The Annals of Statistics
1999, Vol. 27, No. 5, 1638–1665

ASYMPTOTICS OF REWEIGHTED ESTIMATORS OF MULTIVARIATE LOCATION AND SCATTER

BY HENDRIK P. LOPUHAÄ

Delft University of Technology

We investigate the asymptotic behavior of a weighted sample mean and covariance, where the weights are determined by the Mahalanobis distances with respect to initial robust estimators. We derive an explicit expansion for the weighted estimators. From this expansion it can be seen that reweighting does not improve the rate of convergence of the initial estimators. We also show that if one uses smooth S-estimators to determine the weights, the weighted estimators are asymptotically normal. Finally, we will compare the efficiency and local robustness of the reweighted S-estimators with two other improvements of S-estimators: τ-estimators and constrained M-estimators.

1. Introduction.

Let $X_1, X_2, \ldots$ be independent random vectors with a distribution $P_{\mu,\Sigma}$ on $\mathbb{R}^k$, which is assumed to have a density

$$f(x) = |\Sigma|^{-1/2} h\big((x-\mu)^\top \Sigma^{-1}(x-\mu)\big), \tag{1.1}$$

where $\mu \in \mathbb{R}^k$, $\Sigma \in \mathrm{PDS}(k)$, the class of positive definite symmetric matrices of order k, and $h\colon [0,\infty) \to [0,\infty)$ is assumed to be known. Suppose we want to estimate (μ, Σ). The sample mean and sample covariance may provide accurate estimates, but they are also notorious for being sensitive to outlying points. Robust estimates M_n and V_n may protect us against outlying observations, but these estimates will not be very accurate in case no outlying observations are present.

Two concepts that reflect to some extent the sensitivity of estimators are the finite sample breakdown point and the influence function, whereas the asymptotic efficiency may give some indication of how accurate the estimators are. The finite sample (replacement) breakdown point [Hampel (1968), Donoho and Huber (1983)] is roughly the smallest fraction of outliers that can take the estimate over all bounds. It must be seen as a global measure of robustness as opposed to the influence function [Hampel (1968), Hampel (1974)] as a local measure which measures the influence of an infinitesimal pertubation at a point x on the estimate. Affine equivariant M-estimators [Maronna (1976)] are robust alternatives to the sample mean and covariance, defined as the

Received January 1998; revised May 1999.
AMS 1991 *subject classifications*. 62E20, 62F12, 62F35, 62H10, 62H12.
Key words and phrases. Robust estimation of multivariate location and covariance, reweighted least squares, application of empirical process theory.

solution $\theta_n = (T_n, C_n)$ of

$$\sum_{i=1}^{n} \Psi(X_i, \theta) = 0, \tag{1.2}$$

where Ψ attains values in $\mathbb{R}^k \times \mathrm{PDS}(k)$. They have a bounded influence function and a high efficiency. Their breakdown point, however, is at most $1/(k+1)$, due to the increasing sensitivity of covariance M-estimators to outliers contained in lower-dimensional hyperplanes as k gets larger [Tyler (1986)].

Affine equivariant estimators with a high breakdown point where introduced by Stahel (1981), Donoho (1982) and Rousseeuw (1985). The Stahel–Donoho estimator converges at rate $n^{1/2}$ [Maronna and Yohai (1995)], however, the limiting distribution is yet unknown, and Rousseeuw's minimum volume ellipsoid (MVE) estimator converges at rate $n^{1/3}$ to a nonnormal limiting distribution [Kim and Pollard (1990), Davies (1992a)]. Multivariate S-estimators [Davies (1987), Lopuhaä (1989)] are smoothed versions of the MVE estimator, which do converge at rate $\sqrt{n}$ to a limiting normal distribution. Nevertheless, one still has to make a tradeoff between breakdown point and asymptotic efficiency. Further extensions of S-estimators, such as τ-estimators [Lopuhaä (1991)] and constrained M (CM)-estimators [Kent and Tyler (1997)], are able to avoid this tradeoff.

Several procedures have been proposed that combine M-estimators together with a high breakdown estimator with bounded influence [see, e.g., Yohai (1987), Lopuhaä (1989)]. Unfortunately, a similar approach for covariance estimators would fail because of the low breakdown point of covariance M-estimators. Another approach is to perform a one-step "Gauss–Newton" approximation to (1.2),

$$\theta_n = \theta_{0,n} + \left[\sum_{i=1}^{n} D\Psi(X_i, \theta_{0,n})\right]^{-1} \sum_{i=1}^{n} \Psi(X_i, \theta_{0,n}),$$

starting from an initial estimator $\theta_{0,n} = (M_n, V_n)$ with high breakdown point and bounded influence [Bickel (1975), Davies (1992b)]. Such a procedure improves the rate of convergence and the one-step estimator has the same limiting behavior as the solution of (1.2). The breakdown behavior is, however, unknown and might be as poor as that of the covariance M-estimator.

Rousseeuw and Leroy (1987) proposed to use the MVE estimator, omit observations whose Mahalanobis distance with respect to this estimator exceeds some cut-off value and compute the sample mean and sample covariance of the remaining observations. This method looks appealing and found his way for instance in the area of computer vision [see Jolion, Meer and Bataouche (1991) and Matei, Meer and Tyler (1998) for an application of the reweighted MVE, and also Meer, Mintz, Rosenfeld and Kim (1991) for an application of a similar procedure in the regression context]. In Lopuhaä and Rousseeuw (1991) it is shown that such a procedure preserves affine equivariance and the breakdown point of the initial estimators. It can also be seen that reweighting has close connections to (1.2) (see Remark 2.1). It is therefore natural to question whether one-step reweighting also improves the rate of convergence and what the limit behavior of the reweighted estimators is.

We will derive an explicit asymptotic expansion for the reweighted estimators. From this expansion it can be seen immediately that the reweighted estimators converge at the same rate as the initial estimators. A similar result in the regression context can be found in He and Portnoy (1992). We will also show that, if smooth S-estimators are used to determine the weights, the reweighted estimators are $n^{1/2}$ consistent and asymptotically normal. Similar to τ-estimators and CM-estimators, reweighted S-estimators are able to avoid the tradeoff between asymptotic efficiency and breakdown point. However, with all three methods there still remains a tradeoff between efficiency and local robustness. In the last section we will compare the efficiency together with the local robustness of reweighted S-estimators with those of the τ-estimators and the CM-estimators.

2. Definitions. Let $M_n \in \mathbb{R}^k$ and $V_n \in \mathrm{PDS}(k)$ denote (robust) estimators of location and covariance. Estimators M_n and V_n are called *affine equivariant* if,

$$M_n(AX_1+b,\dots,AX_n+b) = AM_n(X_1,\dots,X_n)+b,$$
$$V_n(AX_1+b,\dots,AX_n+b) = AV_n(X_1,\dots,X_n)A^\top,$$

for all nonsingular $k\times k$ matrices A and $b\in\mathbb{R}^k$. We will use M_n and V_n as a diagnostic tool to identify outlying observations, rather than using them as actual estimators of location and covariance. If we think of robust estimators M_n and V_n as reflecting the bulk of the data, then outlying observations X_i will have a large squared Mahalanobis distance,

$$d^2(X_i, M_n, V_n) = (X_i - M_n)^\top V_n^{-1}(X_i - M_n), \tag{2.1}$$

compared to the distances of those observations that belong to the majority. Once the outliers have been identified, one could compute a weighted sample mean and covariance to obtain more accurate estimates. Observations with large $d^2(X_i, M_n, V_n)$ can then be given a smaller weight or, even more drastically, one could assign weight 0 to X_i whenever $d^2(X_i, M_n, V_n)$ exceeds some kind of threshold value $c>0$.

Therefore, let $w\colon [0,\infty)\to[0,\infty)$ be a (weight) function, that satisfies

(W) w is bounded and of bounded variation, and almost everywhere continuous on $[0,\infty)$.

Define a weighted sample mean and covariance as follows

$$T_n = \frac{\sum_{i=1}^n w\big(d^2(X_i, M_n, V_n)\big)X_i}{\sum_{i=1}^n w\big(d^2(X_i, M_n, V_n)\big)} \tag{2.2}$$

$$C_n = \frac{\sum_{i=1}^n w\big(d^2(X_i, M_n, V_n)\big)(X_i - T_n)(X_i - T_n)^\top}{\sum_{i=1}^n w\big(d^2(X_i, M_n, V_n)\big)}. \tag{2.3}$$

A typical choice for w would be the function

$$w(y) = 1_{[0,c]}(y), \tag{2.4}$$

in which case T_n and C_n are simply the sample mean and sample covariance of the X_i with $d^2(X_i, M_n, V_n) \le c$. Note that (W) also permits $w \equiv 1$, in which case T_n and C_n are the ordinary sample mean and covariance matrix based on all observations.

Under additional restrictions on the function w, the finite sample breakdown point of M_n and V_n is preserved [Lopuhaä and Rousseeuw (1991)]. Typical examples for (M_n, V_n) are the MVE estimators and S-estimators. If M_n and V_n are affine equivariant it is easy to see that, for each $i = 1, \dots, n$, the Mahalanobis distance $d^2(X_i, M_n, V_n)$ is invariant under affine transformations of X_i. This means that affine equivariance of M_n and V_n carries over to the weighted estimators T_n and C_n.

REMARK 2.1. Consider the following score equations for multivariate M-estimators:

$$\sum_{i=1}^{n} w\big(d^2(X_i, t, C)\big)(X_i - t) = 0,$$

$$\sum_{i=1}^{n} w\big(d^2(X_i, t, C)\big)\big[(X_i - t)(X_i - t)^\top - C\big] = 0.$$

If we would replace (M_n, V_n) by (T_n, C_n) in (2.2) and (2.3), then (T_n, C_n) would be a fixed point of the above M-score equations. Hence the reweighted estimators can be seen as a one-step iteration towards the fixed point of the M-score equations.

We investigate the asymptotic behavior of T_n and C_n, as $n \to \infty$, under the location-scale model (1.1). This means that most of the constants that will follow can be rewritten by application of the following lemma [see Lopuhaä (1997)].

LEMMA 2.1. *Let* $z\colon [0, \infty) \to \mathbb{R}$ *and write* $x = (x_1, \dots, x_k)^\top$. *Then*

$$\int z(x^\top x)\, dx = \frac{2\pi^{k/2}}{\Gamma(k/2)} \int_0^\infty z(r^2) r^{k-1} dr,$$

$$\int z(x^\top x) x_i^2 dx = \frac{1}{k} \int z(x^\top x)(x^\top x)\, dx,$$

$$\int z(x^\top x) x_i^2 x_j^2 dx = \frac{1 + 2\delta_{ij}}{k(k+2)} \int z(x^\top x)(x^\top x)^2 dx,$$

for $i, j = 1, \dots, k$, *where* δ_{ij} *denotes the Kronecker delta.*

In order to avoid smoothness conditions on the function w, we assume some smoothness of the function h:

(H1) h is continuously differentiable.

We also need that h has a finite fourth moment:

(H2) $\int (x^\top x)^2 h(x^\top x)\, dx < \infty.$

This is a natural condition, which, for instance, is needed to obtain a central limit theorem for C_n. Note that by Lemma 2.1, condition (H2) implies that

$$\int_0^\infty h(r^2) r^{k-1+j} dr < \infty \quad \text{for } j = 0, 1, \ldots, 4. \tag{2.5}$$

Finally we will assume that the initial estimators M_n and V_n are affine equivariant and consistent, that is, $(M_n, V_n) \to (\mu, \Sigma)$ in probability.

