1-1-2015

Gini Covariance Matrix and its Affine Equivariant Version

Lauren Anne Weatherall
University of Mississippi

GINI COVARIANCE MATRIX AND ITS AFFINE EQUIVARIANT VERSION

DISSERTATION

A Dissertation
presented in partial fulfillment of requirements
for the degree of Doctor of Philosophy
in the Department of Mathematics
The University of Mississippi

by

LAUREN ANNE WEATHERALL

May 2015

ABSTRACT

Gini's mean difference (GMD) and its derivatives have been widely used as alternative measures of variability over one century in many research fields especially in finance, economics and social welfare. In this dissertation, we generalize the univariate GMD to the multivariate case and propose a new covariance matrix so called the Gini covariance matrix (GCM). The extension is natural, which is based on the covariance representation of GMD with the notion of multivariate spatial rank function. In order to gain the affine equivariance property for GCM, we utilize the transformation-retransformation (TR) technique and obtain TR version GCM that turns out to be a symmetrized M-functional. Indeed, both GCMs are symmetrized approaches based on the difference of two independent variables without reference of a location, hence avoiding some arbitrary definition of location for non-symmetric distributions. We study the properties of both GCMs. They possess the so-called independence property, which is highly important, for example, in independent component analysis. Influence functions of two GCMs are derived to assess their robustness. They are found to be more robust than the regular covariance matrix but less robust than Tyler and Dümbgen M-functionals. Under elliptical distributions, the relationship between the scatter parameter and the two GCM are obtained. With this relationship, principal component analysis (PCA) based on GCM is possible.

Estimation of two GCMs is presented. We study asymptotical behavior of the estimators. $\sqrt{n}$-consistency and asymptotical normality of estimators are established. Asymptotic relative efficiency (ARE) of TR-GCM estimator with respect to the sample covariance matrix is compared to that of Tyler and Dümbgen M-estimators. With little loss on efficiency ($< 2\%$) in the normal case, it gains high efficiency for heavy-tailed distributions. Finite sample behavior of Gini estimators is explored under various models using two criteria. As a by-product, a closely related scatter Kotz functional and its estimator are also studied.

The proposed Gini covariance balances well between efficiency and robustness. In applications, we implement the Gini-based PCA to two real data sets from UCI machine learning repository. Relying on some graphical and numerical summaries, Gini-based PCA demonstrates its competitive performance.

DEDICATION

This dissertation is dedicated to my parents Beecher and Lori Weatherall. They have supported me throught my entire educational career. Without their love and encouragement, this would not have been possible. This work is also dedicated to my grandparents Joseph and Anne Barnett and Doris Weatherall. They have encouraged me to keep going and never give up.

1 Corinthians 13: 11 - 13 “When I was a child, I spake as a child, I understood as a child, I thought as a child: but when I became a man, I put away childish things. For now we see through a glass, darkly; but then face to face: now I know in part; but then shall I know even as also I am known. And now abideth faith, hope, love; but the greatest of these is love.”

ACKNOWLEDGEMENTS

I am especially thankful to my advisor Dr. Xin Dang, who accepted me as her student when I needed research guidance the most. Without her continuous support and inspiration during the course of my research, this work would have never been possible. I am sincerely grateful for Dr. Dang's vast knowledge, patience, and eagerness to help me learn. Her dedication made my research worthwhile, and I can not thank her enough for this. It is because of her that I am able to successfully complete this work.

I would also like to thank Dr. Hailin Sang, Dr. Martial Longla, and Dr. John Bentley for dedicating their time to serve on my committee. Their support and guidance helped navigate me through both the writing and presentation processes. A special thanks to Dr. Sang for his joint effort on this research and research paper. To Dr. Bentley, thank you for providing interesting ideas to add to this work. Dr. Gerard Buskes, thank you for providing me financial support through the Graduate Assistance in Areas of National Need (GAANN) Fellowship. I greatly appreciate the GAANN Fellowship, and, because of this assistance, the stress of tuition and other research expenses was allieviated. I would like to thank Dr. Donald Cole for his support. He has helped and encouraged me throughout my graduate career. Lanzhen Song gave me the wonderful opportunity of piloting a newly structured statistics class while in graduate school. I am grateful for the experience and her confidence in me as an instructor. The office staff and custodians Colleen Gilman, Casandra Jenkins, Shelia Lewis, and Kay Phillips were always there to help in any way needed.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

1 INTRODUCTION

1.1 MOTIVATION

Sample covariance matrix plays an essential role in classical multivariate statistical inference methods including multivariate analysis of variance, principal components analysis, factor analysis, and canonical correlation analysis. These moment-based techniques are optimal (most efficient) under the normality distributional assumption. They are, however, extremely sensitive to outlying observations and susceptible to small perturbations in data. A straightforward treatment is to replace the sample covariance matrix with a robust one. A variety of robust estimates of scatter matrix have been proposed. Some include the following: M-estimates (see [17]), S-estimates (see [2]), MCD-estimates (see [24]), depth (projection)-based estimates (see [38, 36]), as well as sign and rank covariance estimates (see [31, 35]). But those methods may trade off too much efficiency for robustness. For example, Tyler M-estimates only have 50% efficiency under normal distributions (see [30]). In this dissertation, we propose a new scatter estimator that has a good balance between efficiency and robustness.

The new scatter estimator is motivated by the univariate Gini mean difference (GMD). Rather than the assumption on the finite second moment as the variance, the GMD only requires existence of the finite mean of the distribution (see [33]). Hence, the GMD is more robust than the variance, and it is often used for heavy-tailed asymmetric distributions. On the other hand, it is highly efficient. The relative efficiency of the sample GMD with respect to sample standard deviation under normal distributions is about 98% (see [7]). With a little loss in efficiency, the GMD gains robustness against departures from normal distributions.

We extend the univariate GMD to the multivariate case and propose the so called Gini Covariance Matrix (GCM).

This chapter lays out preliminary concepts related to my dissertation work. We start with univariate scale measures and introduce the GMD and its covariance representations. For the multivariate case, we focus on the family of elliptical distributions. Then existing multivariate scatter functionals along with their estimators are introduced. Further, efficiency and robustness are discussed. A *terminology* section introduces terms and theorems seen throughout the dissertation. At the end of this chapter, the contributions of this dissertation are itemized and the structure of remaining chapters is described.

1.2 STANDARD DEVIATION VS GINI MEAN DIFFERENCE

A fundamental problem in statistical analyses is to determine the variability of a data set. When working with univariate data, measures of scale are ways to describe the variability that we are interested in.

Definition 1.2.1. *In order for a parameter σ to be a univariate measure of scale, it must be* scale equivariant *as well as* location and sign invariant*. That is, for any $a, b \in \mathbb{R}$, $\sigma(ax+b) = |a|\sigma(x)$.*

There are two types of measures of scale. One is based on measures about the deviation from a measure of center of the distribution. The other one is based on measures about distance of two independent random variables. This second type of scale measure does not need a center reference. They are also called symmetrized scales. Next, we look at commonly used dispersion measures and their comparison.

Definition 1.2.2. *For a univariate random variable X with cumulative distribution function $F(x)$, the expected value (or mean) is defined as $\mathbb{E}[X] := \int_{-\infty}^{\infty} xd(F(x))$.*

The standard deviation can be described as the square root of the expected squared distance from a random variable to its mean. Therefore, if X is a random variable from a univariate distribution F, then the standard deviation of X (or F) is written as

$$\sigma_s(X) = \sigma_s(F) := \sqrt{\text{var}(X)} = \sqrt{\mathbb{E}(X - \mathbb{E}(X))^2}. \tag{1.2.3}$$

As seen in equation (1.2.3), the standard deviation requires the assumption that there exists a finite second moment. Another measure of dispersion uses *expected absolution deviation to the mean* of the distribution so called the Kotz scale, which only requires the first moment. It is written as

$$\sigma_k(X) = \sigma_k(F) := \mathbb{E}|X - \mathbb{E}(X)|. \tag{1.2.4}$$

On the other hand, *median absolute deviation about the median* (MAD) uses the median as the measure of center. It is written as follows:

$$\sigma_m(X) = \sigma_m(F) = Med|X - Med(X)|,$$

where $Med(X) = Med(F) = \inf_x\{x|F(x) \geq 1/2\}$. All of the measures of scale above require a measure of center.

The second type of scale measures are based on a pair of random variables. They measure distance between two independent random variables X_1 and X_2 without referencing a center. Interestingly, the standard deviation can also be written in this way. The standard deviation is

$$\sigma_s(X) = \sigma_s(F) := \sqrt{\frac{1}{2}\mathbb{E}(X_1 - X_2)^2}. \tag{1.2.5}$$

Another common scale of this type is called the Gini Mean Difference (GMD). It is defined as the expected absolute difference between two independent random variables X_1 and X_2 from F. The GMD of X (or F) is

$$\sigma_g = \sigma_g(X) = \sigma_g(F) = \mathbb{E}|X_1 - X_2|. \tag{1.2.6}$$

Clearly, σ_g is the symmetrized version of the Kotz scale σ_k of (1.2.4).

The GMD was first introduced by Corrado Gini in 1912 as an alternative measure of variability. Since then, the GMD and its derivatives such as Gini index have been widely used in a variety of research fields especially in finance, economics and social welfare (see [34]). The GMD is more robust than the variance, and it is often used for heavy-tailed, asymmetric distributions. On the other hand, it is highly efficient. With a little loss in efficiency, the GMD gains robustness against departures from normal distributions. In this dissertation, our goal is to extend the GMD to the multivariate case. To do so, we must look at the other representations of the GMD.

1.2 Other Representations of GMD

There are several covariance formulations for the GMD. Some of these representations depend on the following: if we let X_1 and X_2 be two independent variables from F with mean μ, then

$$|X_1 - X_2| = (X_1 + X_2) - 2\min\{X_1, X_2\} \tag{1.2.7}$$

Using this equation, the GMD can be written as

$$\sigma_g = 2\mu - 2\mathbb{E}[\min\{X_1, X_2\}]. \tag{1.2.8}$$

In other words, the Gini Mean Difference can be written as twice the difference in the expected value of a random variable and the expected value of the minimum of two random variables from the distribution (see [4]).

While the variance is the covariance of X with itself ($\sigma^2 = \mathrm{Cov}(X, X)$), the GMD is four times the covariance of X with its cumulative distribution $F(X)$.

Claim 1.2.9. *Another formulation of the GMD can be written as*

$$\sigma_g := 4\mathbb{E}\{X(F(X) - \mathbb{E}[F(X)])\} = 4\,Cov(X, F(X)). \tag{1.2.10}$$

Proof of Claim 1.2.9. Let X_1 and X_2 be independently and identically distributed (i.i.d.) from F with density f. Taking the expectation of Formula (1.2.7), we have

$$\mathbb{E}|X_1 - X_2| = \mathbb{E}(X_1 + X_2) - 2\mathbb{E}[\min\{X_1, X_2\}] = 2\mathbb{E}(X_1) - 2\mathbb{E}[\min\{X_1, X_2\}].$$

Let $Y = \min(X_1, X_2)$, then Y has density $2(1 - F(y))f(y)$. From Formula (1.2.8) we have

$$\begin{aligned}\mathbb{E}|X_1 - X_2| = 2\int_{-\infty}^{\infty} xf(x)dx - 2\int_{-\infty}^{\infty} x2(1 - F(x))f(x)dx \\ = 2\int_{-\infty}^{\infty} x(2F(x) - 1)f(x)dx \\ = 4\int_{-\infty}^{\infty} x(F(x) - \frac{1}{2})f(x)dx \\ = 4\mathbb{E}[X(F(X) - \frac{1}{2})] \\ = 4Cov(XF(X))\end{aligned}$$

The last equality holds since $\mathbb{E}[F(X)] = \frac{1}{2}$. This concludes the proof of the claim.

$\square$

From Formula (1.2.10), there are two natural extensions of the Gini covariance of two variables X and Y as follows.

$$\text{Cov}_g(X, Y) = 4\text{Cov}(X, F_Y(Y)), \;\; \text{Cov}_g(Y, X) = 4\text{Cov}(Y, F_X(X)).$$

These formulations may be useful, but there is a major drawback. When using these formulas, there is asymmetry between X and Y. In general, $\text{Cov}_g(X, Y) \neq \text{Cov}_g(Y, X)$. It is even possible for $\text{Cov}_g(X, Y)$ and $\text{Cov}_g(Y, X)$ to have different signs in some cases, which brings extreme difficulty in interpretation (see [27]).

From $\sigma_g = 2\int_{-\infty}^{\infty} x(2F(x)-1)f(x)dx$, another GMD formulation is

$$\sigma_g = 2\mathbb{E}[X(2F(X)-1)] = 2\mathrm{Cov}(X, 2F(X)-1) = 2\mathrm{Cov}(X, r(X)) = 2\mathbb{E}(Xr(X)), \quad (1.2.11)$$

allowing an insightful interpretation: $\sigma_g(X)$ is twice of the covariance of X and the *centered* rank function $r(X) = 2F(X) - 1$. The median of F has centered rank 0 and $r(X) \in [-1, 1]$. Center-oriented rank is of vital importance for a rank concept in high dimension where the natural ordering in one dimension no longer exists.

A nice generalization of the rank function in high dimension and the representation of GMD in (1.2.11) shall yield a natural extension of GMD for a multivariate distribution F. We call this extenstion the Gini Covariance Matrix (GCM) which has the following form

$$\boldsymbol{\Sigma}_g := \boldsymbol{\Sigma}_g(\boldsymbol{X}) := 2\mathbb{E}\boldsymbol{X}\boldsymbol{r}^T(\boldsymbol{X}) \quad (1.2.12)$$

where $\boldsymbol{r}(\boldsymbol{X})$ is the spatial rank function discussed in Chapter 2. We use boldface lowercase letters to represent vectors and random vectors from this point on.

In this dissertation, we focus on elliptical distributions. We explore the relationship of the GCM and the scatter parameter of elliptical models.

1.3 ELLIPTICAL MODELS

Definition 1.3.1. *Let d be a positive integer. A d-dimensional random vector $\boldsymbol{x}$ has an absolutely continuous elliptical distribution if and only if its density function is of the form*

$$f(\boldsymbol{x}|\boldsymbol{\mu}, \boldsymbol{\Sigma}) = |\boldsymbol{\Sigma}|^{-1/2} g\{(\boldsymbol{x}-\boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu})\}, \quad (1.3.2)$$

for some positive definite symmetric matrix scatter parameter $\boldsymbol{\Sigma}$, a vector $\boldsymbol{\mu}$, and some nonnegative function g (free of $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$) with $\int_0^{\infty} t^{d/2-1} g(t)dt < \infty$.

The parameter $\boldsymbol{\mu}$ is the symmetric center, and it equals the first moment, if it exists. The scatter parameter $\boldsymbol{\Sigma}$ is proportional to the covariance matrix when it exists. The k^{th} moment of $\boldsymbol{x}$ exits if $t^{(d+k)/2-1}g(t)$ is integrable. In addition, the variates $r = \|\boldsymbol{\Sigma}^{-1/2}(\boldsymbol{x}-\boldsymbol{\mu})\|$ and $\boldsymbol{u} = \{\boldsymbol{\Sigma}^{-1/2}(\boldsymbol{x}-\boldsymbol{\mu})\}/r$ are independent with $\boldsymbol{u}$ being uniformly distributed on the unit sphere and r having density

$$f_r(r) = \frac{2\pi^{d/2}}{\Gamma(d/2)} r^{d-1} g(r^2), \tag{1.3.3}$$

where $\Gamma(a) = \int_0^\infty t^{a-1}e^{-t}dt$. The derivation of (1.3.3) goes through a spherical coordinate transformation (see [6]). Note that if the covariance matrix of $\boldsymbol{x}$ exists, it equals $\{\mathbb{E}r^2/d\}\boldsymbol{\Sigma}$. The family of elliptical distributions is denoted as $\mathcal{E}(\boldsymbol{\mu}, \boldsymbol{\Sigma}, g)$. If $\boldsymbol{\mu} = \mathbf{0}$ and $\boldsymbol{\Sigma} = \boldsymbol{I}_d$, the identity matrix, we say that the distribution is *spherically symmetric* and we denote it as $F_0(g)$.

The family of elliptical distributions contains a quite rich collection of models. Perhaps the most widely used one is the Gaussian family for which

$$g(t) = (2\pi)^{-d/2}e^{-t/2}. \tag{1.3.4}$$

To model data with heavy-tailed regions, t distributions are commonly used. For t distributions, we have

$$g(t) = \frac{\Gamma[(\nu+d)/2]}{\Gamma(\nu/2)(\nu\pi)^{d/2}}(1+t/\nu)^{-(d+\nu)/2}, \tag{1.3.5}$$

where ν is the degree freedom. ν determines the fatness of the tail regions. For $\nu = 1$, it is called d-variate Cauchy distribution which has very heavy tails; its first moment doesn't exist. When $\nu \to \infty$, it yields the Gaussian distribution.

A quite flexible elliptical family is the Kotz type family, for which the density is of the form in Formula (1.3.2) with

$$g(t) = c(d, \alpha, \beta, \gamma)t^{\alpha-1}e^{-\gamma t^{\beta}}.$$

The parameters are $\beta, \gamma > 0$, $\alpha > 1 - d/2$ and $c(d, \alpha, \beta, \gamma)$ is the normalization constant ([19, 23]). Clearly, when $\beta = 1$, $\alpha = 1$ and $\gamma = 1/2$, the distribution reduces to the normal distribution. The heaviness (or lightness) of the tail regions of distributions mainly depends on β. In particular, we are interested in the special case of $\beta = 1/2$, $\alpha = 1$ and $\gamma = 1$, that is,

$$g(t) = \frac{\Gamma(d/2)}{2\pi^{d/2}\Gamma(d)} e^{-\sqrt{t}}. \tag{1.3.6}$$

We call it the Kotz distribution. For $d = 1$, the Kotz distribution reduces to the Laplace distribution. Not only can the Kotz distribution be viewed as a multivariate generalization of the Laplace distribution, it also has a close connection with our proposed Gini Covariance Matrix as we see in later sections (see [15]).

1.3 Decomposition of a Scatter Matrix

We now consider the scatter matrix ($\mathbf{\Sigma}$) decomposition and discuss the shape matrix. The scatter parameter matrix $\mathbf{\Sigma}$ has an eigen-decomposition

$$\mathbf{\Sigma} = \boldsymbol{U}\mathbf{\Lambda}\boldsymbol{U}^T, \tag{1.3.7}$$

where $\mathbf{\Lambda}$ is the diagonal matrix of corresponding eigenvalues ($\mathbf{\Lambda} = diag(\lambda_1, ..., \lambda_d)$) and $\boldsymbol{U}$ is the matrix of unit eigenvectors. Note that $det(\mathbf{\Sigma}) = det(\mathbf{\Lambda})$, which is called the Wilks generalized variance. The geometric mean of the eigenvalues to the power d is the Wilks generalized variance (this is one example of a "global" measurement of multivariate scatter). A second "global" multivariate scatter measurement is $trace(\mathbf{\Sigma}) = trace(\mathbf{\Lambda})$, which is the sum of the eigenvalues. In this dissertation, we use the second type of "global" measurement as the size of a matrix, and the shape matrix is defined as

$$\boldsymbol{W} = \frac{d}{Tr(\mathbf{\Sigma})}\mathbf{\Sigma}. \tag{1.3.8}$$

It is easy to see that W is standardized to have trace $= d$. We will show how important the consideration of the shape matrix W is, particularly when comparing different scatter estimators. It is important because scatter estimators estimate $\boldsymbol{\Sigma}$ with different factors. With shape matrix estimators, they all estimate the same quantity and can be compared easily with no correction factor. For detailed and comprehensive accounts on elliptical models (see [6]).

1.4 MULTIVARIATE SCATTER FUNCTIONALS

Definition 1.4.1. *Let F be a cumulative distribution function on $\mathbb{R}^d$ and $T : F \to T(F) \in \mathcal{M}^+$, where $\mathcal{M}^+$ is the set of $d \times d$ positive definite matrices. $T(F)$ is a* scatter functional *if, for $\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x} + \boldsymbol{b}$ with any nonsingular matrix $\boldsymbol{A}$ and d-vector $\boldsymbol{b}$, $T(F_{\boldsymbol{y}}) = \kappa_1 \boldsymbol{A} T(F_{\boldsymbol{x}}) \boldsymbol{A}^T$, with $\kappa_1 = \kappa_1(\boldsymbol{A}, \boldsymbol{b}, F_{\boldsymbol{x}})$ a positive scalar function of $\boldsymbol{A}$, $\boldsymbol{b}$, and $F_{\boldsymbol{x}}$.*

In the multivariate case, scatter functionals should possess the affine equivariance property, which is similar to the scale equivariant and location invariant properties in the univariate case.

One of the most common affine equivariant scatter matrices is the covariance matrix, defined as

$$\boldsymbol{\Sigma}_C(\boldsymbol{x}) := \boldsymbol{\Sigma}_C(F) = \mathbb{E}(\boldsymbol{x} - \boldsymbol{\mu})(\boldsymbol{x} - \boldsymbol{\mu})^T, \tag{1.4.2}$$

where $\boldsymbol{\mu}$ is the finite first moment of F. The covariance matrix is the multivariate extension of variance in Formula (1.2.3). A more general covariance matrix can be written as a weighted covariance matrix $\boldsymbol{\Sigma}_M$ called the M-functional (for more details see [17]).

1.4 M-Functionals and Symmetrized M-functionals

The multivariate M-functional $\boldsymbol{\Sigma}_M(F)$ introduced by [17] is defined as the solution of the following equation

$$\mathbb{E}u(r)(\boldsymbol{x}-\boldsymbol{\mu})(\boldsymbol{x}-\boldsymbol{\mu})^T - v(r)\boldsymbol{\Sigma}_M = \mathbf{0}, \tag{1.4.3}$$

where $r = \sqrt{(\boldsymbol{x}-\boldsymbol{\mu})^T\boldsymbol{\Sigma}_M^{-1}(\boldsymbol{x}-\boldsymbol{\mu})}$ which is the Mahalanobis distance (see [16]) of $\boldsymbol{x}$ with respect to $\boldsymbol{\mu}$, $\boldsymbol{\Sigma}_M$, and u , v are nonnegative real-valued functions. $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}_M$ are implicitly defined for some choices of u and v, so we can not get an explicit solution.

When $u(t) = 1$ and $v(t) = 1$, we get the regular covariance matrix $\boldsymbol{\Sigma}_C$. If $u(t) = 1/t^2$, $v(t) = 1$, and $tr(\boldsymbol{\Sigma}_T) = d$, we have the Tyler M-functional $\boldsymbol{\Sigma}_T$. Clearly, $\boldsymbol{\Sigma}_T$ is a shape matrix. In the case of $u(t) = 1/t$ and $v(t) = 1$, we get the Kotz functional, denoted $\boldsymbol{\Sigma}_K$. The rationale for such a name is because it equals the scatter parameter under the Kotz distribution in Formula (1.3.6). For $d = 1$, the Kotz functional is the square of the kotz scale in Formula (1.2.4).

As in the univariate case, we can have scatter matrices based on two independent random vectors $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$ without reference to the location parameter $\boldsymbol{\mu}$. The symmetrized M-functional is defined in terms of the difference of two independent random vectors as the solution to the following:

$$\mathbb{E}u(r_{12})(\boldsymbol{x}_1-\boldsymbol{x}_2)(\boldsymbol{x}_1-\boldsymbol{x}_2)^T - v(r_{12})\boldsymbol{\Sigma}_{SM} = \mathbf{0}, \tag{1.4.4}$$

where $r_{12} = \sqrt{(\boldsymbol{x}_1-\boldsymbol{x}_2)^T\boldsymbol{\Sigma}_{SM}^{-1}(\boldsymbol{x}_1-\boldsymbol{x}_2)}$.

For $u(t) = 1$ and $v(t) = 2$, the covariance matrix is obtained. If $u(t) = 1/t^2$, $v(t) = 1$, and $tr(\boldsymbol{\Sigma}_D) = d$, we obtain the symmetrized Tyler M-functional called Dümbgen M-functional (see [5]). In the case of $u(t) = 1/t$ and $v(t) = 1$, the symmetrized Kotz matrix is obtained. As we will see later, the symmetrized Kotz matrix is the affine equivariant version of our proposed Gini Covariance matrix.

If the components of $\boldsymbol{x}$ are independent, then $\boldsymbol{\Sigma}_{SM}$ is diagonal. Such a property holds naturally for the covariance matrix, but this may not be true for any M-functional. This independence property is highly important, for example, in independent component analysis (see [11]).

1.5 EMPIRICAL DISTRIBUTION AND INDUCED ESTIMATORS

In general, if a parameter $\boldsymbol{\theta}$ of the distribution F can be written as $T(F)$, a functional of F, an estimate of $\boldsymbol{\theta}$ can be given by $\hat{\boldsymbol{\theta}} = T(F_n)$, where F_n is the empirical distribution of a random sample $\boldsymbol{x}_1$, $\boldsymbol{x}_2$,...,$\boldsymbol{x}_n$ taken from F. The *empirical distribution function* of the sample is the distribution that puts $1/n$ mass probability at each sample point $\boldsymbol{x}_i$. For example, in the univariate case we have

$$F_n(x) = \frac{1}{n}\sum_{i=1}^{n} I(x_i \leq x),$$

where $I(A) = 1$ if A is true and otherwise 0. The variance can be expressed as $\sigma_s^2 = T(F) = \int x^2 dF(x) - (\int x dF(x))^2$, therefore an estimate is expressed as $\hat{\sigma}_s^2 = T(F_n) = \frac{1}{n}\sum_{i=1}^{n} x_i^2 - (\frac{1}{n}\sum_{i=1}^{n} x_i)^2 = 1/n \sum_i x_i^2 - \bar{x}^2$. In the multivariate case, the covariance matrix is defined in Formula (1.4.2). The sample version is

$$\hat{\boldsymbol{\Sigma}}_C = \boldsymbol{\Sigma}_C(F_n) = \frac{1}{n}\sum_{i=1}^{n}(\boldsymbol{x}_i - \bar{\boldsymbol{x}})(\boldsymbol{x}_i - \bar{\boldsymbol{x}})^T. \tag{1.5.1}$$

Similarly, from (1.4.3) we have scatter M-estimators $\hat{\boldsymbol{\Sigma}}_M = \boldsymbol{\Sigma}_M(F_n)$. Assuming that the location parameter $\boldsymbol{\mu}$ is known, $\hat{\boldsymbol{\Sigma}}_M$ is the solution of

$$\frac{1}{n}\sum_{i=1}^{n} u(r_i)(\boldsymbol{x}_i - \boldsymbol{\mu})(\boldsymbol{x}_i - \boldsymbol{\mu})^T - v(r_i)\hat{\boldsymbol{\Sigma}}_M = \boldsymbol{0}, \tag{1.5.2}$$

where $r_i = \sqrt{(\boldsymbol{x}_i - \boldsymbol{\mu})^T \hat{\boldsymbol{\Sigma}}_M^{-1} (\boldsymbol{x}_i - \boldsymbol{\mu})}$. Assuming a known location parameter is to avoid some restrictive regularity conditions for the simultaneous M-estimators. Uniqueness of the joint solution so far has been proved only under symmetric distributions, which is unrealistic for the sample distribution case F_n. However, the symmetrized M-estimators do not suffer from such difficulties. They are defined without reference of the center. Symmetrized M-estimators are the solution of

$$\frac{2}{n(n-1)} \sum_{1 \le i < j \le n}^{n} u(r_{ij}^2)(\boldsymbol{x}_i - \boldsymbol{x}_j)(\boldsymbol{x}_i - \boldsymbol{x}_j)^T - v(r_{ij}) \hat{\boldsymbol{\Sigma}}_{SM} = \mathbf{0}, \tag{1.5.3}$$

where $r_{ij} = \sqrt{(\boldsymbol{x}_i - \boldsymbol{x}_j)^T \hat{\boldsymbol{\Sigma}}_{SM}^{-1} (\boldsymbol{x}_i - \boldsymbol{x}_j)}$ (for more details on symmetric M-estimators refer to [28]).

So far, we have looked at various scatter functionals and their estimators. In some situations, the use of one scatter functional may be more appropriate than another. We shall consider these estimators in regards to two aspects: robustness and efficiency.

1.6 ROBUSTNESS

When statistical inferences are made, they are based on observations as well as assumptions about the underlying distribution or prior information. Ideally, these assumptions would be met; however, they do not always hold in real world examples. The term robustness can have many meanings and some of these may be inconsistent. Huber and Ronchetti (2009, p.2) [10] define robustness in the following sense:

> Robustness signifies insensitivity to small deviations from the assumptions.

In most cases we are interested in *distributional robustness*, which means that the true distribution differs slightly from the assumed model. Robust methods are used to insure the following: if there are small deviations of the assumptions, then there will only be small changes in the expected result under the assumed model.

We are interested in the behavior of $T(F)$ or $T(F_n)$ when F is approximately known. So, we consider F_ϵ in some neighborhood of the true distribution F. A *contamination neighborhood* is defined as

$$\mathcal{F}_\epsilon(F,\epsilon) = \{(1-\epsilon)F + \epsilon G = F_\epsilon . G \in \mathcal{G}\}, \tag{1.6.1}$$

where $\mathcal{G}$ is a set of distributions of interest.

Definition 1.6.2. *The* Gâteaux derivative *of T at F in the direction of G is defined*

$$L_F(G) = \lim_{\epsilon \to 0} \frac{T(F+\epsilon(F-G)) - T(F)}{\epsilon} = \frac{\partial T(F_\epsilon)}{\partial \epsilon}|_{\epsilon=0}. \tag{1.6.3}$$

A special case of the above definition is when $\mathcal{G}$ is a set of point mass distributions. In this case we get the influence function (IF) of T at F. We will consider the IF for scatter matrices.

Definition 1.6.4. *Let F be a cumulative distribution function on $\mathbb{R}^d$ and $T : F \to T(F) \in \mathcal{M}^+$, where $\mathcal{M}^+$ is the set of $d \times d$ positive definite matrices. Then the* influence function *(IF) of T at F is written as*

$$IF(\boldsymbol{x}; T, F) = \lim_{\epsilon \to 0} \frac{T[(1-\epsilon)F + \epsilon\delta_{\boldsymbol{x}}] - T(F)}{\epsilon}, \tag{1.6.5}$$

for $\boldsymbol{x} \in \mathbb{R}^d$ and where $\delta_{\boldsymbol{x}}$ denotes the point mass distribution at $\boldsymbol{x}$.

