DISCUSSION PAPERS IN STATISTICS AND ECONOMETRICS

SEMINAR OF ECONOMIC AND SOCIAL STATISTICS
UNIVERSITY OF COLOGNE

No. 5/07

Asymptotic Distributions of Robust Shape Matrices and Scales

by

Gabriel Frahm

8th version
November 19, 2008

DISKUSSIONSBEITRÄGE ZUR STATISTIK UND ÖKONOMETRIE

SEMINAR FÜR WIRTSCHAFTS- UND SOZIALSTATISTIK
UNIVERSITÄT ZU KÖLN

Albertus-Magnus-Platz, D-50923 Köln, Deutschland

This page intentionally left blank.

DISCUSSION PAPERS IN STATISTICS AND ECONOMETRICS

SEMINAR OF ECONOMIC AND SOCIAL STATISTICS
UNIVERSITY OF COLOGNE

No. 5/07

Asymptotic Distributions of Robust Shape Matrices and Scales[1]

by

Gabriel Frahm[2]

8th version
November 19, 2008

Abstract

It has been frequently observed in the literature that many multivariate statistical methods require the covariance or dispersion matrix Σ of an elliptical distribution only up to some scaling constant. If the topic of interest is not the scale but only the shape of the elliptical distribution, it is not meaningful to focus on the asymptotic distribution of an estimator for Σ or another matrix $\Gamma \propto \Sigma$. In the present work, robust estimators for the shape matrix and the associated scale are investigated. Explicit expressions for their joint asymptotic distributions are derived. It turns out that if the joint asymptotic distribution is normal, the presented estimators are asymptotically independent for one and only one specific choice of the scale function. If it is non-normal (this holds for example if the estimators for the shape matrix and scale are based on the minimum volume ellipsoid estimator) only the presented scale function leads to asymptotically uncorrelated estimators. This is a generalization of a result obtained by Paindaveine (2008) in the context of local asymptotic normality theory.

Keywords: local asymptotic normality, M-estimator, R-estimator, robust covariance matrix estimator, scale-invariant function, S-estimator, shape matrix, Tyler's M-estimator.

AMS Subject Classification: Primary 62H20, Secondary 62E20.

[1]This paper has been previously circulated under the title *Asymptotic covariance matrices of different shape matrix estimators*.

[2]Department of Economic and Social Statistics, Statistics & Econometrics, University of Cologne, Albertus-Magnus-Platz, 50923 Cologne, Germany.

This page intentionally left blank.

Asymptotic Distributions of Robust Shape Matrices and Scales

Gabriel Frahm [1]

University of Cologne
Statistics and Econometrics
Meister-Ekkehart-Str. 9
50937 Cologne
Germany

Abstract

It has been frequently observed in the literature that many multivariate statistical methods require the covariance or dispersion matrix Σ of an elliptical distribution only up to some scaling constant. If the topic of interest is not the scale but only the shape of the elliptical distribution, it is not meaningful to focus on the asymptotic distribution of an estimator for Σ or another matrix $\Gamma \propto \Sigma$. In the present work, robust estimators for the shape matrix and the associated scale are investigated. Explicit expressions for their joint asymptotic distributions are derived. It turns out that if the joint asymptotic distribution is normal, the presented estimators are asymptotically independent for one and only one specific choice of the scale function. If it is non-normal (this holds for example if the estimators for the shape matrix and scale are based on the minimum volume ellipsoid estimator) only the presented scale function leads to asymptotically uncorrelated estimators. This is a generalization of a result obtained by Paindaveine (2008) in the context of local asymptotic normality theory.

Key words: local asymptotic normality, M-estimator, R-estimator, robust covariance matrix estimator, scale-invariant function, S-estimator, shape matrix, Tyler's M-estimator.

1 Motivation

After the seminal paper by Maronna (1976), covariance matrix estimation has become a popular branch of robust statistics. Several techniques have been developed for calculating the asymptotic distributions of robust covariance matrix estimators such as the radial distribution approach of Tyler (1982) and the approach based on

[1] Email: frahm@statistik.uni-koeln.de, phone: +49 221 470-4267.

influence functions (Hampel et al., 1986). Moreover, in recent years deep insights have been gained from the viewpoint of local asymptotic normality (LAN) theory (Hallin et al., 2006, Hallin and Paindaveine, 2006a,b).

Let X be a d-dimensional random vector possessing an elliptically symmetric distribution, i.e. it can be represented by $X = \mu + \Lambda\mathcal{R}U$, where U is a k-dimensional random vector, uniformly distributed on the unit hypersphere, $\mathcal{R}$ is a nonnegative random variable being stochastically independent of U, $\mu \in \mathbb{R}^d$, and $\Lambda \in \mathbb{R}^{d\times k}$ (Cambanis et al., 1981, Fang et al., 1990, p. 42). It is assumed that $\mathcal{R}$ and U are unobservable quantities. The positive-semidefinite matrix $\Sigma := \Lambda\Lambda'$ is called the *dispersion matrix* and $\mathcal{R}$ is the *generating variate* of X. If $\mathbb{E}(\mathcal{R}^2) < \infty$, the covariance matrix of X is given by $\mathbb{V}\mathrm{ar}(X) = \mathbb{E}(\mathcal{R}^2)/k \cdot \Sigma$, whereas in case $\mathbb{E}(\mathcal{R}^2) = \infty$, the linear dependence structure of X can be further described by the dispersion matrix Σ although $\mathbb{V}\mathrm{ar}(X)$ is not defined.

In general I will assume that Σ is positive-definite, i.e. $\mathrm{r}(\Lambda) = d \leq k$. In the robust statistics literature (Tyler, 1982, Bilodeau and Brenner, 1999, Ch. 13) and in the context of LAN theory (Hallin and Paindaveine, 2006a, Paindaveine, 2008) it is often supposed that the distribution of $\mathcal{R}$ is absolutely continuous. Then the density of X can be written as $p(x) = \sqrt{\det \Sigma^{-1}}\, g\{(x-\mu)'\Sigma^{-1}(x-\mu)\}$, where the so-called *density generator* $g : \mathbb{R}^+ \to \mathbb{R}_0^+$ depends on x only through the quadratic form $(x-\mu)'\Sigma^{-1}(x-\mu)$. It can be shown (Frahm, 2004, p. 9) that the density function of $\mathcal{R}$ is given by $f(r) \propto r^{d-1}g(r^2)$.

Tatsuoka and Tyler (2000) wrote that 'The assumption of an elliptically symmetric distribution is often made simply because of its mathematical tractability'. Nevertheless, the class of elliptically symmetric distributions is a natural extension of the multivariate normal distribution. Moreover, the elliptical distribution assumption is fundamental in multivariate analysis and the results presented in this work generally require that the data are elliptically symmetric distributed. However, there is one exception where the data are only assumed to be *generalized elliptically distributed* (Frahm, 2004, Ch. 3). This will be treated in more detail below.

Note that $X = \mu + \Lambda\mathcal{R}U = \mu + V\mathcal{S}U$ with $\mathcal{S} := \mathcal{R}/\tau$, $V := \tau\Lambda$, and $\tau > 0$. That means if X possesses the dispersion matrix Σ, there always exists an equivalent representation of X with dispersion matrix $\tau^2\Sigma$ and so this can be only identified if the distribution of $\mathcal{R}$ is somehow restricted. However, many multivariate statistical methods like principal components analysis, canonical correlation analysis, linear discriminant analysis, and multivariate regression require the covariance or dispersion matrix only up to some scaling constant. This has been frequently observed in the literature (Croux and Haesbroeck, 1999, Hallin and Paindaveine, 2006a, Oja, 2003, Paindaveine, 2008, Taskinen et al., 2006). If the topic of interest is not the scale but only the *shape* of the distribution of X, it is not meaningful to focus on the asymptotic covariance matrix (ACM) of an estimator for Σ, $\mathbb{V}\mathrm{ar}(X)$ or another matrix $\Gamma \propto \Sigma$ (i.e. $\Gamma = \tau^2\Sigma$, where τ is a *constant* and thus not determined by Σ).