Write $\Theta = \mathbb{R}^k \times \text{PDS}(k)$, $\theta = (m, V)$ and $d(x, \theta) = d(x, m, V)$. Multiplying the numerator and denominator in (2.2) and (2.3) by $1/n$ leaves T_n and C_n unchanged. This means that if we define

$$\begin{aligned}
\boldsymbol{\Psi}_1(x, \theta) &= w(d(x, \theta)),\\
\boldsymbol{\Psi}_2(x, \theta) &= w(d(x, \theta))x,\\
\boldsymbol{\Psi}_3(x, \theta, t) &= w(d(x, \theta))(x - t)(x - t)^\top,
\end{aligned}$$

and write $\theta_n = (M_n, V_n)$, then T_n and C_n can be written as

$$T_n = \frac{\int \boldsymbol{\Psi}_2(x, \theta_n)\, dP_n(x)}{\int \boldsymbol{\Psi}_1(x, \theta_n)\, dP_n(x)},$$

$$C_n = \frac{\int \boldsymbol{\Psi}_3(x, \theta_n, T_n)\, dP_n(x)}{\int \boldsymbol{\Psi}_1(x, \theta_n)\, dP_n(x)},$$

where P_n denotes the empirical measure corresponding to $X_1, X_2, \ldots, X_n$.

For each of the functions $\boldsymbol{\Psi}_j$, $j = 1, 2$ we can write

$$\begin{aligned}
\int \boldsymbol{\Psi}_j(x, \theta_n)\, dP_n(x) &= \int \boldsymbol{\Psi}_j(x, \theta_n)\, dP(x) + \int \boldsymbol{\Psi}_j(x, \theta_0)\, d(P_n - P)(x)\\
&\quad + \int \Big(\boldsymbol{\Psi}_j(x, \theta_n) - \boldsymbol{\Psi}_j(x, \theta_0)\Big)\, d(P_n - P)(x),
\end{aligned} \tag{2.6}$$

where $\theta_0 = (\mu, \Sigma)$. From here we can proceed as follows. The first term on the right-hand side can be approximated by a first-order Taylor expansion which is linear in $\theta_n - \theta_0$. The second term can be treated by the central limit theorem. The third term contains most of the difficulties but is shown to be of smaller order. For this we will use results from empirical process theory as treated in Pollard (1984). A similar decomposition holds for $\boldsymbol{\Psi}_3$.

We will first restrict ourselves to the case $(\mu, \Sigma) = (0, I)$, that is, $f(x) = h(x^\top x)$ and $(M_n, V_n) \to (0, I)$ in probability. In that case it is more convenient to reparametrize things and to write $V = (I + A)^2$, so that V_n can be written as

$$V_n = (I + A_n)^2 \quad \text{with } \|A_n\| = o_P(1),$$

where, throughout the paper, $\|\cdot\|$ will denote Euclidean norm. In order to obtain the linear Taylor approximations for the first term, we define for $j = 1, 2$,

$$\lambda_{j,P}(m, A) = \int \boldsymbol{\Psi}_j(x, m, (I + A)^2)\, dP(x)$$

and

$$\lambda_{3,P}(m, A, t) = \int \boldsymbol{\Psi}_3(x, m, (I + A)^2, t)\, dP(x).$$

Then the first term on the right-hand side of (2.6) can be written as

$$\int \boldsymbol{\Psi}_j(x, \theta_n)\, dP(x) = \lambda_{j,P}(M_n, A_n),$$

where $(M_n, A_n) \to (0, 0)$ in probability. We will first investigate the expansions of $\lambda_{j,P}(m, A)$, $j = 1, 2$, as $(m, A) \to (0, 0)$, and $\lambda_{3,P}(m, A, t)$, as $(m, A, t) \to (0, 0, 0)$.

3. Expansions of $\boldsymbol{\lambda}_{j,P}$. Denote by $\mathrm{tr}(A)$ the trace of a square matrix A. The following lemma gives the expansions of $\lambda_{j,P}$, as $m \to 0$, $A \to 0$ and $t \to 0$.

LEMMA 3.1. *Let w satisfy* (W) *and let $f(x) = h(x^\top x)$ satisfy* (H1) *and* (H2). *Then the following hold*:

(i) *As* $(m, A) \to (0, 0)$,

$$\lambda_{1,P}(m, A) = c_1 + c_0 \mathrm{tr}(A) + o(\|(m, A)\|),$$
$$\lambda_{2,P}(m, A) = c_2 m + o(\|(m, A)\|).$$

(ii) *As* $(m, A, t) \to (0, 0, 0)$,

$$\lambda_{3,P}(m, A, t) = c_3 I + c_4 \{\mathrm{tr}(A) I + 2A\} + o(\|(m, A, t)\|).$$

The constants are given by

$$c_0 = \frac{2\pi^{k/2}}{\Gamma(k/2)} \int_0^\infty \frac{2}{k} w(r^2) h'(r^2) r^{k+1}\, dr, \tag{3.1}$$

$$c_1 = \frac{2\pi^{k/2}}{\Gamma(k/2)} \int_0^\infty w(r^2) h(r^2) r^{k-1}\, dr > 0, \tag{3.2}$$

(3.3) $$c_2 = \frac{2\pi^{k/2}}{\Gamma(k/2)} \int_0^\infty w(r^2)\left[h(r^2) + \frac{2}{k}h'(r^2)r^2\right] r^{k-1}dr,$$

(3.4) $$c_3 = \frac{2\pi^{k/2}}{\Gamma(k/2)} \int_0^\infty \frac{1}{k} w(r^2)h(r^2)r^{k+1}dr > 0,$$

(3.5) $$c_4 = \frac{2\pi^{k/2}}{\Gamma(k/2)} \int_0^\infty w(r^2)\left[\frac{r^2}{k}h(r^2) + \frac{2r^4}{k(k+2)}h'(r^2)\right] r^{k-1}dr.$$

REMARK 3.1. Note that in case $w \equiv 1$, the constants are given by $c_0 = -1$, $c_1 = 1$, $c_2, c_4 = 0$ and $c_3 = E\|X_1\|^2/k$.

PROOF. First note that because w is bounded, property (2.5) together with partial integration implies that the constants c_0, c_1, c_2, c_3 and c_4 are finite.

(i) After transformation of coordinates, for $\lambda_{1,P}$ we may write

$$\lambda_{1,P}(m, A) = |I + A| \int w(x^\top x) f(m + x + Ax)\, dx.$$

Note that

(3.6) $$|I + A| = 1 + \mathrm{tr}(A) + o(\|A\|) \quad \text{for } A \to 0.$$

The derivative of $\phi_1(x, m, A) = f(m + x + Ax)$ with respect to (m, A) at a point (m_0, A_0) is the linear map

$$D\phi_1(x, m_0, A_0)\colon (m, A) \mapsto 2h'\big(\|m_0 + x + A_0x\|^2\big)(m_0 + x + A_0x)^\top(m + Ax)$$

[see Dieudonné (1969), Chapter 8]. The conditions on f imply that $D\phi_1(x, 0, 0)$ is continuous. Therefore by Taylor's formula,

$$\phi_1(x, m, A) = \phi_1(x, 0, 0) + D\phi_1(x, 0, 0)(m, A) + \int_0^1 [D\phi_1(x, \zeta m, \zeta A) - D\phi_1(x, 0, 0)]\, d\zeta (m, A).$$

Together with (3.6) this means that

$$\lambda_{1,P}(m, A) = (1 + \mathrm{tr}(A))\lambda_{1,P}(0, 0) + \int w(x^\top x) D\phi_1(x, 0, 0)(m, A)\, dx + R_1(m, A) + o(\|A\|),$$

where

$$R_1(m, A) = \int w(x^\top x) \int_0^1 [D\phi_1(x, \zeta m, \zeta A) - D\phi_1(x, 0, 0)](m, A) d\zeta dx.$$

According to Lemma 2.1, we have $\lambda_{1,P}(0, 0) = c_1$. By symmetry and the fact that $x^\top Ax = \mathrm{tr}(Axx^\top)$, it follows from Lemma 2.1 that

$$\int w(x^\top x) D\phi_1(x, 0, 0)(m, A)\, dx = 2\mathrm{tr}\left[A \int w(x^\top x)h'(x^\top x)xx^\top dx\right] = c_0 \mathrm{tr}(A).$$

Obviously $o(\|A\|) = o(\|(m, A)\|)$ and that $R_1(m, A) = o(\|(m, A)\|)$ can be seen as follows. Write $R_1(m, A)$ as

$$\begin{aligned}&2\int w(x^\top x)\int_0^1 h'(\|\zeta m + x + \zeta Ax\|^2)(\zeta m + x + \zeta Ax)^\top (m + Ax)\, d\zeta dx\\&-2\int w(x^\top x)h'(\|x\|^2)x^\top (m + Ax)\, dx.\end{aligned} \tag{3.7}$$

By a change of variables $y = \zeta m + x + \zeta Ax$ in (3.7),

$$R_1(m, A) = 2\int\int_0^1 h'(y^\top y)y^\top r_{1,\zeta}(y, m, A)\, d\zeta dy,$$

where

$$\begin{aligned}r_{1,\zeta}(y, m, A) = {}&|I + \zeta A|^{-1}w\big(\|(I + \zeta A)^{-1}(y - \zeta m)\|^2\big)\big(m + A(I + \zeta A)^{-1}(y - \zeta m)\big)\\&-w(\|y\|^2)(m + Ay).\end{aligned}$$

Note that for $\|A\|$ sufficiently small, $|I + \zeta A| \geq \frac{1}{2}$ and (2.5) implies $\int h'(y^\top y)\|y\|^2 dy < \infty$. Therefore, since w is bounded and a.e. continuous, it follows by dominated convergence that $R_1(m, A) = o(\|m\|) + o(\|A\|) = o(\|(m, A)\|)$. This proves the first part of (i). Similarly for $\lambda_{2,P}$, we may write

$$\lambda_{2,P}(m, A) = |I + A|\int w(x^\top x)(m + x + Ax)f(m + x + Ax)dx.$$

The derivative of $\phi_2(x, m, A) = (m + x + Ax)f(m + x + Ax)$ with respect to (m, A) at $(0, 0)$ is the continuous linear map

$$D\phi_2(x, 0, 0)\colon (m, A) \mapsto \big[h(x^\top x) + 2h'(x^\top x)xx^\top\big](m + Ax).$$

Note that by symmetry $\lambda_{2,P}(0, 0) = \int w(x^\top x)h(x^\top x)x dx = 0$. Hence, similarly to the reasoning above, it follows that

$$\lambda_{2,P}(m, A) = \int w(x^\top x)\big[h(x^\top x) + 2h'(x^\top x)xx^\top\big]m dx + R_2(m, A) + o(\|A\|),$$

where

$$R_2(m, A) = \int w(x^\top x)\int_0^1 [D\phi_2(x, \zeta m, \zeta A) - D\phi_2(x, 0, 0)](m, A)d\zeta dx.$$

According to Lemma 2.1,

$$\int w(x^\top x)\big[h(x^\top x) + 2h'(x^\top x)xx^\top\big]m dx = c_2 m.$$

Similar to $R_1(m, A)$, using that (2.5) implies $\int f(y)\|y\|dy < \infty$ and $\int h'(y^\top y)\|y\|^3 dy < \infty$, it follows by dominated convergence that $R_2(m, A) = o(\|(m, A)\|)$.

