, The State University of New Jersey

in partial fulfillment of the requirements

for the degree of

Doctor of Philosophy

Graduate Program in Statistics

Written under the direction of

David E. Tyler

and approved by

New Brunswick, New Jersey

October, 2008

ABSTRACT OF THE DISSERTATION

Some Properties of Robust Statistics Under Asymmetric Models

by JUE WANG

Dissertation Director: David E. Tyler

Properties of robust statistics have been extensively studied in the univariate setting when the underlying model is presumed to be symmetric, and in the multivariate case when the underlying model is presumed to be elliptically symmetric. Much less attention has been given to the behavior of robust statistics under asymmetric models. The goal of this dissertation is thus to obtain theoretical results for robust statistics under asymmetric models. To this end, local asymmetric alternatives to symmetric and elliptically symmetric distributions are considered. A key tool used in obtaining the theories presented in this dissertation is the LeCam's lemmas on contiguity.

The classes of robust univariate statistic considered here are the M-estimates, one-step version of the M-estimates, the W-estimates and the trimmed means. The classes of robust multivariate statistics considered are the M-estimates, the S-estimates, the CM-estimates and the MM-estimates, which are all treated under the unified framework of M-estimates with auxiliary scale, as well as their one-step versions. Asymptotic distributions of these statistics are obtained under local mixture models and skew-symmetric models. The asymptotic properties for the MM-estimates, even under elliptical symmetry, are the first such results for the multivariate MM-estimates.

Under asymmetry, different robust statistics for location are not consistent with each

other, i.e. they are estimating different notions of central tendency. Likewise, in the multivariate setting, under non-elliptical distributions, the different scatter statistics are again not consistent with each other and are reflecting different structures of the underlying distribution. This suggests the difference in location statistics can be used to detect asymmetry and the comparison of different scatter statistics can be used to detect deviations from elliptical symmetry.

Consequently, new classes of tests for symmetry and for elliptical symmetry are introduced in this dissertation based upon the comparisons of different location statistics and different scatter statistics respectively. Furthermore, the asymptotic null distributions of the proposed test statistics are derived as well as their local power functions under contiguous mixture distributions. The local power functions help provide some guidelines for choosing the proper tuning constant of the proposed tests.

Acknowledgements

Many people have contributed to the thesis. Above all is my advisor Professor David E. Tyler, the most distinguished professor I have ever met, who opened to me the beautiful world of statistics, who embarked me on robustness, cheered me on and rendered invaluable guidance and inspirations every step of the way, whose academic spirit and personal charm will influence the rest of my life. No one could have done more.

I am infinitely grateful to Fang Zhang, an outstanding Fudan alumnus and a true friend of twenty years, a rising entrepreneur with a generous heart, a genuine mind, and a shining smile; who never gives up upon me, who never loses confidence in me even at times when I am having doubt on myself; without whose persistent and enthusiastic support, I could not move on, I could not have completed this thesis.

My sincere appreciation goes to my colleague Ren Guo, a brilliant mathematician of unlimited future, whose complete devotion to friendship, whose eternal passion for academic research has been a unique source to sustain me throughout the many years of my own course.

The paper is also a dedication to my parents, AnMin and GuiLian, who gave me twice my life. For their endless and unconditional love I own immeasurably.

Dedication

- To the students of my generation who made heroic sacrifice in June of 1989 for the future of China
- To the students of my generation who are still holding the simple yet beautiful dream

Table of Contents

List of Tables

List of Figures

Chapter 1

Introduction

1.1 Preliminaries

The validity of many widely applied statistical procedures in multivariate analysis depends on a key assumption that the data comes from a normal distribution. Under normality assumption, the sample mean and sample covariance-matrix jointly are sufficient statistics based on which the inferences about location and scatter of the underlying distribution can be correctly described and summarized. Even though they form the basis of many standard data analysis techniques, sample mean and sample covariance matrix are unfortunately non-robust in that they are very sensitive to and can be grossly distorted by tiny amount of perturbation in the underlying normal population. At the presence of even one outlier, inference based on sample mean and sample covariance becomes unreliable and skeptical.

A variety of alternative choices on the estimates of multivariate location vector and scatter-matrix have since been proposed in an attempt to protect against the sensitivity to non-normality, but these robust estimates may not be very accurate in case no outlying observations are present. To balance extreme sensitivity and accuracy, quantitative measures on these two descriptive quantities are also developed during the evolution of robust estimators. The sensitivity or robustness of estimators are measured both globally by the concept of finite sample breakdown point [8, 12], which indicates the smallest proportion of outliers that can take the estimate over all bounds, and locally by influence function [12], which reflects the impact of an infinitesimal contamination having on the estimate; whereas the accuracy of estimators are customarily quantified and compared by their asymptotic efficiency.

An early class of robust estimates was the M-estimates proposed by Maronna (1976).

They have a high efficiency over a broad range of population model, and a bounded influence function, that is, they are locally robust to, hence not greatly disturbed by, a small perturbation in the data. Their global robustness, however, have a low breakdown point [24] which is disappointingly no more than $1/(p+1)$, where p is the dimension of the data, due to the increasing sensitivity of M-estimates of scatter to outliers in a higher dimensional data [37].

Subsequently, different versions of robust estimates with high breakdown points were proposed by Stahel (1981), Donoho (1982), Rousseeuw (1985) and others. Today the most widely known high breakdown multivariate estimates of location and scatter are probably the S-estimates introduced by Davies (1987) and Lopuhaä (1989). Complimentary to M-estimates, many of the high breakdown point statistics come with an ironic disadvantage that they are inefficient at normal model [22] or tend to have poor local robustness properties [38]. To some extend, S-estimates and M-estimates behave oppositely to each other, choosing between the two estimates is to make a trade off between asymptotic efficiency and breakdown point.

In the hope of bringing the efficiency of M-estimates together with the high breakdown property of S-estimates and with bounded influence, more flexible variations of S-estimates have recently been discussed. Notably among them were the Constrained M-estimates, or CM-estimates for shorthand, introduced by Kent and Tyler (1996), and one-step W-estimates studied by Lopuhaä (1999) - that is to perform a one-step appropriately chosen re-weight on an initial estimator of high breakdown and bounded influence (naturally S-estimates for example). Both CM-estimates and one-step W-estimates on an S-estimate improve the efficiency and local robustness while maintaining the high breakdown.

Tatsuoka and Tyler (2000) noted that multivariate estimates discussed so far can be embedded within a larger class of estimates which they called multivariate M-estimates with auxiliary scale. M-, S-, and CM-estimates are identified as special cases of this super class with *simultaneously* defined auxiliary scale statistics. Alternatively, Tatsuoka and Tyler [33] proposed to use a *preliminary* scale statistic. This gives the rise to the class of MM-estimates, which are essentially redescending M-estimates of high asymptotic efficiencies with high breakdown points, and when properly tuned, MM-estimates can achieve low gross error sensitivity (high local robustness) as well.

Under the elliptical symmetric model, the consistency and asymptotic normality were proved by Maronna (1976) for M-estimates (with unbounded ρ-function), by Davies (1987) for S-estimates, and by Kent and Tyler (1996) for CM-estimates under certain regularity requirements. The uniqueness and consistency for MM-estimates along with breakdown properties were established by Tatsuoka and Tyler (2000) under broader classes of symmetric distributions. The asymptotic normality and local robustness of MM-estimates, however, were left as open questions. One of the primary purposes of this dissertation is to study the local robustness of MM-estimates in terms of influence function, and to derive the forms of the asymptotic distributions of M-, S-, MM-, and one-step W-estimates under certain non-elliptical models, particularly under the mixture of two elliptical distributions of different shapes. Before studying their asymptotic under asymmetry, these robust estimates are to be formally introduced and a nice property shared by all of them, namely affine equivariance, is to be discussed.

1.2 Class of Affine Equivariant Estimates of Location Vector and Scatter/Shape Matrix

Affine Equivariance

Let $F_{\mathbf{z}}$ represent the distribution of $\mathbf{z} \in \mathbb{R}^p$, and $F_{\mathbf{x}}$ represent the distribution of $\mathbf{x} = B\mathbf{z} + b$, with B being a nonsingular matrix of order p and $b \in \mathbb{R}^p$. The location and scatter functionals of F, denoted by $\boldsymbol{\mu}(F) \in \mathbb{R}^p$ and $V(F) \in \mathcal{PDS}(p)$ - the set of all positive definite symmetric matrices of order p, are affine equivariant iff

$$\boldsymbol{\mu}(F_{\mathbf{x}}) = B\boldsymbol{\mu}(F_{\mathbf{z}}) + b \quad \text{and} \quad V(F_{\mathbf{x}}) = B\,V(F_{\mathbf{z}})\,B' \tag{1.2.1}$$

Given a scatter functional $V(F)$, the *scale functional* of distribution F is defined as $\sigma(F) = |V(F)|^{1/2p}$. A scale functional $\sigma(F)$ is affine equivariant iff

$$\sigma(F_{\mathbf{x}}) = |\det(B)|^{1/p}\sigma(F_{\mathbf{z}}) \tag{1.2.2}$$

The sample versions of location, scatter and scale estimates, denoted by $\widehat{\boldsymbol{\mu}}(F)$, $\widehat{V}(F)$, and $\widehat{\sigma}(F)$ respectively are affine equivariant iff they satisfy respective equations of (1.2.1) and (1.2.2).

M-estimates of location and scatter are originally defined as solutions to the M-estimating equations

$$\widehat{\boldsymbol{\mu}} = \frac{\sum_{i=1}^{n} u(s_i)\mathbf{x}_i}{\sum_{i=1}^{n} u(s_i)} \quad \text{and} \quad \widehat{V} = \sum_{i=1}^{n} u(s_i)\left(\mathbf{x}_i - \widehat{\boldsymbol{\mu}}\right)\left(\mathbf{x}_i - \widehat{\boldsymbol{\mu}}\right)' \tag{1.2.3}$$

where $s_i = \left(\mathbf{x}_i - \widehat{\boldsymbol{\mu}}\right)\widehat{V}^{-1}\left(\mathbf{x}_i - \widehat{\boldsymbol{\mu}}\right)'$, and $u(\cdot)$ is some appropriately chosen weight function.

S-estimates are defined as the solution of the optimization problem where $t_i = [(\mathbf{x}_i - \widehat{\boldsymbol{\mu}})\widehat{V}^{-1}(\mathbf{x}_i - \widehat{\boldsymbol{\mu}})]^{1/2}$:

$$\min|\widehat{V}| \quad \text{subject to} \quad \frac{1}{n}\sum \rho(t_i) = \epsilon_0\rho(+\infty)$$

where ϵ is a fixed value between 0 and 1/2, and $\rho(t)$ is a bounded non-increasing function for $t > 0$. For a differentiable ρ with derivative ψ, the S-estimates satisfy the simultaneous S-estimating equations [20]

$$\widehat{\boldsymbol{\mu}} = \frac{\sum_{i=1}^{n} u(t_i)\mathbf{x_i}}{\sum_{i=1}^{n} u(t_i)} \quad \text{and} \quad \widehat{V} = \frac{\sum_{i=1}^{n} u(t_i)\big(\mathbf{x}_i - \widehat{\boldsymbol{\mu}}\big)\big(\mathbf{x}_i - \widehat{\boldsymbol{\mu}}\big)'}{\sum_{i=1}^{n} w(t_i)} \tag{1.2.4}$$

where $u(t) = \psi(t)/t$ and $w(t) = \left[\psi(t)t - \rho(t) + \epsilon_0\rho(+\infty)\right]/p$

One-step W-estimates are re-weighted version of sample mean and sample covariance matrix with weights relying on initial affine equivariant estimates of location $\widehat{\boldsymbol{\mu}}_o$ and scatter $\widehat{V}_o$,

$$\widehat{\boldsymbol{\mu}} = \frac{\sum_{i}^{n} u_1(s_{o,i})\mathbf{x}_i}{\sum_{i}^{n} u_1(s_{o,i})}, \quad \text{and} \quad \widehat{V} = \frac{\sum_{i}^{n} u_2(s_{o,i})(\mathbf{x}_i - \widehat{\boldsymbol{\mu}}_o)(\mathbf{x}_i - \widehat{\boldsymbol{\mu}}_o)'}{\sum_{i}^{n} u_2(s_{o,i})} \tag{1.2.5}$$

where $s_{o,i} = (\mathbf{x}_i - \widehat{\boldsymbol{\mu}}_o)'V_o^{-1}(\mathbf{x}_i - \widehat{\boldsymbol{\mu}}_o)$.

MM-estimates

Let $\widehat{\sigma}$ be a preliminary scale-estimate satisfying the equivariant property (1.2.2), and suppose $(\widehat{\boldsymbol{\mu}}, \widehat{\Gamma})$ is any solution which minimizes

$$\sum_{i=1}^{n} \rho\left\{\frac{(\mathbf{x}_i - \widehat{\boldsymbol{\mu}})'\widehat{\Gamma}^{-1}(\mathbf{x}_i - \widehat{\boldsymbol{\mu}})}{\widehat{\sigma}^2(F)}\right\}$$

over all $\widehat{\boldsymbol{\mu}} \in \mathbb{R}^p$ and $\widehat{\Gamma} \in \mathcal{PDS}(p)$ with $|\widehat{\Gamma}| = 1$. Then $\widehat{\boldsymbol{\mu}}$ and $\widehat{V} = \widehat{\sigma}^2\widehat{\Gamma}$ are called MM-estimates of location and scatter [39].

It will be shown in Section 5.5 that for a differentiable ρ function, the MM-estimating equations are

$$\widehat{\boldsymbol{\mu}} = \frac{\sum_{i=1}^n u(s_i)\mathbf{x_i}}{\sum_{i=1}^n u(t_i)} \quad \text{and} \quad \widehat{V} = \frac{\sum_{i=1}^n u(s_i)\big(\mathbf{x}_i - \widehat{\boldsymbol{\mu}}\big)\big(\mathbf{x}_i - \widehat{\boldsymbol{\mu}}\big)'}{\sum_{i=1}^n \psi(s_i)/p + \log|\widehat{\Gamma}|} \tag{1.2.6}$$

where $s_i = (\mathbf{x}_i - \widehat{\boldsymbol{\mu}})\widehat{V}^{-1}(\mathbf{x}_i - \widehat{\boldsymbol{\mu}})$, $u(s) = 2\rho'(s)$ and $\psi(s) = u(s)t$.

1.3 Exploring Difference in Robust Statistics, and Motivation

In the univariate setting, if the underlying population is symmetric, then all location estimates are basically estimating the same characteristic of the distribution, namely the center of symmetry. On the contrary, if the data arises from a distribution which is asymmetric, different location estimates represent different aspects of the central tendency; the quantified measures yielded from variety of location estimates may not be the same. This simple observation underlies the origination of Pearson's skewness measure, and subsequent variations following it – Bowley's and Yule's coefficients for examples. These classical skewness coefficients are in essence the standardized differences between two different location estimates which are zero at symmetric models but not necessarily equal to zero at asymmetric distributions.

These experiences with the univariate data motivate one to apply the similar ideas in the multivariate analysis of scatter matrix. A scatter functional $V(F)$ on a distribution F can be decomposed into what Kent and Tyler (1996) call "shape" and "scale" components. A shape component is any function of $V(F)$ which is invariant under a positive scalar multiple, such as $V/\mathrm{tr}(V)$; and a scale component is any equivariant function of $V(F)$ under the same transformation, such as $\mathrm{tr}(V)$. Under the spherical distributions $E_p(\mathbf{0}, I_p)$, all scatter estimates should have the same shape component which is the identity matrix, though they may have varying scales. It follows, all scatter estimates are proportional to each other if sampling from an elliptical symmetric distribution $E_p(\boldsymbol{\mu}, \Sigma)$. This claim will not hold true if the data comes from a distribution other than an elliptical symmetric distribution, even one which is symmetric [40].

This suggests an idea that the departure of underlying distribution from an elliptical symmetric may well be detected by comparing the shapes of different scatter matrices as they are measuring different quantities of a non-elliptical distribution; a new test of elliptical symmetry against non-elliptical asymmetry may be constructed based on this comparison.

On the other hand, the classical skewness measures at the univariate setting are sometimes equals to zero hence fail to capture the skewness of certain non-symmetric models. This is simply because these statistics are comparing pre-determined, fixed two location estimates whose values may be coincident even though the underlying population is clearly asymmetric. This unpleasant drawback can be overcome by looking at values of a class of location estimates instead of particularly two. A class of location functional indexed by a tuning constant associated with the defining weight function can be viewed as a transformation of distribution, which is constant if and only if the underlying distribution is symmetry. For a sample of course, there will be some statistical variability, and no sample transform will be exactly constant. By comparing a class of location estimates, though, one can construct a consistent test of asymmetry or skewness by considering the difference between maximum and minimum values of estimates or by some other measure of non-constancy. This idea can too be extended to the multivariate analysis. The best shape test of non-elliptical distributions ought to be built up on the maximum difference between classes of scatter matrices rather than on the difference of particularly two scatter estimates.

This paper is the first attempt to implement these ideas.

1.4 Organizations of Dissertation

This thesis has essentially two self-contained but closely related parts, one for univariate location estimates and the other for multivariate scatter-matrices. Though primary studies are focused on multivariate scatters, the ideas there are originated from and preluded by the reflecting cases in the univariate setting.

The paper is organized as follow. Chapter 3 alone is solely attributed to univariate location problems where the asymptotic distributions of some traditional skewness measures, and of a new procedure in the form of the difference between two robust location estimates are studied and formulated under three different non-symmetric models; the concept of separating class is introduced; and the efficiencies of the new procedure at testing skewness or detecting mixture are derived and compared against classical skewness coefficients.

The multivariate scatter estimates are studied through chapter 4, 5 and 6. Chapter 4 lays out the theoretical foundations for the asymptotic properties of the robust scatter estimates under asymmetric models of skew-elliptical or elliptical mixture. The specifics of asymptotic normality for the individual members themselves aforementioned from the general class of Multivariate M-estimates with auxiliary scale are presented in Chapter 5 with emphasis on the MM-estimates, of which the influence function for local robustness is deduced and comparisons of efficiencies at elliptical mixture models are completed across families of MM-estimates and through different underlying spherical distributions. A new set of statistics based on two robust scatter estimates is formally proposed in Chapter 6 for the multivariate shape analysis. The asymptotic relative efficiency of the new statistic at testing elliptical mixture against elliptical symmetric model is compared within M-estimates and within MM-estimates.

The main technical tools implemented to obtain the results in both the univariate and multivariate cases are the LeCam's lemmas related to the concepts of contiguity which, for completeness and smoothness, are themselves constituents of Chapter 2.

The thesis is concluded by Chapter 7 where conclusions and open research questions are enumerated along with further discussions. Formal proofs as well as tables and figures are reserved in the appendices of each respective chapter for the neatness of the layout.

Chapter 2

Background: Concept of Contiguity, LeCam's Lemmas and Asymptotic Normality

Most descriptive statistics and test procedures employed in a classical multivariate analysis are based on a key assumption that the underlying distribution is a normal where the location (mean) vector together with the scatter matrix is a sufficient statistic. Recently, Robust statistics researches have been flourishing in a particular generation of the normal distribution, namely class of elliptical distributions, denoted by $E_p\big(\mathbf{b}, \Sigma\big)$, whose density is of the form

$$f(\mathbf{x}) = |\Sigma|^{-1/2} h\left[(\mathbf{x}-\mathbf{b})'\Sigma^{-1}(\mathbf{x}-\mathbf{b})\right],$$

where $\mathbf{x}$, $\mathbf{b} \in \mathbb{R}^p$, $\Sigma \in \mathcal{PDS}(p)$, the set of positive definite symmetric matrices of order p, and $h : [0, \infty) \to [0, \infty)$ is a fixed function depending $\mathbf{x}$ only through $(\mathbf{x}-\mathbf{b})'\Sigma^{-1}(\mathbf{x}-\mathbf{b})$.

Under elliptical distribution, several classes of robust estimates of multivariate location and scatter estimates have been implemented, trying to address the concerns of non-normality, especially a distribution with longer tails (than a Normal) on either direction or with outlying points. Though comparison of these newly proposed estimates haven't been completed and studies within elliptical distribution are still promising, vigorous research beyond elliptical has yet to provide any encouraging result, to bear fruit.

The distribution space outside elliptical is unlimited, yet one can focus on the first steps of a natural extension beyond symmetry of Elliptical. A skew-elliptical distribution, and a mixture of two elliptical distributions with different shapes are examples of such asymmetric distributions. To investigate the characteristics of a traditional statistical procedure under these asymmetric framework, a basic yet powerful technical instrument is the **contiguity** of probability measures that leads to (local) asymptotic normality of statistical models.

A comprehensive treatment of contiguity can be found, for example, in [13, 29]. For the purpose of reference and completeness, main results are enumerated in this section.

Definition 2.0.1. *Let $(\Omega, \mathcal{A})$ be some measurable space, with $\mathcal{M}(\mathcal{A})$ denoting the set of all finite measures. A bounded sequence $Q_n \in \mathcal{M}(\mathcal{A})$ is called contiguous to another bounded sequence $P_n \in \mathcal{M}(\mathcal{A})$ if for any sequence of events $A_n \in \mathcal{A}$,*

$$\lim_{n\to\infty} P_n(A_n) = 0 \implies \lim_{n\to\infty} Q_n(A_n) = 0$$

The concept of contiguity is essentially instrumental for the asymptotic methods, which are the primary technologies implemented in this dissertation to investigate the statistical properties of some common Robust-procedures under certain non-symmetric distributions. Imagine that Q_n is a sequence of asymmetric distributions of interests, while P_n is a sequence of symmetric ones. Once the contiguity of Q_n to P_n is established, the asymptotic behaviors of a statistic S_n that is based on sample $\mathbf{z}_1, \cdots, \mathbf{z}_n \in \mathbb{R}^p$ are well founded in the elegant LeCam third Lemma. Moreover, the powerful LeCam first Lemma renders a surprisingly simple tool to examine the contiguity.

Lemma 2.0.2. *(LeCam first Lemma) Let $P_n, Q_n \in \mathcal{M}(\mathcal{A})$ be two sequences of probabilities. If the log-likelihood function $L_n = \log dQ_n/dP_n$ is under P_n asymptotically normal,*

$$L_n(P_n) \xrightarrow{d} N\big(-\tau^2/2, \tau^2\big)$$

for some $\tau \in [0, \infty)$, then Q_n is contiguous to P_n.

Lemma 2.0.3. *(LeCam third Lemma) Let P_n, Q_n be two sequences of probabilities with log likelihood $L_n = \log dQ_n/dP_n$, and S_n a sequence of statistics on $(\Omega, \mathcal{A})$ taking values in some finite-dimensional $\big(\bar{\mathbb{R}}^m, \bar{\mathbb{B}}^m\big)$, where $\bar{\mathbb{B}}$ is a Borel σ-algebra on $\bar{\mathbb{R}}$, such that for some $\mathbf{a}$, $\mathbf{c} \in \mathbb{R}^m$, $\tau \in [0, \infty)$, and $A \in \mathbb{R}^m_m$*

$$\begin{pmatrix} S_n \\ L_n \end{pmatrix}(P_n) \xrightarrow{d} N\left[\begin{pmatrix} \mathbf{a} \\ -\tau^2/2 \end{pmatrix}, \begin{pmatrix} A & \mathbf{c} \\ \mathbf{c}' & \tau^2 \end{pmatrix}\right]$$

Then

$$\begin{pmatrix} S_n \\ L_n \end{pmatrix}(Q_n) \xrightarrow{d} N\left[\begin{pmatrix} \mathbf{a}+\mathbf{c} \\ \tau^2/2 \end{pmatrix}, \begin{pmatrix} A & \mathbf{c} \\ \mathbf{c}' & \tau^2 \end{pmatrix}\right]$$

The proofs of LeCam's lemmas can be found, for example, in [29].

Contiguity and LeCam's Lemmas together suggest a feasible and efficient platform on which the behavioral differences in statistical aspects of a robust procedure (statistic), if any, when the underlying distribution changes from elliptical-symmetric P to asymmetric Q, could be tracked, analyzed and compared locally and asymptotically. One can set up a sequence of (local) hypotheses, $H_{0,n}$ verses $H_{1,n}$ indexed by $n \in \mathbb{N}$, in which Q_n is the joint distribution of a sample of size n when it is from alternative hypothesis while P_n is the joint distribution when sampling from null-hypothesis. The hypotheses are set-up in a way that Q_n is contiguous to P_n, and for a fixed n, Q_n and P_n are the (local) realizations of asymmetric Q and symmetric P respectively.

If location/shape functions exhibit interesting statistical differences between P and Q (expressed locally and asymptotically as between Q_n and P_n), then analyses on the differences across class of location/shape functions may lead to the suggestion or development of a new robust procedure at testing and estimating the location/shape of a multivariate data that would work equally powerful even when the underlying distribution is not limited to a symmetric one, rather, includes certain asymmetric ones.

In this spirit, the asymptotics of a location/shape estimate will be investigated at various platforms of asymmetric distributions.

Chapter 3

Univariate Location Estimates at Skewed or Mixture Model

3.1 Introduction

The concept of skewness - shape and asymmetry - of a univariate distribution has long history in the literature, beginning with different measures on how to quantitate skewness. Most of today's classical functionals measuring skewness were introduced in the early 1920. Using notations μ, M, σ^2 for the population mean, mode, and variance respectively of a univariate random variable $X \sim F$, and letting Q_p represent $(p*100)$th percentile (quartile) of F, some of the early proposed measures of skewness were

$(\mu - M)/\sigma$	Pearson (1895)
$E(X-\mu)^3/\sigma^3$	Charlier (1905) and Edgeworth (1904)
$(Q_{0.75} + Q_{0.25} - 2Q_{0.5})/(Q_{0.75} - Q_{0.25})$	Bowley (1920)
$(\mu - Q_{0.5})/\sigma$	Yule (1911)

David and Johnson (1954) suggested a generalization of Bowley's coefficient

$$\gamma_p(F) = \frac{[Q_{1-p} - Q_{0.5}] - [Q_{0.5} - Q_p]}{[Q_{1-p} - Q_{0.5}] + [Q_{0.5} - Q_p]}, \quad p \in \left(0, \frac{1}{2}\right)$$

Integrating the numerator and denominator of $\gamma_p(F)$ yields another skewness coefficient

$$\frac{\int_0^{1/2} [Q_{1-p} + Q_p - 2Q_{0.5}]\, dp}{\int_0^{1/2} [Q_{1-p} - Q_p]\, dp} = \frac{\mu - Q_{0.5}}{E|X - Q_{0.5}|}$$

Except the most prominent "measure of skewness", $E(X-\mu)^3/\sigma^3$, these skewness coefficients are essentially test statistics measuring a standardized distance between two separate location parameters. In particular, the Yule's skewness coefficient $(\mu - Q_{0.5})/\sigma$ and Pearson's skewness measure $(\mu - M)/\sigma$ are members of the family of statistics in the form $(\mu_1 - \mu_2)/\sigma$, where $\mu_i' s$ are affine equivariant location estimates, whereas Bowley's coefficient $(Q_{0.75} + Q_{0.25} - 2Q_{0.5})/(Q_{0.75} - Q_{0.25})$ can be viewed as a special case of the family that represents a comparison of two different measures of central tendency or location measure, namely $(Q_{1-p} + Q_p)/2$ and the median.

The sample versions of these measures have long been served as classical test statistics in the detection of asymmetry as well. The performances of these skewness coefficients in the context of hypothesis testing under certain alternative non-symmetric models have been studied and compared rigorously. Abundant research papers focusing on this subject can be found in the literature. Yet most of the comparisons are either limited within these skewness measures themselves or isolated made with respect to a newly-proposed but eventually short-lived test statistic. Little attention has been paid to from the point of view that they are distinct but plain members of a broader class; little is known as what is the asymptotic distribution of an arbitrary member from this class under some common models - symmetric or asymmetric; and what is the relative efficiency between two members of this class that are randomly chosen to test skewness or detect asymmetry. This chapter is devoted to answer these questions.

Let $\widehat{\mu}_i$ be the consistent estimator of μ_i based on empirical sample, one part of this dissertation is intended to investigate the statistical properties of statistic $\widehat{T} = \widehat{\mu}_2 - \widehat{\mu}_1$ and its relative efficacy as a test statistic compared with traditional procedures when samples are from a skewed distribution of one of the following populations

Mixture $(1 - \epsilon)f(x) + \epsilon g(x - \theta)$

Asymmetry $f(x)I_{[x\leq 0]} + \frac{1}{\tau} f\left(\frac{x}{\tau}\right) I_{[x>0]}$

Skew-Symmetry $2f(x)G(\theta x)$

where f and g are symmetric about 0; $\epsilon \in (0, 1)$, $\theta \neq 0$, $\tau > 0$ are fixed.

3.2 Difference of Two Location Estimates

These classical statistics were first introduced as "measures of skewness". Since these measures are zero for any symmetric distribution, estimates of these measures have served as test statistics in testing for asymmetry. A disappointing problem associated with these test statistics is that the corresponding skewness measure being zero doesn't necessarily imply the underlying distribution is symmetry, and so such tests are not consistent for detecting asymmetry. In general, two different location measures can be equal even though the underlying distribution is clearly non-symmetric. Consequently, a test for asymmetry based upon the difference of two particular location estimates can fail to detect asymmetry.

Many examples can be constructed to illustrate the aforementioned phenomena. Consider, for example, the mixture of two normal distributions given by

$$\frac{3}{5}\,N(0,1) + \frac{2}{5}\,N\left(\frac{5}{2}z_{0.8}, \frac{9}{4}\left(z_{0.8}/z_{0.05}\right)^2\right), \tag{3.2.1}$$

where z_α is the $100\alpha\%$ quantile of a standard normal. This distribution is asymmetric, as is evident from the plot of its density displayed in Figure 3.1.a. Nevertheless, both the mean and the median of this normal mixture are equal to $z_{0.8}$, and hence Yule's measure of skewness is zero. An example of an asymmetric distribution for which the mean and mode coincide, and hence has a Pearson's skewness measure equal to zero, is one with a density given by

$$f(x) = \frac{2\gamma}{1+2\gamma}\left((1+\gamma x)\mathrm{I}_{[-\frac{1}{\gamma}\leq x<0]} + e^{-x}\mathrm{I}_{[x\geq 0]}\right), \tag{3.2.2}$$

with $\gamma = 1/\sqrt{6}$. This density is plotted in Figure 3.1.b, and has both a mean and a mode equal to zero. An example of an asymmetric distribution for which Bowley's skewness coefficient has a value of zero can be obtain by considering a half Normal and half Cauchy density, This density is given by

$$f(x) = \frac{1}{2\pi}e^{-\frac{x^2}{2}}\mathrm{I}_{[x\leq 0]} + \frac{\mathrm{I}_{[x>0]}}{\pi\ z_{0.75}\ \left(1+\left(x/z_{0.75}\right)^2\right)}, \tag{3.2.3}$$

which is plotted in Figure 3.1.c. For this density, the first, second and third quantile agree with those of a standard normal distribution. Finally, consider a standard Weibull(k) distribution, whose density (Figure 3.1.d) is given by

$$f(x) = kx^{(k-1)}e^{-x^k} \quad \text{for} \quad x \geq 0, \tag{3.2.4}$$

with $k > 0$ being a shape parameter. This distribution has mean $\mu = \Gamma(1 + 1/k)$, median $Q_{0.5} = \sqrt[k]{\ln 2}$, mode $M = \sqrt[k]{(1 - 1/k)}$ and quartiles $Q_p = (-\ln(1-p))^{1/k}$ for $p \in (0, 1)$. Thus, for certain values of the shape parameter k, the Weibull distribution, although nonsymmetric, can have a Yule coefficient of zero, a Pearson coefficient of zero, or a Bowley coefficient of zero. For $k = 3.4395$, one obtains $\mu = Q_{0.5} = 0.89892$, whereas for $k = 3.31247$, one obtains $\mu = M = 0.89718$, and for $k = 3.2883$, one obtains $(Q_{0.75} + Q_{0.25})/2 = Q_{0.5} = 0.89452$.

One way to overcome the ambiguity regarding skewness measures is to examine a family of location measures rather than simply considering two fixed ones. The median and mean are two special members of the family of Huber M-estimates, or more properly M-functionals, see e.g. [12, 15]. For a given tuning constant $c > 0$, a Huber M-functional of location is defined implicitly to be the solution $\mu_c(F)$ to the "M-estimating" equation

$$E_F[\psi_c(X - \mu_c)] = 0, \quad \text{where} \quad \psi_c(r) = \max[-c, \min(r, c)]. \tag{3.2.5}$$

Equivalently, the M-estimating equation can be expressed as

$$\mu_c = \frac{E_F[u_c(X - \mu_c)X]}{E_F[u_c(X - \mu_c)]}, \tag{3.2.6}$$

where the weight function $u_c(r) = \psi_c(r)/r$. This expression gives the intuitive interpretation of μ_c as an adaptively weighted mean.

It is shown in [41] that the set of all Huber functions, $\Psi_H = \{\psi_c, c > 0\}$, constitutes a *semi-separating* class. We say that a class of functions Ψ forms a *semi-separating* class if whenever the equality

$$\int_0^{+\infty} \psi_c(r) dF(r) = \int_0^{+\infty} \psi_c(r) dG(r), \tag{3.2.7}$$

holds for all $\psi_c \in \Psi$, where F and G are two distribution functions, then $F(r) = G(r)$ for all $r \geq 0$. It then follows that if all Huber location functionals $\mu_c(F)$ are the same for any $c > 0$, then F is symmetric about $\mu = \mu_c(F)$, i.e. $F(x - \mu) = 1 - F(-(x - \mu))$. We will call a class of locational functionals possessing this property to be a *symmetry identifying class.*

The property of being a semi-separating class can be shown to extend to many other classes of ψ-functions. For example, the class

$$\psi_c(x) = xe^{-|x/c|^\alpha} \quad \text{for} \quad c > 0, \tag{3.2.8}$$

which corresponds to the Gaussian weight functions $u_c(x) = e^{-|x/c|^\alpha}$, and the class

$$\psi_c(x) = x[1-(x/c)^2]_+^2 \quad \text{for} \quad c > 0, \tag{3.2.9}$$

which yields Tukey's biweight functions $u_c(x) = [1-(x/c)^2]_+^2$, both form semi-separating classes. Unlike Huber's ψ, however, these two classes of ψ-functions are redescenders and hence do not necessarily admit unique solutions. However, suppose one considers the one-step M-functionals of location

$$\mu_{1,c} = \mu_o + \frac{E_F[\psi_c\,(X-\mu_o)]}{E_F[\psi_c'\,(X-\mu_o)} \tag{3.2.10}$$

or the W-functionals of location

$$\omega_c = \frac{E_F[u_c(X-\mu_o)X]}{E_F[u_c(X-\mu_o)]}, \tag{3.2.11}$$

based on these ψ-functions, where μ_o is some fixed location functional, e.g. the median. It follows that if $\mu_{1,c}$ or ω_c is constant over $c > 0$, when $\Psi = \{\psi_c, c > 0\}$ is a semi-separating class, then F is symmetric about μ_o.

Classes of location measures which uniquely imply symmetry can also be found outside of the M-functionals and the W-functionals of location. The symmetry of F is also implied whenever the α-trimmed mean defined as,

$$T_\alpha(F) = \frac{1}{1-2\alpha}\int_\alpha^{1-\alpha} F^{-1}(s)ds \tag{3.2.12}$$

remains constant for all $\alpha \in (0, 1/2)$, as well as by the constancy of the α-Winsorized mean

$$W_\alpha(F) = \int_\alpha^{1-\alpha} F^{-1}(s)ds + \alpha\left(F^{-1}(\alpha) + F^{-1}(1-\alpha)\right) \tag{3.2.13}$$

over $\alpha \in (0, 1/2)$, and by the constancy of $(Q_{1-p} + Q_p)/2$ over $p \in (0, 1/2)$, which appears in the generalized Bowley's coefficients. The latter case is equivalent to the generalized Bowley's coefficient being equal to zero for all $p \in (0, 1/2)$.

3.3 Symmetry Identifying Transformations

A symmetry identifying class of location functionals indexed by a real value can be viewed as a transformation of a distribution, which is constant if and only if the underlying distribution is symmetric. For a sample of course, there will be some statistical variability, and no sample transform will be exactly constant. By considering the values of such a symmetry identifying class of location estimates, though, one can construct consistent tests of asymmetry by considering the difference between the maximum and minimum values of the estimates or by some other measure of non-constancy. The statistical theory for such an approach is fairly challenging, and we leave that to future research. In any event, plotting a family of such location statistics provides a transformation of the data for which it may be visually easier to notice a departure from constancy than it would be to notice asymmetry in a plot of a density estimate.

Furthermore, if a distribution is not symmetric, then summarizing its central tendency by a single value may not be sufficiently descriptive. So, aside from using a symmetry identifying transformation for detecting asymmetry, it also serves as a graphical descriptive summary of central tendency.

For illustrative purposes, we consider the computationally simple one step M-estimates, W-estimates and α-trimmed means. For a sample $x_1, \cdots, x_n$, the one step M-estimates consider here are given by

$$\hat{\mu}_{c,1} = \hat{m} + s\frac{\sum_1^n \psi_c\left(\frac{x_i-\hat{m}}{s}\right)}{\sum_1^n \psi_c'\left(\frac{x_i-\hat{m}}{s}\right)}, \tag{3.3.1}$$

where $\hat{m}$ and s denote respectively the sample median and MAD, i.e. the median absolute deviation about the median. The W-estimates considered here are defined by

$$\hat{\omega}_c = \frac{\sum_1^n u_c\left(\frac{x_i-\hat{m}}{s}\right)x_i}{\sum_1^n u_c\left(\frac{x_i-\hat{m}}{s}\right)}. \tag{3.3.2}$$

The sample versions of trimmed means are given by

$$\hat{T}(\alpha) = \frac{1}{n-2m}\sum_{i=m+1}^{n-m} x_{(i)}, \quad \text{where} \quad m = [(n-1)\alpha]. \tag{3.3.3}$$

The given definitions for the one step M-estimates (3.3.1) and the W-estimates (3.3.2) are not simply the sample versions of (3.2.10) and (3.2.11) respectively. As defined in the previous section, the M-functionals, the one-step M-functionals and the W-functionals are location measures in the sense that they are equivariant under translation and reflection. That is, if $T(F_X)$ represents a location measure for the random variable $X \sim F_X$, then for $Y = X+a$, with $a \in \Re$, $T(F_Y) = T(F_X)+a$, and for $Y = -X$, $T(F_Y) = -T(F_X)$. However, unlike the α-trimmed means and the α-Winsorized mean, they are not necessarily scale equivariant, meaning that if $Y = b\ X$, with $b \in \Re$, then $T(F_Y)$ is not necessarily equal to $b\ T(F_X)$. So, to make them scale equivariant, one usually introduces a scaling term such as the *MAD*. Note, however, such a scaling term is not necessary when considering a class of location estimates over the range $c > 0$, since the scaling term s and the tuning constant c are confounded. In other words, any of these classes of location estimates is scale invariant.