Therefore I will concentrate on robust estimators for the *shape matrix* of X (Oja, 2003, Paindaveine, 2008). The associated estimators for the scale are investigated concomitantly. I will derive explicit expressions for their joint asymptotic distributions. The paper is organized as follows. Section 2 introduces the notation and provides some helpful prerequisites about homogeneous functions. The question of how to choose an appropriate scale is investigated in Section 3. This section also contains the main results concerning the joint asymptotic distributions of estimators for the shape matrix and scale. In Section 4 it is shown how to calculate the asymptotic distributions of such estimators on the basis of some well-known robust covariance matrix estimators, namely M-, R-, and S-estimators.

2 Prerequisites

2.1 Notation

The following notation will be used in the sequel. The $d^2 \times d^2$ identity matrix is symbolized by I_{d^2}. Let $\mathbf{e}_{ij}$ be the $d \times d$ matrix with 1 in the ijth position and zeros elsewhere. The $d^2 \times d^2$ matrix J_{d^2} is defined as $J_{d^2} := \sum_{i=1}^{d} \mathbf{e}_{ii} \otimes \mathbf{e}_{ii}$, where '$\otimes$' denotes the Kronecker product (Schott, 1997, p. 253). The $n \times m$ matrix A' denotes the transpose of an $m \times n$ matrix A. In contrast, if f is an $\mathbb{R}$-valued function on an open subset of $\mathbb{R}$, then $f'(x)$ stands for the derivative of f at $x \in \mathbb{R}$. Further, the *commutation matrix* K_{d^2} is the $d^2 \times d^2$ matrix given by $K_{d^2} := \sum_{i,j=1}^{d} \mathbf{e}_{ij} \otimes \mathbf{e}_{ji}$ (Schott, 1997, p. 277).

For any symmetric $d \times d$ matrix A, the d^2-dimensional vector $\operatorname{vec} A$ is obtained by stacking the columns of A on top of each other, whereas $\operatorname{vech} A$ denotes the $d\,(d+1)/2$-dimensional vector obtained by stacking only the elements of the lower triangular part of A. Further, the *duplication matrix* is the $d^2 \times d\,(d+1)/2$ matrix D_d such that $D_d \operatorname{vech} A = \operatorname{vec} A$ (Schott, 1997, p. 283). Then it holds that $D_d^{+} \operatorname{vec} A = \operatorname{vech} A$, where the $d\,(d+1)/2 \times d^2$ matrix D_d^{+} is the Moore-Penrose inverse of D_d (Schott, 1997, p. 284). Let I_0 be defined as the $\{d\,(d+1)/2 - 1\} \times d\,(d+1)/2$ matrix $I_0 := [\,0 \;\; I_{d\,(d+1)/2-1}\,]$ and $N_d := I_0 D_d^{+}$, so that $\operatorname{vech}_0 A := N_d \operatorname{vec} A$ is the vech of A deprived of its first component A_{11} (Hallin and Paindaveine, 2006a).

I will frequently calculate the differential of an $\mathbb{R}^m$-valued function f, i.e. $\mathrm{d}f = \mathcal{J}_f \partial x$, where $\mathcal{J}_f := \partial f(x)/\partial x' \in \mathbb{R}^{m \times n}$ denotes the Jacobi matrix of f at $x \in \mathbb{R}^n$. Suppose that x represents the vec of a symmetric matrix. Then each off-diagonal element in the lower triangular part of that matrix represents an implicit function of the corresponding off-diagonal element in the upper triangular part and vice versa. However, I will not take the symmetry into consideration when calculating the partial derivatives of f. Otherwise, to adjust for the redundancies caused by the symmetry it would be necessary to apply the operator $(I_{d^2} + J_{d^2})/2$ on the partial differentials ∂x when calculating the total differential $\mathrm{d}f$. Hence, to avoid additional notation and tedious calculations of implicit derivatives, the Jacobi matrix $\mathcal{J}_f$ is understood to be the matrix of partial derivatives of f which are obtained by ignoring the symmetry condition. In the present context this poses no problem since $\mathcal{J}_f$ is always used only in combination with ∂x.

2.2 Homogeneous Functions

Consider a differentiable $\mathbb{R}^m$-valued function h of $x \in \mathbb{R}^n$. The function h is said to be *homogeneous* of degree $\nu \in \mathbb{R}$ if $h(\alpha x) = \alpha^\nu h(x)$ for all $x \in \mathbb{R}^n$ and $\alpha > 0$. Due to the Euler relation it holds that $\mathcal{J}_h x = \nu h(x)$. A function f is said to be *scale-invariant* if it is homogeneous of degree 0, i.e. $f(\alpha x) = f(x)$ for all $\alpha > 0$. That means $\mathcal{J}_f x = 0$ and if h is homogeneous of degree 1, it holds that $\mathcal{J}_h x = h(x)$. In the following a homogeneous function is always understood to be homogeneous of degree 1. Note that the partial derivatives of any homogeneous function are scale-invariant.

Let $\mathcal{P}^d$ be the set of all symmetric positive-definite $d \times d$ matrices and $\varphi\colon \mathcal{P}^d \to \mathbb{R}^k$ a scale-invariant function, i.e. $\varphi(\alpha\Gamma) = \varphi(\Gamma)$ for all $\alpha > 0$ and $\Gamma \in \mathcal{P}^d$. Especially, consider a scale-invariant function $\Omega(\Gamma) = \Gamma/\sigma^2(\Gamma)$, where $\sigma^2 : \mathcal{P}^d \to \mathbb{R}^+$ is an homogeneous function, i.e. $\sigma^2(\alpha\Gamma) = \alpha\sigma^2(\Gamma) > 0$. It is supposed that the so-called *scale function* σ^2 is differentiable at any point $\Gamma \in \mathcal{P}^d$ and also that $\sigma^2(I_d) = 1$. Then $\sigma^2(\Gamma)$ is called the *scale* of Γ. The matrix $\Omega(\Gamma)$ will be called the *shape matrix* (with respect to the scale function σ^2) belonging to Γ. I will write $\sigma^2 \equiv \sigma^2(\Gamma)$ and $\Omega \equiv \Omega(\Gamma)$ whenever these quantities cannot be confounded with the corresponding functions.

Note that $\sigma^2(\Omega) = 1$ and $\varphi \circ \Omega = \varphi$, since $\varphi\{\Omega(\Gamma)\} = \varphi\{\Gamma/\sigma^2(\Gamma)\} = \varphi(\Gamma)$. For instance, the correlation matrix produced by Γ is scale-invariant and thus it can be derived from any shape matrix Ω. Hence, whenever Ω_n is an estimator for Ω, an estimator for $\varphi(\Gamma)$ is simply given by $\varphi(\Omega_n)$. This is a formal justification of directing one's attention to shape matrices (Frahm and Jaekel, 2007a, Hallin and Paindaveine, 2006a, Oja, 2003, Paindaveine, 2008, Taskinen et al., 2006). General robustness and efficiency properties of scale-invariant functions have been investigated by Tyler (1983).