(ii) Write $\lambda_{3,P}(m, A, t)$ as

$$|I + A|\int w(x^\top x)(m - t + x + Ax)(m - t + x + Ax)^\top f(m - t + x + Ax)\, dx.$$

The derivative of $\phi_3(m, A, t) = (m - t + x + Ax)(m - t + x + Ax)^\top f(m - t + x + Ax)$ with respect to (m, A, t) at $(0, 0, 0)$ is the continuous linear map $D\phi_3(x, 0, 0, 0)$:

$$(m, A, t) \mapsto h(x^\top x)(m - t + Ax)x^\top + h(x^\top x)x(m - t + Ax)^\top + 2h'(x^\top x)(x^\top Ax)xx^\top.$$

Similarly to the reasoning above, using (3.6), it follows that

$$\begin{aligned}\lambda_{3,P}(m, A, t) = (1 + \mathrm{tr}(A)) &\int w(x^\top x)h(x^\top x)xx^\top dx \\ &+ \int w(x^\top x)h(x^\top x)Axx^\top dx + \int w(x^\top x)h(x^\top x)xx^\top A dx \\ &+ 2\int w(x^\top x)h'(x^\top x)(x^\top Ax)xx^\top dx + R_3(m, A, t) + o(\|A\|),\end{aligned}$$

where

$$R_3(m, A, t) = \int w(x^\top x)\int_0^1 [D\phi_3(x, \zeta m, \zeta A, \zeta t) - D\phi_3(x, 0, 0, 0)](m, A, t)d\zeta dx.$$

Consider the (i, j)th entry of the fourth term on the right-hand side of $\lambda_{3,P}(m, A, t)$:

$$2\int w(x^\top x)h'(x^\top x)(x^\top Ax)x_i x_j dx. \tag{3.8}$$

When $i = j$, then (3.8) is equal to

$$2\int w(x^\top x)x_i^2(x_1^2 a_{11} + \cdots + x_i^2 a_{ii} + \cdots + x_p^2 a_{pp})h'(x^\top x)\, dx$$

and when $i \neq j$, then (3.8) is equal to

$$2\int w(x^\top x)h'(x^\top x)(x_i x_j a_{ij} + x_j x_i a_{ji})x_i x_j dx.$$

With Lemma 2.1 we find that for all $i, j = 1, \dots, k$ the (i, j)th entry (3.8) is equal to

$$2\int w(x^\top x)h(x^\top x)(x^\top x)^2 dx \frac{\delta_{ij}\mathrm{tr}(A) + 2a_{ij}}{k(k+2)}.$$

It follows that

$$\begin{aligned}\lambda_{3,P}(m, A, t) = (1 + \mathrm{tr}(A)) &\int w(x^\top x)h(x^\top x)xx^\top dx \\ &+ \int w(x^\top x)h(x^\top x)Axx^\top dx + \int w(x^\top x)h(x^\top x)xx^\top A dx \\ &+ \frac{2\mathrm{tr}(A)}{k(k+2)}\int w(x^\top x)h'(x^\top x)(x^\top x)^2 dx \cdot I \\ &+ \frac{4A}{k(k+2)}\int w(x^\top x)h'(x^\top x)(x^\top x)^2 dx + R_3(m, A, t) + o(\|A\|).\end{aligned}$$

By Lemma 2.1 we find that

$$\lambda_{3,P}(m, A, t) = c_3 I + c_4 \mathrm{tr}(A) I + 2c_4 A + R_3(m, A, t).$$

Similarly to $R_1(m, A)$ and $R_2(m, A)$, using that (2.5) implies

$$\int f(y)\|y\|^2 dy < \infty \quad \text{and} \quad \int h'(\|y\|^2)\|y\|^4 dy < \infty,$$

it follows by dominated convergence that $R_3(m, A, t) = o(\|(m, A, t)\|)$. □

4. Expansion of T_n and C_n. The main problem in obtaining the limiting behavior of T_n and C_n, is to bound the following expressions:

$$\sqrt{n} \int \Big(\mathbf{\Psi}_j(x, \theta_n) - \mathbf{\Psi}_j(x, \theta_0) \Big) d(P_n - P)(x) \quad \text{for } j = 1, 2, \tag{4.1}$$

$$\sqrt{n} \int \Big(\mathbf{\Psi}_3(x, \theta_n, T_n) - \mathbf{\Psi}_3(x, \theta_0, \mu) \Big) d(P_n - P)(x), \tag{4.2}$$

as $n \to \infty$, where $\theta_n = (M_n, V_n)$ and $\theta_0 = (\mu, \Sigma)$. For this we will use results from empirical process theory as treated in Pollard (1984). These results apply only to *real valued* functions, whereas the functions $\mathbf{\Psi}_2(x, \theta)$ and $\mathbf{\Psi}_3(x, \theta)$ are vector and matrix valued, respectively. This can easily be overcome by considering the real valued components individually.

LEMMA 4.1. *Let $\theta_n = (M_n, V_n)$ and $\theta_0 = (\mu, \Sigma) = (0, I)$. Suppose that w and h satisfy* (W) *and* (H2). *Then the following hold*:

(i) *If $\theta_n \to \theta_0$ in probability, then for $j = 1, 2$,*

$$\int \Big(\mathbf{\Psi}_j(x, \theta_n) - \mathbf{\Psi}_j(x, \theta_0) \Big) d(P_n - P)(x) = o_P(n^{-1/2}).$$

(ii) *If $\theta_n \to \theta_0$ in probability, and $T_n \to 0$ in probability, then*

$$\int \Big(\mathbf{\Psi}_3(x, \theta_n, T_n) - \mathbf{\Psi}_3(x, \theta_0, 0) \Big) d(P_n - P)(x) = o_P(n^{-1/2}).$$

PROOF. Consider the classes $\mathscr{F} = \{w(d(x, \theta)) \colon \theta \in \Theta\}$, $\mathscr{F}_j = \{w(d(x, \theta))x_j \colon \theta \in \Theta\}$ and $\mathscr{F}_{ij} = \{w(d(x, \theta))x_i x_j \colon \theta \in \Theta\}$, for $i, j = 1, \dots, k$. Denote by $\mathscr{G}$, $\mathscr{G}_j$ and $\mathscr{G}_{ij}$ the corresponding classes of graphs of functions in $\mathscr{F}$, $\mathscr{F}_j$ and $\mathscr{F}_{ij}$, respectively. Because w is of bounded variation, it follows from Lemma 3 in Lopuhaä (1997) that $\mathscr{G}$, $\mathscr{G}_j$ and $\mathscr{G}_{ij}$ all have polynomial discrimination for $i, j = 1, \dots, k$. Since w is bounded and h satisfies (H2), $\mathscr{F}$, $\mathscr{F}_j$ and $\mathscr{F}_{ij}$, all have square integrable envelopes. As $\theta_n \to \theta_0$ in probability, from Pollard (1984) we get that

$$\int \Big(w(d(x, \theta_n)) - w(d(x, \theta_0)) \Big) d(P_n - P)(x) = o_P(n^{-1/2}), \tag{4.3}$$

$$\int \Big(w(d(x,\theta_n)) - w(d(x,\theta_0)) \Big) x_i \, d(P_n - P)(x) = o_P(n^{-1/2}), \tag{4.4}$$

$$\int \Big(w(d(x,\theta_n)) - w(d(x,\theta_0)) \Big) x_i x_j \, d(P_n - P)(x) = o_P(n^{-1/2}). \tag{4.5}$$

for every $i, j = 1, 2, \ldots, k$. Case (i) follows directly from (4.3) and (4.4). For case (ii), split $\mathbf{\Psi}_3(x, \theta_n, T_n) - \mathbf{\Psi}_3(x, \theta_0, 0)$ into

$$\Big(w(d(x,\theta_n)) - w(d(x,\theta_0)) \Big) \left\{ xx^\top - xT_n^\top - T_n x^\top + T_n T_n^\top \right\}$$
$$+ w(d(x,\theta_0)) \left\{ xT_n^\top + T_n x^\top - T_n T_n^\top \right\}.$$

Note that by the central limit theorem, $\int w(d(x,\theta_0))\, d(P_n - P)(x) = O_P(n^{-1/2})$ and $\int w(d(x,\theta_0)) x d(P_n - P)(x) = O_P(n^{-1/2})$. Because w is bounded and continuous, and h satisfies (H2) and because $T_n \to 0$ in probability, together with (4.4) and (4.5), it follows that if we integrate with respect to $d(P_n - P)(x)$ all terms are $o_P(n^{-1/2})$, which proves (ii). □

We are now able to prove the following theorem, which describes the asymptotic behavior of T_n and C_n.

THEOREM 4.1. *Let $X_1, \ldots, X_n$ be independent with density $f(x) = h(x^\top x)$. Suppose that $w\colon [0,\infty) \to [0,\infty)$ satisfies* (W) *and h satisfies* (H1) *and* (H2). *Let M_n and $V_n = (I + A_n)^2$ be affine equivariant location and covariance estimators such that $(M_n, A_n) = o_P(1)$. Let T_n and C_n be defined by* (2.2) *and* (2.3). *Then*

$$T_n = \frac{c_2}{c_1} M_n + \frac{1}{nc_1} \sum_{i=1}^n w(X_i^\top X_i) X_i + o_P(1/\sqrt{n}) + o_P\big(\|(M_n, A_n)\|\big)$$

and

$$C_n = \frac{c_3}{c_1} I + \frac{c_4}{c_1} \{\mathrm{tr}(A_n) I + 2A_n\}$$
$$+ \frac{1}{nc_1} \sum_{i=1}^n \left\{ w(X_i^\top X_i) X_i X_i^\top - c_3 I \right\} + o_P(1/\sqrt{n}) + o_P\big(\|(M_n, A_n, T_n)\|\big),$$

where c_1, c_2, c_3 and c_4 are defined in (3.1), (3.3), (3.4) *and* (3.5).