The influence function provides effects of infinitesimal pertibations of T. For the estimator $T(F_n)$, we have the finite sample version of the influence function called the *sensitivity curve.*

Definition 1.6.6. *Let $\boldsymbol{x}_1$, $\boldsymbol{x}_2$,...,$\boldsymbol{x}_n$ be a random sample with the empirical distribution F_n. If an outlier, say $\boldsymbol{x}_0$ is added to this sample, with the empirical distribution F_{n+1}, then the*

Standardized Sensitivity Curve (SC) *is written as follows.*

$$SC(\boldsymbol{x}_0; T, F_n) = \frac{T(F_{n+1}) - T(F_n)}{1/(n+1)}. \tag{1.6.7}$$

The influence function can be viewed as the "limit version" of the sensitivity curve. That is, $SC(\boldsymbol{x}_0; T, F_n) \overset{n\to\infty}{\longrightarrow} IF(\boldsymbol{x}_0; T, F)$.

The influence function has two main uses. The first is that it is used for assessment of robustness of a functional or an estimate. Ideally, the IF or SC should be bounded, implying bounded change rate due to infinitesimal perturbations. The IF also allows heuristic assessment of the asymptotic properties of $T(F_n)$. If $T(F_n)$ is $\sqrt{n}$-consistent, then IF gives an explicit formula for the asymptotic variance (AV) of $T(F_n)$.

Definition 1.6.8. *For an estimator* $T(F_n)$*, if* $\lim_{n\to\infty} k_n \, Var(T(F_n)) = t^2$*, where* k_n *is a sequence of constants, then* t^2 *is called the* asymptotic variance $AV(T(F_n))$.

For example, if $T(F_n)$ is univariate, then

$$AV(T(F_n)) = \mathbb{E}[IF(X; T, F)]^2 = \int IF(X; T, F)^2 dF(x). \tag{1.6.9}$$

We consider the scatter matrix estimator T under spherical distributions. The asymptotic variance-covariance matrix is defined as

$$AV(vec(T(F_{0,n})) = \mathbb{E}[vec(IF(\boldsymbol{x}; T, F_0))vec(IF(\boldsymbol{x}; T, F_0))^T]. \tag{1.6.10}$$

Here, *vec* denotes the *column vectorization* of a matrix (the operation of stacking each column of a matrix into a vector).

This asymptotic variance will be used to assess the statistical accuracy of a scatter estimator, which is discussed in the next section.

1.7 EFFICIENCY

To assess the performance of a statistic or statistical method, *statistical efficiency* is usually examined. Efficiency generally refers to a particular measure of accuracy of an estimator. Basically, an estimator is more efficient than another if it requires a smaller sample size for the same level of accuracy. Most often, efficiency is in terms of the variance or mean square error. If V_n and W_n are two univariate unbiased estimators of a parameter from the distribution F, then the *Asymptotic Relative Efficiency* (ARE) of V_n with respect to W_n is

$$ARE(V_n, W_n) = \frac{AV(W(F_n))}{AV(V(F_n))}, \tag{1.7.1}$$

where $AV(W(F_n))$ and $AV(V(F_n))$ are the asymptotic variances of W_n and V_n, respectively. If $ARE(V_n, W_n) > 1$, then V_n is asymptotically more efficient than W_n in distribution F.

For the affine equivariant scatter estimators, we will compare asymptotic covariance matrices for efficiency. Asymptotic covariance matrices are given by Formula (1.6.10). Under the normal distribution, $\hat{\mathbf{\Sigma}}_C$ defined by (1.5.1) is more efficient than Tyler, but $\hat{\mathbf{\Sigma}}_C$ is not robust. Tyler M-estimator is robust with bounded SC but is less efficient with ARE 0.5 compared to the covariance matrix. Our proposed Gini Covariance estimators are between those two in terms of robustness and efficiency. That is, our estimator is more efficient than the Tyler M-estimator, but less efficient than the Kotz or Dümbgen estimators.

1.8 TERMINOLOGY

In this section, some terminology that is used in later chapter is defined. First, definitions related to convergence and consistency are given.

Definition 1.8.1. *Let X be a random variable with cdf $F(x, \theta)$ and $\theta \in \Theta$. Let $X_1, X2, ..., X_n$ be a sample from the distribution of X and $\hat{\theta}$ denote an estimate of θ. We say $\hat{\theta}$* converges in probability *to θ, denoted by $\hat{\theta} \xrightarrow{p} \theta$. if $P(\|\hat{\theta} - \theta\| \geq \epsilon) \to 0$ as $n \to \infty$.*

Definition 1.8.2. *A sequence of random variables* $X_1, X_2, ...,$ converges in distribution *to a random variable* X *if* $\lim_{n\to\infty} F_{X_n}(x) = F_X(x)$ *at all points* x *where* $F_X(x)$ *is continuous. We can also write* $F_{X_n}(x) \xrightarrow{D} F_X(x)$ *as* $n \to \infty$.

Definition 1.8.3. *Let* X *be a random variable with cdf* $F(x, \theta)$ *and* $\theta \in \Theta$. *Let* $X_1, X2, ..., X_n$ *be a sample from the distribution of* X *and* T_n *denote a statistic. We say* T_n *is a* consistent *estimator of* θ *if* $T_n \xrightarrow{p} \theta$.

Definition 1.8.4. *If* $T_n \xrightarrow{p} \theta$, *then* T_n *is* $\sqrt{n}$*-consistent if* $T_n = \theta + O_p(1/\sqrt{n})$.

Definition 1.8.5. *Let* $X_1, X2, ..., X_n$ *be a sample with cdf* $F(X, \theta)$. *If* $\hat{\theta} = T(\hat{F}_n)$ *is an estimator of* θ *and* $\hat{F}_n$ *is the empirical distribution function, then the estimator is* Fisher consistent *if* $T(F(X, \theta)) = \theta$.

Next, an explicit formula for U-statistics is given.

Definition 1.8.6. *Let* $X_1, ..., X_n$ *be independently and identically distributed (i.i.d.) from an unknown population* P *in a nonparametric family* $\mathcal{P}$. *Let* $\theta = \mathbb{E}[h(X_1, ..., X_m)]$ *with a positive integer* m *and a symmetric Borel function* h *that satisfies* $\mathbb{E}[h(X_1, ..., X_m)] < \infty$ *for any* $P \in \mathcal{P}$. *A symmetric unbiased estimator of* θ *is*

$$U_n = \binom{n}{m}^{-1} \sum_c h(X_{i1}, ..., X_{im}), \tag{1.8.7}$$

where $\sum_c$ *denotes the summation over the* $\binom{n}{m}$ *combinations of* m *distinct elements* $(i_1, ..., i_m)$ *from* $(1, ..., n)$. *In Formula* (1.8.7), U_n *is called a* U-statistic *with* kernel h *of order* m.

Theorem 1.8.1 (Hoeffdin's theorem)**.** *For a U-statistic* U_n *given by Formula* (1.8.6) *with* $\mathbb{E}[h(X_1, ..., X_m)]^2 < \infty$,

$$Var(U_n) = \binom{n}{m}^{-1} \sum_{k=1}^{m} \binom{m}{k} \binom{(n-m)}{(m-k)} \eta_k,$$

where

$$\eta_k = Var(h_k(X_1, ..., X_k)).$$

Definition 1.8.8. *We say "f is little-o of h as x approaches x_0" and write $f(x) = o(h(x))$ as $x \to x_0$ to mean* $\lim_{x \to x_0} \frac{f(x)}{h(x)} = 0$.

Definition 1.8.9. *Let f and g be two functions defined on some subset of the real numbers. We write*

$$f(x) = O(g(x)) \text{ as } x \to \infty$$

if and only if there is a positive constant M such that for all sufficiently large x, $|f(x)| \leq M|g(x)|$.

Definition 1.8.10. *The* Frobenius Norm *is a matrix norm of an $m \times n$ matrix A defined as*

$$\|A\|_F = \sqrt{\sum_{j=1}^{m} \sum_{i=1}^{n} |a_{ij}|^2}.$$

Theorem 1.8.2 (Brower's Fixed Point)**.** *Any continuous function $g : \mathbb{B}^n \to \mathbb{B}^n$ has a fixed point B_0 such that $g(B_0) = B_0$, where $\mathbb{B}^n = \{\boldsymbol{x} \in \mathbb{R}^n : x_1^2 + ... + x_n^2 \leq 1\}$.*

Definition 1.8.11. *If* $\boldsymbol{A}$ *is an* $m \times n$ *matrix and* $\boldsymbol{B}$ *is a* $p \times q$ *matrix, then the* kronecker product $\boldsymbol{A} \otimes \boldsymbol{B}$ *is the* $mp \times nq$ *block matrix*

$$\boldsymbol{A} \otimes \boldsymbol{B} = \begin{bmatrix} a_{11}\boldsymbol{B} & \cdots & a_{1n}\boldsymbol{B} \\ \vdots & \ddots & \vdots \\ a_{m1}\boldsymbol{B} & \cdots & a_{mn}\boldsymbol{B} \end{bmatrix}.$$

1.9 OVERVIEW

1.9 Contributions

The contributions of this dissertation are:

- *A new covariance matrix.* The Gini Covariance Matrix (GCM) serves as a multivariate extension of the Gini Mean Difference(GMD). The generalization stems from one covariance formulation of the GMD and the spatial rank function.

- *An affine equivariant version of this GCM.* Through the transformation-retransformation (TR) technique, a new affine equivariant version of the GCM is given. This version is a symmetrized M-functional.

- *Developed properties of the GCM and TR GCM.* Their influence functions are derived. They are not robust in a strict sense, but they are more robust than that of covariance matrix.

- *Properties of estimatiors of GCM and TR GCM.* The asymptotic normality is established and asymptotic efficiency is studied. The TR GCM is highly efficient in normal distributions with ARE greater than 98% compared to the regular covariance matrix. Also it is more robust than the covariance matrix in heavy tailed t distribution and Kotz distribution.

- *Study of Kotz functional and Kotz estimator.*

1.9 Dissertation Structure

The remaining chapters of this dissertation are organized as follows. Chapter 2 briefly discusses three types of sign and rank function concepts: marginal, Oja, and spatial signs and ranks. In Chapter 3, the two versions of Gini Covariance Matrix (GCM) and some properties are developed. Chapter 3 also discusses the independence property and gives the robustness in terms of the influence function. Estimation of both versions of the GCM is presented in Chapter 4. In Chapter 5, finite sample efficiency is discussed compared to other robust methods. Chapter 6 presents applications by using the TR Gini in Principal Component Analysis. Finally, Chapter 7 gives the conclusions and possible future work. R codes are listed in the Appendix.

2 MULTIVARIATE SIGN AND RANK FUNCTIONS

In order to extend the univariate Gini Mean Difference in Formula (1.2.11) to the multivariate case, we need a notion of multivariate centered rank. In this chapter, we discuss three rank functions. These functions can be defined as the gradient of some objective functions. Sign functions are defined in a similar manner. For objective functions $H(\boldsymbol{x})$ and $D(\boldsymbol{x}, F)$, the corresponding sign function $\boldsymbol{s}(\boldsymbol{x})$ and rank function $\boldsymbol{r}(\boldsymbol{x}, F)$ are defined as

$$\boldsymbol{s}(\boldsymbol{x}) = \nabla_{\boldsymbol{x}} H(\boldsymbol{x}) \text{ and } \boldsymbol{r}(\boldsymbol{x}, F) = \nabla_{\boldsymbol{x}} D(\boldsymbol{x}, F).$$

The solution of $\boldsymbol{r}(\boldsymbol{x}, F) = \boldsymbol{0}$ provides a notion of multivariate medians $\boldsymbol{m}(F)$, which minimizes $D(\boldsymbol{x}, F)$. The corresponding sign and rank covariance matrices are

$$\mathbb{E}\boldsymbol{s}(\boldsymbol{x} - \boldsymbol{m}(F))\boldsymbol{s}^T(\boldsymbol{x} - \boldsymbol{m}(F)) \text{ and } \mathbb{E}\boldsymbol{r}(\boldsymbol{x}, F)\boldsymbol{r}^T(\boldsymbol{x}, F),$$

respectively. Here $\boldsymbol{r}(\boldsymbol{x}, F)$ is already centered, i.e. $\mathbb{E}\boldsymbol{r}(\boldsymbol{x}, F) = \boldsymbol{0}$ (we will see the proof of this claim in the next section).

In this chapter, the following sign and rank functions are discussed: vectors of marginal signs and ranks, Oja signs and ranks, and spatial signs and ranks. The objective functions and properties of these sign and rank functions are also discussed.

2.1 MARGINAL SIGNS AND RANKS

The marginal sign and rank functions in $\mathbb{R}^d$ are obtained by defining the objective functions as follows:

$$H_1(\boldsymbol{x}) = \|\boldsymbol{x}\|_1 = |x_1| + ... + |x_d| \text{ and}$$

$$D_1(\boldsymbol{x}, F) = \mathbb{E}_{\boldsymbol{y}}\|\boldsymbol{x} - \boldsymbol{y}\|_1,$$

where $\boldsymbol{x} \in \mathbb{R}^d$, $\boldsymbol{y}$ a random vector from continuous distribution F, and we let $\mathbb{E}_{\boldsymbol{y}}$ denote taking the expectation with $\boldsymbol{y}$. Visuri *et al.*([31]) states the corresponding marginal sign and marginal rank functions are

$$\boldsymbol{s}_1(\boldsymbol{x}) = \nabla_{\boldsymbol{x}} H(\boldsymbol{x}) = [\text{sign}(x_1), ...\text{sign}(x_d)]^T \text{ and}$$

$$\boldsymbol{r}_1(\boldsymbol{x}, F) = \nabla_{\boldsymbol{x}} D(\boldsymbol{x}, F) = \mathbb{E}\boldsymbol{s}_1(\boldsymbol{x} - \boldsymbol{y})$$

We have that $\boldsymbol{r}_1$ is centered.

Claim 2.1.1. $\mathbb{E}\boldsymbol{r}_1(\boldsymbol{x}, F) = \mathbf{0}$.

Proof of Claim 2.1.1. Since $\boldsymbol{x}$ and $\boldsymbol{y}$ independent from F, we have

$$\begin{aligned}
\mathbb{E}_{\boldsymbol{x}}\boldsymbol{r}_1(\boldsymbol{x}, F) &= \mathbb{E}_{\boldsymbol{x}}\mathbb{E}_{\boldsymbol{y}}[\boldsymbol{s}_1(\boldsymbol{x} - \boldsymbol{y})] \\
&= \mathbb{E}_{\boldsymbol{y}}\mathbb{E}_{\boldsymbol{x}}[\boldsymbol{s}_1(\boldsymbol{x} - \boldsymbol{y})] \\
&= -\mathbb{E}_{\boldsymbol{y}}\mathbb{E}_{\boldsymbol{x}}[\boldsymbol{s}_1(\boldsymbol{y} - \boldsymbol{x})] \\
&= -\mathbb{E}_{\boldsymbol{y}}\boldsymbol{r}_1(\boldsymbol{y}, F) \\
&= -\mathbb{E}_{\boldsymbol{x}}\boldsymbol{r}_1(\boldsymbol{x}, F).
\end{aligned}$$

This concludes the proof of the claim.

□

If F is continuous, the solution of $\boldsymbol{r}_1(\boldsymbol{x}, F) = \mathbf{0}$ exists uniquely, and the solution is the marginal median, which consists of the component-wise medians

$$\boldsymbol{m}_1(F) = [\text{med}(y_1), ..., \text{med}(y_d)]^T.$$

We have the vectors of centered marginal signs $\boldsymbol{s}_1(\boldsymbol{x} - \boldsymbol{m}_1(F))$ and the centered marginal ranks $\boldsymbol{r}_1(\boldsymbol{x}, F)$. Let the sign covariance matrix (SCM) and the Spearman's rank covariance matrix (RCM) be defined as

$$SCM_1 = \mathbb{E}\{\boldsymbol{s}_1(\boldsymbol{x} - \boldsymbol{m}_1(F))\boldsymbol{s}_1^T(\boldsymbol{x} - \boldsymbol{m}_1(F))\}, \tag{2.1.2}$$

$$RCM_1 = \mathbb{E}\{\boldsymbol{r}_1(\boldsymbol{x}, F)\boldsymbol{r}_1^T(\boldsymbol{x}, F)\} = \mathbb{E}\{\boldsymbol{s}_1(\boldsymbol{x}_1 - \boldsymbol{x}_2)\boldsymbol{s}_1^T(\boldsymbol{x}_1 - \boldsymbol{x}_2)\}, \tag{2.1.3}$$

where $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$ are i.i.d. copies of the r.v. $\boldsymbol{x}$.

The marginal sign and rank functions are scale invariant but not rotation equivariant. The orientation and shape information (i.e. eigenvalues and eigenvectors) are not invariant to these transformations. Using these sign and rank covariance matrices leads to a component-wise approach that will ignore correlation between variables. The marginal rank function also suffers from this same problem (see [31]).

2.2 OJA SIGNS AND RANKS

The volume of a simplex defined by $\boldsymbol{x}, \boldsymbol{y}_1, ..., \boldsymbol{y}_d$ distinct points is

$$V = \frac{1}{d!}\left| det \begin{pmatrix} 1 & ... & 1 & 1 \\ \boldsymbol{y}_1 & ... & \boldsymbol{y}_d & \boldsymbol{x} \end{pmatrix} \right|.$$

Consider the objective functions,

$$H_2(\boldsymbol{x}, F) = \mathbb{E}\left| det \begin{pmatrix} 1 & 1 & ... & 1 & 1 \\ \boldsymbol{0} & \boldsymbol{y}_1 & ... & \boldsymbol{y}_{d-1} & \boldsymbol{x} \end{pmatrix} \right|,$$

and

$$D_2(\boldsymbol{x}, F) = \mathbb{E}\left| det \begin{pmatrix} 1 & ... & 1 & 1 \\ \boldsymbol{y}_1 & ... & \boldsymbol{y}_d & \boldsymbol{x} \end{pmatrix} \right|,$$

where $\boldsymbol{y}_1, \boldsymbol{y}_2, ..., \boldsymbol{y}_d$ are i.i.d. random variables with distribution F. The Oja sign and rank functions are defined as follows

$$\boldsymbol{s}_2(\boldsymbol{x}, F) = \nabla_{\boldsymbol{x}} H_2(\boldsymbol{x}, F) \text{ and}$$

$$\boldsymbol{r}_2(\boldsymbol{x}, F) = \nabla_{\boldsymbol{x}} D_2(\boldsymbol{x}, F).$$

The solution of $\boldsymbol{r}_2(\boldsymbol{x}, F) = \boldsymbol{0}$ is $\boldsymbol{m}_2(F)$, the affine equivariant Oja median. It was introduced by Oja ([20])and minimizes $D_2(\boldsymbol{x}, F)$. The Oja sign covariance matrix and the Oja rank covariance matrix are defined as follows:

$$SCM_2 = \mathbb{E}\left(\boldsymbol{s}_2(\boldsymbol{x}, F) - \boldsymbol{m_2}(F)\right)\left(\boldsymbol{s}_2(\boldsymbol{x}, F) - \boldsymbol{m_2}(F)\right)^T, \quad (2.2.1)$$

$$RCM_2 = \mathbb{E}\boldsymbol{r}_2(\boldsymbol{x}, F)\boldsymbol{r}_2^T(\boldsymbol{x}, F). \quad (2.2.2)$$

The Oja sign and rank covariance matrices are affine equivariant in the sense that if $\boldsymbol{x}_i^* = \boldsymbol{A}\boldsymbol{x}_i + \boldsymbol{b}$, where $\boldsymbol{A}$ is a non-singular $d \times d$ matrix, and $\boldsymbol{b}$ is a d-variate vector, then the sign and rank covariance matrix on the transformed data satisfies $SCM_2^* = \boldsymbol{A}^* SCM_2 \boldsymbol{A}^{*T}$, and $RCM_2^* = \boldsymbol{A}^* RCM_2 \boldsymbol{A}^{*T}$, where $\boldsymbol{A}^* = |(det(\boldsymbol{A}))(\boldsymbol{A}^{-1})^T|$.

The major problem of Oja signs and ranks is the computation. The sample version of D_2 is given by the following formula:

$$D_2(\boldsymbol{x}, F_n) = \binom{n}{d}^{-1} \sum_{1 \le i_1 < ... < i_d \le n} \left| det \begin{pmatrix} 1 & ... & 1 & 1 \\ \boldsymbol{y}_{i_1} & ... & \boldsymbol{y}_{i_d} & \boldsymbol{x} \end{pmatrix} \right|.$$

From the sample version of D_2 we can see that the calculation for the sample Oja signs and ranks is very heavy. There are $\binom{n}{d}$ combinations that need to be computed. For each combination we need to calculate the determinant of a $(d+1) \times (d+1)$ matrix, whose computation is difficult for large d. Therefore, using the Oja sign and rank functions is not ideal for moderate n, even for a large sample size it is not feasible.

2.3 SPATIAL SIGNS AND RANKS

For spatial sign and rank functions, we consider the following objective functions:

$$H_3(\boldsymbol{x}) = \|\boldsymbol{x}\| = \sqrt{x_1^2 + ... + x_d^2} \text{ and}$$

$$D_3(\boldsymbol{x}, F) = \mathbb{E}_{\boldsymbol{y}} \left[\|\boldsymbol{x} - \boldsymbol{y}\| | \boldsymbol{x} \right],$$

then the spatial sign ($\boldsymbol{s}_3$) and spatial rank ($\boldsymbol{r}_3$) functions can be written

$$\boldsymbol{s}_3(\boldsymbol{x}) = \nabla_{\boldsymbol{x}} H_3(\boldsymbol{x}) = \frac{\boldsymbol{x}}{\|\boldsymbol{x}\|}, \tag{2.3.1}$$

and

$$\boldsymbol{r}_3(\boldsymbol{x}, F) = \nabla_{\boldsymbol{x}} D_3(\boldsymbol{x}, F) = \mathbb{E}\boldsymbol{s}_3(\boldsymbol{x} - \boldsymbol{y}) = \mathbb{E}_{\boldsymbol{y}} \frac{\boldsymbol{x} - \boldsymbol{y}}{\|\boldsymbol{x} - \boldsymbol{y}\|}, \tag{2.3.2}$$

where we define $\boldsymbol{s}(\boldsymbol{0}) = \boldsymbol{0},\ \boldsymbol{x} \neq \boldsymbol{0}$.

The spatial rank function is the *expected direction* to $\boldsymbol{x}$ from random variable $\boldsymbol{y}$. It is also centered, since $\mathbb{E}\boldsymbol{r}_3(\boldsymbol{x}, F) = \boldsymbol{0}$. The solution of $\boldsymbol{x}$ in $\boldsymbol{r}_3(\boldsymbol{x}, F) = \boldsymbol{0}$ is called the *spatial* *median* $\boldsymbol{m}_3(F)$, which minimizes $D_3(\boldsymbol{x}, F) = \mathbb{E}_{\boldsymbol{y}}\left[\|\boldsymbol{x} - \boldsymbol{y}\|\right]$. By Hölder's inequality, we have $\|\boldsymbol{r}_3(\boldsymbol{x}, F)\| \leq 1$ for all $\boldsymbol{x}$. Under weak assumptions on F, $\boldsymbol{r}_3(\boldsymbol{x}, F)$ maps $\boldsymbol{x}$ to a vector inside the unit ball with magnitude $\|\boldsymbol{r}_3(\boldsymbol{x}, F)\|$, and the center of the unit ball is the spatial median of F (see [13]).

The rank covariance matrix is the covariance matrix of the spatial rank defined by

$$\begin{aligned} RCM_3 &= \mathbb{E}_{\boldsymbol{x}}\{\boldsymbol{r}_3(\boldsymbol{x}, F)\boldsymbol{r}_3^T(\boldsymbol{x}, F)\} = \mathbb{E}_{\boldsymbol{x}_1}\left([\mathbb{E}_{\boldsymbol{x}_2}\boldsymbol{s}_3(\boldsymbol{x}_1, \boldsymbol{x}_2)]\,[\mathbb{E}_{\boldsymbol{x}_3}\boldsymbol{s}_3(\boldsymbol{x}_1, \boldsymbol{x}_3)]^T\right) \\ &= \mathbb{E}_{\boldsymbol{x}_1}\left(\mathbb{E}_{\boldsymbol{x}_2, \boldsymbol{x}_3}\left[\boldsymbol{s}_3(\boldsymbol{x}_1, \boldsymbol{x}_2)\boldsymbol{s}_3^T(\boldsymbol{x}_1, \boldsymbol{x}_3)\right]\right) = \mathbb{E}\frac{(\boldsymbol{x}_1 - \boldsymbol{x}_2)(\boldsymbol{x}_1 - \boldsymbol{x}_3)^T}{\|\boldsymbol{x}_1 - \boldsymbol{x}_2\|\|\boldsymbol{x}_1 - \boldsymbol{x}_3\|}, \end{aligned}$$

where $\boldsymbol{x}, \boldsymbol{x}_1, \boldsymbol{x}_2, \boldsymbol{x}_3$ are independently and identically distributed from F. RCM_3 and its modified version have been studied (see [31, 35]).

The corresponding sign covariance matrix is

$$SCM_3 = \mathbb{E}\{\boldsymbol{s}_3(\boldsymbol{x} - \boldsymbol{m}_3(F))\boldsymbol{s}_3^T(\boldsymbol{x} - \boldsymbol{m}_3(F))\} = \mathbb{E}\frac{(\boldsymbol{x} - \boldsymbol{m}_3(F))(\boldsymbol{x} - \boldsymbol{m}_3(F))^T}{\|\boldsymbol{x} - \boldsymbol{m}_3(F)\|}. \qquad (2.3.3)$$

SCM_3 requires the use of the spatial median $\boldsymbol{m}_3$. The symmetrized sign covariance matrix is defined by the difference approach:

$$SSCM_3 = \mathbb{E}\{\boldsymbol{s}_3(\boldsymbol{x}_1 - \boldsymbol{x}_2)\boldsymbol{s}_3^T(\boldsymbol{x}_1 - \boldsymbol{x}_2)\} = \mathbb{E}\frac{(\boldsymbol{x}_1 - \boldsymbol{x}_2)(\boldsymbol{x}_1 - \boldsymbol{x}_2)^T}{\|\boldsymbol{x}_1 - \boldsymbol{x}_2\|^2}, \qquad (2.3.4)$$

where $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$ are i.i.d from the distribution F. The resulting matrix is known as the symmetrized spatial sign covariance matrix ($SSCM_3$), which has been studied by Croux & Oja and Taskinen *et al.* (see [1, 29]). This symmetrized version uses the pairwise approach in order to avoid needing the median. SCM_3 and $SSCM_3$ use only the directional information of F. Since RCM_3 takes information from three independent random vectors in its definition, the sample RCM_3 is more efficient than the sample $SSCM_3$.

The spatial rank function has the following nice properties. The rank function $\boldsymbol{r}_3(\boldsymbol{x}, F)$ characterizes the distribution F (up to a location shift) (see [13]). This means that if we know the rank function, then we know the distribution (up to a location shift). These properties are better than those of the marginal rank function. Also, the computation of the sample spatial rank is more feasible than the Oja rank function.

We propose a new covariance matrix based on the spatial rank function. In order to simplify notations, from this point forward $\boldsymbol{s}(\boldsymbol{x})$ and $\boldsymbol{r}(\boldsymbol{x}, F)$ represent the spatial sign and spatial rank functions, respectively.

3 GINI COVARIANCE MATRICES

In this chapter two new covariance matrices are proposed. These matrices are based on the spatial rank function discussed in the previous chapter. The Gini Covariance Matrix (GCM) can be seen as a multivariate extension of the univariate Gini Mean Difference (GMD). In order to overcome the fact that the GCM is not "fully" affine equivariant, an affine equivariant version is derived through the transformation-retransformation technique. These concepts are introduced in the next few sections.

3.1 GINI COVARIANCE MATRIX

Let $\boldsymbol{x}$ be a d-variate random vector with distribution F. If the first moment of $\boldsymbol{x}$ exists, then the Gini Covariance Matrix of $\boldsymbol{x}$ (GCM) defined in Formula (1.2.12) may be written using different notations. When necessary, we can use the following notations:

$$\boldsymbol{\Sigma}_g := 2\mathbb{E}\boldsymbol{x}\boldsymbol{r}^T(\boldsymbol{x}), \tag{3.1.1}$$

$$\boldsymbol{\Sigma}_g := 2\mathbb{E}_{\boldsymbol{x}_1}\mathbb{E}_{\boldsymbol{x}_2}[\boldsymbol{s}^T(\boldsymbol{x}_1 - \boldsymbol{x}_2)|\boldsymbol{x}_1], \tag{3.1.2}$$

and

$$\boldsymbol{\Sigma}_g := 2\mathbb{E}\boldsymbol{x}_1\boldsymbol{s}^T(\boldsymbol{x}_1 - \boldsymbol{x}_2), \tag{3.1.3}$$

where $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$ are independent copies of $\boldsymbol{x}$. Recall that the function $\boldsymbol{r}(\boldsymbol{x})$ is the spatial rank function. The Gini covariance matrix is a direct generalization from (1.2.11). Moreover, the Gini Covariance Matrix can be written as

$$\boldsymbol{\Sigma}_g = 2\mathbb{E}\frac{\boldsymbol{x}_1(\boldsymbol{x}_1 - \boldsymbol{x}_2)^T}{\|\boldsymbol{x}_1 - \boldsymbol{x}_2\|} = \mathbb{E}\frac{(\boldsymbol{x}_1 - \boldsymbol{x}_2)(\boldsymbol{x}_1 - \boldsymbol{x}_2)^T}{\|\boldsymbol{x}_1 - \boldsymbol{x}_2\|}. \tag{3.1.4}$$

The second equality in Formula (3.1.4) follows from

$$\mathbb{E}\frac{\boldsymbol{x}_1(\boldsymbol{x}_1 - \boldsymbol{x}_2)^T}{\|\boldsymbol{x}_1 - \boldsymbol{x}_2\|} = -\mathbb{E}\frac{\boldsymbol{x}_2(\boldsymbol{x}_1 - \boldsymbol{x}_2)^T}{\|\boldsymbol{x}_1 - \boldsymbol{x}_2\|}.$$

From (3.1.4), $\boldsymbol{\Sigma}_g$ is semi-positive definite; it has the basic requirement of a covariance matrix. Equation (3.1.4) recovers L_1 metric representation of the Gini mean difference (1.2.6) when $d = 1$.