Figure 3.2 gives the plots of the one-step Huber M-estimates $\hat{\mu}_{1,c}$, the W-estimates $\hat{\omega}_c$ based on the Gaussian weights (3.2.8) and the bisquare weights (3.2.9) respectively as functions of the tuning constant $c > 0$, as well as the α-trimmed mean as a function of α, for a random sample of size $n = 10000$ from the normal mixture given in (3.2.1). In each of these plots, the location estimates vary from the median, on the left, and the mean, on the right. As noted in section (3.2), the population mean and median are equal for this example, but the distribution is not symmetric. Hence, we note the curves are not constant.

3.4 Asymptotic Distributions of New Test Statistics

When the observations X_i are i.i.d according to P, the difference between a location parameter(functional) $T(P)$ and its consistent estimate $\widehat{T}$ derived from the empirical distribution can be approximated adequately, under certain regularity conditions [15], by the influence function as

$$S_n = \sqrt{n}\left(\widehat{T} - T\right) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n} \mathrm{IF}(x_i; T, P) + O_p(n^{-1/2}) \tag{3.4.1}$$

hence two location estimates are asymptotically joint normal

$$\sqrt{n}\begin{pmatrix}\widehat{T}_1 - T_1 \\ \widehat{T}_2 - T_2\end{pmatrix} \longrightarrow N\left[\begin{pmatrix}0\\0\end{pmatrix}, \begin{pmatrix}\sigma_{11} & \sigma_{12}\\ \sigma_{12} & \sigma_{22}\end{pmatrix}\right],$$

where $\sigma_{ij} = E_P\Big(\mathrm{IF}(X;T_i,P)\mathrm{IF}(X;T_j,P)\Big)$. Their difference follows asymptotic normal as well

$$\sqrt{n}\left[\left(\widehat{T}_1 - \widehat{T}_2\right) - \left(T_1 - T_2\right)\right] \longrightarrow N\left[0, E_P\left(\mathrm{IF}(X;T_i,P) - \mathrm{IF}(X;T_j,P)\right)^2\right]$$

We notice that the difference of the two influence functions is actually the influence function of $T_1 - T_2$.

When P is symmetric, location parameters are constant, thus $\sqrt{n}\widehat{\mu}_n/\widehat{\sigma} \to N(0,1)$, where $\widehat{\mu}_n = \widehat{T}_1 - \widehat{T}_2$, and $\widehat{\sigma}^2$ is a consistent estimator of the asymptotic variance. As a test statistic, $\widehat{\mu}_n$ preserves its normality when the underlying distribution changes from symmetry to a mixture or asymmetry, but it will have a non-zero asymptotic mean in most cases.

3.4.1 On Mixture

Let F and G be two **symmetric** distributions with pdf f and g respectively. Let $P_{1,n}$ denote the joint distribution of $x_1, \cdots, x_n$ when they are i.i.d. according to

$$(1-\epsilon)F(x) + \epsilon\, G(x - \theta_n), \text{ with } \theta_n = \theta/\sqrt{n}$$

Let $P_{0,n}$ denote their joint distribution when they are i.i.d. $P \sim (1-\epsilon)F + \epsilon\, G$.

Theorem 3.4.1. *$P_{1,n}$ are contiguous to $P_{0,n}$.*

Corollary 3.4.2. *Under $P_{1,n}$, $S_n \xrightarrow{d} N\left(\epsilon\theta u, \upsilon^2\right)$, where $u = -\int_{-\infty}^{\infty}[\mathrm{IF}(x;T,P)]g'(x)dx$, and $\upsilon^2 = E_P[\mathrm{IF}(x;T,P)]^2$. In particular, when $g(x) = \frac{1}{b}f\left(\frac{x}{b}\right)$ having non-zero $g'(x)$,*

1. *For $T_1 = \mu - Q_{0.5}$, $u = \frac{(1-\epsilon)(b-1)}{(1-\epsilon)b+\epsilon}$, $\upsilon^2 = (1-\epsilon+\epsilon b^2)E_F X^2 - \frac{1-\epsilon+\epsilon b}{\left(1-\epsilon+\frac{\epsilon}{b}\right)f(0)}E_F|X| + \frac{1}{4\left[\left(1-\epsilon+\frac{\epsilon}{b}\right)f(0)\right]^2}$*
2. *For $T_2 = \frac{E(X-\mu)^3}{\sigma^3}$, $u = \frac{3(1-\epsilon)(b^2-1)}{(1-\epsilon+\epsilon b^2)^{3/2}}\left(E_F X^2\right)^{-1/2}$, $\upsilon^2 = \frac{(1-\epsilon+\epsilon b^6)}{(1-\epsilon+\epsilon b^2)^3}\frac{E_F X^6}{(E_F X^2)^3} - 6\frac{(1-\epsilon+\epsilon b^4)}{(1-\epsilon+\epsilon b^2)^2}\frac{E_F X^4}{(E_F X^2)^2} + 9$*
3. *For $T_3 = \frac{Q_{(1-p)}+Q_p-2Q_{0.5}}{Q_{(1-p)}-Q_p}$, $\upsilon^2 = \left[\frac{1}{\left[(1-\epsilon)f(q)+\frac{\epsilon}{b}f\left(\frac{q}{b}\right)\right]^2} - \frac{2}{\left[(1-\epsilon)f(q)+\frac{\epsilon}{b}f\left(\frac{q}{b}\right)\right]\left(1-\epsilon+\frac{\epsilon}{b}\right)f(0)}\right]\frac{p}{2q^2} + \frac{1}{4q^2\left[\left(1-\epsilon+\frac{\epsilon}{b}\right)f(0)\right]^2}$,*
   $u = \frac{\frac{1}{b}f\left(\frac{q}{b}\right)}{\left[(1-\epsilon)f(q)+\frac{\epsilon}{b}f\left(\frac{q}{b}\right)\right]q} - \frac{\frac{1}{b}}{\left[(1-\epsilon)+\frac{\epsilon}{b}\right]q}$, where $(1-\epsilon)F(q) + \epsilon F\left(\frac{q}{b}\right) = 1-p$.

Corollary 3.4.2 implies that a location estimator would have an asymptotic mean of 0 if the underlying distribution is a mixture of a symmetry with a point mass or with a Uniform. To construct a meaningful test statistic, Corollary 3.4.3 through Corollary 3.4.6 assume $g'(x) \neq 0$.

Corollary 3.4.3. *If $T(P)$ is an M-estimator defined implicitly by $E_P(\psi(x-T(P)) = 0$, then under $P_{1,n}$, $S_n \xrightarrow{d} N\left(\frac{\epsilon\theta}{(1-\epsilon)u+\epsilon}, \frac{E_P[\psi(x)]^2}{\{E_P[\psi'(x)]\}^2}\right)$, where $u = \frac{E_F[\psi'(x)]}{E_G[\psi'(x)]}$. In particular, $u = \frac{E_F[\psi'(x)]}{E_F[\psi'(bx)]}$ if $G(x) = F(x/b)$.*

Proof. Since M-estimate has influence function $\frac{\psi(x)}{E_P[\psi'(x)]}$, where $E_P[\psi'(x)] = (1-\epsilon)E_F[\psi'(x)] + \epsilon E_G[\psi'(x)]$, the proof is a direct application of Corollary 3.4.2. A discussion on the difference of two Huber-Type M-estimators on special cases is in Appendix B ☐

Corollary 3.4.4. *If $T(P)$ is a W-estimator defined implicitly as $T = \frac{E_P[w(x-T)x]}{E_P w(x-T)}$, then $\widehat{T}$ has the same asymptotic distribution under $P_{1,n}$ as an M-estimator defined by $\psi(x) = w(x)x$.*

Proof. W-estimators defined by $w(x)$ possess the same influence function as M-estimators defined by $\psi(x) = w(x)x$ (page 116 of [12]), the conclusion follows Corollary 3.4.3. ☐

Corollary 3.4.5. *If $T(P)$ is a one-step M-estimator defined as*

$$T = T_0 + S_0 \frac{E_P\left[\psi\left(\frac{x-T_0}{S_0}\right)\right]}{E_P\left[\psi'\left(\frac{x-T_0}{S_0}\right)\right]},$$

where ψ is odd, T_0 and S_0 are preliminary estimates of location and scale, then under $P_{1,n}$,

$$S_n \xrightarrow{d} N\left(\frac{\epsilon\theta}{(1-\epsilon)u+\epsilon}, S_0^2 \frac{E_P\left[\psi^2\left(\frac{x-T_0}{S_0}\right)\right]}{\left\{E_P\left[\psi'\left(\frac{x-T_0}{S_0}\right)\right]\right\}^2}\right).$$

where $u = \frac{E_F\left[\psi'\left(\frac{x-T_0}{S_0}\right)\right]}{E_G\left[\psi'\left(\frac{x-T_0}{S_0}\right)\right]}$. In particular, $u = \frac{E_F\left[\psi'\left(\frac{x-T_0}{S_0}\right)\right]}{E_F\left[\psi'\left(\frac{bx-T_0}{S_0}\right)\right]}$ if $G(x) = F\left(\frac{x}{b}\right)$.

Proof. The results follow the fact that when ψ is odd and P is symmetric, the one-step M-estimator has influence function $\mathrm{IF}(x; T, P) = \frac{S_0\psi\left(\frac{x-T_0}{S_0}\right)}{E_P\left[\psi'\left(\frac{x-T_0}{S_0}\right)\right]}$ (page 141 of [15]). ☐

Corollary 3.4.6. *If* $T(P) = \frac{1}{1-2\alpha}\int_{\alpha}^{1-\alpha} P^{-1}(s)ds$ *is the α-trimmed mean, then under $P_{1,n}$, S_n is asymptotically normal with mean equals to* $\epsilon\theta\frac{2G(q)-1}{1-2\alpha}$ *and variance equals to*

$$\frac{2(1-\epsilon)\left(q^2F(-q)+E_FX^2 1_{[0<X\leq q]}\right)+2\epsilon\left(q^2G(-q)+E_GX^2 1_{[0<X\leq q]}\right)}{(1-2\alpha)^2}$$

where $P(q) = (1-\epsilon)F(q) + \epsilon G(q) = 1-\alpha$

Proof. The symmetry of P implies $P^{-1}(\alpha) = -P^{-1}(1-\alpha)$ and a simplified influence function of $T(P)$ (page 58 of [15]),

$$\mathrm{IF}(x;T,P) = \frac{1}{1-2\alpha}\max\{P^{-1}(\alpha), \min[x, P^{-1}(1-\alpha)]\}$$

$\square$

3.4.2 On Asymmetry

Gupta [11] has considered a class of distribution functions $f(x)\mathrm{I}_{[x\leq 0]} + \frac{1}{\tau}f\left(\frac{x}{\tau}\right)\mathrm{I}_{[x>0]}$, when testing the hypothesis of symmetry

$$H_0 : \tau = 1 \text{ vs } H_1 : \tau > 1$$

Mira [25] introduced an equivalent sequence of alternatives

$$H_{1,n} : \tau = 1 + \frac{\eta}{\sqrt{n}}$$

Let $F_{1,n}$ and $F_{0,n}$ denote the joint distribution of $x_1, \cdots, x_n$ when they are i.i.d. from $H_{1,n}$ and from $F \sim H_0$ respectively

Theorem 3.4.7. *$F_{1,n}$ are contiguous to $F_{0,n}$.*

Corollary 3.4.8. *If* $\frac{d^2f(x)}{dx^2} \neq 0$, $S_n \xrightarrow{d} N\left(\eta u, \upsilon^2\right)$ *under* $F_{1,n}$, *with* $\upsilon^2 = E_F\left[\mathrm{IF}\left(x;T,F\right)\right]^2$, *and* $u = -E_F\left[\mathrm{IF}\left(x;T,F\right)\left(1+\frac{xf'(x)}{f(x)}\right)I_{[x>0]}\right]$. *In particular,*

1. *For* $T_1 = \mu - Q_{0.5}$, $u = \frac{1}{2}E_F|x|$, $\upsilon^2 = E_F\left(x^2\right) - \frac{E_F|x|}{f(0)} + \frac{1}{4f^2(0)}$
2. *For* $T_2 = E(x-\mu)^3/\sigma^3$, $u = \frac{3}{2}\left(\frac{E_F|x|^3 - Ex^2E_F|x|}{(Ex^2)^{3/2}}\right)$, $\upsilon^2 = \frac{E_Fx^6}{(Ex^2)^3} - 6\frac{Ex^4}{(Ex^2)^2} + 9$
3. *For* $T_3 = \frac{Q_{1-p}+Q_p-2Q_{0.5}}{Q_{1-p}-Q_p}$, $u = \frac{1}{2}$, $\upsilon^2 = \frac{1}{4q^2}\left(\frac{2p}{f^2(q)} - \frac{4p}{f(q)f(0)} + \frac{1}{f^2(0)}\right)$, *where* $q = F^{-1}(1-p)$

3.4.3 On Skew-Symmetry

The exact definition of Skew-Symmetric distribution has yet to be unified. Let F and G be two **symmetric** distributions with pdf f and g respectively, one class of Skew-Symmetric distribution that was discussed by Azzalini [1] has its pdf ψ of this form $\psi = 2f(x)G(\alpha x)$. The test of skewness($\alpha \neq 0$) versus symmetry ($\alpha = 0$) again can be examined locally by a sequence of hypotheses. Let $\Psi_{1,n}$ denote the joint distribution of $x_1, \cdots, x_n$ when they are i.i.d. according to

$$2f(x)G(\alpha_n x), \text{ with } \alpha_n = 1/\sqrt{n}$$

Let $\Psi_{0,n}$ denote their joint distribution when they are i.i.d. from F

Theorem 3.4.9. *$\Psi_{1,n}$ are contiguous to $\Psi_{0,n}$, provided $g'(0) = 0$.*

Corollary 3.4.10. *Under $H_{1,n}$, $S_n \xrightarrow{d} N\Big(2g(0)u, \upsilon^2\Big)$, where $u = E_F\Big(x\mathrm{IF}(x;T,F)\Big)$, and $\upsilon^2 = E_F[\mathrm{IF}(x;T,P)]^2$. In particular,*

1. *For $T_1 = \mu - Q_{0.5}$, $u = Ex^2 - \frac{E|x|}{2f(0)}$, $\upsilon^2 = Ex^2 - \frac{E|x|}{f(0)} + \frac{1}{4f^2(0)}$*
2. *For $T_2 = \frac{E(x-\mu)^3}{\sigma^3}$, $u = \frac{Ex^4 - 3\left(Ex^2\right)^2}{(Ex^2)^{3/2}}$, $\upsilon^2 = \frac{Ex^6}{(Ex^2)^3} - 6\frac{Ex^4}{(Ex^2)^2} + 9$*
3. *For $T_3 = \frac{Q_{1-p}+Q_p-2Q_{0.5}}{Q_{1-p}-Q_p}$, $u = \frac{E\left(|x|I_{[|x|>q]}\right)}{2qf(q)} - \frac{E|x|}{2qf(0)}$, $\upsilon^2 = \frac{1}{4q^2}\left(\frac{2p}{f^2(q)} - \frac{4p}{f(q)f(0)} + \frac{1}{f^2(0)}\right)$, where $q = F^{-1}(1-p)$*

3.5 Power and Efficacy Comparing With Classical Skewness Tests

3.5.1 On Mixture

The performance of a statistic on the hypothesis testing of mixtures

$$H_0 : F(x) \quad \text{vs} \quad H_1 : (1-\epsilon)F(x) + \epsilon F(x-\theta)$$

can be evaluated by its efficacy on the sequence of equivalent hypotheses

$$H_0 : (1-\epsilon)F + \epsilon G \quad \text{vs} \quad H_{1,n} : (1-\epsilon)F(x) + \epsilon G\left(x - \theta_n\right), \text{ where } \theta_n = \frac{\theta}{\sqrt{n}}$$

with $g(x) = \frac{1}{b} f\left(\frac{x}{b}\right)$, and $b = 1$. Under this setting, all the test statistics discussed in Corollary 3.4.2 would have zero Pitman's efficacy [19, 28], yet we can still compare them by looking at their asymptotic relative efficiency as b approaching 1. Letting $b \to 1$, the ARE of $(\mu - M_{0.5})$ with respect to $E(x-\mu)^3/\sigma^3$ is then obtained as

$$\frac{E_F X^6 - 6\left(E_F X^4\right)\left(E_F X^2\right) + 9\left(E_F X^2\right)^3}{36\left(E_F X^2\right)^2 \left(E_F X^2 - \frac{E_F|X|}{f(0)} + \frac{1}{4f^2(0)}\right)},$$

while the ARE of $(\mu - Q_{0.5})$ with respect to $\left(Q_{(1-p)} + Q_p - 2Q_{0.5}\right) / \left(Q_{(1-p)} - Q_p\right)$ is

$$\left(\frac{f(q)}{f'(q)}\right)^2 \frac{\left(\frac{1}{f^2(q)} - \frac{2}{f(q)f(0)}\right)\frac{p}{2q^2} + \frac{1}{4q^2 f^2(0)}}{E_F X^2 - \frac{E_F|X|}{f(0)} + \frac{1}{4f^2(0)}}, \qquad \text{where } q = F^{-1}(1-p).$$

The values on the ARE's on different symmetric distributions are presented in Table 3.1, and Table 3.2.

Within the family of Huber M-estimates

$$\{\mu_c : E\psi_c(x - \mu_c) = 0 \text{ for } \psi_c(r) = \max[-c, \min(r, c)],\}$$

where the limiting cases $c \to \infty$ and $c \to 0$ corresponding to mean μ and median $Q_{0.5}$ respectively, we are particularly interested in the relationship between traditional skewness test $\mu - Q_{0.5}$ and an arbitrary test statistic $\mu_{c_1} - \mu_{c_2}$ in terms of asymptotic efficiency.

Figure 3.3 gives the contour plot of ARE $(\mu_{c_1} - \mu_{c_2}, \mu - Q_{0.5})$ as functions of (c_1, c_2) when F are Normal, , Student T_3, Laplace, and Triangular respectively. Except under Laplace distribution, the $\mu - Q_{0.5}$ is not the most powerful test within this Huber family.

3.5.2 On Asymmetry

In a similar pattern, we compare the efficiency of different statistics on testing asymmetry

$$H_0 : f(x) \quad \text{vs} \quad H_{1,n} : f(x)\mathrm{I}_{[x\le 0]} + \frac{1}{\tau} f\left(\frac{x}{\tau}\right), \tau > 0$$

by comparing their efficacies at a sequence of equivalent local alternatives

$$H_0 : f(x) \quad \text{vs} \quad H_{1,n} : f(x)\mathrm{I}_{[x\le 0]} + \frac{1}{\tau_n} f\left(\frac{x}{\tau_n}\right), \text{ where } \tau_n = 1 + n^{-1/2}$$

whose analytical solutions are available by Corollary 3.4.8 and listed in Table 3.3

3.5.3 On Skew-Symmetry

Again, the powers of statistics at testing skewness

$$H_0 : f(x) \quad \text{vs} \quad H_1 : 2f(x)G(\alpha x)$$

are compared by their Pitman's efficacies [19, 28] at testing a sequence of local hypotheses

$$H_0 : f(x) \quad \text{vs} \quad H_{1,n} : 2f(x)G\left(\frac{x}{\sqrt{n}}\right)$$

When $f(x) = \phi(x)$ is the Standard Normal, the three statistics of interests T_1, T_2, and T_3 are all having zero asymptotic mean by Corollary 3.4.10; to achieve comparison, a mixture of Normal is replacing the standard one

$$f(x) = (1-\epsilon)\phi(x) + \frac{\epsilon}{b}\phi\left(\frac{x}{b}\right)$$

Letting $b \to 1$, we obtain AREs between the three test statistics. The AREs under other suitable distributions, particularly for the class of $G = F$ with $f'(0) = 0$, are calculated directly from statistics' asymptotic means and variances given in Corollary 3.4.10, and are listed in Table 3.4

When testing Skewness, the difference of mean and median (T_1) are not the best test in terms of relative efficiency compare to the difference of other members of Huber-family (Figure 3.4), some of which are highly efficient that they beat the "Measure of Skewness"as well. Note that the ARE of $\mu_{c1}-\mu_{c2}$ versus $\mu - Q_{0.5}$ at testing Skew-Normal is the same as at testing Mixture of Normal (Figure 3.3).

3.6 Appendix A: Separating Class

Claim: Suppose that the solution $T(F)$ to the equation $E_F\psi_c(x - T(F)) = 0$, where $\psi_c(x) = xe^{-|x/c|^\alpha}$, is constant in $c > 0$, then F is symmetric.

Proof. WLOG, let $T(F) = 0$.

$$E_F\psi(x) = \int_{-\infty}^{+\infty} xe^{-|x/c|^\alpha} f(x)dx = \int_0^{+\infty} xe^{-(x/c)^\alpha} \left(f(x) - f(-x)\right) dx = 0$$

Let $k = |\frac{1}{c}|^\alpha$, $y = e^{-x^\alpha}$, then $x = (-\ln y)^{1/\alpha}$, $dx = \frac{(-\ln y)^{1/\alpha-1}}{-\alpha y}$, and for all $k > 0$,

$$\int_0^1 y^k h(y) = 0, \text{ where } h(y) = \frac{(-\ln y)^{2/\alpha-1}}{y} \left(f[(-\ln y)^{1/\alpha}] - f[-(-\ln y)^{1/\alpha}]\right)$$

This implies $h(y) = 0$ for all $y \in (0, 1)$, it follows $f(x) = f(-x)$. □

It is clear from this proof that if ψ_c constitutes a separating class, then the fact $T(F)$, for which $E_F\psi_c\,(x - T(F)) = 0$, remains constant in c will imply the symmetry of F, i.e. ψ_c being a separating class is sufficient to conclude the symmetry of F, but this is not a necessary condition. For example, let ψ_c be Huber functions

$$\psi_c(x)/x = \begin{cases} 1 & |x| \le c \\ c/|x| & |x| > c \end{cases} \tag{3.6.1}$$

Then, $E_F\psi_(x) = 0$ for all $c > 0$ implies the symmetry of F, but ψ_c is not a separating class. For, let F be asymmetric distribution with pdf f, let G be a continuous function with derivative $g(x) = dG = \frac{f(x)-f(-x)}{2}$, clearly $G(x) = \frac{F(x)-(1-F(-x))}{2} \neq F(x)$, but the symmetry of $dF - dG = \frac{f(x)+f(-x)}{2}$ guarantees $\int \psi_c dF = \int \psi_c dG$ for all $c > 0$.

Claim: If $E_F\psi_c(x) = 0$ for all $c > 0$, where $\psi(x) = x[1 - (x/c)^2]^2 I(|x| \le c)$, then F is symmetric.

Proof.

$$E_F\psi_c(x) = \int_{-c}^{c} x\left[1 - \left(\frac{x}{c}\right)^2\right]^2 f(x)dx = \int_0^c x\left[1 - \left(\frac{x}{c}\right)^2\right]^2 \left(f(x) - f(-x)\right) dx = 0$$

Letting $k = c^2,\ y = x^2$ gives

$$E_F\psi_k(x) = \int_0^k \left(1 - \frac{y}{k}\right)^2 G(y)dy = 0, \quad \text{where } G(y) = f(\sqrt{y}) - f(-\sqrt{y})$$

Taking derivatives,

$$0 = \frac{\partial^2 E_F \psi_k(x)}{\partial k^2} \propto \int_0^k y^2 G(y) dy \equiv H(k)$$

then, $H^{'}(k) = k^2 G(k) = 0, \quad \text{for all } k > 0$. It follows $f(x) = f(-x)$ □

Claim: If the $\alpha-$trimmed mean $T(F) = \frac{1}{1-2\alpha}\int_\alpha^{1-\alpha} F^{-1}(s)ds$ is a constant for all $\alpha \in \left(0, \frac{1}{2}\right)$, then F is symmetric. Similar argument holds for $\alpha-$Winsorized mean

$$W(F) = \int_\alpha^{1-\alpha} F^{-1}(s)ds + \alpha F^{-1}(\alpha) + \alpha F^{-1}(1-\alpha)$$

Proof. WLOG, let $T(F) = 0$.

$$0 = \frac{\partial T(F)}{\partial \alpha} = \frac{2}{1-2\alpha} T(F) - F^{-1}(1-\alpha) - F^{-1}(\alpha)$$

which gives $F^{-1}(1-\alpha) = -F^{-1}(\alpha)$, i.e. $F(-x) = 1 - F(x)$. □

3.7 Appendix B: Proofs of Theorems and Corollaries

Proof of Theorem 3.4.1 and Corollary 3.4.2

Proof. The log-likelihood is given by $L_n = \log(P_{1,n}/P_{0,n}) = \sum_{i=1}^n h(x_i)$, where

$$\begin{aligned} h(x) &= \log\left(1 + \epsilon\, \frac{g(x-\theta_n) - g(x)}{dP}\right) \\ &= \epsilon\left(\frac{g(x-\theta_n) - g(x)}{dP}\right) - \frac{\epsilon^2}{2}\left(\frac{g(x-\theta_n) - g(x)}{dP}\right)^2 + O(n^{-3/2}) \end{aligned}$$

Now $\frac{g(x-\theta_n)-g(x)}{dP} = -\frac{g'(x)}{dP}\frac{\theta}{\sqrt{n}} + \frac{1}{2}\frac{g''(x)}{dP}\frac{\theta^2}{n} + O(n^{-3/2})$, also $E_P\left(\frac{g'(x)}{dP}\right) = E_G\left(g'/g\right) = 0$, $E_P\left(\frac{g''(x)}{dP}\right) = E_G\left(g''/g\right) = 0$, so under $P_{0,n}$,

$$\begin{aligned} L_n &= \epsilon\theta\left\{\frac{1}{\sqrt{n}}\sum_{i=1}^n \frac{-g'(x_i)}{dP}\right\} + \frac{\epsilon\theta^2}{2}\left\{\frac{1}{n}\sum_{i=1}^n \frac{g''(x_i)}{dP}\right\} - \frac{\epsilon^2\theta^2}{2}\left\{\frac{1}{n}\sum_{i=1}^n\left(\frac{g'(x_i)}{dP}\right)^2\right\} + O(n^{-1/2}) \\ &= \epsilon\theta\left\{\frac{1}{\sqrt{n}}\sum_{i=1}^n \frac{-g'(x_i)}{dP}\right\} - \frac{\epsilon^2\theta^2\tau^2}{2} + O(n^{-1/2}), \text{ where } \tau^2 = E_P\left(\frac{g'(x)}{dP}\right)^2 \\ &\xrightarrow{d} \epsilon\theta\, N(0,\tau^2) - \frac{\epsilon^2\theta^2\tau^2}{2} \sim N\left(-\epsilon^2\theta^2\tau^2/2,\ \epsilon^2\theta^2\tau^2\right). \end{aligned}$$

By LeCam first Lemma [29], $P_{1,n}$ are contiguous to $P_{0,n}$. The limiting distribution of S_n under $P_{1,n}$ follows LeCam third Lemma, $\sqrt{n}S_n \xrightarrow{d} N(u,\, \upsilon^2)$, where by the same

Lemma, $v^2 = E_P[\mathrm{IF}(x;T,P)]^2$, and

$$u = E_P(S_n L_n) = \epsilon\theta\, E_P\left[\frac{-g'(x)}{dP}\mathrm{IF}(x;T,P)\right]$$
$$= -\epsilon\theta\int_{-\infty}^{+\infty} g'(x)\mathrm{IF}(x;T,P)dx = -\epsilon\theta\, E_G\left[\frac{\partial \mathrm{IF}(x;T,P)}{\partial x}\right]$$

if $g'(x) \neq 0$. The parameters of particular statistics when $g(x) = \frac{1}{b}f\left(\frac{x}{b}\right)$ are obtained by noting that $T_1 = \mu - Q_{0.5}$ has influence function $x - \frac{\mathrm{sign}(x)/2}{(1-\epsilon)f(0)+\epsilon g(0)}$, and $T_2 = \frac{E(x-\mu)^3}{\sigma^3}$ has influence function $\frac{x^3-3xE_P(x^2)}{[E_P(x^2)]^{3/2}}$, while $\mathrm{IF}(x;T_3,P)$ is the odd step-function with values for $x > 0$ given by $\frac{1}{2q}\left(\frac{\mathrm{I}_{[x>q]}}{(1-\epsilon)f(q)+\epsilon g(q)} - \frac{1}{(1-\epsilon)f(0)+\epsilon g(0)}\right)$ (Groeneveld [10]), where q is the solution to $(1-\epsilon)F(q) + \epsilon G(q) = 1 - p$ □

Proof and Discussion of Corollary 3.4.3

Proof. The influence function of an M-estimate implicitly defined by a $\psi(\cdot)$ function is of the form $\frac{\psi(x)}{E_P[\psi'(x)]}$, the proof is a direct application of Corollary 3.4.2. In the case when $T(P)$ is the difference of two Huber-Type M-estimators $T_i(P)$ that are associated with functions $\psi_i(x) = \max[-k_i, \min(x,k_i)]$, $k_1 < k_2$, and suppose $F(x/\sigma) = G(x)$, then $\sqrt{n}\left(T_1(\widehat{P}_n) - T_2(\widehat{P}_n)\right)$ converges to a normal distribution $N(\mu,\tau^2)$ under Q_n with

$$\mu = \epsilon\,\theta\left\{\frac{F(k_1/\sigma)-1/2}{(1-\epsilon)\,F(k_1)+\epsilon\,F(k_1/\sigma)-1/2} - \frac{F(k_2/\sigma)-1/2}{(1-\epsilon)\,F(k_2)+\epsilon\,F(k_2/\sigma)-1/2}\right\}$$
$$\tau^2 = \frac{(1-\epsilon)\,\left\{k_1^2F(-k_1)+I_1(x,k_1)\right\}+\epsilon\,\left\{k_1^2F(-k_1/\sigma)+I_1(\sigma x,k_1)\right\}}{2\left\{(1-\epsilon)\,F(k_1)+\epsilon\,F(k_1/\sigma)-1/2\right\}^2}$$
$$+\frac{(1-\epsilon)\,\left\{k_2^2F(-k_2)+I_1(x,k_2)\right\}+\epsilon\,\left\{k_2^2F(-k_2/\sigma)+I_1(\sigma x,k_2)\right\}}{2\left\{(1-\epsilon)\,F(k_2)+\epsilon\,F(k_2/\sigma)-1/2\right\}^2}$$
$$-\frac{(1-\epsilon)\,\left\{k_1k_2F(-k_2)+k_1I_2(x,k_1,k_2)+I_1(x,k_1)\right\}+\epsilon\,\left\{k_1k_2F(-k_2/\sigma)+k_1I_2(\sigma x,k_1,k_2)+I_1(\sigma x,k_1)\right\}}{\left\{(1-\epsilon)\,F(k_1)+\epsilon\,F(k_1/\sigma)-1/2\right\}\left\{(1-\epsilon)\,F(k_2)+\epsilon\,F(k_2/\sigma)-1/2\right\}}$$

where $I_1(x,k) = E_F\left(x^2\cdot 1_{[0<x<k]}\right)$, and $I_2(x,k_1,k_2) = E_F\left(x\cdot 1_{[k_1<x<k_2]}\right)$ □

Proof of Theorem 3.4.7 and Corollary 3.4.8

Proof. Let $q_x(\tau) = \left(\frac{\frac{1}{\tau}f\left(\frac{x}{\tau}\right)-f(x)}{f(x)}\right)\mathrm{I}_{[x>0]}$. Then

$$q_x'(1) = -\left(1+\frac{xf'(x)}{f(x)}\right)\mathrm{I}_{[x>0]}\,,\ \text{ and }\ q_x''(1) = \left(2+4\frac{xf'(x)}{f(x)}+\frac{x^2f''(x)}{f(x)}\right)\mathrm{I}_{[x>0]}$$

Re-write the log-likelihood by its Taylor expansion,

$$\begin{aligned}
L_n = \log(F_{1,n}/F_{0,n}) &= \sum_{i=1}^{n} \log\left(\frac{f(x_i)\mathrm{I}_{[x_i\le 0]} + \frac{1}{\tau} f\left(\frac{x_i}{\tau}\right)\mathrm{I}_{[x_i>0]}}{f(x_i)}\right) \\
&= \sum_{i=1}^{n} \log\left(1 + \left[\frac{\frac{1}{\tau} f\left(\frac{x_i}{\tau}\right) - f(x_i)}{f(x_i)}\right]\mathrm{I}_{[x_i>0]}\right) \\
&= \sum_{i=1}^{n}\left(q_{x_i}(\tau) - \frac{1}{2} q_{x_i}^2(\tau) + \cdots\right) \\
&= \frac{\eta}{\sqrt{n}} \sum_{i=1}^{n}\left[q_{x_i}^{'}(1)\right] - \frac{\eta^2}{2n}\sum_{i=1}^{n}\left[q_{x_i}^{'}(1)\right]^2 + \frac{\eta^2}{2n}\sum_{i}^{n}\left[q_{x_i}^{''}(1)\right] + O\left(n^{-\frac{1}{2}}\right)
\end{aligned}$$

S_n and L_n will be asymptotically joint normal under H_0, this is because that $E_F\left[q_{x_i}^{'}(1)\right] = E_F\left[q_{x_i}^{''}(1)\right] = 0$ leads to

$$\begin{aligned}
\begin{pmatrix} S_n \\ L_n \end{pmatrix} &= \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\begin{pmatrix} \mathrm{IF}[x_i;T,F] \\ \eta\, q_{x_i}^{'}(1) \end{pmatrix} + \begin{pmatrix} 0 \\ -\frac{\eta^2}{2n}\sum_{i=1}^{n}\left[q_{x_i}^{'}(1)\right]^2 \end{pmatrix} + O\left(n^{-\frac{1}{2}}\right) \\
&\xrightarrow{H_0} N\left\{\begin{pmatrix} 0 \\ -\frac{\eta^2\gamma^2}{2} \end{pmatrix}, \begin{pmatrix} E_F[\mathrm{IF}(x;T,F)]^2 & c \\ c & \eta^2\gamma^2 \end{pmatrix}\right\}
\end{aligned}$$

where $\gamma^2 = E_F\left[q_x^{'}(1)\right]^2$, $c = -\eta E_F\left[\mathrm{IF}\Big(x;T,F\Big)\left(1 + \frac{xf'(x)}{f(x)}\right)\mathrm{I}_{[x>0]}\right]$. Therefore, $F_{1,n}$ are contiguous to $F_{0,n}$ by LeCam first Lemma. Subsequently LeCam third Lemma gives the asymptotic distribution of S_n under $F_{1,n}$. □

Proof of Theorem 3.4.9 and Corollary 3.4.10

Proof. The first terms of the log-likelihood ratio

$$\begin{aligned}
L_n = \ln\left(\frac{\Psi_{1,n}}{\Psi_{0,n}}\right) &= \sum_{i=1}^{n}\ln\Big(2G(\alpha_n x_i)\Big) = \sum_{i=1}^{n}\ln\left(1 + \frac{2g(0)x_i}{\sqrt{n}} + \frac{g'(0)x_i^2}{n} + o(1)\right) \\
&= \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\Big(2g(0)x_i\Big) - \frac{1}{2n}\sum_{i=1}^{n}\Big(2g(0)x_i\Big)^2 + \frac{1}{n}\sum_{i=1}^{n}\Big(g'(0)x_i^2\Big) + o(1) \\
&\xrightarrow{H_0} N\left(-\frac{\sigma^2}{2} + \left[g'(0)E_F x^2\right], \sigma^2\right), \quad \text{where } \sigma^2 = E_F\Big(2g(0)x\Big)^2
\end{aligned}$$

$\Psi_{1,n}$ is contiguous to $\Psi_{0,n}$ iff $g'(0) = 0$. At contiguity, LeCam third Lemma gives the asymptotic distribution of S_n under $H_{1,n}$: $\sqrt{n}S_n \xrightarrow{d} N(u,\, v^2)$, where $v^2 = E_F[\mathrm{IF}(x;T,F)]^2$, and $u = E_F(S_n L_n) = 2g(0)E_F\Big(x\mathrm{IF}(x;T,F)\Big)$. This completes the proof. □

3.8 Appendix C: Tables and Figures

$f(x)$	$\text{ARE}\left(\mu - Q_{0.5}, \frac{E(x-\mu)^3}{\sigma^3}\right)$
Normal, $\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}$	$\frac{1}{3\pi-6} \approx 0.292$
Logistic, $\frac{e^x}{(1+e^x)^2}$	$\frac{23\pi^2}{35(\pi^2+12-24\ln 2)} \approx 1.2391$
Laplace, $\frac{1}{2}e^{-\lvert x\rvert}$	3.5
Triangular, $1-\lvert x\rvert,\ \lvert x\rvert \leq 1$	$\frac{9}{70}$
T_ν $\frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{\nu\pi}\,\Gamma\left(\frac{\nu}{2}\right)}\left(\frac{\nu+x^2}{\nu}\right)^{-\frac{\nu+1}{2}}$	$\frac{\nu^2-\nu+10}{\left(3\pi\alpha_\nu^*-6\frac{\nu-3}{\nu-1}\right)(\nu-4)(\nu-6)}$ $\text{ARE} = \begin{cases} 5.3685, \nu = 7 \\ 2.5482, \nu = 8 \\ 1.6829, \nu = 9 \end{cases}$
Slash, $S = \frac{Z}{U^{1/k}}$, $\frac{k}{\sqrt{2\pi}}\int_0^1 t^k e^{-\frac{(tx)^2}{2}}dt$	$\frac{k^4-10k^3+42k^2-68k+80}{\left\{3\pi\left(\frac{k+1}{k}\right)^2-6\frac{k^2-k-4}{(k-2)(k-1)}\right\}k(k-2)(k-4)(k-6)}$ $\text{ARE} = \begin{cases} 1.2800, k = 7 \\ 0.6945, k = 8 \\ 0.5209, k = 9 \end{cases}$

Table 3.1: ARE on Mixture (Part 1)

where $\alpha_\nu^* = \frac{\nu-2}{\nu}\alpha_\nu = \left(\frac{\nu-2}{2}\right)\left(\frac{\Gamma\left(\frac{\nu}{2}\right)}{\Gamma\left(\frac{\nu+1}{2}\right)}\right)^2 \xrightarrow{\nu} 1$, $\phi(x)$ is the pdf of a Normal