PROOF. First consider the denominator of T_n and C_n, and write this as

$$\int \mathbf{\Psi}_1(x, \theta_n)\, dP_n(x) = \int \mathbf{\Psi}_1(x, \theta_n)\, dP(x) + \int \mathbf{\Psi}_1(x, \theta_0)\, d(P_n - P)(x)$$
$$+ \int \Big(\mathbf{\Psi}_1(x, \theta_n) - \mathbf{\Psi}_1(x, \theta_0) \Big)\, d(P_n - P)(x),$$

where $\theta_0 = (0, I)$. According to Lemma 3.1, the first term on the right-hand side is $c_1 + o_p(1)$. The second term on the right-hand side is $O_p(1/\sqrt{n})$, according to the central limit theorem. The third term is $o_p(1/\sqrt{n})$, according to Lemma 4.1. It follows that

$$\int \boldsymbol{\Psi}_1(x, \theta_n)\, dP_n(x) = c_1 + o_p(1). \tag{4.6}$$

Similarly, write the numerator of T_n as

$$\begin{aligned}\int \boldsymbol{\Psi}_2(x, \theta_n)\, dP_n(x) &= \int \boldsymbol{\Psi}_2(x, \theta_n)\, dP(x) + \int \boldsymbol{\Psi}_2(x, \theta_0)\, d(P_n - P)(x)\\ &\quad + \int \Big(\boldsymbol{\Psi}_2(x, \theta_n) - \boldsymbol{\Psi}_2(x, \theta_0)\Big)\, d(P_n - P)(x).\end{aligned}$$

According to Lemma 3.1, the first term on the right-hand side is $c_2 M_n + o_p(\|(M_n, A_n)\|)$ and the third term is $o_p(1/\sqrt{n})$, according to Lemma 4.1. The second term is equal to

$$\int \boldsymbol{\Psi}_2(x, \theta_0)\, d(P_n - P)(x) = \frac{1}{n} \sum_{i=1}^{n} w(X_i^\top X_i) X_i,$$

because by symmetry $Ew(X_1^\top X_1)X_1 = 0$. Together with (4.6) this proves the expansion for T_n. The argument for C_n is completely similar using that, according to Lemma 2.1,

$$Ew(X_1^\top X_1) X_1 X_1^\top = c_3 I$$

and that the expansion for T_n implies $T_n = o_P(1)$. □

The result for the general case with $X_1, \dots, X_n$ being a sample from $P_{\mu,\Sigma}$ follows immediately from Theorem 4.1, using affine equivariance of M_n and V_n and basic properties for positive definite symmetric matrices. For $\Sigma \in \mathrm{PDS}(k)$, write $\Sigma = B^2$, with $B \in \mathrm{PDS}(k)$, and write $V_n = B_n^2$, with $B_n \in \mathrm{PDS}(k)$.

COROLLARY 4.1. *Let $X_1, \dots, X_n$ be a sample from $P_{\mu,\Sigma}$. Suppose that w: $[0, \infty) \to [0, \infty)$ satisfies* (W) *and h satisfies* (H1) *and* (H2). *Let M_n and $V_n = B_n^2$ be affine equivariant location and covariance estimates such that $(M_n - \mu, B_n - B) = o_P(1)$. Let T_n and C_n be defined by* (2.2) *and* (2.3). *Then*

$$\begin{aligned}T_n &= \mu + \frac{c_2}{c_1}(M_n - \mu) + \frac{1}{nc_1} \sum_{i=1}^{n} w\left(d^2(X_i, \mu, \Sigma)\right)(X_i - \mu)\\ &\quad + o_P(1/\sqrt{n}) + o_P\big(\|(M_n - \mu, B_n - B)\|\big)\end{aligned}$$

and

$$\begin{aligned}C_n &= \frac{c_3}{c_1}\Sigma + \frac{c_4}{c_1}\left\{\mathrm{tr}\left(B^{-1}(B_n - B)\right)\Sigma + 2B^{-1}(B_n - B)\Sigma\right\}\\ &\quad + \frac{1}{nc_1} \sum_{i=1}^{n} \left\{w\left(d^2(X_i, \mu, \Sigma)\right)(X_i - \mu)(X_i - \mu)^\top - c_3\Sigma\right\}\end{aligned}$$

$$+o_P(1/\sqrt{n}) + o_P(\|(M_n - \mu, B_n - B, T_n - \mu)\|),$$

where c_1, c_2, c_3 *and* c_4 *are defined in* (3.2), (3.3), (3.4) *and* (3.5).

If w has a derivative with $w' < 0$, then from (3.3) and (3.5) it follows by partial integration that

$$c_2 = -\frac{4\pi^{k/2}}{k\Gamma(k/2)} \int_0^\infty w'(r^2)h(r^2)r^k dr > 0,$$

$$c_4 = -\frac{4\pi^{k/2}}{k(k+2)\Gamma(k/2)} \int_0^\infty w'(r^2)h(r^2)r^{k+3} dr > 0.$$

Similarly, for w as defined in (2.4), we have $c_2, c_4 > 0$. In these cases it follows immediately from Corollary 4.1 that if the initial estimators M_n and V_n converge to μ and Σ, respectively, at a rate slower than $\sqrt{n}$, the reweighted estimators T_n and C_n converge to μ and $(c_3/c_1)\Sigma$, respectively, at the same rate. A typical example might be to do reweighting on basis of MVE estimators of location and scatter. However, these estimators converge at rate $n^{1/3}$ [see Davies (1992a)]. Reweighting does not improve the rate of convergence.

On the other hand, note that the constants c_2/c_1 and c_4/c_3 can be interpreted as the relative efficiency of the (unbiased) reweighted estimators with respect to the initial estimators. From (3.2) and (3.3) it can be seen that at the multivariate normal, in which case $h'(y) = -\frac{1}{2}h(y)$, we always have

$$\frac{c_2}{c_1} = 1 - \frac{2\pi^{k/2}}{c_1 k\Gamma(k/2)} \int_0^\infty w(r^2)h(r^2)r^{k+1} dr < 1,$$

for nonnegative weight functions w.

If we take w as defined in (2.4), then by partial integration it follows that

$$\frac{c_2}{c_1} = \frac{2\pi^{k/2}}{c_1 k\Gamma(k/2)} h(c^2)c^k < 1,$$

for any unimodal distribution with $h(y)$ nonincreasing for $y \geq 0$. For c_4/c_3 we find similar behavior. For w as defined in (2.4) both ratios are plotted in Figure 1 as a function of the cutoff value c, at the standard normal (solid line) and at the symmetric contaminated normal (SCN) $(1-\varepsilon)N(\mu,\Sigma) + \varepsilon N(\mu, 9\Sigma)$ for $\varepsilon = 0.1$ (dotted), $\varepsilon = 0.3$ and $\varepsilon = 0.5$ (dashed).

We observe that reweighting leads to an important gain in efficiency despite the fact that there is no improvement in the rate of convergence. In order to end up with $\sqrt{n}$ consistent estimators T_n and C_n, we have to start with $\sqrt{n}$ consistent estimators M_n and V_n. For this one could use smooth S-estimators. The resulting limiting behavior is treated in the next section.

REMARK 4.1. The influence function IF for the reweighted estimators can be obtained in a similar way as the expansions in Corollary 4.1. A formal definition of the IF can be found in Hampel (1974). If the functionals corresponding to the initial estimators are affine equivariant, for the location-scale

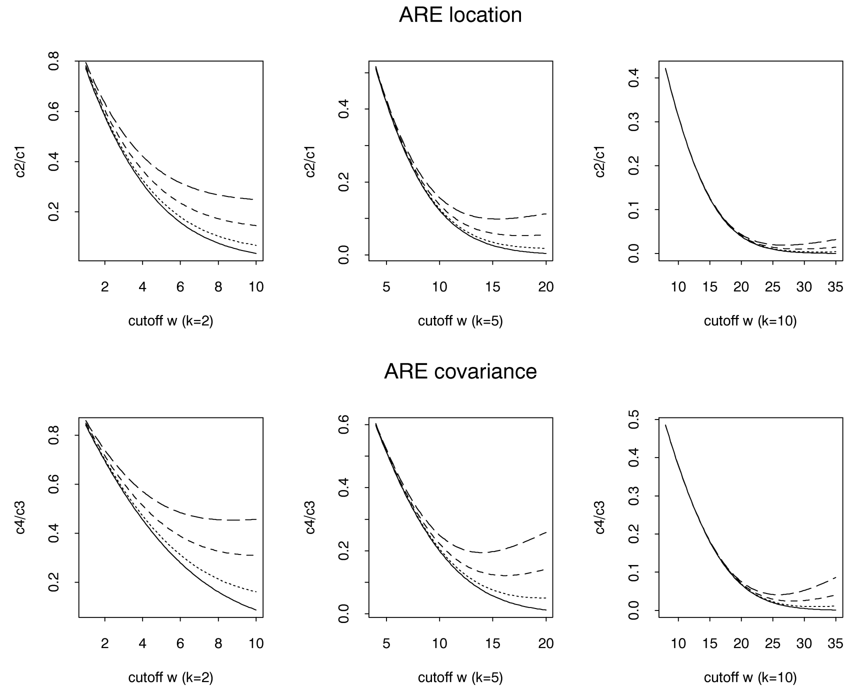

FIG. 1. *Ratios c_2/c_1 and c_4/c_3.*

model it suffices to give the IF at spherically symmetric distributions, that is, $(\mu, \Sigma) = (0, I)$. In that case one can show that

$$\mathrm{IF}(x; T, P) = \frac{c_2}{c_1}\mathrm{IF}(x; M, P) + \frac{w(x^\top x)x}{c_1},$$

$$\mathrm{IF}(x; C, P) = \frac{c_4}{c_1}\mathrm{IF}(x; V, P) + \frac{c_4}{2c_1}\mathrm{tr}\,(\mathrm{IF}(x; V, P))\, I + \frac{w(x^\top x)xx^\top - c_3 I}{c_1},$$

where $\mathrm{IF}(x; M, P)$ and $\mathrm{IF}(x; V, P)$ denote the influence functions of the initial estimators, and P has density f. Hence if $w(u^2)u^2$ is bounded, reweighting also preserves bounded influence of the initial estimators.

5. Reweighted *S*-estimators. Multivariate S-estimators are defined as the solution (M_n, V_n) to the problem of minimizing the determinant $|V|$ among all $m \in \mathbb{R}^k$ and $V \in \mathrm{PDS}(k)$ that satisfy

$$\frac{1}{n}\sum_{i=1}^{n} \rho(d(X_i, m, V)) \le b, \tag{5.1}$$

where $\rho\colon \mathbb{R} \to \mathbb{R}$ and $b \in \mathbb{R}$. These estimators arise as an extension of the MVE estimators ($\rho(y) = 1 - 1_{[0,c]}(y)$). With $\rho(y) = y^2$ and $b = k$ one obtains the LS estimators [see Grübel (1988)]. Results on properties of S-estimators and S-functionals can be found in Davies (1987) and Lopuhaä (1989). To inherit the high breakdown point of the MVE estimators as well as the limiting behavior of the LS estimators, that is, $\sqrt{n}$ rate of convergence and a limiting normal distribution, one must choose a bounded smooth function ρ. The exact asymptotic expansion of the S-estimator (M_n, V_n) is derived in Lopuhaä (1997). From this expansion a central limit theorem for reweighted S-estimators can easily be obtained.

To describe the limiting distribution of a random matrix, we consider the operator vec$(\cdot)$ which stacks the columns of a matrix M on top of each other, that is,

$$\text{vec}(M) = (M_{11}, \dots, M_{1k}, \dots, M_{k1}, \dots, M_{kk})^\top.$$

We will also need the commutation matrix $D_{k,k}$, which is a $k^2 \times k^2$ matrix consisting of $k \times k$ blocks: $D_{k,k} = (\Delta_{ij})_{i,j=1}^k$, where each (i, j)th block is equal to a $k \times k$-matrix Δ_{ji}, which is 1 at entry (j, i) and 0 everywhere else. By $A \otimes B$ we denote the Kronecker product of matrices A and B, which is a $k^2 \times k^2$ matrix with $k \times k$ blocks, the (i, j)th block equal to $a_{ij}B$.