Before we explore properties of the Gini Covariance Matrix, it is worthwhile to present that another useful extension of GMD from the covariance representation is $2\mathbb{E}\boldsymbol{x}^T\boldsymbol{r}(\boldsymbol{x}) = \mathbb{E}\|\boldsymbol{x}_1 - \boldsymbol{x}_2\|$. It coincides with the multivariate Gini mean difference defined in Koshevoy & Mosler (see [14]).

3.2 PROPERTIES OF THE GINI COVARIANCE MATRIX

Consider elliptical distributions discussed in Chapter 1. The following theorem states the relationship of the Gini Covariance Matrix and the scatter matrix $\boldsymbol{\Sigma}$ in elliptical distributions.

Theorem 3.2.1. *If $\boldsymbol{x}$ has elliptical distribution F with first moment $\boldsymbol{\mu}$ and scatter parameter $\boldsymbol{\Sigma}$ having the spectral decomposition $V\Lambda V^T$,$\Lambda = diag(\lambda_1, ..., \lambda_d)$, then there exists a constant $c(F)$ and a vector $\boldsymbol{u} = (u_1, ...u_d)$ uniformly distributed on the unit sphere for which $\boldsymbol{\Sigma}_g = V\Lambda_g V^T$,*

$$\Lambda_g = diag(\lambda_{g,1}, ..., \lambda_{g,d}) = c(F)\mathbb{E}\left[\frac{\Lambda^{1/2}\boldsymbol{u}\boldsymbol{u}^T\Lambda^{1/2}}{\sqrt{\boldsymbol{u}^T\Lambda\boldsymbol{u}}}\right].$$

Note that from theorem 3.2.1 have

$$\lambda_{g,i} = c(F)\mathbb{E}\left[\frac{\lambda_i u_i^2}{\sqrt{\sum_{j=1}^d \lambda_j u_j^2}}\right], \tag{3.2.1}$$

where λ_i's are eigenvalues of $\mathbf{\Sigma}$.

The main consequence of Theorem 3.2.1 is that the same orthogonal matrix V diagonalizes $\mathbf{\Sigma}$ and $\mathbf{\Sigma}_g$. In other words, the Gini Covariance Matrix $\mathbf{\Sigma}_g$ has the same eigenvectors as $\mathbf{\Sigma}$. Therefore, the Gini covariance matrix can be used for principal component analysis.

Proof of Theorem 3.2.1. Let $\boldsymbol{z} = V^T(\boldsymbol{x} - \boldsymbol{\mu})$. Notice that $\boldsymbol{z} = r\Lambda^{1/2}\boldsymbol{u}$ where $r = \|\Lambda^{-1/2}\boldsymbol{z}\|$ and $\boldsymbol{u} = \Lambda^{-1/2}\boldsymbol{z}/r$. r and $\boldsymbol{u}$ are independent, and $\boldsymbol{u}$ is uniformly distributed on the unit sphere. The pairwise difference $\boldsymbol{z}_1 - \boldsymbol{z}_2 = V^T(\boldsymbol{x}_1 - \boldsymbol{x}_2)$ follows a centered distribution with diagonal scatter matrix 2Λ (see [1]). We can write $\boldsymbol{z}_1 - \boldsymbol{z}_2 = \sqrt{2}r\Lambda^{1/2}\boldsymbol{u}$ with $r = \|\frac{1}{\sqrt{2}}\Lambda^{-1/2}(\boldsymbol{z}_1 - \boldsymbol{z}_2)\|$, $\boldsymbol{u} = \frac{1}{\sqrt{2}}\Lambda^{-1/2}(\boldsymbol{z}_1 - \boldsymbol{z}_2)/r$, with μ and r independent. Then

$$\begin{aligned}
\mathbf{\Sigma}_g &= \mathbb{E}\left[\frac{(\boldsymbol{x}_1 - \boldsymbol{x}_2)(\boldsymbol{x}_1 - \boldsymbol{x}_2)^T}{\|\boldsymbol{x}_1 - \boldsymbol{x}_2\|}\right] = V\mathbb{E}\left[\frac{(\boldsymbol{z}_1 - \boldsymbol{z}_2)(\boldsymbol{z}_1 - \boldsymbol{z}_2)^T}{\|\boldsymbol{z}_1 - \boldsymbol{z}_2\|}\right]V^T \\
&= V\mathbb{E}\left[\frac{2r^2\Lambda^{1/2}\boldsymbol{u}\boldsymbol{u}^T\Lambda^{1/2}}{\sqrt{2r^2\boldsymbol{u}^T\Lambda^{1/2}\Lambda^{1/2}\boldsymbol{u}}}\right]V^T = \sqrt{2}\mathbb{E}rV\mathbb{E}\left[\frac{\Lambda^{1/2}\boldsymbol{u}\boldsymbol{u}^T\Lambda^{1/2}}{\sqrt{\boldsymbol{u}^T\Lambda\boldsymbol{u}}}\right]V^T.
\end{aligned}$$

The last equality is due to independence. Denote $\sqrt{2}\mathbb{E}r$ as $c(F)$, the proof is complete.

□

Remark 3.2.2. *In the case of an elliptical distribution F having $\mathbf{\Sigma} = \boldsymbol{I}_d$, i.e, $\lambda_1 = ... = \lambda_d = 1$, $\lambda_{g,i} = c(F)\mathbb{E}(u_i^2/\|\boldsymbol{u}\|) = c(F)\mathbb{E}u_i^2 = c(F)/d$ for all $i = 1, ..., d$ and hence $\mathbf{\Sigma}_g = \frac{c(F)}{d}\boldsymbol{I}_d$. In other words, for spherical distributions F_0, their Gini covariance matrix is the identity matrix multiplied by a constant. Dividing this factor can make GCM Fisher consistent at F_0.*

Remark 3.2.3. *For any elliptical distribution F, the constant $c(F) = c(F_0) = \mathbb{E}\|\boldsymbol{x}_1 - \boldsymbol{x}_2\| = \sqrt{2}\mathbb{E}_{F_0}\|\boldsymbol{x}_1\| = \sqrt{2}\mathbb{E}r$, where $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$ are independent random vectors from F_0, r has density in equation* (1.3.3)*, and $\mathbb{E}_{F_0}$ denotes taking expectation with respect to an r.v from F_0.*

Now, constants for various multivariate distributions are calculated.

Remark 3.2.4. *If F is a multivariate normal distribution $\mathcal{N}_d(\boldsymbol{\mu}, \boldsymbol{\Sigma})$,*

$$c(F) = \sqrt{2}\mathbb{E}_F\sqrt{(\boldsymbol{x}_1 - \boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}_1 - \boldsymbol{\mu})} = \sqrt{2}\mathbb{E}(D^{1/2}),$$

where D has a χ^2 distribution with the degree of freedom d. Hence

$$c(F) = \frac{2\Gamma[(d+1)/2]}{\Gamma(d/2)}.$$

For a univariate normal distribution $\mathcal{N}(\mu, \sigma^2)$, the Gini covariance is equal to the Gini mean difference $2\sigma/\sqrt{\pi}$.

Proof of Remark 3.2.4. Using the pdf for the Gaussian distribution $g(t) = (2\pi)^{-d/2}e^{-t/2}$ from (1.3.4) and (1.3.3), we have

$$f_r(r) = \frac{2\pi^{d/2}}{\Gamma(d/2)}r^{d-1}g(r^2) = \frac{2}{\Gamma(d/2)2^{d/2}}r^{d-1}e^{-r^2/2}.$$

Then,

$$\begin{aligned}
\mathbb{E}r &= \int_0^\infty \frac{2}{\Gamma(d/2)2^{d/2}} r^{d-1} e^{-r^2/2} r\, dr \\
&= \frac{2}{\Gamma(d/2)2^{d/2}} \int_0^\infty (r^2)^{(d-1)/2} e^{(-r^2)/2} r\, dr \\
&= \frac{1}{\Gamma(d/2)2^{d/2}} \int_0^\infty u^{(d-1)/2} e^{-u/2}\, du \\
&= \frac{1}{\Gamma(d/2)2^{d/2}} 2^{(d+1)/2} \Gamma[(d+1)/2] \\
&= \frac{\sqrt{2}\Gamma[(d+1)/2]}{\Gamma(d/2)}.
\end{aligned}$$

$c(F) = \sqrt{2}\mathbb{E}r$, and this concludes the proof.

□

Remark 3.2.5. *If F is a multivariate $\mathcal{T}_d(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \nu)$ distribution with $\nu > 1$,*

$$c(F) = \frac{\nu^{1/2}\Gamma[(\nu-1)/2]}{\sqrt{2}\Gamma(\nu/2)} \frac{2\Gamma[(d+1)/2]}{\Gamma(d/2)}.$$

Proof of Remark 3.2.5. Using (1.3.3) and the pdf for the $\mathcal{T}_d(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \nu)$ distribution (1.3.5)

$$g(t) = \frac{\Gamma[(\nu+d)/2]}{\Gamma(\nu/2)(\nu\pi)^{d/2}} (1 + t/\nu)^{-(d+\nu)/2}, \text{ we have}$$

$$f_r(r) = \frac{\Gamma[(\nu+d)/2]}{\Gamma(\nu/2)(\nu\pi)^{d/2}} \frac{2\pi^{d/2}}{\Gamma(d/2)} r^{d-1} (1 + r^2/\nu)^{-(d+\nu)/2}.$$

Then,

$$\begin{aligned}
\mathbb{E}r &= \frac{\Gamma[(\nu+d)/2]}{\Gamma(\nu/2)(\nu\pi)^{d/2}} \frac{2\pi^{d/2}}{\Gamma(d/2)} \int_0^\infty r^d (1 + r^2/\nu)^{-(d+\nu)/2}\, dr \\
&= \frac{2\Gamma[(\nu+d)/2]}{\Gamma(d/2)\Gamma(\nu/2)(\nu)^{d/2}} \frac{\nu^{(d+1)/2}\Gamma[(d+1)/2]\Gamma[(\nu-1)/2]}{2\Gamma[(d+\nu)/2]} \\
&= \frac{\nu^{1/2}\Gamma[(\nu-1)/2]}{\Gamma(\nu/2)} \frac{\Gamma[(d+1)/2]}{\Gamma(d/2)}
\end{aligned}$$

Formula (3.2.6) is obtained by a formula integration in Mathematica. Since $c(F) = \sqrt{2}\mathbb{E}\|\boldsymbol{x}\| = \sqrt{2}\mathbb{E}r$, we have

$$c(F) = \frac{\nu^{1/2}\Gamma[(\nu-1)/2]}{\sqrt{2}\Gamma(\nu/2)} \frac{2\Gamma[(d+1)/2]}{\Gamma(d/2)}.$$

We know that when $\nu \to \infty$, $\mathcal{T}_d(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \nu) \to N_d(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, and hence we expect that $c(F)$ in $\mathcal{T}_d(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \nu)$ approaches $c(F)$ from $N_d(\mu, \Sigma)$. Now, using Stirling formula $\Gamma(\nu) \approx \sqrt{2\pi} e^{-\nu} \nu^{\nu - 1/2}$ for large ν, we have as $\nu \to \infty$,

$$\begin{aligned}
\frac{\nu^{1/2}\Gamma[(\nu-1)/2]}{\sqrt{2}\Gamma(\nu/2)} &\approx \frac{\nu^{1/2}\sqrt{2\pi}\, e^{-(\nu-1)/2}[\nu-1)/2]^{(\nu-1)/2-1/2}}{\sqrt{2}\sqrt{2\pi}\, e^{-\nu/2}(\nu/2)^{\nu/2-1/2}} \\
&= e^{1/2}\left(\frac{\nu-1}{\nu}\right)^{\nu/2-1} \\
&\to 1 \quad \text{as } \nu \to \infty.
\end{aligned}$$

□

Remark 3.2.6. *If F is the Kotz distribution* (1.3.6)*, then $c(F) = \sqrt{2}d$. By Remark 3.2.2, for the spherical Kotz distribution ($\boldsymbol{\Sigma} = \boldsymbol{I}_d$), the Gini covariance matrix is $\sqrt{2}\boldsymbol{I}_d$. The Fisher consistency correction factor is $1/\sqrt{2}$, correcting from taking the pair difference, and it is free of d.*

Proof of Remark 3.2.6. The pdf of the Kotz distribution is

$$g(r) = \frac{\Gamma(d/2)}{2\pi^{d/2}\Gamma(d)} e^{-\sqrt{r}}.$$

Therefore, the density of $r = \|\boldsymbol{x}\|$ is

$$f_r(r) = \frac{2\pi^{d/2}}{\Gamma(d/2)} r^{d-1} g(r^2) = \frac{1}{\Gamma(d)} r^{d-1} e^{-r}.$$

Clearly, r has a gamma distribution with parameters $\alpha = d$ and $\beta = 1$. Hence $c(F) = \sqrt{2}\mathbb{E}\|\boldsymbol{x}\| = \sqrt{2}\mathbb{E}r = \sqrt{2}d$.

□

As any spatial procedure, spatial signs and spatial ranks are orthogonally equivariant in the sense that for any $d \times d$ orthogonal matrix O ($O^T = O^{-1}$), d-dimensional vector $\boldsymbol{b}$ and nonzero scalar c, letting $\boldsymbol{x}^* = cO\boldsymbol{x} + \boldsymbol{b}$ and $\boldsymbol{x}^*$ has distribution $F_{\boldsymbol{x}^*}$,

$$\boldsymbol{s}(\boldsymbol{x}^*) = \text{sign}(c)O\boldsymbol{s}(\boldsymbol{x}), \quad \text{and} \quad \boldsymbol{r}(\boldsymbol{x}^*, F_{\boldsymbol{x}^*}) = \text{sign}(c)O\boldsymbol{r}(\boldsymbol{x}, F_{\boldsymbol{x}}).$$

Therefore, we have the orthogonal equivariance property of GCM.

Theorem 3.2.2. *For any distribution F, the Gini covariance matrix $\boldsymbol{\Sigma}_g$ is orthogonally equivariant. That is, $\boldsymbol{\Sigma}_g(cO\boldsymbol{x} + \boldsymbol{b}) = |c|O\boldsymbol{\Sigma}_g(\boldsymbol{x})O^T$.*

Proof of Theorem 3.2.2. We have

$$\begin{aligned}
\boldsymbol{\Sigma}_g(\boldsymbol{x}^*) &= \mathbb{E}\boldsymbol{x}^*\boldsymbol{r}^T(\boldsymbol{x}^*, F_{\mathbf{x}^*}) \\
&= \mathbb{E}(cO\boldsymbol{x} + \boldsymbol{b})[\text{sign}(c)O\boldsymbol{r}^T(\boldsymbol{x}, F_{\boldsymbol{x}})] \\
&= c\text{sign}(c)O\mathbb{E}[\boldsymbol{x}\boldsymbol{r}^T(\boldsymbol{x}, F_{\boldsymbol{x}})]O^T \\
&= |c|O\boldsymbol{\Sigma}_g(\boldsymbol{x})O^T.
\end{aligned}$$

□

Orthogonal equivariance ensures that under rotation, translation and homogeneous scale change, the quantities are transformed accordingly. However, it does not allow heterogeneous scale changes. The above formulas do not hold for a general $d \times d$ nonsingular matrix A. Hence, they are not "fully" affine equivariant. Therefore, there is a need to develop an affine equivariant version of the Gini Covariance Matrix.

3.3 AFFINE EQUIVARIANT GINI COVARIANCE MATRIX

In order to achieve full affine equivariance, we use the transformation - retransformation (TR) technique, which serves as standardization of multivariate data. More details can be found in Serfling [26]. The well-known scatter functional population version of Tyler M functional from Tyler [30] is a TR version of spatial sign covariance matrix (2.3.3). In the same spirit, Dümbgen [5] considered symmetrized TR spatial sign covariance matrix of (2.3.4). Oja & Randles [21] constructed nonparametric tests based on TR spatial rank covariance matrix (2.3.3) . The affine equivariant counterpart of the Gini covariance matrix is denoted as $\boldsymbol{\Sigma}_G$. First, the original random vector $\boldsymbol{x}$ is transformed or standardized via $\boldsymbol{z} = \boldsymbol{\Sigma}_G^{-1/2}(\boldsymbol{x} - \boldsymbol{\mu})$. $\boldsymbol{z}$ follows the spherical distribution F_0 with scatter matrix $\boldsymbol{I}_{d\times d}$. Also, $\boldsymbol{z}_1 - \boldsymbol{z}_2 = \boldsymbol{\Sigma}_G^{-1/2}(\boldsymbol{x}_1 - \boldsymbol{x}_2)$. By Remark 3.2.2, we have

$$\mathbb{E}\frac{\boldsymbol{\Sigma}_G^{-1/2}(\boldsymbol{x}_1 - \boldsymbol{x}_2)(\boldsymbol{x}_1 - \boldsymbol{x}_2)^T\boldsymbol{\Sigma}_G^{-1/2}}{\sqrt{(\boldsymbol{x}_1 - \boldsymbol{x}_2)^T\boldsymbol{\Sigma}_G^{-1}(\boldsymbol{x}_1 - \boldsymbol{x}_2)}} = \mathbb{E}\frac{(\boldsymbol{z}_1 - \boldsymbol{z}_2)(\boldsymbol{z}_1 - \boldsymbol{z}_2)^T}{\|\boldsymbol{z}_1 - \boldsymbol{z}_2\|} = \frac{c(F)}{d}\boldsymbol{I}_d. \tag{3.3.1}$$

Hence, the TR version of the Gini covariance matrix is the solution of

$$\boldsymbol{\Sigma}_G = \frac{d}{c(F)}\mathbb{E}\frac{(\boldsymbol{x}_1 - \boldsymbol{x}_2)(\boldsymbol{x}_1 - \boldsymbol{x}_2)^T}{\sqrt{(\boldsymbol{x}_1 - \boldsymbol{x}_2)^T\boldsymbol{\Sigma}_G^{-1}(\boldsymbol{x}_1 - \boldsymbol{x}_2)}}. \tag{3.3.2}$$

We discuss the conditions needed for existence and uniqueness of the sample version given in the next chapter.

Theorem 3.3.1. *The matrix valued functional $\boldsymbol{\Sigma}_G(\cdot)$ is a scatter matrix in the sense that for any nonsingular $d \times d$ matrix A and d-vector $\boldsymbol{b}$, $\boldsymbol{\Sigma}_G(A\boldsymbol{x} + \boldsymbol{b}) = A\boldsymbol{\Sigma}_G(\boldsymbol{x})A^T$.*

Proof of Theorem 3.3.1. Let $\boldsymbol{x}$ be a random vector from F. Multiplying A on the left and A^T on the right on both sides of Equation (3.3.1), we have

$$A\boldsymbol{\Sigma}_G A^T = \frac{d}{c(F)}\mathbb{E}\frac{A(\boldsymbol{x}_1 - \boldsymbol{x}_2)(\boldsymbol{x}_1 - \boldsymbol{x}_2)^T A^T}{\sqrt{(\boldsymbol{x}_1 - \boldsymbol{x}_2)^T\boldsymbol{\Sigma}_G^{-1}(\boldsymbol{x}_1 - \boldsymbol{x}_2)}}.$$

Since A is nonsingular, A^{-1} and $(A^T)^{-1}$ exist. Hence,

$$A\boldsymbol{\Sigma}_G A^T = \frac{d}{c(F)}\mathbb{E}\frac{A(\boldsymbol{x}_1-\boldsymbol{x}_2)(\boldsymbol{x}_1-\boldsymbol{x}_2)^T A^T}{\sqrt{(\boldsymbol{x}_1-\boldsymbol{x}_2)^T A^T (A\boldsymbol{\Sigma}_G A^T)^{-1} A(\boldsymbol{x}_1-\boldsymbol{x}_2)}}.$$

Thus, $A\boldsymbol{\Sigma}_G A^T$ is the affine equivariant version of the Gini Covariance Matrix of $A\boldsymbol{x}+\boldsymbol{b}$.

□

Remark 3.3.3. *For $d=1$, the affine equivariant Gini mean difference is $\sigma_G^{1/2} = \mathbb{E}|x_1 - x_2|/c(F) = \sigma$. Hence, σ_G is Fisher consistent to the squared scale parameter for the local-scale family.*

Proof of Remark 3.3.3. For $d=1$, we have

$$\begin{aligned}
\sigma_G &= \frac{1}{c(F)}\mathbb{E}\frac{(x_1-x_2)^2}{\sqrt{(x_1-x_2)\sigma_G^{-1}(x_1-x_2)}} \\
\sigma_G^{1/2} &= \frac{\mathbb{E}|x_1-x_2|}{c(F)} \\
&= \sigma_g/c(F) = \sigma.
\end{aligned}$$

□

Sirkiä *et al.* (see [28]) studied a general symmetrized M-functional M that solves the function in the following defintion.

Definition 3.3.4. *Let $\boldsymbol{z}_{12}(\boldsymbol{M}) = \boldsymbol{M}^{-1/2}(\boldsymbol{x}_1-\boldsymbol{x}_2)$, $r_{12}(\boldsymbol{M}) = \|\boldsymbol{z}_{12}(\boldsymbol{M})\|$, and $\boldsymbol{s}_{12}(\boldsymbol{M}) = r_{12}(\boldsymbol{M})^{-1}\boldsymbol{z}_{12}(\boldsymbol{M})$, where $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$ are independent random vectors from F, then a general symmetrized M-functional M solves*

$$\mathbb{E}[w_1(r_{12}(M))\boldsymbol{s}_{12}(M)\boldsymbol{s}_{12}^T(M) - w_2(r_{12}(M))\boldsymbol{I}] = \boldsymbol{0}, \tag{3.3.5}$$

where w_1 and w_2 are real-valued functions on $[0,\infty]$.

The affine equivariant version of the Gini covariance matrix $\boldsymbol{\Sigma}_G$ is a special case of the symmetrized M-functional defined as the solution of

$$\mathbb{E}[r_{12}(\boldsymbol{\Sigma}_G)\boldsymbol{s}_{12}(\boldsymbol{\Sigma}_G)\boldsymbol{s}_{12}^T(\boldsymbol{\Sigma}_G) - \frac{c(F)}{d}\boldsymbol{I}] = \mathbf{0}. \tag{3.3.6}$$

We see that the weight functions for $\boldsymbol{\Sigma}_G$ are $w_1(t) = t$ and $w_2(t) = c(F)/d$.

Note that the weight functions $w_1(t)$ and $w_2(t)$ of (3.3.5) are related to $u(t)$ and $v(t)$ of Formula (1.4.4) by

$$w_1(t) = t^2 u(t), \quad \text{and} \quad w_2(t) = v(t). \tag{3.3.7}$$

The weight functions $w_1(t) = t$ and $w_2(t) = 1$ define the Kotz function $\boldsymbol{\Sigma}_K$. Note that for $F \in \mathcal{E}(\boldsymbol{\mu}, \boldsymbol{\Sigma}, g)$, $\boldsymbol{\Sigma}_K(F) = \{\mathbb{E}r/d\}^2\boldsymbol{\Sigma}$ where $r = \|\boldsymbol{\Sigma}^{-1/2}(\boldsymbol{x} - \boldsymbol{\mu})\|$. In other words, $\boldsymbol{\Sigma}_K(F) = \{c(F)/\sqrt{2}d\}^2\boldsymbol{\Sigma}$. For the Kotz distribution F, $\boldsymbol{\Sigma}_K(F) = \boldsymbol{\Sigma}$. It is worthwhile to notice that our TR Gini covariance matrix can be viewed as the symmetrized Kotz functional, but for our affine equivariant GCM, $\boldsymbol{\Sigma}_G = \boldsymbol{\Sigma}$ for all elliptical distributions F.

3.4 MORE PROPERTIES OF THE TWO GINI COVARIANCE MATRICES

In this section, we study more properties of the two Gini Covariance Matrices. We explore the independent property of the two Gini covariance matrices, then their influence functions are derived. We also derive the influence function of the Kotz functional.

3.4 Independence Property

Corollary 3.4.1 (Sirkiä *et al.* [28]). *A symmetrised scatter matrix functional $M_s(\cdot)$ has the independence property, that is, when F is the cdf of a random vector with independent components, $M_s(\cdot)$ is diagonal.*

The independence property is highly important, for example, in independent component analysis (see [22]). This property holds for the regular covariance matrix, but it may not hold for general M-functionals. Sirkiä *et al.* [28] concluded in Corollary 1 (3.4.1) that any symmetrized affine equivariant scatter functional has the independence property. In this dissertation, we reduce this to symmetrized orthogonally equivariant scatter matrices, and the result for $\mathbf{\Sigma}_g$ follows. The result for $\mathbf{\Sigma}_G$ follows directly from the result of Sirkiä *et al.* (see [28]), as $\mathbf{\Sigma}_G$ is symmetrized and affine equivariant.

Theorem 3.4.2. *Let F be the cdf of a random vector with independent components. Then $\mathbf{\Sigma}_g(F)$ and $\mathbf{\Sigma}_G(F)$ are diagonal.*

Proof of Theorem 3.4.2. Let M be an orthogonally equivariant covariance matrix. Let $\boldsymbol{x}$ be a random $d-$vector with independent and symmetric components. The vector of centers of symmetry is denoted as $\boldsymbol{\mu}$. Let I_i^- be the $d \times d$ diagonal matrix with the i^{th} diagonal element being -1 and all other diagonal elements being 1. Then $\boldsymbol{x} - \boldsymbol{\mu}$ and $I_i^-(\boldsymbol{x} - \boldsymbol{\mu})$ have the same distribution. Also because $M(\cdot)$ is orthogonally equivariant, we have

$$M(F_{\boldsymbol{x}}) = M(F_{\boldsymbol{x}-\boldsymbol{\mu}}) = M(F_{I_i^-(\boldsymbol{x}-\boldsymbol{\mu})}) = I_i^- M(F_{\boldsymbol{x}}) I_i^-$$

for all $i = 1, 2, ..., d$. This implies that all off-diagonal elements of $M(F_{\boldsymbol{x}})$ are equal to 0.

Now Theorem 3.4.2 is implied by the fact that $\boldsymbol{x}_1 - \boldsymbol{x}_2$ always has symmetric components and both $\mathbf{\Sigma}_g(\cdot)$ and $\mathbf{\Sigma}_G(\cdot)$ are orthogonally equivariant. □

In the next section, we study robustness properties of the two Gini covariance matrices along with the Kotz functional through the influence function approach.

3.4 Influence Function

The influence function (IF) introduced by Hampel (see [8]) is a standard heuristic tool for measuring the effect of infinitesimal perturbations on a functional T. The influence function for the Gini Covariance Matrix by means of Formula (1.6.5) can be obtained via the following Lemma.