3 Asymptotic Distributions

3.1 The Choice of the Scale Function

In most cases asymptotic normality of robust estimators μ_n and Γ_n for the mean vector and covariance matrix can be guaranteed by the usual regularity conditions given in the robust statistics literature. Typically μ_n and Γ_n are also asymptotically independent. In the present work it is shown that the asymptotic independence of an estimator Ω_n for the shape matrix and an associated estimator σ_n^2 for the scale can only be guaranteed for one and only one scale function σ^2. A similar result in the context of LAN theory has been obtained by Paindaveine (2008) (see below).

Let Γ_n be some estimator for $\Gamma \propto \Sigma$ where n represents the sample size. The corresponding shape matrix estimator is given by $\Omega_n := \Gamma_n/\sigma^2(\Gamma_n)$. At a first glance the choice of the scale function σ^2 might be considered as arbitrary and the following variants can be often observed in the literature (Paindaveine, 2008):

(**S1**) Frahm (2004, p. 64), Hallin et al. (2006), Hallin and Paindaveine (2006b), Hettmansperger and Randles (2002) as well as Randles (2000) simply choose $\sigma^2(\Gamma) = \Gamma_{11}$ so that $\Omega_{11} = 1$.

(**S2**) Dümbgen (1998), Frahm and Jaekel (2007b) as well as Tyler (1987a) take the scale function $\sigma^2(\Gamma) = (\operatorname{tr}\Gamma)/d$ so that $\operatorname{tr}\Omega = d$.

(**S3**) Dümbgen and Tyler (2005), Hallin and Paindaveine (2008a,b), Paindaveine (2008), Salibian-Barrera et al. (2006), Taskinen et al. (2006) as well as Tatsuoka and Tyler (2000) postulate $\sigma^2(\Gamma) = (\det\Gamma)^{1/d}$ so that $\det\Omega = 1$.

Paindaveine (2008) considers the latter normalization as *canonical* since this is the only one where the Fisher information matrix with respect to the mean vector, shape matrix and scale is block diagonal if the distribution of X or, more precisely, the corresponding experiment is LAN (van der Vaart, 1998, Ch. 7).

The scale functions defined by **S2** and **S3** correspond to the arithmetic and geometric means of the eigenvalues of Γ, respectively. Hence, another possible scale function is given by the harmonic mean of the eigenvalues of Γ, i.e.

(**S4**) $\sigma^2(\Gamma) = d/(\operatorname{tr}\Gamma^{-1})$ so that $\operatorname{tr}\Omega^{-1} = d$.

It is worth to point out that shape matrices are not affine equivariant, since

$$\Omega(V\Gamma V') = \frac{V\Gamma V'}{\sigma^2(V\Gamma V')} = \frac{\sigma^2(\Gamma)}{\sigma^2(V\Gamma V')} \cdot V\Omega(\Gamma)V'$$

for any nonsingular $d \times d$ matrix V and generally $\sigma^2(\Gamma)$ does not correspond to $\sigma^2(V\Gamma V')$. This is not surprising because even after an affine-linear transformation of the data, the shape matrix has to satisfy the scaling condition $\sigma^2(\Omega) = 1$ and so the equality $\Omega(V\Gamma V') = V\Omega(\Gamma)V'$ cannot be guaranteed in general. However, a natural requirement is that the equivariance property holds at least for all transformations V with $\sigma^2(VV') = 1$. That means if not the scale but only the shape of the distribution of X is affected by V, the shape matrix should remain equivariant.

More generally, it can be required (Tyler, 2002) that

$$\Omega(V\Gamma V') = \frac{V\Omega(\Gamma)V'}{\sigma^2(VV')},$$

i.e. $\sigma^2(V\Gamma V') = \sigma^2(VV')\,\sigma^2(\Gamma)$. Interestingly, from the scale functions considered in **S1–S4** only the canonical one (**S3**) satisfies this kind of affine equivariance property. This is another argument in favor of the determinant-based normalization proposed by Paindaveine (2008).

The previous arguments as well as a thorough discussion in Hallin and Paindaveine (2006a) show that the choice of the scale function must be driven by statistical considerations and should be handled carefully.

Lemma 1 *Let $\Omega(\Gamma) = \Gamma/\sigma^2(\Gamma)$ be a $d \times d$ shape matrix and σ^2 a scale function. Then*

$$\mathcal{J}_\Omega := \frac{\partial \operatorname{vec}\Omega(\Gamma)}{\partial(\operatorname{vec}\Gamma)'} = \frac{1}{\sigma^2}\left\{I_{d^2} - \operatorname{vec}\Omega\,\mathcal{J}_{\sigma^2}\right\},$$

where

$$\mathcal{J}_{\sigma^2} := \frac{\partial\sigma^2(\Gamma)}{\partial(\operatorname{vec}\Gamma)'} = \frac{\partial\sigma^2(\Omega)}{\partial(\operatorname{vec}\Omega)'}.$$

Proof. By the product rule it follows that

$$\mathcal{J}_\Omega = \frac{1}{\sigma^2}\cdot\frac{\partial\operatorname{vec}\Gamma}{\partial(\operatorname{vec}\Gamma)'} - \frac{\operatorname{vec}\Gamma}{\sigma^4}\cdot\mathcal{J}_{\sigma^2} = \frac{1}{\sigma^2}\left\{I_{d^2} - \operatorname{vec}\Omega\,\mathcal{J}_{\sigma^2}\right\}.$$

Since the partial derivatives of an homogeneous function are scale-invariant, it holds that $\mathcal{J}_{\sigma^2} = \partial\sigma^2(\Omega)/\partial(\operatorname{vec}\Omega)'$. ■

In the following I will write $\Psi := I_{d^2} - \operatorname{vec}\Omega\,\mathcal{J}_{\sigma^2}$ for notational convenience.

3.2 Main Results

Let $\mathcal{Q}$ be a symmetric random $d \times d$ matrix. A symmetric random $d \times d$ matrix $\mathcal{M}$ is said to possess a *radial distribution* if $\mathcal{O}\mathcal{M}\mathcal{O}' \sim \mathcal{M}$ for any orthogonal $d \times d$ matrix $\mathcal{O}$ (Tyler, 1982). In the following let $\mathcal{N}$ be a symmetric random $d \times d$ matrix with finite second moments. It is supposed that $\mathcal{N}$ is of the radial type with respect to a symmetric positive-definite $d \times d$ matrix Γ. That means $T\mathcal{N}T'$ has a radial distribution whenever the $d \times d$ matrix T is such that $T'T = \Gamma^{-1}$. Further, let (Γ_n) be a sequence of symmetric positive-definite random $d \times d$ matrices and (σ_n^2) an associated sequence with $\sigma_n^2 := \sigma^2(\Gamma_n)$, where σ^2 is a scale function. Moreover, consider the sequence (Ω_n) of symmetric positive-definite random $d \times d$ matrices with $\Omega_n := \Gamma_n / \sigma_n^2$.