THEOREM 5.1. *Let $X_1, \dots, X_n$ be a sample from $P_{\mu,\Sigma}$. Suppose that w: $[0, \infty) \to [0, \infty)$ satisfies* (W) *and h satisfies* (H1) *and* (H2). *Let (M_n, V_n) be S-estimators defined by* (5.1), *where ρ and b satisfy the conditions of Theorem* 2 *in Lopuhaä* (1997). *Then T_n and C_n are asymptotically independent, $\sqrt{n}(T_n - \mu)$ has a limiting normal distribution with zero mean and covariance matrix $\alpha\Sigma$, and $\sqrt{n}(C_n - (c_3/c_1)\Sigma)$ has a limiting normal distribution with zero mean and covariance matrix $\sigma_1(I + D_{k,k})(\Sigma \otimes \Sigma) + \sigma_2 \text{vec}(\Sigma)\text{vec}(\Sigma)$, where*

$$\alpha = \frac{2\pi^{k/2}}{\Gamma(k/2)} \int_0^\infty \frac{1}{k} a^2(r) h(r^2) r^{k+1} dr,$$

$$\sigma_1 = \frac{2\pi^{k/2}}{\Gamma(k/2)} \int_0^\infty \frac{1}{k(k+2)} l^2(r) h(r^2) r^{k-1} dr,$$

$$\sigma_2 = \frac{2\pi^{k/2}}{\Gamma(k/2)} \int_0^\infty \left[\frac{1}{k(k+2)} l^2(r) + m^2(r) + \frac{2}{k} l(r) m(r) \right] h(r^2) r^{k-1} dr,$$

with

$$a(u) = \frac{w(u^2)}{c_1} - \frac{c_2 \psi(u)}{c_1 \beta_2 u},$$

$$l(u) = \frac{w(u^2)u^2}{c_1} - \frac{2kc_4 \psi(u) u}{c_1 \beta_3},$$

$$m(u) = -\frac{c_3}{c_1} - \frac{(k+2)c_4(\rho(u) - b)}{kc_1 \beta_1} + \frac{2c_4 \psi(u) u}{c_1 \beta_3},$$

where ψ denotes the derivative of ρ, β_1, β_2, β_3 are given in Lemma 2 *of Lopuhaä* (1997) *and c_1, c_2, c_3 and c_4 are defined in* (3.2), (3.3), (3.4) *and* (3.5).

PROOF. First consider the case $(\mu, \Sigma) = (0, I)$, and write $V_n = (I + A_n)^2$. According to Theorem 2 in Lopuhaä (1997), the S-estimators admit the following expansions:

$$\begin{aligned}
\operatorname{tr}(A_n) &= -\frac{1}{n\beta_1}\sum_{i=1}^{n}\left\{\rho(\|X_i\|) - b\right\} + o_P(1/\sqrt{n}),\\
M_n &= -\frac{1}{n\beta_2}\sum_{i=1}^{n}\frac{\psi(\|X_i\|)}{\|X_i\|}X_i + o_P(1/\sqrt{n}),\\
A_n &= -\frac{1}{n}\sum_{i=1}^{n}\left[\frac{k\psi(\|X_i\|)}{\beta_3\|X_i\|}X_iX_i^{\top} + \left\{\frac{\rho(\|X_i\|) - b}{k\beta_1} - \frac{\psi(\|X_i\|)\|X_i\|}{\beta_3}\right\}I\right]\\
&\quad + o_P(1/\sqrt{n}),
\end{aligned}$$

where the constants β_1, β_2 and β_3 are defined in Lemma 2 in Lopuhaä (1997).

Together with the expansions given in Theorem 4.1, it follows immediately that

$$T_n = \frac{1}{n}\sum_{i=1} a(\|X_i\|)X_i + o_p(1/\sqrt{n}).$$

According to the central limit theorem, using boundedness of $w(u)$ and $\psi(u)u$ together with (H2), $\sqrt{n}T_n$ has a limiting normal distribution with zero mean and covariance matrix

$$Ea^2(\|X_1\|)X_1X_1^{\top} = \alpha I,$$

according to Lemma 2.1. Similarly for C_n we get that

$$C_n - \frac{c_3}{c_1}I = \frac{1}{n}\sum_{i=1}^{n}\left[l(\|X_i\|)\frac{X_iX_i^{\top}}{\|X_i\|^2} + m(\|X_i\|)I\right] + o_p(1/\sqrt{n}).$$

The conditions imposed on b imply that $E\rho(\|X_1\|) = b$, so that $l(\cdot)$ and $m(\cdot)$ satisfy

$$E\left[l(\|X_1) + km(\|X_1\|)\right] = 0.$$

From Lemma 5 in Lopuhaä (1997), again using boundedness of $w(u)$, $\psi(u)u$ and $\rho(u)$ together with (H2), it then follows that $\sqrt{n}(C_n - (c_3/c_1)I)$ has a limiting normal distribution with zero mean and covariance matrix $\sigma_1(I + D_{k,k}) + \sigma_2\operatorname{vec}(I)\operatorname{vec}(I)$.

That T_n and C_n are asymptotically independent can be seen as follows. If we write $X_1 = (X_{11}, \ldots, X_{1k})$, then the limiting covariance between an element of the vector $\sqrt{n}T_n$ and an element of the matrix $\sqrt{n}(C_n - (c_3/c_1)I)$ is given by

$$E\left[a(\|X_1\|)X_{1i}\Big(l(\|X_1\|)X_{1s}X_{1t} + m(\|X_1\|)\delta_{st}\Big)\right] \tag{5.2}$$

for $i = 1, \ldots, k$ and $s, t = 1, \ldots, k$. Hence by symmetry, (5.2) is always equal to zero, which implies that T_n and C_n are asymptotically uncorrelated. From the expansions for T_n and C_n it follows again by means of the central limit theorem, that the vector $\sqrt{n}(T_n, \text{vec}(C_n - (c_3/c_1)I)$ is asymptotically normal, so that T_n and C_n are also asymptotically independent.

Next consider the general case where $X_1, \dots, X_n$ are independent with distribution $P_{\mu,\Sigma}$, where $\Sigma = B^2$. Because of affine equivariance it follows immediately that $\sqrt{n}(T_n - \mu)$ converges to a normal distribution with zero mean and covariance matrix $B(\alpha I)B = \alpha\Sigma$, and $\sqrt{n}(C_n - (c_3/c_1)\Sigma)$ converges to a normal distribution with mean zero and covariance matrix $E\text{vec}(BMB)\text{vec}(BMB)^\top$, where M is the random matrix satisfying $E\text{vec}(M)\text{vec}(M)^\top = \sigma_1(I + D_{k,k}) + \sigma_2\text{vec}(I)\text{vec})I)^\top$. It follows from Lemma 5.2 in Lopuhaä (1989) that

$$E\text{vec}(BMB)\text{vec}(BMB)^\top = \sigma_1(I + D_{k,k})(\Sigma \otimes \Sigma) + \sigma_2\text{vec}(\Sigma)\text{vec}(\Sigma)^\top. \qquad \square$$

REMARK 5.1. When b in Theorem 5.1 is different from $b_h = \int \rho(\|x\|) h(x^\top x)\, dx$, the location S-estimator M_n still converges to μ, whereas the covariance S-estimator V_n converges to a multiple of Σ. In that case it can be deduced by similar arguments that $\sqrt{n}(T_n - \mu)$ and $\sqrt{n}(C_n - \gamma\Sigma)$ are still asymptotically normal with the same parameters α, σ_1, where

$$\gamma = \frac{c_3}{c_1} + \frac{c_4(k+2)(b - b_h)}{kc_1\beta_1}.$$

6. Comparison with other improvements of S-estimators. In this section we will investigate the efficiency and robustness of the reweighted S-estimator and compare this with two other improvements of S-estimators: τ-estimators proposed in Lopuhaä (1991) and CM-estimators proposed by Kent and Tyler (1997). We follow the approach taken by Kent and Tyler (1997), who consider an asymptotic index for the variance for the location and covariance estimator separately and an index for the local robustness also for the location and covariance estimator separately. For the underlying distribution we will consider the multivariate normal (NOR) distribution $N(\mu, \Sigma)$ and the symmetric contaminated normal (SCN) distribution $(1 - \varepsilon)N(\mu, \Sigma) + \varepsilon N(\mu, 9\Sigma)$ for $\varepsilon = 0.1, 0.3$ and 0.5.

The asymptotic variance of the location estimators is of the type $\alpha\Sigma$. We will compare the efficiency of the location estimators by comparing the corresponding values of the scalar α. To compare the local robustness of the location estimators, we compare the corresponding values of the gross-error-sensitivity (GES) defined to be

$$G_1 = \sup_{x \in \mathbb{R}^k} \|\text{IF}(x; P)\|,$$

where IF denotes the influence function of the corresponding location functionals at P.

The asymptotic variance of the covariance estimators is of the type

$$\sigma_1(I + D_{k,k})(\Sigma \otimes \Sigma) + \sigma_2\text{vec}(\Sigma)\text{vec}(\Sigma). \tag{6.1}$$

Kent and Tyler (1997) argue that the asymptotic variance of any *shape* component of the covariance estimator only depends on the asymptotic variance of the covariance estimators via the scalar σ_1. A shape component of a matrix C is any function $H(C)$ that satisfies $H(\lambda C) = H(C)$, $\lambda > 0$. We will compare the efficiency of the covariance estimators by comparing the corresponding values of the scalar σ_1. Note that if C_n is asymptotically normal with asymptotic variance of type (6.1), also λC_n is asymptotically normal with asymptotic variance of type (6.1) with the same value for σ_1. Kent and Tyler (1997) also motivate a single scalar G_2 for the local robustness of a covariance estimator and show that

$$G_2 = \frac{\mathrm{GES}(C;P)}{\left(1+\frac{2}{k}\right)\left(1-\frac{1}{k}\right)^{1/2}},$$

where $\mathrm{GES}(C;P)$ is the gross-error-sensitivity of the functional

$$\frac{C(P)}{\mathrm{trace}(C(P))}, \tag{6.2}$$

which is a shape component for the covariance functional $C(P)$. We will compare the robustness of the covariance estimators by comparing the corresponding values of the scalar G_2. Note that since (6.2) is a shape component for $C(P)$, the values of G_2 for $C(P)$ and $\lambda C(P)$ are the same.

6.1. *Reweighted biweight S-estimator.* For the reweighted S-estimator we define the initial S-estimator by

$$\rho(u) = \begin{cases} \dfrac{y^2}{2} - \dfrac{y^4}{2c^2} + \dfrac{y^6}{6c^4}, & |y| \le c, \\[2ex] \dfrac{c^2}{6}, & |y| > c. \end{cases} \tag{6.3}$$

Its derivative $\psi(y) = \rho'(y)$ is known as Tukey's biweight function. We take $b = E_\Phi \rho(\|X\|)$ in (5.1), so that the initial S-estimator is consistent for (μ, Σ) in the normal model. The cut-off value is choosen in such a way that the resulting S-estimator has 50% breakdown point. Finally we take weight function w defined in (2.4).

The initial S-estimator may not be consistent at the SCN, but we will always have that the reweighted biweight S-estimators (T_n, C_n) are consistent for $(\mu, \gamma\Sigma)$ (see Remark 5.1). As mentioned before, the values of σ_1 and G_2 are the same for C_n and its asymptotically unbiased version $\gamma^{-1}C_n$.