Lemma 3.4.3. *The influence function of the Gini Covariance Matrix* $\boldsymbol{\Sigma}_g$ *is defind at* $\boldsymbol{x} \in \mathbb{R}^d$ *by*

$$IF(\boldsymbol{x}; \boldsymbol{\Sigma}_g, F) = 2\mathbb{E}\frac{(\boldsymbol{x}_1 - \boldsymbol{x})(\boldsymbol{x}_1 - \boldsymbol{x})^T}{\|\boldsymbol{x}_1 - \boldsymbol{x}\|} - 2\boldsymbol{\Sigma}_g,$$

where $\boldsymbol{x}_1$ *is from the continuous distribution* F*.*

Proof of Lemma 3.4.3. The proof is straightforward. Recall $\boldsymbol{\Sigma}_g(F) = 2\mathbb{E}\boldsymbol{x}_1\mathbb{E}\left[\boldsymbol{s}^T(\boldsymbol{x}_1 - \boldsymbol{x}_2)|\boldsymbol{x}_1\right]$. Let $F_\varepsilon = (1-\varepsilon)F + \varepsilon\delta_{\boldsymbol{x}}$ and $\boldsymbol{x}_1$, $\boldsymbol{x}_2$ have distribution F_ε. Then, we have the following:

$$\begin{aligned}
\boldsymbol{\Sigma}_g(F_\varepsilon) &= 2\left\{\mathbb{E}_{F_\varepsilon}\mathbb{E}_{F_\varepsilon}\left[\boldsymbol{x}_1\boldsymbol{s}^T(\boldsymbol{x}_1 - \boldsymbol{x}_2)|\boldsymbol{x}_1\right]\right\} \\
&= 2\mathbb{E}_{F_\varepsilon}\left[(1-\varepsilon)\mathbb{E}_F\left[\boldsymbol{x}_1\boldsymbol{s}^T(\boldsymbol{x}_1 - \boldsymbol{x}_2)|\boldsymbol{x}_1\right] + \varepsilon\left[\boldsymbol{x}_1\boldsymbol{s}^T(\boldsymbol{x}_1 - \boldsymbol{x})|\boldsymbol{x}_1\right]\right] \\
&= 2\{(1-\varepsilon)^2\mathbb{E}_F\mathbb{E}_F\left[\boldsymbol{x}_1\boldsymbol{s}^T(\boldsymbol{x}_1 - \boldsymbol{x}_2)|\boldsymbol{x}_1\right] + \varepsilon(1-\varepsilon)\mathbb{E}_F\left[\boldsymbol{x}\boldsymbol{s}^T(\boldsymbol{x} - \boldsymbol{x}_2)\right] \\
&+ \varepsilon(1-\varepsilon)\mathbb{E}_F\left[\boldsymbol{x}_1\boldsymbol{s}^T(\boldsymbol{x}_1 - \boldsymbol{x})\right] + \varepsilon^2\left[\boldsymbol{x}\boldsymbol{s}^T(\boldsymbol{x} - \boldsymbol{x})\right]\} \\
&= 2\{(1-\varepsilon)^2\mathbb{E}_F\mathbb{E}_F\left[\boldsymbol{x}_1\boldsymbol{s}^T(\boldsymbol{x}_1 - \boldsymbol{x}_2)|\boldsymbol{x}_1\right] + \varepsilon(1-\varepsilon)\mathbb{E}_F\left[-\boldsymbol{x}\boldsymbol{s}^T(\boldsymbol{x}_2 - \boldsymbol{x})\right] \\
&+ \varepsilon(1-\varepsilon)\mathbb{E}_F\left[\boldsymbol{x}_1\boldsymbol{s}^T(\boldsymbol{x}_1 - \boldsymbol{x})\right] + \varepsilon^2\left[\boldsymbol{x}\boldsymbol{s}^T(\mathbf{0})\right]\} \\
&= 2\{(1-\varepsilon)^2\mathbb{E}_F\mathbb{E}_F\left[\boldsymbol{x}_1\boldsymbol{s}^T(\boldsymbol{x}_1 - \boldsymbol{x}_2)|\boldsymbol{x}_1\right] + \varepsilon(1-\varepsilon)\mathbb{E}_F\left[-\boldsymbol{x}\boldsymbol{s}^T(\boldsymbol{x}_1 - \boldsymbol{x})\right] \\
&+ \varepsilon(1-\varepsilon)\mathbb{E}_F\left[\boldsymbol{x}_1\boldsymbol{s}^T(\boldsymbol{x}_1 - \boldsymbol{x})\right]\} \text{ since } \boldsymbol{s}(\mathbf{0}) = \mathbf{0} \\
&= 2\{(1-\varepsilon)^2\mathbb{E}_F\mathbb{E}_F\left[\boldsymbol{x}_1\boldsymbol{s}^T(\boldsymbol{x}_1 - \boldsymbol{x}_2)|\boldsymbol{x}_1\right] + \varepsilon(1-\varepsilon)\mathbb{E}_F\left[(\boldsymbol{x}_1 - \boldsymbol{x})\boldsymbol{s}^T(\boldsymbol{x}_1 - \boldsymbol{x})\right]\}.
\end{aligned}$$

Therefore, we get the IF as follows:

$$\begin{aligned}
IF(\boldsymbol{x};\boldsymbol{\Sigma}_g,F) &= \lim_{\varepsilon\to 0}\frac{1}{\varepsilon}\left[\boldsymbol{\Sigma}(F_\varepsilon)-\boldsymbol{\Sigma}(F)\right] \\
&= 2\lim_{\varepsilon\to 0}\frac{1}{\varepsilon}\{(1-2\varepsilon+\varepsilon^2-1)\mathbb{E}_F\mathbb{E}_F\left[\boldsymbol{x}_1\boldsymbol{s}^T(\boldsymbol{x}_1-\boldsymbol{x}_2)|\boldsymbol{x}_1\right] \\
&+ \varepsilon(1-\varepsilon)\mathbb{E}_F\left[(\boldsymbol{x}_1-\boldsymbol{x})\boldsymbol{s}^T(\boldsymbol{x}_1-\boldsymbol{x})\right]\} \\
&= 2\lim_{\varepsilon\to 0}\frac{1}{\varepsilon}\{\varepsilon(-2+\varepsilon)\mathbb{E}_F\mathbb{E}_F\left[\boldsymbol{x}_1\boldsymbol{s}^T(\boldsymbol{x}_1-\boldsymbol{x}_2)|\boldsymbol{x}_1\right] \\
&+ \varepsilon(1-\varepsilon)\mathbb{E}_F\left[(\boldsymbol{x}_1-\boldsymbol{x})\boldsymbol{s}^T(\boldsymbol{x}_1-\boldsymbol{x})\right]\} \\
&= 2\lim_{\varepsilon\to 0}\{(-2+\varepsilon)\mathbb{E}_F\mathbb{E}_F\left[\boldsymbol{x}_1\boldsymbol{s}^T(\boldsymbol{x}_1-\boldsymbol{x}_2)|\boldsymbol{x}_1\right] \\
&+ (1-\varepsilon)\mathbb{E}_F\left[(\boldsymbol{x}_1-\boldsymbol{x})\boldsymbol{s}^T(\boldsymbol{x}_1-\boldsymbol{x})\right]\} \\
&= 2\{-2\mathbb{E}_F\mathbb{E}_F\left[\boldsymbol{x}_1\boldsymbol{s}^T(\boldsymbol{x}_1-\boldsymbol{x}_2)|\boldsymbol{x}_1\right]+\mathbb{E}_F\left[(\boldsymbol{x}_1-\boldsymbol{x})\boldsymbol{s}^T(\boldsymbol{x}_1-\boldsymbol{x})\right]\} \\
&= 2\mathbb{E}_F\left[(\boldsymbol{x}_1-\boldsymbol{x})\boldsymbol{s}^T(\boldsymbol{x}_1-\boldsymbol{x})\right]-2\boldsymbol{\Sigma}_g.
\end{aligned}$$

This concludes the proof of the lemma.

□

Remark 3.4.1. *For $d = 1$, we obtain the influence function for the Gini mean difference, that is, $IF(x;\sigma_g,F) = 2\mathbb{E}_{x_1}|x_1 - x| - 2\sigma_g$, which is approximately linear in x in contrast to a quadratic form in $IF(x;\sigma^2,F) = \mathbb{E}_{x_1}(x_1 - x)^2 - \sigma^2$.*

The influence function of the affine equivariant GCM is more complicated than that of GCM. Hampel *et al.* [8] showed that, for any scatter functional $M(\cdot)$, the influence function of M for a spherical distribution $F_0(g)$, symmetric at the origin with $\boldsymbol{\Sigma}(F_0) = \boldsymbol{I}_d$, is given by

$$IF(\boldsymbol{x};M,F_0) = \alpha_M(\|\boldsymbol{x}\|)\frac{\boldsymbol{x}\boldsymbol{x}^T}{\|\boldsymbol{x}\|^2} - \beta_M(\|\boldsymbol{x}\|)\boldsymbol{I}_d, \tag{3.4.2}$$

where α_M and β_M are two real valued functions depending on F_0. Then the influence function of M at an elliptical distribution $F \in \mathcal{E}(\boldsymbol{\mu}, \boldsymbol{\Sigma}, g)$ is

$$IF(\boldsymbol{x}; M, F) = \boldsymbol{\Sigma}^{1/2} IF(\boldsymbol{\Sigma}^{-1/2}(\boldsymbol{x} - \boldsymbol{\mu}); M, F_0)\boldsymbol{\Sigma}^{1/2}.$$

The following theorem gives the influence function of TR GCM, which can be obtained as a special case of Theorem 2 in Sirkiä *et al.* [28] with $w_1(r) = r$ and $w_2(r) = c(F)/d$.

Theorem 3.4.4. *The influence function of the affine equivariant version of Gini covariance matrix $\boldsymbol{\Sigma}_G$ for a spherical distribution F_0 is of the form* (3.4.2) *with*

$$\begin{aligned}
\alpha_{\Sigma_G}(\|\boldsymbol{x}\|) &= \frac{2d(d+2)}{(d+1)c(F_0)} \mathbb{E}_{\boldsymbol{x}_1|\boldsymbol{x}_2}\left[(\|\boldsymbol{x}_1 - \|\boldsymbol{x}\|\boldsymbol{e}_1\|) - \frac{d(\boldsymbol{x}_1)_2^2}{\|\boldsymbol{x}_1 - \|\boldsymbol{x}\|\boldsymbol{e}_1\|}\right], \\
\beta_{\Sigma_G}(\|\boldsymbol{x}\|) &= 4 - \frac{2d}{(d+1)c(F_0)} \mathbb{E}_{\boldsymbol{x}_1|\boldsymbol{x}_2}\left[(\|\boldsymbol{x}_1 - \|\boldsymbol{x}\|\boldsymbol{e}_1\|) + \frac{(d+2)(\boldsymbol{x}_1)_2^2}{\|\boldsymbol{x}_1 - \|\boldsymbol{x}\|\boldsymbol{e}_1\|}\right],
\end{aligned}$$

where $\boldsymbol{x}_1$ is an r.v. from F and $(\boldsymbol{x}_1)_2$ denotes the second coordinate of $\boldsymbol{x}_1$, $\boldsymbol{e}_1 = (1, 0, ..., 0)^T$, and $c(F_0) = \mathbb{E}_{F_0}\|\boldsymbol{x}_1 - \boldsymbol{x}_2\|$ as Remark 3.2.3.

Proof of Theorem 3.4.4. From Theorem 2 of Sirkiä *et al.* [28] we have that a symmetrised M-functional $\boldsymbol{M}(\cdot)$ will have an influence function (at spherical F_0) of the form

$$\begin{aligned}
\alpha_{\boldsymbol{M}}(\|\boldsymbol{x}\|) &= \frac{1}{\eta_1} \mathbb{E}_{\boldsymbol{x}_1|\boldsymbol{x}_2}\left[w_1((\|\boldsymbol{x}_1 - \|\boldsymbol{x}\|\boldsymbol{e}_1\|)) - \frac{d(\boldsymbol{x}_1)_2^2}{\|\boldsymbol{x}_1 - \|\boldsymbol{x}\|\boldsymbol{e}_1\|}\right], \\
\beta_{\boldsymbol{M}}(\|\boldsymbol{x}\|) &= \frac{1}{d}\alpha_{\boldsymbol{M}}(\|\boldsymbol{x}\|) + \frac{1}{\eta_2} \mathbb{E}_{\boldsymbol{x}_1|\boldsymbol{x}_2}\left[w_2(\|\boldsymbol{x}_1 - \|\boldsymbol{x}\|\boldsymbol{e}_1\|) - \frac{1}{d} w_2(\|\boldsymbol{x}_1 - \|\boldsymbol{x}\|\boldsymbol{e}_1\|)\right],
\end{aligned}$$

if $\eta_2 \neq 0$, where

$$\begin{aligned}
\eta_1 &= \frac{\mathbb{E}[w_1'(\|\boldsymbol{x}_1 - \boldsymbol{x}_2\|)\|\boldsymbol{x}_1 - \boldsymbol{x}_2\| + d w_1(\|\boldsymbol{x}_1 - \boldsymbol{x}_2\|)]}{2d(d+2)}, \\
\eta_2 &= \frac{\mathbb{E}[w_1'(\|\boldsymbol{x}_1 - \boldsymbol{x}_2\|)\|\boldsymbol{x}_1 - \boldsymbol{x}_2\| - d^2 w_2'(\|\boldsymbol{x}_1 - \boldsymbol{x}_2\|)\|\boldsymbol{x}_1 - \boldsymbol{x}_2\|]}{4d}.
\end{aligned}$$

The affine equivariant version of Gini covariance matrix $\boldsymbol{\Sigma}_G$ has functions of the following form: $w_1(t) = t$ with $w_1^{'}(t) = 1$ and $w_2(t) = c(F)/d = c(F_0)/d$ with $w_2^{'}(t) = 0$. Therefore, using the formulas given above we get

$$\begin{aligned}
\eta_1 &= \frac{1}{2d(d+2)}\mathbb{E}\left[w_1^{'}(\|\boldsymbol{x}_1-\boldsymbol{x}_2\|)\|\boldsymbol{x}_1-\boldsymbol{x}_2\| + dw_1(\|\boldsymbol{x}_1-\boldsymbol{x}_2\|)\right] \\
&= \frac{(d+1)\mathbb{E}[\|\boldsymbol{x}_1-\boldsymbol{x}_2\|]}{2d(d+2)} = \frac{(d+1)c(F_0)}{2d(d+2)}, \\
\eta_2 &= \frac{1}{4d}\mathbb{E}\left[w_1^{'}(\|\boldsymbol{x}_1-\boldsymbol{x}_2\|)\|\boldsymbol{x}_1-\boldsymbol{x}_2\| - d^2 w_2^{'}(\|\boldsymbol{x}_1-\boldsymbol{x}_2\|)\|\boldsymbol{x}_1-\boldsymbol{x}_2\|\right] \\
&= \frac{\mathbb{E}\|\boldsymbol{x}_1-\boldsymbol{x}_2\|}{4d} = \frac{c(F_0)}{4d}.
\end{aligned}$$

Thus,

$$\begin{aligned}
\alpha_{\Sigma_G}(\|\boldsymbol{x}\|) &= \frac{2d(d+2)}{(d+1)c(F_0)}\mathbb{E}_{\boldsymbol{x}_1|\boldsymbol{x}_2}\left[(\|\boldsymbol{x}_1 - \|\boldsymbol{x}\|\boldsymbol{e}_1\|) - \frac{d(\boldsymbol{x}_1)_2^2}{\|\boldsymbol{x}_1 - \|\boldsymbol{x}\|\boldsymbol{e}_1\|}\right], \\
\beta_{\Sigma_G}(\|\boldsymbol{x}\|) &= \frac{2(d+2)}{(d+1)c(F_0)}\mathbb{E}_{\boldsymbol{x}_1|\boldsymbol{x}_2}\left[(\|\boldsymbol{x}_1 - \|\boldsymbol{x}\|\boldsymbol{e}_1\|) - \frac{d(\boldsymbol{x}_1)_2^2}{\|\boldsymbol{x}_1 - \|\boldsymbol{x}\|\boldsymbol{e}_1\|}\right] \\
&+ \frac{4d}{c(F_0)}\mathbb{E}_{\boldsymbol{x}_1|\boldsymbol{x}_2}\left[\frac{c(F_0)}{d} - \|\boldsymbol{x}_1 - \|\boldsymbol{x}\|\boldsymbol{e}_1\|\right] \\
&= 4 - \frac{2d}{(d+1)c(F_0)}\mathbb{E}_{\boldsymbol{x}_1|\boldsymbol{x}_2}\left[(\|\boldsymbol{x}_1 - \|\boldsymbol{x}\|\boldsymbol{e}_1\|) + \frac{(d+2)(\boldsymbol{x}_1)_2^2}{\|\boldsymbol{x}_1 - \|\boldsymbol{x}\|\boldsymbol{e}_1\|}\right].
\end{aligned}$$

□

Remark 3.4.3. *For $d = 1$, the influence function for σ_G is*

$$IF(x;\sigma_G;F_0) = \alpha_{\sigma_G}(|x|) - \beta_{\sigma_G}(|x|) = \frac{4}{c(F_0)}\mathbb{E}_{x_1}|x_1 - |x|| - 4,$$

where $c(F_0) = \mathbb{E}_{F_0}|x_1 - x_2| = \sigma_g(F_0)$. $IF(x;\sigma_G;F_0)$ is approximately linear in $|x|$.

Theorem 3.4.5. *The influence function of Kotz functional $\boldsymbol{\Sigma}_K$ at a spherical distribution F_0 is of the form* (3.4.2) *with*

$$\alpha_{\Sigma_K}(\|\boldsymbol{x}\|) = \frac{\sqrt{2}}{c(F_0)}\frac{d(d+2)}{(d+1)}\|\boldsymbol{x}\|,$$
$$\beta_{\Sigma_K}(\|\boldsymbol{x}\|) = \frac{\sqrt{2}d}{c(F_0)}\left[2 - \frac{\|\boldsymbol{x}\|}{d+1}\right].$$

Proof of Theorem 3.4.5. We have the influence function of M-functional M in the form $IF(\boldsymbol{x}; M, F_0) = -2\dot{W}$ where $W = M^{-1/2}$, $\dot{W} = IF(\boldsymbol{x}; W, F_0)$, and

$$\begin{aligned} \frac{1}{d}tr(W) &= -\frac{\frac{1}{d}w_1(\|\boldsymbol{x}\|) - w_2(\|\boldsymbol{x}\|)}{\mathbb{E}_y[(\frac{1}{d}w_1'(\|\boldsymbol{y}\|) - w_2'(\|\boldsymbol{y}\|))\|\boldsymbol{y}\|]}, \\ \dot{W} - \frac{1}{d}tr(W)\boldsymbol{I}_d &= -\frac{d+2}{2}\frac{w_1(\|\boldsymbol{x}\|)(\frac{\boldsymbol{x}\boldsymbol{x}^T}{\|\boldsymbol{x}\|^2} - \frac{1}{d}\boldsymbol{I}_d)}{\mathbb{E}_y[w_1(\|\boldsymbol{y}\|) + \frac{1}{d}w_1'(\|\boldsymbol{y}\|)\|\boldsymbol{y}\|]}, \end{aligned}$$

where $\boldsymbol{y}$ is a random vector from the distribution F_0 (see pages 220-222 of [10]).

With $w_1(t) = t$ and $w_2(t) = 1$ along with $w_1'(t) = 1$ and $w_2'(t) = 0$ for $\boldsymbol{\Sigma}_K$, solving for $\dot{W}$ in the above equations we get

$$\dot{W} = \frac{-d(d+2)}{2(d+1)\mathbb{E}\|\boldsymbol{y}\|}\|\boldsymbol{x}\|\frac{\boldsymbol{x}\boldsymbol{x}^T}{\|\boldsymbol{x}\|^2} + \frac{d+2}{2(d+1)\mathbb{E}\|\boldsymbol{y}\|}\|\boldsymbol{x}\|\boldsymbol{I}_d - \frac{\|\boldsymbol{x}\| - d}{\mathbb{E}\|\boldsymbol{y}\|}\boldsymbol{I}_d.$$

From Remark (3.2.2) $c(F_0) = \sqrt{2}\mathbb{E}\|\boldsymbol{y}\|$. Therefore, we obtain

$$IF(\boldsymbol{x}; \boldsymbol{\Sigma}_K, F_0) = -2\dot{W} = \frac{\sqrt{2}d(d+2)}{(d+1)c(F_0)}\|\boldsymbol{x}\|\frac{\boldsymbol{x}\boldsymbol{x}^T}{\|\boldsymbol{x}\|^2} + \frac{\sqrt{2}d}{(d+1)c(F_0)}\|\boldsymbol{x}\|I_d - \frac{2\sqrt{2}d}{c(F_0)}\boldsymbol{I}_d.$$

The conclusion follows. □

3.4 Plots of Some Examples

α_M and β_M functions from the influence function of various scatter functionals are plotted under the bivariate standard normal distribution. For TR Gini, numerical integration in R-package "cubature" is used for computing the expectations in the functions. For each $\|\boldsymbol{x}\| = r$, which is a sequence from 0 to 10 in increments of 0.1, we need to compute $\alpha_{\Sigma_G}(r)$ and $\beta_{\Sigma_G}(r)$. Here, $\alpha_{\Sigma_G}(r)$ is calculated as follows:

$$\begin{aligned}\alpha_{\Sigma_G}(r) &= \frac{8}{3c(F_0)}\mathbb{E}_{\boldsymbol{x}_1}\left[\|\boldsymbol{x}_1 - r\boldsymbol{e}_1\| - \frac{2x_{12}^2}{\|\boldsymbol{x}_1 - r\boldsymbol{e}_1\|}\right] \\ &= \frac{2\sqrt{2}}{3\Gamma(3/2)\pi}\int\int\left(\sqrt{(x_{11}-r)^2 + x_{12}^2} - \frac{2x_{12}^2}{\sqrt{(x_{11}-r)^2 + x_{12}^2}}\right)e^{-x_{11}^2/2}e^{-x_{12}^2/2}\,dx_{11}\,dx_{12}.\end{aligned}$$

$\beta_{\Sigma_G}(r)$ is calculated as follows:

$$\begin{aligned}\beta_{\Sigma_G}(r) &= 4 - \frac{4}{3c(F_0)}\mathbb{E}_{\boldsymbol{x}_1}\left[\|\boldsymbol{x}_1 - r\boldsymbol{e}_1\| + \frac{4x_{12}^2}{\|\boldsymbol{x}_1 - r\boldsymbol{e}_1\|}\right] \\ &= 4 - \frac{2\sqrt{2}}{3\pi^{3/2}}\int\int\left(\sqrt{(x_{11}-r)^2 + x_{12}^2} + \frac{4x_{12}^2}{\sqrt{(x_{11}-r)^2 + x_{12}^2}}\right)e^{-(x_{11}^2+x_{12}^2)/2}\,dx_{11}\,dx_{12}.\end{aligned}$$

The integration limits for x_{11} and x_{12} are set to be $[-10, 10]$. One must be careful dealing with the term $\sqrt{(x_{11}^2 - r)^2 + x_{12}^2}$ in the denominator. To avoid division by zero, we take the integration range of x_{11} to be $[-10, r-10^{-5}]\cup[r+10^{-5}, 10]$ when $x_{12} \in [-10^{-5}, 10^{-5}]$. In such way, the denominator is always greater than 10^{-5}.

Figures 3.4.1 and 3.4.2 display function $\alpha_M(r)$ and $\beta_M(r)$, respectively, for covariance matrix, Tyler M functional, Dübgen functional, Kotz functional and TR Gini covariance matrix under the bivariate standard normal distribution. From Formula (3.4.2), the function α is the influence of $\boldsymbol{x}$ on an off-diagonal element of M ($IF(\boldsymbol{x}; M_{ij}, F_0) = \alpha_M(\|\boldsymbol{x}\|)u_i u_j$, where u_i and u_j are the i^{th} and j^{th} component of $\boldsymbol{u} = \boldsymbol{x}/\|\boldsymbol{x}\|$). This implies that for boundedness of the influence at off-diagonal elements, the necessary and sufficient condition is that the α is bounded (see [8]). As we can see from Figures 3.4.1 and 3.4.2, the α functions of Tyler and

Dübgen M functionals are bounded. The α function of the covariance matrix is quadratic in the radius r, while that of the TR Gini covariance matrix is approximately linear for large r and that of Kotz functional is linear. This suggests that the TR Gini covariance matrix and Kotz functional will provide more protection against moderate outliers than the covariance matrix, but they are not robust in the strict sense. The Kotz functional and its symmetrized version TR Gini covariance matrix are an L_1 method. They are more robust than an L_2 method approach, very efficient as we will see in the next chapter, but not highly robust.

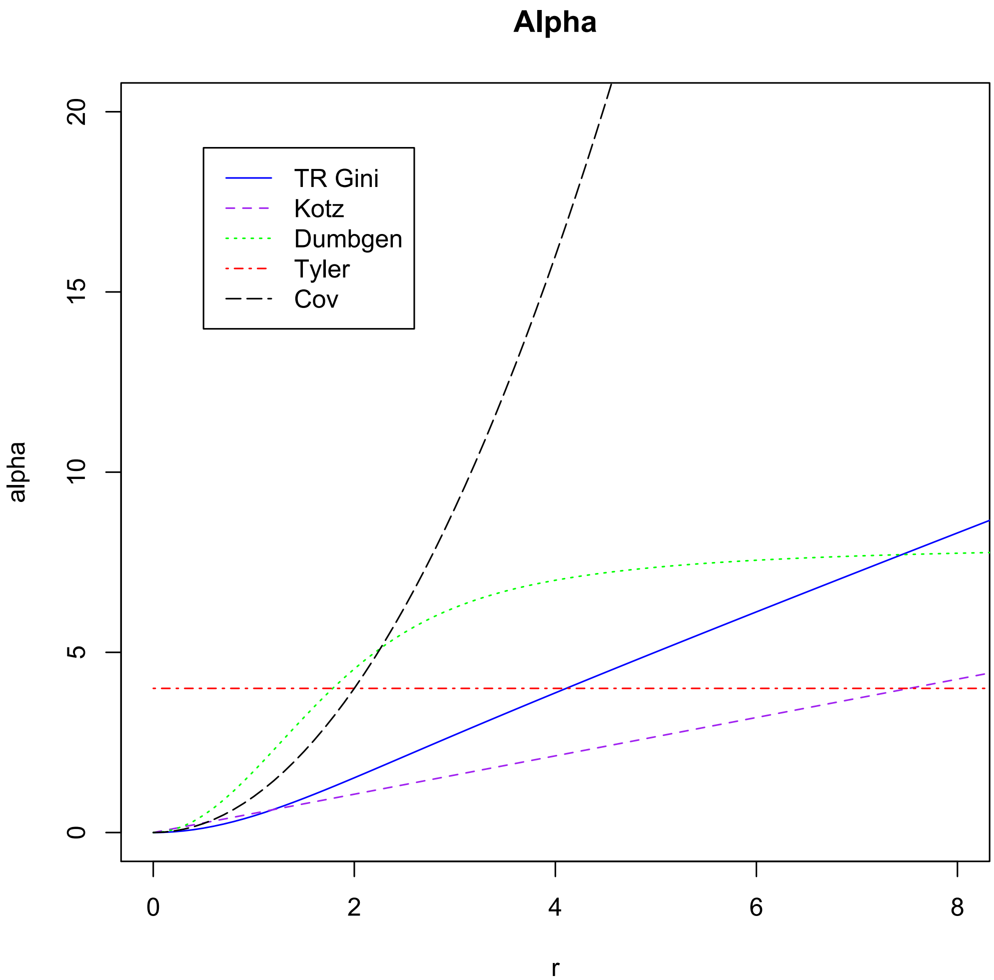
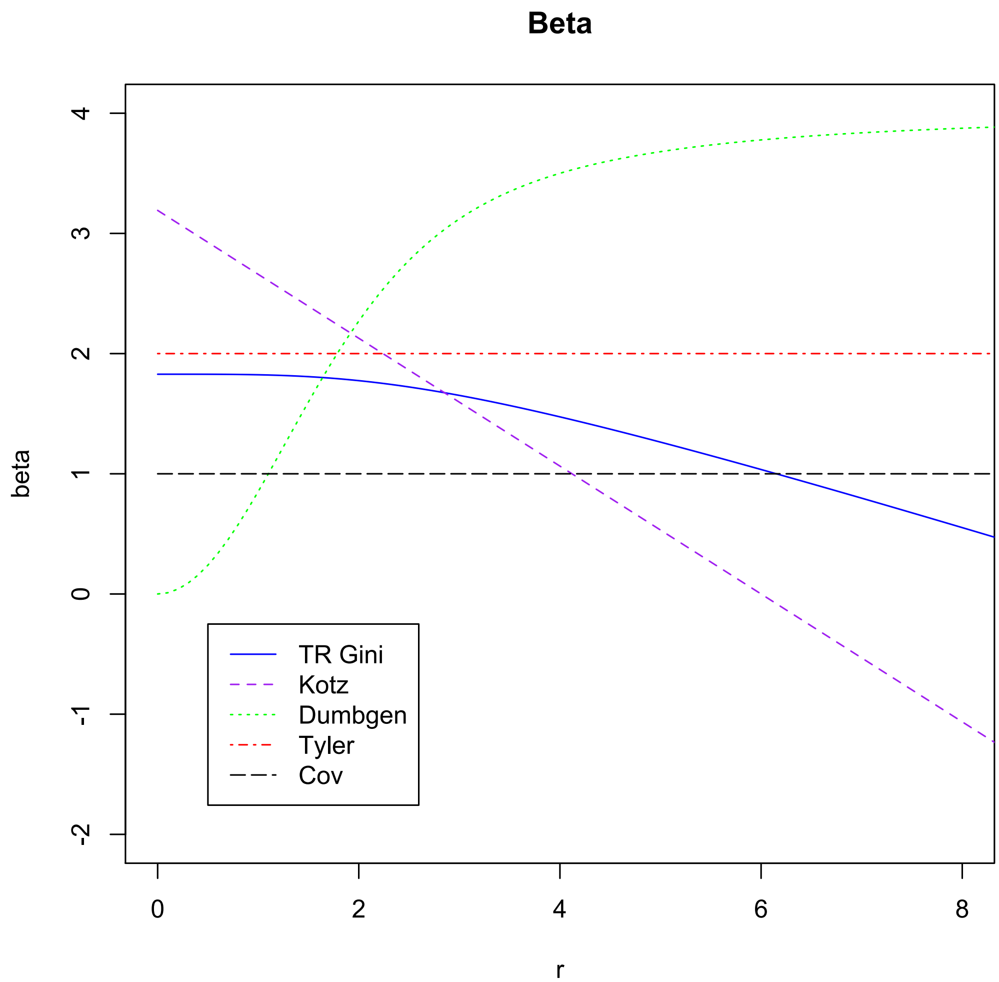


Figure 3.4.1: Function $\alpha_M(r)$

$\alpha_M(r)$ of the influence functions for covariance matrix, Tyler M functional, Dübgen functional, Kotz functional and TR Gini covariance matrix under the bivariate standard normal distribution.



s

Figure 3.4.2: Function $\beta_M(r)$

$\beta_M(r)$ of the influence functions for covariance matrix, Tyler M functional, Dübgen functional, Kotz functional and TR Gini covariance matrix under the bivariate standard normal distribution.

4 ESTIMATION OF THE TWO GINI COVARIANCE MATRICES

In this chapter, we discuss estimators of both GCMs. The sample Gini Covariance Matrix has the form of a U-statistic; therefore, its $\sqrt{n}$-consistency and asymptotic normality follow from well-established U-statistics theory. The TR GCM estimator is a symmetrized M-estimator. We prove $\sqrt{n}$-consistency of TR Gini estimator and establish its asymptotic normality. Asymptotic efficiency of TR GCM estimator is compared with other estimators.

4.1 SAMPLE GINI COVARIANCE MATRIX

Suppose $\mathcal{X} = \{\boldsymbol{x}_1, ..., \boldsymbol{x}_n\}$ is a random sample from a continuous distribution F in $\mathbb{R}^d$. The sample counterpart of the Gini covariance matrix is defined using the empirical distribution F_n in (3.1.4).

$$\hat{\boldsymbol{\Sigma}}_g = \boldsymbol{\Sigma}_g(\mathcal{X}) = \boldsymbol{\Sigma}_g(F_n) = 2\frac{1}{n}\sum_{i=1}^{n} \boldsymbol{x}_i \boldsymbol{r}(\boldsymbol{x}_i)^T = \frac{1}{\binom{n}{2}} \sum_{1 \le i < j \le n} \frac{(\boldsymbol{x}_i - \boldsymbol{x}_j)(\boldsymbol{x}_i - \boldsymbol{x}_j)^T}{\|\boldsymbol{x}_i - \boldsymbol{x}_j\|}. \qquad (4.1.1)$$

Clearly, the sample Gini covariance matrix $\boldsymbol{\Sigma}_g(F_n)$ is a matrix-valued U-statistic U_n of the form in Formula (1.8.7) for estimation of $\boldsymbol{\Sigma}_g(F)$ with kernel $h(\boldsymbol{x}_1, \boldsymbol{x}_2) = \frac{(\boldsymbol{x}_1 - \boldsymbol{x}_2)(\boldsymbol{x}_1 - \boldsymbol{x}_2)^T}{\|\boldsymbol{x}_1 - \boldsymbol{x}_2\|}$. The size of the kernel is 2. A straightforward generalization of univariate results given by Serfling [25] establishes $\sqrt{n}$-consistency of $\boldsymbol{\Sigma}_g(F_n)$. This means that for F having a finite second moment,

$$\sqrt{n}(\hat{\boldsymbol{\Sigma}}_g - \boldsymbol{\Sigma}_g) = \sqrt{n}(U_n - \boldsymbol{\Sigma}_g) = \sqrt{n}\left[\frac{1}{n}\sum_{i=1}^{n} IF(\boldsymbol{x}_i; \boldsymbol{\Sigma}_g, F)\right] + R_n, \qquad (4.1.2)$$

where the remainder term satisfies $R_n \xrightarrow{p} \mathbf{0}$. We have the following theorem.

