Symmetrised M-estimators of multivariate scatter

Seija Sirkiä[*], Sara Taskinen[*] and Hannu Oja[+]

[*] *Dept. of Mathematics and Statistics, University of Jyväskylä*
[+] *Tampere School of Public Health, University of Tampere*

August 25, 2006

Abstract

In this paper we introduce a family of symmetrised M-estimators of multivariate scatter. These are defined to be M-estimators only computed on pairwise differences of the observed multivariate data. Symmetrised Huber's M-estimator and Dümbgen's estimator serve as our examples. The influence functions of the symmetrised M-functionals are derived and the limiting distributions of the estimators are discussed in the multivariate elliptical case to consider the robustness and efficiency properties of estimators. The symmetrised M-estimators have the important independence property; they can therefore be used to find the independent components in the independent component analysis (ICA).

Keywords: Efficiency, elliptical distribution, influence function, M-estimator, robustness, scatter matrix.

1 Introduction

A fundamental problem in multivariate analysis is to develop robust affine equivariant alternatives to the sample mean vector and sample covariance

[*]Corresponding author: Seija Sirkiä, Department of Mathematics and Statistics, P.O. Box 35, University of Jyväskylä, FIN-40014 Jyväskylä, Finland; tel: +358-14-260 2980, fax: +358-14-2602981, email: ssirkia@maths.jyu.fi

matrix. The sample mean vector and sample covariance matrix are the maximum likelihood estimates of the symmetry center (location parameter) $\boldsymbol{\mu}$ and the covariance matrix (scatter parameter) Σ in the multivariate normal model. The multivariate normal distribution is a member in a larger family of elliptically symmetric distributions. A k-variate random vector $\boldsymbol{x}$ is elliptically symmetric with location vector $\boldsymbol{\mu}$ and symmetric scatter matrix $\Sigma > 0$ if its density function is, for some function ρ, of the form

$$f(\boldsymbol{x}) = |\Sigma|^{-1/2} \exp\{-\rho(||\Sigma^{-1/2}\boldsymbol{x} - \boldsymbol{\mu}||\}. \tag{1}$$

(Throughout the paper $C^{1/2}$ means a symmetric square root of a positive definite symmetric matrix C.) The distributions in this elliptical family are denoted by $\mathcal{E}(\mu, \Sigma, \rho)$. Scatter matrix Σ is proportional to the covariance matrix (if it exists) and it determines the shape of its concentric elliptical contours. Note that if $\boldsymbol{x}$ is a random variable having an elliptical distribution $\mathcal{E}(\mu, \Sigma, \rho)$, then the standardized variable $\boldsymbol{z} = \Sigma^{-1/2}(\boldsymbol{x} - \boldsymbol{\mu})$ has a spherical distribution with symmetry center $\mathbf{0}$, and $\boldsymbol{z}$ can be decomposed as $\boldsymbol{z} = r\boldsymbol{u}$, where $r = ||\boldsymbol{z}||$ and $\boldsymbol{u} = ||\boldsymbol{z}||^{-1}\boldsymbol{z}$ are independent with $\boldsymbol{u}$ being uniformly distributed on the unit sphere.

Assume first that $\boldsymbol{x}_1, ..., \boldsymbol{x}_n$ is a random sample from an elliptical distribution $\mathcal{E}(\mu, \Sigma, \rho)$. In this paper, we are interested in the scatter matrix estimation only, and we therefore assume that the location center is known. Without loss of generality, we assume that $\boldsymbol{\mu} = \mathbf{0}$. For any $k \times k$ matrix $C > 0$, write $\boldsymbol{z}_i(C) = C^{-1/2}\boldsymbol{x}_i,\ r_i(C) = ||\boldsymbol{z}_i(C)||$ and $\boldsymbol{u}_i(C) = ||\boldsymbol{z}_i(C)||^{-1}\boldsymbol{z}_i(C)$, $i = 1, ..., n$. Then the maximum likelihood (ML) estimator minimizes the objective function

$$\frac{1}{n}\sum_{i=1}^{n} \rho(r_i(C)) + \frac{1}{2}\log|C| \tag{2}$$

or solves the estimating equation

$$\frac{1}{n}\sum_{i=1}^{n} w(r_i(C))\boldsymbol{u}_i(C)\boldsymbol{u}_i^T(C) = I_k \tag{3}$$

where $w(r) = \rho'(r)r$. Huber [6] proved the consistency and asymptotic normality of the estimators (2) and (3) under weaker conditions: the observations were no more assumed to come from the specific elliptical target population $\mathcal{E}(\mu, \Sigma, \rho)$. Huber [7] later called this estimator the maximum likelihood type estimator, or M-estimator, either based on the criterion function (2) or on the estimating equation (3). Maronna [13], Huber [7] and Kent and Tyler [9], for example, considered the existence and uniqueness of the estimate.

Maronna [13] defined a more general class of M-estimators for an elliptical population with the estimating equation

$$\frac{1}{n}\sum_{i=1}^{n} w_1(r_i(C))\boldsymbol{u}_i(C)\boldsymbol{u}_i^T(C) = \frac{1}{n}\sum_{i=1}^{n} w_2(r_i(C))I_k \tag{4}$$

Maronna [13] and Huber [7] proved the existence and uniqueness of solutions under some general assumptions on weight functions w_1 and w_2 and the observed sample. Maronna [13] proved the consistency and asymptotic normality utilizing Huber's [6] results. The influence functions and upper limit for breakdown point were also derived. Later, Tyler [20] studied the breakdown properties of M-estimators in detail.

Tyler [21] considered a limiting form of Huber's type M-estimator. His estimator $\widehat{C}$ solves

$$\frac{k}{n}\sum_{i=1}^{n} \boldsymbol{u}_i(C)\boldsymbol{u}_i^T(C) = I_k.$$

This corresponds to choosing $w_1(r) = k$ and $w_2(r) = 1$ in Maronna's definition. It is remarkable that, in the elliptical model $\mathcal{E}(\mu, \Sigma, \rho)$, the finite sample distribution (and the limiting distribution) of $\hat{C}$ does not depend on ρ at all. The estimator is then the most robust estimator among the set of consistent and asymptotically normal estimators in the sense that it minimizes maximum asymptotic variance over the elliptical model. For these results, see Tyler [21].

In this paper we introduce a family of so called symmetrised M-estimators of scatter which are defined to be M-estimators computed on pairwise differences of the observed data. A special case, the symmetrised version of Tyler's M-estimator, has been earlier proposed by Dümbgen [2]. In the univariate case taking the pairwise differences is a well known operation; it makes the distribution symmetric, with location at 0. In the multivariate case taking pairwise differences makes all univariate projections symmetric with location at 0, thus it is not necessary to impose any arbitrary definition of the location in a situation with a non-symmetric distribution. Pairwise differences are also useful for other reasons. Estimators of scatter in the elliptically symmetric family usually require the location either be known or estimated simultaneously. With the estimators at hand, this is no longer needed as the location center of the pairwise differences is always the origin.

Maybe the most interesting point is that a symmetrised M-estimator of scatter, or indeed any symmetrised scatter matrix, has the so called independence property: the scatter functional is a diagonal matrix if the the components of the random vector are independent. The covariance matrix naturally has this property but for example regular M-estimators do not. In the literature, this property has not received much attention so far.

The independence property is highly important, for example, in independent component analysis or ICA. See Hyvärinen et al [8]. Briefly, the ICA problem consists of finding an original random vector, or source, $\boldsymbol{s}$ with independent components when only an unknown linear mixture $\boldsymbol{x} = A\boldsymbol{s}$ is observed. Previously proposed solutions to the ICA problem are usually based on an idea justified by the central limit theorem that linear mixtures of non-normal random variables are closer to the normal distribution than any of the original ones. The solution to ICA is then found as a solution to the optimization problem concerning some measure of non-gaussianity. Oja et al. [14] proposed a method that is based on the use of two different scatter matrices that both have the independence property. Thus, the concept of independence is used itself to solve the ICA problem.

The plan of this paper is as follows. In Section 2 the scatter matrix estimators based on pairwise differences are introduced and their basic properties are discussed. The influence functions are derived in Section 3 and the limiting distribution as well as numerical values for the asymptotic and finite-sample efficiencies are given in Section 4. The paper is concluded with some final comments in Section 5.

2 Definitions and basic properties

Throughout the paper we assume that the k-variate random variables $\boldsymbol{x}$ with cumulative distribution function (cdf) $F_{\boldsymbol{x}}$ are continuous implying that $P(\boldsymbol{a}^T\boldsymbol{x} + b = 0) = 0$ for all k-vectors $\boldsymbol{a}$ and scalars b. A scatter functional is denoted by $C(\cdot)$ and is defined as follows.