$f(x)$	ARE $\left(\mu - Q_{0.5}, \frac{Q_{1-p}+Q_p-2Q_{0.5}}{Q_{1-p}-Q_p}\right)$
Normal, $\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}$	$\frac{\left(e^{q^2}-2e^{\frac{q^2}{2}}\right)p+\frac{1}{2}}{\frac{\pi}{2}-1}\frac{\pi}{q^4}$, min=0.4561 ($p$=0.012)
Logistic, $\frac{e^x}{(1+e^x)^2}$	$\frac{0.5-4p^2+4p^3}{(2p-1)^2p(1-p)^2\left(\ln\frac{1-p}{p}\right)^2\left(\frac{\pi^2}{3}-8\ln 2+4\right)}$, min=0.1859 ($p$=0.0495)
Laplace, $\frac{1}{2}e^{-\lvert x\rvert}$	$\frac{e^q-1}{q^2}=\frac{\frac{1}{2p}-1}{(\ln 2p)^2}$, min=1.544 ($p$=0.1016)
Triangular, $1-\lvert x\rvert$, $\lvert x\rvert \leq 1$	$\frac{12p}{1-\sqrt{2p}}$ min=0 (p=0)
T_ν, $\frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{\nu\pi}\,\Gamma\left(\frac{\nu}{2}\right)}\left(\frac{\nu+x^2}{\nu}\right)^{-\frac{\nu+1}{2}}$	$\frac{\left[\left(1+\frac{q^2}{\nu}\right)^{\nu+1}-2\left(1+\frac{q^2}{\nu}\right)^{\frac{\nu+1}{2}}\right]p+\frac{1}{2}}{\frac{\pi}{2}\alpha_\nu-2\frac{\nu}{\nu-1}+\frac{\nu}{\nu-2}}\frac{\pi}{q^4}\left(\frac{\nu+q^2}{\nu+1}\right)^2\alpha_\nu$, $\text{Min}=\begin{cases}3.140(p=0.111), \nu=3\\2.009(p=0.085), \nu=4\\1.578(p=0.070), \nu=5\end{cases}$
Slash, $S=\frac{Z}{U^{1/k}}$, $f(x)=\frac{k}{\sqrt{2\pi}}\int_0^1 t^k e^{-\frac{(tx)^2}{2}}dt$	$\frac{\frac{p}{2}-\sqrt{2\pi}\left(\frac{k+1}{k}\right)f(q)p+\frac{\pi}{2}\left(\frac{k+1}{k}\right)^2f^2(q)}{\left(k\phi(q)-(k+1)f(q)\right)\left(\frac{\pi}{2}\left(\frac{k+1}{k}\right)^2+\frac{k}{k-2}-2\frac{k+1}{k-1}\right)}$ $\text{Min}=\begin{cases}0.553(p=0.064), k=3\\0.645(p=0.045), k=4\\0.630(p=0.035), k=5\end{cases}$

Table 3.2: ARE on Mixture (Part 2)

where $\alpha_\nu^*=\frac{\nu-2}{\nu}\alpha_\nu=\left(\frac{\nu-2}{2}\right)\left(\frac{\Gamma\left(\frac{\nu}{2}\right)}{\Gamma\left(\frac{\nu+1}{2}\right)}\right)^2\xrightarrow{\nu}1$, $\phi(x)$ is the pdf of a Normal

$f(x)$	$T_1 = \mu - Q_{0.5}$			$T_2 = \frac{E(x-\mu)^3}{\sigma^3}$			$T_3 = \frac{Q_{1-p}+Q_p-2Q_{0.5}}{Q_{1-p}-Q_p}$	
	c	υ^2	Efficacy	c	υ^2	Efficacy	Max Efficacy	p
Normal, $\frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$	$\frac{1}{\sqrt{2\pi}}\eta$	$\frac{\pi}{2} - 1$	0.2788	$\frac{3}{\sqrt{2\pi}}\eta$	6	0.2387	0.2501	0.053
Logistic, $\frac{e^x}{(1+e^x)^2}$								
Laplace, $\frac{1}{2} e^{-\lvert x \rvert}$	$\frac{1}{2}\eta$	1	0.25	$\frac{15}{2}\eta$	594	0.0947	0.1619	0.1016
T_ν, $\frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{\nu\pi}\,\Gamma\left(\frac{\nu}{2}\right)} \left(\frac{\nu+x^2}{\nu}\right)^{-\frac{\nu+1}{2}}$								
Slash, $S = \frac{Z}{U^{1/k}}$								

Table 3.3: Efficacy on Asymmetry

$f(x)$	$\text{ARE}\left(\mu - Q_{0.5}, \frac{E(x-\mu)^3}{\sigma^3}\right)$	$\text{ARE}\left(\mu - Q_{0.5}, \frac{Q_{1-p}+Q_p-2Q_{0.5}}{Q_{1-p}-Q_p}\right)$
Normal, $\frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$	$\frac{1}{3\pi-6} \approx 0.292$	$\frac{\left(e^{q^2} - 2e^{\frac{q^2}{2}}\right)p + \frac{1}{2}}{\frac{\pi}{2}-1} \frac{\pi}{q^4}$, min=0.4561 ($p$=0.012)
Logistic, $\frac{e^x}{(1+e^x)^2}$	$\frac{345\left(\frac{\pi^2}{3} - 4\ln 2\right)^2}{7\left(\frac{\pi^2}{3} - 8\ln 2 + 4\right)} \approx 0.7658$	
Laplace, $\frac{1}{2} e^{-\lvert x \rvert}$	3.5	$\frac{e^q - 1}{q^2} = \frac{\frac{1}{2p} - 1}{(\ln 2p)^2}$, min=1.544 ($p$=0.1016)
T_ν, $\frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{\nu\pi}\,\Gamma\left(\frac{\nu}{2}\right)} \left(\frac{\nu+x^2}{\nu}\right)^{-\frac{\nu+1}{2}}$	$\frac{(\nu^2-\nu+10)(\nu-4)}{\left(3\pi\alpha_\nu^* - 6\frac{\nu-3}{\nu-1}\right)(\nu-1)^2(\nu-6)}$, $\text{ARE} = \begin{cases} 1.3421, \nu = 7 \\ 0.8321, \nu = 8 \\ 0.6574, \nu = 9 \end{cases}$	
Slash, $S = \frac{Z}{U^{1/k}}$, $f(x) = \frac{k}{\sqrt{2\pi}} \int_0^1 t^k e^{-\frac{(tx)^2}{2}} dt$	$\frac{\left(k^4 - 10k^3 + 42k^2 - 68k + 80\right)(k-4)}{\left\{3\pi\left(\frac{k+1}{k}\right)^2 - 6\frac{k^2-k-4}{(k-2)(k-1)}\right\} k(k-1)^2(k-2)(k-6)}$, $\text{ARE} = \begin{cases} 0.3200, k = 7 \\ 0.2268, k = 8 \\ 0.2035, k = 9 \end{cases}$	

Table 3.4: ARE on Skew-Symmetry

where $\alpha_\nu^* = \frac{\nu-2}{\nu}\alpha_\nu = \left(\frac{\nu-2}{2}\right)\left(\frac{\Gamma\left(\frac{\nu}{2}\right)}{\Gamma\left(\frac{\nu+1}{2}\right)}\right)^2 \xrightarrow{\nu} 1$

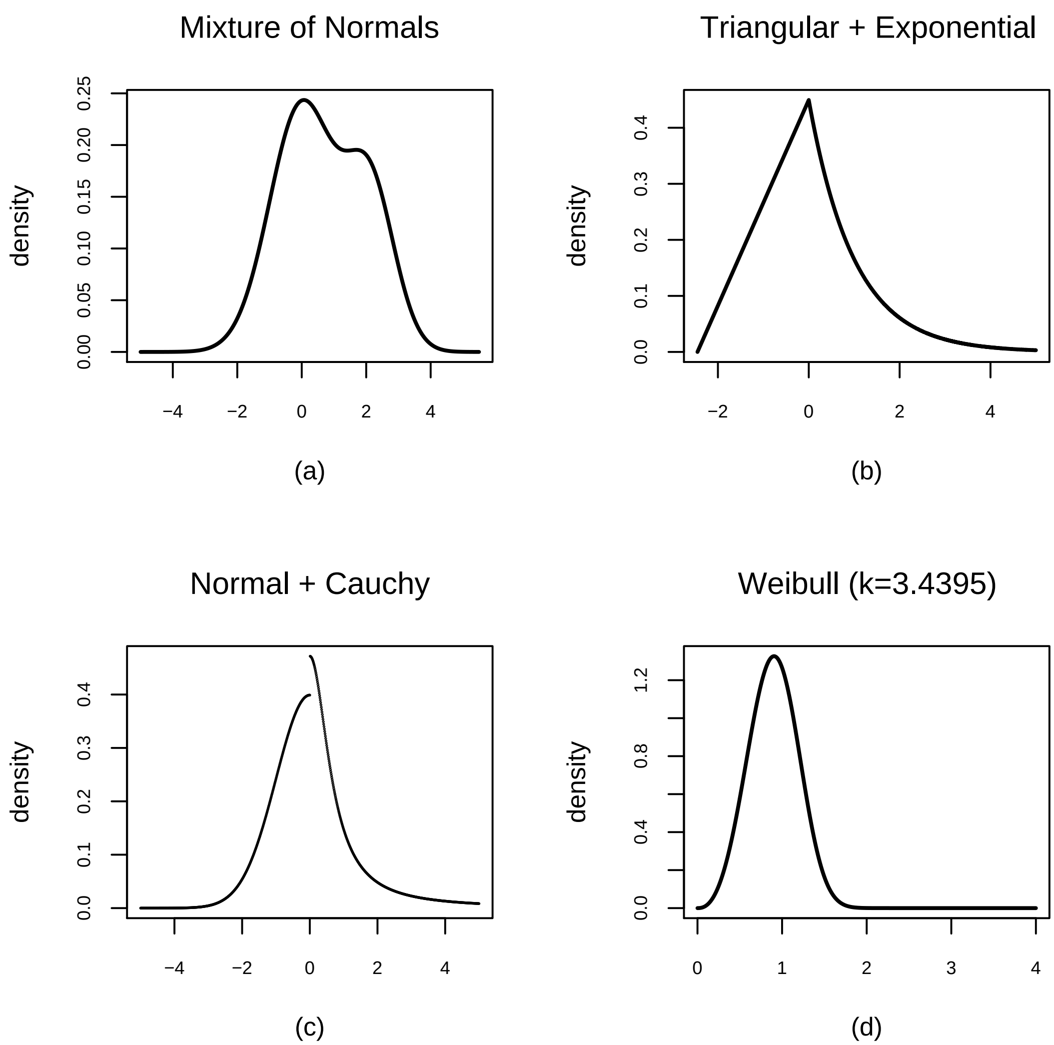

Figure 3.1: Asymmetric Distributions that classical skewness measures fail to capture

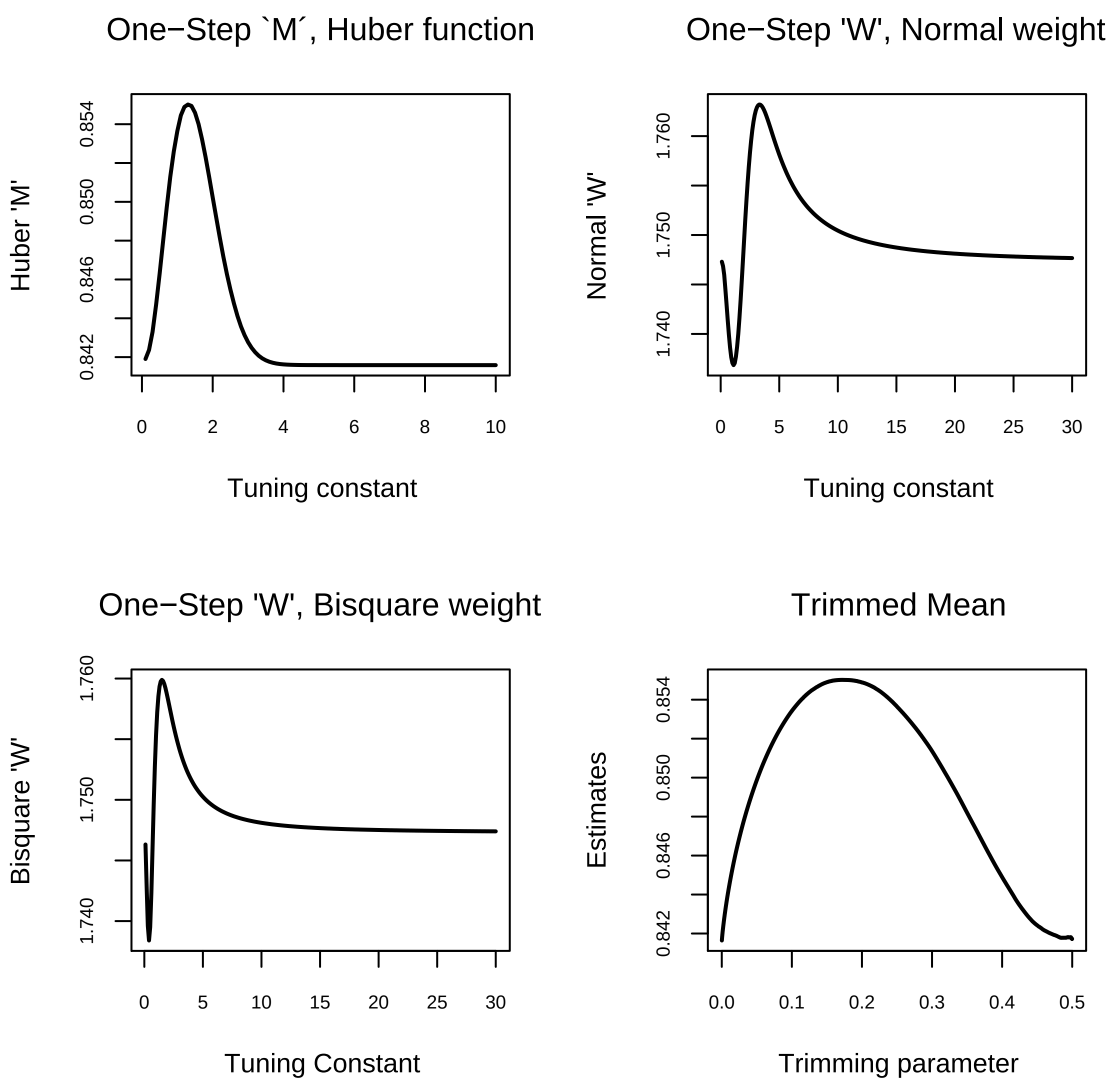

Figure 3.2: Different location estimates on a mixture of Normal Distributions

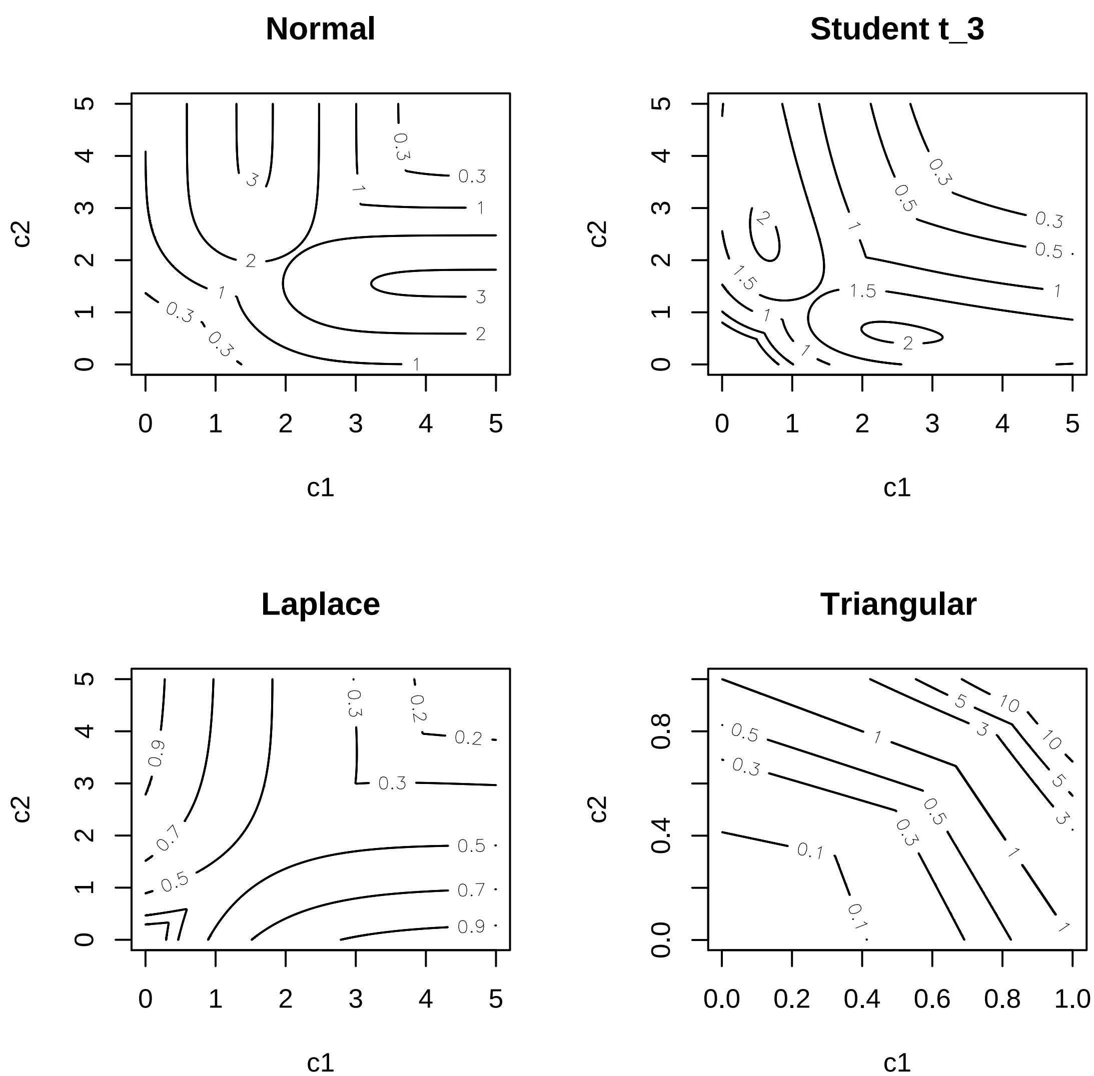

Figure 3.3: ARE of $\mu_{c1} - \mu_{c2}$ vs $\mu_0 - \mu_\infty$ ($\mu - Q_{0.5}$) within Huber family when testing mixture

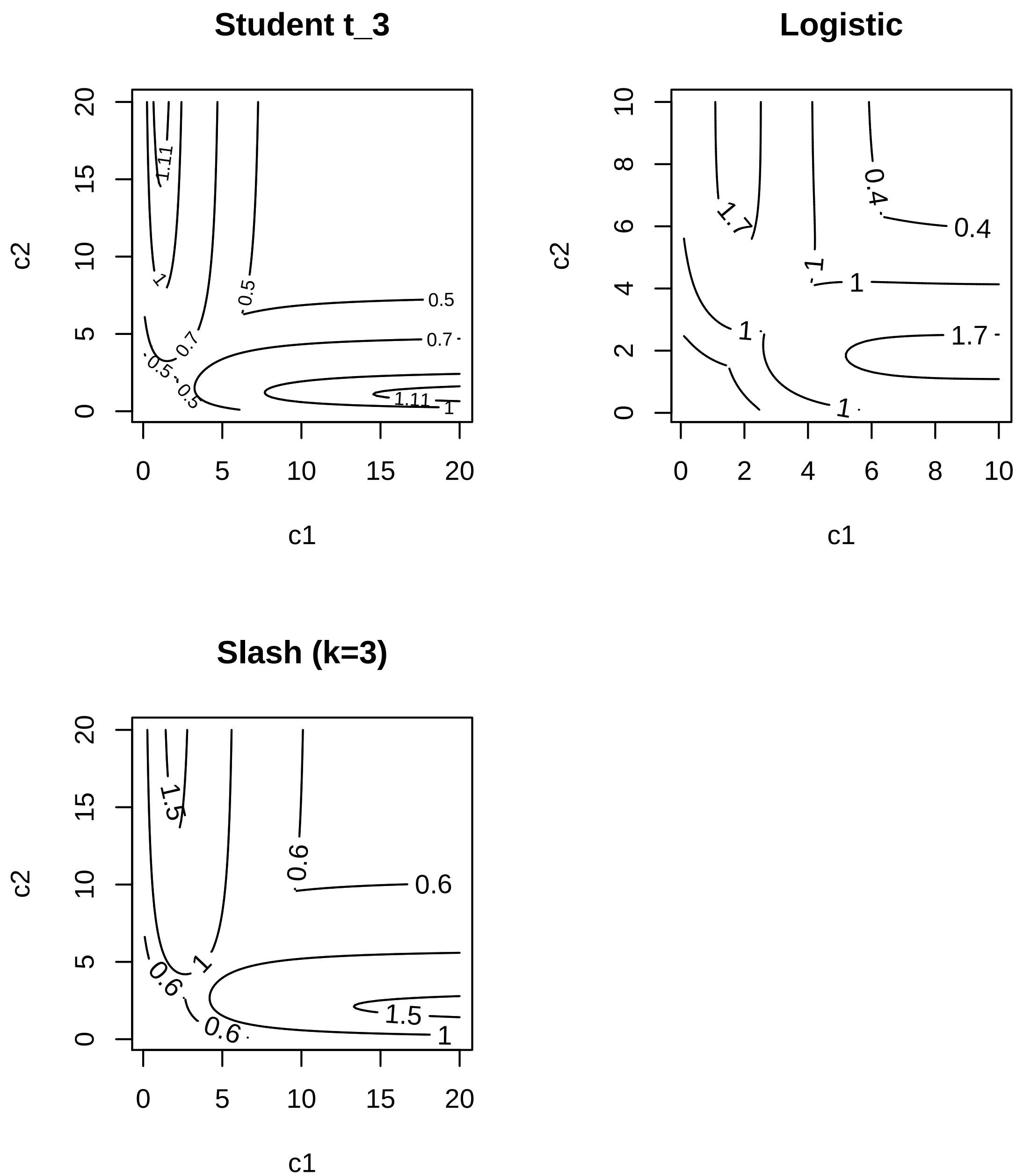

Figure 3.4: ARE of $\mu_{c1} - \mu_{c2}$ vs $\mu_0 - \mu_\infty$ ($\mu - Q_{0.5}$) within Huber family when testing Skew-Symmetry

Chapter 4

Beyond Multivariate Elliptical: Skew-Elliptical Distributions and Mixture of Elliptical Distributions

A key assumption that validates most descriptive statistics and test procedures employed in a classical multivariate analysis is unrealistic, that the underlying distribution has to be a normal distribution for which the location (mean) vector together with the scatter matrix is a sufficient statistic. Recently, Robust statistics researches have been flourishing in a particular generation of the normal distribution, namely class of elliptical distributions, denoted by $E_p\big(\mathbf{b}, \Sigma\big)$, whose density is of the form

$$f(\mathbf{x}) = |\Sigma|^{-1/2} h\left[(\mathbf{x}-\mathbf{b})'\Sigma^{-1}(\mathbf{x}-\mathbf{b})\right],$$

where $\mathbf{x}, \mathbf{b} \in \mathbb{R}^p$, $\Sigma \in \mathcal{PDS}(p)$, the set of positive definite symmetric matrices of order p, and $h : [0, \infty) \to [0, \infty)$ is a fixed function depending $\mathbf{x}$ only through $(\mathbf{x}-\mathbf{b})'\Sigma^{-1}(\mathbf{x}-\mathbf{b})$.

Several classes of robust estimates of multivariate location and scatter estimates have been implemented in an attempt to address the concerns of non-normality, especially a distribution with longer tails (than a Normal) on either direction or with outlying points. These robust estimates have desirable properties within the framework of elliptical symmetric distributions. Though comparison of these newly proposed estimates haven't been completed yet and studies within elliptical models are still promising, more vigorous researches beyond elliptical are necessary to provide some encouraging result, to bear meaningful fruit.

The non-elliptical distribution space is too broad, yet one can focus on the first steps of a natural extension beyond symmetry of elliptical. A skew-elliptical distribution, and a mixture of two elliptical distributions with different shapes are examples of such asymmetric distributions. To investigate the characteristics of a statistical procedure under these asymmetric framework, a basic yet powerful technical instrument is the **contiguity** of probability measures that leads to (local) asymptotic normality of statistical models, which together with LeCam's Lemmas suggest a feasible and efficient platform on which the behavioral differences in statistical aspects of a robust procedure (statistic), if any, when the underlying distribution changes from elliptical-symmetric P to asymmetric Q, could be tracked, analyzed and compared locally and asymptotically. One can set up a sequence of (local) hypotheses, H_0 verses $H_{1,n}$ indexed by $n \in \mathbb{N}$, in which Q_n is the joint distribution of a sample of size n when it is from alternative hypothesis while P_n is the joint distribution when sampling from null-hypothesis. The hypotheses are set-up in a way that Q_n is contiguous to P_n, and for a fixed n, Q_n and P_n are the (local) realizations of asymmetric Q and symmetric P respectively.

If scatter/shape functions exhibit interesting statistical differences between P and Q (expressed locally and asymptotically as between Q_n and P_n), then analyses on the differences across class of scatter/shape functions may lead to the suggestion or development of a new robust procedure at testing and estimating the shape of a multivariate data that would work equally powerful even when the underlying distribution is not limited to a symmetric one, rather, includes certain asymmetric ones.

In this spirit, the asymptotics of a scatter/shape estimate will be investigated separately on a Skew-Elliptical distribution, and on a mixture of multivariate distributions.

4.1 Asymptotic Distributions of Location/Shape Estimators

Let $\boldsymbol{\mu}(F) \in \mathbb{R}^p$ and $V(F) \in \mathbb{R}_p^p$ be affine equivariant Location and Shape functionals of distribution F, respectively, with $\widehat{\boldsymbol{\mu}}$ and $\widehat{V}$ denoting their consistent estimators based on sample. In this dissertation, the derivations of asymptotics of $\widehat{\boldsymbol{\mu}}$ and $\widehat{V}$ on asymmetric distribution Q rely heavily on the influence functions of $\boldsymbol{\mu}$ and V at a symmetric distribution P, denoted by $\mathrm{IF}(\mathbf{z}; \boldsymbol{\mu}, P)$ and $\mathrm{IF}(\mathbf{z}; V, P)$. The influence functions are not only of general simple forms, but also natural bridges connecting the estimators $\widehat{\boldsymbol{\mu}}$ and $\widehat{V}$ with their corresponding functionals $\boldsymbol{\mu}$ and V.

Lemma 4.1.1. *(Hampel 1986) Let $F \sim E_p(\mathbf{0}, I)$ denote the Spherical distribution in $\mathbb{R}^p$.*

1. *Affine equivariant location and shape functionals are characterized by three functions,* $w_0, w_1, w_2 : \mathbb{R}^+ \to \mathbb{R}$ *through*

$$\begin{aligned} \mathrm{IF}(\mathbf{z};\boldsymbol{\mu},F) &= w_0(\mathbf{z}'\mathbf{z})\mathbf{z} \\ \mathrm{IF}(\mathbf{z};V,F) &= w_1(\mathbf{z}'\mathbf{z})\mathbf{z}\mathbf{z}' - w_2(\mathbf{z}'\mathbf{z})\, I_p \end{aligned}$$

2. *Suppose* $\mathbf{z}_1, \cdots, \mathbf{z}_n \sim \mathbf{z} \in \mathbb{R}^p$ *are i.i.d. from* F*. Under certain regularity conditions,*

$$\begin{aligned} \sqrt{n}\ (\widehat{\boldsymbol{\mu}} - \boldsymbol{\mu}) &= \frac{1}{\sqrt{n}} \sum_{i=1}^{n} IF(\mathbf{z}_i;\boldsymbol{\mu},F) + o_p(1) \\ \sqrt{n}\ \left(\widehat{V} - V\right) &= \frac{1}{\sqrt{n}} \sum_{i=1}^{n} IF(\mathbf{z}_i;V,F) + o_p(1) \end{aligned}$$

Proof. See for example Hampel etc. [12] □

Taking trace of the influence function of scatter(shape) functional V, the relationship of w_1 and w_2 can be readily revealed,

$$w_2(\mathbf{z}'\mathbf{z}) = \frac{1}{p}\Big[w_1\big(\mathbf{z}'\mathbf{z}\big)\mathbf{z}'\mathbf{z}\Big] - \frac{1}{p}\mathrm{tr}\Big[\mathrm{IF}(\mathbf{z};V,F)\Big]$$

and this relationship is oftentimes applied to re-write the influence function of V in an equivalent form

$$\mathrm{IF}(\mathbf{z};V,F) = w_1(\mathbf{z}'\mathbf{z})\left[\mathbf{z}\mathbf{z}' - \frac{\mathbf{z}'\mathbf{z}}{p}I\right] + \frac{\mathrm{tr}\big[\mathrm{IF}(\mathbf{z};V,F)\big]}{p}I$$

By affine-equivariance, the general expressions of influence functions at an elliptical distribution $F^* \sim E_p\big(\mathbf{b},\Sigma\big)$ can be readily derived as

$$\begin{aligned} \mathrm{IF}\big(\mathbf{x};\boldsymbol{\mu},F^*\big) &= B\left[\mathrm{IF}\big(B^{-1}(\mathbf{x}-\mathbf{b});\boldsymbol{\mu},F\big)\right] = w_0(s)\big(\mathbf{x}-\mathbf{b}\big) \\ \mathrm{IF}\big(\mathbf{x};V,F^*\big) &= B\left[\mathrm{IF}\big(B^{-1}(\mathbf{x}-\mathbf{b});V,F\big)\right]B' = w_1(s)\big(\mathbf{x}-\mathbf{b}\big)\big(\mathbf{x}-\mathbf{b}\big)' - w_2(s)\,\Sigma \end{aligned}$$

here, $\Sigma = BB'$ and $s = (\mathbf{x}-\mathbf{b})'\Sigma^{-1}(\mathbf{x}-\mathbf{b})$.

Throughout this dissertation, location and scatter functionals and their consistent estimates are assumed to have properties described in this Lemma 4.1.1.

4.1.1 On Skew-Elliptical Distributions

Though multiple skewness mechanisms have been suggested in the literature, yielding various approaches to construct a skewed multivariate distribution of mathematical tractability and shape flexibility, most of these proposals are in essence the generalization of, hence resulting in similar probability density function as the popular model implemented by Azzalini and Dalla Valle [2], in which the pdf of a skew-symmetric random vector $\mathbf{x} \in \mathbb{R}^p$ is a multiplicative function

$$2f(\mathbf{x}; \mathbf{b}, \Sigma) G\left(\boldsymbol{\alpha}'(\mathbf{x} - \mathbf{b})\right)$$

where $f(\mathbf{x}; \mathbf{b}, \Sigma)$ is the pdf of a p-dimensional elliptical $F \sim E_p(\mathbf{b}, \Sigma)$, and G is the cdf of a univariate elliptical $E_1(0,1)$

The asymptotic of an affine equivariant estimate on skew-symmetric distribution ($\boldsymbol{\alpha} \neq \mathbf{0}$) versus symmetry ($\boldsymbol{\alpha} = \mathbf{0}$) can be investigated locally under the framework of a sequence of hypotheses

$$H_0 : \boldsymbol{\alpha} = \mathbf{0} \;\text{ vs }\; H_{1,n} : \boldsymbol{\alpha} = \mathbf{a}/\sqrt{n}, \quad \text{where } \mathbf{a}' = (a_1, \cdots, a_p) \in \mathbb{R}^p$$

Let $\Phi_{1,n}$ and $\Phi_{0,n}$ be joint densities of sample $\mathbf{x}_1, \cdots, \mathbf{x}_n \sim \mathbf{x} \in \mathbb{R}^p$ when $\mathbf{x}$ is from $H_{1,n}$, and from $F \sim H_0$ respectively. Without loss of generality, assuming F is the spherical distribution, i.e. $\mathbf{b} = \mathbf{0}$ and $\Sigma = I_p$.

Theorem 4.1.2. *$\Phi_{1,n}$ is contiguous to $\Phi_{0,n}$, provided $g(x) = dG(x)/dx$ depends on x only through x^2.*

Corollary 4.1.3. *Under $H_{1,n}$, location estimate $\widehat{\boldsymbol{\mu}}$ and scatter estimate $\widehat{V}$ are asymptotically independent and*

1. $\sqrt{n}\big(\widehat{\boldsymbol{\mu}} - \boldsymbol{\mu}\big) \xrightarrow{d} N^p\big(c_1\,\mathbf{a}\,,\, c_2\, I_p\big)$ *where*

$$c_1 = 2g(0) E_s\left[w_0(s)s\right]/p$$

$$c_2 = E_s\left[w_0^2(s)s\right]/p, \quad s = \mathbf{x}'\mathbf{x} \sim \frac{\pi^{p/2}}{\Gamma(p/2)} s^{\frac{p}{2}-1} f(s)$$

2. $\sqrt{n}\left(\widehat{V} - V\right) \xrightarrow{d} N_p^p\left(0, \Omega\right)$, *where*

$$\Omega = E_F\left\{\mathrm{Vec}\big[\mathrm{IF}(\mathbf{x}; V, F)\big]\mathrm{Vec}\big[\mathrm{IF}(\mathbf{x}; V, F)\big]'\right\} \in \mathbb{R}_{p^2}^{p^2}$$

Corollary 4.1.3 implies that an equivariant scatter estimate has the same asymptotic on an elliptical F as on the multiplicative skew-elliptical distribution generated from F, consequently, any equivariant scatter estimate would have no test power in detecting such asymmetry. Therefore in the remaining of this dissertation, the aspects of scatter functions on skew-elliptical distributions will not be further explored.

4.1.2 On Mixture of Elliptical Distributions

Let $F\big(\mathbf{0}, I_p\big)$ and $G\big(\mathbf{0}, I_p\big)$ be two **Spherical** distributions. There are two similar but different mechanisms to build a mixture that is not-symmetric

Case A: $H_0 : F\big(\mathbf{0}, I_p\big)$ vs $H_{1,n} : (1 - \delta/\sqrt{n})\, F\big(\mathbf{0}, I_p\big) + (\delta/\sqrt{n})\, G\big(\mathbf{0}, I_p\big)$

The contiguity of $H_{1,n}$ to H_0 is proved by Kankainen (2007) etc. in [16], in which the asymptotic normality of $\widehat{V}$ under alternatives is given explicitly as $\sqrt{n}\,\widehat{V} \xrightarrow{d} N\left(\delta\boldsymbol{\alpha}, \Omega\right)$, where

$$\begin{aligned}\boldsymbol{\alpha} &= E_G\left[\mathrm{IF}(\mathbf{z}; V, F)\right]\\ \Omega &= E_F\Big\{\mathrm{Vec}\big[\mathrm{IF}(\mathbf{z}; V, F)\big]\mathrm{Vec}\big[\mathrm{IF}(\mathbf{z}; V, F)\big]'\Big\}\end{aligned}$$

An immediate implication is that, when G is of the same shape as F but with different scale and center, i.e. $G\big(\mathbf{0}, I_p\big) = F\big(\mathbf{a}, bI_p\big)$, $\widehat{V}$ would have asymptotic mean of $\mathbf{0}$ under local alternatives. Any affine equivariant scatter estimate will not be able to differentiate such mixture that is non-symmetry from a contiguous symmetry. Because of this very reason, in later chapters statistical properties of a scatter-matrix function will not be discussed under asymmetry of this type of mixture distributions.

Case B:

$$H_0 : (1-\epsilon)F\big(\mathbf{0}, I_p\big)+\epsilon G\big(\mathbf{0}, I_p\big) \quad \text{vs} \quad H_{1,n} : (1-\epsilon)F\big(\mathbf{0}, I_p\big)+\epsilon G\left(\mathbf{a}/\sqrt{n},\ \big(I_p + D/\sqrt{n}\big)^{-1}\right)$$

where $\epsilon \in (0, 1)$ is fixed, $\mathbf{a} = (a_1, \cdots, a_p)' \in \mathbb{R}^p$, and D is a $p \times p$ diagonal matrix with diagonal entries d_i. Denote P the distribution of null-hypothesis.

Theorem 4.1.4. *Let Q_n and P_n be the joint probabilities of sample $\mathbf{z}_1, \cdots, \mathbf{z}_n \sim \mathbf{z} \in \mathbb{R}^p$ from $H_{1,n}$, and from H_0, respectively. Q_n is contiguous to P_n.*

Theorem 4.1.5. *Let $\widehat{\boldsymbol{\mu}}$ and $\widehat{V}$ be consistent estimates of affine equivariant location functional $\boldsymbol{\mu}$ and scatter functional V respectively. Under local alternative hypotheses,*

1. *$\widehat{\boldsymbol{\mu}}$ and $\widehat{V}$ are asymptotically independent*
2. *Let S be either $\boldsymbol{\mu}$ or V, and denote by $\widehat{S}$ the consistent estimator of S based on empirical distribution. $\sqrt{n}\left(\hat{S}-S\right) \xrightarrow{d} N\Big(\epsilon\big(\boldsymbol{\alpha}-2\boldsymbol{\alpha}_1\big),\Omega\Big)$, where*

$$\begin{aligned}
\Omega &= E_P\Big\{\mathrm{Vec}\big[\mathrm{IF}(\mathbf{z};S,P)\big]\mathrm{Vec}\big[\mathrm{IF}(\mathbf{z};S,P)\big]'\Big\},\\
\boldsymbol{\alpha} &= E_G\left[\left(\frac{\mathrm{tr}(D)}{2}+\frac{g'(\mathbf{z}'\mathbf{z})}{g(\mathbf{z}'\mathbf{z})}\mathbf{z}'D\mathbf{z}\right)\ \mathrm{IF}(\mathbf{z};S,P)\right]\\
\boldsymbol{\alpha}_1 &= E_G\left[\frac{g'(\mathbf{z}'\mathbf{z})}{g(\mathbf{z}'\mathbf{z})}\left(\mathbf{z}'\mathbf{a}\right)\ \mathrm{IF}(\mathbf{z};S,P)\right]
\end{aligned}$$

It is indicated clearly from the proof of Theorem 4.1.5 that when $S=V$ is a scatter/shape matrix, then $\boldsymbol{\alpha}_1=\mathbf{0}$. This in turn indicates that the location-shift $\mathbf{a}$ by component G of this type mixture has no effect on the asymptotics of an equivariant scatter estimator. By the same argument, $\boldsymbol{\alpha}=\mathbf{0}$ for location $\widehat{\boldsymbol{\mu}}$, this implies that the shape transformation D from component G doesn't change the asymptotic of a location estimator.

4.2 Summary

This dissertation is mainly focusing on the asymptotic behavior of a *scatter* matrix under a multivariate distribution that is beyond elliptical. As having discussed so far, a meaningful asymmetric platform on which the aspect of an equivariant *scatter* matrix can be further investigated is the mixture of two ellipticals:

$$(1-\epsilon)F\big(\mathbf{0},I_p\big)+\epsilon G\left(\mathbf{0},\ \big(I_p+D\big)^{-1}\right)$$

To facilitate the powerful instruments of Contiguity and LeCam's lemmas, the statistical properties of a scatter function under this asymmetric mixture will be explored locally and asymptotically on the sequence of equivalent alternative hypotheses:

$$\begin{aligned}
H_n &: (1-\epsilon)F\big(\mathbf{0},I_p\big)+\epsilon G\big(\mathbf{0},I_p\big) \qquad \text{vs}\\
H_{1,n} &: (1-\epsilon)F\big(\mathbf{0},I_p\big)+\epsilon G\left(\mathbf{0},\ \big(I_p+D/\sqrt{n}\big)^{-1}\right)
\end{aligned} \tag{4.2.1}$$

An affine-equivariant scatter estimate $\widehat{V}$ of scatter function V is asymptotically normal under $H_{1,n}$,

$$\sqrt{n}\left(\hat{V}-V\right)\xrightarrow{d}N\Big(\epsilon\boldsymbol{\alpha},\Omega\Big),$$

with the explicit formulae of $\boldsymbol{\alpha}$ and Ω presented in Theorem 4.1.5.

This sets up the primary framework and distribution-space that underly the discussions in the remaining chapters.