Theorem 1 *Let σ^2 be a scale function and $\Omega \equiv \Omega(\Gamma) = \Gamma/\sigma^2(\Gamma)$ the shape matrix belonging to Γ. Further, let (a_n) be a sequence of real numbers increasing to infinity such that $a_n(\operatorname{vec}\Gamma_n - \operatorname{vec}\Gamma) \to_{\mathrm{d}} \operatorname{vec}\mathcal{Q}$ as $n \to \infty$ with $\mathbb{E}(\operatorname{vec}\mathcal{Q}) = 0$ and*

$$\mathbb{V}\mathrm{ar}(\operatorname{vec}\mathcal{Q}) = \gamma_1(I_{d^2} + K_{d^2})(\Gamma \otimes \Gamma) + \gamma_2(\operatorname{vec}\Gamma)(\operatorname{vec}\Gamma)', \tag{1}$$

where $\gamma_1 \geq 0$ and $\gamma_2 \geq -2\gamma_1/d$. Then it follows that

$$a_n\left(\begin{bmatrix}\sigma_n^2 \\ \operatorname{vec}\Omega_n\end{bmatrix} - \begin{bmatrix}\sigma^2 \\ \operatorname{vec}\Omega\end{bmatrix}\right) \xrightarrow{\mathrm{d}} \xi, \qquad n \longrightarrow \infty,$$

where $\sigma^2 \equiv \sigma^2(\Gamma)$, ξ is a (d^2+1)-dimensional random vector with $\mathbb{E}(\xi) = 0$, and

$$\mathbb{V}\mathrm{ar}(\xi) = \begin{bmatrix}\mathcal{V}(\sigma_n^2) & \mathcal{V}(\sigma_n^2, \Omega_n) \\ \mathcal{V}(\sigma_n^2, \Omega_n)' & \mathcal{V}(\Omega_n)\end{bmatrix}.$$

More specifically,

$$\mathcal{V}(\sigma_n^2) = \sigma^4\left\{2\gamma_1 \mathcal{J}_{\sigma^2}(\Omega \otimes \Omega)\mathcal{J}_{\sigma^2}' + \gamma_2\right\}$$

with $\mathcal{J}_{\sigma^2} = \partial\sigma^2(\Omega)/\partial(\operatorname{vec}\Omega)'$ and $\sigma^4 = \{\sigma^2(\Gamma)\}^2$,

$$\mathcal{V}(\sigma_n^2, \Omega_n)' = 2\gamma_1\sigma^2\,\Psi(\Omega \otimes \Omega)\mathcal{J}_{\sigma^2}',$$

with $\Psi = I_{d^2} - \operatorname{vec}\Omega\,\mathcal{J}_{\sigma^2}$, and

$$\mathcal{V}(\Omega_n) = \gamma_1\Psi(I_{d^2} + K_{d^2})(\Omega \otimes \Omega)\Psi'.$$

Proof. The vector $\{\sigma^2(\Gamma), \operatorname{vec}\Omega(\Gamma)\}$ is differentiable at $\operatorname{vec}\Gamma$ and thus

$$a_n\left(\begin{bmatrix}\sigma_n^2 \\ \operatorname{vec}\Omega_n\end{bmatrix} - \begin{bmatrix}\sigma^2 \\ \operatorname{vec}\Omega\end{bmatrix}\right) \xrightarrow{\mathrm{d}} \xi := \mathcal{J}_{\sigma^2,\Omega}\operatorname{vec}\mathcal{Q}, \qquad n \longrightarrow \infty,$$

where $\mathcal{J}_{\sigma^2,\Omega}$ is defined as $\partial\{\sigma^2(\Gamma), \operatorname{vec}\Omega(\Gamma)\}/\partial(\operatorname{vec}\Gamma)'$. From $\mathbb{E}(\operatorname{vec}\mathcal{Q}) = 0$ it follows that $\mathbb{E}(\xi) = 0$ and the variance of the first element of ξ is given by $\mathcal{V}(\sigma_n^2) = \mathcal{J}_{\sigma^2}\mathbb{V}\mathrm{ar}(\operatorname{vec}\mathcal{Q})\mathcal{J}_{\sigma^2}'$. Since σ^2 is a homogeneous function it holds that $\mathcal{J}_{\sigma^2}\operatorname{vec}\Gamma = \sigma^2$. Note also that $\mathcal{J}_{\sigma^2}(I_{d^2} + K_{d^2}) = 2\mathcal{J}_{\sigma^2}$ and thus

$$\mathcal{V}(\sigma_n^2) = 2\gamma_1\mathcal{J}_{\sigma^2}(\Gamma \otimes \Gamma)\mathcal{J}_{\sigma^2}' + \gamma_2\sigma^4 = \sigma^4\left\{2\gamma_1\mathcal{J}_{\sigma^2}(\Omega \otimes \Omega)\mathcal{J}_{\sigma^2}' + \gamma_2\right\}.$$

Similarly, the covariances between the first element of ξ and its residual elements are given by $\mathcal{V}(\sigma_n^2, \Omega_n) = \mathcal{J}_{\sigma^2}\mathbb{V}\mathrm{ar}(\mathrm{vec}\,\mathcal{Q})\Psi'/\sigma^2$. Since Ω is a scale-invariant function of Γ, due to Euler's relation it holds that $(\mathrm{vec}\,\Gamma)'\Psi' = 0$ and thus

$$\mathcal{V}(\sigma_n^2, \Omega_n) = \gamma_1 \mathcal{J}_{\sigma^2}(I_{d^2} + K_{d^2})(\Gamma \otimes \Gamma)\Psi'/\sigma^2 = 2\gamma_1\sigma^2\mathcal{J}_{\sigma^2}(\Omega \otimes \Omega)\Psi'. \quad (2)$$

The expression for the variances and covariances of the residual elements of ξ, i.e. $\mathcal{V}(\Omega_n)$ follows by a straightforward application of the arguments given above. ■

The next proposition ensures that the preceding theorem is applicable to any case where Γ_n represents an affine equivariant covariance matrix estimator and the data stem from an elliptically symmetric distribution.

Proposition 1 *Let σ^2 be a scale function and $\Omega \equiv \Omega(\Gamma) = \Gamma/\sigma^2(\Gamma)$ the shape matrix belonging to Γ. Further, let (a_n) be a sequence of real numbers increasing to infinity such that $a_n(\mathrm{vec}\,\Gamma_n - \mathrm{vec}\,\Gamma) \to_{\mathrm{d}} \mathrm{vec}\,\mathcal{N}$ as $n \to \infty$. Here $\mathbb{E}(\mathrm{vec}\,\mathcal{N}) = 0$ and $\mathcal{N}$ is of the radial type with respect to the matrix Γ. Then the conditions of Theorem 1 are satisfied.*

Proof. It is only necessary to show that the second moment condition (1) is satisfied. Since $\mathcal{N}$ is of the radial type, this follows immediately from Corollary 1 of Tyler (1982). ■

In the following Γ_n can be interpreted as a covariance matrix estimator. Due to the central limit theorem, in most practical situations it can be found that $a_n = \sqrt{n}$ and the random vector $\mathrm{vec}\,\mathcal{N}$ is multivariate normally distributed. A well-known exception is the *minimum volume ellipsoid* (MVE) estimator (Rousseeuw, 1985). This is only $\sqrt[3]{n}$-consistent and its asymptotic distribution is non-normal (Davies, 1992). Nonetheless, whenever Γ_n is affine equivariant and the data stem from an elliptically symmetric distribution, the limiting random matrix $\mathcal{N}$ is of the radial type (Tyler, 1982). Hence, Proposition 1 is applicable to a wide range of covariance matrix estimators.

An important consequence of Theorem 1 is that the asymptotic distribution of Ω_n is only driven by the number γ_1. That means γ_2 has no impact on the asymptotic distribution of Ω_n. Hence, the asymptotic relative efficiency of some shape matrix estimator Ω_{1n} compared to another shape matrix estimator Ω_{2n} (i.e. both estimators are based on the *same* scale function σ^2 but different covariance matrix estimators) can be simply calculated by the ratio γ_{12}/γ_{11}, where γ_{11} is the γ_1 of Ω_{1n} and γ_{12} is the γ_1 of Ω_{2n} (Tyler, 1983).