Definition 1. *A $k \times k$ matrix valued functional $C(\cdot)$ is a scatter matrix if it is symmetric, positive definite and affine equivariant in the sense that, for any nonsingular $k \times k$ matrix A and k-vector $\boldsymbol{b}$, $C(F_{A\boldsymbol{x}+\boldsymbol{b}}) = AC(F_{\boldsymbol{x}})A^T$.*

We also assume that the scatter matrix functional $C(\cdot)$ can be applied to the empirical distribution function F_n. The resulting estimator is then denoted by $\widehat{C} = C(F_n)$.

Consider now two independent vectors $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$ with the same distribution F and their difference $\boldsymbol{x}_1 - \boldsymbol{x}_2$.

Definition 2. *A symmetrised scatter functional $C_S(\cdot)$ is a symmetrised version of a scatter matrix functional $C(\cdot)$ defined by $C_S(F) = C(F_{\boldsymbol{x}_1-\boldsymbol{x}_2})$ where $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$ are independent random vectors with cdf F.*

A symmetrised scatter functional is indeed a scatter functional as it is affine equivariant, symmetric and positive definite. See Oja et al. [14]. The symmetrised version of a scatter estimator is obtained by replacing $F_{\boldsymbol{x}_1-\boldsymbol{x}_2}$ in the above definition with the empirical distribution function of pairwise differences of the observations or, alternatively, using pairwise differences instead of the observations in the original scatter estimator. Such an estimator is often asymptotically equivalent to a U-statistic. See Chapter 5 in Serfling [16].

Note that in the elliptically symmetric case the distribution of the difference is also elliptically symmetric with the scatter parameter proportional to the original scatter parameter. Thus both the original and symmetrised versions of any given estimator estimate the same population quantity, up to a constant.

The following theorem and its corollary show an important property of the symmetrised scatter functionals.

Theorem 1. *Let $C(\cdot)$ be a scatter matrix functional and F be the cdf of a random vector with symmetric and independent components. Then $C(F)$ is diagonal.*

The following corollary is implied by the fact that $\boldsymbol{x}_1 - \boldsymbol{x}_2$ always has symmetric components.

Corollary 1. *A symmetrised scatter matrix functional $C_S(\cdot)$ has the independence property, that is, when F is the cdf of a random vector with independent components, $C_S(F)$ is diagonal.*

As a special case of symmetrised scatter matrices, we will consider in the following the so called symmetrised M-estimators, which from now on will be denoted by $C(\cdot)$. Let $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n$ be a random sample from a k-variate elliptical distribution. For a simplicity, write $\boldsymbol{z}_{ij}(C) = C^{-1/2}(\boldsymbol{x}_i - \boldsymbol{x}_j)$, $r_{ij}(C) = ||\boldsymbol{z}_{ij}(C)||$ and $\boldsymbol{u}_{ij}(C) = ||\boldsymbol{z}_{ij}(C)||^{-1}\boldsymbol{z}_{ij}(C)$, $1 \leq i < j \leq n$, where C is a positive definite symmetric $k \times k$ matrix.

As in the regular M-estimation, we consider three different types of symmetrised M-estimators:

Definition 3. *Let C be a positive definite symmetric $k \times k$ matrix. The symmetrised M-estimator of scatter $\widehat{C}$ is (i) the choice of C that minimizes*

$$\binom{n}{2}^{-1} \sum\sum_{i<j} \rho(r_{ij}(C)) + \frac{1}{2} \log |C|, \tag{5}$$

or (ii) the choice that solves

$$\binom{n}{2}^{-1} \sum\sum_{i<j} \left\{ w(r_{ij}(C)) \boldsymbol{u}_{ij}(C) \boldsymbol{u}_{ij}^T(C) - I_k \right\} = 0, \tag{6}$$

or (iii) the choice that solves

$$\binom{n}{2}^{-1} \sum\sum_{i<j} \left\{ w_1(r_{ij}(C)) \boldsymbol{u}_{ij}(C) \boldsymbol{u}_{ij}^T(C) - w_2(r_{ij}(C)) I_k \right\} = 0, \tag{7}$$

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

If $w(r) = \rho'(r)r$, then the estimators (5) and (6) coincide. If $w = w_1$ and $w_2(r) = 1$, estimators (6) and (7) are the same. Kent and Tyler [9, 10] considered a class of M-estimates of scatter that minimize the objective function for a given function ρ. The existence and uniqueness of so called redescending M-estimates and constrained M-estimates was proved under very light conditions on function ρ and the observed sample. In [10], the existence and uniqueness of CM-functionals was also shown and the asymptotic distributions were derived. Further, Tatsuoka and Tyler [17] and Kent and Tyler [11] considered the uniqueness of different M-functionals under broad class of symmetric distributions.

The symmetrised M-estimator based on the estimation equation (7) is indeed the same as for the regular M-estimators of scatter by Huber [7], Section 8.4, with the exception that pairwise differences are used. This removes the need for any explicit location vector in the definition. The existence and uniqueness of a solution then follow from Huber's corresponding proofs for the case of known location in Section 8.6, taking into account that what is needed of the sample is now needed of the set of pairwise differences of the sample. The assumptions used in Huber's proof for the existence are, using the current notation, as follows:

(E-1) $w_1(r)/r^2$ is decreasing, and positive when $r > 0$

(E-2) $w_2(r)$ is increasing, and positive when $r \geq 0$

(E-3) $w_1(r)$ and $w_2(r)$ are bounded and continuous

(E-4) $w_1(0)/w_2(0) < k$

(E-5) For any hyperplane H, let $P(H)$ be the fraction of pairwise differences belonging to that hyperplane.

- (i) For all hyperplanes H, $P(H) < 1 - kw_2(\infty)/w_1(\infty)$
- (ii) For all hyperplanes H, $P(H) \leq 1/k$

For the proof of uniqueness the following assumptions are needed:

(U-1) $w_1(r)/r^2$ is decreasing

(U-2) $w_1(r)$ is continuous and increasing, and positive when $r > 0$

(U-3) $w_2(r)$ is continuous and decreasing, non-negative, and positive when $0 \leq r < r_0$ for some r_0

(U-4) For all hyperplanes H $P(H) < 1/2$

Because of assumptions (E-2) and (U-3), to prove both the existence and uniqueness simultaneously the second weight function w_2 has to be constant. This is, however, not a necessary condition, since in case of several solutions, some rule for choosing the solution can be used. Therefore, in this paper w_2 is not assumed to be a constant. Further, for the influence function and asymptotic normality the existence of Taylor expansions of the weight functions and certain expectations are needed. These are assumed implicitly as the question is mostly technical.

To find a solution to the estimating equation an iterative algortihm of the form

$$C \leftarrow \frac{\sum\sum_{i<j}\{w_1(r_{ij}(C))\boldsymbol{z}_{ij}(C)\boldsymbol{z}_{ij}^T(C))\}}{\sum\sum_{i<j}\{w_2(r_{ij}(C))\}},$$

can be used. This is similar to one commonly used to find regular M-estimates (see Huber [7]).

The symmetrised M-functional $C(F)$ corresponding to $\widehat{C}$ is defined as the solution of

$$E\left[w_1(r_{12}(C(F)))\boldsymbol{u}_{12}(C(F))\boldsymbol{u}_{12}^T(C(F)) - w_2(r_{12}(C(F)))I_k\right] = 0, \qquad (8)$$

where $\boldsymbol{x}_1, \boldsymbol{x}_2 \sim F$ are independent. The proofs for the existence and uniqueness of $C(F)$ are as those for the estimator, with the exception that in assumptions (E-5) and (U-4) the probability of a hyperplane $P(H)$ is according to the true distribution of the differences. To guarantee the Fisher-consistency of $C(F)$ to Σ under the specific elliptical distribution F, the M-functional $C(F)$ should be uniquely defined at F and in addition the weight functions w_1 and w_2 should be scaled so that, for that specific F,

$$E[w_1(r_{12}(\Sigma))] = kE[w_2(r_{12}(\Sigma)))]. \qquad (9)$$

In this paper, we will consider the following M-estimators in more detail:

(i) Weight functions $w_1(r) = r^2$ and $w_2(r) = 2$ yield the regular sample covariance matrix.

(ii) The choices $w_1(r) = k$ and $w_2(r) = 1$ and an additional condition that $Tr(C) = k$ give the estimator already introduced by Dümbgen [2]. This is the symmetrised version of Tyler's [21] M-estimator.