Theorem 4.1.1. *Let $\boldsymbol{x}_1, ..., \boldsymbol{x}_n$ be a random sample from d-variate distribution F with finite second moment. Then, $\boldsymbol{\Sigma}_g(F_n)$ is an unbiased $\sqrt{n}$-consistent estimator of $\boldsymbol{\Sigma}_g(F)$. Furthermore,*

$$\sqrt{n}\, vec(\hat{\boldsymbol{\Sigma}}_g - \boldsymbol{\Sigma}_g) \to \mathcal{N}_{d^2}(\mathbf{0}, 4\mathbb{E}[\boldsymbol{\psi}(\boldsymbol{x})\boldsymbol{\psi}(\boldsymbol{x})^T]),$$

where $\boldsymbol{\psi}(\boldsymbol{x}) = vec(\mathbb{E}_{\boldsymbol{x}_1|\boldsymbol{x}} h(\boldsymbol{x}, \boldsymbol{x}_1) - \boldsymbol{\Sigma}_g)$ with $h(\boldsymbol{x}, \boldsymbol{x}_1) = (\boldsymbol{x} - \boldsymbol{x}_1)(\boldsymbol{x} - \boldsymbol{x}_1)^T / \|\boldsymbol{x} - \boldsymbol{x}_1\|$ and $\boldsymbol{x}_1$ is an r.v. from F.

Note that $2\boldsymbol{\psi}(\boldsymbol{x}) = vec(\mathbb{E}[IF(\boldsymbol{x}; \boldsymbol{\Sigma}_g, F)])$. The assumption on finite second moment guarantees existence $\mathbb{E}[\boldsymbol{\psi}(\boldsymbol{x})\boldsymbol{\psi}(\boldsymbol{x})^T]$.

4.2 SAMPLE TR GINI COVARIANCE MATRIX

Replacing F by F_n in (3.3.1), the sample affine equivariant Gini covariance matrix $\hat{\boldsymbol{\Sigma}}_G$ is defined as the solution of

$$\frac{2}{n(n-1)} \sum_{1 \le i < j \le n} \frac{\hat{\boldsymbol{\Sigma}}_G^{-1/2}(\boldsymbol{x}_i - \boldsymbol{x}_j)(\boldsymbol{x}_i - \boldsymbol{x}_j)^T \hat{\boldsymbol{\Sigma}}_G^{-1/2}}{\sqrt{(\boldsymbol{x}_i - \boldsymbol{x}_j)^T \hat{\boldsymbol{\Sigma}}_G^{-1} (\boldsymbol{x}_i - \boldsymbol{x}_j)}} - \frac{c(F)}{d}\boldsymbol{I}_d = \mathbf{0}. \tag{4.2.1}$$

Existence and uniqueness of the solution of (4.2.1) can be established by checking conditions of scatter M-estimators in Huber & Ronchetti (see [10].) Those conditions are also used for symmetrized M-estimators in Sirkiä *et al.* (see [28]). For weight functions w_1 and w_2, a symmetrized M-estimator is the solution of M in the following equation:

$$\frac{2}{n(n-1)} \sum_{1 \le i < j \le n} \{w_1(r_{ij})\boldsymbol{s}_{ij}(M)\boldsymbol{s}_{ij}^T(M) - w_2(r_{ij})\boldsymbol{I}_d\} = \mathbf{0}, \tag{4.2.2}$$

where $r_{ij}(M) = \|M^{-1/2}(\boldsymbol{x}_i - \boldsymbol{x}_j)\|$ and $\boldsymbol{s}_{ij}(M) = r_{ij}^{-1} M^{-1/2}(\boldsymbol{x}_i - \boldsymbol{x}_j)$.

The following conditions ensure existence (E) and uniqueness (U) of symmetrized M-estimators:

E1 $w_1(r)/r^2$ is decreasing, and positive for $r > 0$.

E2 $w_2(r)$ is increasing, and positive for $r \geq 0$.

E3 $w_1(r)$ and $w_2(r)$ are bounded and continuous.

E4 $w_1(0)/w_2(0) < d$.

E5 For any hyperplane H, let $P(H)$ be the fraction of pairwise difference belonging to that hyperplane. $P(H) < 1 - dw_2(\infty)/w_1(\infty)$ and $P(H) \leq 1/d$.

U1 $w_1(r)/r^2$ decreasing.

U2 $w_1(r)$ is continuous and increasing, and positive when $r > 0$.

U3 $w_2(r)$ is continuous and decreasing, non-negative, and positive when $0 \leq r < r_0$ for some r_0.

U4 For all hyperplane H, $P(H) < 1/2$.

E3' The distribution of F has a finite first moment.

In order to prove both uniqueness and existence simultaneously for a symmetrized M-estimator, w_2 must be constant because of assumptions E2 and U3 (see [28]). Note that in the previous conditions, "increasing" and "decreasing" are understood to be "nondecreasing" and "nonincreasing," respectively.

The affine equivalent version of Gini covariance estimator takes the form of a solution for Equation (4.2.2) with $w_1(r) = r$ and $w_2(r) = c(F)/d$. All conditions hold for the TR Gini except for assumption E3. However, if we replace E3 with E3', then Lemma 8.3 in Huber & Ronchetti [10] still holds. Hence our estimator does exist and exists uniquely.

Lemma 8.3 of Huber & Ronchetti [10] provides a condition along with E1, E2, continuity of E3, E4, and E5 such that the solution of (4.2.2) converges. The condition is that there is an $r_0 > 0$ such that

$$\frac{\mathbb{E}w_1(r_0\|\boldsymbol{x}\|)}{\mathbb{E}w_2(r_0\|\boldsymbol{x}\|)} < 1. \tag{4.2.3}$$

This condition holds if (E1-E3) hold. In our case, $w_1(r) = r$ is unbounded, which doesn't satisfy E3. However, if we assume E3' and let $0 < r_0 < c(F_0)/\mathbb{E}\|\boldsymbol{x}\| = \sqrt{2}$, then we have

$$\frac{\mathbb{E}w_1(r_0\|\boldsymbol{x}\|)}{\mathbb{E}w_2(r_0\|\boldsymbol{x}\|)} = \frac{r_0(\mathbb{E}\|\boldsymbol{x}\|)}{c(F_0)} < 1.$$

So, the result of the Lemma 8.3 from Huber & Ronchetti [10] follows.

By left and right multiplying $\hat{\Sigma}_G^{\frac{1}{2}}$ on both sides of Equation (4.2.1), we can see that $\hat{\boldsymbol{\Sigma}}_G$ is the solution of

$$\hat{\boldsymbol{\Sigma}}_G = \frac{2}{n(n-1)}\frac{d}{c(F)}\sum_{1\le i<j\le n}\frac{(\boldsymbol{x}_i - \boldsymbol{x}_j)(\boldsymbol{x}_i - \boldsymbol{x}_j)^T}{\sqrt{(\boldsymbol{x}_i - \boldsymbol{x}_j)^T(\hat{\boldsymbol{\Sigma}}_G^{-1}(\boldsymbol{x}_i - \boldsymbol{x}_j)}} \tag{4.2.4}$$

The solution of Equation of (4.2.4) can be found by a common iterative algorithm. First, set the initial value to be $\hat{\boldsymbol{\Sigma}}_G^{(0)} = \boldsymbol{I}_d$. Then for every integer $t > 0$, we define

$$\hat{\boldsymbol{\Sigma}}_G^{(t+1)} \longleftarrow \frac{2}{n(n-1)}\frac{d}{c(F)}\sum_{1\le i<j\le n}\frac{(\boldsymbol{x}_i - \boldsymbol{x}_j)(\boldsymbol{x}_i - \boldsymbol{x}_j)^T}{\sqrt{(\boldsymbol{x}_i - \boldsymbol{x}_j)^T(\hat{\boldsymbol{\Sigma}}_G^{(t)})^{-1}(\boldsymbol{x}_i - \boldsymbol{x}_j)}}. \tag{4.2.5}$$

The iteration stops when $\|\hat{\boldsymbol{\Sigma}}_G^{(t+1)} - \hat{\boldsymbol{\Sigma}}_G^{(t)}\| < \varepsilon$ for a pre-specified number $\varepsilon > 0$, where $\|\cdot\|$ can take any matrix norm. The Frobenius and Operator norms are used in this dissertation. We should mention that the estimator is a parametric estimator since $c(F)$ is involved in (4.2.5).

I f we assume that the location parameter $\boldsymbol{\mu}$ is known, then the MLE of $\boldsymbol{\Sigma}$ in the Kotz distribution is found to be scatter M-estimator, the solution $\hat{\boldsymbol{\Sigma}}_K$ of the equation

$$\frac{1}{n}\sum_{i=1}^{n}\frac{(\boldsymbol{x}_i - \boldsymbol{\mu})(\boldsymbol{x}_i - \boldsymbol{\mu})^T}{\sqrt{(\boldsymbol{x}_i - \boldsymbol{\mu})^T\hat{\boldsymbol{\Sigma}}_K^{-1}(\boldsymbol{x}_i - \boldsymbol{\mu})}} = \hat{\boldsymbol{\Sigma}}_K. \tag{4.2.6}$$

$\hat{\boldsymbol{\Sigma}}_K$ is a scatter M-estimator with the weight function $w_1(t) = t$.

4.3 $\sqrt{n}$-consistency

We shall establish $\sqrt{n}$-consistency of $\hat{\boldsymbol{\Sigma}}_G$ and $\hat{\boldsymbol{\Sigma}}_K$ under spherical distributions $F_0(g)$. The $\sqrt{n}$-consistency under elliptical distributions $F(\boldsymbol{\mu}, \boldsymbol{\Sigma}, g)$ follows from the fact that both estimators are affine equivariante. There are several issues in the proof of $\sqrt{n}$-consistency of $\hat{\boldsymbol{\Sigma}}_G$ and $\hat{\boldsymbol{\Sigma}}_K$. First of all, they are M-estimators, in other words, they are defined as a solution of equations, and there is no explicit form for the estimators. As a result, a linear expansion of $\hat{\boldsymbol{\Sigma}}_G$ or $\hat{\boldsymbol{\Sigma}}_K$ inevitably involves some inverse function. Secondly it is neccessary to define a manageable derivative for a matrix such that a Taylor expansion of $\hat{\boldsymbol{\Sigma}}_G$ or $\hat{\boldsymbol{\Sigma}}_K$ provides a rather explicit bound for the remainder term. The third issue is only for $\hat{\boldsymbol{\Sigma}}_G$. One needs to deal with paired differences, which are no longer independent. The results of Maronna [17] can not be directly applied to the Kotz estimator since $w_1(t) = t$ is unbounded. Here we follow the approach of Dümbgen (see [5]). In order to prove Theorem ahead, we will need three lemmas.

Let $\mathcal{P}$ denote the set of all distributions with a finite first moment. Let $H(\boldsymbol{x}) := \boldsymbol{x}\boldsymbol{s}^T(\boldsymbol{x}) = \|\boldsymbol{x}\|^{-1}\boldsymbol{x}\boldsymbol{x}^T$ and $H(F) := \mathbb{E}_F H(\boldsymbol{x})$. For $M \in \mathcal{M}^+$, $H(M^{-1/2}F) = \mathbb{E}_F H(M^{-1/2}\boldsymbol{x})$. Then for the distribution F with $\boldsymbol{\mu}(F) = 0$, $\boldsymbol{\Sigma}_K(F)$ is the solution of M with $H(M^{-1/2}F) = \boldsymbol{I}$. Let $\boldsymbol{x}_{12} = \boldsymbol{x}_1 - \boldsymbol{x}_2$, where $\boldsymbol{x}_1, \boldsymbol{x}_2$ are independent random vectors from a distribution F. The distribution of $\boldsymbol{x}_{12}$ is denoted as F^s. Then $H(\boldsymbol{x}_{12}) := \|\boldsymbol{x}_{12}\|^{-1}\boldsymbol{x}_{12}\boldsymbol{x}_{12}^T$ for $\boldsymbol{x}_1 \neq \boldsymbol{x}_2$, $H(F^s) = \int H(\boldsymbol{x}_{12})F(d(\boldsymbol{x})) = \mathbb{E}_F H(\boldsymbol{x}_{12})$. From Equation (3.3.1), we can see that the TR Gini covariance matrix Σ_G of F is the solution of $H(M^{-1/2}F^s) = \frac{c(F)}{d}\boldsymbol{I}$.

Definition 4.3.1. *For any $\boldsymbol{x} \in \mathbb{R}^d/\{\mathbf{0}\}$, the function $H(M^{-1/2}\boldsymbol{x})$ is differentiable with respect to $M \in \mathcal{M}^+$ in the sense of*

$$D(\boldsymbol{x}, B) := \frac{\partial}{\partial t}|_{t=0} H(\boldsymbol{I} + tB)^{-1/2}\boldsymbol{x}) = \frac{1}{2}\frac{\boldsymbol{x}^T B\boldsymbol{x}}{\|\boldsymbol{x}\|^2}H(\boldsymbol{x}) - \frac{1}{2}BH(\boldsymbol{x}) - \frac{1}{2}H(\boldsymbol{x})B.$$

$D(\boldsymbol{x}, B)$ *is called the* Gâteaux differential *of* $H(M^{-1/2}\boldsymbol{x})$ *at the identity matrix* $\boldsymbol{I}$ *in the direction of* B. [1]

Under the assumption of finite first moment of Q, $D(Q, B) = \mathbb{E}_Q D(\boldsymbol{x}, B)$ exists for all $B \in \mathcal{M}$. It is easy to prove that $D(Q, \cdot)$ is a linear mapping from $\mathcal{M}$ to $\mathcal{M}$.

Definition 4.3.2. *$D(Q, \cdot)$ is* nonsingular *if for any $B_1, B_2 \in \mathcal{M}$, $D(Q, B_1) = D(Q, B_2)$, then $B_1 = B_2$.*

The next lemma states that it is a one-to-one mapping.

Lemma 4.3.1. *For a distribution $Q \in \mathcal{P}$ having a first moment, $D(Q, \cdot)$ is nonsingular on $\mathcal{M}$.*

Proof of Lemma 4.3.1. Let $A = B_1 - B_2$, then

$$\int \frac{\boldsymbol{x}^T A \boldsymbol{x}}{\|\boldsymbol{x}\|^2} \frac{\boldsymbol{x}\boldsymbol{x}^T}{\|\boldsymbol{x}\|} - A \frac{\boldsymbol{x}\boldsymbol{x}^T}{\|\boldsymbol{x}\|} - \frac{\boldsymbol{x}\boldsymbol{x}^T}{\|\boldsymbol{x}\|} A \, dQ(\boldsymbol{x}) = \boldsymbol{0}.$$

Let $\lambda_1 \geq ... \geq \lambda_d$ be the eigenvalues of A and U is a matrix formed by eigenvectors of A. Left multiplying U and right multiplying U^T and multiplying $\boldsymbol{I} = U^T U$ in the middle, we have

$$\int \frac{\boldsymbol{x}^T A \boldsymbol{x}}{\|\boldsymbol{x}\|^2} \frac{U\boldsymbol{x}\boldsymbol{x}^T U^T}{\|\boldsymbol{x}\|} - UAU^T \frac{U\boldsymbol{x}\boldsymbol{x}^T U^T}{\|\boldsymbol{x}\|} - \frac{U\boldsymbol{x}\boldsymbol{x}^T U^T}{\|\boldsymbol{x}\|} UAU^T \, dQ(\boldsymbol{x}) = \boldsymbol{0}.$$

Let $\boldsymbol{y} = U\boldsymbol{x} = r\boldsymbol{v}$, where $r = \|\boldsymbol{y}\|$ and $\boldsymbol{v} = \|\boldsymbol{y}\|^{-1}\boldsymbol{y}$. We obtain

$$\int r(\boldsymbol{v}^T \Lambda \boldsymbol{v} \boldsymbol{v} \boldsymbol{v}^T - \Lambda \boldsymbol{v}\boldsymbol{v}^T - \boldsymbol{v}\boldsymbol{v}^T \Lambda) \, dQ^*(\boldsymbol{y}) = \boldsymbol{0}, \quad (4.3.3)$$

where $\Lambda = \text{diag}(\lambda_1, ..., \lambda_d)$ and $Q^*(\boldsymbol{y})$ be the distribution of $\boldsymbol{y}$. The diagonal elements of the matrix on the left side of (4.3.3) are $\int (\boldsymbol{v}^T \Lambda \boldsymbol{v} - 2\lambda_i) r v_i^2 \, dQ^*(\boldsymbol{y})$ for $i = 1, ..., d$. By

[1] Although B is a matrix, we may regard it as a vector in $\mathbb{R}^{d^2}$.

the fact that $\lambda_d \leq \boldsymbol{v}^T \Lambda \boldsymbol{v} \leq \lambda_1$ and $rv_i^2 > 0$, we have $\int (\lambda_d - 2\lambda_d) rv_d^2 \, dQ^*(\boldsymbol{y}) \leq 0$ and $\int (\lambda_1 - 2\lambda_1) rv_1^2 \, dQ^*(\boldsymbol{y}) \geq 0$. Hence $\lambda_i(A) = 0$ for all i.

□

The inverse operator of $D(Q, \cdot)$ is denoted as $D^{-1}(Q, \cdot)$. For $M \in \mathcal{M}$, let $\|M\| := \max\{|\lambda_1|, |\lambda_d|\}$. This is the spectral radius or the norm of the operator M. In general, we use the norm for a linear operator as $\|L\| = \max_{\|\boldsymbol{y}\|=1} \|L\boldsymbol{y}\|$. The following lemma establishes a bound on the score function.

Lemma 4.3.2. *For $Q \in \mathcal{P}$, there exists a constant $\kappa \in \mathbb{R}^+$ (not depending on Q) such that*

$$\|H((\boldsymbol{I} + B)^{-1/2} Q) - H(Q) - D(Q, B)\| \leq \kappa \|H(Q)\| \|B\|^2 + O(\|B\|^3)$$

for $B \in \mathcal{M}$ with $\|B\| \leq 1/2$.

Proof of Lemma 4.3.2. For $A \in \mathcal{M}$ with $\|A\| < \sqrt{2} - 1$, define

$$K(\boldsymbol{x}, A) = H((I - A)\boldsymbol{x}) - H(\boldsymbol{x}) - 2D(\boldsymbol{x}, A).$$

Then for $\boldsymbol{z} = \|\boldsymbol{x}\|^{-1} \boldsymbol{x}$,

$$\begin{aligned} K(\boldsymbol{x}, A) &= \|\boldsymbol{x}\| \left[\frac{(I-A)\boldsymbol{z}\boldsymbol{z}^T(I-A)}{\sqrt{\boldsymbol{z}^T (I-A)^2 \boldsymbol{z}}} - \boldsymbol{z}\boldsymbol{z}^T - 2D(\boldsymbol{z}, A) \right] \\ &= \|\boldsymbol{x}\| \frac{H(\boldsymbol{z}) - AH(\boldsymbol{z}) - H(\boldsymbol{z})A + AH(\boldsymbol{z})A - \sqrt{1 - 2\boldsymbol{z}^T A \boldsymbol{z} + \boldsymbol{z}^T A^2 \boldsymbol{z}} [H(\boldsymbol{z}) + 2D(\boldsymbol{z}, A)]}{\sqrt{\boldsymbol{z}^T (I-A)^2 \boldsymbol{z}}}. \end{aligned}$$

Using the Taylor expansion of $\sqrt{1-x}$ as a function of $x = 2\boldsymbol{z}^T A \boldsymbol{z} + \boldsymbol{z}^T A^2 \boldsymbol{z}$

$$\sqrt{1 - 2\boldsymbol{z}^T A \boldsymbol{z} + \boldsymbol{z}^T A^2 \boldsymbol{z}} = 1 - \boldsymbol{z}^T A \boldsymbol{z} + \frac{1}{2} \boldsymbol{z}^T A^2 \boldsymbol{z} - \frac{1}{2} (\boldsymbol{z}^T A \boldsymbol{z})^2 + O(\|A\|^3),$$

we have

$$K(\boldsymbol{x}, A) = \frac{\|\boldsymbol{x}\|}{\sqrt{\boldsymbol{z}^T(I-A)^2\boldsymbol{z}}}[AH(\boldsymbol{z})A + c_1(\boldsymbol{z}, A)H(\boldsymbol{z}) + c_2(\boldsymbol{z}, A)(AH(\boldsymbol{z}) + H(\boldsymbol{z})A)],$$

where $c_1(\boldsymbol{z}, A) = \frac{3}{2}(\boldsymbol{z}^T A\boldsymbol{z})^2 - \frac{1}{2}\boldsymbol{z}^T A^2\boldsymbol{z} + O(\|A\|^3)$ and $c_2(\boldsymbol{z}, A) = -\boldsymbol{z}^T A\boldsymbol{z} + O(\|A\|^2)$. Let $\boldsymbol{u}$ be any unit vector then we can find a unit vector $\boldsymbol{v} = \frac{A\boldsymbol{u}}{\|A\boldsymbol{u}\|}$ such that $A\boldsymbol{u} = \|A\boldsymbol{u}\|\boldsymbol{v}$, then

$$|\boldsymbol{u}^T(AH(\boldsymbol{z})A + c_1(\boldsymbol{z}, A)H(\boldsymbol{z}))\boldsymbol{u}| \leq 3\|A\|^2(\boldsymbol{v}^T H(\boldsymbol{z})\boldsymbol{v} + \boldsymbol{u}^T H(\boldsymbol{z})\boldsymbol{u}) + O(\|A\|^3);$$

$$|\boldsymbol{u}^T c_2(\boldsymbol{z}, A)(AH(\boldsymbol{z}) + H(\boldsymbol{z})A)\boldsymbol{u}| \leq \|A\|^2(|\boldsymbol{v}^T H(\boldsymbol{z})\boldsymbol{u}| + |\boldsymbol{u}^T H(\boldsymbol{z})\boldsymbol{v}|) + O(\|A\|^3).$$

Further, there are orthonormal vectors $\tilde{\boldsymbol{u}}$, $\tilde{\boldsymbol{v}}$ such that

$$\begin{aligned} \boldsymbol{u} &= ((1 + \boldsymbol{u}^T\boldsymbol{v})/2)^{1/2}\tilde{\boldsymbol{u}} + ((1 - \boldsymbol{u}^T\boldsymbol{v})/2)^{1/2}\tilde{\boldsymbol{v}}, \\ \boldsymbol{v} &= ((1 + \boldsymbol{u}^T\boldsymbol{v})/2)^{1/2}\tilde{\boldsymbol{u}} - ((1 - \boldsymbol{u}^T\boldsymbol{v})/2)^{1/2}\tilde{\boldsymbol{v}}. \end{aligned}$$

Hence

$$\begin{aligned} |\boldsymbol{u}^T H(\boldsymbol{z})\boldsymbol{v}| &= \tfrac{1}{2}|(1 + \boldsymbol{u}^T\boldsymbol{v})\tilde{\boldsymbol{u}}^T H(\boldsymbol{z})\tilde{\boldsymbol{u}} - (1 - \boldsymbol{u}^T\boldsymbol{v})\tilde{\boldsymbol{v}}^T H(\boldsymbol{z})\tilde{\boldsymbol{v}}| \\ &\leq \tfrac{1}{2}(1 + \boldsymbol{u}^T\boldsymbol{v})\tilde{\boldsymbol{u}}^T H(\boldsymbol{z})\tilde{\boldsymbol{u}} + \tfrac{1}{2}(1 - \boldsymbol{u}^T\boldsymbol{v})\tilde{\boldsymbol{v}}^T H(\boldsymbol{z})\tilde{\boldsymbol{v}} \\ &\leq \tilde{\boldsymbol{u}}^T H(\boldsymbol{z})\tilde{\boldsymbol{u}} + \tilde{\boldsymbol{v}}^T H(\boldsymbol{z})\tilde{\boldsymbol{v}}. \end{aligned}$$

Similarly, $|\boldsymbol{v}^T H(\boldsymbol{z})\boldsymbol{u}| \leq \tilde{\boldsymbol{u}}^T H(\boldsymbol{z})\tilde{\boldsymbol{u}} + \tilde{\boldsymbol{v}}^T H(\boldsymbol{z})\tilde{\boldsymbol{v}}$. Therefore

$$\begin{aligned} \|K(Q, A)\| &\leq \max_{\|\boldsymbol{u}\|=1} \int |\boldsymbol{u}^T K(\boldsymbol{x}, A)\boldsymbol{u}|\, dQ(\boldsymbol{x}) \\ &\leq \lambda_1((I - A)^{-2})(9\|A|\|^2 + O(\|A\|^3)) \max_{\|\boldsymbol{u}\|=1} \boldsymbol{u}^T H(Q)\boldsymbol{u} \\ &= \lambda_1((I - A)^{-2})(9\|A|\|^2 + O(\|A\|^3))\|H(Q)\|. \end{aligned}$$

Now let $B \in \mathcal{M}$ with $\|B\| \leq 1/2$ and let $A = I - (I + B)^{-1/2}$. Then it follows from the spectral representation of B and A and a Taylor expansion of the function $t \mapsto 1-(1+t)^{-1/2}$, $\lambda_1((I-A)^{-2}) \leq 1+\|B\|$, $\|2A-B\| \leq (3/2)\|B\|^2+O(\|B\|^3)$ and $\|A\| \leq (5/4)\|B\|+O(\|B\|^2)$.

Moreover, we have

$$\|D(Q,A)\| \leq \frac{3}{2}\|A\|\|H(Q)\| \tag{4.3.4}$$

for any $A \in \mathcal{M}$. Hence

$$\begin{aligned}
&\|H((I+B)^{-1/2}Q) - H(Q) - D(Q,B)\| \\
&\leq \|K(Q,A)\| + \|D(Q,2A-B)\| \\
&\leq (1+\|B\|)(9\|A\|^2 + O(\|A\|^3))\|H(Q)\| + 3/2\|H(Q)\|\|2A-B\| \\
&\leq 5\|B\|^2\|H(Q)\| + O(\|B\|^3).
\end{aligned} \tag{4.3.5}$$

□

The following lemma provides a basic linear expansion of $\mathbf{\Sigma}_G(\cdot)$ and $\mathbf{\Sigma}_K(\cdot)$ at F_0.

Lemma 4.3.3. *Suppose $\|D^{-1}(F_0,\cdot)\| \leq b$ for some $b \in \mathbb{R}^+$ and $F_0, Q \in \mathcal{P}$. Then there exist finite constants $\kappa_1(b), \kappa_2(b), \kappa_3(b), \kappa_4(b)$ such that*

$$\begin{aligned}
&\|\mathbf{\Sigma}_G(Q) - \boldsymbol{I} + D^{-1}(F_0, H(Q^s - \frac{c(Q)}{c(F_0)}F_0^s))\| \\
&\leq \kappa_1(b)\|H(Q^s - F_0^s)\|^2 + \kappa_2(b)|c(Q) - c(F_0)|^2, \\
&\|\mathbf{\Sigma}_K(Q) - c(F_0)^2/(2d^2)\boldsymbol{I} + D^{-1}(F_0, H(Q - \frac{c(Q)}{c(F_0)}F_0))\| \\
&\leq \kappa_3(b)\|H(Q - F_0)\|^2 + \kappa_4(b)|c(Q) - c(F_0)|^2.
\end{aligned}$$

Proof of Lemma 4.3.3. Let $L = D^{-1}(F_0^s,\cdot)$. Assume that $\|L\| \leq b < \infty$, $Q \in \mathcal{P}$ and $\|H(Q - F_0)\| \leq \epsilon < \infty$. Let $g(B) = L(H((I+B)^{-1/2}Q^s) - \frac{c(Q)}{d}I)$, which is a continuous mapping from $\mathcal{M}^\rho = \{B \in \mathcal{M} : \|B\| \leq \rho\}$ into $\mathcal{M}$, where $\rho \in (0, 1/2]$ is some constant.