Corollary 1 *Suppose that the conditions of Theorem 1 are satisfied and σ^2 corresponds to the scale function given by* ***S3***. *Then it holds that*

$$\mathcal{V}(\sigma_n^2) = \sigma^4\left(\frac{2\gamma_1}{d} + \gamma_2\right), \qquad \mathcal{V}(\sigma_n^2, \Omega_n)' = 0,$$

and

$$\mathcal{V}(\Omega_n) = \gamma_1 \left(I_{d^2} + K_{d^2}\right)(\Omega \otimes \Omega) - \frac{2\gamma_1}{d} \cdot (\mathrm{vec}\,\Omega)(\mathrm{vec}\,\Omega)'\,. \tag{3}$$

In particular, if $\mathrm{vec}\,\mathcal{Q}$ *is multivariate normally distributed, the quantities* σ_n^2 *and* Ω_n *are asymptotically independent.*

Proof. Note that

$$\mathcal{J}_{\sigma^2} = \frac{\sigma^2}{d \det \Gamma} \cdot \frac{\partial \det \Gamma}{\partial (\mathrm{vec}\,\Gamma)'} = \frac{\sigma^2}{d} \cdot (\mathrm{vec}\,\Gamma^{-1})' = (\mathrm{vec}\,\Omega^{-1})'/d\,.$$

Due to Theorem 1 the asymptotic variance $\mathcal{V}(\sigma_n^2)$ is given by

$$\mathcal{V}(\sigma_n^2) = \sigma^4 \left\{2\gamma_1 \mathcal{J}_{\sigma^2}(\Omega \otimes \Omega)\mathcal{J}_{\sigma^2}' + \gamma_2\right\}$$

and note that $(\Omega \otimes \Omega)\mathcal{J}_{\sigma^2}' = \mathrm{vec}\,\Omega/d$. Moreover, $\mathcal{J}_{\sigma^2}\mathrm{vec}\,\Omega = 1$, which means that $\mathcal{V}(\sigma_n^2) = \sigma^4(2\gamma_1/d + \gamma_2)$. Further,

$$\mathcal{V}(\sigma_n^2, \Omega_n)' = 2\gamma_1\sigma^2\Psi(\Omega \otimes \Omega)\mathcal{J}_{\sigma^2}' = 2\gamma_1\sigma^2\, \Psi\mathrm{vec}\,\Omega/d\,.$$

Due to Euler's relation it holds that $\Psi\mathrm{vec}\,\Omega = 0$ and thus $\mathcal{V}(\sigma_n^2, \Omega_n)' = 0$. That means σ_n^2 and Ω_n are asymptotically uncorrelated or even independent if $\mathrm{vec}\,\mathcal{Q}$ is multivariate normally distributed. Finally, the expression for $\mathcal{V}(\Omega_n)$ follows by a straightforward calculation after noting that $\mathcal{J}_{\sigma^2}(\Omega \otimes \Omega)\mathcal{J}_{\sigma^2}' = 1/d$. ■

Theorem 2 *Suppose that the conditions of Theorem 1 are satisfied with* $\gamma_1 > 0$. *Then the scale function given by* ***S3*** *is the only one where* σ_n^2 *and* Ω_n *are asymptotically uncorrelated.*

Proof. Paindaveine (2008) shows that the determinant-based scale function given by **S3** is the only one where the Fisher information $\mathcal{I}_{\sigma^2,\Omega}$ is a block diagonal matrix if the considered family of elliptically symmetric distributions is LAN. Suppose that the data are multivariate normally distributed. Then Theorem 1 applies to the sample covariance matrix and it is clear that the family of multivariate normal distributions is LAN. The Fisher information is the inverse of the ACM of σ_n^2 and Ω_n (which can be obtained after re-shaping Ω_n to avoid singularity (Hallin and Paindaveine, 2006a,b)). Hence, there is no other scale function such that (2) vanishes. Since the latter is only an algebraic statement, the same must hold for any other distribution under the conditions of Theorem 2. ■

Theorem 2 extends the main result of Paindaveine (2008) which has been obtained in the context of LAN theory. Similarly, it can be shown that the canonical scale function is the only one which admits the simple representation of the ACM of a shape matrix estimator given by Eq. 3. In fact, this ACM exhibits the same desirable form as the ACM of any affine equivariant covariance matrix estimator according to Theorem 2 and Eq. 1. The operators Ψ and $\mathcal{J}_{\sigma^2}$ corresponding to the remaining scale functions defined by **S1**, **S2**, and **S4** are now given for convenience without an explicit derivation.

ad S1. $\mathcal{J}_{\sigma^2} = \mathbf{e}_1'$, where $\mathbf{e}_1$ is the $d^2 \times 1$ vector with 1 in the first position and zeros elsewhere, so that $\Psi = I_{d^2} - \operatorname{vec}\Omega\, \mathbf{e}_1'$.

ad S2. $\mathcal{J}_{\sigma^2} = (\operatorname{vec} I_d)'/d$ and thus $\Psi = I_{d^2} - (\operatorname{vec}\Omega)(\operatorname{vec} I_d)'/d$ (see also Theorem 5 in Sirkiä et al., 2007).

ad S4. It can be shown that $\mathcal{J}_{\sigma^2} = d/(\operatorname{tr}\Gamma^{-1})^2 \cdot (\operatorname{vec}\Gamma^{-2})' = (\operatorname{vec}\Omega^{-2})'/d$, where $\Gamma^{-2} := \Gamma^{-1}\Gamma^{-1}$ and $\Omega^{-2} := \Omega^{-1}\Omega^{-1}$, i.e. $\Psi = I_{d^2} - (\operatorname{vec}\Omega)(\operatorname{vec}\Omega^{-2})'/d$.

If a shape matrix estimator Ω_{1n} defined via a scale function σ_1^2 is *re-normalized* by applying some other scale function σ_2^2 to Ω_{1n}, its ACM simply corresponds to

$$\mathcal{V}(\Omega_{2n}) = \gamma_1 \Psi_2 (I_{d^2} + K_{d^2})(\Omega_2 \otimes \Omega_2)\Psi_2', \tag{4}$$

where $\Psi_2 = I_{d^2} - \operatorname{vec}\Omega_2\, \mathcal{J}_{\sigma_2^2}$ and Ω_2 is the shape matrix belonging to Γ with respect to the scale function σ_2^2. That means the first normalization has no impact on the asymptotic distribution of Ω_{2n}.

4 Robust Covariance Matrix Estimation

In the following I will present some well-known robust covariance matrix estimators (i.e. M-, R-, and S-estimators) which satisfy the aforementioned conditions and calculate the joint asymptotic distributions of the corresponding estimators for the shape matrix and scale. It is neither possible nor reasonable to study here all existing robust covariance matrix estimators (for some contemporary overviews see, e.g., Zuo, 2006, Maronna et al., 2006, Ch. 6), but the essential concept might become clear from the subsequent discussion.

Let Γ_n be an affine equivariant estimator which is consistent for Γ. Due to the general result of Tyler (1982), in most practical situations Γ_n is asymptotically normally distributed with ACM $\mathcal{V}(\Gamma_n) = \gamma_1(I_{d^2} + K_{d^2})(\Gamma \otimes \Gamma) + \gamma_2(\operatorname{vec}\Gamma)(\operatorname{vec}\Gamma)'$, where $\gamma_1 \geq 0$ and $\gamma_2 \geq -2\gamma_1/d$ usually depend on the generating variate $\mathcal{R}$. In the following I will only present the numbers γ_1 and γ_2. The $\sqrt{n}$-convergence to the normal law is implicitly assumed. Hence, Theorem 2 implies that the canonical scale function is the only one where the estimators for the shape matrix and scale are asymptotically independent. As a counterexample consider the MVE-estimator. This is not $\sqrt{n}$-consistent and asymptotically normally distributed (Davies, 1992). However, since the MVE-estimator is affine equivariant and the rate of convergence does not matter, the corresponding MVE-estimators for the shape matrix and scale remain asymptotically uncorrelated (under the elliptical distribution assumption).