(iii) The weight functions for symmetrised Huber's M-estimator are given by $w_2(r) = 1$ and

$$w_1(r) = \begin{cases} r^2/\sigma^2, & r^2 \leq c^2 \\ c^2/\sigma^2, & r^2 > c^2, \end{cases}$$

where c is a tuning constant defined so that $q = Pr(\chi_k^2 \leq c^2/2)$ for a chosen q. The scaling factor σ is such that $E[w(\|\boldsymbol{x}_1 - \boldsymbol{x}_2\|)] = k$, where $\boldsymbol{x}_1, \boldsymbol{x}_2 \sim N_k(\boldsymbol{0}, I_k)$. For the relationship between the tuning constant and asymptotic breakdown point of regular M-estimator, see Tyler [20].

Note that the assumptions (E-1) to (E-5) and (U-1) to (U-4) hold for the symmetrised Huber's M-estimator but (E-4) does not hold for the Dümbgen's estimator. However, as the symmetrised version of Tyler's M-estimator, it does exist and is unique.

3 Influence functions

The influence function measures the robustness of a functional T against a single outlier, that is, the effect of contamination by a distribution with its whole probability mass located at a single point $\boldsymbol{x}$ (see Hampel et al. [4]). Consider hereafter the contaminated distribution

$$F_\epsilon = (1-\epsilon)F + \epsilon\Delta_{\boldsymbol{x}},$$

where $\Delta_{\boldsymbol{x}}$ is the cdf of a distribution with probability mass one at a singular point $\boldsymbol{x}$. Then the influence function of T is defined as

$$IF(\boldsymbol{x}; T, F) = \lim_{\epsilon\to 0} \frac{T(F_\epsilon) - T(F)}{\epsilon}.$$

Hampel et al. [4] showed that, for any scatter functional $C(F)$, the influence function of C at a spherical F_0, symmetric around the origin and with $C(F_0) = I_k$, is given by

$$IF(\boldsymbol{x}; C, F_0) = \alpha_C(\|\boldsymbol{x}\|)\frac{\boldsymbol{x}\boldsymbol{x}^T}{\|\boldsymbol{x}\|^2} - \beta_C(\|\boldsymbol{x}\|)I_k, \tag{10}$$

where α_C and β_C are two real valued functions (depending on F_0). It can be seen that function α_C measures the effect of an outlier on the off-diagonal element of C, while the influence function of a diagonal element of C depends on both α_C and β_C. For robust estimator α_C and β_C should be continuous and bounded.

It is worth noting that the influence function of a symmetrised M-functional is not the same as the influence function of the corresponding regular M-functional on the symmetrised distribution. Instead, for the symmetrised M-functional defined in (8), we obtain the following.

Theorem 2. *Assume that a symmetrised M-functional $C(\cdot)$ is Fisher-consistent. Then at spherical F_0, its influence function is given by*

$$\alpha_C(\|\boldsymbol{x}\|) = \frac{1}{\eta_1} E_{\boldsymbol{x}_1}\left[w_1(\|\boldsymbol{x}_1 - \|\boldsymbol{x}\|\boldsymbol{e}_1\|)\left(1 - \frac{k(\boldsymbol{x}_1)_2^2}{\|\boldsymbol{x}_1 - \|\boldsymbol{x}\|\boldsymbol{e}_1\|^2}\right)\right]$$

$$\beta_C(\|\boldsymbol{x}\|) = \frac{1}{k}\alpha_C(\|\boldsymbol{x}\|) + \frac{1}{\eta_2} E_{\boldsymbol{x}_1}\left[w_2(\|\boldsymbol{x}_1 - \|\boldsymbol{x}\|\boldsymbol{e}_1\|) - \frac{1}{k} w_1(\|\boldsymbol{x}_1 - \|\boldsymbol{x}\|\boldsymbol{e}_1\|)\right],$$

if $\eta_2 \neq 0$, where $(\boldsymbol{x}_1)_2$ denotes the second component of $\boldsymbol{x}_1$, $\boldsymbol{e}_1 = (1, 0, \ldots, 0)^T$ and

$$\eta_1 = \frac{E[w_1'(\|\boldsymbol{x}_1 - \boldsymbol{x}_2\|)\|\boldsymbol{x}_1 - \boldsymbol{x}_2\| + k w_1(\|\boldsymbol{x}_1 - \boldsymbol{x}_2\|)]}{2k(k+2)},$$

$$\eta_2 = \frac{E[w_1'(\|\boldsymbol{x}_1 - \boldsymbol{x}_2\|)\|\boldsymbol{x}_1 - \boldsymbol{x}_2\| - k^2 w_2'(\|\boldsymbol{x}_1 - \boldsymbol{x}_2\|)\|\boldsymbol{x}_1 - \boldsymbol{x}_2\|]}{4k},$$

where $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$ are independent and have the distribution F_0. Additionally, if the symmetrised M-functional $C(\cdot)$ has a fixed trace $Tr(C(F)) = k$ then α_C is as before and $\beta_C = \alpha_C/k$.

Note that $\beta_C = \alpha_C/k$ holds for other estimators with fixed trace, like Tyler's M-estimator, as well and not just for symmetrised M-estimators. Figure 1 illustrates functions α_C and β_C for the sample covariance matrix, Dümbgen's estimator, Tyler's M-estimator, the regular Huber's M-estimator with $q = 0.94$ and the symmetrised Huber's M-estimator with $q = 0.72$ at the bivariate standard normal distribution. The values for q in both Huber's M-estimators are chosen so that the asymptotic relative efficiency with respect to the regular sample covariance matrix is 0.95 under normal distribution model (see Section 4). The influence function for regular Huber's M-estimator is given in Huber [7] and for Tyler's M-estimator, $\alpha_C(r) = k+2$ (Ollila et al. [15]).

As seen in Figure 1, functions α_C and β_C of regular Huber's M- and Tyler's M-estimator are clearly continuous and bounded. The boundedness of the influence functions of symmetrised Huber's M- and Dümbgen's estimators is implied by the following theorem.

Theorem 3. *The functions α_C and β_C of a symmetrised M-estimator are bounded when the weight functions w_1 are w_2 are bounded.*

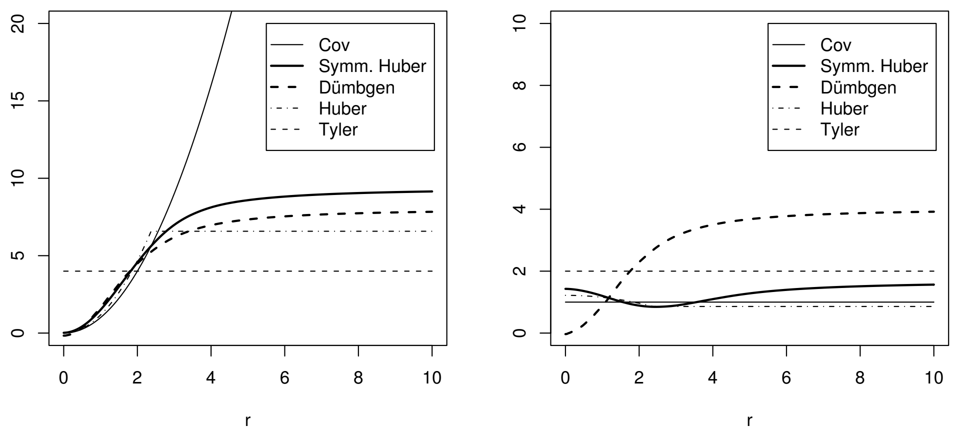

Figure 1: Functions α_C and β_C for sample covariance matrix, symmetrised Huber's M-, Dümbgen's estimator, regular Huber's M- and Tyler's M-estimator at the bivariate standard normal distribution.

4 Limiting distributions and efficiencies

In this section we give the asymptotic distribution of symmetrised M-estimators and consider their efficiency properties. In Maronna [13], the consistency and asymptotic normality of regular M-estimators of location and scatter were proven partly based on Huber's [6] results. In Huber's approach, it is assumed that the objective function is BASED ON?? the sum of i.i.d observations, therefore his results cannot be applied here. Arcones et al. [1] extended Huber's results to the case where the objective function is a U-process. In their paper, the asymptotic normality of estimators based on such a U-process is proven under some technical conditions. These results can be applied to the estimator defined via the objective function (5). The proofs for the estimators based on the estimating equations (6) are still to be done.

Next we give the result of asymptotic normality of symmetrised M-estimators assuming that our estimator is $\sqrt{n}$-consistent. In this paper, the problem of consistency is left open.