4.3 Appendix

Proof of Theorem 4.1.2

Proof. First note that $g(x)$ is a function of x^2 implies $g'(0) = \frac{dg(x)}{dx}|_{x=0} = 0$. The log-likelihood ratio statistics for testing $H_{1,n}$ against H_0 is given by

$$\begin{aligned}
L_n = \log \frac{\Phi_{1,n}}{\Phi_{0,n}} &= \sum_i^n \log \left[2G\left(\frac{1}{\sqrt{n}} \mathbf{a}'\mathbf{x}_i \right) \right] = \sum_i^n \log \left[1 + \frac{2g(0)}{\sqrt{n}} \mathbf{a}'\mathbf{x}_i + O(n^{-1}) \right] \\
&= \frac{1}{\sqrt{n}} \sum_i^n \left(2g(0)\mathbf{a}'\mathbf{x}_i \right) - \frac{1}{2n} \sum_i^n \left(2g(0)\mathbf{a}'\mathbf{x}_i \right)^2 + o(1) \\
&\xrightarrow[n]{\text{under } H_0} N\left(-\tau^2/2, \tau^2 \right)
\end{aligned}$$

since $2g(0)\mathbf{a}'E_F(\mathbf{x}) = 0$. Here $\tau^2 = E_F\big(2g(0)\mathbf{a}'\mathbf{x}\big)^2$. The contiguity of $\Phi_{1,n}$ to $\Phi_{0,n}$ follows LeCam first Lemma. □

Proof of Corollary 4.1.3

Proof. This is a direct application of Theorem 4.1.2 and LeCam third Lemma, from which $\sqrt{n}\left(\widehat{V} - V\right) \xrightarrow{d} N_p^p(A, \Omega)$ under $H_{1,n}$, where $\Omega = E_F\left\{ \mathrm{Vec}\big[\mathrm{IF}(\mathbf{x}; V, F)\big] \mathrm{Vec}\big[\mathrm{IF}(\mathbf{x}; V, F)\big]' \right\}$. Since $\mathrm{IF}\big(\mathbf{x}; V, F\big) = w_1(\mathbf{x}'\mathbf{x})\,(\mathbf{x}\mathbf{x}') - w_2(\mathbf{x}'\mathbf{x})I$ for some w_1 and w_2,

$$\begin{aligned}
A = \lim_n \left\{ \mathrm{Cov}\left[L_n, \sqrt{n}\left(\widehat{V} - V \right) \right] \right\} &= 2g(0)E_F\left[\big(\mathbf{a}'\mathbf{x}\big)\, \mathrm{IF}(\mathbf{x}; V, F) \right] \\
&= 2g(0)E_F\left[\big(\mathbf{a}'\mathbf{x}\big)\Big(w_1(\mathbf{x}'\mathbf{x})\mathbf{x}\mathbf{x}' - w_2(\mathbf{x}'\mathbf{x})I \Big) \right] \\
&= 2g(0)E_s\left(w_1(s)s^{\frac{3}{2}} \right) E_{\mathbf{u}}\left[(\mathbf{a}'\mathbf{u})\mathbf{u}\mathbf{u}' \right] - 2g(0)E_s\left(w_2(s)s^{\frac{1}{2}} \right) E_{\mathbf{u}}\left(\mathbf{a}'\mathbf{u} \right) I
\end{aligned}$$

where $s = \mathbf{x}'\mathbf{x} \sim \frac{\pi^{p/2}}{\Gamma(p/2)} s^{\frac{p}{2}-1} f(s)$, and $\mathbf{u} = \mathbf{x}/\sqrt{s}$ is uniformly distributed on the the unit sphere S^{p-1}. It follows $E_{\mathbf{u}}\left[\big(\mathbf{a}'\mathbf{u}\big)\mathbf{u}\mathbf{u}' \right] = 0$ and $E_{\mathbf{u}}\big(\mathbf{a}'\mathbf{u}\big)I = 0$, hence $A = 0$.

Similarly, $\sqrt{n}\,(\widehat{\boldsymbol{\mu}} - \boldsymbol{\mu}) \xrightarrow{d} N^p(\mathbf{c}, \Omega_0)$ under $H_{1,n}$. With $\mathrm{IF}(\mathbf{x}; \boldsymbol{\mu}, F) = w_0(\mathbf{x}'\mathbf{x})\,(\mathbf{x})$,

$$\begin{aligned}
\Omega_0 &= E_F\left[w_0^2(\mathbf{x}'\mathbf{x})\mathbf{x}\mathbf{x}' \right] = E_s\left[w_0^2(s)s \right] E_{\mathbf{u}}\left[\mathbf{u}\mathbf{u}' \right] = p^{-1} E_s\left[w_0^2(s)s \right] I \\
\mathbf{c} &= \lim_n \left\{ \mathrm{Cov}\left[L_n, \sqrt{n}(\widehat{\boldsymbol{\mu}} - \boldsymbol{\mu}) \right] \right\} = 2g(0)E_F\left[\big(\mathbf{a}'\mathbf{x}\big)\, \mathrm{IF}(\mathbf{x}; \boldsymbol{\mu}, F) \right] \\
&= 2g(0)E_F\left[\big(\mathbf{a}'\mathbf{x}\big)\Big(w_0(\mathbf{x}'\mathbf{x})\mathbf{x} \Big) \right] = 2g(0)E_s\left[w_0(s)s \right] E_{\mathbf{u}}\left[\big(\mathbf{a}'\mathbf{u}\big)\mathbf{u} \right]
\end{aligned}$$

Since $E_{\mathbf{u}}\left[\mathbf{u}\mathbf{u}' \right] = p^{-1} I_p$, $E_{\mathbf{u}}\left[(\mathbf{a}'\mathbf{u})\mathbf{u} \right] = \mathbf{a}/p$. This completes the proof. □

Proof of Theorem 4.1.4

Proof. Let $\Sigma_n^{-1} = I + D/\sqrt{n}$, $\mathbf{a_n} = \mathbf{a}/n$, $\mathbf{d} = (d_1, \cdots, d_p)'$, and $\mathbf{d_n} = \mathbf{d}/n$. The log-likelihood ratio statistic for testing $H_{1,n}$ against H_0 is given by $L_n = \log(dQ_n/dP_n) = \sum_{i=1}^{n} l(\mathbf{z}_i)$, where

$$\begin{aligned}
l(\mathbf{z}) &= \log\left(\frac{(1-\epsilon)f(\mathbf{z}'\mathbf{z}) + \epsilon|\Sigma_n|^{-\frac{1}{2}} g\big[(\mathbf{z}-\mathbf{a_n})'\Sigma_n^{-1}(\mathbf{z}-\mathbf{a_n})\big]}{(1-\epsilon)f(\mathbf{z}'\mathbf{z}) + \epsilon g(\mathbf{z}'\mathbf{z})}\right) \\
&= \log\left(1 + \epsilon\,\frac{|\Sigma_n|^{-\frac{1}{2}} g\big[(\mathbf{z}-\mathbf{a_n})'\Sigma_n^{-1}(\mathbf{z}-\mathbf{a_n})\big] - g(\mathbf{z}'\mathbf{z})}{dP}\right) \\
&= \log\big(1+\epsilon h\big) = \epsilon h - \frac{(\epsilon h)^2}{2} + \cdots
\end{aligned}$$

where $h \in \mathbb{R}$ is a function of $\mathbf{a_n}$ and $\mathbf{d_n}$, in the form

$$\begin{aligned}
h &= \frac{|\Sigma_n|^{-\frac{1}{2}}\, g\big[(\mathbf{z}-\mathbf{a_n})'\Sigma_n^{-1}(\mathbf{z}-\mathbf{a_n})\big] - g(\mathbf{z}'\mathbf{z})}{dP} \\
&= \left\{\prod_{j=1}^{p}\left(1+\frac{d_j}{\sqrt{n}}\right)^{1/2} g\left(\sum_{j=1}^{p}\left(1+\frac{d_j}{\sqrt{n}}\right)\left(z_j - \frac{a_j}{\sqrt{n}}\right)^2\right) - g(\mathbf{z}'\mathbf{z})\right\}/(dP)
\end{aligned}$$

whose Taylor expansion around the neighborhood of $\mathbf{0} \in \mathbb{R}^{2p}$ is simply,

$$h\begin{pmatrix}\mathbf{a_n}\\ \mathbf{d_n}\end{pmatrix} = \frac{1}{\sqrt{n}}\left(\mathbf{a}', \mathbf{d}'\right)\begin{pmatrix}\dfrac{\partial h}{\partial \mathbf{a_n}}\\ \dfrac{\partial h}{\partial \mathbf{d_n}}\end{pmatrix} + \frac{1}{n}\left(\mathbf{a}', \mathbf{d}'\right)\begin{pmatrix}\dfrac{\partial^2 h}{\partial \mathbf{a_n}^2} & \dfrac{\partial^2 h}{\partial \mathbf{d_n}\partial \mathbf{a_n}}\\ \dfrac{\partial^2 h}{\partial \mathbf{a_n}\partial \mathbf{d_n}} & \dfrac{\partial^2 h}{\partial \mathbf{d_n}^2}\end{pmatrix}\begin{pmatrix}\mathbf{a}\\ \mathbf{d}\end{pmatrix} + o\left(\frac{1}{n}\right)$$

Denoting $\mathbf{1}$ as the $p \times 1$ column vector of all ones, and $\odot$ as the Hadamard product, straightforward yet tedious calculations give the rather simple forms of the above derivatives of h evaluated at $\mathbf{0}$,

$$\frac{\partial h}{\partial \mathbf{a_n}} = \Big[-2g(\mathbf{z}'\mathbf{z})\mathbf{z}\Big]/dP, \qquad \frac{\partial^2 h}{\partial \mathbf{a_n}^2} = \Big[2g'(\mathbf{z}'\mathbf{z})I_p + 4g''(\mathbf{z}'\mathbf{z})\mathbf{z}\mathbf{z}'\Big]/dP$$

$$\frac{\partial h}{\partial \mathbf{d_n}} = \left[\frac{g(\mathbf{z}'\mathbf{z})}{2}\,\mathbf{1} + g'(\mathbf{z}'\mathbf{z})\big(\mathbf{z}\odot\mathbf{z}\big)\right]/dP$$

$$\begin{aligned}
\frac{\partial^2 h}{\partial \mathbf{d_n}^2} &= \begin{bmatrix} -\frac{g(\mathbf{z}'\mathbf{z})}{4} + g'(\mathbf{z}'\mathbf{z})z_1^2 + g''(\mathbf{z}'\mathbf{z})z_1^4 \;, & \frac{g(\mathbf{z}'\mathbf{z})}{4} + \frac{g'(\mathbf{z}'\mathbf{z})}{2}(z_1^2+z_2^2) + g''(\mathbf{z}'\mathbf{z})(z_1^2\, z_2^2) \;, & \cdots \\ \vdots \;, & -\frac{g(\mathbf{z}'\mathbf{z})}{4} + g'(\mathbf{z}'\mathbf{z})z_2^2 + g''(\mathbf{z}'\mathbf{z})z_2^4 \;, & \cdots \\ \vdots \;, & \vdots \;, & \ddots \end{bmatrix}/dP \\
&= \left[\frac{g(\mathbf{z}'\mathbf{z})}{4}\Big(\mathbf{1}\mathbf{1}' - 2I_p\Big) + \frac{g'(\mathbf{z}'\mathbf{z})}{2}\Big(\big(\mathbf{z}\odot\mathbf{z}\big)\mathbf{1}' + \mathbf{1}\big(\mathbf{z}\odot\mathbf{z}\big)'\Big) + g''(\mathbf{z}'\mathbf{z})\Big(\big(\mathbf{z}\odot\mathbf{z}\big)\big(\mathbf{z}\odot\mathbf{z}\big)'\Big)\right]/dP
\end{aligned}$$

$$\frac{\partial^2 h}{\partial \mathbf{d_n}\partial \mathbf{a_n}} = \left[\frac{\partial^2 h}{\partial \mathbf{a_n}\partial \mathbf{d_n}}\right]' = \Big[-g'(\mathbf{z}'\mathbf{z})\Big(\mathbf{z}\mathbf{1}' + 2\mathrm{Diag}\big(\mathbf{z}\big)\Big) - 2g''(\mathbf{z}'\mathbf{z})\mathbf{z}\big(\mathbf{z}\odot\mathbf{z}\big)'\Big]/dP$$

All these derivatives are of expectation $\mathbf{0}\,(\in \mathbb{R}^p$ or $\in \mathbb{R}^p_p)$, over the null-distribution P. This claim is trivially true for $(\partial h/\partial \mathbf{a_n})$ and $\big(\partial^2 h/\partial \mathbf{d_n}\partial \mathbf{a_n}\big)$ because of the symmetry of $\mathbf{z}$. To see this claim remains true for the other three derivatives, one needs to utilize the properties that if $\mathbf{z} \sim G(\mathbf{0}, I_p)$, then $s = \mathbf{z}'\mathbf{z}$ has density $\frac{\pi^{p/2}}{\Gamma(p/2)} s^{\frac{p}{2}-1} g(s)$ such that $E_s\left(\frac{g'(s)}{g(s)}s\right) = -\frac{1}{2}p$, $E_s\left(\frac{g''(s)}{g(s)}s^2\right) = \frac{1}{4}p(p+2)$, $E_s\left(\frac{g''(s)}{g(s)}s\right) = -\frac{p}{2}E_s\left(\frac{g'(s)}{g(s)}\right)$; and $\mathbf{u} = \mathbf{z}/\sqrt{s}$ is independent of s with identities $E\big[\mathbf{u}\mathbf{u}'\big] = \frac{1}{p}I$, $E\big[\mathbf{u}\odot\mathbf{u}\big] = \frac{1}{p}\mathbf{1}$, $E\Big[\big(\mathbf{u}\odot\mathbf{u}\big)\big(\mathbf{u}\odot\mathbf{u}\big)'\Big] = \frac{1}{p(p+2)}\mathbf{1}\mathbf{1}' + \frac{2}{p(p+2)}I$, so that

$$
\begin{aligned}
E_P\left[\frac{\partial h}{\partial \mathbf{d_n}}\right] &= \frac{1}{2}\mathbf{1} + E_G\left[\frac{g'(\mathbf{z}'\mathbf{z})}{g(\mathbf{z}'\mathbf{z})}\big(\mathbf{z}\odot\mathbf{z}\big)\right] = \frac{1}{2}\mathbf{1} + E_s\left[\frac{g'(s)}{g(s)}s\right]E_{\mathbf{u}}\Big[\mathbf{u}\odot\mathbf{u}\Big] = \mathbf{0}\in\mathbb{R}^p \\
E_P\left[\frac{\partial^2 h}{\partial \mathbf{a_n}^2}\right] &= E_G\left[2\frac{g'(\mathbf{z}'\mathbf{z})}{g(\mathbf{z}'\mathbf{z})}I_p + 4\frac{g''(\mathbf{z}'\mathbf{z})}{g(\mathbf{z}'\mathbf{z})}\mathbf{z}\mathbf{z}'\right] \\
&= 2E_s\left[\frac{g'(s)}{g(s)}\right]I + 4E_s\left[\frac{g''(s)}{g(s)}s\right]E_{\mathbf{u}}\left[\mathbf{u}\mathbf{u}'\right] = 0\in\mathbb{R}^p_p \\
E_P\left[\frac{\partial^2 h}{\partial \mathbf{d_n}^2}\right] &= 4\Big[\mathbf{1}\mathbf{1}' - 2I_p\Big] + \frac{1}{2}E_G\left\{\frac{g'(\mathbf{z}'\mathbf{z})}{g(\mathbf{z}'\mathbf{z})}\Big[\big(\mathbf{z}\odot\mathbf{z}\big)\mathbf{1}' + \mathbf{1}\big(\mathbf{z}\odot\mathbf{z}\big)'\Big]\right\} \\
&\quad + E_G\left\{\frac{g''(\mathbf{z}'\mathbf{z})}{g(\mathbf{z}'\mathbf{z})}\Big[\big(\mathbf{z}\odot\mathbf{z}\big)\big(\mathbf{z}\odot\mathbf{z}\big)'\Big]\right\} \\
&= 4\Big[\mathbf{1}\mathbf{1}' - 2I_p\Big] + \frac{1}{2}E_s\left[\frac{g'(s)}{g(s)}s\right]E_{\mathbf{u}}\Big[\big(\mathbf{u}\odot\mathbf{u}\big)\mathbf{1}' + \mathbf{1}\big(\mathbf{u}\odot\mathbf{u}\big)'\Big] \\
&\quad + E_s\left[\frac{g''(s)}{g(s)}s^2\right]E_{\mathbf{u}}\Big[\big(\mathbf{u}\odot\mathbf{u}\big)\big(\mathbf{u}\odot\mathbf{u}\big)'\Big] = 0\in\mathbb{R}^p_p
\end{aligned}
$$

To this end, let

$$
\beta(\mathbf{z}) = \big(\mathbf{a}', \mathbf{d}'\big)\begin{pmatrix}\dfrac{\partial h}{\partial \mathbf{a_n}} \\ \dfrac{\partial h}{\partial \mathbf{d_n}}\end{pmatrix} = \left[-2\,g'(\mathbf{z}'\mathbf{z})\mathbf{z}'\mathbf{a} + \frac{\operatorname{tr}(D)}{2}g(\mathbf{z}'\mathbf{z}) + g'(\mathbf{z}'\mathbf{z})\mathbf{z}'D\mathbf{z}\right]/dP
$$

and observe that,

$$
\begin{aligned}
E_P\Big[\beta(\mathbf{z})\Big] &= -2E_G\left[\frac{g'(\mathbf{z}'\mathbf{z})}{g(\mathbf{z}'\mathbf{z})}\mathbf{z}'\mathbf{a}\right] + \frac{\operatorname{tr}(D)}{2} + E_G\left[\frac{g'(\mathbf{z}'\mathbf{z})}{g(\mathbf{z}'\mathbf{z})}\mathbf{z}'D\mathbf{z}\right] \\
&= -2E_s\left[\frac{g'(s)}{g(s)}\sqrt{s}\right]E_{\mathbf{u}}\big[\mathbf{u}'\mathbf{a}\big] + \frac{\operatorname{tr}(D)}{2} + \left[E_s\frac{g'(s)}{g(s)}s\right]E_{\mathbf{u}}\left(\mathbf{u}'D\mathbf{u}\right) = 0\in\mathbb{R}^1
\end{aligned}
$$

The Taylor expansion of the log-likelihood ratio hence becomes

$$
\begin{aligned}
L_n = \sum_i^n l(\mathbf{z}_i) &= \frac{1}{\sqrt{n}}\sum_i^n\Big[\epsilon\beta(\mathbf{z}_i)\Big] - \frac{1}{2n}\sum_i^n\Big[\epsilon\beta(\mathbf{z}_i)\Big]^2 + o(n^{-1}) \\
&\xrightarrow[n]{\text{under } H_0} N\big(-\tau^2/2, \tau^2\big)
\end{aligned}
$$

where $\tau^2 = E_P\big[\epsilon\beta(\mathbf{z})\big]^2$. The contiguity of Q_n to P_n follows LeCam first Lemma. $\square$

Proof of Theorem 4.1.5

Proof. The limiting distribution of $\sqrt{n}\left(\widehat{S}-S\right)$ under local alternative hypotheses is a multivariate normal distribution according to LeCam third Lemma because of the contiguity of $H_{1,n}$ to H_0. Using notations in the proof of Theorem 4.1.4, the asymptotic mean $\boldsymbol{\mu}$ and Variance-Covariance matrix Ω are obtained as

$$
\begin{aligned}
\boldsymbol{\mu} &= E_P\Big[\epsilon\beta(\mathbf{z})\,\mathrm{IF}(\mathbf{z};S,P)\Big]\\
&= \epsilon\int\left(-2\,g'(\mathbf{z}'\mathbf{z})\mathbf{z}'\mathbf{a}+\frac{\mathrm{tr}(D)}{2}g(\mathbf{z}'\mathbf{z})+g'(\mathbf{z}'\mathbf{z})\mathbf{z}'D\mathbf{z}\right)\mathrm{IF}(\mathbf{z};S,P)d\mathbf{z}\\
&= -2\epsilon E_G\left[\frac{g'(\mathbf{z}'\mathbf{z})}{g(\mathbf{z}'\mathbf{z})}\big(\mathbf{z}'\mathbf{a}\big)\mathrm{IF}(\mathbf{z};S,P)\right]+\epsilon E_G\left[\left(\frac{\mathrm{tr}(D)}{2}+\frac{g'(\mathbf{z}'\mathbf{z})}{g(\mathbf{z}'\mathbf{z})}\mathbf{z}'D\mathbf{z}\right)\mathrm{IF}(\mathbf{z};S,P)\right]\\
&= -2\epsilon\boldsymbol{\alpha}_1+\epsilon\boldsymbol{\alpha},\qquad\text{say.}\\
\Omega &= E_P\Big\{\mathrm{Vec}\big[\mathrm{IF}(\mathbf{z};S,P)\big]\mathrm{Vec}\big[\mathrm{IF}(\mathbf{z};S,P)\big]'\Big\}.
\end{aligned}
$$

For the case when $S=V$ is an affine equivariant scatter matrix, the structure of its influence function (Lemma 4.1.1), $\mathrm{IF}(\mathbf{z};V,F)=w_1(\mathbf{z}'\mathbf{z})\mathbf{z}\mathbf{z}'-w_2(\mathbf{z}'\mathbf{z})\,I_p$, leads to

$$
\begin{aligned}
\boldsymbol{\alpha}_1 &= E_G\left[\frac{g'(\mathbf{z}'\mathbf{z})}{g(\mathbf{z}'\mathbf{z})}\big(\mathbf{z}'\mathbf{a}\big)\mathrm{IF}(\mathbf{z};V,P)\right]\\
&= E_s\left[\frac{g'(s)}{g(s)}w_1(s)s^{\frac{3}{2}}\right]E_{\mathbf{u}}\Big[\big(\mathbf{a}'\mathbf{u}\big)\mathbf{u}\mathbf{u}'\Big]-E_s\left[\frac{g'(s)}{g(s)}w_2(s)s^{\frac{1}{2}}\right]E_{\mathbf{u}}\Big[\mathbf{a}'\mathbf{u}\Big]I
\end{aligned}
$$

where the last step comes from the symmetry of $\mathbf{u}=\mathbf{z}/\sqrt{\mathbf{z}'\mathbf{z}}$ whenever $\mathbf{z}=\sqrt{s}\mathbf{u}\sim G\big(\mathbf{0},I\big)$ is from a spherical distribution. □

Chapter 5

Robust Shape/Scatter Estimators at Mixture Models

5.1 M-estimates

The multivariate M-functionals of location and scatter were first independently introduced by Maronna (1976) and Huber (1977). Let $\rho(s)$ be a given function on $s \geq 0$ and let $\mathcal{PDS}(p)$ denote the set of positive definite symmetric matrices of order p. Let $\mathbf{x} \in \mathbb{R}^p$ be a random vector from distribution F. The M-functionals of multivariate location and scatter are defined to be any pair $\boldsymbol{\mu}(F) \in \mathbb{R}^p$ and $V(F) \in \mathcal{PDS}(p)$ respectively which minimizes the objective function

$$L(\boldsymbol{\mu}, V; \mathbf{x}) = E\Big[\rho\big((\mathbf{x}-\boldsymbol{\mu})V^{-1}(\mathbf{x}-\boldsymbol{\mu})\big)\Big] + \frac{1}{2}\log|V| \tag{5.1.1}$$

over all $\boldsymbol{\mu} \in \mathbb{R}^p$ and $V \in \mathcal{PDS}(p)$, where the expectation is taken over F. If the expectation is taken over the empirical distribution F_n of a random sample $\mathbf{x}_1, \cdots, \mathbf{x}_n$ from F, the solutions to the minimization problem (5.1.1), denoted by $\widehat{\boldsymbol{\mu}}(F_n)$ and $\widehat{\boldsymbol{V}}(F_n)$ are conventionally called M-estimates of location and scatter, respectively.

If ρ is differentiable, then setting the derivatives of (5.1.1) with respect to $\boldsymbol{\mu}$ and V to $\mathbf{0}$ yields the simultaneous estimating equations

$$\boldsymbol{\mu} = E\big[u(s)\mathbf{x}\big]/E\big[u(s)\big] \tag{5.1.2}$$

$$V = E\Big[u(s)\big(\mathbf{x}-\boldsymbol{\mu}\big)\big(\mathbf{x}-\boldsymbol{\mu}\big)'\Big] \tag{5.1.3}$$

where $s = (\mathbf{x}-\boldsymbol{\mu})V^{-1}(\mathbf{x}-\boldsymbol{\mu})$ and $u(s) = 2\rho'(s)$.

The existence and uniqueness problems of solutions $(\widehat{\boldsymbol{\mu}}, \widehat{V})$ to the implicitly defined M-estimating equations (5.1.2) and (5.1.3) were initially attempted by Maronna (1976), and Huber (1981), but completely solved by Kent and Tyler (1991). Certain conditions on the function $\rho(s)$, weight function $u(s)$, as well as on the data $\{\mathbf{x}_1, \cdots, \mathbf{x}_n\}$ are needed to ensure the existence and uniqueness. In general, the function $\rho(s)$ should be unbounded (Maronna (1976)), and $su(s)$ has to be non-decreasing in s (Kent and Tyler (1991)).

When the M-estimates of multivariate location and scatter are defined through implicit equations, more general form than that given by (5.1.2) and (5.1.3) is proposed in the literatures. Maronna (1976) allows the two weight functions $u(s)$ in (5.1.2) and (5.1.3) differ from each other, hence creating a more general class of multivariate M-estimates that needs not to be related a minimization problem of form (5.1.1).

This dissertation primarily concerns with the analysis of the shape of a multivariate distribution via robust scatter matrix. With out loss of generosity, the location (center) of a symmetric distribution, elliptical distribution for example, can be assumed to be known and to be $\mathbf{0}$. In this scatter-only setting, the M-functional of the scatter of a multivariate distribution has a simpler definition.

Definition 5.1.1. *A scatter* M-functional $V \in \mathcal{PDS}(p)$ *at* F *is defined implicitly as the solution to* $E_F(\psi(\mathbf{z}, V)) = 0$*, where* $\psi(\mathbf{z}, V) = u\big(\mathbf{z}'V^{-1}\mathbf{z}\big)\mathbf{z}\mathbf{z}' - V$ *for some suitable function* u*. A scatter* M-estimate $\widehat{V}$ *is a consistent estimator of* V *based on sample* $\{\mathbf{z}_1, \cdots, \mathbf{z}_n\}$ *satisfying* $1/n\sum_{i=1}^n \psi(\mathbf{z}_i, \widehat{V}) = 0$ *under regularity conditions.*

Proposition 5.1.2. *A scatter M-functional* V *at* $F \sim E_p(\mathbf{0}, I)$ *defined by* $\psi(\mathbf{z}, V) = u(\mathbf{z}'V^{-1}\mathbf{z})\mathbf{z}\mathbf{z}' - V$ *has influence function*

$$
\begin{aligned}
IF\left(\mathbf{z}; V, F\right) &= \frac{u\left(\mathbf{z}'\mathbf{z}/\lambda\right)}{2h+1}\,\mathbf{z}\mathbf{z}' - \frac{\frac{h}{(2h+1)}u\left(\mathbf{z}'\mathbf{z}/\lambda\right)\mathbf{z}'\mathbf{z} + \lambda}{(p+2)h+1}\,I_p \\
&= \frac{u\left(\mathbf{z}'\mathbf{z}/\lambda\right)}{2h+1}\left(\mathbf{z}\mathbf{z}' - \frac{\mathbf{z}'\mathbf{z}}{p}I\right) + \frac{u\left(\mathbf{z}'\mathbf{z}/\lambda\right)\mathbf{z}'\mathbf{z}/p - \lambda}{(p+2)h+1}I
\end{aligned}
$$

where $h = E\left(u'(\mathbf{z}'\mathbf{z}/\lambda)(\mathbf{z}'\mathbf{z}/\lambda)^2\right)/p(p{+}2)$*, and* λ *is the solution to* $E\left(u\left(\mathbf{z}'\mathbf{z}/\lambda\right)\mathbf{z}'\mathbf{z}/\lambda\right) = p$*.*

As summarized in section (4.2), the asymptotic of an M-estimate of scatter on a non-elliptical distribution will be investigated under the platform on a sequence of hypotheses defined in (4.2.1), i.e.

$$
\begin{aligned}
&H_0 : (1-\epsilon)F\big(\mathbf{0}, I_p\big) + \epsilon G\big(\mathbf{0}, I_p\big) \quad \text{vs} \qquad (5.1.4)\\
&H_{1,n} : (1-\epsilon)F\big(\mathbf{0}, I_p\big) + \epsilon G\left(\mathbf{0},\ \left(I_p + D/\sqrt{n}\right)^{-1}\right)
\end{aligned}
$$

Using the notations in Theorem 4.1.4, assuming that $G(\mathbf{0}, I) = F\left(\mathbf{0}, \nu^2 I\right)$, i.e. $g(\mathbf{x}'\mathbf{x}) = \nu^{-p} f(\mathbf{x}'\mathbf{x}/\nu^2)$ and the null hypothesis is a mixture $P \sim (1-\epsilon)F\big(\mathbf{0}, I_p\big) + \epsilon F\big(\mathbf{0}, \nu^2 I_p\big)$, the asymptotic distribution of an M-estimate of scatter is readily obtained via Theorem 4.1.5.

Theorem 5.1.3. *Let V be an M-functional of Definition 5.1.1 with influence function defined in Proposition 5.1.2, and $\widehat{V}$ be an affine equivariant estimator of V based on sample. Under sequence of alternative hypotheses $H_{1,n}$ of 5.1.4, $V \to \lambda I$, where λ is the solution to equation $(1-\epsilon)E_s\left[u(s/\lambda)(s/\lambda)\right] + \epsilon E_s\left[u(s\nu^2/\lambda)(s\nu^2/\lambda)\right] = p$; and*

1. $\sqrt{n}\left(\widehat{V} - \lambda I\right) \xrightarrow{d} N(-\epsilon\lambda\, \mathbf{m}, \Omega\,)$, *where* $\mathbf{m} = \alpha D - \beta\, \mathrm{tr}(D) I$, $\Omega = \gamma(I + K_p) + \eta\, \mathrm{Vec}(I)\mathrm{Vec}(I)'$, *and*

$$\alpha = \frac{1}{2h+1}\left\{\frac{E_s\left[u\left(s\nu^2/\lambda\right)\left(s\nu^2/\lambda\right)\right]}{p} + \frac{2E_s\left[u'\left(s\nu^2/\lambda\right)\left(s\nu^2/\lambda\right)^2\right]}{p(p+2)}\right\}$$

$$\beta = \frac{1}{(2h+1)[(p+2)h+1]}\left\{\frac{hE_s\left[u\left(s\nu^2/\lambda\right)\left(s\nu^2/\lambda\right)\right]}{p} - \frac{E_s\left[u'\left(s\nu^2/\lambda\right)\left(s\nu^2/\lambda\right)^2\right]}{p(p+2)}\right\}$$

$$h = \frac{1}{p(p+2)}E_s\left[(1-\epsilon)u'\left(s/\lambda\right)\left(s/\lambda\right)^2 + \epsilon u'\left(s\nu^2/\lambda\right)\left(s\nu^2/\lambda\right)^2\right]$$

$$\gamma = \frac{1}{(2h+1)^2 p(p+2)}\left\{(1-\epsilon)E_s\left[u(s/\lambda)s\right]^2 + \epsilon E_s\left[u\left(s\nu^2/\lambda\right)s\nu^2\right]^2\right\}$$

$$\eta = \frac{[1-2h^2(p+2)]\gamma - \lambda^2}{[(p+2)h+1]^2}, \qquad s \sim \frac{\pi^{p/2}}{\Gamma(p/2)} s^{\frac{p}{2}-1} f(s)$$

2. *If $\nu^2 = 1$, i.e. $F\big(\mathbf{0}, I_p\big) = G\big(\mathbf{0}, I_p\big)$, then $\sqrt{n}\left(\widehat{V} - \lambda I\right) \xrightarrow{d} N_p^p\left(-\epsilon\lambda\, D, \Omega\,\right)$.*

Corollary 5.1.4. *Assuming notations in Theorem 5.1.3, under the sequence of alternative hypotheses $H_{1,n}$ of 5.1.4,*

1. $n\left[\log\frac{\left(\frac{1}{p}\mathrm{tr}\,\widehat{V}\right)^p}{|\widehat{V}|}\right] \xrightarrow{d} \frac{\gamma}{\lambda^2}\chi_q^2(\delta)$, *a non-central Chi-square distribution with degree of freedom $q = \frac{1}{2}(p+2)(p-1)$ and non-centrality $\delta = \frac{(\epsilon\lambda\alpha)^2}{2\gamma}\left[\mathrm{tr}\left(D^2\right) - \frac{1}{p}\mathrm{tr}^2(D)\right]$*

2. $\frac{n}{2}\log\left[\frac{\mathrm{tr}\,\widehat{V}\cdot \mathrm{tr}\left(\widehat{V}^{-1}\right)}{p^2}\right]^p \xrightarrow{d} \frac{\gamma}{\lambda^2}\chi_q^2(\delta)$

From the proof of Corollary 5.1.4, the limiting distribution of $n\left[p\log\left(\mathrm{tr}\,\widehat{V}/p\right) - \log|V|\right]$ is proportional to $Z'AZ$, where $Z \in \mathbb{R}^{p^2}$ is a random vector of normal distribution and $A = \left(\frac{1}{2}(I + K_p) - \frac{1}{p}\,\mathrm{Vec}(I)\mathrm{Vec}(I)'\right)$ is an idempotent matrix of order p^2, i.e. $A = A^2$. This idempotent has eigenvalue 0 occurring $\frac{1}{2}p(p-1)+1$ times. The associated Eigenvectors are the linear combinations of the identity and skew-symmetric matrices. This property leads to two immediate consequences. Firstly, the non-centrality parameter δ will depend on the influence function only through ω_1 whenever $IF(\mathbf{z}; V_1^{-1}V_2, F) = \omega_1 \mathbf{z}\mathbf{z}' - \omega_2 I$. Secondly, the statistics in theorem 5.1.3 are unable to separate a distribution G from an elliptical F, if G is such that $\sqrt{n}\left(\widehat{V} - \lambda I\right)$ has an asymptotic mean proportional to I or to $B = -B'$.

5.2 S-estimates

S-estimate was first introduced by Rousseeuw and Yohai (1984) in the context of multiple regression as a generalization of LMS estimator. Davies (1987) gave a slightly different definition in an attempt to address the low breakdown property inherited within M-estimates. Today's universal accepted version of S-estimates of location and scatter is suggested by Lopuhaä (1989).

Definition 5.2.1. *The* S-functionals *of multivariate location and scatter are defined to be any pair of $\boldsymbol{\mu}(F) \in \mathbb{R}^p$ and $V(F) \in \mathcal{PDS}(p)$ respectively which minimizes* $\det(V)$, *subject to the constraint*

$$E\Big[\rho\big(\sqrt{(\mathbf{x}-\boldsymbol{\mu})V^{-1}(\mathbf{x}-\boldsymbol{\mu})}\big)\Big] = \epsilon_0\rho(\infty), \tag{5.2.1}$$

where ϵ_0 is a fixed value between 0 and 1, and for $s \geq 0$, $\rho(s)$ is non-decreasing, left continuous everywhere, right continuous at zero with $0 = \rho(0) < \rho(\infty) < \infty$

When ρ is differentiable with derivative ψ, Lopuhaä [20] showed the S-functionals of location and scatter satisfy the simultaneous S-estimating equations

$$\boldsymbol{\mu} = E\big[u_2(t)\mathbf{x}\big]/E\big[u_2(t)\big] \tag{5.2.2}$$

$$V = E\Big[u_2(t)\big(\mathbf{x}-\boldsymbol{\mu}\big)\big(\mathbf{x}-\boldsymbol{\mu}\big)'\Big]/E\big[u_3(t)\big] \tag{5.2.3}$$

where $t = \sqrt{(\mathbf{x}-\boldsymbol{\mu})V^{-1}(\mathbf{x}-\boldsymbol{\mu})}$ and

$$u_2(t) = \psi(t)/t$$
$$u_3(t) = \left[\psi(t)t - \rho(t) + \epsilon_0\rho(\infty)\right]/p$$

It follows that S-functionals $(\boldsymbol{\mu}, V)$ meet first-order conditions (5.1.3) and (5.1.2) of M-functionals defined by Huber (1981). Consequently, their influence functions exist under certain regularity condition [20], and satisfy the following relationship [12, 15]

$$\begin{pmatrix} \mathrm{IF}(\mathbf{x};\boldsymbol{\mu},F) \\ \mathrm{Vec}\left[\mathrm{IF}(\mathbf{x};V,F)\right] \end{pmatrix} = \begin{pmatrix} \Lambda_{\boldsymbol{\mu_0}} & \\ & \Lambda_{V_0} \end{pmatrix}^{-1} \begin{pmatrix} \Psi_1\big(\mathbf{x},\boldsymbol{\mu}_0,V_0\big) \\ \mathrm{Vec}\left[\Psi_2\big(\mathbf{x},\boldsymbol{\mu}_0,V_0\big)\right] \end{pmatrix} \tag{5.2.4}$$

where $(\boldsymbol{\mu}_0, V_0)$ are solutions to problem (5.2.1), Ψ_1 and Ψ_2 are functions

$$\Psi_1 = u_2(t)(\mathbf{x}-\boldsymbol{\mu}) \ , \qquad \Lambda_{\boldsymbol{\mu}_0} = \frac{\partial \Psi_1}{\partial \boldsymbol{\mu}}|_{\boldsymbol{\mu}_0,V_0}$$
$$\Psi_2 = u_2(t)\big(\mathbf{x}-\boldsymbol{\mu}\big)\big(\mathbf{x}-\boldsymbol{\mu}\big)' - u_3(t)V \ , \qquad \Lambda_{V_0} = \frac{\partial \mathrm{Vec}\,(\Psi_2)}{\partial \mathrm{Vec}(V)}|_{\boldsymbol{\mu}_0,V_0}$$

The relationship (5.2.4) of influence functions implies that the influence function of scatter V is independent of that of location $\boldsymbol{\mu}$. Hence when the scatter-functional V is the only interest, one can assume, without loss of generality, location is known and is at $\mathbf{0}$.