Throughout this section it is supposed that the unknown location vector $\mu \in \mathbb{R}^d$

can be substituted by some $\sqrt{n}$-consistent estimate (here, too, it has been already demonstrated by Rousseeuw (1985) that the MVE-estimator for the location is only $\sqrt[3]{n}$-consistent and its asymptotic distribution is non-normal). In most cases – under mild regularity conditions concerning the distribution of X (see, e.g., Hallin and Paindaveine, 2006b, Tyler, 1987a, Bilodeau and Brenner, 1999, Ch. 13) – it can be shown that the resulting covariance matrix estimator is asymptotically normally distributed possessing an ACM of that form which is required in Theorem 1. Hence, in the following $X_1, \ldots, X_n$ will represent *centered* i.i.d. random vectors for the sake of simplicity and without loss of generality.

4.1 M-Estimation

An *M-estimator* for Γ (Maronna, 1976) is defined as a solution of

$$\Gamma_n = \frac{1}{n} \sum_{t=1}^{n} w\Big(X_t'\Gamma_n^{-1}X_t\Big) X_t X_t' ,$$

where $w\colon \mathbb{R}^+ \to \mathbb{R}_0^+$ satisfies a set of general conditions (Maronna, 1976, Bilodeau and Brenner, 1999, Section 13.4.1). The estimator Γ_n is strongly consistent for the matrix $\Gamma = \mathbb{E}\{w(X'\Gamma^{-1}X)XX'\}$ which is related to the dispersion matrix of X by $\Gamma = \tau^2\Sigma$, where $\tau > 0$ is such that $\mathbb{E}\{\psi(\mathcal{R}^2/\tau^2)\} = d$ with $\psi(t) := tw(t)$. The numbers γ_1 and γ_2 can be calculated by $\gamma_1 = (d+2)^2\psi_1/(d+2\psi_2)^2$ and

$$\gamma_2 = \frac{(\psi_1 - 1) - 2\,(\psi_2 - 1)\,\psi_1\{d + (d+4)\,\psi_2\}/(d+2\psi_2)^2}{\psi_2^2} ,$$

where $\psi_1 := \mathbb{E}\{\psi^2(\mathcal{R}^2/\tau^2)\}/\{d\,(d+2)\}$ and $\psi_2 := \mathbb{E}\{\psi'(\mathcal{R}^2/\tau^2)\,\mathcal{R}^2\}/(d\tau^2)$ (Tyler, 1982, Bilodeau and Brenner, 1999, p. 223).

If X possesses a continuous elliptical distribution and Σ_n is the corresponding *ML-estimator* for the dispersion matrix Σ, it holds that $\gamma_1 = \{d\,(d+2)/4\}/\mathbb{E}\{h^2(\mathcal{R}^2)\}$ and $\gamma_2 = -2\gamma_1\,(1-\gamma_1)/\{2 + d\,(1-\gamma_1)\}$, where $h(t) := t\,\partial \log g(t)/\partial t$. If $X \sim \mathcal{N}_d(0,\Sigma)$ and Σ_n represents the sample covariance matrix, it holds that $\gamma_1 = 1$ and $\gamma_2 = 0$. Otherwise the sample covariance matrix is an M-estimator where $\psi(t) = t$. That means $\mathbb{E}(\mathcal{R}^2/\tau^2) = \mathbb{E}\{\psi(\mathcal{R}^2/\tau^2)\} = d$, $\psi_1 = d/(d+2) \cdot \mathbb{E}(\mathcal{R}^4)/\mathbb{E}^2(\mathcal{R}^2)$, and $\psi_2 = 1$ so that $\gamma_1 = \psi_1$ and $\gamma_2 = \gamma_1 - 1$ if $\mathcal{R}$ has a finite fourth moment.

Now special attention is devoted to Tyler's M-estimator (Tyler, 1983, 1987a)

$$T_n = \frac{d}{n} \sum_{t=1}^{n} \frac{X_t X_t'}{X_t' T_n^{-1} X_t} = \frac{d}{n} \sum_{t=1}^{n} \frac{S_t S_t'}{S_t' T_n^{-1} S_t} , \tag{5}$$

where $S_t := X_t/\|X_t\|$, $\|\cdot\|$ denotes the Euclidean norm, and it is only supposed that $\mathbb{P}(\mathcal{R} > 0) = 1$. Note that T_n is not affected by the realizations of the generating variate $\mathcal{R}$, since $S = X/\|X\| = \mathcal{R}\Lambda U/\|\mathcal{R}\Lambda U\| = \Lambda U/\|\Lambda U\|$ (a.s.).

That means Tyler's M-estimator is *distribution-free* in the context of elliptically symmetric distributions. This has been already observed by Tyler (1987b). Frahm and Jaekel (2007a,b) pointed out that the distribution-free property even holds within the class of generalized elliptical distributions. A random vector is said to be generalized elliptically distributed if its generating variate $\mathcal{R}$ can be negative and might depend on U (Frahm, 2004, p. 46). This feature allows for the modeling of various kinds of asymmetries (Kring et al., 2007, Frahm, 2004, Section 3.4). For instance it can be shown that any *skew-elliptical distribution* (Liu and Dey, 2004) belongs to the class of generalized elliptical distributions (Frahm, 2004, p. 47).

Tyler's M-estimator (5) is unique up to a scaling constant. Hence, in fact T_n is a genuine *shape matrix* estimator since it can be only calculated with some suitable scale function σ^2 such that $\sigma^2(T_n) = 1$. Originally, Tyler (1987a,b) applied the trace-based scale function given by **S2**, whereas in Tatsuoka and Tyler (2000) the authors prefer to use the canonical normalization **S3**. For the purpose of calculating the asymptotic distribution, Tyler (1987a,b) focuses on $\overline{T}_n := d/(\operatorname{tr}\Sigma^{-1}T_n) \cdot T_n$, that means he defines the scale of T_n via Σ by $\sigma^2(T_n) = \operatorname{tr}\Sigma^{-1}T_n/d$. This leads to $\sigma^2(\overline{T}_n) = \sigma^2(\Sigma) = 1$ for any positive-definite $d \times d$ matrix Σ.

Note that in contrast to some normalization according to **S1**–**S4**, the shape matrix estimator $\overline{T}_n$ indeed is affine equivariant and consequently its ACM (Tyler, 1987b) exhibits the simple structure suggested by Eq. 1, viz

$$\mathcal{V}(\overline{T}_n) = \frac{d+2}{d} \cdot (I_{d^2} + K_{d^2})(\Sigma \otimes \Sigma) - \frac{2\,(d+2)}{d^2} \cdot (\operatorname{vec}\Sigma)(\operatorname{vec}\Sigma)'\,. \tag{6}$$

Since Σ represents a shape matrix with respect to Tyler's scale function, this ACM in fact corresponds to the ACM given by Eq. 3 with $\gamma_1 = (d+2)/d$. Furthermore, the Jacobian of Tyler's scale function is given by $\mathcal{J}_{\sigma^2} = (\operatorname{vec}\Sigma^{-1})'/d$ and this actually corresponds to the Jacobian of the *canonical* scale function (see the proof of Corollary 1). That means by using Tyler's scale function in association with some other affine equivariant covariance matrix estimator, the corresponding estimators for the shape matrix and scale become asymptotically uncorrelated. This seems to contradict Theorem 2. However, note that Tyler's σ^2 in general does not meet the natural requirement $\sigma^2(I_d) = 1$ and unfortunately $\overline{T}_n$ cannot be applied in practical situations, since σ^2 is determined by the unknown parameter Σ.