Theorem 4. *Let $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n$ be a random sample from a k-variate spherical distribution F_0 and denote $\widehat{C} = C(F_n)$ where F_n is the empirical distribution* *function of the sample. Assume that* $\widehat{C}$ *is* $\sqrt{n}$*-consistent, then*

$$\sqrt{n}\, vec(\widehat{C} - I_k) \to_d N_k(\mathbf{0}, E[vec(IF(\boldsymbol{x}; C, F_0))vec(IF(\boldsymbol{x}; C, F_0))^T]).$$

According to Tyler [18], the covariance matrix of a scatter matrix in the spherical case may be written as

$$ASV(\widehat{C}_{12}; F_0)(I_{k^2} + I_{k,k}) + ASC(\widehat{C}_{11}, \widehat{C}_{22}; F_0)vec(I_k)vec(I_k)^T,$$

where $I_{k,k}$ is a $k^2 \times k^2$ matrix with (i,j)-block being equal to a $k \times k$ matrix that has 1 at entry (j,i) and zero elsewhere, $ASV(\widehat{C}_{12}; F_0)$ denotes the asymptotic variance of any off-diagonal element and $ASC(\widehat{C}_{11}, \widehat{C}_{22}; F_0)$ the covariance of any two diagonal elements. Here, these variances and covariances are as in the following corollary to Theorem 4.

Corollary 2. *In the k-variate spherical case the asymptotic variance of any off-diagonal element of a symmetrised M-estimator* $\widehat{C}$ *is*

$$ASV(\widehat{C}_{12}; F_0) = \frac{1}{k(k+2)} E_{\boldsymbol{x}_2}\left[\frac{1}{\eta_1^2} E_{\boldsymbol{x}_1}\left[w_1(\|\boldsymbol{x}_1 - \boldsymbol{x}_2\|)\left(1 - \frac{\|\boldsymbol{x}_1\|^2}{\|\boldsymbol{x}_1 - \boldsymbol{x}_2\|}\right)\right]^2\right],$$

the asymptotic variance of any diagonal element, if $\eta_2 \neq 0$*, is*

$$\begin{aligned} ASV(\widehat{C}_{11}; F_0) = {} & \frac{2(k-1)}{k} ASV(\widehat{C}_{12}; F_0) \\ & + \frac{1}{k^2} E_{\boldsymbol{x}_2}\left[\frac{1}{\eta_2^2} E_{\boldsymbol{x}_1}\left[w_1(\|\boldsymbol{x}_1 - \boldsymbol{x}_2\|) - k w_2(\|\boldsymbol{x}_1 - \boldsymbol{x}_2\|)\right]^2\right] \end{aligned}$$

and the asymptotic covariance between any two distinct diagonal elements is

$$ASC(\widehat{C}_{11}, \widehat{C}_{22}; F_0) = ASV(\widehat{C}_{11}; F_0) - 2ASV(\widehat{C}_{12}; F_0).$$

Note that due to the affine equivariance of $\widehat{C}$ and properties of *vec*-operator and Kronecker product, the limiting distribution of $\sqrt{n}\, vec(\widehat{C} - C)$ at elliptical F is multivariate normal with zero mean and covariance matrix

$$ASV(\widehat{C}_{12}; F_0)(I_{k^2} + I_{k,k})(C \otimes C) + ASC(\widehat{C}_{11}, \widehat{C}_{22}; F_0)vec(C)vec(C)^T,$$

where $ASV(\widehat{C}_{12}; F_0)$ and $ASC(\widehat{C}_{11}, \widehat{C}_{22}; F_0)$ are as in Corollary 2.

It should be noted that the assumption about condition (9) is essential. The limiting distribution when that condition does not hold could also be given but this is not sensible as in practice the real underlying distribution is unknown. Because of this, the estimators are tuned under one reference distribution, usually the normal distribution. Different scatter matrix estimators are thus comparable only in the normal distribution case. For other elliptical distributions, a correction factor is needed in order to have Fisher consistency towards Σ and further to make scatter matrices comparable.

In the following we compare limiting efficiencies of different M-estimators. To circumvent the problem of Fisher consistency, we compare different shape matrix estimators instead of scatter matrices. The shape matrix functional $V(\cdot)$ associated with the scatter functional $C(\cdot)$ is defined by

$$V(F) = \frac{k}{Tr(C(F))} C(F).$$

Note that both Dümbgen's and Tyler's estimators estimate the shape without any modifications. At elliptical F, all shape estimators estimate the same population quantity and are comparable without any correction factors. Moreover, in most applications it is enough to estimate the scatter only up to a constant. The limiting distribution of $\sqrt{n}\, vec(\widehat{V} - V)$ at elliptical distribution is given by the following theorem; the result follows from Theorem 1 in Tyler [19].

Theorem 5. *Let $\widehat{C}$ be a scatter matrix and $\widehat{V} = (k/Tr(\widehat{C}))\widehat{C}$ the associated shape matrix. The limiting distribution of $\sqrt{n}\, vec(\widehat{V} - V)$ at elliptical F is multinormal with asymptotic covariance matrix*

$$\tau_1 \left(I_{k^2} - \frac{1}{k} vec(V) vec(I_k)^T \right) (I_{k^2} + K_k)(V \otimes V) \left(I_{k^2} - \frac{1}{k} vec(I_k) vec(V)^T \right),$$

where $\tau_1 = ASV(\widehat{V}_{12}; F_0)$.

The limiting distribution of the shape matrix estimator is thus characterized by one single number, that is, the variance of any off-diagonal element of $\widehat{V}$ at spherical F_0 (Ollila et al. [15]). The asymptotic relative efficiencies of shape matrix estimators are in turn ratios of these variances. The variance of an off-diagonal element of the Tyler's estimate is $(k+2)/k$, for the shape estimate based on regular Huber's estimate see Huber [7], Section 8.7. To find the variances of the off-diagonal elements of the symmetrised M-estimators considered here we used a combination of numeric integration and Monte Carlo simulation.

Table 1 lists the limiting efficiencies of shape estimators based on the regular Huber's M-estimator and symmetrised Huber's M-estimator with respect to the shape estimator based on the regular sample covariance matrix, or in other words the regular shape estimator. The efficiencies are considered under different t-distributions with selected values of dimensions k and degrees of freedom ν, with $\nu = \infty$ referring to the normal case. In order to make the estimators comparable, the tuning parameter q was chosen so that for both estimators, the resulting efficiency in the normal case with respect to the regular shape estimator is 0.95. These values for q are 0.94, 0.91, 0.87 and 0.84 for the regular Huber's M-estimator and 0.72, 0.56, 0.39 and 0.30 for the symmetrised Huber's M-estimator, for dimensions 2, 3, 4 and 5, respectively. In Table 2, the limiting efficiencies of the Tyler's M-estimator and its symmetrised version, the Dümbgen's estimator, with respect to the regular shape estimator are given. The distributions considered here are the same as in Table 1.

Table 1: Asymptotic relative efficiencies of the shape estimators based on regular Huber's M-estimator (H) and symmetrised Huber's M-estimator (S) relative to the regular shape estimator at different t-distribution cases with selected values of k and ν.

	$k=2$		$k=3$		$k=4$		$k=5$	
ν	H	S	H	S	H	S	H	S
5	2.21	2.37	2.27	2.53	2.32	2.59	2.36	2.65
6	1.54	1.63	1.58	1.71	1.60	1.75	1.63	1.78
8	1.22	1.26	1.24	1.33	1.27	1.34	1.27	1.36
15	1.04	1.06	1.04	1.10	1.05	1.10	1.06	1.11
∞	0.95	0.95	0.95	0.95	0.95	0.95	0.95	0.95

Table 1 shows that when shape estimators based on the regular and symmetrised Huber's M-estimators are tuned to the same efficiency with respect to the regular shape estimator, the symmetrised Huber's M is more efficient in the considered t-distribution cases. From Table 2 it can be seen that when the Tyler's M-estimator is symmetrised, the increase in efficiency is considerable under these distributions. The price to pay is a loss of some robustness, see discussion in Section 5.

Table 3 lists the results of a small simulation study concerning finite sample efficiencies of the same shape estimators as in Tables 1 and 2 with

Table 2: Asymptotic relative efficiencies of the shape estimators based on Tyler's M-estimator (T) and Dümbgen's estimator (D) relative to the regular shape estimator at different t-distribution cases with selected values of k and ν.

	$k=2$		$k=3$		$k=4$		$k=5$	
ν	T	D	T	D	T	D	T	D
5	1.50	2.39	1.80	2.49	2.00	2.59	2.14	2.68
6	1.00	1.61	1.20	1.67	1.33	1.72	1.43	1.77
8	0.75	1.23	0.90	1.27	1.00	1.30	1.07	1.33
15	0.59	1.02	0.71	1.04	0.79	1.06	0.84	1.07
∞	0.50	0.91	0.60	0.92	0.67	0.93	0.71	0.94

respect to the regular shape estimator. 1500 samples of three different sample sizes and two different dimensions were drawn from t-distributions with 5 and 8 degrees of freedom and from the normal distribution. For every estimator and distribution, the mean squared errors

$$\text{MSE}(\widehat{V}_{ij}) = \frac{1}{1500}\sum_{k=1}^{1500}(\widehat{V}_{ij}^{(k)} - I_{ij})^2$$

were computed for every off-diagonal element, that is, $i \neq j$ (I_{ij} is then of course equal to zero). Since the off-diagonal elements have equal variances and are uncorrelated, a further mean of their MSE's was taken. The listed finite sample efficiencies are then ratios of these means. The corresponding asymptotic relative efficiencies (denoted by $n=\infty$) from Tables 1 and 2 are also listed for easy reference.