The expressions for α and σ_1 can be found in Theorem 5.1. For the indices of local robustness we find

$$G_{1,rw} = \sup_{s>0} |a(s)|s \quad \text{and} \quad G_{2,rw} = \frac{1}{k+2} \sup_{s>0} |l(s)|,$$

with a and l defined in Theorem 5.1. Graphs of α, G_1, σ_1 and G_2 as a function of the cut-off value c of the weight function (2.4) are given in Figures 2, 3 and

4, for $k = 2, 5, 10$ at the NOR (solid lines) and at the SCN for $\varepsilon = 0.1$ (dotted), $\varepsilon = 0.3$ and $\varepsilon = 0.5$ (dashed). The values α and σ_1 for the initial S-estimator at the NOR are displayed by a horizontal line.

For $k = 2$ we see that one can improve the efficiency of both the location biweight S-estimator ($\alpha_s = 1.725$) and the covariance biweight S-estimator ($\sigma_{1,s} = 2.656$) at the NOR. This is also true for both at the SCN for $\varepsilon = 0.1$ ($\alpha_s = 1.953$ and $\sigma_{1,s} = 3.021$), $\varepsilon = 0.3$ ($\alpha_s = 2.637$ and $\sigma_{1,s} = 4.129$), and for the location biweight S-estimator at the SCN with $\varepsilon = 0.5$ ($\alpha_s = 3.988$). At the SCN with $\varepsilon = 0.5$ we have $\sigma_{1,s} = 6.342$ for the covariance biweight S-estimator. One cannot improve the local robustness of the location biweight S-estimator ($G_{1,s} = 2.391$). The behavior of the scalar G_2 in the case $k = 2$ is special, since

$$G_2(c) \sim \frac{k\Gamma(k/2)c^{2-k}}{(k+2)2\pi^{k/2}h(0)} \quad \text{as } c \downarrow 0.$$

In contrast, Kent and Tyler (1997) observed that at the NOR the CM-estimators can improve both the efficiency and local robustness of the location biweight S-estimator, and similarly for the covariance biweight S-estimator for $k \leq 5$. For the value $c = 5.207$ of the weight function w, the scalar G_1 for the reweighted biweight S-estimator attains its minimum value $G_{1,rw} = 2.569$ at the NOR. For this value of c we have $\alpha_{rw} = 1.374$, $\sigma_{1,rw} = 2.111$ and $G_{2,rw} = 1.563$ at the NOR. In comparison, for the G_1-optimal CM-estimator we have $G_{1,cm} = 1.927$, $\alpha_{cm} = 1.130$, $\sigma_{1,cm} = 1.243$ and $G_{2,cm} = 1.369$.

For $k = 5$ we observe a similar behavior for the reweighted biweight S-estimator. One can improve the efficiency of both the location biweight S-estimator ($\alpha_s = 1.182$) and the covariance biweight S-estimator ($\sigma_{1,s} = 1.285$) at the NOR. This is also true for both at the SCN for $\varepsilon = 0.1$ ($\alpha_s = 1.318$ and $\sigma_{1,s} = 1.437$) and for the location estimator at the SCN with $\varepsilon = 0.3$ ($\alpha_s = 1.713$) and $\varepsilon = 0.5$ ($\alpha_s = 2.444$). One cannot improve the local robustness of the biweight S-estimator ($G_{1,s} = 2.731$ and $G_{2,s} = 1.207$). For the value $c = 10.53$, the scalar G_1 attains its minimum value $G_{1,rw} = 3.643$ at the NOR. For this value of c we have $\alpha_{rw} = 1.179$, $\sigma_{1,rw} = 1.393$ and $G_{2,rw} = 1.769$ at the NOR. In comparison, for the G_1 optimal CM-estimator we have $G_{1,cm} = 2.595$, $\alpha_{cm} = 1.072$, $\sigma_{1,cm} = 1.068$ and $G_{2,cm} = 1.271$.

For $k = 10$ one can improve the efficiency of both the location biweight S-estimator ($\alpha_s = 1.072$) and the covariance biweight S-estimator ($\sigma_{1,s} = 1.093$) at the NOR. This is also true for both at the SCN for $\varepsilon = 0.1$ ($\alpha_s = 1.191$ and $\sigma_{1,s} = 1.215$), $\varepsilon = 0.3$ ($\alpha_s = 1.534$ and $\sigma_{1,s} = 1.565$) and for the location biweight S-estimator at the SCN with $\varepsilon = 0.5$ ($\alpha_s = 2.151$). One cannot improve the local robustness of the biweight S-estimator ($G_{1,s} = 3.482$ and $G_{2,s} = 1.142$). The scalar G_1 attains its minimum value $G_{1,rw} = 4.670$ at the NOR for $c = 18.25$. For this value of c we have $\alpha_{rw} = 1.114$, $\sigma_{1,rw} = 1.218$ and $G_{2,rw} = 1.705$ at the NOR. In comparison, for the G_1 optimal CM-estimator we have $G_{1,cm} = 3.426$, $\alpha_{cm} = 1.043$, $\sigma_{1,cm} = 1.054$ and $G_{2,cm} = 1.218$.