An alternative way for obtaining the desired ACM of Tyler's M-estimator is given as follows. Note that T_n is simply an M-estimator with $\psi(t) = d$. That means $\psi_1 = d/(d+2)$ and $\psi_2 = 0$ so that $\gamma_1 = (d+2)/d$ and γ_2 is not defined (since σ^2 cannot be estimated by T_n). Hence, due to Theorem 1, the ACM of T_n generally corresponds to $\mathcal{V}(T_n) = (d+2)/d \cdot \Psi(I_{d^2} + K_{d^2})(\Omega \otimes \Omega)\Psi'$. Moreover, due to Corollary 1 the ACM of Tyler's M-estimator, based on the *canonical* scale function, corresponds to (6) where Σ has to be substituted by Ω.

4.2 R-Estimation

The R-estimator for the shape matrix has been introduced by Hallin et al. (2006). Consider Tyler's M-estimator T_n which is normalized according to **S1**, i.e. the upper left element corresponds to 1. The R-estimator is based on a *discretized version* of T_n. Suppose that x is an element of T_n. Then the discretization can be made by $x^{\#} := \operatorname{sgn} x/\sqrt{n}\,\lceil \sqrt{n}\,|x|\rceil$ (Hallin et al., 2006), where $\lceil y\rceil$ denotes the smallest integer not smaller than $y \in \mathbb{R}$. The corresponding discretized version of Tyler's M-estimator is denoted by $T_n^{\#}$. Hallin and Paindaveine (2006b) also define $U_t := (T_n^{\#})^{-1/2}X_t/\|(T_n^{\#})^{-1/2}X_t\|$. Here $A^{-1/2}$ denotes a positive-definite $d \times d$ matrix such that $A^{-1/2}A^{-1/2\,\prime} = A^{-1}$, where A^{-1} is the inverse of a symmetric positive-definite $d \times d$ matrix A. Further, R_t represents the rank of $\|(T_n^{\#})^{-1/2}X_t\|$ with respect to the sample $X_1, \ldots, X_n$.

Let $f_{\mathcal{S}}\colon \mathbb{R}^+ \to \mathbb{R}_0^+$ be the density function of some imaginary generating variate $\mathcal{S}$, whereas $f_{\mathcal{R}}$ refers to the true generating variate $\mathcal{R}$. Consider the cumulative distribution function $F_{\mathcal{S}}(x) = \int_0^x f_{\mathcal{S}}(r)\,dr$ and $F_{\mathcal{R}}$ respectively. Here both $\mathcal{R}$ and $\mathcal{S}$ are absolutely continuous and satisfy some weak regularity conditions which guarantee local asymptotic normality (Hallin and Paindaveine, 2006b). As already mentioned before, the density function of $\mathcal{S}$ is given by $f_{\mathcal{S}}(r) \propto r^{d-1}g_{\mathcal{S}}(r^2)$, where $g_{\mathcal{S}}$ is the density generator of $\mathcal{S}$. However, in the following consider the function $f_{\mathcal{S}}^*(r) := r^{-(d-1)}f_{\mathcal{S}}(r) = g_{\mathcal{S}}(r^2)$ and for $0 < p < 1$ define $K_{\mathcal{S}}(p) := \psi_{\mathcal{S}}\{F_{\mathcal{S}}^{-1}(p)\}\,F_{\mathcal{S}}^{-1}(p)$, where $F_{\mathcal{S}}^{-1}$ is the quantile function of $\mathcal{S}$ and $\psi_{\mathcal{S}}(x) := -f_{\mathcal{S}}^{*\prime}(x)/f_{\mathcal{S}}^*(x)$. Now, the so-called *cross-information coefficient* (Hallin et al., 2006) is given by

$$\mathcal{I}_{\mathcal{R},\mathcal{S}} := \int_0^1 K_{\mathcal{R}}(p)\,K_{\mathcal{S}}(p)\,dp\,. \tag{7}$$

Also define

$$\Delta_n := M_d\left(T_n^{\#} \otimes T_n^{\#}\right)^{-1/2} \sum_{t=1}^{n}\left\{K_{\mathcal{S}}\left(\frac{R_t}{n+1}\right)\operatorname{vec}\left(U_tU_t'\right) - \frac{\overline{K}_{\mathcal{S}}}{d}\cdot \operatorname{vec} I_d\right\}$$

with $\overline{K}_{\mathcal{S}} := 1/n\sum_{t=1}^{n} K_{\mathcal{S}}(t/(n+1))$. The $\{d\,(d+1)/2 - 1\} \times d^2$ matrix M_d symbolizes the Moore-Penrose inverse of N_d' (where N_d is such that $N_d\mathrm{vec}\,A = \mathrm{vech}_0 A$). Further, let $\Psi_n := I_{d^2} - \operatorname{vec} T_n^{\#}\,\mathrm{e}_1'$ and $Q_n := N_d\Psi_n(I_{d^2} + K_{d^2})(T_n^{\#} \otimes T_n^{\#})\Psi_n' N_d'$. Now the R-estimator Ω_n is defined in terms of the vech_0 operator, viz

$$\mathrm{vech}_0\Omega_n = \mathrm{vech}_0 T_n^{\#} + \frac{d\,(d+2)}{2n}\cdot \widehat{\mathcal{I}}_{\mathcal{R},\mathcal{S},n}^{-1}\,Q_n\Delta_n\,,$$

where $\widehat{\mathcal{I}}_{\mathcal{R},\mathcal{S},n}$ represents some consistent estimator for the cross-information coefficient (7) (Hallin et al., 2006). The upper left element of Ω_n is set to 1.

Thereafter, following the arguments of Hallin and Paindaveine (2006a) and Paindaveine (2008), one can apply a re-normalization by using the canonical scale function and the ACM of the resulting R-estimator readily follows by applying Eq. 4 with $\gamma_1 = d\,(d+2)\,\mathcal{I}_{\mathcal{S},\mathcal{S}}/\mathcal{I}^2_{\mathcal{S},\mathcal{R}}$. Especially, if $\mathcal{S} \sim \mathcal{R}$ it holds that $\gamma_1 = d\,(d+2)/\mathcal{I}_{\mathcal{R},\mathcal{R}}$ with $\mathcal{I}_{\mathcal{R},\mathcal{R}} = \int_0^1 K^2_{\mathcal{R}}(p)\,dp = \mathbb{E}(\psi^2_{\mathcal{R}}(\mathcal{R})\,\mathcal{R}^2)$. From $\psi_{\mathcal{R}}(r)\,r = -2r^2 g'(r^2)/g(r^2)$ it follows that $\psi^2_{\mathcal{R}}(r)\,r^2 = 4h^2(r^2)$, where h has been already defined in Section 4.1. Recall that the function h is used for calculating the ACM of an ML-estimator. That means if $\mathcal{S} \sim \mathcal{R}$, the R-estimator has the same limiting distribution as the corresponding ML-estimator and thus it becomes asymptotically efficient.

4.3 S-Estimation

The S-estimator for the dispersion matrix (Davies, 1987) can be defined as $\Gamma_n = \arg\min_{\Upsilon \in \mathcal{P}^d} \det \Upsilon$ subject to

$$\frac{1}{n} \sum_{t=1}^{n} \rho\Big(\sqrt{X_t' \Upsilon^{-1} X_t}\Big) = \alpha\rho(\infty)\,,$$

where $0 < \alpha < 1$ and $\rho\colon \mathbb{R}^+ \to \mathbb{R}^+_0$ has to be bounded, increasing, and sufficiently smooth (Croux and Haesbroeck, 1999, Tyler, 2002, Bilodeau and Brenner, 1999, Section 13.4.2). The chosen constraint guarantees that Γ_n is consistent for $\Gamma = \tau^2\Sigma$, where $\tau > 0$ is such that $\mathbb{E}\{\rho(\mathcal{R}/\tau)\} = \alpha\rho(\infty)$.