The results of the small sample study show that the convergence to the limiting efficiency is reasonably fast in the case of $\nu = 8$. In the case of $\nu = 5$ the convergence is clear but much slower. Especially for small sample sizes, the loss in efficiency is remarkable, but also in the case $n = 200$, the efficiencies are far from the asymptotical ones. In the normal case the finite sample and the limiting efficiencies are naturally the same.

Table 3: Finite sample efficiencies of the shape estimators based on symmetrised Huber's M- (S), Huber's M- (H), Dümbgen's estimator (D), regular Huber's M- (H) and Tyler's M-estimator (T) with respect to the regular shape matrix

		$k=3$				$k=5$			
	n	S	H	D	T	S	H	D	T
$\nu=5$	20	1.25	1.28	1.25	0.98	1.31	1.36	1.33	1.15
	50	1.48	1.48	1.50	1.12	1.59	1.60	1.60	1.38
	200	1.78	1.77	1.80	1.35	1.93	1.93	1.95	1.72
	∞	2.53	2.27	2.49	1.80	2.65	2.36	2.68	2.14
$\nu=8$	20	1.10	1.13	1.07	0.83	1.12	1.15	1.12	0.92
	50	1.19	1.18	1.17	0.83	1.20	1.20	1.19	0.99
	200	1.22	1.21	1.21	0.84	1.26	1.24	1.26	1.04
	∞	1.33	1.24	1.27	0.90	1.36	1.27	1.33	1.07
$\nu=\infty$	20	0.96	0.99	0.91	0.62	0.96	0.94	0.99	0.71
	50	0.95	0.92	0.96	0.61	0.96	0.94	0.95	0.72
	200	0.95	0.92	0.93	0.59	0.95	0.94	0.93	0.72
	∞	0.95	0.95	0.92	0.60	0.95	0.95	0.94	0.71

5 Final remarks

We have shown that the use of pairwise differences in M-estimation of scatter may lead to increase in efficiency. Another benefit is the fact that the location need not be estimated nor known. This offers one solution to the problem of finding simultaneous M-estimates of location and scatter. Previously proposed algorithms either have restrictions on the weight functions or do not have rigorous proofs of convergence, see Hettmansperger and Randles [5]. Symmetrising the scatter estimator of an existing pair of estimators or combining a symmetrised scatter estimator with a location estimator gives a pair of affine equivariant location and scatter estimators with certainly converging algorithms.

Symmetrised M-estimators of scatter were shown to have a property that regular M-estimators in general do not. It is well known that when the components of a random vector are independent, the regular sample covariance matrix is diagonal. This is not true for M-estimators of scatter in general but the symmetrisation of marginal distributions inherent in the symmetrised M-estimators of scatter ensures it. This independence property can be used in the so called independent component analysis (see Oja et al. [14]) to find a random vector with independent components when only an unknown linear mixing of it is observed. A new class of estimators using robust or non-parametric estimators of scatter such as the symmetrised M-estimators of scatter may be therefore used to solve this problem. However, it should be noted that the forms of the influence function and the limiting distribution derived in Section 4 apply only to the elliptic distribution family. Within this family the only random variables with independent components are in fact those with the spherical normal distribution and in that special case any scatter functional is diagonal. The detailed analysis of symmetrised scatter functionals under the so called IC-model, or the one containing distributions with independent components and their affine transformations, is an open question and will be studied in the future.

The breakdown properties of Dümbgen's estimator have been studied by Dümbgen and Tyler [3] and the breakdown point was found to be $1 - \sqrt{1 - 1/k}$ in case of special kind of contamination. If the type of contamination is restricted, the breakdown point becomes $1/k$, that is, the same as for Tyler's M-estimator. The breakdown behaviour of symmetrised M-estimators in general is left open but it is apparent that when an estimator is symmetrised, its breakdown point drops as a single outlier affects $n - 1$ pairwise differences. However, in the light of the efficiency studies in Section 4, it could be argued that a symmetrised version of a highly robust estimator loses some of the robustness but gains efficiency instead.

References

[1] Arcones, M.A., Chen, Z. and Gine, E. (1994), Estimators related to U-processes with applications to multivariate medians: asymptotic normality, *Annals of Statistics*, **22**, 1460–1477.

[2] Dümbgen, L. (1998), On Tyler's M-Functional of Scatter in High Dimension, *Annals of the Institute of Statistical Mathematics*, **50**, 471–491.

[3] Dümbgen, L. and Tyler, D. (2005), On the Breakdown Properties of Some Multivariate M-Functionals, *Scandinavian Journal of Statistics*, **32**, 247–264.

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

[5] Hettmansperger, T.P. and Randles, R. (2002), A Practical Affine Equivariant Multivariate Median, *Biometrika*, **89**, 851–860.

[6] Huber, P. J. (1967), The Behavior of Maximum-likelihood Estimates under Non-standard Conditions. In *Proceedings of the fifth Berkeley Symposium on Mathematical Statistics and Probability*, Berkeley CA, University of California Press, 221–33.

[7] Huber, P.J. (1981), *Robust Statistics*, Wiley, New York.

[8] Hyvärinen, A., Karhunen, J. and Oja, E. (2001), *Independent Component Analysis.* John Wiley & Sons.

[9] Kent, J.T. and Tyler, D.E. (1991), Redescending M-estimates of multivariate location and scatter, *Annals of Statistics*, **19**, 2102–2119.

[10] Kent, J.T. and Tyler, D.E. (1996), Constrained M-estimation for multivariate location and scatter, *Annals of Statistics*, **24**, 1346–1370.

[11] Kent, J.T. and Tyler, D.E. (2001), Regularity and uniqueness for constrained M-estimates and redescending M-estimates, *Annals of Statistics*, **29**, 252–265.

[12] Lehmann, E.L. (1998), *Nonparametrics: Statistical methods based on ranks*, Prentice-Hall, New Jersey.

[13] Maronna, R.A. (1976), Robust M-estimators of multivariate location and scatter. *Annals of Statistics*, **4**, 51–67.

[14] Oja, H., Sirkiä, S. and Eriksson J. (2006), Scatter Matrices and Independent Component Analysis, *Austrian Journal of Statistics*, **35**, 175–189.

[15] Ollila, E., Hettmansperger, T.P. and Oja, H. (2003), Affine equivariant multivariate sign methods, Under revision.

[16] Serfling, R.J. (1980), *Approximate Theorems of Mathematical Statistics*, Wiley, New York.

[17] Tatsuoka, K.S. and Tyler, D.E. (2000), On the uniqueness of S-functionals and M-functionals under nonelliptical distributions, *Annals of Statistics*, **28**, 1219–1243.

[18] Tyler, D.E. (1982). Radial estimates and the test for sphericity, *Biometrika*, **69**, 429–436.

[19] Tyler, D.E. (1983), Robustness and efficiency properties of scatter matrices, *Biometrika*, **70**, 411–420.

[20] Tyler, D.E. (1986), Breakdown properties of the M-estimators of multivariate scatter, Report, Department of Statistics, Rutgers University.

[21] Tyler, D.E. (1987), A distribution-free M-estimator of multivariate scatter, *The Annals of Statistics*, **15**, 234–251.

Appendix: Proofs of the results

Proof of Theorem 1. Assume that $\boldsymbol{x}$ is a random k-vector with independent and symmetric components and let $\boldsymbol{\mu}$ be the vector of symmetry centers. Let I_i^- be a $k \times k$ diagonal matrix that has -1 the ith diagonal element and $+1$ as all other diagonal elements. Now because $\boldsymbol{x} - \boldsymbol{\mu}$ and $I_i^-(\boldsymbol{x} - \boldsymbol{\mu})$ have the same distribution and $C(\cdot)$ is a scatter matrix functional it holds

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

for all $i = 1, \ldots, k$, which implies that all off-diagonal elements of $C(F_{\boldsymbol{x}})$ are equal to their opposite, that is, are equal to zero.

To prove Theorem 2, we need the following Lemma.

