Journal of Multivariate Analysis **71**, 161–190 (1999)
Article ID jmva.1999.1839, available online at http://www.idealibrary.com on IDEAL®

Influence Function and Efficiency of the Minimum Covariance Determinant Scatter Matrix Estimator

Christophe Croux

Université Libre de Bruxelles, Brussels, Belgium
E-mail: ccroux@ulb.ac.be

and

Gentiane Haesbroeck

University of Liège, Liège, Belgium
E-mail: G.Haesbroeck@ulg.ac.be

Received June 2, 1998

The minimum covariance determinant (MCD) scatter estimator is a highly robust estimator for the dispersion matrix of a multivariate, elliptically symmetric distribution. It is relatively fast to compute and intuitively appealing. In this note we derive its influence function and compute the asymptotic variances of its elements. A comparison with the one step reweighted MCD and with S-estimators is made. Also finite-sample results are reported. © 1999 Academic Press

AMS 1991 subject classifications: 62F35, 62G35.

Key words and phrases: influence function; minimum covariance determinant estimator; robust estimation; scatter matrix.

1. INTRODUCTION

The classical average and covariance matrix are key ingredients in almost all multivariate statistical methods. However, they are extremely sensitive to outliers. It is therefore important to consider robust alternatives for these estimators, which can be used to obtain robust procedures in principal component analysis, factor analysis, etc. Among the robust alternatives to the classical estimators, the minimum volume ellipsoid (MVE) introduced by Rousseeuw (1985) is probably the most widely known and used (e.g., by He and Wang, 1996). It consists of taking as location estimator the center of the smallest regular ellipsoid containing half the points of the data set. The scatter estimator is then defined by the shape

0047-259X/99 $30.00

matrix of that ellipsoid. However, it was shown that the MVE estimator is not $\sqrt{n}$ consistent (Davies, 1992a), making it less attractive for efficiency reasons. A robust estimator which has the normal rate of convergence is the minimum covariance determinant (MCD) estimator of Rousseeuw (1985). The location and scatter estimates are given by the average and covariance matrix computed on that half of the data which attain the smallest determinant of their covariance matrix. Until quite recently, the major drawback of the MCD estimator was its computation time. Rousseeuw and Van Driessen (1999) however, proposed a new algorithm to compute MCD, which turns out to be extremely fast, even in high dimensions. Theoretical properties of the MCD estimator have been investigated in Butler (1982) and Butler, Davies, and Jhun (1993), but the asymptotic distribution of the scatter part of the MCD remained unknown. In the particular case of one dimension the influence function of the MCD scale was computed by (Croux and Rousseeuw, 1992).

The main focus of this paper is to derive the influence function of the MCD scatter matrix in arbitrary dimensions. Second, the influence function is used to calculate asymptotic variances and efficiencies for the MCD and its reweighted version. Third, the finite-sample behavior is investigated by means of a simulation study. Furthermore, a thorough comparison with multivariate S-estimators is included.

The outline of the paper is as follows. Section 2 defines the MCD-functional and gives its influence function while in Section 3 asymptotic variances are computed. Section 4 uses results of Lopuhaä (1997) to show that one step reweighted covariances, starting from MCD based weights, combine the breakdown properties of the MCD estimator while achieving a better asymptotic efficiency. However, this reweighting does not yield an estimator as efficient as the S-estimator of scatter. Finite-sample results based on simulations are reported in Section 5 while Section 6 contains some conclusions.

2. INFLUENCE FUNCTION

For a finite sample of observations $\{x_1, ..., x_n\}$ in $\mathbb{R}^p$ the MCD is determined by selecting that subset $\{x_{i_1}, ..., x_{i_h}\}$ of size h, with $1 \leqslant h \leqslant n$, which minimizes the generalized variance (which is the determinant of the covariance matrix computed from the subset) among all possible subsets of size h. The location estimator is then defined as

$$\hat{\mu}_n = \frac{1}{h} \sum_{j=1}^{h} x_{i_j}$$

and the estimator of scatter by

$$\hat{\Sigma}_n = c_p \frac{1}{h} \sum_{j=1}^{h} (x_{i_j} - \hat{\mu}_n)(x_{i_j} - \hat{\mu}_n)^t,$$

where c_p is a consistency factor. The choice $h = [(n+p+1)/2]$ is commonly preferred since it yields the highest possible breakdown point (Lopuhaä and Rousseeuw 1991). Note that $h \approx \frac{n}{2}$ at least if the number of observations is much higher than the dimension. The breakdown point of a multivariate scale estimator is defined as the smallest fraction of observations that you need to replace to arbitrary position before the estimated scatter explodes (its biggest eigenvalue tends to infinity) or implodes (its smallest eigenvalue tends to zero). Another default value is $h \approx 0.75n$, yielding a better compromise between efficiency/stability and high breakdown. To derive the influence function of an estimator, we need a proper definition for its functional form. In Butler, Davies, and Jhun (1993; BDJ from now on) a definition appropriate for continuous distributions is given, which we will generalize afterwards to arbitrary distributions G. This is necessary since we will apply the MCD to point contaminated distributions.