Lopuhaä [20] showed that the influence function of S-functional is the same as that of corresponding M-functional V satisfying $E_F\left[u_2(t)\mathbf{z}\mathbf{z}' - u_3(t)V\right] = 0$, where $t = \sqrt{\mathbf{z}'V^{-1}\mathbf{z}}$, $u_2(t) = \psi(t)/t$ and $u_3(t) = \left(\psi(t)t - \rho(t) + \epsilon_0\rho\left(\infty\right)\right)/p$

Proposition 5.2.2. *An S-estimate $V(F)$ of definition 5.2.1 equals to λI for some $\lambda > 0$ when $F \sim E_p(\mathbf{0}, I)$, its influence function at F is of the form*

$$\begin{aligned}
\mathrm{IF}(\mathbf{x};V,F) &= \frac{u_2(\sqrt{\mathbf{x}'\mathbf{x}/\lambda})}{2h_1+h_3}\mathbf{x}\mathbf{x}' - \frac{\left(\frac{h_1-h_2}{2h_1+h_3}\right)u_2(\sqrt{\mathbf{x}'\mathbf{x}/\lambda})(\mathbf{x}'\mathbf{x}) + \lambda u_3(\sqrt{\mathbf{x}'\mathbf{x}/\lambda})}{(2h_1+h_3)+p(h_1-h_2)} I_p \\
&= \frac{u_2(\sqrt{\mathbf{x}'\mathbf{x}/\lambda})}{2h_1+h_3}\mathbf{x}\mathbf{x}' - \left[-\frac{4h_1+h_3}{p(2h_1+h_3)h_3}u_2(\sqrt{\mathbf{x}'\mathbf{x}/\lambda})(\mathbf{x}'\mathbf{x}) + \frac{2\lambda}{h_3}u_3(\sqrt{\mathbf{x}'\mathbf{x}/\lambda})\right] I_p \\
&= \frac{\psi(\sqrt{\mathbf{x}'\mathbf{x}/\lambda})}{(2h_1+h_3)\sqrt{\mathbf{x}'\mathbf{x}/\lambda}}\mathbf{x}\mathbf{x}' - \left[\frac{\psi(\sqrt{\mathbf{x}'\mathbf{x}/\lambda})\sqrt{\mathbf{x}'\mathbf{x}/\lambda}}{p(2h_1+h_3)} - \frac{2}{ph_3}\left(\rho(\sqrt{\mathbf{x}'\mathbf{x}/\lambda}) - b_0\right)\right]\lambda I_p \\
&= \frac{u_2(\sqrt{\mathbf{x}'\mathbf{x}/\lambda})}{2h_1+h_3}\left(\mathbf{x}\mathbf{x}' - \frac{\mathbf{x}'\mathbf{x}}{p}I\right) + \frac{2\lambda}{ph_3}\left(\rho(\sqrt{\mathbf{x}'\mathbf{x}/\lambda}) - \epsilon_0\rho(\infty)\right) I_p
\end{aligned}$$

where

$$\begin{aligned}
h_1 &= \frac{1}{2p(p+2)} E_F\left[u_2'(\sqrt{\mathbf{z}'\mathbf{z}/\lambda})(\sqrt{\mathbf{z}'\mathbf{z}/\lambda})^3\right] \\
&= \frac{1}{2p(p+2)} E_F\left[\psi'(\sqrt{\mathbf{z}'\mathbf{z}/\lambda})\left(\mathbf{z}'\mathbf{z}/\lambda\right) - \psi(\sqrt{\mathbf{z}'\mathbf{z}/\lambda})\sqrt{\mathbf{z}'\mathbf{z}/\lambda}\right] \\
h_2 &= \frac{1}{2p} E_F\left[u_3'(\sqrt{\mathbf{z}'\mathbf{z}/\lambda})(\sqrt{\mathbf{z}'\mathbf{z}/\lambda})\right] = \frac{1}{2p^2} E_F\left[\psi'(\sqrt{\mathbf{z}'\mathbf{z}/\lambda})\left(\mathbf{z}'\mathbf{z}/\lambda\right)\right] \\
h_3 &= E_F\left[u_3(\sqrt{\mathbf{z}'\mathbf{z}/\lambda})\right] = \frac{1}{p} E_F\left[\psi(\sqrt{\mathbf{z}'\mathbf{z}/\lambda})\sqrt{\mathbf{z}'\mathbf{z}/\lambda}\right]
\end{aligned}$$

The influence function deduced by Lopuhaä is presented as a function of $t = \sqrt{(\mathbf{x}-\boldsymbol{\mu})V^{-1}(\mathbf{x}-\boldsymbol{\mu})}$. This representation makes subsequent computation surprisingly complicated in the scatter-only problem, and unnecessarily confusing as the ρ functions that define S-estimates are commonly denoted in terms of $s = t^2$. A representation of $\mathrm{IF}(\mathbf{x}, V, F)$ as a function of s is much needed especially when S-estimates based on empirical distribution are to compute. Let $u(s) = 2\rho'(s)$, and implementing the relationship

$$\begin{aligned}
\psi(t) &= u\left(t^2\right)t & \qquad \psi'(t) &= u(s) + 2u'(s)s \\
\psi(t)t &= u(s)s & \qquad \psi'(t)t^2 &= u(s)s + 2u'(s)s^2
\end{aligned}$$

the influence function of scatter V can be represented equivalently in terms of s,

Proposition 5.2.3. *An S-functional V of definition 5.2.1 has influence function*

$$\begin{aligned}
\mathrm{IF}(\mathbf{x}; V, F) &= \frac{u(\mathbf{x}'\mathbf{x}/\lambda)}{(2h_1+h_3)}\mathbf{x}\mathbf{x}' - \left[\frac{u\left(\mathbf{x}'\mathbf{x}/\lambda\right)\left(\mathbf{x}'\mathbf{x}/\lambda\right)}{p(2h_1+h_3)} - \frac{2}{ph_3}\Big(\rho(\mathbf{x}'\mathbf{x}/\lambda) - \epsilon_0\rho(\infty)\Big)\right]\lambda I_p \\
&= \frac{u(\mathbf{x}'\mathbf{x}/\lambda)}{2h_1+h_3}\left(\mathbf{x}\mathbf{x}' - \frac{\mathbf{x}'\mathbf{x}}{p}I\right) + \frac{2\lambda}{ph_3}\Big(\rho(\mathbf{x}'\mathbf{x}/\lambda) - \epsilon_0\rho(\infty)\Big)I_p
\end{aligned}$$

where λ is necessarily a solution to equation $E_F\rho\left(\mathbf{z}'\mathbf{z}/\lambda\right) = \epsilon_0\rho(\infty)$, and

$$\begin{aligned}
h_1 &= \frac{1}{p(p+2)} E_F\left[u'(\mathbf{z}'\mathbf{z}/\lambda)(\mathbf{z}'\mathbf{z}/\lambda)^2\right] \\
h_3 &= \frac{1}{p} E_F\left[u(\mathbf{z}'\mathbf{z}/\lambda)\left(\mathbf{z}'\mathbf{z}/\lambda\right)\right]
\end{aligned}$$

Again the asymptotic of an S-estimate of scatter on a mixture of two elliptical distributions will be investigated under the platform on a sequence of hypotheses defined in (4.2.1) with particular interests at $G(\mathbf{0}, I) = F\left(\mathbf{0}, \nu^2 I\right)$, i.e.

$$\begin{aligned} H_0 &: (1-\epsilon)F\big(\mathbf{0}, I_p\big) + \epsilon F\big(\mathbf{0}, \nu^2 I_p\big) \quad \text{vs} \\ H_{1,n} &: (1-\epsilon)F\big(\mathbf{0}, I_p\big) + \epsilon F\left(\mathbf{0},\, \nu^2\big(I_p + D/\sqrt{n}\big)^{-1}\right) \end{aligned} \tag{5.2.5}$$

Denote by P the distribution of null-hypothesis, an application of Theorem 4.1.5 gives readily the asymptotics of S-estimates of scatter.

Theorem 5.2.4. *Let V be an S-functional of Definition 5.2.1 with influence function defined in Proposition 5.2.3, and $\widehat{V}$ be an affine equivariant estimator of V based on sample. Under sequence of alternative hypotheses $H_{1,n}$ of (5.2.5), $V \to \lambda I$, where λ is the solution to $E_P\left[\rho(\mathbf{x}'\mathbf{x}/\lambda)\right] = \epsilon_0 \rho\big(\infty\big)$, and*

1. $\sqrt{n}\left(\widehat{V} - \lambda I\right) \xrightarrow{d} N(-\epsilon\lambda\,\mathbf{m}, \Omega\,)$, *where* $\mathbf{m} = \alpha D - \beta\,\mathrm{tr}(D) I$, $\Omega = \gamma(I + K_p) + \eta\,\mathrm{Vec}(I)\mathrm{Vec}(I)'$, *and*

$$\begin{aligned}
\alpha &= \frac{1}{2h_1 + h_3}\left\{\frac{E_s\left[u\left(s\nu^2/\lambda\right)\left(s\nu^2/\lambda\right)\right]}{p} + \frac{2E_s\left[u'\left(s\nu^2/\lambda\right)\left(s\nu^2/\lambda\right)^2\right]}{p(p+2)}\right\} \\
\beta &= \frac{2}{p^2(2h_1 + h_3)}\left\{\frac{E_s\left[u'\left(s\nu^2/\lambda\right)\left(s\nu^2/\lambda\right)^2\right]}{p+2} - \frac{h_1\,E_s\left[u\left(s\nu^2/\lambda\right)\left(s\nu^2/\lambda\right)\right]}{h_3}\right\} \\
h_1 &= \frac{1}{p(p+2)}E_s\left[(1-\epsilon)u'\left(s/\lambda\right)\left(s/\lambda\right)^2 + \epsilon u'\left(s\nu^2/\lambda\right)\left(s\nu^2/\lambda\right)^2\right] \\
h_3 &= \frac{1}{p}E_s\left[(1-\epsilon)u\left(s/\lambda\right)\left(s/\lambda\right) + \epsilon u\left(s\nu^2/\lambda\right)\left(s\nu^2/\lambda\right)\right] \\
\gamma &= \frac{1}{(2h_1+h_3)^2 p(p+2)}\left\{(1-\epsilon)E_s\left[u(s/\lambda)s\right]^2 + \epsilon E_s\left[u\left(s\nu^2/\lambda\right)s\nu^2\right]^2\right\} \\
\eta &= -\frac{2\lambda^2}{p}\gamma + \frac{4\lambda^2}{p^2 h_3}\left[(1-\epsilon)E_s\rho^2\big(s/\lambda\big) + \epsilon E_s\rho^2\big(s\nu^2/\lambda\big) - \epsilon_0^2\rho^2\big(\infty\big)\right] \\
s &\sim \frac{\pi^{p/2}}{\Gamma(p/2)} s^{\frac{p}{2}-1} f(s)
\end{aligned}$$

2. *If $\nu^2 = 1$, i.e. $F\big(\mathbf{0}, I_p\big) = G\big(\mathbf{0}, I_p\big)$, then $\sqrt{n}\left(\widehat{V} - \lambda I\right) \xrightarrow{d} N_p^p\left(-\epsilon\lambda\,D, \Omega\,\right)$.*

Corollary 5.2.5. *Assuming notations in Theorem 5.2.4, under the sequence of alternative hypotheses $H_{1,n}$ of (5.2.5),*

1. $n\left[\log\frac{\left(\frac{1}{p}\,\mathrm{tr}\,\widehat{V}\right)^p}{|\widehat{V}|}\right] \xrightarrow{d} \frac{\gamma}{\lambda^2}\chi_q^2(\delta)$, *a non-central Chi-square distribution with degree of freedom* $q = \frac{1}{2}(p+2)(p-1)$ *and non-centrality* $\delta = \frac{(\epsilon\lambda\alpha)^2}{2\gamma}\left[\mathrm{tr}\left(D^2\right) - \frac{1}{p}\mathrm{tr}^2(D)\right]$

2. $\frac{n}{2}\ \log\left[\frac{\mathrm{tr}\,\widehat{V}\cdot\mathrm{tr}\left(\widehat{V}^{-1}\right)}{p^2}\right]^p \xrightarrow{d} \frac{\gamma}{\lambda^2}\,\chi_q^2(\delta)$

5.3 One-Step W-estimates

Sample mean $\bar{\mathbf{x}}$ and sample covariance-matrix S_n are the primary and most commonly employed descriptive statistics of location and scatter in the inference of multivariate analysis. As members of M-estimates, sample mean and sample covariance-matrix are unfortunately having low breakdown point (which is no more than $1/(p+1)$, where p is the dimension of distribution), hence are very sensitive to outliers. To develop high breakdown estimates that serve as better summary statistics in the presence of contamination, Tukey [34] first introduced the weighted sample mean and weighted sample covariance-matrix as the simplest alternatives to $\bar{\mathbf{x}}$ and S_n,

$$\widehat{\boldsymbol{\mu}} = \frac{\sum_i^n u_1(s_{o,i})\mathbf{x}_i}{\sum_i^n u_1(s_{o,i})}, \quad \text{and} \quad \widehat{V} = \frac{\sum_i^n u_2(s_{o,i})(\mathbf{x}_i-\bar{\mathbf{x}})(\mathbf{x}_i-\bar{\mathbf{x}})'}{\sum_i^n u_2(s_{o,i})} \tag{5.3.1}$$

where $s_{o,i} = (\mathbf{x}_i-\bar{\mathbf{x}})'S_n^{-1}(\mathbf{x}_i-\bar{\mathbf{x}})$.

The location and scatter estimates defined in equation (5.3.1) are called One-Step W-estimates as they are in fact the first step of an iterative algorithm for adaptively weighted sample mean and sample covariance-matrix,

$$\widehat{\boldsymbol{\mu}}_{(k+1)} = \frac{\sum_i^n u_1(s_{k,i})\mathbf{x}_i}{\sum_i^n u_1(s_{k,i})}, \quad \text{and} \quad \widehat{V}_{(k+1)} = \frac{\sum_i^n u_2(s_{k,i})\left(\mathbf{x}_i-\widehat{\boldsymbol{\mu}}_{(k)}\right)\left(\mathbf{x}_i-\widehat{\boldsymbol{\mu}}_{(k)}\right)'}{\sum_i^n u_2(s_{k,i})}$$

where $s_{k,i} = \left(\mathbf{x}_i-\widehat{\boldsymbol{\mu}}_{(k)}\right)' V_{(k)}^{-1}\left(\mathbf{x}_i-\widehat{\boldsymbol{\mu}}_{(k)}\right)$.

Because of the computational expensive and inefficient nature of an iterative algorithm, One-Step W-estimates have become and remained popular in the statistical application since their inceptions. In their original definition of (5.3.1), sample mean and sample covariance-matrix are assumed to be the initial location and scatter estimates. By taking any affine-equivariant location $\widehat{\boldsymbol{\mu}}_o$ and scatter $\widehat{V}_o$ statistics as the initial estimates, the concept of One-Step W-estimates can be extended [40],

$$\widehat{\boldsymbol{\mu}} = \frac{\sum_i^n u_1(s_{o,i})\mathbf{x}_i}{\sum_i^n u_1(s_{o,i})}, \quad \text{and} \quad \widehat{V} = \frac{\sum_i^n u_2(s_{o,i})(\mathbf{x}_i-\widehat{\boldsymbol{\mu}}_o)(\mathbf{x}_i-\widehat{\boldsymbol{\mu}}_o)'}{\sum_i^n u_2(s_{o,i})} \tag{5.3.2}$$

where now $s_{o,i} = (\mathbf{x}_i-\widehat{\boldsymbol{\mu}}_o)'V_o^{-1}(\mathbf{x}_i-\widehat{\boldsymbol{\mu}}_o)$.

In a scatter-only problem, one can assume the location is known to be $\mathbf{0}$ and population version of a One-Step W-functional of scatter is defined analogously as replacing the $\sum$ is equation (5.3.2) by the expectation over the distribution F of the population.

Definition 5.3.1. *Given an affine equivariant scatter functional V_1 and a weight function $u(\cdot)$, the* One-Step W-functional *of scatter matrix V_2 is defined as*

$$V_2 = \frac{E_F\left[u\left(\mathbf{z}'V_1^{-1}\mathbf{z}\right)\mathbf{z}'\mathbf{z}\right]}{E_F\left[u\left(\mathbf{z}'V_1^{-1}\mathbf{z}\right)\right]}$$

Proposition 5.3.2. *Let V_1 and V_2 be as in definition 5.3.1. Suppose $V_i = \lambda_i I$ at elliptical $F \sim E_p(\mathbf{0}, I)$, then the influence function of V_2, denoted by* $\mathrm{IF}(\mathbf{z}; V_2, F)$ *is of the form*

$$-\frac{2h_4}{\lambda_1^2 h_1}\mathrm{IF}(\mathbf{z}; V_1, F) + \left(\frac{\lambda_2 h_3 - h_4}{\lambda_1^2 h_1}\right) \operatorname{tr}\left[\mathrm{IF}(\mathbf{z}; V_1, F)\right] I + \frac{u\left(\mathbf{z}'\mathbf{z}/\lambda_1\right)\left(\mathbf{z}\mathbf{z}' - \lambda_2 I\right)}{h_1},$$

where it is necessarily for $\lambda_2 = E_F\left[u(\mathbf{z}'\mathbf{z}/\lambda_1)(\mathbf{z}'\mathbf{z})\right]/(ph_1)$, and

$$\begin{aligned}
h_1 &= E_F\left[u(\mathbf{z}'\mathbf{z}/\lambda_1)\right] \\
h_3 &= E_F\left[u'(\mathbf{z}'\mathbf{z}/\lambda_1)(\mathbf{z}'\mathbf{z})\right]/p \\
h_4 &= E_F\left[u'(\mathbf{z}'\mathbf{z}/\lambda_1)(\mathbf{z}'\mathbf{z})^2\right]/p(p+2)
\end{aligned}$$

The asymptotic distribution of $\sqrt{n}\left(\widehat{V}_2 - V_2\right)$ under the sequence of hypothesis $H_{1,n}$ of (5.2.5) can too be deduced directly from Theorem 4.1.5. But it is more interesting to understand the relationship between the limiting distribution of One-Step estimate $\widehat{V}_2$ and that of the original scatter estimate $\widehat{V}_1$. To this end, let Z_1 be the limiting distribution of $\sqrt{n}\left(\widehat{V}_1 - V_1\right)$ under $H_{1,n}$, and Z_2 be the limiting distribution of $n^{-1/2}\sum_i^n \psi(\mathbf{x}_i)$, where $\psi(\mathbf{x}) = u(\mathbf{x}'\mathbf{x}/\lambda_1)(\mathbf{x}'\mathbf{x} - \lambda_2 I)/h_1$. The contiguity of $H_{1,n}$ to H_0 guarantees that both Z_1 and Z_2 are normally distributed. For brevity, let $\alpha = -2h_4/\lambda_1^2 h_1$ and β be the shorthand of $(\lambda_2 h_3 - h_4)/\lambda_1^2 h_1$. The connection between a scatter matrix and its influence function presented in Lemma (4.1.1) gives

$$\begin{aligned}
&\text{Vec}\left[\sqrt{n}\left(\widehat{V}_2 - V_2\right)\right] = \frac{1}{\sqrt{n}}\sum_i^n \text{Vec}\left[\text{IF}(\mathbf{x}_i; V_2, P)\right] + o_p(1) \\
&= \left(\alpha I_{p^2} + \beta\,\text{Vec}(I)\,[\text{Vec}(I)]'\right)\text{Vec}\left[\frac{1}{\sqrt{n}}\sum_i^n \text{IF}(\mathbf{x}_i; V_1, P)\right] + \text{Vec}\left[\frac{1}{\sqrt{n}}\sum_i^n \psi(\mathbf{x}_i)\right] + o_p(1) \\
&= \left(\alpha I_{p^2} + \beta\,\text{Vec}(I)\,[\text{Vec}(I)]'\,,\, I_{p^2}\right)\text{Vec}\begin{pmatrix} \frac{1}{\sqrt{n}}\sum_i^n \text{IF}(\mathbf{x}_i; V_1, P) = \sqrt{n}\left(\widehat{V}_1 - V_1\right) + o_p(1) \\ \frac{1}{\sqrt{n}}\sum_i^n \psi(\mathbf{x}_i) \end{pmatrix} \\
&\xrightarrow{H_{1,n}} \left(\alpha I_{p^2} + \beta\,\text{Vec}(I)\,[\text{Vec}(I)]'\,,\, I_{p^2}\right)\text{Vec}\begin{pmatrix} Z_1 \\ Z_2 \end{pmatrix} \equiv (A, I_{p^2})\text{Vec}\begin{pmatrix} Z_1 \\ Z_2 \end{pmatrix} \quad \text{say.}
\end{aligned}$$

Following LeCam third Lemma and Theorem 4.1.5, Z_1 and Z_2 are jointly normal

$$\text{Vec}\begin{pmatrix} Z_1 \\ Z_2 \end{pmatrix} = N\left[\text{Vec}\begin{pmatrix} \epsilon\,\mathbf{m}_1 \\ \epsilon\,\mathbf{m}_2 \end{pmatrix}, \begin{pmatrix} \Omega_1 & \Sigma \\ \Sigma & \Omega_2 \end{pmatrix}\right]$$

with P denoting the mixture of $(1-\epsilon)F(\mathbf{0}, I) + \epsilon G(\mathbf{0}, I)$, where $G(\mathbf{0}, I) = F(\mathbf{0}, \nu^2 I)$, the parameters are

$$\begin{aligned}
&\mathbf{m}_1 = E_G\left[\left(\frac{\text{tr}(D)}{2} + \frac{g'(\mathbf{x}'\mathbf{x})}{g(\mathbf{x}'\mathbf{x})}\mathbf{x}'D\mathbf{x}\right)\text{IF}(\mathbf{x}; V_1, P)\right], \quad \mathbf{m}_2 = E_G\left[\left(\frac{\text{tr}(D)}{2} + \frac{g'(\mathbf{x}'\mathbf{x})}{g(\mathbf{x}'\mathbf{x})}\mathbf{x}'D\mathbf{x}\right)\psi(\mathbf{x})\right] \\
&\Omega_1 = E_P\left\{\text{Vec}\left[\text{IF}(\mathbf{x}; V_1, P)\right]\text{Vec}\left[\text{IF}(\mathbf{x}; V_1, P)\right]'\right\}, \qquad \Omega_2 = E_P\left\{\text{Vec}\left[\psi(\mathbf{x})\right]\text{Vec}\left[\psi(\mathbf{x})\right]'\right\} \\
&\Sigma = E_P\left\{\text{Vec}\left[\text{IF}(\mathbf{x}; V_1, P)\right]\text{Vec}\left[\psi(\mathbf{x})\right]'\right\}
\end{aligned}$$

All these results come to the fruition of this assertion

$$\sqrt{n}\,\text{Vec}\left(\widehat{V}_2 - V_2\right) \xrightarrow{H_{1,n}} N_p^p\left(\epsilon A\text{Vec}(\mathbf{m}_1) + \epsilon\text{Vec}(\mathbf{m}_2), A\Omega_1 A + \Omega_2 + 2A\Sigma\right)$$

This facilitates the following theorem,

Theorem 5.3.3. *Let V be an M-functional of definition 5.1.1 or an S-functional of definition 5.2.1 with influence function* $\text{IF}(\mathbf{x}; V, F) = w(\mathbf{x}'\mathbf{x})\mathbf{x}\mathbf{x}' - \pi(\mathbf{x}'\mathbf{x})I$. *Let V_2 be a One-Step W-functional of definition 5.3.1, based on a weight function $u(\cdot)$ and V. Denote by $\widehat{V}_2$ and $\widehat{V}$ respectively the consistent estimates of V_2 and V at empirical distribution. Under sequence of alternative hypotheses $H_{1,n}$ of (5.2.5),* *1.* $\sqrt{n}\left(\widehat{V}_2-\lambda_2 I\right)\xrightarrow{d} N(-\epsilon\mathbf{m},\Omega)$*, where* $\mathbf{m}=\alpha_2\,D+\beta_2\,\mathrm{tr}(D)I$*,* $\Omega=\gamma_2(I+K_p)+\eta_2\,\mathrm{Vec}(I)\mathrm{Vec}(I)'$*, and*

$$\alpha_2=\left(-\lambda_2^*-\frac{2k_4^*-2k_4}{k_1\lambda}\right)$$
$$\beta_2=\left(\frac{\lambda_2 k_3^*-k_4^*}{k_1\lambda}-\frac{\lambda_2 k_3-k_4}{k_1\lambda}\alpha+\frac{p\lambda_2 k_3-(p+2)k_4}{k_1\lambda}\beta\right)$$
$$\gamma_2=\left(\frac{2k_4}{k_1\lambda^2}\right)^2\gamma+\frac{1}{p(p+2)k_1^2}\left\{(1-\epsilon)E_s\left[u^2(s/\lambda)s^2\right]+\epsilon E_s\left[u^2\left(s\nu^2/\lambda\right)\left(s\nu^2\right)^2\right]\right\}$$
$$-\frac{4k_4}{p(p+2)k_1^2\lambda^2}\left\{(1-\epsilon)E_s\left[w(s)u(s/\lambda)s^2\right]+\epsilon E_s\left[w\left(s\nu^2\right)u\left(s\nu^2/\lambda\right)\left(s\nu^2\right)^2\right]\right\}$$

here α*,* β*,* γ *and* λ *are defined as in theorem 5.2.4 or 5.1.3,*

$$k_1=E_s\left[(1-\epsilon)u\left(s/\lambda\right)+\epsilon u\left(s\nu^2/\lambda\right)\right]$$
$$k_3=(1-\epsilon)E_s\left[u'\left(s/\lambda\right)s\right]/p+\epsilon k_3^*,\qquad k_3^*=E_s\left[u'\left(s\nu^2/\lambda\right)s\nu^2\right]/p$$
$$k_4=(1-\epsilon)\frac{E_s\left[u'\left(s/\lambda\right)s^2\right]}{p(p+2)}+\epsilon k_4^*,\qquad k_4^*=\frac{E_s\left[u'\left(s\nu^2/\lambda\right)\left(s\nu^2\right)^2\right]}{p(p+2)}$$
$$\lambda_2=(1-\epsilon)E_s\left[u\left(s/\lambda\right)s\right]/(pk_1)+\epsilon\lambda_2^*,\qquad \lambda_2^*=E_s\left[u\left(s\nu^2/\lambda\right)s\nu^2\right]/(pk_1)$$
$$s\sim\frac{\pi^{p/2}}{\Gamma(p/2)}s^{\frac{p}{2}-1}f(s)$$

2. If $\nu^2=1$*, i.e.* $F\big(\mathbf{0},I_p\big)=G\big(\mathbf{0},I_p\big)$*, then* $\sqrt{n}\left(\widehat{V}_2-\lambda_2 I\right)\xrightarrow{d} N_p^p\left(-\epsilon\lambda_2\,D,\Omega\right)$*.*

5.4 Invariant Scatter Estimates

Almost all scatter estimates appeared in the literature are of affine-equivariance. They are popular due to their nice properties. Given such a scatter functional $V(F)$, the scaled shape functional $\mathbf{h}(V)=pV/\mathrm{tr}(V)$ is apparently not affine-equivariant, because $|\mathbf{h}(V(F))|=1$ for all spherical $F\sim E_p\left(\mathbf{0},\nu^2 I_p\right)$. The corresponding non affine-equivariant scatter estimate $\mathbf{h}(\widehat{V})$, however, has a slight advantage over the affine-equivariant ones in the application of testing hypothesis on the mixtures of elliptical distributions (5.2.5).

The computation of asymptotic mean and covariance matrix of a scatter estimate is enormously intensive (Theorem 4.1.5), involves heavily with the influence function of the scatter-functional which is typically of the form $\omega_1(|\mathbf{z}|)\mathbf{z}\mathbf{z}'-\omega_2(|\mathbf{z}|)I$, and in all known cases ω_2 is a much more complicated function of $|\mathbf{z}|$ than ω_1. The scaled scatter estimate $\mathbf{h}(\widehat{V})$ on the other hand has a simpler influence function than original non-scaled affine-equivariant one.

Proposition 5.4.1. *Let V be an affine equivariant scatter matrix functional with influence function $IF(\mathbf{z}; V, F) = \omega_1(|\mathbf{z}|)\mathbf{z}\mathbf{z}' - \omega_2(|\mathbf{z}|)I$, and having value of λI at spherical distribution $F \sim E(\mathbf{0}, I)$. Then the matrix functional $S = \frac{p}{\operatorname{tr}(V)}V$ has influence function $IF(\mathbf{z}; S, F) = \frac{\omega_1}{\lambda}\left(\mathbf{z}\mathbf{z}' - \frac{(\mathbf{z}'\mathbf{z})}{p}I\right)$.*

Given an equivariant scatter statistic $\widehat{V}$, a corresponding scaled scatter estimate $\widehat{S} = p\widehat{V}/\operatorname{tr}(\widehat{V})$ provides a computationally simple choice for computing its asymptotic distribution. In fact,

Theorem 5.4.2. *Suppose $\sqrt{n}\left(\widehat{V} - V\right)$ is asymptotically normal under the sequence of hypothesis $H_{1,n}$ of (5.2.5), with mean $\mathbf{m} = -\epsilon\lambda\left[\alpha D - \beta \operatorname{tr}(D)I\right]$, and covariance-matrix $\Omega = \gamma(I + K_p) + \eta \operatorname{Vec}(I)\operatorname{Vec}(I)'$*

1. *$\sqrt{n}\left(\widehat{S} - I\right) \xrightarrow{d} N(-\epsilon\,\mathbf{m}^*, \Omega^*)$, where $\Omega^* = \frac{\gamma}{\lambda^2}\left[(I + K_p) - \frac{2}{p}\operatorname{Vec}(I)\operatorname{Vec}(I)'\right]$, and $\mathbf{m}^* = \alpha\left[D - \frac{\operatorname{tr}(D)}{p}I\right]$*
2. *$\frac{n}{2}\log\left[\frac{\operatorname{tr}\widehat{S}\cdot\operatorname{tr}\left(\widehat{S}^{-1}\right)}{p^2}\right]^p \sim n\left[\log\frac{\left(\frac{1}{p}\operatorname{tr}\widehat{S}\right)^p}{|\widehat{S}|}\right] \sim n\left[\frac{\left(\frac{1}{p}\operatorname{tr}\widehat{V}\right)^p}{|\widehat{V}|}\right] \xrightarrow{d} \frac{\gamma}{\lambda^2}\chi^2_q(\delta)$, a non-central Chi-square distribution with degree of freedom $q = \frac{1}{2}(p+2)(p-1)$ and non-centrality $\delta = \frac{(\epsilon\lambda\alpha)^2}{2\gamma}\left[\operatorname{tr}\left(D^2\right) - \frac{1}{p}\operatorname{tr}^2(D)\right]$*

5.5 MM-Estimates

MM estimates of multivariate location and scatter were recently introduced by Tatsuoka and Tyler (2000) to combine the efficiency of M-estimates with the high breakdown properties of S-estimates. In their paper, however, they do not derive the asymptotic distribution of the MM-estimates. Thus the results presented in this section are the first steps to formally address this problem.

Definition 5.5.1. *For a distribution F in $\Re^p$ and an equivariant scale functional $\sigma(F) > 0$, the multivariate location and scatter MM-functionals are defined to be $\mu(F)$*

and $V(F) = \sigma^2(F)\Gamma(F)$ *respectively, where* (μ, Γ) *is any solution which minimizes*

$$E_F\left[\rho\left\{\frac{(\mathbf{x}-\mu)'\Gamma^{-1}(\mathbf{x}-\mu)}{\sigma^2(F)}\right\}\right] \tag{5.5.1}$$

over all $\mu \in \mathbb{R}^p$ *and* $\Gamma \in \mathcal{PDS}(p)$ *with* $\det(\Gamma) = 1$

5.5.1 Relationship to M-functional

The objective function L of the minimization problem (5.5.1) can be constructed as

$$L(\mu, \Gamma) = E_F\left[\rho\left\{\frac{(\mathbf{x}-\mu)'\Gamma^{-1}(\mathbf{x}-\mu)}{\sigma^2(F)}\right\}\right] + \lambda \log(\det(\Gamma))$$

A solution (μ, Γ) of (5.5.1) must satisfy the equations

$$\frac{\partial L}{\partial \mu} = E_F\left[-\frac{u\,(t)}{\sigma^2}\Gamma^{-1}(\mathbf{x}-\mu)\right] = 0 \tag{5.5.2}$$

$$\frac{\partial L}{\partial \Gamma} = E_F\left[-\frac{u\,(t)}{2\sigma^2}\Gamma^{-1}(\mathbf{x}-\mu)(\mathbf{x}-\mu)'\Gamma^{-1}\right] + \lambda\Gamma^{-1} = 0 \tag{5.5.3}$$

where $u(t) = 2\rho'(t)$ and $t = (\mathbf{x}-\mu)'\Gamma^{-1}(\mathbf{x}-\mu)/\sigma^2$. Multiplying (5.5.3) by Γ and taking the trace gives $\lambda = E_F\left(u(t)t\right)/(2p)$. Denote $\psi(t) = u(t)t$, a solution (μ, Γ) of (5.5.1) is also a solution to equations

$$E_F\left[u\,(t)\,(\mathbf{x}-\mu)\right] = 0 \tag{5.5.4}$$

$$E_F\left[u\,(t)\,(\mathbf{x}-\mu)(\mathbf{x}-\mu)'\right] - \frac{\sigma^2}{p}E_F\left[\psi\,(t)\right]\Gamma - \log|\Gamma|\Gamma = 0 \tag{5.5.5}$$

The term $-\log|\Gamma|\Gamma$ is being added to equation (5.5.5) because merely substitute λ into (5.5.3) would only render a system of dependent equations that has no unique solution.

5.5.2 Influence Function of Shape-Functional Γ

The influence function for μ and Γ can be treated separately. Hence Γ with $\det(\Gamma) = 1$ must satisfy $E\Psi(\mathbf{z}, \Gamma) = 0$, where

$$\Psi(\mathbf{z}, \Gamma) = u\left(\frac{\mathbf{z}'\Gamma^{-1}\mathbf{z}}{\sigma^2}\right)\mathbf{z}\mathbf{z}' - \frac{\sigma^2}{p}\psi\left(\frac{\mathbf{z}'\Gamma^{-1}\mathbf{z}}{\sigma^2}\right)\Gamma - \log|\Gamma|\Gamma = 0 \tag{5.5.6}$$

By [12], the influence function of Γ is proportional to Ψ:

$$\mathrm{Vec}\left\{\mathrm{IF}(\mathbf{z}; \Gamma, F)\right\} = \mathbf{M}^{-1}\mathrm{Vec}\{\Psi(\mathbf{z}, \Gamma_0)\}$$

where $\mathbf{M} = -E_F\left[\frac{\partial \text{Vec}\{\Psi(\mathbf{z},\Gamma)\}}{\partial \text{Vec}(\Gamma)}\right]_{\Gamma=\Gamma_0} \in \mathbb{R}^{p^2}_{p^2}$ and Γ_0 is such that $E_F[\Psi(\mathbf{z},\Gamma_0)] = 0$. When $F \sim E(\mathbf{0}, I_p)$, $\Gamma_0 = I$ clearly is a solution, so

$$\mathbf{M} = E\left[\frac{1}{\sigma^2}u'\left(\frac{\mathbf{z}'\mathbf{z}}{\sigma^2}\right)\mathbf{z}\mathbf{z}' \otimes \mathbf{z}\mathbf{z}' - \frac{1}{p}\psi'\left(\frac{\mathbf{z}'\mathbf{z}}{\sigma^2}\right)\text{Vec}(I_p)\text{Vec}(\mathbf{z}\mathbf{z}')' + \frac{\sigma^2}{p}\psi\left(\frac{\mathbf{z}'\mathbf{z}}{\sigma^2}\right)I_{p^2}\right] + \text{Vec}(I_p)\text{Vec}(I_p)$$

Let $\mathbf{u} = \mathbf{z}/|\mathbf{z}|$. Using the identities that

$$E_F\left(\mathbf{u}\mathbf{u}'\right) = p^{-1}I_p \quad \text{and}$$
$$E_F\left(\mathbf{u}\mathbf{u}' \otimes \mathbf{u}\mathbf{u}'\right) = \frac{1}{p(p+2)}\left[\left(I_{p^2} + K_p\right) + \text{Vec}(I_p)\text{Vec}(I_p)'\right],$$

, the closed form of M is obtained as

$$\begin{aligned}
\mathbf{M} &= \frac{\sigma^2}{p(p+2)}E\left[u'\left(\frac{\mathbf{z}'\mathbf{z}}{\sigma^2}\right)\left(\frac{\mathbf{z}'\mathbf{z}}{\sigma^2}\right)^2\right]\left[\left(I_{p^2} + K_p\right) + \text{Vec}(I_p)\text{Vec}(I_p)\right] \\
&- \frac{\sigma^2}{p^2}E\left[\psi'\left(\frac{\mathbf{z}'\mathbf{z}}{\sigma^2}\right)\left(\frac{\mathbf{z}'\mathbf{z}}{\sigma^2}\right)\right]\text{Vec}(I_p)\text{Vec}(I_p)' + \frac{\sigma^2}{p}E\left[\psi\left(\frac{\mathbf{z}'\mathbf{z}}{\sigma^2}\right)\right]I_{p^2} + \text{Vec}(I_p)\text{Vec}(I_p) \\
&= (h_1 + h_3)I_{p^2} + h_1K_p + (h_1 - h_2)\text{Vec}(I_p)\text{Vec}(I_p)'
\end{aligned}$$

where

$$\begin{aligned}
h_1 &= \frac{\sigma^2}{p(p+2)}E\left[u'\left(\frac{\mathbf{z}'\mathbf{z}}{\sigma^2}\right)\left(\frac{\mathbf{z}'\mathbf{z}}{\sigma^2}\right)^2\right] \\
h_2 &= \frac{\sigma^2}{p^2}E\left[\psi'\left(\frac{\mathbf{z}'\mathbf{z}}{\sigma^2}\right)\left(\frac{\mathbf{z}'\mathbf{z}}{\sigma^2}\right)\right] - 1 = \frac{\sigma^2}{p^2}E\left[u'\left(\frac{\mathbf{z}'\mathbf{z}}{\sigma^2}\right)\left(\frac{\mathbf{z}'\mathbf{z}}{\sigma^2}\right)^2 + \psi\left(\frac{\mathbf{z}'\mathbf{z}}{\sigma^2}\right)\right] = \left(1 + \frac{2}{p}\right)h_1 + \frac{h_3}{p} - 1 \\
h_3 &= \frac{\sigma^2}{p}E\left[\psi\left(\frac{\mathbf{z}'\mathbf{z}}{\sigma^2}\right)\right] = \frac{\sigma^2}{p}E\left[u\left(\frac{\mathbf{z}'\mathbf{z}}{\sigma^2}\right)\left(\frac{\mathbf{z}'\mathbf{z}}{\sigma^2}\right)\right]
\end{aligned}$$

It follows,

$$\begin{aligned}
\mathbf{M}^{-1} &= \frac{1}{2h_1 + h_3}\left[\frac{h_1 + h_3}{h_3}I_{p^2} - \frac{h_1}{h_3}K_p - \frac{h_1 - h_2}{(2h_1 + h_3) + p(h_1 - h_2)}\text{Vec}(I_p)\text{Vec}(I_p)'\right] \\
&= \frac{1}{2h_1 + h_3}\left[\frac{h_1 + h_3}{h_3}I_{p^2} - \frac{h_1}{h_3}K_p - \left(\frac{1}{p} - \frac{2h_1 + h_3}{p^2}\right)\text{Vec}(I_p)\text{Vec}(I_p)'\right],
\end{aligned}$$

this gives,

$$\begin{aligned}
\text{Vec}\left\{\text{IF}(\mathbf{z};\Gamma,F)\right\} &= \mathbf{M}^{-1}\left\{u\left(\frac{\mathbf{z}'\mathbf{z}}{\sigma^2}\right)\text{Vec}(\mathbf{z}\mathbf{z}') - \frac{\sigma^2}{p}\psi\left(\frac{\mathbf{z}'\mathbf{z}}{\sigma^2}\right)\text{Vec}(I_p)\right\} \\
&= \frac{1}{2h_1 + h_3}\left[u\left(\frac{\mathbf{z}'\mathbf{z}}{\sigma^2}\right)\text{Vec}(\mathbf{z}\mathbf{z}') - \frac{1}{p}u\left(\frac{\mathbf{z}'\mathbf{z}}{\sigma^2}\right)(\mathbf{z}'\mathbf{z})\text{Vec}(I)\right]
\end{aligned}$$

which in turn gives

$$\text{IF}(\mathbf{z};\Gamma,F) = \frac{u\left(\mathbf{z}'\mathbf{z}/\sigma^2\right)}{2h_1 + h_3}\left(\mathbf{z}\mathbf{z}' - \frac{\mathbf{z}'\mathbf{z}}{p}I\right)$$

The MM-Scatter functional $V(F) = \sigma^2(F)\Gamma(F)$ satisfies equation (5.5.5), is thus an affine equivariant with influence function $\text{IF}(\mathbf{z};V,F) = \sigma^2\text{IF}(\mathbf{z};\Gamma,F)$

5.5.3 Asymptotic Distribution on Mixture of Elliptical Distributions

For the purpose of consistency and easy comparison, the characteristics of an MM-estimate of scatter matrix on an asymmetric non-elliptical will be examined asymptotically on a mixture of two elliptical distributions in the context of a sequence of hypotheses defined in (4.2.1), i.e.

$$\begin{aligned} &H_0 : (1-\epsilon)F\big(\mathbf{0}, I_p\big) + \epsilon G\big(\mathbf{0}, \nu^2 I_p\big) \quad \text{vs} \\ &H_{1,n} : (1-\epsilon)F\big(\mathbf{0}, I_p\big) + \epsilon G\left(\mathbf{0},\, \nu^2\big(I_p + D/\sqrt{n}\big)^{-1}\right) \end{aligned} \tag{5.5.7}$$

Denote by P the distribution of null-hypothesis, when interests are particularly at $G(\mathbf{0}, I) = F\left(\mathbf{0}, \nu^2 I\right)$, the asymptotics of an MM-estimate is a straightforward application of Theorem 4.1.5.