Lemma 1. *Assume that $\boldsymbol{x}_1$ is a random vector with a spherical distribution F_0 and for an arbitrary $\boldsymbol{x}$ write $\boldsymbol{x} = r\boldsymbol{u}$, where $r = ||\boldsymbol{x}||$ and $\boldsymbol{u} = ||\boldsymbol{x}||^{-1}\boldsymbol{x}$. Then*

$$\begin{aligned} M(\boldsymbol{x}) := E_{\boldsymbol{x}_1}\left[w_1(\|\boldsymbol{x}_1 - \boldsymbol{x}\|) \frac{(\boldsymbol{x}_1 - \boldsymbol{x})(\boldsymbol{x}_1 - \boldsymbol{x})^T}{\|\boldsymbol{x}_1 - \boldsymbol{x}\|^2} - w_2(\|\boldsymbol{x}_1 - \boldsymbol{x}\|) I_k \right] \\ = (m(r) - kg(r))\boldsymbol{u}\boldsymbol{u}^T + g(r) I_k, \end{aligned}$$

where

$$m(r) = E_{\boldsymbol{x}_1}\left[w_1(\|\boldsymbol{x}_1 - r\boldsymbol{e}_1\|) - kw_2(\|\boldsymbol{x}_1 - r\boldsymbol{e}_1\|)\right]$$

and

$$g(r) = E_{\boldsymbol{x}_1}\left[w_1(\|\boldsymbol{x}_1 - r\boldsymbol{e}_1\|)\frac{(\boldsymbol{x}_1)_2^2}{\|\boldsymbol{x}_1 - r\boldsymbol{e}_1\|^2} - w_2(\|\boldsymbol{x}_1 - r\boldsymbol{e}_1\|)\right],$$

where in turn $(\boldsymbol{x}_1)_2$ *and* $\boldsymbol{e}_1$ *are as in Theorem 2.*

Proof of Lemma 1. First consider the case $\boldsymbol{x} = r\boldsymbol{e}_1$. Due to spherical symmetry of $\boldsymbol{x}_1$ $M(r\boldsymbol{e}_1)$ is diagonal and all diagonal elements except the first are equal to

$$g(r) := E_{\boldsymbol{x}_1}\left[w_1(\|\boldsymbol{x}_1 - r\boldsymbol{e}_1\|)\frac{(\boldsymbol{x}_1)_2^2}{\|\boldsymbol{x}_1 - r\boldsymbol{e}_1\|^2} - w_2(\|\boldsymbol{x}_1 - r\boldsymbol{e}_1\|)\right].$$

The trace of $M(r\boldsymbol{e}_1)$ is equal to

$$m(r) := E_{\boldsymbol{x}_1}\left[w_1(\|\boldsymbol{x}_1 - r\boldsymbol{e}_1\|) - kw_2(\|\boldsymbol{x}_1 - r\boldsymbol{e}_1\|)\right].$$

Now, the first diagonal element is equal to $m(r) - (k-1)g(r)$ and so

$$M(r\boldsymbol{e}_1) = (m(r) - kg(r))\boldsymbol{e}_1\boldsymbol{e}_1^T + g(r)I_k.$$

For an arbitrary $\boldsymbol{x}$ there exists an orthogonal matrix A such that $\boldsymbol{x} = A(r\boldsymbol{e}_1)$. Note that $\boldsymbol{x}_1$ and $A\boldsymbol{x}_1$ have the same distribution and that $\|A\boldsymbol{s}\| = \|\boldsymbol{s}\|$ for any vector $\boldsymbol{s}$. We then have that

$$M(\boldsymbol{x}) = AM(r\boldsymbol{e}_1)A^T = (m(r) - kg(r))\boldsymbol{u}\boldsymbol{u}^T + g(r)I_k.$$

Proof of Theorem 2. Inserting $F_\epsilon = (1-\epsilon)F_0 + \epsilon\delta_{\boldsymbol{x}}$ to equation (8) (and adopting the convention that $0/0 = 0$) and taking the derivative with respect to ϵ at $\epsilon = 0$ yields to

$$\begin{aligned}
&\frac{\partial}{\partial\epsilon}E\left[w_1(r_{12}(C(F_\epsilon)))\boldsymbol{u}_{12}(C(F_\epsilon))\boldsymbol{u}_{12}^T(C(F_\epsilon)) - w_2(r_{12}(C(F_\epsilon)))I_k\right]\Bigg|_{\epsilon=0} \\
&- 2E\left[w_1(r_{12}(C(F_\epsilon)))\boldsymbol{u}_{12}(C(F_\epsilon))\boldsymbol{u}_{12}^T(C(F_\epsilon)) - w_2(r_{12}(C(F_\epsilon)))I_k\right]\Bigg|_{\epsilon=0} \\
&+ 2E\left[w_1(r_{\boldsymbol{x}}(C(F_\epsilon)))\boldsymbol{u}_{\boldsymbol{x}}(C(F_\epsilon))\boldsymbol{u}_{\boldsymbol{x}}^T(C(F_\epsilon)) - w_2(r_{\boldsymbol{x}}(C(F_\epsilon)))I_k\right]\Bigg|_{\epsilon=0} = 0,
\end{aligned}$$

where $r_{\boldsymbol{x}}(C(F_\epsilon)) = \|C(F_\epsilon)^{-1/2}(\boldsymbol{x}_1 - \boldsymbol{x})\|_{C(F_\epsilon)}$ and $\boldsymbol{u}_{\boldsymbol{x}}(C(F_\epsilon)) = r_{\boldsymbol{x}}(C(F_\epsilon))^{-1}\, C(F_\epsilon)^{-1/2}(\boldsymbol{x}_1 - \boldsymbol{x})$. In the following, write $IF(\boldsymbol{x}; C, F_0) = IF(\boldsymbol{x})$ for simplicity. Note that the second term above is equal to zero. Now (assuming that the order of differentiation and integration can be changed) one has that

$$
\begin{aligned}
&E\left[w_1(\|\boldsymbol{x}_1 - \boldsymbol{x}_2\|)\frac{(\boldsymbol{x}_1 - \boldsymbol{x}_2)(\boldsymbol{x}_1 - \boldsymbol{x}_2)^T(\boldsymbol{x}_1 - \boldsymbol{x}_2)^T IF(\boldsymbol{x})(\boldsymbol{x}_1 - \boldsymbol{x}_2)}{\|\boldsymbol{x}_1 - \boldsymbol{x}_2\|^4}\right] \\
&- E\left[w_1(\|\boldsymbol{x}_1 - \boldsymbol{x}_2\|)\frac{(\boldsymbol{x}_1 - \boldsymbol{x}_2)(\boldsymbol{x}_1 - \boldsymbol{x}_2)^T IF(\boldsymbol{x}) + IF(\boldsymbol{x})(\boldsymbol{x}_1 - \boldsymbol{x}_2)(\boldsymbol{x}_1 - \boldsymbol{x}_2)^T}{2\|\boldsymbol{x}_1 - \boldsymbol{x}_2\|^2}\right] \\
&- E\left[w_1'(\|\boldsymbol{x}_1 - \boldsymbol{x}_2\|)\|\boldsymbol{x}_1 - \boldsymbol{x}_2\|\frac{(\boldsymbol{x}_1 - \boldsymbol{x}_2)(\boldsymbol{x}_1 - \boldsymbol{x}_2)^T(\boldsymbol{x}_1 - \boldsymbol{x}_2)^T IF(\boldsymbol{x})(\boldsymbol{x}_1 - \boldsymbol{x}_2)}{2\|\boldsymbol{x}_1 - \boldsymbol{x}_2\|^4}\right] \\
&+ E\left[w_2'(\|\boldsymbol{x}_1 - \boldsymbol{x}_2\|)\|\boldsymbol{x}_1 - \boldsymbol{x}_2\|\frac{(\boldsymbol{x}_1 - \boldsymbol{x}_2)^T IF(\boldsymbol{x})(\boldsymbol{x}_1 - \boldsymbol{x}_2)}{2\|\boldsymbol{x}_1 - \boldsymbol{x}_2\|^2} I_k\right] \\
&+ 2E\left[w_1(\|\boldsymbol{x}_1 - \boldsymbol{x}\|)\frac{(\boldsymbol{x}_1 - \boldsymbol{x})(\boldsymbol{x}_1 - \boldsymbol{x})^T}{\|\boldsymbol{x}_1 - \boldsymbol{x}\|^2} - w_2(\|\boldsymbol{x}_1 - \boldsymbol{x}\|)I_k\right] = 0.
\end{aligned}
$$