Since $H(F_0^s) = \frac{c(F_0)}{d} I$, we have

$$\begin{aligned} g(B) &= L(H(Q^s - \tfrac{c(Q)}{c(F_0)} F_0^s)) + L(H((I+B)^{-1/2} Q^s) - H(Q^s)) \\ &= L(H(Q^s - \tfrac{c(Q)}{c(F_0)} F_0^s)) + B + LD(Q^s - F_0^s, B) \\ &\quad + L(H((I+B)^{-1/2} Q^s) - H(Q^s) - D(Q^s, B)) \\ &= L(H(Q^s - \tfrac{c(Q)}{c(F_0)} F_0^s)) + B + R(B), \end{aligned}$$

where $R(B) = LD(Q^s - F_0^s, B) + L(H((I+B)^{-1/2} Q^s) - H(Q^s) - D(Q^s, B))$. By inequalities (4.3.5) and (4.3.4), we have

$$\begin{aligned} \|R(B)\| &< b\|D(Q^s - F_0^s, \cdot)\|\|B\| + 5b\|H(Q^s)\|\|B\|^2 \\ &\leq 3/2b\|H(Q^s - F_0^s)\|\|B\| + 5b\|H(Q^s)\|\|B\|^2 \qquad (4.3.6) \\ &= b(3/2\epsilon + 5\mu\rho)\|B\|. \end{aligned}$$

Since $\|L(H(Q^s - F_0^s))\| \leq \sqrt{2} b\epsilon$, if $b\epsilon, b\mu\rho$ are sufficiently small, then

$$\|R(B)\| \leq \|B\|/2 \quad \text{and} \quad \|B - g(B)\| \leq \rho \quad \text{for all } B \in \mathcal{M}^\rho.$$

By Brouwer's Fixed Point theorem (theorem 1.8.2), there is $B_0 \in \mathcal{M}^\rho$ such that $g(B_0) = 0$, which is equivalent to $H((I+B_0)^{-1/2} Q^s) = \frac{c(Q)}{d} \boldsymbol{I}$. By the definition of the TR gini covariance matrix, $\boldsymbol{\Sigma}_G(Q) = I + B_0$. Then by the triangle inequality, $\|B_0\| \leq \|LH(Q^s - \frac{c(Q)}{c(F_0)} F_0^s)\| + \|R(B_0)\| \leq b\|H(Q^s - \frac{c(Q)}{c(F_0)} F_0^s)\| + \|B_0\|/2$, hence $\|B_0\| \leq 2b\|H(Q^s - \frac{c(Q)}{c(F_0)} F_0^s)\|$.

$$\begin{aligned} \|B_0 + LH(Q^s - \tfrac{c(Q)}{c(F_0)} F_0^s)\| &= \|R(B_0)\| \\ &\leq 3b^2 \|H(Q^s - \tfrac{c(Q)}{c(F_0)} F_0^s)\|\|H(Q^s - F_0^s)\| \\ &\quad + 20b^2 \|H(Q^s)\|\|H(Q^s - \tfrac{c(Q)}{c(F_0)} F_0^s)\|^2. \end{aligned}$$

Since $\|H(Q^s - \frac{c(Q)}{c(F_0)}F_0^s)\| \leq \|H(Q^s - F_0^s)\| + |c(F_0) - c(Q)|$, this completes the proof of the expansion of $\Sigma_G(\cdot)$ in Lemma 4.3.3. The proof of the expansion of $\Sigma_K(\cdot)$ follows the same idea.

□

This lemma provides explicit bounds for the remainder of linear expansions of $\boldsymbol{\Sigma}_G(Q)$ and $\boldsymbol{\Sigma}_K(Q)$ at $\boldsymbol{\Sigma}_G(F_0) = \boldsymbol{I}$ and $\boldsymbol{\Sigma}_K(F_0) = c(F_0)^2/(2d^2)\boldsymbol{I}$, respectively.

Now we are ready to state the $\sqrt{n}$-consistency property of $\hat{\boldsymbol{\Sigma}}_G$ and $\hat{\boldsymbol{\Sigma}}_K$.

Theorem 4.3.4. *Let $\boldsymbol{x}_1, ..., \boldsymbol{x}_n$ be a random sample from a spherical distribution F_0 in $\mathbb{R}^d$. Under the assumption of finite second moment of F_0, $\hat{\boldsymbol{\Sigma}}_G$ is $\sqrt{n}$-consistent estimator of $\boldsymbol{\Sigma}_G(F_0) = \boldsymbol{I}_d$ and $\hat{\boldsymbol{\Sigma}}_K$ is $\sqrt{n}$-consistent estimator of $\boldsymbol{\Sigma}_K(F_0) = c(F_0)^2/(2d^2)\boldsymbol{I}_d$.*

Proof of Theorem 4.3.4. We apply the results of Lemma 4.3.3 by letting $Q = F_n$, where F_n is the empirical distribution based on a random sample of size n from F_0. Recall that from Remark 3.2.3,

$$c(Q) = \mathbb{E}_{F_n}\|\boldsymbol{x}_1 - \boldsymbol{x}_2\| = \binom{n}{2}^{-1} \sum_{i \neq j} \|\boldsymbol{x}_i - \boldsymbol{x}_j\|.$$

Clearly, $c(Q)$ is a U-statistic with the kernel of size 2. If the second moment exists, then the U-statistics theorem gives

$$\sqrt{n}(c(Q) - c(F_0)) = \sqrt{n}\left[\frac{1}{n}\sum_{i=1}^{n} 2(\mathbb{E}_{\boldsymbol{x}}\|\boldsymbol{x}_i - \boldsymbol{x}\| - c(F_0))\right] + o_p(1),$$

with $\boldsymbol{x} \sim F_0$ and $\boldsymbol{x}_i$ is fixed. Also by Formula (4.1.2), we have

$$\sqrt{n}H(Q^s - F_0^s) = \sqrt{n}(\hat{\boldsymbol{\Sigma}}_g - \boldsymbol{\Sigma}_g) = \sqrt{n}\left[\frac{1}{n}\sum_{i=1}^{n} IF(\boldsymbol{x}_i; \boldsymbol{\Sigma}_g, F_0)\right] + o_p(1).$$

Now applying Lemma 4.3.3, we have

$$\|\hat{\boldsymbol{\Sigma}}_G - \boldsymbol{I} + D^{-1}(F_0, H(Q^s - \frac{c(F_n)}{c(F_0)}F_0^s))\| = O(n^{-1}), \tag{4.3.7}$$

which implies $\sqrt{n}$-consistency of $\hat{\mathbf{\Sigma}}_G$ under the assumption that $\|D^{-1}(F_0, \cdot)\|$ is bounded, or at least bounded in probability.

$\square$

Remark 4.3.8. *If F_0 is the spherically distributed Kotz distribution, then $\mathbf{\Sigma}_K(F_0) = \boldsymbol{I}_d$ and both $\hat{\mathbf{\Sigma}}_G$ and $\hat{\mathbf{\Sigma}}_K$ are consistent scatter estimators.*

4.3 Asymptotic Normality

Theorem 4.3.5. *Let $\boldsymbol{x}_1, ..., \boldsymbol{x}_n$ be a random sample from a spherical distribution F_0 in $\mathbb{R}^d$. If the covariance matrix of F_0 exists, then*

$$\sqrt{n}\ vec(\hat{\mathbf{\Sigma}}_G - \boldsymbol{I}_d) \to N_{d^2}(\mathbf{0}, \mathbb{E}[vec(IF(\boldsymbol{x}; \mathbf{\Sigma}_G, F_0))vec(IF(\boldsymbol{x}; \mathbf{\Sigma}_G, F_0))^T]).$$

Proof of Theorem 4.3.5. Once we proved $\sqrt{n}$-consistency of $\hat{\mathbf{\Sigma}}_G$, we are able to use the Lemma 2 of Sirkiä *et al.* [28], that we state as follows:

Lemma 4.3.6 (Sirkiä *et al.* [28])**.** *Let $\boldsymbol{x}_1, ...\boldsymbol{x}_n$ be a sample from a spherically symmetric distribution and let $\boldsymbol{x}_{ij} = \boldsymbol{x}_i - \boldsymbol{x}_j$, $r_{ij} = \|\boldsymbol{x}_{ij}\|$, and $\boldsymbol{u}_{ij} = r^{-1}\boldsymbol{x}_{ij}$. Assume that the symmetrised M-estimator $\hat{\boldsymbol{M}}$ is $\sqrt{n}-$consistent. Then*

$$\sqrt{n}(\hat{\boldsymbol{M}} - \boldsymbol{I}_d) = \sqrt{n}\left[\binom{n}{2}^{-1}\sum_{i<j}\left(\frac{w_1(r_{ij})}{2\eta_1}\left(\boldsymbol{u}_{ij}\boldsymbol{u}_{ij}^T - \frac{\eta_2 - \eta_1}{k\eta_2}\boldsymbol{I}_k\right) - \frac{w_2(r_{ij})}{2\eta_2}\boldsymbol{I}_d\right)\right] + o_p(1),$$

where

$$\begin{aligned}
\eta_1 &= \frac{\mathbb{E}[w_1'(\|\boldsymbol{x}_1 - \boldsymbol{x}_2\|)\|\boldsymbol{x}_1 - \boldsymbol{x}_2\| + dw_1(\|\boldsymbol{x}_1 - \boldsymbol{x}_2\|)]}{2d(d+2)}, \\
\eta_2 &= \frac{\mathbb{E}[w_1'(\|\boldsymbol{x}_1 - \boldsymbol{x}_2\|)\|\boldsymbol{x}_1 - \boldsymbol{x}_2\| - d^2 w_2'(\|\boldsymbol{x}_1 - \boldsymbol{x}_2\|)\|\boldsymbol{x}_1 - \boldsymbol{x}_2\|]}{4d}.
\end{aligned}$$

This lemma is applied to the TR Gini estimator where $\boldsymbol{w}_1(t) = t$, $\boldsymbol{w}_2(t) = \frac{c(F_0)}{d}$. Then

$$\begin{aligned}\eta_1 &= \frac{(d+1)\mathbb{E}\|\boldsymbol{x}_1 - \boldsymbol{x}_2\|}{2d(d+2)} = \frac{(d+1)c(F_0}{2d(d+2}, \\ \eta_2 &= \frac{\mathbb{E}\|\boldsymbol{x}_1 - \boldsymbol{x}_2\|}{4d} = \frac{c(F_0)}{4d}.\end{aligned}$$

Therefore,

$$\sqrt{n}(\hat{\boldsymbol{\Sigma}}_G - \boldsymbol{I}_d) = \sqrt{n}\left[\binom{n}{2}^{-1}\sum_{i<j}\boldsymbol{\psi}(\boldsymbol{x}_i, \boldsymbol{x}_j)\right] + o_p(1), \tag{4.3.9}$$

where

$$\boldsymbol{\psi}(\boldsymbol{x}_i, \boldsymbol{x}_j) = \frac{d(d+2)}{(d+1)c(F_0)}\frac{(\boldsymbol{x}_i - \boldsymbol{x}_j)(\boldsymbol{x}_i - \boldsymbol{x}_j)^T}{\|\boldsymbol{x}_i - \boldsymbol{x}_j\|} + \frac{d\|\boldsymbol{x}_i - \boldsymbol{x}_j\|}{(d+1)c(F_0)}\boldsymbol{I}_d - 2\boldsymbol{I}_d.$$

Formula (4.3.9) shows that $\sqrt{n}\, vec(\hat{\boldsymbol{\Sigma}}_G - \boldsymbol{I}_d)$ has the same limiting distribution as the d^2-variate U-statistic $\sqrt{n}\boldsymbol{U}_n$ of kernel size 2, where

$$\boldsymbol{U}_n = \binom{n}{2}^{-1}\sum_{i<j} vec(\boldsymbol{\psi}(\boldsymbol{x}_i, \boldsymbol{x}_j)).$$

Application of the U-statistics theorem yields asymptotical normality of $\boldsymbol{U}$ and hence

$$\sqrt{n}\, vec(\hat{\boldsymbol{\Sigma}}_G - \boldsymbol{I}_d) \to N_{d^2}(\boldsymbol{0}, 4\mathbb{E}\boldsymbol{\phi}(\boldsymbol{x})\boldsymbol{\phi}(\boldsymbol{x})^T),$$

where $\boldsymbol{\phi}(\boldsymbol{x}) = vec(\mathbb{E}_{\boldsymbol{x}_1}\boldsymbol{\psi}(\boldsymbol{x}, \boldsymbol{x}_1)) = 1/2 vec(IF(\boldsymbol{x}, \boldsymbol{\Sigma}_G, F_0))$. This concludes the proof of Theorem 4.3.5.

□

Using Formula (3.4.2) and Theorem 3.4.4, the covariance matrix of the limiting distribution, $\mathbb{E}[vec(IF(\boldsymbol{x}; \boldsymbol{\Sigma}_G, F_0))vec(IF(\boldsymbol{x}; \boldsymbol{\Sigma}_G, F_0))^T]$, can be written as

$$ASV(\hat{\boldsymbol{\Sigma}}_{G_{12}}; F_0)(\boldsymbol{I}_{d^2} + \boldsymbol{1}_{d,d}) + ASC(\hat{\boldsymbol{\Sigma}}_{G_{11};F_0}, \hat{\boldsymbol{\Sigma}}_{G_{22}}; F_0)vec(\boldsymbol{I}_d)vec(\boldsymbol{I}_d)^T,$$

where $\mathbf{1}_{d,d}$ is $d^2 \times d^2$ matrix with (i,j)-block being equal to a $d \times d$ matrix that has 1 at entry (j,i) and 0 elsewhere. $ASV(\hat{\mathbf{\Sigma}}_{G_{12}}; F_0)$ denotes the asymptotic variance of an off-diagonal element and $ASC(\hat{\mathbf{\Sigma}}_{G_{11}}, \hat{\mathbf{\Sigma}}_{G_{22}}; F_0)$ denotes the covariance of any two diagonal elements. Theorems 4.3.5 and 3.4.4 imply

$$ASV(\hat{\mathbf{\Sigma}}_{G_{12}}; F_0) = \tfrac{4d(d+2)}{(d+1)^2 c^2(F_0)} \mathbb{E}_{\boldsymbol{x}_2} \left[\mathbb{E}_{\boldsymbol{x}_1} (\|\boldsymbol{x}_1 - \|\boldsymbol{x}_2\| \boldsymbol{e}_1\| - \tfrac{d(\boldsymbol{x}_1)_2^2}{\|\boldsymbol{x}_1 - \|\boldsymbol{x}_2\| \boldsymbol{e}_1\|}) \right]^2;$$

$$ASV(\hat{\mathbf{\Sigma}}_{G_{11}}; F_0) = \tfrac{2(d-1)}{d} ASV(\hat{\mathbf{\Sigma}}_{G_{12}}; F_0) + 16 \left[\tfrac{\mathbb{E}_{\boldsymbol{x}_2} \left[\mathbb{E}_{\boldsymbol{x}_1} (\|\boldsymbol{x}_1 - \|\boldsymbol{x}_2\| \boldsymbol{e}_1\|) \right]^2}{c^2(F_0)} - 1 \right];$$

$$ASC(\hat{\mathbf{\Sigma}}_{G_{11}}, \hat{\mathbf{\Sigma}}_{G_{22}}; F_0) = ASV(\hat{\mathbf{\Sigma}}_{G_{11}}; F_0) - 2ASV(\hat{\mathbf{\Sigma}}_{G_{12}}; F_0).$$

Using the affine equivariance property of $\hat{\mathbf{\Sigma}}_G$ and Kronecker product $\otimes$, the limiting distribution of $\sqrt{n}\, vec(\hat{\mathbf{\Sigma}}_G - \mathbf{\Sigma})$ at the elliptical distribution F is multivariate normal with zero mean and covariance matrix

$$ASV(\hat{\mathbf{\Sigma}}_{G_{12}}; F_0)(\boldsymbol{I}_{d^2} + \mathbf{1}_{d,d})(\mathbf{\Sigma} \otimes \mathbf{\Sigma}) + ASC(\hat{\mathbf{\Sigma}}_{G_{11}}, \hat{\mathbf{\Sigma}}_{G_{22}}; F_0) vec(\mathbf{\Sigma}) vec(\mathbf{\Sigma})^T. \tag{4.3.10}$$

Checking the conditions (N1-N4) of MLE proposed by Huber [9], we are able to establish the normality of Kotz estimator $\hat{\mathbf{\Sigma}}_K$, assuming a known location parameter.

Theorem 4.3.7. *Let $\boldsymbol{x}_1, ..., \boldsymbol{x}_n$ be a random sample from a spherical distribution F_0 in $\mathbb{R}^d$. If the second moment of F_0 exists and the first moment is known, then*

$$\sqrt{n}\, vec(\hat{\mathbf{\Sigma}}_K - \mathbf{\Sigma}_K) \to N_{d^2}(\mathbf{0}, \mathbb{E}[vec(IF(\boldsymbol{x}; \mathbf{\Sigma}_K, F_0)) vec(IF(\boldsymbol{x}; \mathbf{\Sigma}_K, F_0))^T]).$$

With the results of Theorems 3.4.5 and 4.3.7, we have

$$ASV(\hat{\mathbf{\Sigma}}_{K_{12}}; F_0) = \frac{2d(d+2)\mathbb{E}[\|\boldsymbol{x}\|^2]}{(d+1)^2 (c(F_0))^2}. \tag{4.3.11}$$

4.3 Asymptotic Efficiency

Although our TR Gini covariance estimator is Fisher consistent to the scatter matrix since it is corrected by $c(F_0)/d$, we must consider its shape estimator in order to compare its limiting efficiency with that of the Tyler and Dübgen M-estimators. The shape matrix associated with the scatter functional $\boldsymbol{\Sigma}$ is given in Formula (1.3.8). Tyler and Dümbgen estimators estimate the shape matrix W. At elliptical distributions, all estimators of the shape matrix estimate the same population quantity; therefore, they are comparable without any correction factors. Theorem 5 of Sirkiä *et al.* [28] states the following:

Theorem 4.3.8. *Let $\hat{\boldsymbol{\Sigma}}$ be a scatter estimator with associated shape estimator*

$$\hat{W} = (d/Tr(\hat{\boldsymbol{\Sigma}})\hat{\boldsymbol{\Sigma}}.$$

The limiting distribution of $\sqrt{n}\, vec(\hat{W} - W)$ at elliptical F is multinormal with covariance matrix

$$\tau \left(\boldsymbol{I}_{d^2} - \frac{1}{d} vec(W) vec(\boldsymbol{I}_d)^T \right) (\boldsymbol{I}_{d^2} + \mathbf{1}_{d,d})(W \otimes W) \left(\boldsymbol{I}_{d^2} - \frac{1}{d} vec(\boldsymbol{I}_d) vec(W^T) \right),$$

where $\tau = ASV(\hat{W}_{12}, F_0) = ASV(\hat{\boldsymbol{\Sigma}}_{12}, F_0)$.

A single number τ, which is the variance of off-diagonal elements of $\hat{\boldsymbol{\Sigma}}$ or $\hat{W}$ at F_0, characterizes the limiting distribution of the shape estimator. Hence, the asymptotic relative efficiencies of shape matrix estimators are defined as the ratios of the corresponding τ values.

Listed in Table 4.3.1 are the limiting efficiencies of shape estimators with respect to the shape estimator based on the regular sample covariance matrix (i.e. the regular shape estimator). The efficiencies are considered under the Kotz(d) distributions and $\mathcal{T}_d(\nu)$ distributions at different d dimension with different degrees of freedom ν, with $\nu = \infty$ referring to the normal case. The variance of the off-diagonal element of the regular shape estimator at F_0 equal to $1+\kappa(F_0)$, where $\kappa(F_0)$ is the kurtosis of F_0. That is, τ of the regular shape estimator is $(\nu-2)/(\nu-4)$ in the $\mathcal{T}_d(\nu)$-distributions for $\nu > 4$ and $(d+3)/(d+1)$ in the $Kotz(d)$ distribution (see [32, 37]). In the normal case, $\tau = 1$ corresponds to that of the $\mathcal{T}_d(\nu)$-distribution case when $\nu \to \infty$. τ of the Tyler estimator is always $(d+2)/d$ for any distribution in $\mathbb{R}^d$. From (4.3.11), the asymptotic variance of off-diagonal elements of the Kotz estimator under F_0 is equal to $2d(d+2)\mathbb{E}(\|\boldsymbol{x}\|)^2/((d+1)(c(F_0))^2)$ with $\boldsymbol{x}$ from F_0. For example, ASV of the Kotz estimator under the $Kotz(d)$ distribution is $(d+2)/(d+1)$. The variances of off-diagonal elements of the TR Gini covariance matrix are computed through a combination of numerical integration and Monte Carlo simulation. More specifically, to calculate $ASV(\hat{\boldsymbol{\Sigma}}_{G_{12}}; F_0)$, we need to evaluate double expectations that involves a $(d+1)$-dimensional integration as follows, where $\|\boldsymbol{x}_2\| = r$ follows the distribution with density in Equation 1.3.3.

$$ASV(\hat{\boldsymbol{\Sigma}}_{G_{12}}; F_0) = \tfrac{4d(d+2)}{(d+1)^2 c^2(F_0)} \mathbb{E}_{\boldsymbol{x}_2} \left[\mathbb{E}_{\boldsymbol{x}_1} (\|\boldsymbol{x}_1 - \|\boldsymbol{x}_2\| \boldsymbol{e}_1\| - \tfrac{d(\boldsymbol{x}_1)_2^2}{\|\boldsymbol{x}_1 - \|\boldsymbol{x}_2\| \boldsymbol{e}_1\|}) \right]^2;$$

The inside expectation ($\mathbb{E}_{\boldsymbol{x}_1}$) uses Monte Carlo simulation, while the outside expectation ($\mathbb{E}_{\boldsymbol{x}_2}$) uses numerical integration.

From Table 4.3.1, it can be seen that the ARE of each shape estimator increases as dimension d increases, but decreases as ν increases for $\mathcal{T}_d(\nu)$ distributions. The Kotz estimator is more efficient than the Tyler estimator for all distributions considered. The increases of the symmetrized estimators (TR Gini and Dümbgen) in efficiency comparing to their counterparts (Kotz and Tyler, respectively) are considerable for all cases. TR Gini estimator is also always more efficient than the Tyler estimator. In the normal case, TR Gini estimator has the limiting efficiency 0.98 at $d = 2$ and 0.99 at $d = 3, 4,$ and 5. With little loss on efficiency in the normal case, it gains high efficiency at heavy-tailed distributions.

		$\mathcal{T}_d(5)$	$\mathcal{T}_d(6)$	$\mathcal{T}_d(8)$	$\mathcal{T}_d(15)$	$\mathcal{T}_d(\infty)$	$Kotz(d)$
$d=2$	Tyler	1.50	1.00	0.75	0.59	0.50	0.83
	Dümbgen	2.36	1.57	1.26	1.01	0.91	1.22
	Kotz	2.25	1.56	1.22	1.00	0.88	1.25
	TR Gini	1.97	1.41	1.19	1.03	0.98	1.12
$d=3$	Tyler	1.80	1.20	0.90	0.71	0.60	0.90
	Dümbgen	2.38	1.66	1.27	1.04	0.92	1.18
	Kotz	2.31	1.60	1.25	1.03	0.91	1.20
	TR Gini	2.01	1.46	1.20	1.04	0.99	1.10
$d=4$	Tyler	2.00	1.33	1.00	0.79	0.67	0.93
	Dümbgen	2.39	1.69	1.30	1.06	0.93	1.15
	Kotz	2.34	1.63	1.27	1.05	0.92	1.17
	TR Gini	2.08	1.48	1.21	1.07	0.99	1.09
$d=5$	Tyler	2.14	1.43	1.07	0.84	0.71	0.95
	Dümbgen	2.50	1.71	1.31	1.07	0.94	1.13
	Kotz	2.37	1.65	1.29	1.06	0.93	1.14
	TR Gini	2.15	1.49	1.22	1.07	0.99	1.07

Table 4.3.1: ARE of Shape Estimators

Asymptotic relative efficiencies of the shape estimators based on Tyler M-estimator, Dümbgen, Kotz M-estimator and TR Gini covariance estimator relative to the regular shape estimator at different distributions F_0 at different d-dimension.

5 FINITE SAMPLE EFFICIENCY

In this chapter we look at the finite sample efficiency of the AFGC. Monte-Carlo simulation is used to compare it with other estimators. The efficiency is studied under various distributions using different criteria to compare.

5.1 CONSIDERED DISTRIBUTIONS

There are three cases considered for simulations. These cases involve the following distributions: t_ν distribution, Kotz distribution, and contaminated normal distributions.

Case I: Heavy-tailed $t_\nu(\mathbf{0}, \Sigma_{d\times d})$ distributions for degrees of freedom $\nu = 5, 6, 8, 15$ and ∞ at dimensions $d = 2$ and $d = 5$. Note that $\nu = \infty$ corresponds to the standard normal distribution. The R Package "mvtnorm" is used to generate samples from multivariate $\mathcal{T}$-distributions and multivariate normal distributions.

Case II: Kotz Distribution with dimension $d = 2$ and $d = 5$. We discuss how to generate random samples from the Kotz distribution later, and the R codes are provided in the Appendix.

Case III: Contaminated Normal distributions with ε to be 0, 0.1, 0.2 and $d = 2$ and $d = 5$.

- Contaminated Normal distributions on shifted locations. i.e $(1-\varepsilon)\mathcal{N}(\mathbf{0}, \Sigma_{d\times d}) + \varepsilon\mathcal{N}(\mathbf{10}_d, \Sigma_{d\times d})$, where $\mathbf{10}_d$ is the d-vector with all elements 10, $\Sigma_{d\times d} = \mathrm{diag}(4, \mathbf{1}_{d-1}^T)$.
- Contaminated Normal distributions with different Σ. i.e $(1-\varepsilon)\mathcal{N}(\mathbf{0}, \Sigma_{d\times d}) + \varepsilon\mathcal{N}(\mathbf{0}, \Sigma^*_{d\times d})$, where $\Sigma_{d\times d} = \mathrm{diag}(4, \mathbf{1}_{d-1}^T)$, $\Sigma^*_{d\times d} = 10 \times \mathrm{diag}(\mathbf{1}_{d-1}^T, 4)$.

5.1 Generating Random Sample from the Kotz Distribution

Although the Kotz distribution can be viewed as a multivariate generalization of Laplace distribution, a random sample of size d from a univariate standard Laplace distribution $\boldsymbol{z} = (z_1, ..., z_d)^T$ doesn't have a Kotz distribution since $f(\boldsymbol{z}) \propto e^{-\sum |z_i|} \neq e^{-\sqrt{z_1^2 + ... z_d^2}}$. In other words, a marginal distribution of a spherical Kotz random vector is not Laplace distribution. Hence we can not generate multivariate Kotz random sample from univariate Laplace random sample.

Here, we use the property of the spherical distribution F_0: if $\boldsymbol{z}$ is from a spherical distribution, then $r = \|\boldsymbol{z}\|$ and $\boldsymbol{u} = \boldsymbol{z}/r$ are independent and $\boldsymbol{u}$ is uniformly distributed on the unit sphere. First, by applying this property on the standard normal distribution, $\boldsymbol{u}$ can be generated by d i.i.d. standard normal variables by $\boldsymbol{z}/\|\boldsymbol{z}\|$. Then this property is applied to the Kotz distribution: $\boldsymbol{y} = r\boldsymbol{u}$ is from the spherical Kotz distribution, in which r is distributed from the Gamma distribution with the shape parameter being d and the scale parameter being 1. If a random sample from Kotz$(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ is required, then by taking $\boldsymbol{\Sigma}$'s Cholesky decomposition L, we have $\boldsymbol{x} = L\boldsymbol{y} + \boldsymbol{\mu}$ is from Kotz$(\boldsymbol{\mu}, \boldsymbol{\Sigma})$.

5.2 METHODS

To compare robustness and efficiency of the AFGC, we consider the following estimators:

Tyler: Tyler's M-estimator is obtained by the function *HR.Mest* in the R package ICSNP. It simultaneously yields affine equivariant spatial median (see Hettmansberger and Randles (2003)) and Tyler's shape matrix.

Dümbgen: Dümbgen's symmetrized M-estimator is the Tyler's M-estimator on the paired differences data with the location parameter is set to be $\mathbf{0}$.

Kotz: Kotz estimator is computed through a common iterative procedure.

MRCM: The function *spatial.rank* in the R package ICSNP is used for computing the spatial rank vector and rank covariance matrix (RCM). Then the spectral decomposition of RCM is performed. The eigenvalues are re-estimated through univariate scale estimators of the projected data on each eigenvector.

Mcd: Minimum covariance determinant estimator is computed by the R package rrcov. The MCD method looks for the h observations (out of n) whose classical covariance matrix has the lowest possible determinant. Then MCD scatter estimator is the covariance matrix based on those h observations.

Sest: Re-weighted S-estimator (Sest) is calculated by the R package riv using Tukey biweighted ρ function with $c = 2.661$ for $d = 2$ and $c = 4.652$ for $d = 5$. Such c values provided as output of the function *slc* yield the breakdown point close to 1/2.

Cov: Non-robust sample covariance matrix.

MRCM, MCD and S-estimator are highly robust with the breakdown point close to 1/2. Here, the breakdown point (BP) is another quantitative robustness measure proposed by Donoho & Huber (see [3]). Roughly speaking, the breakdown point is the minimum fraction of "bad" data points that can render the estimator beyond any boundary. For scatter estimators, the breakdown means either the smallest eigenvalue is 0 or the largest eigenvalue can be arbitrarily large. For M-estimators such as Tyler, its breakdown point is not greater than $1/d$ (see [30]). The BP of symmetrized M-estimators is even lower, which is understandable since one single outlier affects $n - 1$ pairwise differences.

5.3 CRITERIA

There are two different criteria used to examine efficiency and robustness: the Mean Squared Error (MSE) and Mean Log Condition Number (MLCN). We discuss these in the following subsections.

5.3 Mean Squared Error

The Mean Squared Error (MSE) criteria measures squared errors on the off diagonal elements of each estimator. For each estimator $\hat{\boldsymbol{\Sigma}}$, the mean squared errors (MSE) of off-diagonal elements are computed:

$$MSE(\hat{\boldsymbol{\Sigma}}_{ij}) = \frac{1}{5000}\sum_{m=1}^{5000}(\hat{\boldsymbol{\Sigma}}_{ij}^{(m)} - \boldsymbol{\Sigma}_{ij})^2,$$

where $\boldsymbol{\Sigma}$ is the true scatter parameter and $\hat{\boldsymbol{\Sigma}}^{(m)}$ is the estimator in the m^{th} simulated data. Since the off-diagonal elements have equal variances and are uncorrelated, the average of their MSEs is computed:

$$MSE(\hat{\boldsymbol{\Sigma}}) = \frac{2}{d(d-1)}\sum_{1\le i<j\le d} MSE(\hat{\boldsymbol{\Sigma}}_{ij})$$

Note that the MSE is affected by the Fisher correction factor of each estimator. To avoid this issue, MSE is used for shape estimators. Then, the finite sample relative efficiencies is the ratio of MSE for the regular shape matrix to that of each estimator. For sample size $n \to \infty$, the relative efficiency of each estimator should converge to the ARE in the Table 4.3.1.

5.3 Mean Log Condition Number

We use another criterion called Mean Log Condition Number (MLCN) to assess the efficiency.