Theorem 5.5.2. *In particular if $G(\mathbf{0}, I) = F(\mathbf{0}, bI)$, then P is the mixture of $(1-\epsilon)F(\mathbf{0}, I) + \epsilon\, F(\mathbf{0}, bI)$, let $s \sim \frac{\pi^{p/2}}{\Gamma(p/2)} s^{\frac{p}{2}-1} f(s)$, then under $H_{1,n}$,*

1. $\sqrt{n}\left(\widehat{V} - \sigma^2 I\right) \xrightarrow{d} N\left[-\epsilon\sigma^2\alpha\left(D - \frac{\operatorname{tr} D}{p} I\right),\, \gamma(I + K_p) - \frac{2\gamma}{p}\operatorname{Vec}(I)[\operatorname{Vec}(I)]'\right]$, *where*

$$\begin{aligned}
\alpha &= \frac{1}{2h_1 + h_3}\left\{ \frac{E_s\left[u\left(bs/\sigma^2\right)\left(bs/\sigma^2\right)\right]}{p} + \frac{2E_s\left[u'\left(bs/\sigma^2\right)\left(bs/\sigma^2\right)^2\right]}{p(p+2)} \right\} \\
\gamma &= \frac{1}{p(p+2)(2h_1+h_3)^2}\left\{(1-\epsilon)E_s\left[u^2\left(s/\sigma^2\right)s^2\right] + \epsilon E_s\left[u^2\left(bs/\sigma^2\right)(bs)^2\right]\right\} \\
h_1 &= \frac{1}{p(p+2)}E_s\left[(1-\epsilon)u'\left(s/\sigma^2\right)\left(s/\sigma^2\right)^2 + \epsilon u'\left(bs/\sigma^2\right)\left(bs/\sigma^2\right)^2\right] \\
h_3 &= \frac{1}{p}E_s\left[(1-\epsilon)u\left(s/\sigma^2\right)\left(s/\sigma^2\right) + \epsilon u\left(bs/\sigma^2\right)\left(bs/\sigma^2\right)\right]
\end{aligned}$$

2. $$\begin{aligned} & n\left(1 - \frac{|\widehat{V}|}{\left(\frac{1}{p}\operatorname{tr}\widehat{V}\right)^p}\right) \sim n\left(\log\frac{\left(\frac{1}{p}\operatorname{tr}\widehat{V}\right)^p}{|\widehat{\Gamma}|}\right) \sim \frac{n}{2}\log\left(\frac{\operatorname{tr}\widehat{V}}{p}\cdot\frac{\operatorname{tr}(\widehat{\Gamma}^{-1})}{p}\right)^p \sim \\ & n\left(1 - \frac{|\widehat{\Gamma}|}{\left(\frac{1}{p}\operatorname{tr}\widehat{\Gamma}\right)^p}\right) \sim n\left(\log\frac{\left(\frac{1}{p}\operatorname{tr}\widehat{\Gamma}\right)^p}{|\widehat{\Gamma}|}\right) \sim \frac{n}{2}\log\left(\frac{\operatorname{tr}\widehat{\Gamma}}{p}\cdot\frac{\operatorname{tr}(\widehat{\Gamma}^{-1})}{p}\right)^p \xrightarrow{d} \frac{\gamma}{(\sigma^2)^2}\chi_q^2(\delta), \end{aligned}$$
   where $\delta = \frac{\left(\epsilon\sigma^2\alpha\right)^2}{2\gamma}\left(\operatorname{tr}(D^2) - \frac{1}{p}\operatorname{tr}^2(D)\right)$, *and* $q = \frac{1}{2}(p+2)(p-1)$

5.5.4 Efficiency at Mixture of Elliptical Distributions

In this section, we will investigate the efficiency of an MM-estimate at a mixture of elliptical distributions $(1-\epsilon)F(\mathbf{0}, \Sigma) + \epsilon F(\mathbf{0}, b\Sigma)$.

By its very construction, an MM-functional $V = \sigma^2\Gamma$ is naturally separated into what Kent and Tyler (1996) call "shape" and "scale" components. A shape component of a matrix V is any function $H(V)$ which is invariant under a positive scalar multiple, $H(\lambda V) = H(V)$, $\lambda > 0$; and a scale component is any equivariant function under the same transformation. Kent and Tyler (1996) argue that scale component is an ill-defined nuisance parameter, hence investigation of a scatter estimate should be focused only on its shape component. Tyler (1983) was the first to give an explicit formula of the asymptotic covariance matrix of a shape component $H(\widehat{V})$

$$\gamma\, h'(V)(I + K_p)\,(V \otimes V)\,[h'(V)]' \tag{5.5.8}$$

where γ is the only scalar whose value depends on the ρ function that defines scatter $\widehat{V}$. Motivated by this singleton form of the variance of a shape component along with other interesting arguments, Kent and Tyler (1996) suggest a convenient way to compare the asymptotic relative efficiency of scatter estimates. That is to simply look at the variance-matrix of its shape component and compare the corresponding values of γ for varying choices of ρ function.

From theorem 5.5.2 and using the argument of affine equivariance, the asymptotic covariance matrix of the shape component an MM-estimate is in the form of

$$\gamma\,(I + K_p)\,(\Sigma \otimes \Sigma) - \frac{2\gamma}{p}\mathrm{Vec}(\Sigma)\mathrm{Vec}(\Sigma)' \tag{5.5.9}$$

which in fact belongs to type (5.5.8) with $h(V) = V/\,(p|V|)$. Subsequently, the ARE's of MM-estimates will be examined through γ in conformity with Kent and Tyler(1996), and with Lopuhaä (1999).

Huber Estimate

The Huber weight function for a scatter-matrix estimate is given as

$$u_c(s) = \begin{cases} 1 & (s < c) \\ c/s & (s \geq c) \end{cases}$$

which implies the ρ-function is of the form,

$$\rho_c(s) = c\rho(s/c) = \frac{1}{2}\begin{cases} s & (s < c) \\ c\log(s/c) + c & (s \geq c) \end{cases}$$

and the derivative is simply $u_c'(s) = -c/s^2,\ s \geq c$.

Consider the values of the asymptotic variance coefficient γ of Huber MM-estimates of scatter as a function of tuning constant c. The graphs of this function are plotted in Figure 5.1 for Normal-distributions and in Figures 5.2 & 5.3 for multivariate T-distribution $t_{v,p}$. Since $\gamma(c)$ is confounded with the scale-functional σ^2, without loss of generosity, σ is set to be 1. Note that the limiting case of Huber estimate, $c \to \infty$, corresponds to the sample covariance matrix, whose γ under $t_{v,p}$ converges to $(v-2)/(v-4)$. This claim comes from Theorem 5.5.2, that $u(s) = 1$ for all s gives $s/p \sim F_{p,v}$, hence

$$\gamma(\infty) = \frac{1}{p(p+2)}\frac{Es^2}{\left(\frac{1}{p}Es\right)^2} = \left(\frac{p}{p+2}\right)\frac{\frac{(p+2)v^2}{p(v-2)(v-4)}}{\frac{v^2}{(v-2)^2}} = \frac{v-2}{v-4}$$

On the other end, the limiting case of $c \to 0$ has asymptotic $\gamma = (p+2)/p$. It follows immediately that the Huber estimates do not attain its minimum variance at $c = \infty$ under $t_{v,p}$ distribution.

Tukey's Biweight Estimate

Tukey's Biweight estimates are defined by

$$\rho_c(s) = c\rho\left(s/c\right) = \begin{cases} \frac{s}{2} - \frac{s^2}{2c} + \frac{s^3}{6c^2}, & 0 < s \leq c \\ \frac{c}{6}, & s > c \end{cases}$$

It follows

$$u_c(s) = 2\rho_c'\left(s\right) = \left(1 - \frac{s}{c}\right)^2 \mathrm{I}_{[0 \leq s \leq c]}$$
$$u_c'\left(s\right) = \frac{2}{c}\left(\frac{s}{c} - 1\right)\mathrm{I}_{[0 \leq s \leq c]}$$

As a function of tuning constant c, values of asymptotic variance coefficient $\gamma(c)$ of Tukey's MM-estimates are plotted in Figure 5.4 for Normal (or Normal mixture) distributions and in Figure 5.5 for T-distribution $t_{v,p}$ with degree of freedom 1 and 3 on different dimensions. Vigorous computations have shown that γ does not attain its minimum at $c \to \infty$ or maximum at $c \to 0$ under non-normal model.

Welsh's Estimate

Welsh ρ-function is of the form $\rho_c(s) = c\rho(s/c) = \frac{c}{2}\left(1 - e^{-s/c}\right)$. The defining weight functions of a Welsh's scatter-estimate are given as

$$\begin{aligned} u_c\left(s\right) &= 2\rho_c'\left(s\right) = e^{-\frac{s}{c}} \\ u_c'\left(s\right) &= -\frac{1}{c}e^{-\frac{s}{c}} \end{aligned}$$

When the underlying (H_0) distribution is a mixture of two Normals, i.e.$P = (1 - \epsilon)N(\mathbf{0}, I_p) + \epsilon N(\mathbf{0}, bI_p)$, the scalar α of asymptotic mean and γ of asymptotic variance have closed forms,

$$\begin{aligned} \alpha(c) &= \frac{1}{(1-\epsilon)\alpha^* + \epsilon}, \quad \text{where} \quad \alpha^* = \frac{1}{b}\left(\frac{c\sigma^2 + 2b}{c\sigma^2 + 2}\right)^{\frac{p}{2}+2} \\ \gamma(c) &= \left(\sigma^2\right)^2 \frac{\frac{1-\epsilon}{(c\sigma^2+4)^{\frac{p}{2}+2}} + \frac{\epsilon b^2}{(c\sigma^2+4b)^{\frac{p}{2}+2}}}{\left[\frac{1-\epsilon}{(c\sigma^2+2)^{\frac{p}{2}+2}} + \frac{\epsilon b}{(c\sigma^2+2b)^{\frac{p}{2}+2}}\right]^2 (c\sigma^2)^{\frac{p}{2}+2}} \end{aligned}$$

Under a multivariate normal distribution $N_p(\mathbf{0}, I)$, Welsh's estimate has its minimum variance coefficient γ at $c \to +\infty$ which corresponds to the sample variance-covariance matrix. Under a mixture of Normal distributions, however a Welsh's estimate can attain its minimum variance anywhere of the tuning constant c depending the structure of the underlying mixture (Figure 5.6).

5.5.5 When Scale Parameter σ Is Unknown

Intuitively, one would substitute with an consistent estimate $\widehat{s}$ when the auxiliary scale parameter σ is unknown. In fact it is a conjecture that the asymptotics of MM-estimates deduced in section 5.5.3 will remain unchanged when $\widehat{s}$ replaces σ, if the following conditions are met

1. The underlying distribution $F_{\mathbf{x}}$ is symmetric

2. The scale estimate $\widehat{s}$ satisfies the equivariant property of (1.2.2)

3. $\widehat{s}$ converges at $\sqrt{n}$ -rate to $\sigma(F_{\mathbf{x}})$

An easy way to produce such a scale estimate is to set $\widehat{s} = |\widehat{V}_o|^{1/2p}$, where $\widehat{V}_o$ is an affine equivariant scatter estimate with high breakdown points. The proof of this conjecture is to be completed and presented in a subsequent paper.

Some important remarks have to be made before closing on this chapter. It is a well known fact that the efficiency of a scatter estimate improves as the dimension of data increases, therefore the variance coefficient γ of a scatter must be smaller in higher dimensions. The plots of values of γ in Appendix B, however seem disagree with this principal in that γ looks like increasing as p getting larger. A simple explanation is this, the horizontal axis on each of those graphs are not the tuning constant c, rather it is $c_o = c\sigma(F_{\mathbf{x}})$. Due to the very construction of an MM-estimate, the scale functional σ is confounded with the tuning constant c from weight function. For the simplicity of computation, the values of γ are calculated assuming $\sigma(F_{\mathbf{x}}) = 1$ for varying c. In reality, σ depends on underlying distribution $F_{\mathbf{x}}$ and is neither equal to 1 unless at a Normal model nor constant for different dimensional spaces. Thus, to make an accurate inference on γ across dimension p, one has to factor in the value of σ for a particular model.

5.6 Appendix A: Proofs of Theorems and Corollaries

Proof of Proposition 5.1.2

Proof. In addition to its representation, $IF(\mathbf{z}; V, F) = \omega_1 (|\mathbf{z}|)\mathbf{z}\mathbf{z}' - \omega_2(|\mathbf{z}|)\, I_p$, given by Lemma 4.1.1, the influence function of an M-functional has a unique property that it is always proportional to ψ (page 230 in Hampel [12]):

$$\operatorname{Vec}\{IF(\mathbf{z}; V, F)\} \; = \mathbf{M}^{-1} \operatorname{Vec}\{\psi(\mathbf{z}, V_0)\}$$

where V_0 is such $E\psi(\mathbf{z}, V_0) = 0$, and $\mathbf{M} = -E\left[\frac{\partial \operatorname{Vec}\{\psi(\mathbf{z},V)\}}{\partial \operatorname{Vec}(V)}\right]_{V=V_0} \in \mathbb{R}^{p^2}_{p^2}$. Since $V_0 = \lambda I_p$ under F,

$$\mathbf{M} = E\left\{\mu'\left(\frac{\mathbf{z}'\mathbf{z}}{\lambda}\right) \operatorname{Vec}\left(\frac{\mathbf{z}\mathbf{z}'}{\lambda}\right)\left[\operatorname{Vec}\left(\frac{\mathbf{z}\mathbf{z}'}{\lambda}\right)\right]'\right\} + I_{p^2} = h\,\left(I_{p^2} + K_p\right) + h \operatorname{Vec}(I_p)\,[\operatorname{Vec}(I_p)]' + I_{p^2}$$

where the close form of h is given by Tyler [35]. It follows,

$$\mathbf{M}^{-1} = \frac{1}{2h+1}\left[(h+1)\, I_{p^2} - h\, K_p - \frac{h}{(p+2)h+1} \operatorname{Vec}(I_p)\,[\operatorname{Vec}(I_p)]'\right]$$

which in turn gives the result. □

Proof of Theorem 5.1.3

Proof. Let $\mathbf{z} \sim F\big(\mathbf{0}, I\big)$, and $\mathbf{x} = \sigma\mathbf{z}$. Then $\mathbf{x} \sim G\big(\mathbf{0}, I\big)\;\, F\big(\mathbf{0}, \sigma^2 I\big)$. The asymptotic mean of $\sqrt{n}\left(\widehat{V} - V\right)$ is defined in Theorem 4.1.5, which equals to

$$E_G\left[\left(\frac{\operatorname{tr}(D)}{2} + \frac{g'(\mathbf{x}'\mathbf{x})}{g(\mathbf{x}'\mathbf{x})}\mathbf{x}'D\mathbf{x}\right)\, \operatorname{IF}(\mathbf{x}; V, P)\right]$$

The influence function of an M-functional is given by Proposition 5.1.2

$$IF\,(\mathbf{x}; V, F) = \frac{u\,(\mathbf{x}'\mathbf{x}/\lambda)}{2h+1}\left(\mathbf{x}\mathbf{x}' - \frac{\mathbf{x}'\mathbf{x}}{p} I\right) + \frac{u\,(\mathbf{x}'\mathbf{x}/\lambda)\,\mathbf{x}'\mathbf{x}/p - \lambda}{(p+2)h+1} I$$

where $h = E\left(u'(\mathbf{x}'\mathbf{x}/\lambda)(\mathbf{x}'\mathbf{x}/\lambda)^2\right)/p(p{+}2)$, λ is the solution to equation $E\big(u\,(\mathbf{x}'\mathbf{x}/\lambda)\,\mathbf{x}'\mathbf{x}/\lambda\big) = p$, and the expectation is taken over the null-distribution $P \sim (1{-}\epsilon)F\big(\mathbf{0}, I\big){+}\epsilon F\big(\mathbf{0}, \sigma^2 I\big)$.

It is well known that when $\mathbf{z} \sim F\big(\mathbf{0}, I\big)$, the random radius $s = \mathbf{z}'\mathbf{z}$ is independent of $\mathbf{u} = \mathbf{z}/s$, and $\mathbf{u}$ is uniformly distributed on S^{p-1} with identities $E_{\mathbf{u}}\big[\mathbf{u}\mathbf{u}'D\mathbf{u}\mathbf{u}'\big] = \big[2D + \operatorname{tr}(D)I\big]/\big[p(p{+}2)\big]$ and $E_{\mathbf{u}}\big[\mathbf{u}'D\mathbf{u}\big] = \operatorname{tr}(D)/p$. Also notice that change-of-variable in the integration gives $E_G\big[g'(\mathbf{x}'\mathbf{x})/g(\mathbf{x}'\mathbf{x})\big] = E_s\big[f'(s)/f(s)\big]/\sigma^2$ and $E_s\big[f'(s)/f(s)s\big] = -p/2$. With the aid of these properties, one can carry out the tedious calculations.

$$
\begin{aligned}
A_1 &= \frac{\operatorname{tr}(D)}{2} E_G\left[\frac{u\left(\mathbf{x}'\mathbf{x}/\lambda\right)}{2h+1}\left(\mathbf{x}\mathbf{x}' - \frac{\mathbf{x}'\mathbf{x}}{p}I\right)\right] = 0 \\
A_2 &= \frac{\operatorname{tr}(D)}{2} E_G\left[\frac{u\left(\mathbf{x}'\mathbf{x}/\lambda\right)\mathbf{x}'\mathbf{x}/p - \lambda}{(p+2)h+1} I\right] = \frac{\operatorname{tr}(D)}{2\big[(p+2)h+1\big]} E_s\left[u\left(s\sigma^2/\lambda\right)s\sigma^2/p - \lambda\right] I \\
A_3 &= E_G\left[\frac{g'(\mathbf{x}'\mathbf{x})}{g(\mathbf{x}'\mathbf{x})}\left(\mathbf{x}'D\mathbf{x}\right)\frac{u\left(\mathbf{x}'\mathbf{x}/\lambda\right)}{2h+1}\left(\mathbf{x}\mathbf{x}' - \frac{\mathbf{x}'\mathbf{x}}{p}I\right)\right] \\
&= \frac{1}{2h+1} E_s\left[\frac{f'(s)}{f(s)} u\left(s\sigma^2/\lambda\right)s^2\sigma^2\right] E_{\mathbf{u}}\left[\left(\mathbf{u}\mathbf{u}\right)' D\left(\mathbf{u}\mathbf{u}'\right) - \frac{1}{p}\left(\mathbf{u}'D\mathbf{u}\right) I\right] \\
&= \frac{2}{p(p+2)(2h+1)} E_s\left[\frac{f'(s)}{f(s)} u\left(s\sigma^2/\lambda\right)s^2\sigma^2\right]\left[D - \frac{\operatorname{tr}(D)}{p} I\right] \\
A_4 &= E_G\left[\frac{g'(\mathbf{x}'\mathbf{x})}{g(\mathbf{x}'\mathbf{x})}\left(\mathbf{x}'D\mathbf{x}\right)\left(\frac{u\left(\mathbf{x}'\mathbf{x}/\lambda\right)\mathbf{x}'\mathbf{x}/p - \lambda}{(p+2)h+1}\right)\right] I \\
&= \frac{1}{p\big[(p+2)h+1\big]} E_s\left[\frac{f'(s)}{f(s)}\left(\frac{u\left(s\sigma^2/\lambda\right)s^2\sigma^2}{p} - \lambda s\right)\right] E_{\mathbf{u}}\big[\mathbf{u}'D\mathbf{u}\big] I \\
&= \frac{\operatorname{tr}(D)}{p^2\big[(p+2)h+1\big]} E_s\left[\frac{f'(s)}{f(s)} u\left(s\sigma^2/\lambda\right)s^2\sigma^2\right] I + \frac{\lambda \operatorname{tr}(D)}{2\big[(p+2)h+1\big]} I
\end{aligned}
$$

The sum of $A_i's$ is the asymptotic mean, which can be simplified by noting

$$
E_s\left[\frac{f'(s)}{f(s)} u\left(s\sigma^2/\lambda\right)s^2\sigma^2\right] = -\frac{1}{\lambda} E_s\left[u'\left(s\sigma^2/\lambda\right)\left(s\sigma^2\right)^2\right] - \frac{p+2}{2} E_s\left[u\left(s\sigma^2/\lambda\right)\left(s\sigma^2\right)\right]
$$

□

Proof of Corollary 5.1.4

Proof. First let $\mathbf{h}(V) = \log\left[\frac{\left(\frac{1}{p}\operatorname{tr} V\right)^p}{|V|}\right] = p\log\left(\operatorname{tr} V\right) - \log|V| - p\log(p)$. Denote $(\partial\mathbf{a}/\partial\mathbf{b})$ as the matrix of partial derivatives $(\partial a_i/\partial b_j)$, where i varies over rows and j runs over columns. Taking into account of the symmetry of V, the first and second order derivatives of $\mathbf{h}(V)$ with respect to V are conventionally defined as

$$
\begin{aligned}
\mathbf{h}^{(1)}(V) &= \frac{\partial\,\mathbf{h}(V)}{\partial\operatorname{Vec}(V)}\frac{1}{2}(I + J_p) \in \mathbb{R}^1_{p^2} \\
\mathbf{h}^{(2)}(V) &= \frac{1}{2}(I + J_p)\left(\frac{\partial\left[\mathbf{h}^{(1)}(V)\right]'}{\partial\operatorname{Vec}(V)}\right)' \in \mathbb{R}^{p^2}_{p^2}
\end{aligned}
$$

where $J_p = \sum_{i=1}^p \mathbf{e}_i\mathbf{e}_i' \otimes \mathbf{e}_i\mathbf{e}_i'$ and $\mathbf{e}_i \in \mathbb{R}^p$ is a vector of zero but a 1 in position i.

Applying the identities that

$$\frac{\partial |V|}{\partial \mathrm{Vec}(V)} \frac{1}{2}(I + J_p) = |V|[\mathrm{Vec}(V^{-1})]'$$
$$\frac{\mathrm{tr}\, V}{\partial \mathrm{Vec}(V)} \frac{1}{2}(I + J_p) = [\mathrm{Vec}(I)]'$$

to the calculations of the first derivative,

$$\mathbf{h}^{(1)}(V) = \frac{p}{\mathrm{tr}\, V} [\mathrm{Vec}(I)]' - \left[\mathrm{Vec}(V^{-1})\right]'$$

Let V_{ij} be the ij-th element of V. The identity $\partial V^{-1}/\partial V_{ij} = -V^{-1}\left(\partial V/\partial V_{ij}\right) V^{-1}$ can be extended to obtain

$$\begin{aligned} \frac{\partial \mathrm{Vec}(V^{-1})}{\partial \mathrm{Vec}(V)} &= -\left(V^{-1} \otimes V^{-1}\right) \left[\frac{d\,\mathrm{Vec}(V)}{d\,V_{11}}, \frac{d\mathrm{Vec}(V)}{d\,V_{12}}, \cdots, \frac{d\mathrm{Vec}(V)}{d\,V_{pp}}\right] \\ &= -\left(V^{-1} \otimes V^{-1}\right)(I + K_p - J_p) \end{aligned}$$

With the aid of $(I + K_p - J_p)\,\frac{1}{2}(I + J_p) = \frac{1}{2}(I + K_p)$,

$$\mathbf{h}^{(2)}(V) = \frac{1}{2}\left(I + K_p\right)\left(V^{-1} \otimes V^{-1}\right) - \frac{p}{(\mathrm{tr}\, V)^2} \mathrm{Vec}(I)\,[\mathrm{Vec}(I)]'$$

Interests are laid in the evaluations of the derivatives at $V = \lambda I$ when $H_{1,n}$ reaches its limit (i.e. when the null hypothesis is true). This gives $\mathbf{h}^{(1)}(V) = \mathbf{0}$, $\mathbf{h}^{(2)}(V) = \left[\frac{1}{2}(I + K_p) - \frac{1}{p}\mathrm{Vec}(I)\mathrm{Vec}(I)'\right]/\lambda^2$. The Taylor expansion of $\mathbf{h}(\widehat{V})$ at $V = \lambda I$ hence becomes

$$\mathbf{h}(\widehat{V}) - \mathbf{h}(V) = \left[\mathrm{Vec}\left(\widehat{V} - V\right)\right]' \left(\frac{1}{2}\mathbf{h}^{(2)}(V)\right) \left[\mathrm{Vec}\left(\widehat{V} - V\right)\right] + o_p(n^{-1})$$

Denote Z the limiting distribution of $\sqrt{n}\left[\mathrm{Vec}\left(\widehat{V} - V\right)\right]$. It follows,

$$n \log \left[\frac{\left(\frac{1}{p}\,\mathrm{tr}\, V\right)^p}{|V|}\right] = n\left[\mathbf{h}(\widehat{V}) - \mathbf{h}(V)\right] \xrightarrow{d} \frac{\gamma}{\lambda^2} Z' \left[\frac{\lambda^2}{2\gamma}\mathbf{h}^{(2)}(V)\right] Z = \frac{\gamma}{\lambda^2} Z'BZ \text{ say.}$$

From Theorem 5.1.3, Z has asymptotic mean $\boldsymbol{\mu} = -\epsilon\,\lambda\,(\alpha D - \beta\,\mathrm{tr}(D) I)$, and Variance-Covariance matrix $\Omega = \gamma(I + K_p) - \eta \mathrm{Vec}(I)[\mathrm{Vec}(I)]'$ under $H_{1,n}$.

Let $B = \lambda^2/(2\gamma)\mathbf{h}^{(2)}(V) = \left[\frac{1}{4}(I + K_p) - \frac{1}{2p}\mathrm{Vec}(I)\mathrm{Vec}(I)'\right]/\gamma$. Since $B\Omega B = B$, the quadratic form $Z'BZ$ is a χ^2 distribution; since $2\gamma B = (2\gamma B)^2$ is idempotent, this

χ^2 has degree of freedom $= \text{rank}\,(B) = \text{tr}\,(2\gamma B) = \frac{1}{2}\,(p+2)(p-1)$; the non-centrality parameter δ is by definition equal to $\text{Vec}(\boldsymbol{\mu})'\,B\,\text{Vec}(\boldsymbol{\mu}) = \frac{(\epsilon\lambda\alpha)^2}{2\gamma}\left[\text{tr}\left(D^2\right) - \frac{1}{p}\text{tr}^2(D)\right]$

To find the limiting distribution of $\frac{n}{2}\,\log\left[\frac{\text{tr}\,\widehat{V}\cdot\,\text{tr}\left(\widehat{V}^{-1}\right)}{p^2}\right]^p = \frac{n}{2}\left[\mathbf{h}\left(\widehat{V}\right) + \mathbf{h}\left(\widehat{V}^{-1}\right)\right]$, we first observe that at $V = \lambda I$,

$$
\begin{aligned}
\mathbf{h}^{(1)}\left(V^{-1}\right) &= \frac{\partial\,\mathbf{h}\left(V^{-1}\right)}{\partial V} = \frac{\partial \mathbf{h}\left(V^{-1}\right)}{\partial V^{-1}}\frac{\partial V^{-1}}{\partial V} = \mathbf{0} \\
\mathbf{h}^{(2)}\left(V^{-1}\right) &= \frac{\partial^2\,\mathbf{h}\left(V^{-1}\right)}{\partial V^2} = \left[\frac{\partial^2 \mathbf{h}\left(V^{-1}\right)}{\partial (V^{-1})^2}\frac{\partial V^{-1}}{\partial V} + \frac{\partial \mathbf{h}\left(V^{-1}\right)}{\partial V^{-1}}\frac{\partial^2 V^{-1}}{\partial (V^{-1})\partial V}\right]\frac{\partial V^{-1}}{\partial V} \\
&= \lambda^2\left[\frac{1}{2}\,(I+K_p) - \frac{1}{p}\,\text{Vec}(I)\text{Vec}(I)'\right]\left[-\frac{1}{2}\,(I+K_p)\left(V^{-1}\otimes V^{-1}\right)\right]^2 \\
&= \lambda^{-2}\left[\frac{1}{2}\,(I+K_p) - \frac{1}{p}\,\text{Vec}(I)\text{Vec}(I)'\right] = \mathbf{h}^{(2)}\,(V)
\end{aligned}
$$

hence the Delta method again gives the desired asymptotic distribution. □

Proof of Proposition 5.3.2

Proof. Let $\Psi(\mathbf{z}, V_1, V_2) = u\left(\mathbf{z}'V_1^{-1}\mathbf{z}\right)\left(V_2 - \mathbf{z}\mathbf{z}'\right)$, and $\Delta_{\mathbf{x}}$ be a point mass at a given $\mathbf{x} \in \mathbb{R}^p$. Put for brevity

$$
F_\epsilon \sim (1-\epsilon)F + \epsilon\Delta_{\mathbf{x}}\,, \quad \text{and} \quad V_{1,\epsilon} = V_1\left(F_\epsilon\right),\; V_{2,\epsilon} = V_2\left(F_\epsilon\right)
$$

By definition, $V_{1,0} = V_1(F) = \lambda_1 I$, $V_{2,0} = V_2(F) = \lambda_2 I$ are solutions to

$$
E_F\left[\Psi(\mathbf{z}, V_{1,0}, V_{2,0})\right] = 0 \tag{5.6.1}
$$

Then $V_{1,\epsilon}$ together with $V_{2,\epsilon}$ verifies

$$
\mathbf{0} = E_{F_\epsilon}\Psi\left(\mathbf{z}, V_{1,\epsilon}, V_{2,\epsilon}\right) = (1-\epsilon)E_F\Psi\left(\mathbf{z}, V_{1,\epsilon}, V_{2,\epsilon}\right) + \epsilon\Psi\left(\mathbf{x}, V_{1,\epsilon}, V_{2,\epsilon}\right)
$$

Taking VEC operation and differentiating with respect to ϵ yields

$$
\begin{aligned}
\mathbf{0} = &- \text{Vec}\left[E_F\Psi\left(\mathbf{z}, V_{1,\epsilon}, V_{2,\epsilon}\right)\right] + E_F\left[u\left(\mathbf{z}V_{1,\epsilon}^{-1}\mathbf{z}\right)\right]\text{Vec}\left(\frac{\partial V_{2,\epsilon}}{\partial\epsilon}\right) \\
&+ E_F\left[u'\left(\mathbf{z}V_{1,\epsilon}^{-1}\mathbf{z}\right)\right]\text{Vec}\left(V_{2,\epsilon} - \mathbf{z}\mathbf{z}'\right)\text{Vec}\left(-V_{1,\epsilon}^{-1}\mathbf{z}\mathbf{z}'V_{1,\epsilon}^{-1}\right)'\text{Vec}\left(\frac{\partial V_{1,\epsilon}}{\partial\epsilon}\right) \\
&+ \text{Vec}\left[\Psi\left(\mathbf{x}, V_{1,\epsilon}, V_{2,\epsilon}\right)\right] + o(\epsilon)
\end{aligned} \tag{5.6.2}
$$

The first term varnishes at $\epsilon = 0$ by (5.6.1). For simplicity, denote by A_1 and A_2 the influence functions of V_1 and V_2 at $\mathbf{x}$ respectively, taking $\epsilon \downarrow 0$ yields

$$\begin{aligned}
&E_F\left[u\left(\mathbf{z}'\mathbf{z}/\lambda_1\right)\right]\mathrm{Vec}\left(A_2\right) = -\mathrm{Vec}\left[\Psi\left(\mathbf{x}, V_{1,0}, V_{2,0}\right)\right] \\
&+ E_F\left[\frac{u'\left(\mathbf{z}'\mathbf{z}/\lambda_1\right)\left(\mathbf{z}'\mathbf{z}\right)\lambda_2}{p\lambda_1^2}\right]\left(\mathrm{Vec}(I)\left[\mathrm{Vec}(I)\right]'\right)\mathrm{Vec}\left(A_1\right) \\
&- E_F\left[\frac{u'\left(\mathbf{z}'\mathbf{z}/\lambda_1\right)\left(\mathbf{z}'\mathbf{z}\right)^2}{p(p+2)\lambda_1^2}\right]\left(I + K + \mathrm{Vec}(I)\left[\mathrm{Vec}(I)\right]'\right)\mathrm{Vec}\left(A_1\right) \\
= &- \mathrm{Vec}\left[\Psi\left(\mathbf{x}, V_{1,0}, V_{2,0}\right)\right] + \left(\frac{\lambda_2 h_3 - h_4}{\lambda_1^2}\right)\ \mathrm{tr}(A_1)\mathrm{Vec}(I) - \frac{2h_4}{\lambda_1^2}\mathrm{Vec}\left(A_1\right)
\end{aligned}$$

It follows

$$A_2 = -\frac{2h_4}{\lambda_1^2 h_1}A_1 + \left(\frac{\lambda_2 h_3 - h_4}{\lambda_1^2 h_1}\right)\ \mathrm{tr}(A_1)I + \frac{u\left(\mathbf{x}'\mathbf{x}/\lambda_1\right)\left(\mathbf{x}\mathbf{x}' - \lambda_2 I\right)}{h_1}$$

□

Proof of Proposition 5.4.1

Proof. Let $\mathbf{h}(V) = pV/\mathrm{tr}(V)$ and $F_\epsilon = (1-\epsilon)F + \epsilon\delta_{\mathbf{x}}$, where $\delta_{\mathbf{x}}$ is a point-mass at $\mathbf{x}$. Define $\mathbf{h}'(V) = \frac{1}{2}\left[\partial\mathrm{Vec}(\mathbf{h}(V))/\partial\mathrm{Vec}(V)\right](I + J_p)$, where $J_p = \sum_{i=1}^p \mathbf{e}_i\mathbf{e}_i' \otimes \mathbf{e}_i\mathbf{e}_i'$ and $\mathbf{e}_i \in \mathbb{R}^p$ is a vector of zero but a 1 in position i [36]. By definition

$$\begin{aligned}
\mathrm{Vec}\left[\mathrm{IF}\left(\mathbf{x}; \mathbf{h}(V), F\right)\right] &= \lim_{\epsilon\to 0}\frac{\mathrm{Vec}\left[\mathbf{h}\left(V(F_\epsilon)\right) - \mathbf{h}\left(V(F)\right)\right]}{\epsilon} \\
&= \lim_{\epsilon\to 0}\frac{\mathrm{Vec}\left[\mathbf{h}\left(V(F_\epsilon)\right) - \mathbf{h}\left(V(F)\right)\right]}{\mathrm{Vec}\left[V(F_\epsilon) - V(F)\right]}\lim_{\epsilon\to 0}\frac{\mathrm{Vec}\left[V(F_\epsilon) - V(F)\right]}{\epsilon} \\
&= \mathbf{h}'(V(F))\mathrm{Vec}\left[\mathrm{IF}(\mathbf{x}; V, F)\right] \\
&= \left[\frac{p}{\mathrm{tr}(V)}I_{p^2} - \frac{p}{\mathrm{tr}^2(V)}\mathrm{Vec}(V)[\mathrm{Vec}(I)]'\right]\mathrm{Vec}\left[\mathrm{IF}(\mathbf{x}; V, F)\right]
\end{aligned}$$

By affine-equivariance $V(F) = \lambda I_p$ at spherical distribution $F \sim E(\mathbf{0}, I_p)$, hence $\mathbf{h}'(V) = \left[I_{p^2} - \frac{1}{p}\mathrm{Vec}\left(I_p\right)\left[\mathrm{Vec}\left(I_p\right)\right]'\right]/\lambda$, and

$$\begin{aligned}
\mathbf{h}'(V)\mathrm{Vec}\left[\mathrm{IF}(\mathbf{z}; V, F)\right] &= \frac{1}{\lambda}\left[I_{p^2} - \frac{1}{p}\mathrm{Vec}\left(I_p\right)\left[\mathrm{Vec}\left(I_p\right)\right]'\right]\left[\omega_1(|\mathbf{z}|)\mathrm{Vec}\left(\mathbf{z}\mathbf{z}'\right) - \omega_2(|\mathbf{z}|)\mathrm{Vec}(I)\right] \\
&= \frac{\omega_1(|\mathbf{z}|)}{\lambda}\left[\mathrm{Vec}\left(\mathbf{z}\mathbf{z}'\right) - \frac{(\mathbf{z}'\mathbf{z})}{p}\mathrm{Vec}\left(I\right)\right].
\end{aligned}$$

This completes the proof. □

Proof of Theorem 5.4.2

Proof. Continue the notations in the proof of Proposition 5.4.1, since $V \to \lambda I$ and $S = \mathbf{h}(V) \to I$ asymptotically,

$$\begin{aligned}\mathrm{Vec}\left[\sqrt{n}\left(\widehat{S} - I\right)\right] &= \mathbf{h}'(V)\,\mathrm{Vec}\left[\sqrt{n}\left(\widehat{V} - V\right)\right] + o_p(1)\\ &\xrightarrow{H_{1,n}} \frac{1}{\lambda}\left[I_{p^2} - \frac{1}{p}\mathrm{Vec}\left(I_p\right)\left[\mathrm{Vec}\left(I_p\right)\right]'\right]\mathrm{Vec}(Z)\end{aligned}$$

where $Z \sim N(\mathbf{m}, \Omega)$. This proves the first claim. The second claim is proved similarly using the lines in the proof of Corollary 5.1.4. $\square$