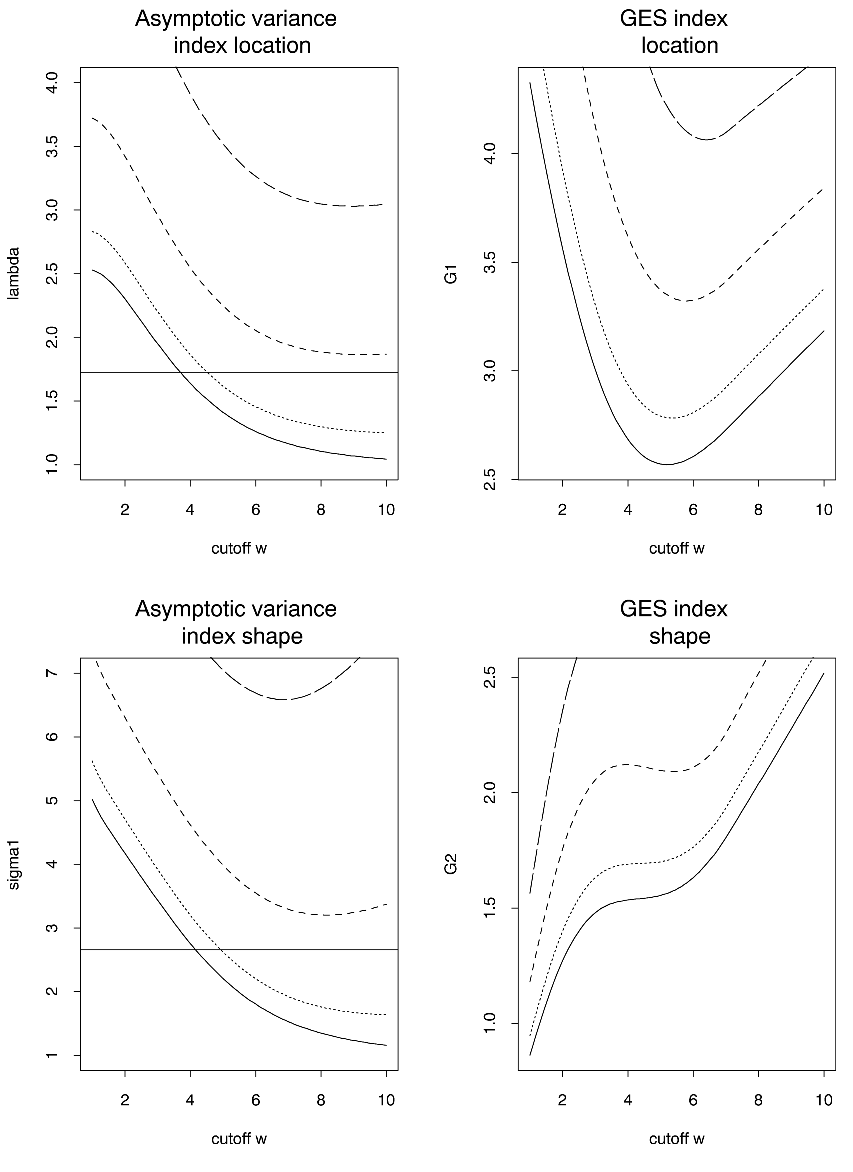

FIG. 2. *Indices α, G_1, σ_1 and G_2 for reweighted S: $k = 2$.*

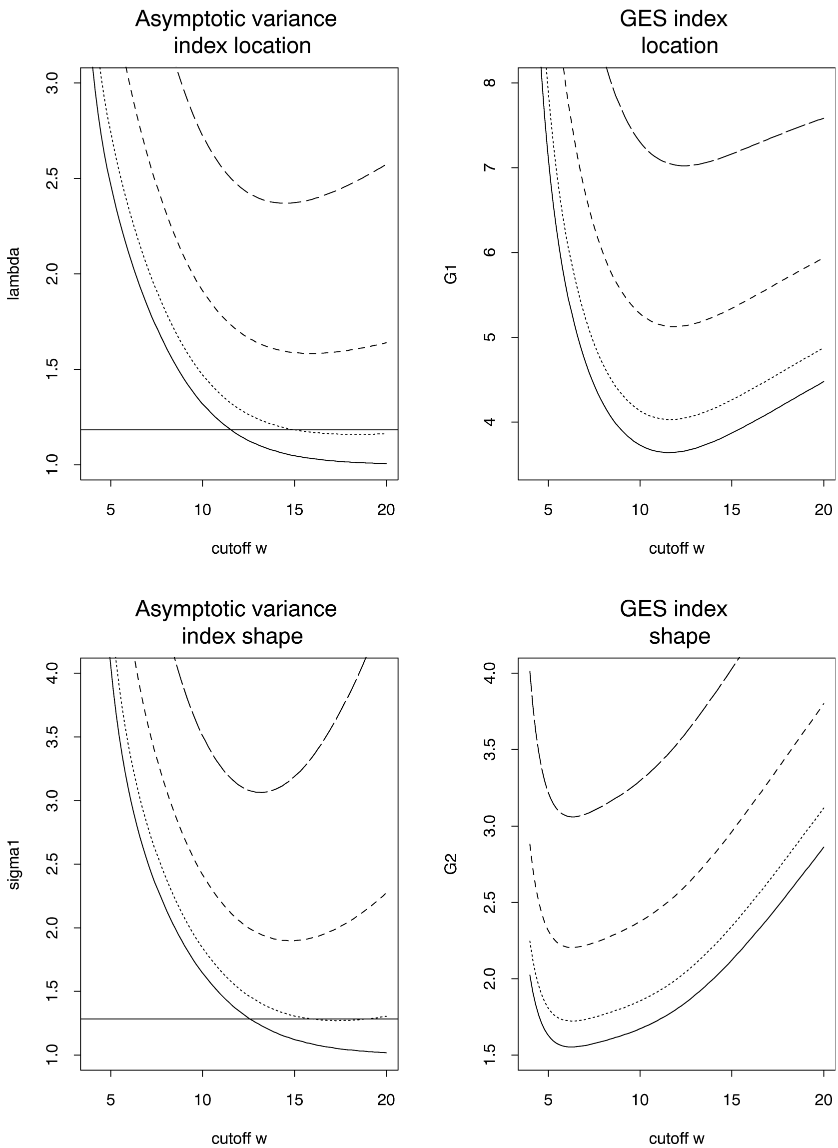

FIG. 3. *Indices α, G_1, σ_1 and G_2 for reweighted S: $k = 5$.*

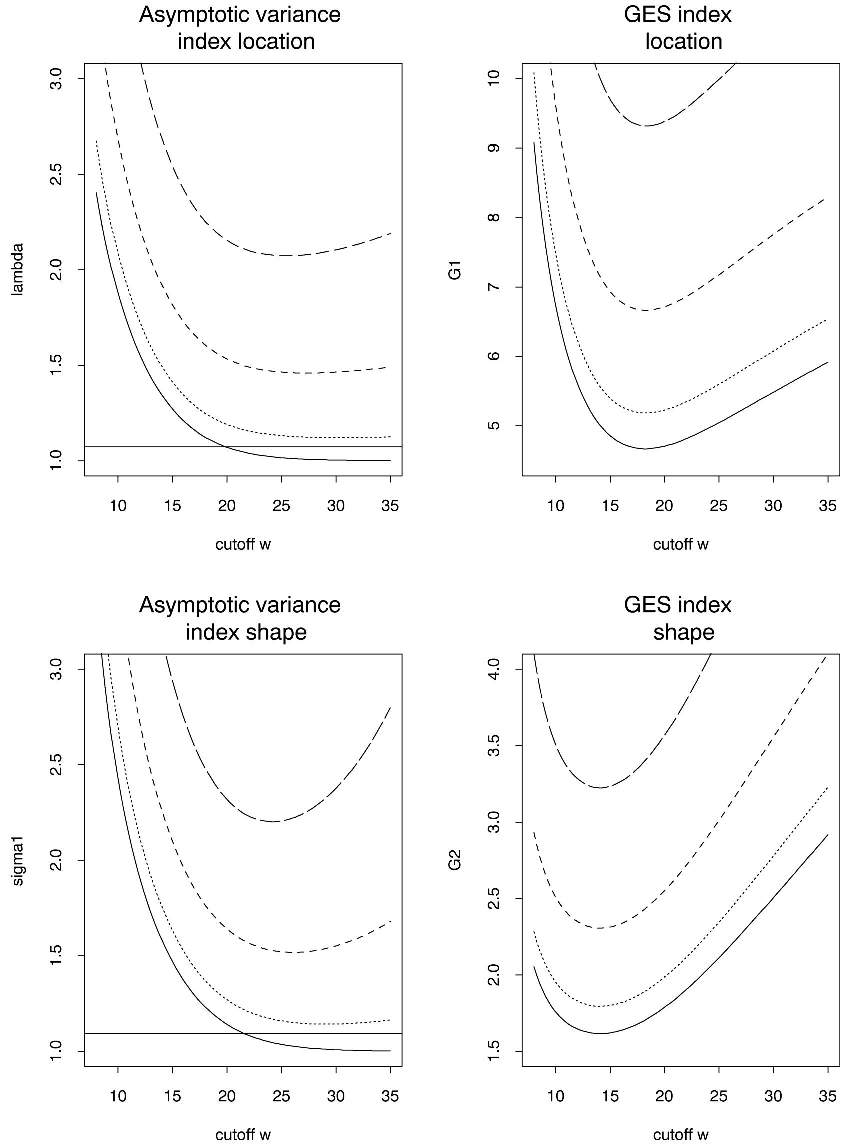

FIG. 4. *Indices α, G_1, σ_1 and G_2 for reweighted S: $k = 10$.*

6.2. *Multivariate τ-estimators.* Multivariate τ-estimators (M_n^τ, C_n^τ) are defined by

$$C_n^\tau = \frac{V_n^\tau}{nb_2} \sum_{i=1}^n \rho_2(d(X_i, M_n^\tau, V_n^\tau)),$$

where (M_n^τ, V_n^τ) minimizes $|V|^{1/k} \sum_{i=1}^n \rho_2(d(X_i, m, V))$ subject to

$$\frac{1}{n} \sum_{i=1}^n \rho_1(d(X_i, m, V)) = b_1. \tag{6.4}$$

For both ρ-functions we take one of the type (6.3). Note that when $\rho_1 = \rho_2$ and $b_1 = b_2$ then (M_n^τ, C_n^τ) are just the ordinary S-estimators. If $b_i = \int \rho_i(\|x\|) h(x^\top x)\, dx$, then $(M_n^\tau, C_n^\tau) \to (\mu, \Sigma)$ in probability. We take $b_i = E_\Phi \rho_i(\|X\|)$, for $i = 1, 2$, so that the τ-estimator is consistent for (μ, Σ) in the normal model. In Lopuhaä (1991) it is shown that M_n^τ and C_n^τ are asymptotically equivalent with the S-estimators defined by the function

$$\tilde{\rho} = A\rho_1 + B\rho_2, \tag{6.5}$$

where $A = E_{0,I}[2\rho_2(\|X\|) - \psi_2(\|X\|)\|X\|]$ and $B = E_{0,I}[\psi_1(\|X\|)\|X\|]$. The breakdown point of the τ-estimators only depends on ρ_1 and the efficiency may be improved by varying c_2. We choose the cut-off value c_1 such that the resulting τ-estimator has 50% breakdown point. At the SCN we still have that M_n^τ is consistent for μ, but C_n^τ is consistent for $\gamma\Sigma$, with $\gamma \neq 1$. However, as mentioned before, the values for σ_1 and G_2 are the same for C_n^τ and $\gamma^{-1} C_n^\tau$.

The behavior of the multivariate τ-estimator is similar to that of the *CM*-estimators. Since the limiting distribution of the multivariate τ-estimator is the same as that of an S-estimator defined with the function $\tilde{\rho}$ in (6.5), the expression for α_τ and $\sigma_{1,\tau}$ can be found in Corollary 5.1 in Lopuhaä (1989). For the indices of local robustness we find

$$G_{1,\tau} = \frac{1}{\tilde{\beta}} \sup_{s>0} |\tilde{\psi}(s)| \quad \text{and} \quad G_{2,\tau} = \frac{k}{(k+2)\tilde{\gamma}_1} \sup_{s>0} |\tilde{\psi}(s)s|,$$

where $\tilde{\beta}$ and $\tilde{\gamma}_1$ are the constants β and γ_1 defined in Corollary 5.2 in Lopuhaä (1989), corresponding with the function $\tilde{\rho}$.

Graphs of α, G_1, σ_1 and G_2 as a function of the cut-off value c_2 of ρ_2 for $c_2 \geq c_1$ are given in Figure 5, for $k = 2$ at the NOR (solid) and at the SCN for $\varepsilon = 0.1$ (dotted), $\varepsilon = 0.3$ and $\varepsilon = 0.5$ (dashed). Note that for $c_2 = c_1$ we have the corresponding values for the initial biweight S-estimator defined with the function ρ_1. We observe the same behavior as with *CM*-estimators, that is, one can improve simultaneously the efficiency and local robustness of both the location biweight S-estimator and the covariance biweight S-estimator. This remains true at the SCN with $\varepsilon = 0.1, 0.3$ and $\varepsilon = 0.5$. For instance, at the

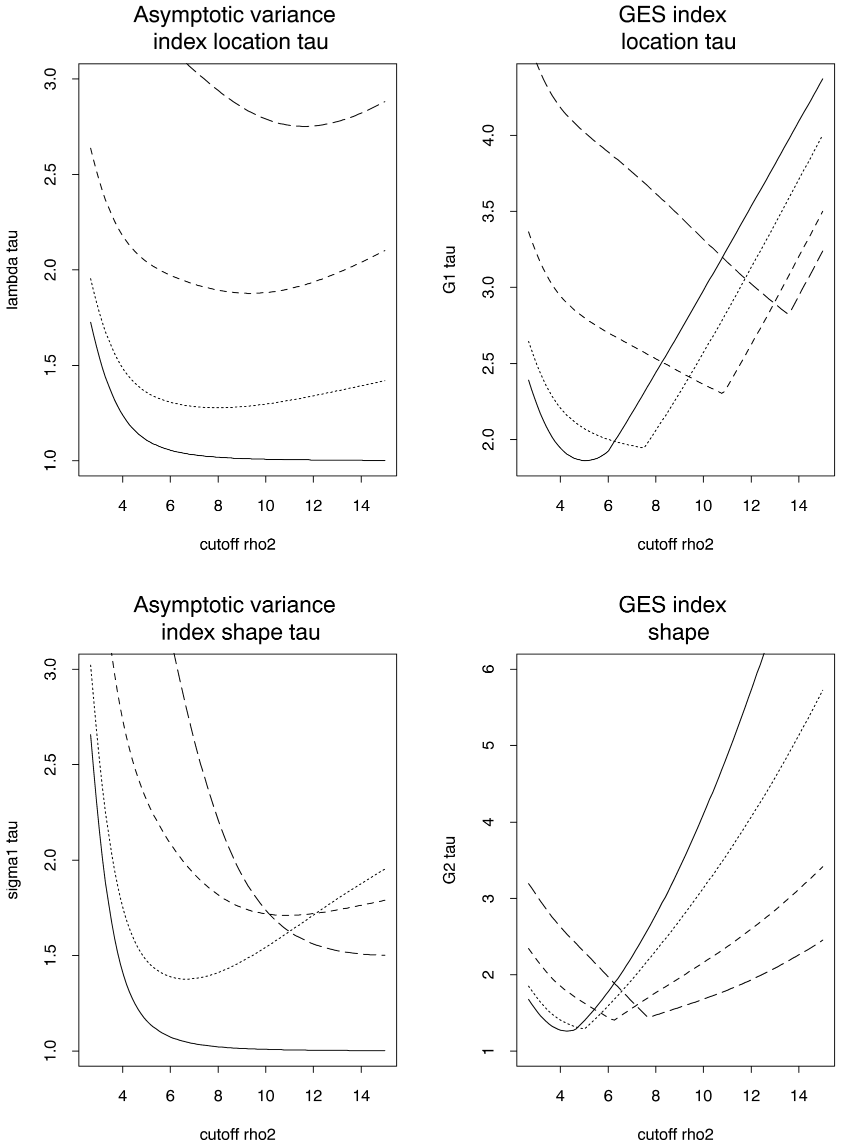

FIG. 5. *Indices α, G_1, σ_1 and G_2 for multivariate τ: $k = 2$.*

value $c_2 = 5.06$ the scalar G_1 for the τ-estimator attains its minimum value $G_{1,\tau} = 1.861$ at the NOR. For this value of c_2 we have $\alpha_\tau = 1.104$, $\sigma_{1,\tau} = 1.153$ and $G_{2,\tau} = 1.415$ at the NOR. These values, are slightly smaller (except for G_2) than the corresponding indices for the G_1-optimal *CM*-estimator.