Denote by $0 < \alpha < 1$ the mass of the data not determining the MCD, which will result in an estimator with (asymptotic) breakdown point $\min(\alpha, 1-\alpha)$. Define

$$\mathcal{D}_G(\alpha) = \{A \mid A \subseteq \mathbb{R}^p \text{ measurable and bounded with } P_G(A) = 1 - \alpha\}, \quad (2.1)$$

and for every $A \in \mathcal{D}_G(\alpha)$, the average and covariance matrix computed over this set will be denoted by

$$T_A(G) = \frac{\int_A y \, dG(y)}{1-\alpha} \quad (2.2)$$

and

$$\Sigma_A(G) = \frac{\int_A (y - T_A(G))(y - T_A(G))^t \, dG(y)}{1-\alpha}. \quad (2.3)$$

The set A is called an MCD-solution if

$$\det(\Sigma_A(G)) \leqslant \det(\Sigma_{\tilde{A}}(G)), \quad (2.4)$$

for every other $\tilde{A} \in \mathcal{D}_G(\alpha)$. The MCD estimators at the theoretical distribution are then defined by

$$T(G) = T_A(G) \qquad \text{and} \qquad \Sigma(G) = c_\alpha \Sigma_A(G) \quad (2.5)$$

for an MCD solution A. The constant c_α can be chosen in such a way that consistency will be obtained at the specified model.

A problem may arise with the above definition when the distribution G is not continuous, in which case $\mathscr{D}_G(\alpha)$ may be empty. Therefore define

$$\tilde{\mathscr{D}}_G(\alpha) = \{(A, x) \mid A \subseteq \mathbb{R}^p \text{ measurable and bounded, } x \in \mathbb{R}^p \backslash A \\ \text{and} \quad \exists 0 \leqslant \delta \leqslant P_G(\{x\}) : P_G(A) + \delta = 1 - \alpha\},$$

which is never empty. Furthermore, for every $(A, x) \in \tilde{\mathscr{D}}_G(\alpha)$ define (with $\delta = 1 - \alpha - P_G(A)$)

$$T_{(A,x)}(G) = \frac{\int_A y \, dG(y) + \delta x}{1 - \alpha}, \tag{2.6}$$

and

$$\Sigma_{(A,x)}(G) = (1-\alpha)^{-1} \left\{ \int_A (y - T_{(A,x)}(G))(y - T_{(A,x)}(G))^t \, dG(y) \right. \\ \left. + \delta (x - T_{(A,x)}(G))(x - T_{(A,x)}(G))^t \right\}. \tag{2.7}$$

(Since A is bounded all the above quantities are well defined). We say that $(A, x) \in \tilde{\mathscr{D}}_G(\alpha)$ is an MCD solution if

$$\det(\Sigma_{(A,x)}(G)) \leqslant \det(\Sigma_{(\tilde{A},\tilde{x})}(G)), \tag{2.8}$$

for every $(\tilde{A}, \tilde{x}) \in \tilde{\mathscr{D}}_G(\alpha)$. The MCD estimators at the population level are then given by

$$T(G) = T_{(A,x)}(G) \qquad \text{and} \qquad \Sigma(G) = c_\alpha \Sigma_{(A,x)}(G), \tag{2.9}$$

for an MCD solution (A, x). Thus, the MCD solution determines a region $B = A \cup \{x\}$, with $P_G(A) \leqslant 1 - \alpha \leqslant P_G(B)$. By giving a lower mass to the atom x, we are in a certain way interpolating between the two sets A and B. For ease of notation, we will write $T_A(G)$ and $\Sigma_A(G)$ instead of $T_{(A,x)}(G)$ and $\Sigma_{(A,x)}(G)$ whenever $P_G(A) = 1 - \alpha$ for a couple $(A, x) \in \tilde{\mathscr{D}}_G(\alpha)$.