Let ψ be the first and ψ' the second derivative of ρ. It is assumed that

$$\mathbb{E}\{\psi'(\mathcal{R}/\tau)\} > 0 \qquad \text{and} \qquad \mathbb{E}\{\psi'(\mathcal{R}/\tau)\,\mathcal{R}^2/\tau + (d+1)\,\psi(\mathcal{R}/\tau)\,\mathcal{R}\} > 0\,.$$

Then the numbers γ_1 and γ_2 are given by

$$\gamma_1 = \frac{d\,(d+2)\,\mathbb{E}\{\psi^2(\mathcal{R}/\tau)\,\mathcal{R}^2\}}{\mathbb{E}^2\{\psi'(\mathcal{R}/\tau)\,\mathcal{R}^2/\tau + (d+1)\,\psi(\mathcal{R}/\tau)\,\mathcal{R}\}}$$

and

$$\gamma_2 = \frac{4\tau^2\mathbb{V}\mathrm{ar}\{\rho(\mathcal{R}/\tau)\}}{\mathbb{E}^2\{\psi(\mathcal{R}/\tau)\,\mathcal{R}\}} - \frac{2\gamma_1}{d}$$

(Davies, 1987, Lopuhaä, 1989, Bilodeau and Brenner, 1999, p. 225).

Acknowledgements

The author would like to thank Karl Mosler for his valuable comments as well as two anonymous referees for their important suggestions and valuable contributions.

References

Bilodeau, M., Brenner, D., 1999. Theory of Multivariate Statistics. Springer.

Cambanis, S., Huang, S., Simons, G., 1981. On the theory of elliptically contoured distributions. Journal of Multivariate Analysis 11, 368–385.

Croux, C., Haesbroeck, G., 1999. Influence function and efficiency of the minimum covariance determinant scatter matrix estimator. Journal of Multivariate Analysis 71, 161–190.

Davies, P., 1987. Asymptotic behaviour of S-estimates of multivariate location parameters and dispersion matrices. Annals of Statistics 15, 1269–1292.

Davies, P., 1992. The asymptotics of Rousseeuw's minimum volume ellipsoid estimator. Annals of Statistics 20, 1828–1843.

Dümbgen, L., 1998. On Tyler's M-functional of scatter in high dimension. Annals of the Institute of Statistical Mathematics 50, 471–491.

Dümbgen, L., Tyler, D., 2005. On the breakdown properties of some multivariate M-functionals. Scandinavian Journal of Statistics 32, 247–264.

Fang, K., Kotz, S., Ng, K., 1990. Symmetric Multivariate and Related Distributions. Chapman & Hall.

Frahm, G., 2004. Generalized Elliptical Distributions: Theory and Applications. Ph.D. thesis, University of Cologne, Department of Economic and Social Statistics, Germany.

Frahm, G., Jaekel, U., 2007a. Distribution-free shape matrix estimation for incomplete data. Discussion paper, University of Cologne, Department of Economic and Social Statistics, Germany.

Frahm, G., Jaekel, U., 2007b. Tyler's M-estimator, random matrix theory, and generalized elliptical distributions with applications to finance. Discussion paper, University of Cologne, Department of Economic and Social Statistics, Germany.

Hallin, M., Oja, H., Paindaveine, D., 2006. Semiparametrically efficient rank-based inference for shape. II. Optimal R-estimation of shape. Annals of Statistics 34, 2757–2789.

Hallin, M., Paindaveine, D., 2006a. Parametric and semiparametric inference for shape: the role of the scale functional. Statistics and Decisions 24, 327–350.

Hallin, M., Paindaveine, D., 2006b. Semiparametrically efficient rank-based inference for shape. I. Optimal rank-based tests for sphericity. Annals of Statistics 34, 2707–2756.

Hallin, M., Paindaveine, D., 2008a. Optimal rank-based tests for homogeneity of scatter. Annals of Statistics 36, 1261–1298.

Hallin, M., Paindaveine, D., 2008b. Optimal tests for homogeneity of covariance, scale, and shape. Journal of Multivariate Analysis, in press (DOI: 10.1016/j.jmva.2008.05.010).

Hampel, F., Ronchetti, E., Rousseeuw, P., Stahel, W., 1986. Robust Statistics. John Wiley.

Hettmansperger, T., Randles, R., 2002. A practical affine equivariant multivariate median. Biometrika 89, 851–860.

Kring, S., Rachev, S., Höchstötter, M., Fabozzi, F., 2007. Multi-tail elliptical distributions. Preprint, University of Karlsruhe, Department of Econometrics, Statistics and Mathematical Finance, Germany.

Liu, J., Dey, D., 2004. Skew-elliptical distributions. In: Genton, M. (Ed.), Skew-Elliptical Distributions and Their Applications: A Journey Beyond Normality. Chapman & Hall, Ch. 3.

Lopuhaä, H., 1989. On the relation between S-estimators and M-estimators of multivariate location and covariance. Annals of Statistics 17, 1662–1683.

Maronna, R., 1976. Robust M-estimators of multivariate location and scatter. Annals of Statistics 4, 51–67.

Maronna, R., Martin, D., Yohai, V., 2006. Robust Statistics. John Wiley.

Oja, H., 2003. Multivariate M-estimates of location and shape. In: Höglund, R., Jäntti, M., Rosenqvist, G. (Eds.), Statistics, Econometrics and Society. Essays in Honor of Leif Nordberg. Statistics Finland.

Paindaveine, D., 2008. A canonical definition of shape. Statistics and Probability Letters 78, 2240–2247.

Randles, R., 2000. A simpler, affine-invariant, multivariate, distribution-free sign test. Journal of the American Statistical Association 95, 1263–1268.

Rousseeuw, P., 1985. Multivariate estimation with high breakdown point. In: Grossmann, W., Pflug, G., Vincze, I., Wertz, W. (Eds.), Mathematical Statistics and Applications. Reidel, Dordrecht, pp. 283–297.

Salibian-Barrera, M., van Aelst, S., Willems, G., 2006. Principal components analysis based on multivariate MM-estimators with fast and robust bootstrap. Journal of the American Statistical Association 101, 1198–1211.

Schott, J., 1997. Matrix Analysis for Statistics. John Wiley.

Sirkiä, S., Taskinen, S., Oja, H., 2007. Symmetrized M-estimators of multivariate scatter. Journal of Multivariate Analysis 98, 1611–1629.

Taskinen, S., Croux, C., Kankainen, A., Ollila, E., Oja, H., 2006. Influence functions and efficiencies of the canonical correlation and vector estimates based on scatter and shape matrices. Journal of Multivariate Analysis 97, 1219–1243.

Tatsuoka, K., Tyler, D., 2000. On the uniqueness of S-functionals and M-functionals under nonelliptical distributions. Annals of Statistics 28, 1219–1243.

Tyler, D., 1982. Radial estimates and the test for sphericity. Biometrika 69, 429–436.

Tyler, D., 1983. Robustness and efficiency properties of scatter matrices. Biometrika 70, 411–420.

Tyler, D., 1987a. A distribution-free M-estimator of multivariate scatter. Annals of Statistics 15, 234–251.

Tyler, D., 1987b. Statistical analysis for the angular central Gaussian distribution on the sphere. Biometrika 74, 579–589.

Tyler, D., 2002. High breakdown point multivariate M-estimation. Estadística 54, 213–247.

van der Vaart, A., 1998. Asymptotic Statistics. Cambridge University Press.

Zuo, Y., 2006. Robust location and scatter estimators in multivariate analysis. In: Fan, J., Koul, H. (Eds.), Frontiers of Statistics (in honor of Professor P.J. Bickel's 65th birthday). Imperial College, pp. 467–490.