Notice next that as $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$ are spherically distributed, also $\boldsymbol{x}_1 - \boldsymbol{x}_2$ is spherical, that is, $\|\boldsymbol{x}_1 - \boldsymbol{x}_2\|$ and $\|\boldsymbol{x}_1 - \boldsymbol{x}_2\|^{-1}(\boldsymbol{x}_1 - \boldsymbol{x}_2)$ are independent. To simplify notations write $r_{12} = \|\boldsymbol{x}_1 - \boldsymbol{x}_2\|$,

$$a = \frac{E\left[w_1'(r_{12})r_{12} + k w_1(r_{12})\right]}{k(k+2)}$$

and

$$b = \frac{E\left[w_1'(r_{12})r_{12} - 2w_1(r_{12}) - (k+2)w_2'(r_{12})r_{12}\right]}{2k(k+2)}.$$

Then using Lemma 1 together with

$$E\left[\frac{(\boldsymbol{x}_1 - \boldsymbol{x}_2)(\boldsymbol{x}_1 - \boldsymbol{x}_2)^T(\boldsymbol{x}_1 - \boldsymbol{x}_2)^T IF(\boldsymbol{x})(\boldsymbol{x}_1 - \boldsymbol{x}_2)}{\|\boldsymbol{x}_1 - \boldsymbol{x}_2\|^4}\right] = \frac{2IF(\boldsymbol{x}) + Tr(IF(\boldsymbol{x}))I_k}{k(k+2)}$$

and

$$E\left[\frac{(\boldsymbol{x}_1 - \boldsymbol{x}_2)^T IF(\boldsymbol{x})(\boldsymbol{x}_1 - \boldsymbol{x}_2)}{\|\boldsymbol{x}_1 - \boldsymbol{x}_2\|^2}\right] = \frac{1}{k}Tr(IF(\boldsymbol{x})),$$

the above equation simplifies to

$$a\, IF(\boldsymbol{x}) = 2(m(\|\boldsymbol{x}\|) - kg(\|\boldsymbol{x}\|))\boldsymbol{u}\boldsymbol{u}^T + 2g(\|\boldsymbol{x}\|)I_k - b\, Tr(IF(\boldsymbol{x}))I_k. \quad (11)$$

Now taking the trace on both sides and solving $Tr(IF(\boldsymbol{x}))$ we get $Tr(IF(\boldsymbol{x})) = 2(a+bk)^{-1}m(\|\boldsymbol{x}\|)$ if $a+bk$ is not zero. Together with (11) this gives

$$IF(\boldsymbol{x}; C, F_0) = \frac{2}{a}(m(\|\boldsymbol{x}\|) - kg(\|\boldsymbol{x}\|))\boldsymbol{u}\boldsymbol{u}^T - \frac{2}{a}\left(\frac{b}{a+bk}m(\|\boldsymbol{x}\|) - g(\|\boldsymbol{x}\|)\right) I_k.$$

Denote next $\eta_1 = a/2$ and $\eta_2 = (a+bk)/2$. The result then follows using Lemma 1. Note that if $a+bk$ is zero then equation (11) is identically true with respect to trace of the influence function. This means that the trace and thus also β_C cannot be found this way. On the other hand, if the trace of $C(F)$ is known to be k then it is also known that $Tr(IF(\boldsymbol{x}; C, F_0)) = 0$. Taking the trace in (11) shows that then also $m(\|\boldsymbol{x}\|) = 0$ and so

$$IF(\boldsymbol{x}; C, F_0) = \frac{2}{a}(-kg(\|\boldsymbol{x}\|))\boldsymbol{u}\boldsymbol{u}^T + \frac{2}{a}g(\|\boldsymbol{x}\|)I_k.$$

Proof of Theorem 3. It suffices to show the finiteness of

$$E_{\boldsymbol{x}_1}\left[w_1(\|\boldsymbol{x}_1 - r\boldsymbol{e}_1\|)\frac{k(\boldsymbol{x}_1)_2^2}{\|\boldsymbol{x}_1 - r\boldsymbol{e}_1\|^2}\right]$$

and since $w_1(\|\boldsymbol{x}_1 - r\boldsymbol{e}_1\|) < K$ for some K it is sufficient to consider only

$$E_{\boldsymbol{x}_1}\left[\frac{(\boldsymbol{x}_1)_2^2}{\|\boldsymbol{x}_1 - r\boldsymbol{e}_1\|^2}\right],$$

which is obviously finite since

$$0 \le (\boldsymbol{x}_1)_2^2 \le \|\boldsymbol{x}_1 - r\boldsymbol{e}_1\|^2.$$

To prove Theorem 4, we need the following Lemma.

Lemma 2. *Let $\boldsymbol{x}_i, \ldots, \boldsymbol{x}_n$ be a sample from a spherically symmetric distribution and write $\boldsymbol{x}_{ij} = \boldsymbol{x}_i - \boldsymbol{x}_j$, $r_{ij} = \|\boldsymbol{x}_{ij}\|$ and $\boldsymbol{u}_{ij} = r_{ij}^{-1}\boldsymbol{x}_{ij}$ for simplicity. Assume that the symmetrised M-estimator $\widehat{C}$ is $\sqrt{n}$-consistent, then*

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

where η_1 and η_2 are as in Theorem 2.

Proof of Lemma 2. At first, write $C^* = \sqrt{n}(\widehat{C} - I_k)$. Since $\widehat{C}$ is $\sqrt{n}$-consistent, C^* is bounded in probability. Further,

$$\widehat{C}^{-1/2} = I_k - \frac{1}{2\sqrt{n}} C^* + o_p(n^{-1/2}),$$

$$\boldsymbol{z}_{ij} = \widehat{C}^{-1/2}\boldsymbol{x}_{ij} = \boldsymbol{x}_{ij} - \frac{1}{2\sqrt{n}} C^*\boldsymbol{x}_{ij} + o_p(n^{-1/2})$$

and

$$\frac{\boldsymbol{z}_{ij}\boldsymbol{z}_{ij}^T}{\|\boldsymbol{z}_{ij}\|^2} = \boldsymbol{u}_{ij}\boldsymbol{u}_{ij}^T + \frac{1}{\sqrt{n}}\boldsymbol{u}_{ij}^T C^*\boldsymbol{u}_{ij}\boldsymbol{u}_{ij}\boldsymbol{u}_{ij}^T - \frac{1}{2\sqrt{n}}(C^*\boldsymbol{u}_{ij}\boldsymbol{u}_{ij}^T + \boldsymbol{u}_{ij}\boldsymbol{u}_{ij}^T C^*) + o_p(n^{-1/2}).$$

Using the Taylor series expansion, we get

$$w_1(\|\boldsymbol{z}_{ij}\|) = w_1(r_{ij}) - \frac{1}{2\sqrt{n}} w_1'(r_{ij}) r_{ij}\, \boldsymbol{u}_{ij}^T C^* \boldsymbol{u}_{ij} + o_p(n^{-1/2})$$

and

$$w_2(\|\boldsymbol{z}_{ij}\|) = w_2(r_{ij}) - \frac{1}{2\sqrt{n}} w_2'(r_{ij}) r_{ij}\, \boldsymbol{u}_{ij}^T C^* \boldsymbol{u}_{ij} + o_p(n^{-1/2})$$

Now (omitting the $o_p(n^{-1/2})$-terms)

$$\begin{aligned}
&\sqrt{n}\left(\binom{n}{2}^{-1} \sum\sum_{i<j} \left\{ w_1(\|\boldsymbol{z}_{ij}\|)\frac{\boldsymbol{z}_{ij}\boldsymbol{z}_{ij}^T}{\|\boldsymbol{z}_{ij}\|^2} - w_2(\|\boldsymbol{z}_{ij}\|) I_k \right\}\right) \\
&= \sqrt{n}\left(\binom{n}{2}^{-1} \sum\sum_{i<j} \left\{ w_1(r_{ij})\boldsymbol{u}_{ij}\boldsymbol{u}_{ij}^T - w_2(r_{ij}) I_k \right\} + \frac{1}{\sqrt{n}} w_1(r_{ij})\boldsymbol{u}_{ij}^T C^* \boldsymbol{u}_{ij}\boldsymbol{u}_{ij}\boldsymbol{u}_{ij}^T \right. \\
&\quad - \frac{1}{2\sqrt{n}} w_1(r_{ij})(C^*\boldsymbol{u}_{ij}\boldsymbol{u}_{ij}^T + \boldsymbol{u}_{ij}\boldsymbol{u}_{ij}^T C^*) - \frac{1}{2\sqrt{n}} w_1'(r_{ij}) r_{ij}\, \boldsymbol{u}_{ij}^T C^*\boldsymbol{u}_{ij}\boldsymbol{u}_{ij}\boldsymbol{u}_{ij}^T \\
&\quad \left. + \frac{1}{2\sqrt{n}} w_2'(r_{ij}) r_{ij}\, \boldsymbol{u}_{ij}^T C^* \boldsymbol{u}_{ij}) \right) = 0.
\end{aligned}$$