5.7 Appendix B: Figures

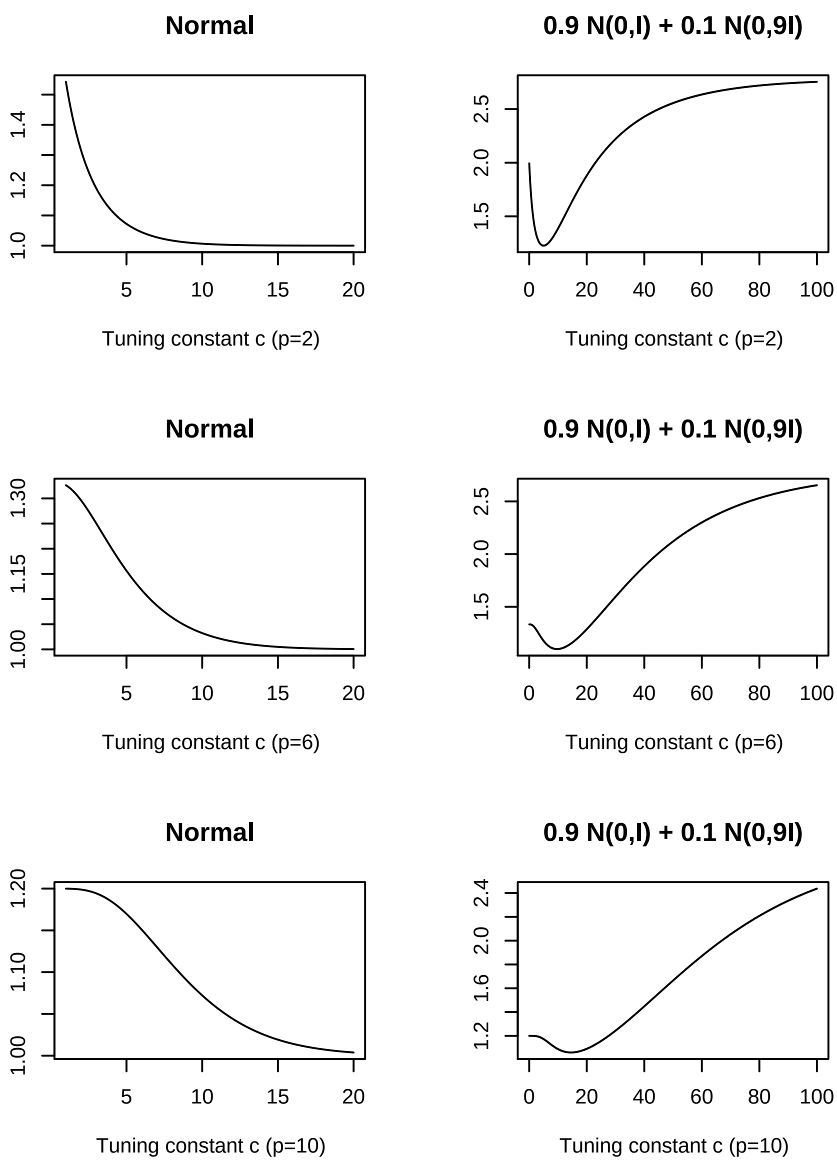

Figure 5.1: Variance coefficient γ of Huber MM-Estimates at Normal distributions

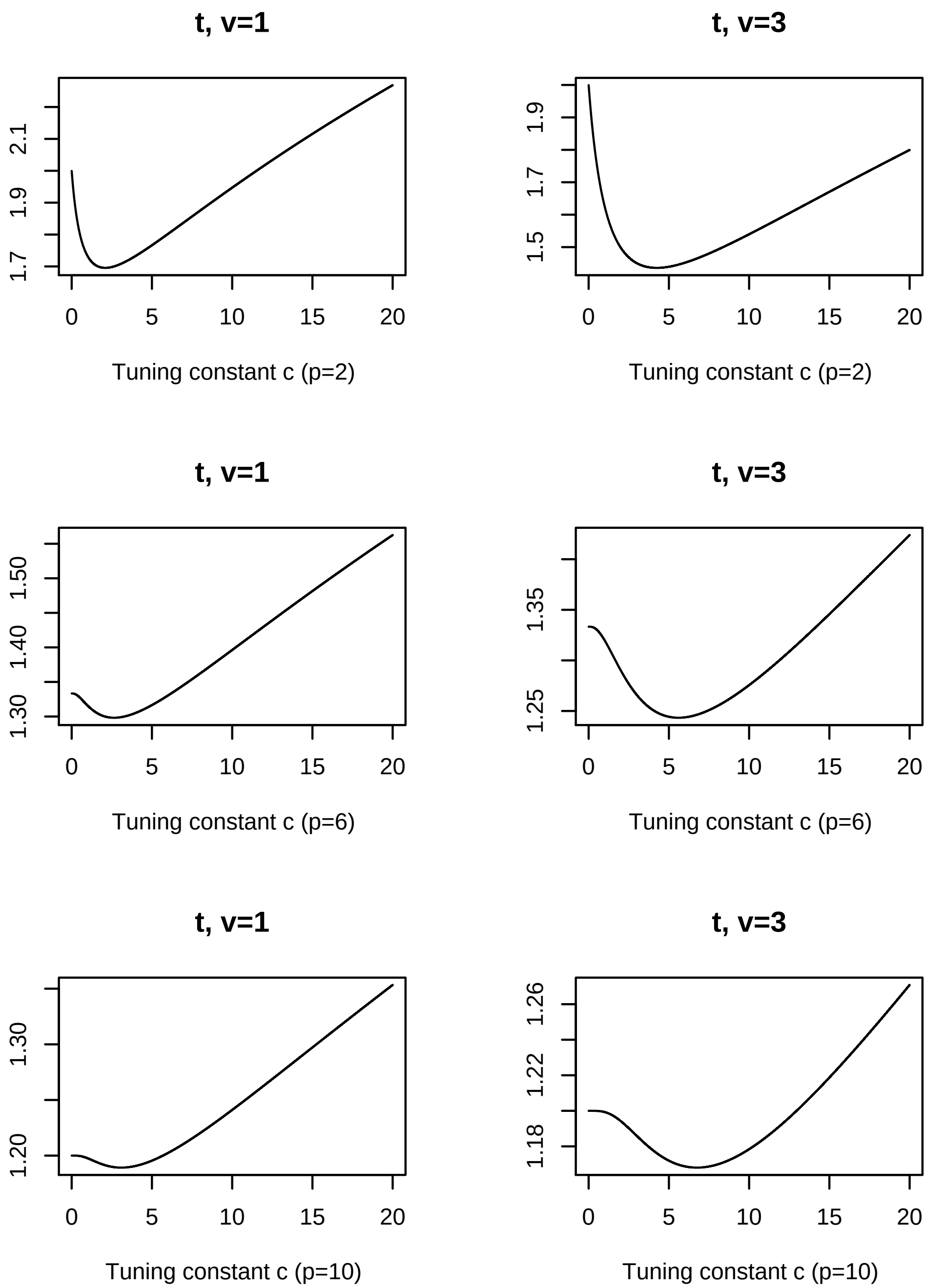

Figure 5.2: Variance coefficient γ of Huber MM-Estimates at $t_{v,p}$ (part 1)

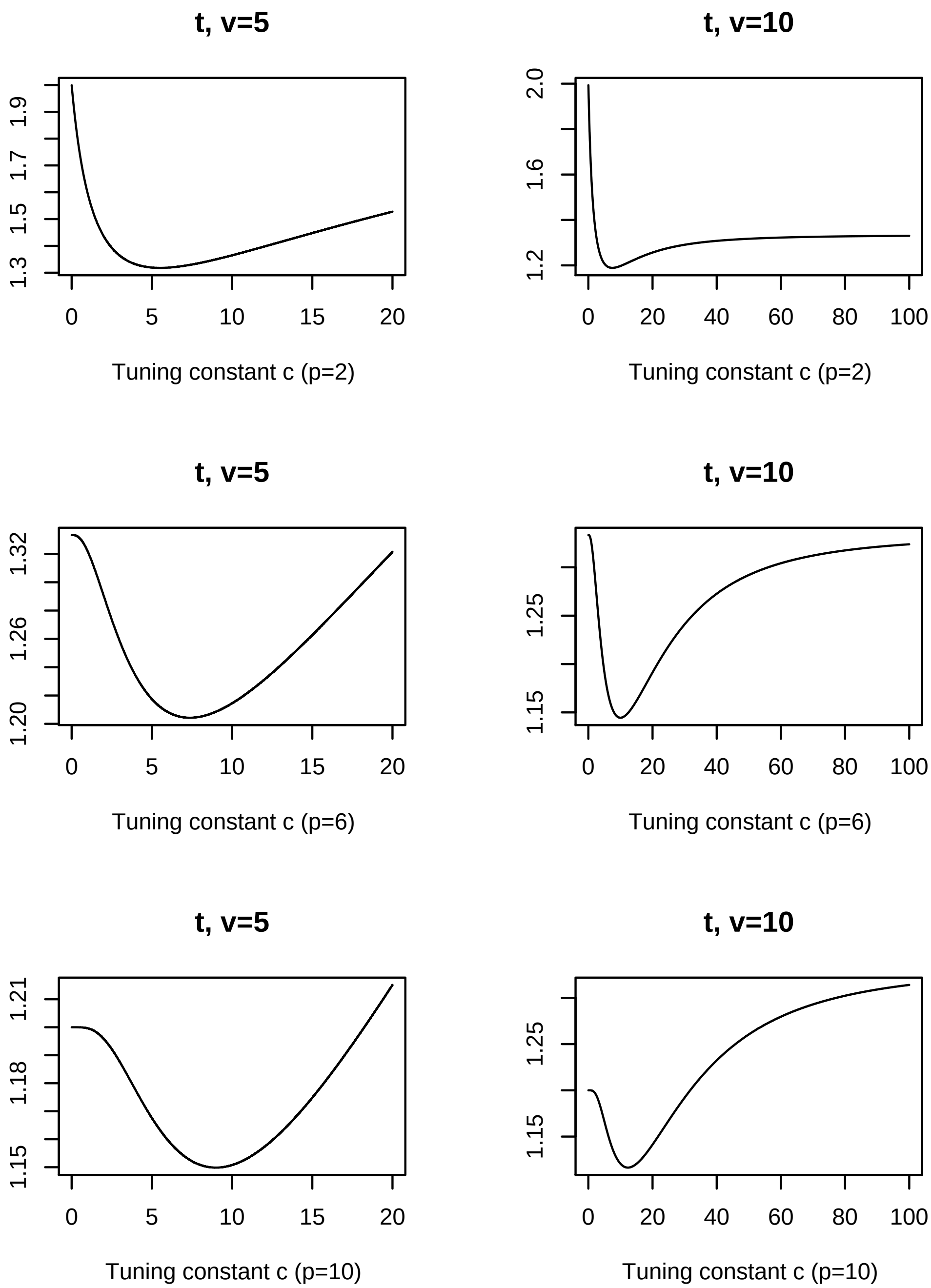

Figure 5.3: Variance coefficient γ of Huber MM-Estimates at $t_{v,p}$ (part 2)

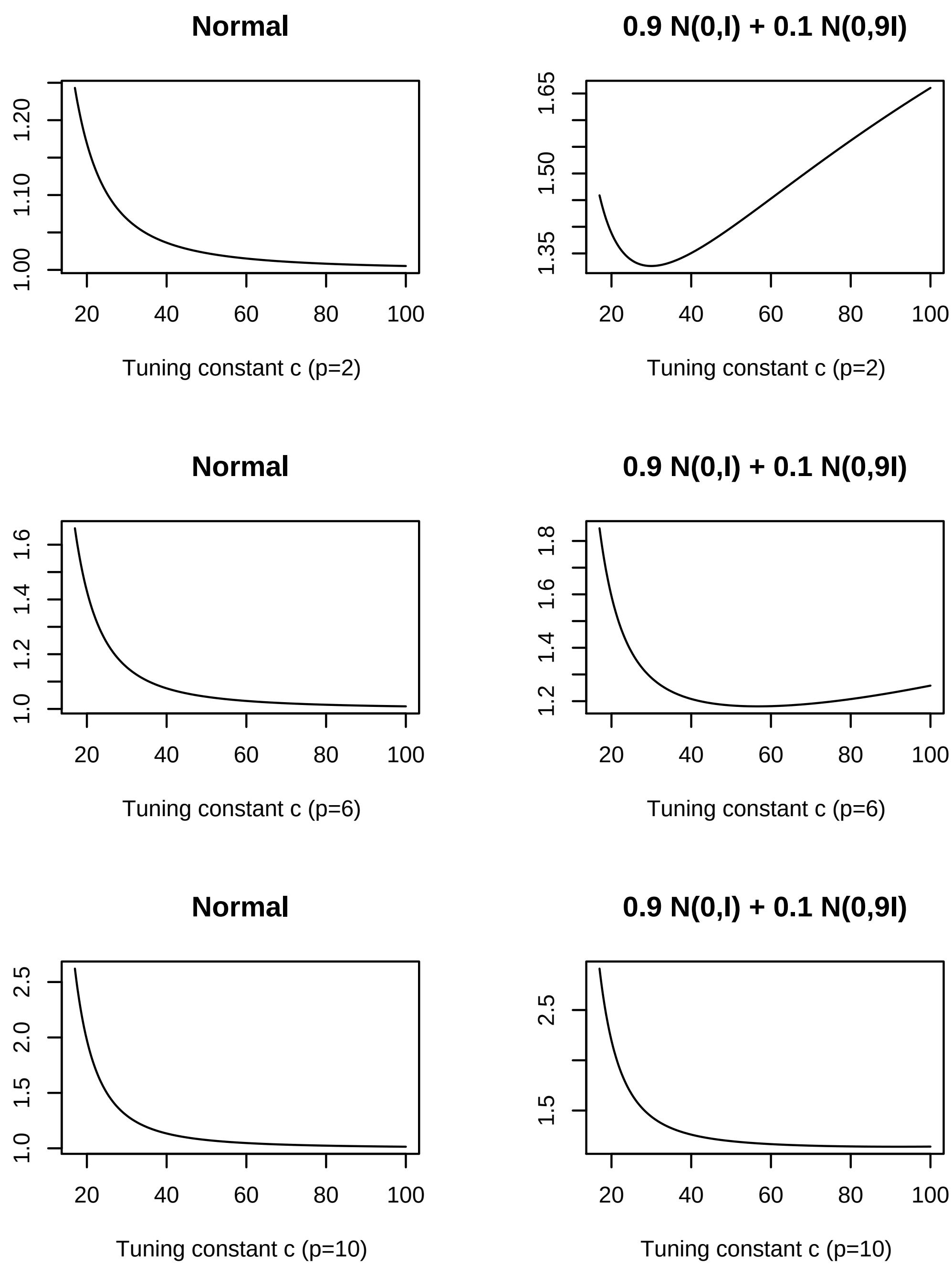

Figure 5.4: Variance coefficient γ of Tukey MM-Estimates at Normal distributions

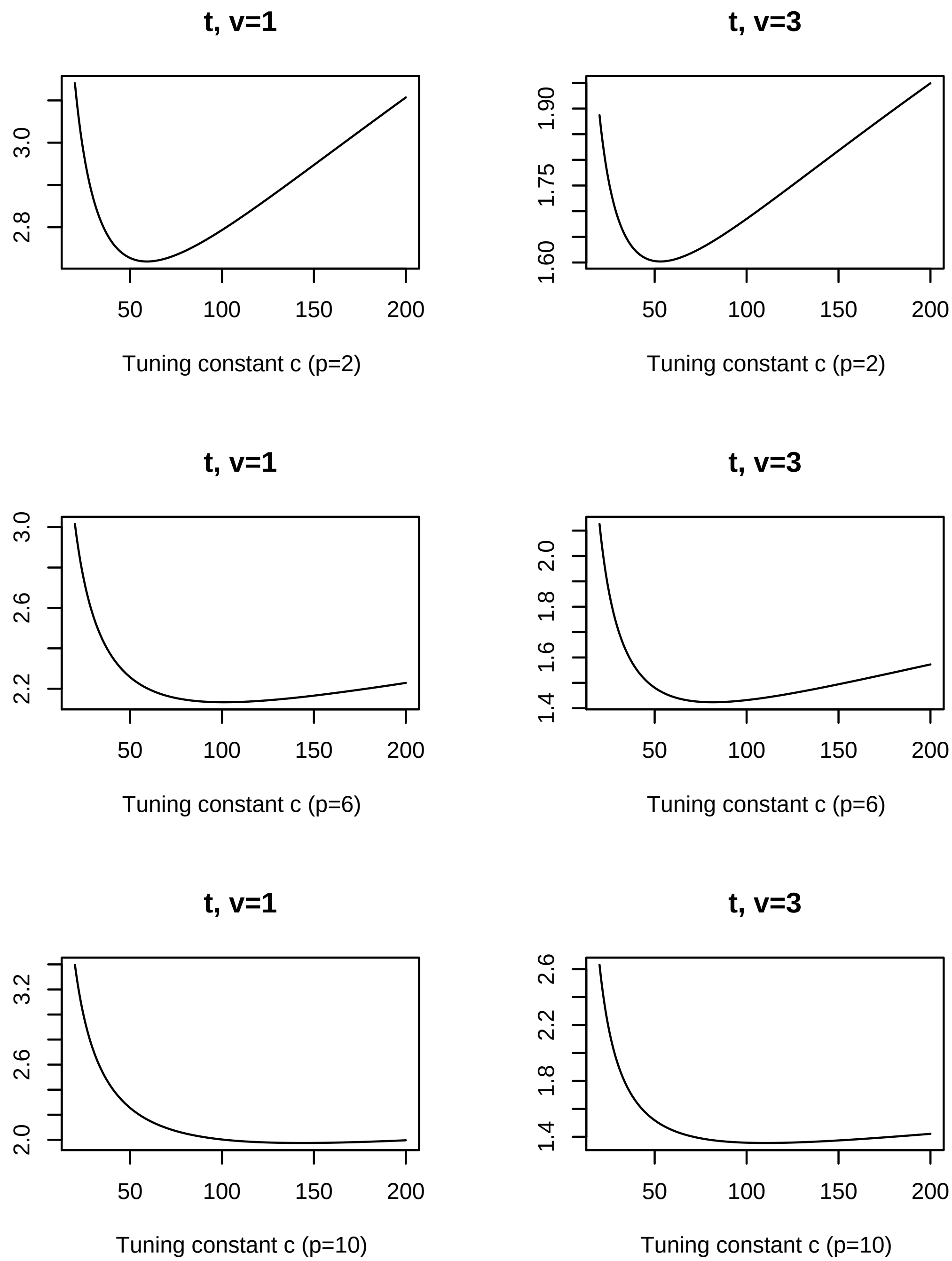

Figure 5.5: Variance coefficient γ of Tukey MM-Estimates at $t_{v,p}$

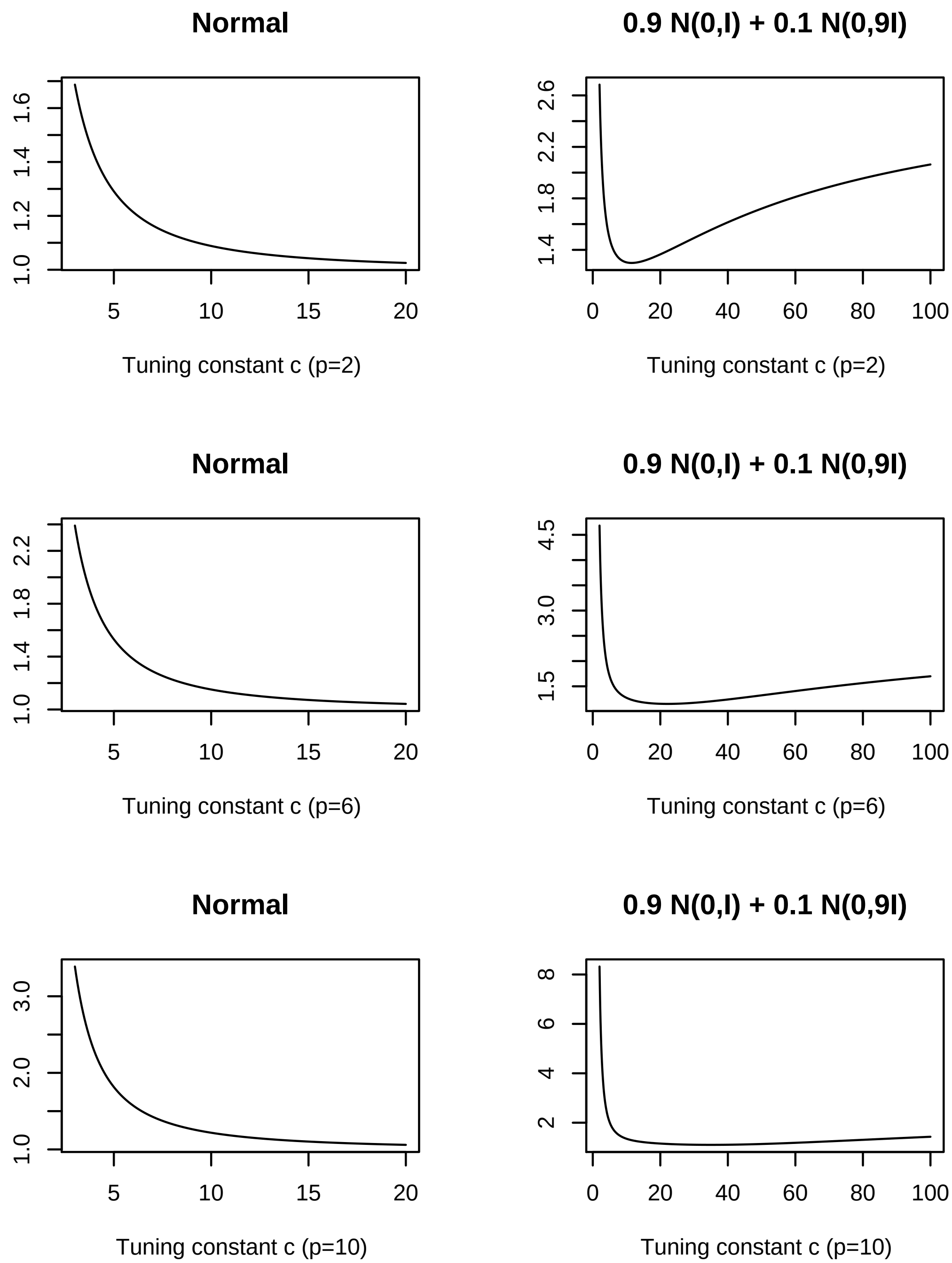

Figure 5.6: Variance coefficient γ of Welsh MM-Estimates at Normal distributions

Chapter 6

Shape Analysis Based on Two Robust Estimators

6.1 Preliminary

As discussed in the Introduction, when the underlying distribution is outside elliptically symmetric, a symmetric one but non-elliptical for example, different scatter functionals are not necessarily representing the same distribution quantity. This triggers an idea that comparing scatter estimates of different types may help discover the departure of underlying distribution from an elliptical symmetric one, that otherwise may not be revealed through single or one type estimate of multivariate scatter. Tyler etc.(2008) gave an interesting argument that the difference between two estimates of scatter matrix $\widehat{V}_i$ can and should be maximally summarized by looking at this new statistic $\widehat{T} = \widehat{V}_1^{-1}\widehat{V}_2$. Various reasoning and interpretations of a statistic of this form were presented there [40], but in the context of detecting mixture though, the reasons why it is excited to use this type of statistics are quite simple. Firstly $\widehat{T}$ is affine equivariant as long as $\widehat{V}_i's$ are; secondly at elliptical symmetric model, scatter estimates are proportional to each other, hence proposition (6.1.1) below implies $\widehat{T}$ would have asymptotic mean of $\mathbf{0}$, whereas scatter matrices are not necessarily proportional to each other at a non-elliptical model which may render a non-zero asymptotic mean for $\widehat{T}$. Consequently, $\widehat{T}$ is likely to have some power at detecting non-elliptical, specifically at a mixture of two elliptical distributions. This chapter is to examine and compare this power through families of scatter estimates. In order to accomplish this goal, the specifics of the asymptotic distribution of $\widehat{T}$ must be first obtained, which is achieved again through the aid of contiguity and usage of influence function.

Proposition 6.1.1. *Let* $\mathbf{z}_1, \cdots, \mathbf{z}_n \sim \mathbf{z} \in \mathbb{R}^p$ *be i.i.d. from* $F \sim E_p(\mathbf{0}, I)$*. Suppose that* $V_k (k = 1, 2)$ *are affine equivariant scatter matrix functionals at* F *possessing influence function* $IF(\mathbf{z}; V_k, F)$*, and* $\widehat{V}_k$ *are respective consistent estimators based on sample. Then,*

1. $IF\left(\mathbf{z}; V_1^{-1} V_2, F\right) = V_1^{-1} \Big(IF(\mathbf{z}; V_2, F) V_2^{-1} - IF(\mathbf{z}; V_1, F) V_1^{-1} \Big) V_2$
2. $\sqrt{n} \left(\widehat{V}_1^{-1} \widehat{V}_2 - V_1^{-1} V_2 \right) = \frac{1}{\sqrt{n}} \sum_{i=1}^n IF(\mathbf{z}_i; V_1^{-1} V_2, F) + o_p(1)$

A random symmetric matrix $Z \in \mathbb{R}_p^p$ is *rotationally invariant* iff $Z \stackrel{d}{=} HZH', \forall H \in O_p$, where $O_p = \{H \in \mathbb{R}_p^p : HH' = I\}$. The following theorem proved by Tyler [35] characterizes the general form of the mean and variance of any rotationally invariant random matrix Z and its variation $V^{1/2} Z V^{1/2}$

Theorem 6.1.2. *Let* $N_i \in \mathbb{R}_p^p \, (i = 1, 2)$ *be two real symmetric random matrices with finite second moments, and* $V_i \in \mathbb{R}_p^p$ *be real symmetric such that* $V_i^{-1/2} N_i V_i^{-1/2}$ *are rotationally invariant. Then, there exist constants* λ_i*,* α_i*,* β_i*, and* γ_i *such that*

$$
\begin{aligned}
E(N_1) &= \lambda_1 V_1 \quad , \quad E(N_2) \;=\; \lambda_2 V_2 \\
\operatorname{var}\{\operatorname{vec}(N_1)\} &= \alpha_1 \, (I + K_p) \, (V_1 \otimes V_1) + \alpha_2 \operatorname{vec}(V_1) \operatorname{vec}(V_1)' \\
\operatorname{var}\{\operatorname{vec}(N_2)\} &= \beta_1 \, (I + K_p) \, (V_2 \otimes V_2) + \beta_2 \operatorname{vec}(V_2) \operatorname{vec}(V_2)' \\
\operatorname{Cov}\{\operatorname{vec}(N_1), \operatorname{vec}(N_2)\} &= \gamma_1 \, (I + K_p) \left(V_1^{1/2} V_2^{1/2} \otimes V_1^{1/2} V_2^{1/2} \right) + \gamma_2 \operatorname{vec}(V_1) \operatorname{vec}(V_2)'
\end{aligned}
$$

where α_1 *represents the variance of any off-diagonal element of* $V_1^{-1/2} N_1 V_1^{-1/2}$*,* α_2 *represents the covariance between any two distinct diagonal elements of* $V_1^{-1/2} N_1 V_1^{-1/2}$*;* γ_1 *represents the covariance between* ij*-th elements of* $V_1^{-1/2} N_1 V_1^{-1/2}$ *and* $V_2^{-1/2} N_2 V_2^{-1/2}$*, and* γ_2 *represents the covariance between* ii*-th element of* $V_1^{-1/2} N_1 V_1^{-1/2}$ *and* jj*-th element of* $V_2^{-1/2} N_2 V_2^{-1/2}$ *(*$i \neq j$ *).*

Corollary 6.1.3. *Let* $\mathbf{z}_1, \cdots, \mathbf{z}_n$ *be a random sample from an elliptical distribution* $\mathbf{z} \sim E_p(\mathbf{u}, \Sigma)$*. Let* $V_i \in \mathbb{R}_p^p \, (i = 1, 2)$ *be affine equivariant scatter matrix functionals,* $\widehat{V}_i$ *be consistent estimates of* V_i *based on the sample empirical distribution. Write* $\widehat{\Delta}_n =$

$\widehat{V}_1^{-1}\widehat{V}_2$ *and* $\Delta = V_1^{-1}V_2$. *Then* $\sqrt{n}\left(\widehat{\Delta}_n - \Delta\right) \xrightarrow{d} \mathbf{N}_p^p\left(\mathbf{0},\, \Omega\right)$, *where*

$$\Omega = \alpha_1(V_2\Delta \otimes V_1^{-1}) + \beta_1(V_2 \otimes \Delta V_1^{-1}) + (\alpha_1+\beta_1)K_p(\Delta\otimes\Delta') + (\alpha_2+\beta_2-2\gamma_2)\mathrm{vec}(\Delta)\mathrm{vec}(\Delta)'$$
$$- \gamma_1\left(V_2^{1/2}V_1^{1/2}\Delta \otimes \Delta V_2^{-1/2}V_1^{-1/2} + \Delta' V_1^{1/2}V_2^{1/2} \otimes V_1^{-1/2}V_2^{-1/2}\Delta'\right)$$
$$- \gamma_1 K_p\left(\Delta V_2^{-1/2}V_1^{1/2}\Delta \otimes V_2^{1/2}V_1^{-1/2} + V_1^{-1/2}V_2^{1/2} \otimes \Delta' V_1^{1/2}V_2^{-1/2}\Delta'\right)$$

6.2 Two Huber M-Estimates

Let $V_i = E\left[u_i\left(\mathbf{x}'V_i\mathbf{x}\right)\right]$ be two implicitly defined M-functional of Definition 5.1.1 with influence functions defined as in Proposition 5.1.2,

$$\mathrm{IF}(\mathbf{x}; V_1, F) = w_1\,\mathbf{x}\mathbf{x}' - \pi_1 I$$
$$\mathrm{IF}(\mathbf{x}; V_2, F) = w_2\,\mathbf{x}\mathbf{x}' - \pi_2 I$$

where w_i and π_i are functions of $\mathbf{x}'\mathbf{x}$.

Let $\widehat{V}_i$ be consistent estimates of V_i, respectively, based on empirical distributions. Let $\Delta = V_1^{-1}V_2$, and $\widehat{\Delta} = \widehat{V}_1^{-1}\widehat{V}_2$. Obviously Δ is affine equivariant whenever $V_i's$ are. As summarized in section (4.2), the distribution of $\widehat{\Delta}$ on a non-elliptical will be investigated locally and asymptotically on the platform of a sequence of hypotheses defined in (4.2.1), i.e.

$$H_0 : (1-\epsilon)F\big(\mathbf{0}, I_p\big) + \epsilon G\big(\mathbf{0}, I_p\big) \;\text{ vs }\; H_{1,n} : (1-\epsilon)F\big(\mathbf{0}, I_p\big) + \epsilon G\left(\mathbf{0},\, \big(I_p + D/\sqrt{n}\big)^{-1}\right)$$

particularly, the case when $G(\mathbf{0}, I) = F(\mathbf{0}, bI)$ is of primary interests, i.e. $g(\mathbf{x}'\mathbf{x}) = b^{-\frac{p}{2}}f\left(\mathbf{x}'\mathbf{x}/b\right)$ and the null hypothesis is a mixture $P \sim (1-\epsilon)F\big(\mathbf{0}, I_p\big) + \epsilon F\big(\mathbf{0}, bI_p\big)$. Using the notations in Theorem 4.1.4, the deviations of asymptotic distribution of $\widehat{\Delta}$ is an easy extension of Theorem 4.1.5.

Theorem 6.2.1. *Under sequence of alternative hypotheses* $H_{1,n}$, $\Delta \to \lambda_2/\lambda_1 I$, *where* λ_i *is the solution to equation* $(1-\epsilon)E_s\left[u_i(s/\lambda_i)(s/\lambda_i)\right] + \epsilon E_s\left[u_i\left(sb/\lambda_i\right)\left(sb/\lambda_i\right)\right] = p$; *and*

1. $\sqrt{n}\left(\widehat{\Delta}-\Delta\right) \xrightarrow{d} N\left(-\epsilon\left(\frac{\lambda_2}{\lambda_1}\right)\mathbf{m},\ \left(\frac{\lambda_2}{\lambda_1}\right)^2\Omega\right)$, *where*

$$\mathbf{m} = (\alpha_2-\alpha_1)D - (\beta_2-\beta_1)\,\mathrm{tr}(D)I$$

$$\Omega = (\gamma_1+\gamma_2-2\gamma_{1,2})\,(I+K_p) + (\eta_1+\eta_2-2\,\eta_{1,2})\,\mathrm{Vec}(I)[\mathrm{Vec}(I)]', \qquad \text{and}$$

$$\alpha_i = \frac{1}{2h_i+1}\left\{\frac{E_s\left[u_i\,(sb/\lambda_i)\,(sb/\lambda_i)\right]}{p} + \frac{2E_s\left[u_i'\,(sb/\lambda_i)\,(sb/\lambda_i)^2\right]}{p(p+2)}\right\}$$

$$\beta_i = \frac{[(p+2)h_i+1]^{-1}}{(2h_i+1)}\left\{\frac{h_iE_s\left[u_i\,(sb/\lambda_i)\,(sb/\lambda_i)\right]}{p} - \frac{E_s\left[u_i'\,(sb/\lambda_i)\,(sb/\lambda_i)^2\right]}{p(p+2)}\right\}$$

$$h_i = \frac{1}{p(p+2)}E_s\left[(1-\epsilon)u_i'\,(s/\lambda_i)\,(s/\lambda_i)^2 + \epsilon u_i'\,(sb/\lambda_i)\,(sb/\lambda_i)^2\right]$$

$$\gamma_i = \frac{1}{(2h_i+1)^2p(p+2)}\left\{(1-\epsilon)E_s\left[u_i\big(s/\lambda_i\big)\,(s/\lambda_i)\right]^2 + \epsilon E_s\left[u_i\big(sb/\lambda_i\big)\,(sb/\lambda_i)\right]^2\right\}$$

$$\gamma_{1,2} = (1-\epsilon)\,\frac{E_s\left[u_1\big(s/\lambda_1\big)u_2\big(s/\lambda_2\big)\,s^2/(\lambda_1\lambda_2)\right]}{(2h_1+1)(2h_2+1)p(p+2)} \quad + \quad \epsilon\,\frac{E_s\left[u_1\big(sb/\lambda_1\big)u_2\big(sb/\lambda_2\big)\,(sb)^2/(\lambda_1\lambda_2)\right]}{(2h_1+1)(2h_2+1)p(p+2)}$$

$$\eta_i = \frac{[1-2h_i^2(p+2)]\gamma_i-1}{[(p+2)h_i+1]^2}, \qquad\qquad s \sim \frac{\pi^{p/2}}{\Gamma(p/2)}s^{\frac{p}{2}-1}f(s)$$

2. *If* $b=1$, *i.e.* $F\big(\mathbf{0},I_p\big) = G\big(\mathbf{0},I_p\big)$, *then* $\sqrt{n}\left(\widehat{\Delta}-\Delta\right) \xrightarrow{d} N_p^p\left(0,\left(\frac{\lambda_2}{\lambda_1}\right)^2\Omega\right)$.

Corollary 6.2.2. *Assuming notations in Theorem 6.2.1, under the sequence of alternative hypotheses* $H_{1,n}$,

1. $n\left[\log\frac{\left(\frac{1}{p}\,\mathrm{tr}\,\widehat{\Delta}\right)^p}{|\widehat{\Delta}|}\right] \xrightarrow{d} \gamma^*\,\chi_q^2(\delta)$, *a non-central Chi-square distribution with degree of freedom* $q=\frac{1}{2}(p+2)(p-1)$ *and non-centrality* $\delta = \frac{(\epsilon\alpha^*)^2}{2\gamma^*}\left[\mathrm{tr}\left(D^2\right) - \frac{1}{p}\mathrm{tr}^2(D)\right]$, *where* $\gamma^* = (\gamma_1+\gamma_2-2\,\gamma_{1,2})$, *and* $\alpha^* = \alpha_2-\alpha_1$

2. $\frac{n}{2}\,\log\left[\frac{\mathrm{tr}\,\widehat{\Delta}\cdot\mathrm{tr}\left(\widehat{\Delta}^{-1}\right)}{p^2}\right]^p \xrightarrow{d} \gamma^*\,\chi_q^2(\delta)$

The weight function for a Huber M-estimate of scatter matrix is given by

$$\beta u_c(s) = \begin{cases} 1 & (s<c) \\ c/s & (s\geq c) \end{cases}$$

The scaling factor β is defined so that $E\left[u\left(\chi_p^2\right)\chi_p^2\right] = p$. This implies the ρ-function is of the form,

$$\rho_c(s) = c\rho(s/c) = \frac{1}{2\beta}\begin{cases} s & (s < c) \\ c\log(s/c) + c & (s \geq c) \end{cases}$$

and the derivative is simply $u_c'(s) = -c/\left(\beta s^2\right),\ s \geq c$.

As the tuning constant c ranges over $[0, +\infty]$, a family of Huber estimates, denoted by $\widehat{V}_c$, is defined, where the limiting case $\lim_{c\to+\infty}\widehat{V}_c$ corresponds to the sample variance-covariance matrix.

The performance of $\widehat{V}_{c_1}^{-1}\widehat{V}_{c_2}$ on the hypothesis testing of symmetric elliptical verse non-elliptical (mixture of two elliptical distributions)

$$H_0 : F(\mathbf{0}, I) \qquad \text{vs} \qquad H_1 : (1-\epsilon)F(\mathbf{0}, I) + \epsilon F(\mathbf{0}, bI)$$

can be evaluated and compared through its asymptotic efficacy on the sequence of equivalent hypotheses defined in (4.2.1), i.e.

$$H_0 : (1-\epsilon)F\big(\mathbf{0}, I_p\big) + \epsilon F\big(\mathbf{0}, bI_p\big) \ \text{ vs } \ H_{1,n} : (1-\epsilon)F\big(\mathbf{0}, I_p\big) + \epsilon F\left(\mathbf{0},\, b\big(I_p + D/\sqrt{n}\big)^{-1}\right)$$

with $b = 1$. Unfortunately, Theorem 6.2.1 indicates that any M-estimate of scatter would have the same asymptotic mean if $b = 1$, i.e. when the underlying distribution of null-hypothesis is a symmetric elliptical. Consequently, the statistics $\widehat{V}_{c_1}^{-1}\widehat{V}_{c_2}$ in Corollary 6.2.2 would have zero non-centrality δ, hence zero Pitman's efficacy [19, 28] no matter which pair of c_1 and c_2 is chosen . It seems the comparison of their ability to distinguish a contamination from a symmetric elliptical is unattainable. However, these statistics do have non-zero non-centrality parameters δ when the underlying distribution of null is a mixture of two different elliptical distributions ($b \neq 1$), they would have varying efficacies at detecting non-elliptical from a mixture of elliptical. This suggests a way to compare these statistics' ability when it comes to separating a non-elliptical, particularly a mixture of two elliptical, from a symmetric elliptical distribution. This goal can be achieved by simply looking at their asymptotic-relative-efficacy as $b \to 1$. More rigorously, let $\delta_{0,+\infty}$ be the non-centrality of $\widehat{V}_0^{-1}\widehat{V}_{+\infty}$ in Corollary 6.2.2, and define

$$ARE(c_1, c_2) = \lim_{b \to 1} \frac{\delta_{c_1,c_2}}{\delta_{0,+\infty}}$$

This quantity ARE measures the asymptotic efficacy of a pair $\widehat{V}_{c_1}^{-1}\widehat{V}_{c_2}$ with respect to that of fixed baseline $\widehat{V}_0^{-1}\widehat{V}_{+\infty}$.