Definition 5.3.1. *The* condition number *(CN) of a matrix $\boldsymbol{A}$ is the ratio of the largest eigenvalue to the smallest:*

$$cond(\boldsymbol{A}) = \frac{\lambda_1(\boldsymbol{A})}{\lambda_d(\boldsymbol{A})}.$$

To get the mean log condition number of an estimator $\hat{\mathbf{\Sigma}}$, first take the log of the condition number of $\mathbf{\Sigma}^{-1}\hat{\mathbf{\Sigma}}$, where $\mathbf{\Sigma}^{-1}$ is the true scatter matrix. Then compute the MLCN which has the following definition:

Definition 5.3.2. *For* $m = 1, ..., M$ *and* $\boldsymbol{x}_1, ..., \boldsymbol{x}_n$ *from distribution* $\epsilon(\mu, \Sigma, g)$, *we obtain the estimator* $\hat{\mathbf{\Sigma}}^{(m)}$. *Then the* mean log condition number *(MLCN) is defined as*

$$MLCN(\hat{\Sigma}) = \frac{1}{M}\sum_{m=1}^{M} LCN(\mathbf{\Sigma}^{-1}\hat{\mathbf{\Sigma}}^{(m)}).$$

For each estimator, the MLCN of $\mathbf{\Sigma}^{-1}\hat{\mathbf{\Sigma}}$ is computed. If $\hat{\mathbf{\Sigma}}$ is a good estimator, it will estimate $\mathbf{\Sigma}$ well and $\mathbf{\Sigma}^{-1}\hat{\mathbf{\Sigma}}$ will be close to the identity matrix. So, the mean log condition number should be close to 0 for good estimators. We compute the average of log condition numbers of each estimators on 5000 repetitions ($M = 5000$). The finite sample efficiency of each estimator is calculated by the ratio of the MLCN of the sample covariance to the MLCN of the estimator.

The MLCN is the criterion used to measure non-sphericity of $\mathbf{\Sigma}^{-1}\hat{\mathbf{\Sigma}}$. Rather than only checking efficiency on the off-diagonal elements of estimator as the MSE does, the MLCN is an overall criterion on the accuracy of an estimator. Since it is based on the condition number, the different Fisher correction factors among estimators does not affect the value of MLCN. However, for MSE criterion, it is important to consider shape estimators for comparison.

5.4 FINITE SAMPLE EFFICIENCY RESULTS

The first simulation studies efficiency under heavy-tailed $\mathcal{T}$, Kotz and normal distributions. 5000 samples of two different sample sizes ($n = 50$ and $n = 200$) at two different dimensions ($d = 2$ and $d = 5$) are drawn from the heavy-tailed $\mathcal{T}$ with $\nu = 5, 8$ and ∞ and Kotz distributions. Note that $\nu = \infty$ corresponds the case of normal distributions. We report the results based on the two criteria.

In Table 5.4.1, the MSE is used to compute relative efficiency in order to see sample convergence speed of RE for the Tyler, Dümbgen, Kotz and TR GCM shape estimators. The asymptotic relative efficiencies ($n = \infty$) from Table 4.3.1 are also listed in Table 5.4.1 for convenient reference.

The results of finite sample study in Table 5.4.1 show that Kotz and TR Gini estimators have a relatively fast convergence to their limiting efficiencies. Even for the case of $n = 50$, their finite sample efficiencies are already close to the asymptotic ones, especially in the normal and Kotz cases. For the Tyler estimator, the convergence is slower, and the loss in efficiency is remarkable for finite sample sizes. In the case of the $\mathcal{T}(5)$ distribution, the convergence to the limiting efficiency is much slower than that of the other cases.

In Table 5.4, the MCLN is used as assessment criterion for relative efficiency. M-estimators and symmetrized M-estimators are more efficient than other robust estimators under heavy-tailed distributions and normal distribution. MRCM has a similar relative efficiency as the Tyler estimator and has highest efficiency among all robust estimators.

The second simulation is to study robustness of estimators under contaminated of normal distributions. Two contaminations are considered, one on location shift and one on scatter.

In Table 5.4.3 under location shift contamination, TR Gini and Kotz are more efficient than others in normal distributions, but they are less efficient under the location contamination. The reason is that both of these estimators need the finite first moment assumption. The location contamination affects their performance, however, they are still more efficient than the covariance matrix. As contamination level and dimension increase, the relative efficiencies of TR Gini and Kotz decrease. Dümbgen and Tyler have a comparable relative efficiency to highly robust estimators MRCM, Mcd and Sest under a low level contamination level $\varepsilon = 0.1$. In the case of $\varepsilon = 0.2$, the relative efficiency of Dümbgen and Tyler decreases but the relative effciency of MRCM, Mcd and Sest increases.

Table 5.4.4 considers scatter contamination. The contamination points have different orientation than the normal data, also they are more scattered than the normal data. The results show that the TR Gini estimator has comparable relative efficiency to others. It even has a higher relative efficiency in some cases. For example, TR Gini is better than both the MCD and the S-estimator.

	n/d	$\mathcal{T}_d(5)$ 2	$\mathcal{T}_d(5)$ 5	$\mathcal{T}_d(8)$ 2	$\mathcal{T}_d(8)$ 5	$\mathcal{T}_d(\infty)$ 2	$\mathcal{T}_d(\infty)$ 5	$Kotz(d)$ 2	$Kotz(d)$ 5
TR Gini	50	1.28	1.37	1.12	1.16	0.98	0.99	1.15	1.09
	200	1.56	1.65	1.19	1.21	0.98	0.99	1.18	1.10
	∞	1.97	2.15	1.19	1.22	0.98	0.99	1.12	1.07
Kotz	50	1.36	1.56	1.15	1.22	0.91	0.96	1.23	1.13
	200	1.67	1.84	1.20	1.26	0.89	0.94	1.24	1.14
	∞	2.25	2.37	1.22	1.29	0.88	0.93	1.25	1.14
Dümbgen	50	1.22	1.37	1.01	1.03	0.82	0.81	1.04	0.94
	200	1.67	1.85	1.18	1.22	0.89	0.91	1.17	1.09
	∞	2.36	2.50	1.26	1.31	0.91	0.94	1.22	1.13
Tyler	50	0.77	1.13	0.61	0.81	0.45	0.58	0.71	0.75
	200	1.06	1.58	0.70	1.00	0.48	0.68	0.79	0.90
	∞	1.50	2.14	0.75	1.07	0.50	0.71	0.83	0.95

Table 5.4.1: Finite Sample RE of Shape Estimators

Finite sample relative efficiencies of the shape estimators with respective to the regular shape matrix at different distributions F_0 using MSE as the criteria.

		$\nu = 5$		$\nu = 8$		$\nu = \infty$	
		$d = 2$	$d = 5$	$d = 2$	$d = 5$	$d = 2$	$d = 5$
50	TR Gini	1.14	1.13	1.06	1.06	0.99	0.99
	Kotz	1.19	1.19	1.07	1.09	0.94	0.98
	Dumbgen	1.19	1.21	1.05	1.08	0.95	0.97
	Tyler	0.96	1.12	0.83	0.98	0.70	0.84
	MRCM	0.96	1.09	0.84	0.96	0.72	0.82
	Mcd	0.71	0.77	0.62	0.67	0.55	0.55
	S-est	0.76	1.02	0.65	0.91	0.56	0.84
	Cov	0.50	1.28	0.43	1.13	0.36	0.96
200	TR Gini	1.23	1.23	1.09	1.09	0.99	1.00
	Kotz	1.28	1.30	1.09	1.11	0.95	0.97
	Dumbgen	1.30	1.33	1.08	1.12	0.95	0.97
	Tyler	1.04	1.23	0.85	1.01	0.72	0.85
	MRCM	1.05	1.19	0.86	0.97	0.73	0.81
	Mcd	0.81	1.00	0.66	0.85	0.61	0.78
	S-est	0.88	1.16	0.72	0.99	0.60	0.88
	Cov	0.26	0.68	0.21	0.56	0.17	0.46

Table 5.4.2: MLCN of Cov and RE of Other Estimators Relative to Cov under $\mathcal{T}_\nu$-distributions

Mean of log condition numbers (MLCN) for Cov and relative efficiencies (RE) for other estimators relative to Cov under $\mathcal{T}_\nu$-distributions with $\mu = \mathbf{0}, \Sigma = \text{diag}(4, \mathbf{1}_{d-1}^T)$.

n		$\varepsilon = 0$		$\varepsilon = 0.1$		$\varepsilon = 0.2$	
		$d = 2$	$d = 5$	$d = 2$	$d = 5$	$d = 2$	$d = 5$
50	TR Gini	0.99	0.99	1.27	1.09	1.10	1.02
	Kotz	0.94	0.98	1.49	1.14	1.17	1.02
	Dümbgen	0.95	0.97	4.30	1.86	2.26	1.09
	Tyler	0.70	0.84	6.03	3.47	3.63	1.50
	MRCM	0.72	0.82	3.78	3.27	3.34	3.06
	Mcd	0.55	0.55	6.77	3.54	8.94	4.33
	Sest	0.56	0.84	6.81	4.81	8.39	5.33
	Cov	0.36	0.96	3.89	5.55	4.46	6.12
200	TR Gini	0.99	1.00	1.27	1.08	1.10	1.02
	Kotz	0.95	0.97	1.48	1.15	1.16	1.02
	Dümbgen	0.95	0.97	4.33	1.99	2.28	1.10
	Tyler	0.72	0.85	7.38	4.19	3.73	1.54
	MRCM	0.73	0.81	3.97	3.72	3.43	3.37
	Mcd	0.61	0.78	14.9	8.88	19.32	9.87
	Tyler	0.72	0.85	7.38	4.19	3.73	1.54
	Sest	0.60	0.88	13.83	9.71	17.8	10.41
	Cov	0.17	0.46	3.84	5.26	4.41	5.83

Table 5.4.3: MLCN of Cov and RE of Other Estimators Relative to Cov under the Location Contamination of Normal Distributions.

Mean of log condition numbers (MLCN) of the sample covariance matrix (Cov) and relative efficiencies (RE) of other estimators relative to Cov under $F = (1-\varepsilon)\mathcal{N}(\mathbf{0}, \Sigma_{d\times d}) + \varepsilon\mathcal{N}(10\mathbf{1}_d, \Sigma_{d\times d})$, where $\Sigma_{d\times d} = \text{diag}(4, \mathbf{1}_{d-1}^T)$.

n		$\varepsilon = 0$		$\varepsilon = 0.1$		$\varepsilon = 0.2$	
		$d = 2$	$d = 5$	$d = 2$	$d = 5$	$d = 2$	$d = 5$
50	TR Gini	0.99	0.99	1.28	1.25	1.19	1.20
	Kotz	0.94	0.98	1.40	1.37	1.30	1.32
	Dümbgen	0.95	0.97	1.46	1.47	1.34	1.40
	Tyler	0.70	0.84	1.18	1.35	1.14	1.34
	MRCM	0.72	0.82	1.19	1.34	1.13	1.34
	Mcd	0.55	0.55	0.94	0.96	0.91	1.01
	S-est	0.56	0.84	0.99	1.33	0.99	1.28
	Cov	0.36	0.96	0.61	1.54	0.59	1.54
200	TR Gini	0.99	1.00	1.33	1.33	1.22	1.22
	Kotz	0.95	0.97	1.44	1.45	1.33	1.34
	Dumbgen	0.95	0.97	1.53	1.56	1.39	1.44
	Tyler	0.72	0.85	1.24	1.44	1.17	1.37
	MRCM	0.73	0.81	1.23	1.39	1.15	1.34
	Mcd	0.61	0.78	1.07	1.33	0.96	1.25
	S-est	0.60	0.88	1.09	1.46	1.05	1.35
	Cov	0.17	0.46	0.31	0.80	0.29	0.76

Table 5.4.4: MLCN of Cov and RE of other estimators relative to Cov under the scatter contamination of normal distributions.

Mean of log condition numbers (MLCN) of the sample covariance matrix (Cov) and relative efficiencies (RE) of other estimators relative to Cov under $F = (1-\varepsilon)\mathcal{N}(\mathbf{0}, \Sigma_{d\times d}) + \varepsilon\mathcal{N}(\mathbf{0}, \Sigma^*_{d\times d})$, where $\Sigma_{d\times d} = \text{diag}(4, \mathbf{1}^T_{d-1})$, and $\Sigma^*_{d\times d} = 10 \times \text{diag}(\mathbf{1}^T_{d-1}, 4)$.

6 APPLICATIONS

In this chapter, we apply the TR Gini to Principal Components Analysis (PCA). Two real data sets are analyzed using PCA. One application uses the correlation matrix, and the other uses the covariance matrix. The results of the PCA from the TR Gini are compared to the results using the sample correlation and sample covariance matrices.

6.1 PRINCIPAL COMPONENTS ANALYSIS

The original purpose of Principal Components Analysis is to reduce a large number of interrelated variables in a data set to a smaller number of variables while still keeping as much of the variation of the original variables as possible. In order to do this, the data is transformed into a new set of variables, which are called *principal components* (PCs). These principal components are uncorrelated and ordered such that the smallest amount of components account for as much of the desired variability of the data set as possible.

The first step in PCA is to find a linear combination of $\boldsymbol{\alpha}_1^T\boldsymbol{x}$ that has maximum variance. The next step is to look for a linear function $\boldsymbol{\alpha}_2^T\boldsymbol{x}$ that is uncorrelated with $\boldsymbol{\alpha}_1^T\boldsymbol{x}$ and has maximum variance. The third component is $\boldsymbol{\alpha}_3^T\boldsymbol{x}$, which has maximum variance and is uncorrelated with $\boldsymbol{\alpha}_1^T\boldsymbol{x}$ and $\boldsymbol{\alpha}_2^T\boldsymbol{x}$ and etc. The solution of PCA turns out to be the corresponding to ordered eigenvalues of $Cov(\boldsymbol{x})$.

For elliptical distributions $\epsilon(\boldsymbol{\mu}, \boldsymbol{\Sigma}, g)$, we have $Cov(\boldsymbol{x}) = c\boldsymbol{\Sigma}$. So, any estimator of $\hat{\boldsymbol{\Sigma}}$ can be used to conduct Principal Components Analysis. Classical PCA is based on the sample covariance matrix or the sample correlation matrix. The sample covariance matrix is most efficient when data is from the normal distribution, but it is not robust because it is sensitive to outliers. Our TR Gini Covariance matrix is highly efficient under normal, but it is more robust than the sample covariance matrix. We expect the Gini based PCA to perform well for data from heavy-tail distributions.

When the variance of the variates largely differs or the variates are measured using different scales, PCA based on the correlation matrix makes more sense. In R the sample correlation matrix is used for PCA by default. In the next section, if the ratio of the largest variance to the smallest variance among variates is greater than 10, we will use the correlation based PCA.

6.2 PCA ON IRIS DATA SET

6.2 Iris Data Set and PCA results

The first data set used to conduct PCA is Anderson's Iris Data Set available within R. This data set is comprised of the following variables: sepal length, sepal width, petal length, and petal width. The measurements of each of these variables is given in centimeters. The data was collected for fifty flowers from each of the following species of iris: *Iris setosa, versicolor,* and *virginica.*

The mean and variance of each of the variables are listed in Table 6.2.1.

	Sepal Length	Sepal Width	Petal Length	Petal Width
Mean	5.843333	3.057333	3.758	1.199333
Variance	0.6856935	**0.1899794**	**3.116278**	0.5810063

Table 6.2.1: Summary Statistics of Variables in Iris Data

Looking at the variance of the variables in table 6.2.1, we can see that there is a large difference in the variance of Sepal Width and Petal Length. The ratio of the highest variance

(Petal Length - 3.116278) to the lowest variance (Sepal Width - 0.189974) is $\frac{3.116278}{0.189974} = 16.40324$. For this reason, we use the correlation matrix when doing PCA.

We see most of the variability is accounted for in the first two components in Tables 6.2.2 and 6.2.3. There is a larger amount of the proportion of variance in Component 1 when using the Regular Correlation Matrix. However, in Component 2, there is a larger amount of the proportion of variance when using the TRGC Correlation Matrix. If we look at the cumulative proportion of variance of the first two components, the TRGC Correlation Matrix does a little better than the Regular Correlation Matrix in accounting for more variation.

	Comp. 1	Comp. 2	Comp. 3	Comp. 4
Standard Deviation	1.7083611	0.9560494	0.38308860	0.143926497
Proportion of Variance	0.7296245	0.2285076	0.03668922	0.005178709
Cumulative Proportion	0.7296245	**0.9581321**	0.99482129	1.00000000

Table 6.2.2: Proportion of Variance of PCs based on Sample Correlation Matrix

	Comp. 1	Comp. 2	Comp. 3	Comp. 4
Standard Deviation	1.696378	0.9931742	0.35158130	0.110886193
Proportion of Variance	0.719425	0.2465988	0.03090235	0.003073937
Cumulative Proportion	0.719425	**0.9660237**	0.99692606	1.000000000

Table 6.2.3: Proportion of Variance of PCs based on TGRC Correlation matrix

We now plot the data projection on the first two principal components in Figures 6.2.1 and 6.2.2. Looking at these two plots, the first species Setosa (represented by □) is clearly separable from the other two species. Setosa mostly has scores around -2 on PC 1, which clearly separates this species from that of the rest. However, Versicolor (•) and Virginica (○) are not as easily separable. The scores for Versicolor on PC 1 range from a little less than 0 to a little larger than 1. As for Virginica the scores on PC 1, these range from a little less than 1 to more than 3. There are overlaps in both PC1 and PC2 for Versicolor and Virginica. Since there is not a clear linear separation between Versicolor and Virginica, we use Support Vector Machine to find the best linear separation between these two species.

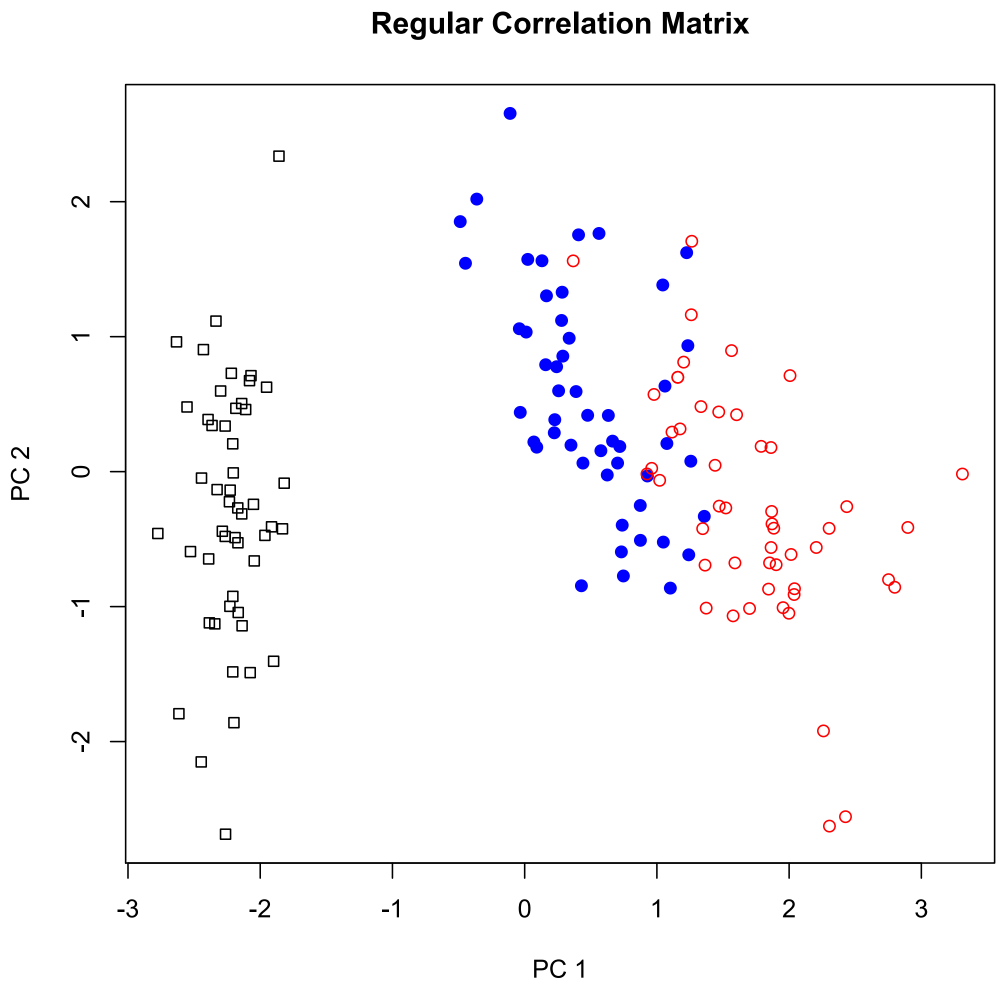


Figure 6.2.1: PCA on Regular Correlation Matrix.

This figure shows the scores of each type of Iris as projected on the first two Pricipal Components. Legend: Setosa (□), Versicolor (•), and Virginica (○).

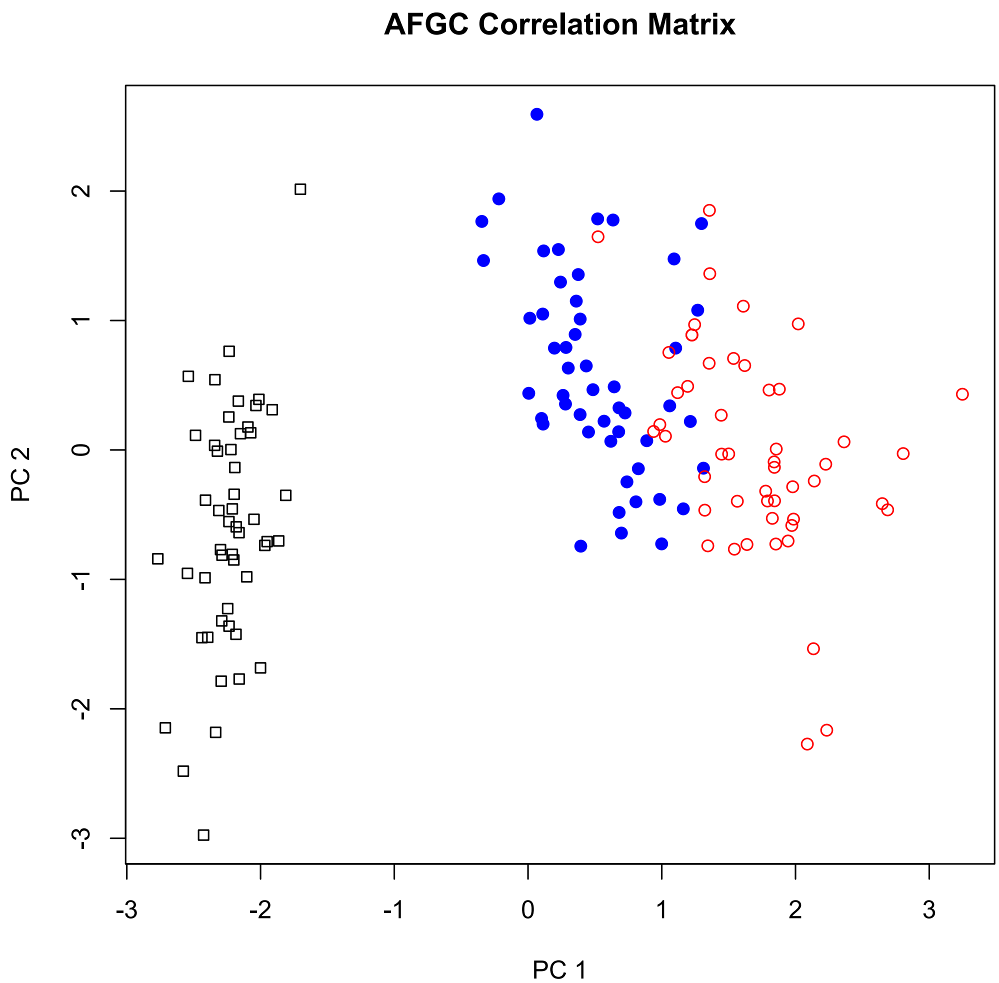


Figure 6.2.2: PCA on AFGC Correlation Matrix.

This figure shows the scores of each type of Iris as projected on the first two Pricipal Components. Legend: Setosa (□), Versicolor (•), and Virginica (∘).

6.2 Support Vector Machine Partition on the Two Major Principal Components

From PCA we saw that the first two Principal Components contained most of the variance of the data set; the cumulative proportions of variance are 0.9581321 and 0.9660237 for the sample correlation matrix and the TRGC correlation matrix, respectively. Also, we can see a clear separation in the graphs between the first species Setosa (□) and the second and third species together Versicolor (●) and Virginica (○). However, we can not see a clear separation between Versicolor (●) and Virginica (○), which makes these two species linearly inseparable. Therefore, we use Support Vector Machine to find an optimal hyperplane to separate these two species.

Support Vector Machines (SVM) is used for classification. It is basically looking to find the optimal separating hyperplane between two classes. This is done by maximizing the margin between the two classes' closest points. We call the points that fall on the boundaries *support vectors*, and the middle of the margin is called the *optimal separating hyperplane* (see [12]). The package *"kernlab"* allows us to use the function *"ksvm"* in R to get the model for the SVM. For example, the model in R is $model = ksvm(x, data, kernel, type, C, cross)$, with the following parameters:

- **x:** symbolic description of the model to be fit. We use $label\ Comp.1 + Comp.2$.
- **data:** data used, which is the scores.
- **kernel:** the kernel function used in training and predicting. We use "vanilladot," which is the linear kernel.
- **type:** what type of ksvm is use. We use the default setting "C-svc" which is for C classification.
- **C:** the cost of contraint violation. This is the "C"-constant of the regularization term in the Lagrange formulation. We use $C = 10$.

- **cross:** for an integer $k > 0$, a k-fold cross validation on training data is performed. We use $cross = 3$.

A detailed description of this function can be found in [18].

The variable "cross" allows for Cross Validation that partitions the data set into three parts. Out of these partitions, cross validation uses two parts as the training data and 1 part as the testing data. The cross-validation error output is actually the testing error. Ultimately, we want the training error and testing error to be small.

	Regular Correlation	AFGC (Corr)
Support Vectors	37	35
Training Error	0.13	0.11
Cross Validation Error	0.129531	0.12

Table 6.2.4: SVM model summary

Table 6.2.4 provides SVM model summaries on two principal components of two species. Using AFGC correlation matrix, the SVM partition of Versicolor and Virginica performs slightly better than when used with the sample correlation.

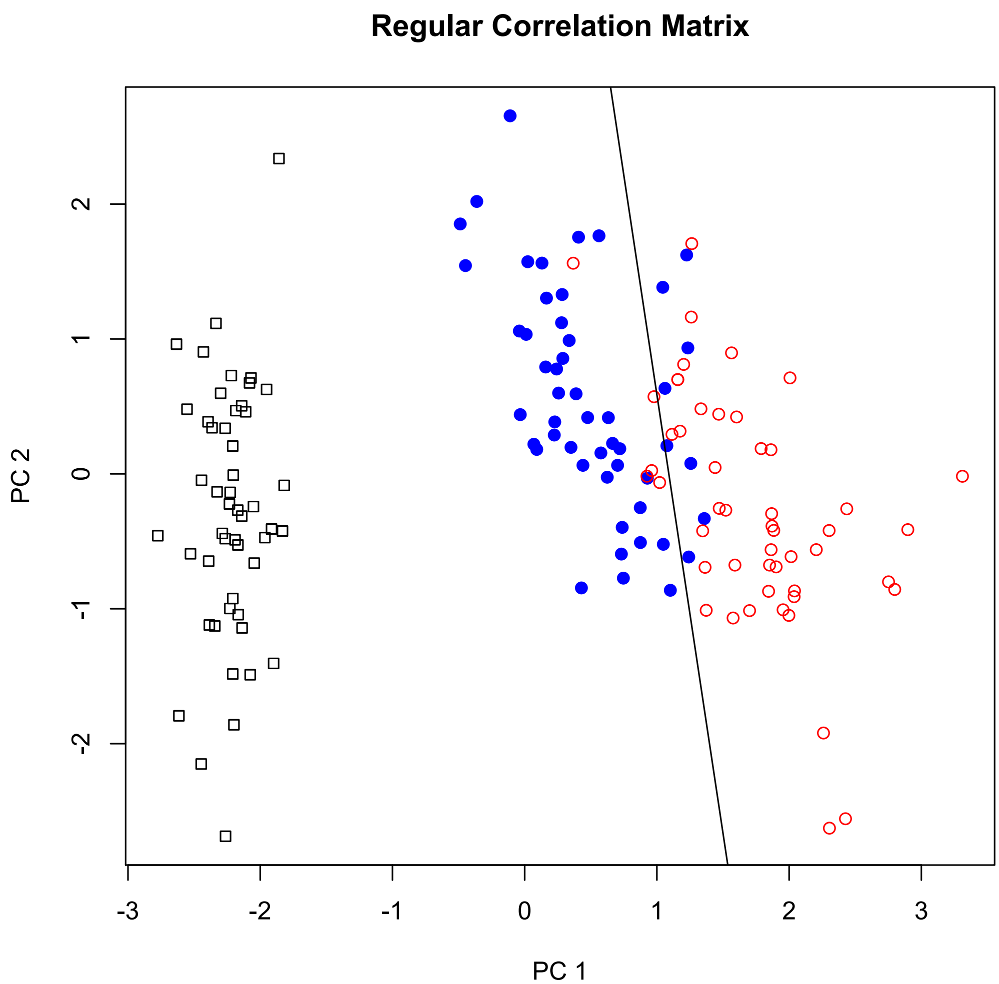


Figure 6.2.3: Support Vector Machine. Setosa (□), Versicolor (•), and Virginica (○).