In this paper we will focus on the problem of estimating the parameters μ and Σ of the distribution $F_{\mu,\Sigma}$ with density

$$f_{\mu,\Sigma}(x) = \frac{g((x-\mu)^t \Sigma^{-1}(x-\mu))}{\sqrt{\det(\Sigma)}} \tag{2.10}$$

with $\mu \in \mathbb{R}^p$ and $\Sigma \in \mathrm{PD}(p)$, the set of all positive definite matrices of size p.

The function g is assumed to be known and to have a strictly negative derivative g', so that $F_{\mu, \Sigma}$ belongs to a parametric class of elliptically symmetric, unimodal distributions. It is shown in Section 2 of BDJ that for this distribution the MCD-problem has a unique solution given by the ellipsoid

$$A(F_{\mu, \Sigma}) = \{z \in \mathbb{R}^p \mid (z-\mu)^t \, \Sigma^{-1} \, (z-\mu) \leqslant q_\alpha\}, \tag{2.11}$$

where $G(t) = P_{F_{0,I}}(Z^tZ \leqslant t)$ and $q_\alpha = G^{-1}(1-\alpha)$, with corresponding MCD-functionals

$$T(F_{\mu, \Sigma}) = \mu$$

$$\Sigma(F_{\mu, \Sigma}) = \left(\frac{c_\alpha}{1-\alpha} \int_{z^tz \leqslant q_\alpha} z_1^2 \; dF_{0,I}(z) \right) \Sigma.$$

To obtain Fisher-consistency at this model, it suffices to set

$$c_\alpha = \frac{1-\alpha}{\int_{z^tz \leqslant q_\alpha} z_1^2 \; dF_{0,I}(z)} = (1-\alpha) \left\{ \frac{\pi^{p/2}}{\Gamma(p/2+1)} \int_0^{\sqrt{q_\alpha}} r^{p+1} g(r^2) \, dr \right\}^{-1}.$$

Table I gives several values of the constants c_α and q_α for different values of p and α at the normal model.

Since the MCD is affine equivariant we will only derive the influence function at the model distribution $F = F_{0, I_p}$. Consider the *contaminated distribution*

$$F_{\varepsilon, x} = (1-\varepsilon) \, F + \varepsilon \Delta_x,$$

where Δ_x is a Dirac measure putting all its mass at $x \in \mathbb{R}^p$. The influence function measures the sensitivity of the MCD-functional to small amounts of contamination in the distribution:

$$IF(x, \Sigma, F) = \lim_{\varepsilon \downarrow 0} \frac{\Sigma(F_{\varepsilon, x}) - \Sigma(F)}{\varepsilon}.$$

More about the use and interpretation of influence functions can be found in Hampel *et al.* (1986). The derivation of the influence function for the MCD relies on the following proposition, which says that the MCD solution at the contaminated distribution $F_{\varepsilon, x}$ is still determined by an ellipsoid. (This can even be shown to be true for any distribution G, and was already proven for empirical distribution functions by BDJ, p. 1392.)

TABLE I

Particular Values of c_α and q_α at the Normal Model

α		$p=2$	$p=3$	$p=5$	$p=10$	$p=30$
0.25	c_α	1.859	1.609	1.412	1.256	1.130
	q_α	2.773	4.108	6.626	12.549	37.799
0.5	c_α	3.259	2.457	1.912	1.531	1.257
	q_α	1.386	2.366	4.351	9.342	29.336

Proposition 1. *Take $0<\varepsilon<\min(\alpha, 1-\alpha)$, $x\in\mathbb{R}^p$ and consider the contaminated distribution $F_{\varepsilon, x}$. For any MCD-solution $(A, y)\in\tilde{\mathscr{D}}_{F_{\varepsilon, x}}(\alpha)$, there exists an open ellipsoid $\mathscr{E}$ such that*

$$\mathscr{E}\in\mathscr{D}_{F_{\varepsilon, x}}(\alpha), \qquad T_{\mathscr{E}}(F_{\varepsilon, x})=T_{(A, y)}(F_{\varepsilon, x}), \quad \textit{and} \quad \Sigma_{\mathscr{E}}(F_{\varepsilon, x})=\Sigma_{(A, y)}(F_{\varepsilon, x})$$

or, in the special case that x lies at the border of $\mathscr{E}$,

$$(\mathscr{E}, x)\in\tilde{\mathscr{D}}_{F_{\varepsilon, x}}(\alpha),$$

$$T_{(\mathscr{E}, x)}(F_{\varepsilon, x})=T_{(A, y)}(F_{\varepsilon, x}), \qquad \textit{and} \qquad \Sigma_{(\mathscr{E}, x)}(F_{\varepsilon, x})=\Sigma_{(A, y)}(F_{\varepsilon, x}).$$

The next theorem gives the influence function of the scatter matrix part of the MCD functional. To facilitate the interpretation, we give separate expressions for the diagonal and off-diagonal elements. All proofs are kept for the Appendix.