As a function of c_1 and c_2, the contour plots of ARE of $\widehat{V}_{c_1}^{-1}\widehat{V}_{c_2}$ with respect to $\widehat{V}_0^{-1}\widehat{V}_{+\infty}$ within the family of Huber M-estimates are presented in Figure 6.1 and 6.2 of Appendix B. The graphs there indicate indubitably that $\widehat{V}_0^{-1}\widehat{V}_{+\infty}$ is not the best statistic at detecting Normal mixture.

6.3 Tukey MM-Estimates vs Tukey

Let $\sigma_i (i = 1, 2)$ be two preliminary scale functionals of affine equivariance, and Γ_i be the minimizers of $E\left[\rho_i\left(\mathbf{x}'\Gamma_i^{-1}\mathbf{x}/\sigma_i^2\right)\right]$ over all $\Gamma \in \mathcal{PDS}(p)$ with $\det(\Gamma) = 1$. Consequently the scatter matrices $V_i = \sigma_i^2\Gamma_i$, $(i = 1, 2)$ are two MM-functionals of definition 5.5.1. Denote again by $\widehat{V}_i$ the sample version of V_i.

Under the sequence of mixture distributions

$$(1-\epsilon)F\big(\mathbf{0}, I_p\big) + \epsilon F\left(\mathbf{0},\, b\big(I_p + D/\sqrt{n}\big)^{-1}\right) \tag{6.3.1}$$

the statistic $\widehat{V}_1^{-1}\widehat{V}_2$ is $\sqrt{n}$ consistent and is convergent to a multi-normal distribution, whose mean and covariance are determined by Theorem 4.1.5 in terms of expectations on influence function of $V_1^{-1}V_2$ which is readily deduced from Proposition 6.1.1 and through the lines in Section 5.5.2.

Theorem 6.3.1. *Let* $s \sim \frac{\pi^{p/2}}{\Gamma(p/2)} s^{\frac{p}{2}-1} f(s)$, $u_i(s) = 2\rho_i'(s)$. *Under a sequence of mixture distributions of (6.3.1),* $V_1^{-1}V_2 \to \sigma_2^2/\sigma_1^2 I$, *and*

1. $\sqrt{n}\left(\widehat{V}_1^{-1}\widehat{V}_2 - V_1^{-1}V_2\right) \xrightarrow{d} N\left(-\epsilon\left(\frac{\sigma_2^2}{\sigma_1^2}\right)\mathbf{m},\ \left(\frac{\sigma_2^2}{\sigma_1^2}\right)^2\Omega\right)$, *where*

$$\mathbf{m} = (\alpha_2 - \alpha_1)\left(D - \frac{\mathrm{tr}(D)}{p}I\right)$$
$$\Omega = (\gamma_1 + \gamma_2 - 2\gamma_{1,2})\left[(I + K_p) - \frac{2}{p}\mathrm{Vec}(I)[\mathrm{Vec}(I)]'\right]$$
$$\alpha_i = \frac{1}{2h_i + k_i}\left\{\frac{2E_s\left[u_i'\left(bs/\sigma_i^2\right)(bs/\sigma_i^2)^2\right]}{p(p+2)} + \frac{E_s\left[u_i\left(bs/\sigma_i^2\right)(bs/\sigma_i^2)\right]}{p}\right\}$$
$$h_i = \frac{1}{p(p+2)}E_s\left[(1-\epsilon)u_i'\left(s/\sigma_i^2\right)\left(s/\sigma_i^2\right)^2 + \epsilon u_i'\left(bs/\sigma_i^2\right)\left(bs/\sigma_i^2\right)^2\right]$$
$$k_i = \frac{1}{p}E_s\left[(1-\epsilon)u_i\left(s/\sigma_i^2\right)\left(s/\sigma_i^2\right) + \epsilon u_i\left(bs/\sigma_i^2\right)\left(bs/\sigma_i^2\right)\right]$$
$$\gamma_i = \frac{(2h_i+k_i)^{-2}}{p(p+2)}\left\{(1-\epsilon)E_s\left[u_i^2\left(s/\sigma_i^2\right)\left(s/\sigma_i^2\right)^2\right] + \epsilon E_s\left[u_i^2\left(bs/\sigma_i^2\right)\left(bs/\sigma_i^2\right)^2\right]\right\}$$
$$\gamma_{1,2} = \frac{(1-\epsilon)E_s\left[u_1\left(\frac{s}{\sigma_1^2}\right)u_2\left(\frac{s}{\sigma_2^2}\right)\frac{s^2}{(\sigma_1^2\sigma_2^2)}\right] + \epsilon E_s\left[u_1\left(\frac{bs}{\sigma_1^2}\right)u_2\left(\frac{bs}{\sigma_2^2}\right)\frac{(bs)^2}{(\sigma_1^2\sigma_2^2)}\right]}{p(p+2)(2h_1+k_1)(2h_2+k_2)}$$

2. *If* $b = 1$, *i.e.* $F\big(\mathbf{0}, I_p\big) = G\big(\mathbf{0}, I_p\big)$, *then* $\alpha_2 = \alpha_1 = 1$, *that implies* $\sqrt{n}\left(\widehat{V}_1^{-1}\widehat{V}_2 - V_1^{-1}V_2\right)$ *has asymptotic mean of* $\mathbf{0}$.

The efficiency of $\widehat{V}_1^{-1}\widehat{V}_2$ can be measured by its Pitman's efficacy [19, 28], denoted by δ, which is proportional to $(\alpha_2-\alpha_1)^2/(\gamma_1+\gamma_2-2\gamma_{1,2})$. It is of primary interests to find and compare the efficiencies at a mixture of distribution of this form

$$(1-\epsilon)F\big(\mathbf{0}, I_p\big) + \epsilon F\left(\mathbf{0}, \big(I_p + D/\sqrt{n}\big)^{-1}\right)$$

which is a special case of (6.3.1) with $b = 1$. Under this mixture model, the efficacy of $\widehat{V}_1^{-1}\widehat{V}_2$ is unfortunately 0 by Theorem 6.3.1. However, one can still obtain the asymptotic relative efficiency (ARE) of $\widehat{V}_1^{-1}\widehat{V}_2$ with respect to $\widehat{V}_o^{-1}\widehat{V}_{+\infty}$ by

$$\mathrm{ARE} = \lim_{b\to 1}\frac{\delta_{\widehat{V}_1^{-1}\widehat{V}_2}}{\delta_{\widehat{V}_0^{-1}\widehat{V}_{+\infty}}}$$

as both $\widehat{V}_1^{-1}\widehat{V}_2$ and $\widehat{V}_0^{-1}\widehat{V}_{+\infty}$ have non-zero efficacies at $b \neq 1$, where $\widehat{V}_0$ and $\widehat{V}_{+\infty}$ are members of Huber M-estimates with $\widehat{V}_{+\infty}$ corresponding to sample covariance matrix and $\widehat{V}_0$ corresponding to the Tyler-estimate.

As illustration, consider Tukey MM-estimates determined by scale functional σ and

Tuykey's ρ function

$$\rho_c(s) = c\rho\left(s/c\right) = \begin{cases} \frac{s}{2} - \frac{s^2}{2c} + \frac{s^3}{6c^2}, & 0 < s \leq c \\ \frac{c}{6}, & s > c \end{cases}$$

which induces the weight function and its derivative

$$\begin{aligned} u_c(s) &= 2\rho_c'\left(s\right) = \left(1 - \frac{s}{c}\right)^2 \mathrm{I}_{[0 \leq s \leq c]} \\ u_c'\left(s\right) &= \frac{2}{c}\left(\frac{s}{c} - 1\right)\mathrm{I}_{[0 \leq s \leq c]} \end{aligned}$$

As the tuning constant c ranges over $(0, +\infty)$, a family of Tukey's MM-estimates $\widehat{V}_c$ is forged. The ARE's of $\widehat{V}_{c_1}^{-1}\widehat{V}_{c_2}$ with respect $\widehat{V}_0^{-1}\widehat{V}_{+\infty}$ are depicted in Figure 6.3 and in Figure 6.4. It is interesting to observe that for any $\widehat{V}$ from the family of Tukey MM-estimates, the best scatter estimate within the family that generates the maximal difference from $\widehat{V}$ is the sample covariance matrix.

A final remark. Any two scatter estimates will not be the best for detecting all departures from an elliptical distribution since the class of non-elliptical distributions is very broad. A good strategy may be to compare many scatters simultaneously and choose the two extreme values. The procedure to find two appropriate scatter estimates that will render the maximum difference between them hence the best test statistic has yet to be discovered which can't be accomplished without first completing the comparison of the AREs across families; the asymptotic distributions of $\widehat{V}_1^{-1}\widehat{V}_2$ when the auxiliary scale is unknown, which turn out to be analogous to current result based on known auxiliary scale, are to be presented in a subsequent paper.

6.4 Appendix A: Proofs of Theorems and Corollaries

Proof of Proposition 6.1.1

Proof. Let $V_i = V_i(F), i = 1, 2$ be two affine equivariant scatter functionals at distribution F, with influence functions denoted by $A_i = \mathrm{IF}(\mathbf{x}; V_i, F)$. Let F_ϵ denote the contaminated distribution $(1-\epsilon)F + \epsilon\delta_{\mathbf{x}}$, where $\delta_{\mathbf{x}}$ is a point mass at $\mathbf{x} \in \mathbb{R}^p$. The relationship between $V_1(F_\epsilon)$ and $V_1(F)$ is defined via the definition of influence function, $V_1(F_\epsilon) = V_1(F) + \epsilon A_1 + o(\epsilon)$, which implies $V_1^{-1}(F_\epsilon) = V_1^{-1} - \epsilon V_1^{-1} A_1 V_1^{-1} + o(\epsilon)$. Then

$$\begin{aligned} V_1^{-1}(F_\epsilon)V_2(F_\epsilon) &= \left[V_1^{-1} - \epsilon V_1^{-1} A_1 V_1^{-1} + o(\epsilon)\right]\left[V_2 + \epsilon A_2 + o(\epsilon)\right] \\ &= V_1^{-1}V_2 - \epsilon V_1^{-1} A_1 V_1^{-1} V_2 + \epsilon V_1^{-1} A_2 + o(\epsilon) \end{aligned}$$

The influence function of $V_1^{-1}V_2$ is readily obtained as

$$\mathrm{IF}\left(\mathbf{x}; V_1^{-1}V_2, F\right) = \lim_{\epsilon \to 0} \frac{V_1^{-1}(F_\epsilon)V_2(F_\epsilon) - V_1^{-1}V_2}{\epsilon} = V_1^{-1}\left(A_2 V_2^{-1} - A_1 V_1^{-1}\right) V_2$$

□

Proof of Corollary 6.1.3

Proof. The assumptions on $\widehat{V}_i$ assure

$$\sqrt{n}\,\mathrm{vec}\left(\widehat{V}_1 - V_1\,,\, \widehat{V}_2 - V_2\right) \xrightarrow{d} \mathrm{vec}\,(N_1\,,\, N_2) \sim N\left(\mathbf{0},\, \Omega_0\right)$$

where, by Theorem 6.1.2,

$$\Omega_0 = \begin{pmatrix} \alpha_1\,(I+K_p)\,(V_1 \otimes V_1) + \alpha_2\,\mathrm{Vec}(V_1)\,[\mathrm{Vec}(V_1)]', & \gamma_1\,(I+K_p)\left(V_1^{\frac{1}{2}} V_2^{\frac{1}{2}} \otimes V_1^{\frac{1}{2}} V_2^{\frac{1}{2}}\right) + \gamma_2\,\mathrm{Vec}(V_1)\,[\mathrm{Vec}(V_2)]' \\ \gamma_1\,(I+K_p)\left(V_2^{\frac{1}{2}} V_1^{\frac{1}{2}} \otimes V_2^{\frac{1}{2}} V_1^{\frac{1}{2}}\right) + \gamma_2\,\mathrm{Vec}(V_2)\,[\mathrm{Vec}(V_1)]', & \beta_1\,(I+K_p)\,(V_2 \otimes V_2) + \beta_2\,\mathrm{Vec}(V_2)\,[\mathrm{Vec}(V_2)]' \end{pmatrix}$$

The inverse of $\widehat{V}_1$ can be approximated as

$$\widehat{V}_1^{-1} = [V_1 + (\widehat{V}_1 - V_1)]^{-1} = V_1^{-1} - V_1^{-1}(\widehat{V}_1 - V_1)V_1^{-1} + O\left(n^{-1/2}\right)$$

which in turn gives the approximation of $\widehat{V}_1^{-1}\widehat{V}_2$,

$$\begin{aligned} \widehat{V}_1^{-1}\widehat{V}_2 &= [V_1 + (\widehat{V}_1 - V_1)]^{-1}\,[V_2 + (\widehat{V}_2 - V_2)] \\ &= V_1^{-1}V_2 + V_1^{-1}(\widehat{V}_2 - V_2) - V_1^{-1}(\widehat{V}_1 - V_1)V_1^{-1}V_2 + O\left(n^{-1}\right) \end{aligned}$$

It follows,

$$\sqrt{n}\left(\widehat{V}_1^{-1}\widehat{V}_2 \,-\, V_1^{-1}V_2\right) \xrightarrow{d} V_1^{-1}N_2 - V_1^{-1}N_1V_1^{-1}V_2 \equiv N \text{ say.}$$

N is a multivariate normal matrix with mean $\mathbf{0}$, its covariance matrix is obtained by noting

$$\text{Vec}(N) = \left(-V_2V_1^{-1} \otimes V_1^{-1}\,,\, I \otimes V_1^{-1}\right)\begin{pmatrix}\text{Vec}(N_1)\\ \text{Vec}(N_2)\end{pmatrix}$$

□

6.5 Appendix B: Figures

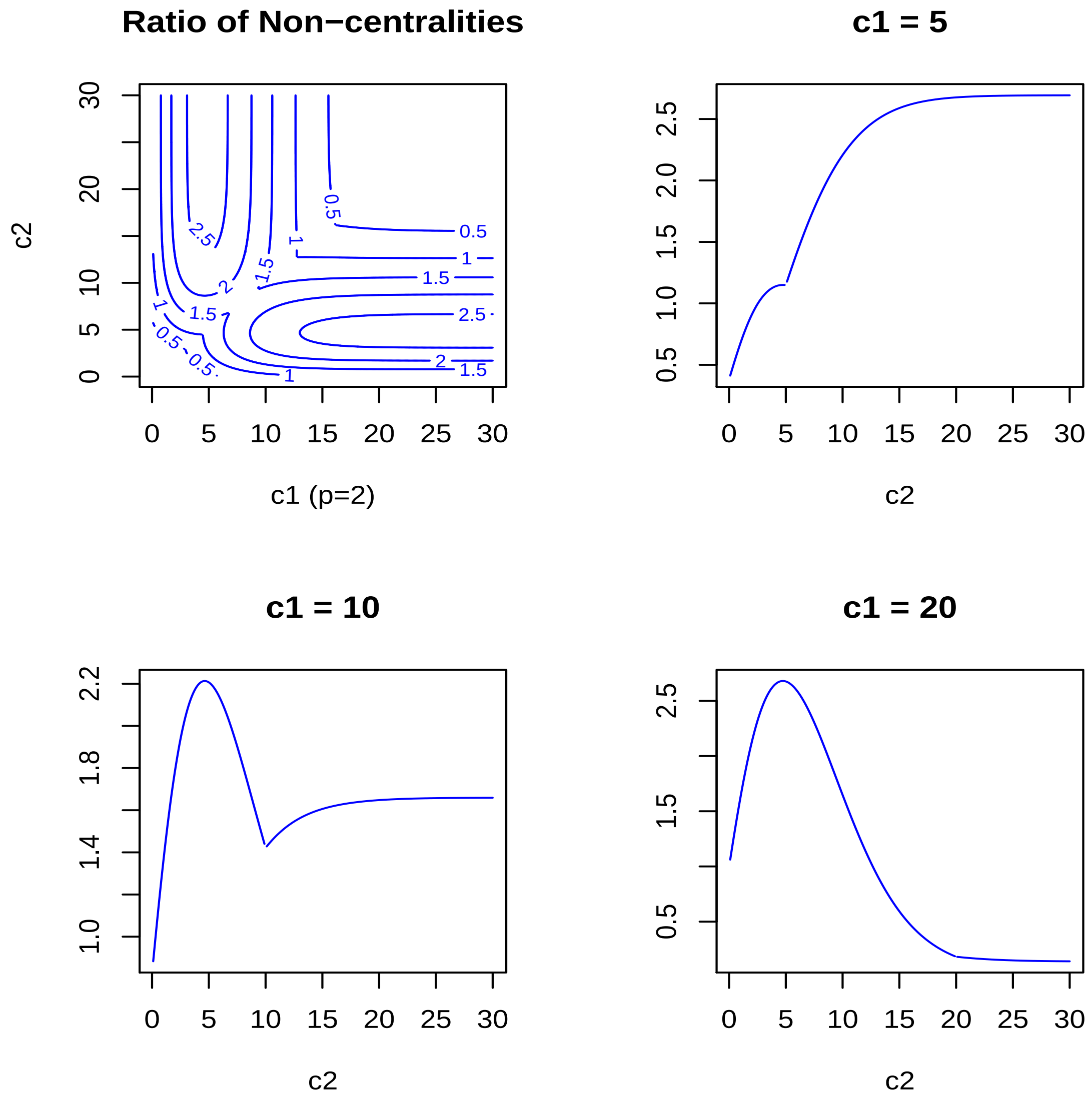

Figure 6.1: ARE of $V_{c_1}^{-1}V_{c_2}$ with respect to $V_0^{-1}V_{+\infty}$ within Family of Huber M-estimates (p=2)

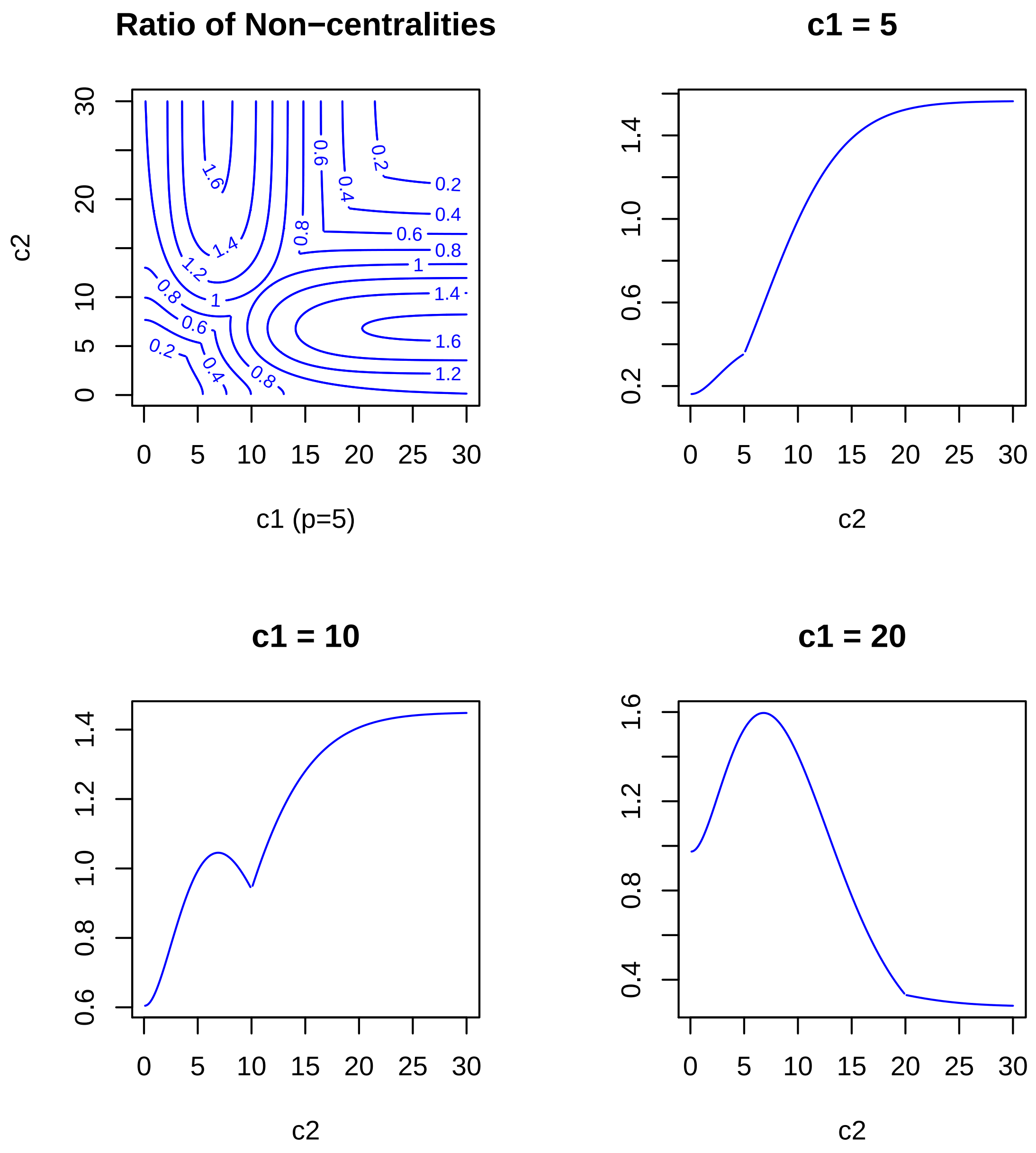

Figure 6.2: ARE of $V_{c_1}^{-1}V_{c_2}$ with respect to $V_0^{-1}V_{+\infty}$ within Family of Huber M-estimates (p=5)

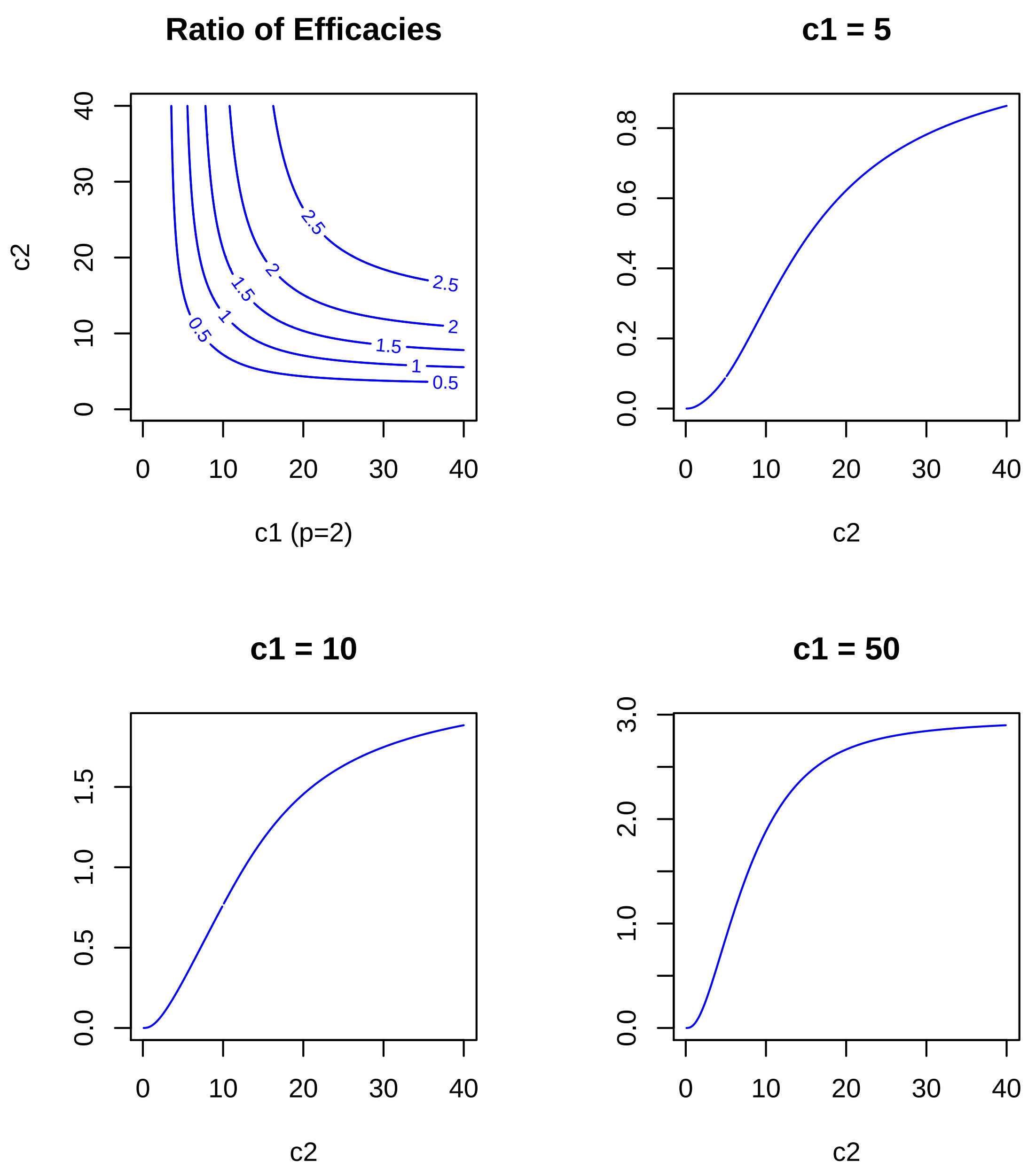

Figure 6.3: ARE of $V_{c_1}^{-1}V_{c_2}$ from Tukey MM-family with respect to $V_0^{-1}V_{+\infty}$ from Huber M-family (p=2)

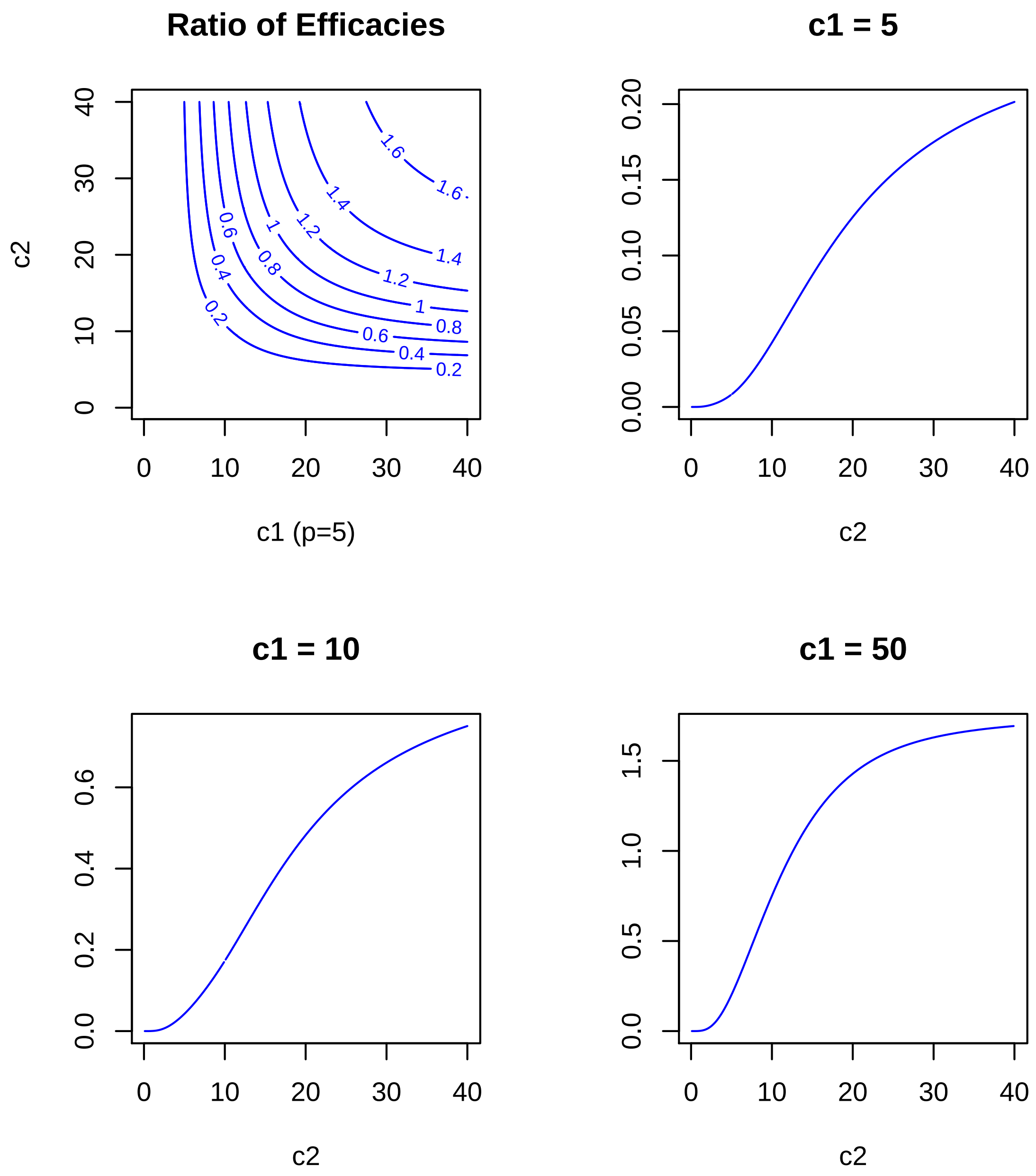

Figure 6.4: ARE of $V_{c_1}^{-1}V_{c_2}$ from Tukey MM-family with respect to $V_0^{-1}V_{+\infty}$ from Huber M-family (p=5)

Chapter 7

Summary and Conclusions

This dissertation presents a pioneering study of the problem of constructing powerful testing procedures based on the maximum difference of two robust statistics under certain asymmetry models. The main results of this thesis are categorized into univariate and multivariate settings.

In the univariate setting, the asymptotic distributions of a class of new statistics $\{T_{c_1,c_2} : T = \widehat{\boldsymbol{\mu}}_{c_1} - \widehat{\boldsymbol{\mu}}_{c_2}, c_i \in [0, +\infty]\}$ are derived for the three most widely cited non-symmetric distributions, where $\widehat{\boldsymbol{\mu}}_{c_i}$ are any two robust location estimates defined by weight functions u_{i,c_i}; most classical skewness measures are special members of this general class, however they are usually not the most efficient and powerful test statistics at testing asymmetry or skewness.

Within this class, the best tests are most likely those based on the difference between two extreme values of location estimates. The statistical theory to find such maximal T is fairly challenging, and we leave that for future research.

On the multivariate shape analysis, families of statistics in the form of

$$\left\{\Delta_{c_1,c_2} : \Delta = \widehat{V}_{c_1}^{-1}\widehat{V}_{c_2}, c_i \in [0, +\infty]\right\}$$

are proposed to detect possible departure of elliptical distributions, where $\widehat{V}_{c_i}$ are any members from the super class of multivariate M-estimates with auxiliary scale. The asymptotic normality of these families is established under skew-elliptical and elliptical mixtures models; asymptotic efficiencies are compared respectively within the family of Huber M-estimates and within the family of Tukey MM-estimates at Normal mixture models . In either family, the Δ based on two extreme members, $\widehat{V}_{+\infty}$ which corresponds to the sample covariance and $\widehat{V}_0$, which in the class of Huber MM-estimates is the so-called Tyler estimate, is not the best test statistic. The influence functions of MM-estimates are also formally derived.

References

[1] Azzalini, A. (1985). *A Class of Distributions which includes the Normal Ones*, Scand. J. Statist. **12**, 171-178.

[2] Azzalini, A., and Dalla Valle, A. (1996) *The Multivariate Skew-Normal Distribution*, Biometrika **83**, 715-726.

[3] Bilodeau, M., and Brenner, D. (1999) *Theory of Multivariate Statistics*, Springer-Verlag, New York.

[4] Bowley, A. L. (1901). *Elements of Statistics* (4th ed., 1920), Scribner's, New York.

[5] David, F. N. and Johnson, N. L. (1956). *Some Tests of Significance with Ordered Variables*, J. Roy. Statist. Soc. Ser. B, **18**, 1-20.

[6] Davies, P. L. (1987) *Asymptotic Behavior of S-estimates of Multivariate Location Parameters and Dispersion Matrices*, Annals of Statistics, **15**, 1269-1292.

[7] Donoho, D. L. (1982) *Breakdown Properties of Multivariate Location Estimates*, Ph.D. qualifying paper, Dept. Statistics, Harvard Univ.

[8] Donoho, D. L., and Huber, P. J. (1983) *The Notion of Breakdown Point*, A Festschrift for Eric Lehmann: In Honour of his Sixty-Fifth Birthday, (P. Bickel, K. Doksum, and J. L. Hodges, eds) 157-184, Wadsworth, Belmont, CA.

[9] Genton, M. G. (Editor) (2004) *Skew-Elliptical Distributions and Their Applications: A Journey Beyond Normality*, Chapman & Hall.

[10] Groeneveld, R.A. (1991) *An Influence Function Approach to Describing the Skewness of a Distribution*, The American Statistician, **45**, 97-102.

[11] Gupta, M. K. (1967). *An Asymptotically Nonparametric test of Symmetry*, Ann. Math. Statist., **38**, 849-866.

[12] Hampel, F. R., Ronchetti, E. M., Rousseeuw, P. J., and Stahel, W. A. (1986) *Robust Statistics: the Approach Based on Influence Functions*, John Wiley & Sons, New York.

[13] Hájek, J. and Zbyněk, Š. (1967). *Theory of Rank Tests*, Academic Press, New York.

[14] Huber, P. J. (1977) *Robust Statistical Procedures*, CBMS-NSF Regional Conference Series in Applied Mathematics, **27**, SIAM, Philadelphia.

[15] Huber, P. J. (1981) *Robust Statistics*, John Wiley & Sons, New York.

[16] Kankainen, A., Taskinen, S., and Oja, H. (2007) *Tests of Multinormality Based on Location Vectors and Scatter Matrices*, Statistical Methods & Applications, **16**, 357-379.

[17] Kent, J. T., and Tyler, D. E. (1991) *Redescending M-Estimates of Multivariate Location and Scatter*, Annals of Statistics, **19**, 2102-2119.

[18] Kent, J. T., and Tyler, D. E. (1996) *Constrained M-Estimation for Multivariate Location and Scatter*, Annals of Statistics, **24**, 1346-1370.

[19] Lehmann, E. L. (1999) *Elements of Large-Sample Theory* Springer-Verlag, New York.

[20] Lopuhaä, H. P. (1989) *On the Relation between S-Estimators and M-Estimators of Multivariate Location and Covariance*, Annals of Statistics, **17**, 1662-1683.

[21] Lopuhaä, H. P. (1999) *Asymptotics of Reweighted Estimators of Multivariate Location and Scatter*, Annals of Statistics, **27**, 1638-1665.

[22] Lopuhaä, H. P., and Rousseeuw, P. J. (1991) *Breakdown Properties of Affine Equivariant Estimators of Multivariate Location and Covariance Matrices*, Annals of Statistics, **19**, 229-248.

[23] MacGillivary, H. L. (1986). *Skewness and Asymmetry: Measures and Ordering*, Ann. Statist., **14**, 994-1011.

[24] Maronna, R. A. (1976) *Robust M-Estimators of Multivariate Location and Scatter*, Annals of Statistics, **4**, 51-67.

[25] Mira, A. (1999) *Distribution-free test for symmetry based on Bonferroni's measure*, J. Applied Statistics, **26**, 959-972.

[26] Oja, H. (1981). *On Location, Scale, Skewness and Kurtosis of Univariate Distributions*, Scand. J. Statist. Soc., **8**, 154-168.

[27] Pearson, K. (1895). *Contributions to the Mathematical Theory of Evolution II. Skew Variation in Homogeneous Material*, Philos. Trans. Roy. Soc. London. Ser. A (1887-1895), **186**, 343-414.

[28] Pitman, E. J. G. (1979) *Some Basic Theory for Statistical Inference* Chapman & Hall, London.

[29] Rieder, H. (1994) *Robust Asymptotic Statistics*, Springer-Verlag Telos.

[30] Rousseeuw, P. J. (1985) *Multivariate Estimation With High Breakdown Point*, Mathematical Statistics and Applications **B** (W. Grossmann, G. Pflug, I. Vincze and W. Wertz, eds.), 283-297, Reidel Publishing Company, Dordrecht.

[31] Rousseeuw, P. J., and Yohai, V. J. (1984) *Robust Regression by Means of S-estimators*, Robust and Nonlinear Time Series Analysis, Lecture Notes in Statistics, **26**, 256-272. Springer, New York.

[32] Tatsuoka, K. S. (1996) *M, CM, and S-estimates: Theory and Computation*, Ph.D Dissertation, Dept. Statistics, Rutgers Univ.

[33] Tatsuoka, K. S., and Tyler, D. E. (2000) *On the Uniqueness of S-Functionals and M-Functionals under Nonelliptical Distributions*, Annals of Statistics, **28**, 1219-1243.

[34] Tukey, J. W. (1977) *Exploratory Data Analysis*, Addison-Wesley.

[35] Tyler, D. E. (1982) *Radial Estimates and the Test of Sphericity*, Biometrika, **69**, 429-436.

[36] Tyler, D. E. (1983) *Robustness and Efficiency Properties of Scatter matrices*, Biometrika, **70**, 411-420.

[37] Tyler, D. E. (1987) *A Distribution-Free M-estimator of Multivariate Scatter*, Annals of Statistics, **15**, 234-251 .

[38] Tyler, D. E. (1991) *Some Issues in the Robust Estimation of Multivariate Location and Scatter*, Directions in Robust Statistics and Diagnostics **2** (W. Stahel and S. Weisberg, eds.) 327 - 336, Springer, New York.

[39] Tyler, D. E. (2002) *High Breakdown Point Multivariate M-Estimations*, Estadística, **54**, 213-274 .

[40] Tyler, D. E., Critchley, F., Dümbgen, L., and Oja, H. (2008) *Invariant Coordinate Selection*, Journal of the Royal Statistical Society: Series B

[41] Wang, J. and Tyler, D. E. (2007). *A Graphical Method for Detecting Asymmetry*, Transactions of 63rd Deming Conference.

[42] Yohai, V. J. (1987) *High Breakdown Point and High Efficiency Robust Estimates for Regression*, Annals of Statistics, **15**, 642-656 .

[43] Yule, G. U. (1911). *Introduction to the Theory of Statistics*, Griffin, London.

Vita

Jue Wang

- **EDUCATION**

2008 Ph.D. in Statistics, Rutgers University.

2002 M. A. in Mathematics, University of South Carolina.

1991 B. S. in Operations Research, Fudan University, Shanghai, China.

- **PROFESSIONAL EXPERIENCE**

2007 - 2008 Intern, Merrill Lynch, New York, NY

Summer 2007 Intern, Travelers Insurance, Hartford, CT

Summer 2006 Intern, Merck Research Laboratories, West Point, PA

2003 - 2007 Teaching/Graduate Assistant, Department of Statistics, Rutgers University.

- **PUBLICATIONS**

Wang, J. and Tyler, D.E. (2007) A Graphical Method for Detecting Asymmetry, *Transactions of the Sixty-third Deming Conference*