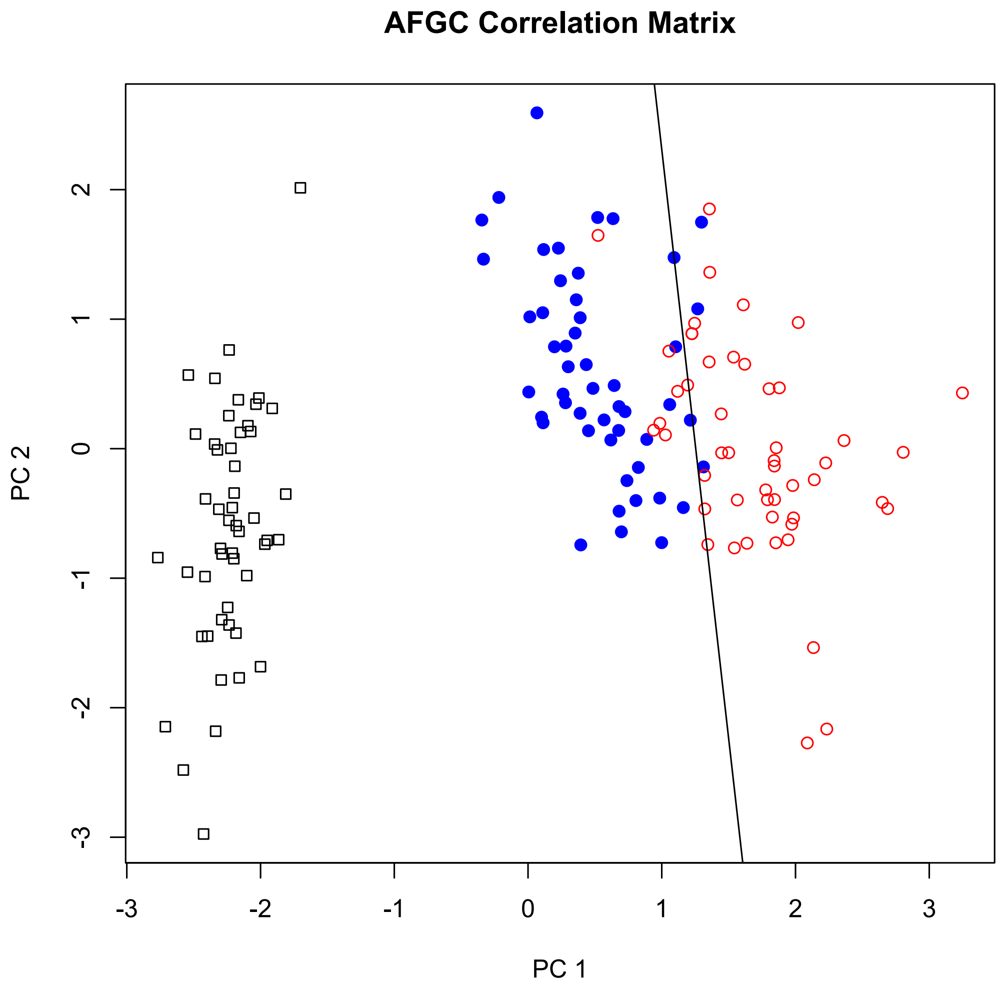


Figure 6.2.4: Support Vector Machine. Setosa (□), Versicolor (•), and Virginica (∘).

6.3 PCA ON E.COLI DATA

In this section, we perform Principal Components Analysis on the E.coli data set using the covariance matrix as the measure of scale rather than the correlation matrix. This data set comes from the UCI Machine Learning Repository. The *"princomp"* function in R does not standardize the variables when using the covariance matrix; it only centers the data. Therefore, when computing the scores for the AFGC covariance matrix, a function was written to center the data.

6.3 E.coli Data Set

The data set used to conduct Pricipal Components Analysis in this section is of E.coli data titled *Protein Localization Sites* by Kenta Nikai. There are seven attributes for this data set:

- **mcg** : McGeoch's method for signal sequence recognition
- **gvh** : von Heijne's method for signal sequence recognition
- **lip** : von Heijne's Signal Peptidase II consensus sequence score
- **chg** : presence of charge on N-terminus of predicted lipoproteins
- **aac** : score of discriminant analysis of the amino acid content of outer membrane and periplasmic proteins
- **alm1** : score of the ALOM membrane spanning region prediction program
- **alm2** : score of the ALOM program after excluding putative cleavable signal regions from the sequence

The variable **chg** contains very little information. All of the sample points have the same value of 0.5 except for one sample point which has a value of 1 for **chg**. So, we may treat this attribute as a constant, and we will delete this variable from the data set before performing PCA.

There are eight different classifications for this data set. These include the following: cytoplasm (cp), inner membrane without signal sequence (im), perisplasm (pp), inner membrane - uncleavable signal sequence (imU), outer membrane (om), outer membrane - lipoprotein (omL), inner membrane - lipoprotein (imL), and inner membrance - cleavable signal sequence (imS).

	mcg	gvh	lip	aac	alm1	alm2
Mean	0.5000595	0.5	0.4954762	0.5000298	0.5001786	0.4997321
Var	0.037882	0.021950	**0.0078314**	0.014976	**0.046549**	0.043853

Table 6.3.1: Summary Statistics of Variables in E.coli Data

Table 6.3.1 lists the mean and variance of each variable. If we look at the variance of the variables in this table, we can see that there is a little difference in the variances, but there is not a really large difference. The ratio of the highest variance (alm1 - 0.04659) to the lowest variance (lip - 0.0078314) is $\frac{0.046549}{0.0078314} = 5.94$. Therefore we can use the covariance matrix to perform PCA on the data.

6.3 PCA using E.coli Data

	Comp. 1	Comp. 2	Comp. 3	Comp. 4	Comp. 5	Comp. 6
SD	0.2990576	0.2056867	0.12078038	0.11332262	0.09192478	0.07002754
PV	0.5183870	0.2452206	0.08455458	0.07443506	0.04897896	0.02842380
CP	**0.5183870**	**0.7636076**	0.84816218	0.92259724	0.97157620	1.00000000

Table 6.3.2: Proportion of Variance of PCs based on Sample Covariance Matrix

	Comp. 1	Comp. 2	Comp. 3	Comp. 4	Comp. 5	Comp. 6
SD	1.8503083	1.2009119	0.70506355	0.64290796	0.38021465	0.28135620
PV	0.5706068	0.2403649	0.08285243	0.06888844	0.02409386	0.01319355
CP	**0.5706068**	**0.8109717**	0.89382414	0.96271258	0.98680645	1.00000000

Table 6.3.3: Proportion of Variance of PCs based on AFGC Covariance Matrix

Standard Deviation (SD), Proportion of Variance (PV), and Cumulative Proportions (CP)

We see most of the variability is accounted for in the first and second components in Tables 6.3.2 and 6.3.3. We can see that there is a higher amount of proportion of variance when using the TRGC Covariance Matrix for each component. If we look at the cumulative proportion of variance for the first two principal components, the TRGC Covariance Matrix does a little better than the Regular Covariance Matrix in accounting for the most cumulative proportion of variance.

Looking at figures 6.3.1 and 6.3.2, we can see a rough separation between three groups. These groups consist of the following:

- Group 1: cp (□)
- Group 2: im (•) and imU (△)
- Group 3: om (×), pp (⋄), and omL (+).

The classes im (•) and imU (△) seem linearly inseparable. This is due to the fact that both im and imU are part of the inner membrane. The class imL (∗) also has one point in this cluster since it is part of the inner membrane. In addition, the classes om (×) and pp (⋄) appear to be mixed together. This is due to the fact that om is the outer membrane and pp is the periplasm, which is the space between the outer membrane and inner membrane.

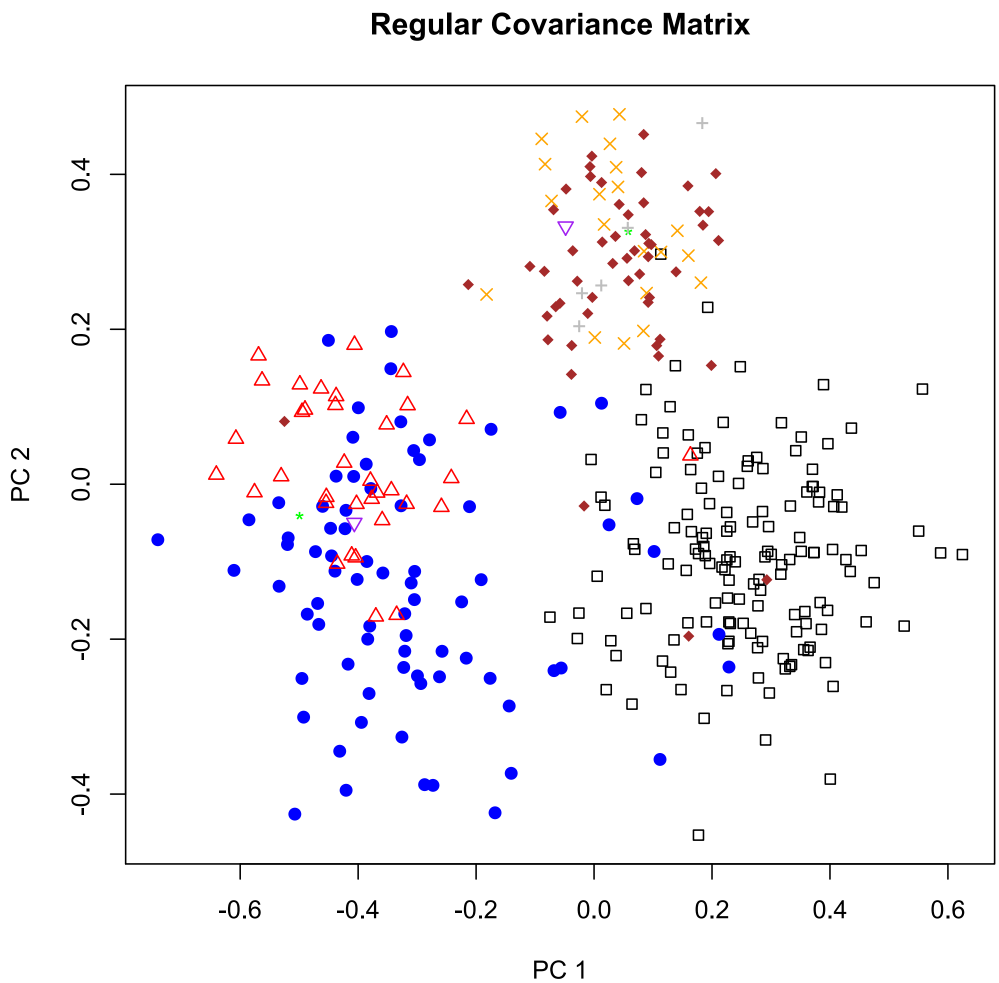


Figure 6.3.1: PCA on E.coli Data using Regular Covariance.

Cp (□), im (•), imS (▽), imL (∗), imU (△), om (×), omL (+), and pp (◇).

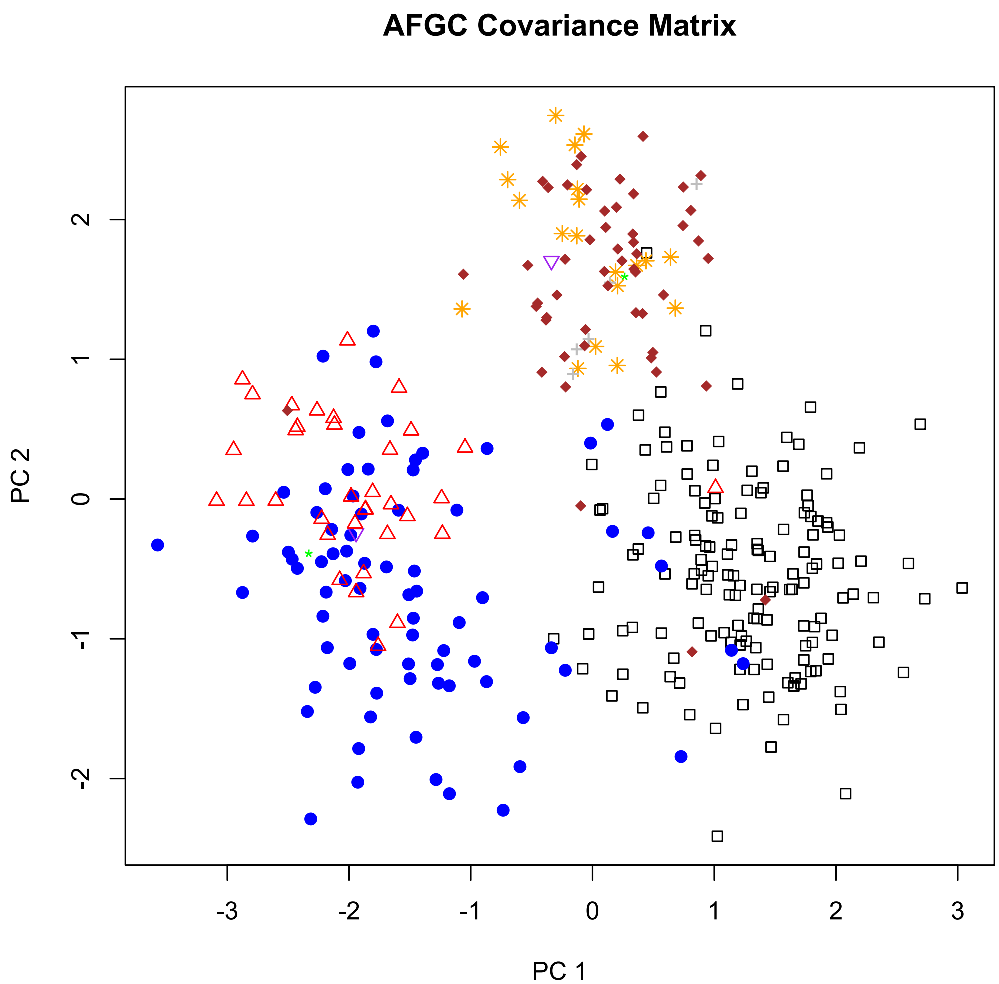


Figure 6.3.2: PCA on E.coli Data using AFGC Covariance.

Cp (□), im (•), imS (▽), imL ($*$), imU (△), om (×), omL (+), and pp (◇).

6.4 DISSCUSSION OF THE APPLICATION TO PCA

Classical PCA uses either the traditional sample correlation matrix or the sample covariance matrix. The choice of matrix depends on the data used. The sample correlation matrix may be used when the scale or variance of the variables largely differ because it standardizes the data to the same scale.

The Iris data set was analyzed using the correlation matrix. In this case the correlation matrix was chosen due to the large difference in variance of the variables. An example of the Support Vector Machine was also shown using the Iris data set. The E.coli data set was analyzed using the covariance matrix. The covariance matrix was chosen because, after the variable **chg** was deleted, the difference in variance of the variables was relatively small. From each of the tables and graphs throughout the chapter, we see that the results for our AFGC is competitive to that of the Regular Covariance and Correlation Matrices. In fact, the AFGC performs a little better than the regular one in most cases.

7 CONCLUSION

7.1 SUMMARY AND CONCLUSIONS

Using rank based statistics as robust methods is an important topic in statistics. In this dissertation, the spatial rank was first studied to extend the univariate Gini Mean Difference (GMD) to the multivariate case. The first version called the Gini Covariance Matrix (GCM) is a direct extension from the univiariate case; however, this version is not affine equivariant. Therefore, there was a need to develop a second version that is affine equivariant. The transformation-retransformation technique was used to obtain the Affine Equivariant Version of the Gini Covariance Matrix (AFGC or TR Gini). The new GCM's use a pairwise difference approach without the need of location parameter.

Their properties have been explored and their influence functions have been derived. It was found that the influence functions of GCM are approximately linear and, therefore, is unbounded. In a strict sense, they are not highly robust. However, they are highly efficient under normal distributions. They have greater than 98% asymptotic relative efficiency with respect to sample covariance matrix. On the other hand, they are more robust than the covariance matrix. The influence function of the covariance matrix is in a quadratic form. GCM will give more protection to moderate outliers than the covariance matrix. The finite sample study also showed that the TR Gini covariance estimator is highly efficient in heavy-tailed distributions. Hence the proposed affine equivariant GCM provides us an option for estimating a scatter matrix with a consideration to balance between efficiency and robustness.

The application section covered Principal Component Analysis (PCA) using the regular sample Covariance and Correlation Matrices compared with those of the TR Gini. Two different data sets were used: the Iris data set and the E.coli data set. For the Iris data, the correlation matrix is used for PCA; however, for the E.coli data, the covariance matrix is used.

BIBLIOGRAPHY

[1] E. Croux, C. Ollila and H Oja, *Sign and rank covariance matrices: statistical properties and application to principal components analysis*, Statistical Data Analysis Based on the L1-Norm and Related Methods Y. Dodge (Eds.) (2002), 257–271.

[2] P. Davies, *Asymptotic behaviour of s-estimates of multivariate location parameters and dispersion matrices*, Annals of Statistics **15** (1987), no. 3, 1269–1292.

[3] D. Donoho and P. Huber, *The notion of breakdown point*, A Festschrift for Erich L. Lehmann (1983), 157–184.

[4] R Dorfman, *A formula for the gini coefficient*, Review of Economics and Statistics **61** (1979), 146–149.

[5] L. Dümbgen, *On Tyler's M-functional of scatter in high dimension*, Annals of the Institute of Statistical Mathematics **50** (1998), no. 3, 471–491. MR 1664575 (2000g:62130)

[6] K. Fang and T. Anderson, *Statistical inference in elliptically contoured and related distributions*, Allerton Press, New York, 1990. MR 1066887 (91e:62006)

[7] C Gerstenberger and D. Vogel, *On the efficiency of gini's mean difference*, preprint, 2014.

[8] F.; Hampel, E.; Ronchetti, P.; Rousseeuw, and W. Stahel, *Robust statistics*, Wiley Series in Probability and Mathematical Statistics: Probability and Mathematical Statistics, John Wiley & Sons Inc., New York, 1986, The approach based on influence functions. MR 829458 (87k:62054)

[9] P. Huber, *The behavior of maximum likelihood estimates under nonstandard conditions*, Proceedings of Fifth Berkeley Symposium on Mathematical Statistics and Probability **1** (1967), 221–233.

[10] P. Huber and E. Ronchetti, *Robust statistics*, second ed., Wiley Series in Probability and Statistics, John Wiley & Sons Inc., Hoboken, NJ, 2009. MR 2488795 (2010j:62004)

[11] A.; Hyvärinen, J.; Karhunen, and E. Oja, *Independent component analysis*, Wiley, 2001.

[12] A.; Karatzoglou, A.; Smola and K. Hornik, *Package 'kernlab'*, Februrary 2015.

[13] V. Koltchinskii, *M-estimation, convexity and quantiles*, Annals of Statistics **25** (1997), 435–477.

[14] G. Koshevoy and K. Mosler, *Multivariate gini indices*, Journal of Multivariate Analysis **60** (1997), 252–276.

[15] S Kotz, *Multivariate distributions at a cross-road*, Statistical Distributions in Scientific Work **1** (1975).

[16] P. Mahalanobis, *On the generalised distance in statistics*, Proceedings of the National Institute of Sciences of India **2** (1936), no. 1, 49–55.

[17] R. Maronna, *Robust m-estimators of multivariate location and scatter*, Annals of Statistics **4** (1976), no. 1, 51–67.

[18] D. Meyer, *Support vector machines *: The interface to libsvm in package e1071*, 2014.

[19] S. Nadarajah, *The kotz-type distribution with applications*, Statistics **37** (2003), 341–358.

[20] H. Oja, *Descriptive statistics for multivariate distributions*, Statistics and Probability Letters **1** (1983), no. 6, 327–332.

[21] H.; Oja and R. Randles, *Multivariate nonparametric tests*, Statistical Science **19** (2004), no. 4, 598–605. MR 2185581

[22] H.; Oja, S.; Sirkiä, and J. Eriksson, *Scatter matrices and independent component analysis*, Austrian Journal of Statistics **35** (2006), 175–189.

[23] K.; Plungpongpun and Naik. D, *Multivariate analysis of variance using a kotz type distribution*, e World Congress on Engineering 2008, vol. 2, July 2008.

[24] P. Rousseeuw, *Multivariate estimation with high breakdown point*, Mathematical Statistics and Applications **B** (1985), no. 1.

[25] R. Serfling, *Approximation theorems of mathematical statistics*, Wiley, 1980.

[26] ———, *Equivariance and invariance properties of multivariate quantile and related functions, and the role of standardisation*, Journal of Nonparametric Statistics **22** (2010), no. 7, 915–936. MR 2738875

[27] R. Serfling and M. Carcea, *A gini autocovariance function for time series modeling*, 2015, doi: 10.1111/jtsa.12130.

[28] S.; Sirkiä, S.; Taskinen, and H. Oja, *Symmetrised m-estimators of scatter*, Journal of Multivariate Analysis **98** (2007), 1611–1629.

[29] S.; Taskinen, I.; Koch, and H. Oja, *Robustifying principal component analysis with spatial sign vectors*, Statistics and Probability Letters **82** (2012), 756–774.

[30] D. Tyler, *A distribution-free M-estimator of multivariate scatter*, Annals of Statistics **15** (1987), no. 1, 234–251. MR 885734 (88g:62112)

[31] S.; Visuri, V.; Koivunen, and H. Oja, *Sign and rank covariance matrices*, Journal of Statistical Planning and Inference **91** (2000), no. 2, 557–575, Prague Workshop on Perspectives in Modern Statistical Inference: Parametrics, Semi-parametrics, Non-parametrics (1998). MR 1814801 (2001m:62074)

[32] J. Wang, *A family of kurtosis orderings for multivariate distributions*, Journal of Multivariate Analysis **100** (2009), 509–517.

[33] S. Yitzhaki, *Gini's mean difference: a superior measure of variability for non-normal distribution*, Metron - International Journal of Statistics **61** (2003), no. 2, 285–316.

[34] S. Yitzhaki and E. Schechtman, *The gini methodology - a primer on a statistical methodology*, Springer, 2013.

[35] K.; Yu, X.; Dang, and Y. Chen, *Robustness of the affine equivariant scatter estimator based on the spatial rank covariance matrix*, Communication in Statistics - Theory and Methods **44** (2015), 914–932.

[36] W.; Zhou and X. Dang, *Projection based scatter depth functions and associated scatter estimators*, Journal of Multivariate Analysis **101** (2010), no. 1, 138–153.

[37] K. Zografos, *On mardias and songs measures of kurtosis is elliptical distributions*, Journal of Multivariate Analysis **99** (2008), 858–879.

[38] Y.; Zuo and H. Cui, *Depth weighted scatter estimators*, Annals of Statistics **33** (2005), no. 1, 381–413.

APPENDICES

Appendix A R Codes

```
## Add packages for the following functions ##
library(MASS)
library(ICSNP)
library(rrcov)
library(MNM)
### End Packages###

##Vector Norm
vec_norm = function(x){drop((t(x)%*%x)^(1/2))}
###
##Square Root of Vector Norm
vec_norm_sqrt = function(x){drop((t(x)%*%x)^(1/4))}
###
### Matrix Square Root Function
mat.sqrt=function(A){
  eg =eigen(A,symmetric=TRUE)
  sqrtA=eg$vectors%*%diag(sqrt(eg$values))
%*%t(eg$vectors)
  return(sqrtA)}
#####


###  GC = E[XR(X)]  Gini Covariance Matrix
GC=function(X){
  n=dim(X)[1]
  p=dim(X)[2]
  r=spatial.rank(X,shape=F)
  GC=(2/n)*t(X)%*%r
  return(GC)
}
###
### Gnc=E(ZZ^T/|Z|) where Z is centered
GnC=function(X){
n=dim(X)[1]
  p=dim(X)[2]
  Xnorm=apply(X,1,vec_norm)
```

```
  R=as.matrix(sweep(X, 1, Xnorm, "/"))
  GnC=(1/n)*t(X)%*%R
  return(GnC)
}
########

### Afine Equivariant GCM (TRGC)
TRGC.shape = function(X,init=NULL,steps=Inf,
  eps=1e-6, maxiter=100,print.it=FALSE,
  na.action=na.fail) {
 X <- na.action(X)
    if (!all(sapply(X, is.numeric)))
        stop("'X' must be numeric")
    X <- as.matrix(X)
    if (is.finite(steps))
        maxiter <- Inf
    p <- dim(X)[2]
    if (p < 2)
        stop("'X' must be at least bivariate")
    iter = 0
    if (is.numeric(init))
        V.0 <- solve(init)
    else V.0 <- solve(cov(X))
    differ = Inf
    while (TRUE) {
            if (any(iter >= steps, differ < eps))
                break
            if (iter >= maxiter) {
            stop("maxiter reached without
convergence")
            }
            sqrtV <- mat.sqrt(V.0)
            V.new <- sqrtV %*% (solve(GC(X %*%
                sqrtV))) %*% sqrtV
            V.new <- V.new/sum(diag(V.new))
            differ = norm(V.new - V.0,"F")
            V.0 <- V.new
            iter = iter + 1
     }
     if (print.it) {
        if (iter < steps)
            print(paste("convergence was reached after",
       iter, "iterations"))
     else print(paste("algorithm stopped after", steps,
"iterations"))
```

```
    }
    V.shape <- solve(V.new)
    V <- V.shape*p/sum(diag(V.shape))
    colnames(V) <- colnames(X)
    rownames(V) <- colnames(X)
    return(V)}
###########

### Generate a random variable from Kotz distribution
rkotz = function(n, d, mu=rep(0,d),
Sigma=diag(rep(1,d))){
if (length(mu)!=d || dim(Sigma)[1]!=d)
print("Warning: Dimension doesn't match")
r = rgamma(n,shape=d,scale=1)
z = matrix(rnorm(n*d),n,d,byrow=TRUE)
rz = apply(z,1,vec_norm)
y = r*z/rz   # y is from Kotz(0,I)
L = chol(Sigma)
x = L%*% t(y)+mu
return(t(x))
}

###Kotz MLE shape
Kotz.shape = function(X, location=NULL,init=NULL,
steps=Inf, eps=1e-6, maxiter=100, print.it=FALSE,
na.action=na.fail)
 {
 X <- na.action(X)
    if (!all(sapply(X, is.numeric)))
        stop("'X' must be numeric")
    X <- as.matrix(X)
    if (is.numeric(location)) {
        data.centered <- as.matrix(sweep(X, 1,
  location, "-"))
    }
    else {
        data.centered <- as.matrix(sweep(X, 1,
   colMeans(X), "-"))
    }
    if (is.finite(steps))
        maxiter <- Inf
    p <- dim(X)[2]
    if (p < 2)
        stop("'X' must be at least bivariate")
    center.ind <- apply(data.centered, 1, setequal,
```

```
              y = rep(0, p))
    n.del <- sum(center.ind)
    if (n.del != 0) {
        data.centered <- data.centered[center.ind
              == F, ]
     if (n.del > 1) {
    warning(paste(n.del, "observations equal
             to the location center were removed"))
     }
        else {
    warning("One observations equal to the location
center was removed")
        }
    }
        n <- dim(data.centered)[1]
    iter = 0
    if (is.numeric(init))
        V.0 <- solve(init)
    else V.0 <- solve(cov(X))
    differ = Inf
    while (TRUE) {
            if (any(iter >= steps, differ < eps))
                break
            if (iter >= maxiter) {
                stop("maxiter reached without
                 convergence")}
            sqrtV <- mat.sqrt(V.0)
            V.new <- sqrtV %*% (solve(GnC(data.centered
                  %*% sqrtV))) %*% sqrtV
            differ = norm(V.new - V.0,"F")
            V.0 <- V.new
            iter = iter + 1
     }
     if (print.it) {
        if (iter < steps)
            print(paste("convergence was reached after",
                iter, "iterations"))
    else print(paste("algorithm stopped after", steps,
"iterations"))
    }
    V.shape <- solve(V.new)
    V <- V.shape*p/sum(diag(V.shape))
    colnames(V) <- colnames(X)
    rownames(V) <- colnames(X)
    return(V)
```

```
}
 ###

###Principal Components Analysis###
##PCA with Correlation Matrix
pcr=princomp(x,cor="TRUE")
summary(pcr)
##Scores for Components 1 and 2
scr=pcr$score[,1:2]

###PCA for TR Gini###
Cg=TRGC.shape(x)
##PCA with TR Gini as Cov Matrix
pcgr=princomp(cor="TRUE",covmat=Cg)
summary(pcgr)

##Scores for TR Gini##
## Change Cov Matrix to Corrlation Matrix
Crg=cov2cor(Cg)
Eig2=eigen(Crg)
eval2=Eig2$values
sd2=sqrt(Eig2$values)
loadings2=Eig2$vectors ## Vector Loadings

## Function to center the data
standardize=function(x){ (x-mean(x))}
X2=apply(x,2,FUN=standardize)

## Calculate scores for Compents 1 and 2
scores2=X2%*%loadings2
scgr=scores2[,1:2]

 ### Modified mad_k #begin#
 mad.k=function(x,k){
 n=length(x)
 if(k>=n){cat("error of mad.k")}
 diff=sort(abs(x-median(x)))
 mad.k=qnorm(3/4)^(-1)*(diff[floor((n+k)/2)]
+diff[floor((n+k+1)/2)])/2
 #mad.k=(diff[floor((n+k)/2)]
             +diff[floor((n+k+1)/2)])/2
 return(mad.k)
 }
 ### Modified mad_k #end#
```

```
## Find Modified RCM #begin#
mrcm=function(x){
 n=dim(x)[1]
 p=dim(x)[2]
 k=p-1 #### maximum break down requirement
 r=spatial.rank(x,shape=F)
 rcm=1/n*t(r)%*%r
 s=eigen(rcm)$vectors
 lambda=rep(0,p)
 for (i in 1:p){
  lambda[i]=mad.k(s[,i]%*%t(x),k=k)
 }
 mrcm=s%*%diag(lambda^2)%*%t(s)
}
## Find Modified RCM #end#

## Find Regular_RCM(rrcm)##
rrcm=function(x){
  n=dim(x)[1]
  p=dim(x)[2]
  r=spatial.rank(x,shape=F)
  rrcm=1/n*t(r)%*%r
}
## Find Regular_RCM(rrcm) End##

## Begin Log conditional number (matrix) ##
log_cn=function(mtx,true_inv=invsgm){
    d=dim(mtx)[1]
    temp=true_inv%*%mtx
    eig=eigen(temp)
    log_cn=log(eig$values[1]/eig$values[d])
    return(log_cn)
}
## End  Log conditional number (matrix) ##
```

VITA

Lauren Anne Weatherall was born in Grenada, MS to Beecher and Lori Weatherall. She was valedictorian of her class at Grenada High School in 2005. Lauren received a Bachelor's of Science degree in 2008 and a Master's of Science degree in 2010 from the University of Mississippi. She is currently a Ph.D. candidate in Mathematics at The University of Mississippi.