Then proceeding as in the proof of Theorem 2, we get

$$\begin{aligned}
&\sqrt{n}\left(\binom{n}{2}^{-1} \sum\sum_{i<j} \left\{ w_1(r_{ij})\boldsymbol{u}_{ij}\boldsymbol{u}_{ij}^T - w_2(r_{ij}) I_k \right\}\right) = E\left[\frac{w_1(r_{ij})}{k}\right] C^* \\
&\quad - E\left[\frac{w_2'(r_{ij}) r_{ij}}{2k}\right] Tr(C^*) I_k + E\left[\frac{w_1'(r_{ij}) r_{ij} - 2w_1(r_{ij})}{2k(k+2)}\right] (2C^* + Tr(C^*) I_k) \\
&= aC^* + bTr(C^*) I_k.
\end{aligned}$$

Now taking the trace on both sides and solving $Tr(C^*)$ yields to $Tr(C^*) = (a+bk)^{-1}\sqrt{n}\big(\binom{n}{2}^{-1}\sum\sum_{i<j}\{w_1(r_{ij}) - w_2(r_{ij})k\}\big)$. Therefore,

$$\begin{aligned} C^* &= \sqrt{n}(\widehat{C} - I_k) \\ &= \sqrt{n}\left[\binom{n}{2}^{-1}\sum\sum_{i<j}\left\{\frac{w_1(r_{ij})}{a}\left(\boldsymbol{u}_{ij}\boldsymbol{u}_{ij}^T - \frac{b}{a+bk}\right) - \frac{w_2(r_{ij})}{a+bk}I_k\right\}\right] \end{aligned}$$

and the result follows since $\eta_1 = a/2$ and $\eta_2 = (a+bk)/2$. Note that, for the Dümbgen's estimator, the terms including η_2 reduce to zero and

$$\sqrt{n}(\widehat{C} - I_k) = \sqrt{n}\,(k+2)\binom{n}{2}^{-1}\sum\sum_{i<j}\left\{\boldsymbol{u}_{ij}\boldsymbol{u}_{ij}^T - k^{-1}I_k\right\} + o_p(1).$$

Proof of Theorem 4. Lemma 2 shows that $\sqrt{n}\,vec(\widehat{C} - I_k)$ has the same limiting distribution as $\sqrt{n}\,\boldsymbol{U}$, where

$$\boldsymbol{U} = \binom{n}{2}^{-1}\sum\sum_{i<j}\boldsymbol{\varphi}(\boldsymbol{x}_i, \boldsymbol{x}_j)$$

with

$$\begin{aligned} \varphi(\boldsymbol{x}_1, \boldsymbol{x}_2) &= \frac{1}{2}\left(\frac{w_1(r_{12})}{\eta_1}vec(\boldsymbol{u}_{12}\boldsymbol{u}_{12}^T) - \left(\frac{(\eta_2-\eta_1)w_1(r_{12})}{k\eta_1\eta_2} + \frac{w_2(r_{12})}{\eta_2}\right)vec(I_k)\right) \\ &= \varphi(\boldsymbol{x}_2, \boldsymbol{x}_1) \end{aligned}$$

is a k^2-variate U-statistic. A straightforward generalisation of univariate result given in Lehmann [12], Appendix, states that

$$\sqrt{n}(\boldsymbol{U} - \boldsymbol{\theta}) \to_d N_{k^2}(\boldsymbol{0}, 4E[\boldsymbol{\psi}(\boldsymbol{x})\boldsymbol{\psi}(\boldsymbol{x})^T]),$$

where $\boldsymbol{\theta} = E[\boldsymbol{\varphi}(\boldsymbol{x}_1, \boldsymbol{x}_2)]$ and $\boldsymbol{\psi}(\boldsymbol{x}) = E_{\boldsymbol{x}_1}[\boldsymbol{\varphi}(\boldsymbol{x}_1, \boldsymbol{x})] - \boldsymbol{\theta}$. The subindex $\boldsymbol{x}_1$ means again that the expectation is with respect to $\boldsymbol{x}_1$. Here $\boldsymbol{\theta} = \boldsymbol{0}$ which follows from noting that r_{12} and $\boldsymbol{u}_{12}$ are independent, $E[\boldsymbol{u}_{12}\boldsymbol{u}_{12}^T] = k^{-1}I_k$ and $E[w_2(r_{12}) - k^{-1}w_1(r_{12})] = 0$. The idea of finding $\boldsymbol{\psi}(\boldsymbol{x})$ is the same as in Lemma 1, resulting in

$$\boldsymbol{\psi}(\boldsymbol{x}) = \frac{1}{2}IF(\boldsymbol{x}; C, F_0),$$

from which the result follows.

Proof of Corollary 2. Using the same notation as in Theorems 2 and 4 and noting that $\|\boldsymbol{x}\|$ and $\|\boldsymbol{x}\|^{-1}\boldsymbol{x}$ are independent it is possible to write the asymptotic covariance matrix of $\widehat{C}$ as

$$\begin{aligned}&E[\alpha_C(\|\boldsymbol{x}\|)^2]E[vec(\boldsymbol{u}\boldsymbol{u}^T)vec(\boldsymbol{u}\boldsymbol{u}^T)^T]\\&-E[\alpha_C(\|\boldsymbol{x}\|)\beta_C(\|\boldsymbol{x}\|)]E[vec(\boldsymbol{u}\boldsymbol{u}^T)vec(I_k)^T+vec(I_k)vec(\boldsymbol{u}\boldsymbol{u}^T)^T]\\&+E[\beta_C(\|\boldsymbol{x}\|)^2]vec(I_k)vec(I_k)^T.\end{aligned}$$

Since $E[\boldsymbol{u}\boldsymbol{u}^T]=k^{-1}I_k$ and

$$E[vec(\boldsymbol{u}\boldsymbol{u}^T)vec(\boldsymbol{u}\boldsymbol{u}^T)^T]=(k(k+2))^{-1}(I_{k^2}+I_{k,k}+vec(I_k)vec(I_k)^T),$$

the covariance matrix is equal to

$$\begin{aligned}&\frac{1}{k(k+2)}E\left[\alpha_C(\|\boldsymbol{x}\|)^2\right](I_{k^2}+I_{k,k})\\&+E\left[\left(\frac{1}{k}\alpha_C(\|\boldsymbol{x}\|)-\beta_C(\|\boldsymbol{x}\|)\right)^2-\frac{2}{k}\left(\frac{1}{k(k+2)}\alpha_C(\|\boldsymbol{x}\|)^2\right)\right]vec(I_k)vec(I_k)^T\end{aligned}$$

Proof of Theorem 5. The result follows from Theorem 1 in Tyler [19]. Write $\tau_1=ASV(\widehat{C}_{12};F_0)=ASV(\widehat{V}_{12};F_0)$ and let $V(C)=k\,vec(C)/Tr(C)$. Then $V(aC)=V(C)$ for all $a>0$ and as $V'(C)=\frac{1}{2}\{dV(C)/dvec(V)\}(I_{k^2}+J_k)$, where $J_k=\sum_{i=1}^k \boldsymbol{e}_i\boldsymbol{e}_i^T\otimes\boldsymbol{e}_i\boldsymbol{e}_i^T$, one has that

$$\begin{aligned}V'(C)&=\frac{1}{2}\left[\frac{k}{Tr(C)}(I_{k^2}+I_{k,k}+J_k)-\frac{k}{Tr^2(C)}vec(C)vec(I_k)^T\right](I_{k^2}+J_k)\\&=\frac{k}{2\,Tr(C)}\left[I_{k^2}-\frac{1}{k}vec(V)vec(I_k)^T\right](I_{k^2}+I_{k,k})=:\frac{k}{2\,Tr(C)}W(I_{k^2}+I_{k,k}).\end{aligned}$$

Now Theorem 1 in [19] implies that the limiting distribution of $\sqrt{n}\,vec(\widehat{V}-V)$ is multivariate normal with mean zero and asymptotic covariance matrix

$$\begin{aligned}&ASC\{\sqrt{n}\,vec(\widehat{V}-V)\}=2\tau_1\{V'(C)\}(V\otimes V)\{V'(C)\}^T\\&=\frac{\tau_1k^2}{2\,Tr^2(C)}W(I_{k^2}+I_{k,k})(C\otimes C)(I_{k^2}+I_{k,k})W^T\\&=\frac{\tau_1}{2}W(I_{k^2}+I_{k,k})(V\otimes V)(I_{k^2}+I_{k,k})W^T=\tau_1W(I_{k^2}+K_k)(V\otimes V)W^T.\end{aligned}$$

The last equality follows, since $(V\otimes V)I_{k,k}=I_{k,k}(V\otimes V)$ and $(I_{k^2}+I_{k,k})^2=2(I_{k^2}+I_{k,k})$.