For dimension $k = 5$ we observe a similar behavior. One can improve simultaneously the efficiency and local robustness of both the location and covariance biweight S-estimator at the NOR and also at the SCN for $\varepsilon = 0.1, 0.3$ and 0.5. However, the decrease of the scalars G_1 and G_2 is only little. For instance, the scalar G_2 for the covariance τ-estimator attains its minimum value $G_{2,\tau} = 1.203$ at the NOR for $c_2 = 4.94$. For this value of c_2 we have $\alpha_\tau = 1.492$, $\sigma_{1,\tau} = 1.230$ and $G_{1,\tau} = 2.676$ at the NOR. These values are almost the same as the corresponding indices for the G_2-optimal *CM*-estimator: $\alpha_{cm} = 1.153$, $\sigma_{1,cm} = 1.237$, $G_{1,cm} = 2.682$ and $G_{2,cm} = 1.204$. For the G_1-optimal τ-estimator the indices α, σ_1 and G_1 are slightly smaller. The scalar G_1 attains its minimum value $G_{1,\tau} = 2.588$ at the NOR for $c_2 = 6.14$. For this value of c_2 we have $\alpha_\tau = 1.069$, $\sigma_{1,\tau} = 1.099$ and $G_{2,\tau} = 1.275$ at the NOR, which are almost the same as the corresponding indices for the G_1-optimal *CM*-estimator.

Similar to what Kent and Tyler (1997) observed, we found that in dimension $k = 10$ one can no longer improve both the efficiency and the local robustness of the covariance biweight S-estimator. It is still possible to improve the efficiency of the location and covariance biweight S-estimator as well as the local robustness of the location biweight S-estimator. Again this remains true at the SCN with $\varepsilon = 0.1, 0.3$ and 0.5. The scalar G_1 attains its minimum value $G_{1,\tau} = 3.425$ at the NOR for $c_2 = 7.87$. For this value of c_2 we have $\alpha_\tau = 1.041$, $\sigma_{1,\tau} = 1.052$ and $G_{2,\tau} = 1.224$ at the NOR, which are almost the same as the corresponding indices for the G_1-optimal *CM*-estimator.

6.3. *Comparing GES at given efficiency.* Another comparison between the three methods can be made by comparing the scalar of local robustness G_1 at a given level of efficiency α for the location estimators at the NOR, and similarly comparing the scalar of local robustness G_2 at a given level of efficiency σ_1 for the covariance estimators at the NOR. In Figure 6 we plotted the graphs of G_1 and G_2 as a function of $\alpha \leq \alpha_s$ and $\sigma_1 \leq \sigma_{1,s}$, respectively, at the NOR for $k = 2, 5, 10$. The graphs for the τ-estimators, *CM*-estimators and reweighted biweight S-estimators are represented by the solid, dotted and dashed lines, respectively. We observe that the local robustness of the τ-estimators and *CM*-estimators is considerably smaller than that of the reweighted estimators at the same level of efficiency.

In dimension $k = 2$, the minimum value $G_{1,\tau} = 1.861$ of the location τ-estimator corresponds with efficiency $\alpha = 1.104$. For this value of efficiency we have $G_{1,cm} = 1.932$ and $G_{1,s} = 2.881$ for the location *CM*-estimator and reweighted location biweight S-estimator, respectively. The minimum value $G_{2,\tau} = 1.260$ for the covariance τ-estimator corresponds with efficiency $\sigma_1 = 1.308$. For this value of efficiency we have $G_{2,cm} = 1.350$ and $G_{2,s} = 2.107$ for

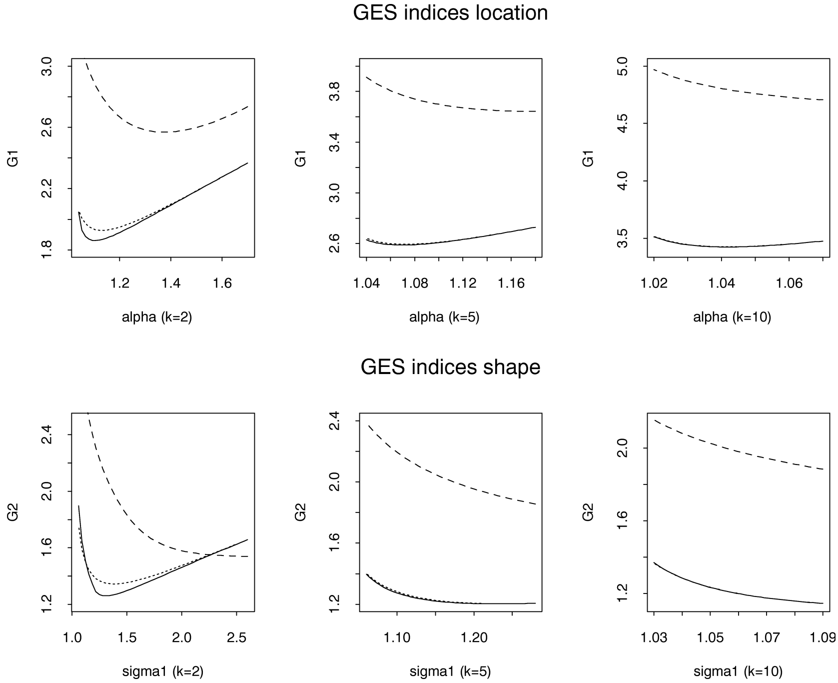

FIG. 6. *Indices G_1 and G_2 at given levels of efficiency.*

the covariance *CM*-estimator and reweighted covariance biweight S-estimator, respectively.

In dimension $k = 5$, the minimum value $G_{1,\tau} = 2.588$ of the location τ-estimator corresponds with efficiency $\alpha = 1.069$. For this value of efficiency we have $G_{1,cm} = 2.595$ and $G_{1,s} = 3.773$ for the location *CM*-estimator and reweighted location biweight S-estimator, respectively. The minimum value $G_{2,\tau} = 1.203$ for the covariance τ-estimator corresponds with efficiency $\sigma_1 = 1.231$. For this value of efficiency we have $G_{2,cm} = 1.204$ and $G_{2,s} = 1.910$ for the covariance *CM*-estimator and reweighted covariance biweight S-estimator, respectively.

In dimension $k = 10$, the minimum value $G_{1,\tau} = 3.425$ of the location τ-estimator corresponds with efficiency $\alpha = 1.041$. For this value of efficiency we have $G_{1,cm} = 3.426$ and $G_{1,s} = 4.780$ for the location *CM*-estimator and reweighted location biweight S-estimator, respectively. The minimum value $G_{2,\tau} = 1.142$ for the covariance τ-estimator corresponds with efficiency $\sigma_1 = 1.093$, and is the same as that for the initial covariance biweight S-estimator. Hence the corresponding value for the covariance *CM*-estimator is the same. For this value of efficiency we have $G_{2,s} = 1.875$ for the reweighted covariance biweight S-estimator.

Acknowledgments. I thank an Associate Editor and two anonymous referees for their comments and suggestions concerning Figure 1 and the comparisons now made in Section 6.

REFERENCES

BICKEL, P. J. (1975). One-step Huber estimates in the linear model. *J. Amer. Statist. Assoc.* **70** 428–434.

DAVIES, P. L. (1987). Asymptotic behavior of S-estimates of multivariate location parameters and dispersion matrices. *Ann. Statist.* **15** 1269–1292.

DAVIES, P. L. (1992a). The asymptotics of Rousseeuw's minimum volume ellipsoid estimator. *Ann. Statist.* **20** 1828–1843.

DAVIES, P. L. (1992b). An efficient Fréchet differentiable high breakdown multivariate location and dispersion estimator. *J. Multivariate Anal.* **40** 311–327.

DIEUDONNÉ, J. (1969). *Foundations of Modern Analysis*. Academic Press, New York.

DONOHO, D. L. (1982). Breakdown properties of multivariate location estimators. Ph.D. qualifying paper, Harvard Univ.

DONOHO, D. L. and HUBER, P. J. (1983). The notion of breakdown point. In *A Festschrift for Erich L. Lehmann* (P. J. Bickel, K. A. Doksum, J. L. Hodges, Jr., eds.) 157–184. Wadsworth, Belmont, CA.

GRÜBEL, R. (1988). A minimal characterization of the covariance matrix. *Metrika* **35** 49–52.

HAMPEL, F. R. (1968). Contributions to the theory of robust estimation. Ph.D. dissertation, Univ. California, Berkeley.

HAMPEL, F. R. (1974). The influence curve and its role in robust estimation, *J. Amer. Statist. Assoc.* **69** 383–393.

HE, X. and PORTNOY, S. (1992). Reweighted estimators converge at the same rate as the initial estimator. *Ann. Statist.* **20** 2161–2167.

JOLION, J., MEER, P. and BATAOUCHE, S. (1991). Robust clustering with applications in computer vision. *IEEE Trans. Pattern Anal. Machine Intelligence* **17** 791–802.

KENT, J. T. and TYLER, D. E. (1997). Constrained M-estimation for multivariate location and scatter. *Ann. Statist.* **24** 1346–1370.

KIM, J. and POLLARD, D. (1990). Cube root asymptotics. *Ann. Statist.* **18** 191-219.

LOPUHAÄ, H. P. (1989). On the relation between S-estimators and M-estimators of multivariate location and covariance. *Ann. Statist.* **17** 1662–1683.

LOPUHAÄ, H. P. (1991). Multivariate τ-estimators for location and scatter. *Canad. J. Statist.* **19** 307–321.

LOPUHAÄ, H. P. (1997). Asymptotic expansion of S-estimators of location and covariance. *Statist. Neerlandica* **51** 220–237.

LOPUHAÄ, H. P. and ROUSSEEUW, P. J. (1991). Breakdown properties of affine equivariant estimators of multivariate location and covariance matrices. *Ann. Statist.* **19** 229–248.

MARONNA, R. A. (1976). Robust M-estimators of multivariate location and scatter. *Ann. Statist.* **4** 51–67.

MARONNA, R. A. and YOHAI, V. J. (1995). The behavior of the Stahel–Donoho robust multivariate estimator. *J. Amer. Statist. Assoc.* **90** 330–341.

MATEI, B., MEER, P. and TYLER, D. (1998). Performance assessment by resampling: rigid motion estimators. In *Empirical Evaluation Methods in Computer Vision* (K. W. Bowyer, P. J. Philips, eds.) 72–95. IEEE CS Press, Los Alamitos, California.

MEER, P., MINTZ, D., ROSENFELD, A. and KIM, D. Y. (1991). Robust regression methods for computer vision: a review. *Internat. J. Comput. Vision* **6** 59–70.

POLLARD, D. (1984). *Convergence of Stochastic Processes*. Springer, New York.

ROUSSEEUW, P. J. (1985). Multivariate estimation with high breakdown point. In *Mathematical Statistics and Applications* **B** (W. Grossmann, G. Pflug, I. Vincze and W. Wertz, eds.) 283–297. Reidel, Dordrecht.

ROUSSEEUW, P. J. and LEROY, A. M. (1987). *Robust Regression and Outlier Detection*. Wiley, New York.

STAHEL, W. A. (1981). Robust estimation: infinitesimal optimality and covariance matrix estimators. Ph.D. thesis (in German). ETH, Zurich.

TYLER, D. E. (1986). Breakdown properties of the M-estimators of multivariate scatter. Technical report, Dept. Statistics, Rutgers Univ.

YOHAI, V. J. (1987). High breakdown point and high efficiency robust estimates for regression. *Ann. Statist.* **15** 642–656.

FACULTY ITS
DEPARTMENT OF MATHEMATICS
DELFT UNIVERSITY OF TECHNOLOGY
MEKELWEG 4, 2628 CD DELFT
THE NETHERLANDS
E-MAIL: h.p.lopuhaa@twi.tudelft.nl