Theorem 1. *With the notations from above, we have that*

$$IF(x, \Sigma_{ii}, F)=\frac{1}{b_1}\left\{\frac{c_\alpha}{1-\alpha}x_i^2\, I(\|x\|^2\leqslant q_\alpha)+\frac{b_2}{b_1-pb_2}\frac{c_\alpha}{1-\alpha}\|x\|^2\, I(\|x\|^2\leqslant q_\alpha)\right.$$
$$\left.+\frac{b_1}{b_1-pb_2}\left[\frac{c_\alpha}{1-\alpha}\frac{q_\alpha}{p}(1-\alpha-I(\|x\|^2\leqslant q_\alpha))-1\right]\right\} \tag{2.12}$$

$$IF(x, \Sigma_{ij}, F)=\frac{x_i x_j}{-2c_3}\, I(\|x\|^2\leqslant q_\alpha) \qquad \text{if } i\neq j, \tag{2.13}$$

where the constants b_1, b_2, c_2, c_3, and c_4 are determined by the relations

$$c_2=\frac{\pi^{p/2}}{\Gamma(p/2+1)}\int_0^{\sqrt{q_\alpha}} r^{p+1}g'(r^2)\,dr$$

$$c_3=\begin{cases}\dfrac{\pi^{p/2}}{(p+2)\,\Gamma(p/2+1)}\displaystyle\int_0^{\sqrt{q_\alpha}} r^{p+3}g'(r^2)\,dr & \textit{if} \quad p\geqslant 2\\ 0 & \textit{otherwise}\end{cases}$$

$$c_4 = \frac{3\pi^{p/2}}{(p+2)\Gamma(p/2+1)} \int_0^{\sqrt{q_\alpha}} r^{p+3} g'(r^2)\, dr$$

$$b_1 = \frac{c_\alpha(c_3 - c_4)}{1-\alpha}$$

$$b_2 = \frac{1}{2} + \frac{c_\alpha}{1-\alpha}\left[c_3 - \frac{q_\alpha}{p}\left(c_2 + \frac{1-\alpha}{2}\right)\right].$$

It follows immediately that the MCD scale estimator has a bounded influence function, which is redescending to zero for the off-diagonal elements, but not for the on-diagonal elements. In Fig. 1 the influence functions $IF(x, \Sigma_{11}, F)$ and $IF(x, \Sigma_{12}, F)$ are pictured for $p=2$ and $\alpha=0.25$ at the normal model. The functions are smooth, except for a jump at the circle with radius $\sqrt{q_\alpha}$. If we denote θ the angle of x with the first axis, then the discontinuity in $IF(x, \Sigma_{11}, F)$ is upwards for $\theta < \pi/4$ and downwards for $\theta > \pi/4$. The maximal value is attained at $(\sqrt{q_\alpha}, 0)$. The off-diagonal influence function is maximal at $(\sqrt{q_\alpha/2}, \sqrt{q_\alpha/2})$.

Remark 1. For $p > 1$ one can see that $c_4 = 3c_3$ and the obtained influence function can be rewritten in the more compact form

$$IF(x, \Sigma, F) = \frac{-1}{2c_3} xx^t I(\|x\|^2 \leqslant q_\alpha) + w(\|x\|)\, I_p,$$

where w is a certain real valued function.

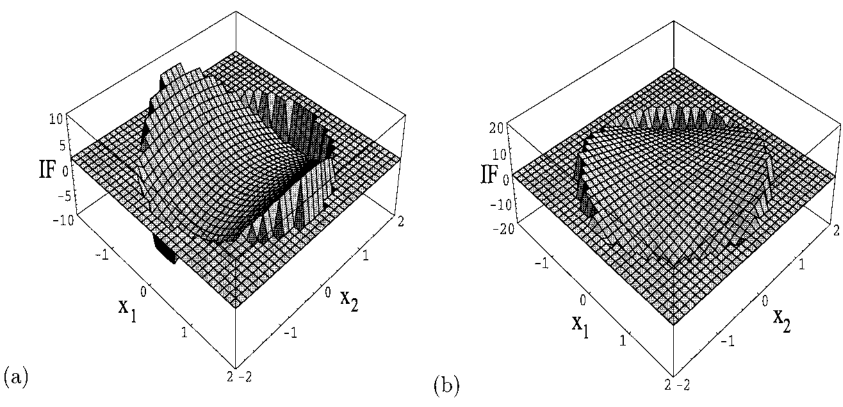

FIG. 1. Influence function of the MCD scale estimator at the normal model, with $p=2$, and $\alpha=0.25$, (a) for the first diagonal element of the scatter matrix (b) for an off-diagonal element of the scatter matrix.

Remark 2. The classical estimator of a scatter matrix is the covariance matrix $C(G)$, defined at the population level by

$$C(G)=c_0 \int_{\mathbb{R}^p} (x-\mu(G))(x-\mu(G))^t \, dG(x), \tag{2.14}$$

where $\mu(G)=\int_{\mathbb{R}^p} x dG(x)$ and the consistency factor c_0 equals $E_F(X_i^2)^{-1}$ where X_i is a component of the vector X. It appears as a limit case of the MCD for α, the mass of the data "discarded" by the MCD, tending to zero. Note that C is only well defined for distributions with a second moment, whereas the MCD functionals are properly defined for arbitrary distributions. It is not so difficult to check that

$$\lim_{\alpha \downarrow 0} IF(x, \Sigma, F)=c_0 \, xx^t - I = IF(x, C, F).$$

Remark 3. In the case $p=1$, (2.12) can be rewritten as

$$IF(x, \Sigma, F)=\left\{\int_{-\sqrt{q_\alpha}}^{\sqrt{q_\alpha}} t^2 \, dF(t)\right\}^{-1} \{\mathrm{I}(|x| \leqslant \sqrt{q_\alpha})(x^2-q_\alpha)+(1-\alpha) \, q_\alpha\}-1,$$

with $\sqrt{q_\alpha}=F^{-1}(1-\alpha/2)$. The parameter $S=\sqrt{\Sigma}$ measures here the dispersion of the univariate distribution F. This above expression corresponds with the results of Croux and Rousseeuw (1992).

Remark 4. The computation of the influence function of the location part of MCD is relatively simple, and the result is implicitly contained in BDJ:

$$IF(x, T, F)=\left(\frac{-2}{1-\alpha} \int_{z^t z \leqslant q_\alpha} zz^t g'(z^t z) \, dz\right)^{-1} \frac{x}{1-\alpha} I(\|x\|^2 \leqslant q_\alpha).$$

Observe that $IF(x, T, F)$ becomes zero when x is outside the ball of radius $\sqrt{q_\alpha}$, illustrating that MCD-location "rejects" huge outliers. This is in contrast with MCD-scatter, where these outliers still have a (bounded) influence on the diagonal elements of the scatter matrix.

Attention in this paper is restrained to the scale case, but also highly robust affine equivariant estimation of location is of interest (e.g. for application in MANOVA models). It is however well known (Rousseeuw and Leroy, 1987, p. 271) that with an *orthogonally* equivariant location estimator T_0 and an *affine* equivariant scatter matrix estimator Σ_0, an affine equivariant location estimator T_1 is easily obtained by setting

$$T_1(x_1, ..., x_n)=\Sigma_0^{1/2} \, T_0(\Sigma_0^{-1/2} \, x_1, ..., \Sigma_0^{-1/2} \, x_n).$$

Lopuhaä (1992) proved that T_1 inherits the robustness properties of the initial estimators Σ_0 and T_0. Furthermore, orthogonally equivariant location estimators which have high breakdown point, good efficiency properties and which are very fast to compute have been proposed in the literature (e.g., Hössjer and Croux, 1994).

Remark 5. Strictly speaking, Theorem 1 only gives an almost sure expression for the influence function. In case that $\|x\|^2 = q_\alpha$, a zero probability event, one has $IF(x, \Sigma, F) = c_\alpha xx^t - I$. This follows quite immediately from the special case of Proposition 1.

3. ASYMPTOTIC VARIANCES

It follows from BDJ (by taking $g(x) = xx^t I(x \in E \cap C)$ in the proof of Theorem 4, page 1397) that the MCD scatter matrix is asymptotically normal, but no expression for the asymptotic variance has been derived there. In this section we will compute asymptotic variances by means of

$$\mathrm{ASV}(\Sigma_{ij}, F) = \int_{\mathbb{R}^p} IF^2(x, \Sigma_{ij}, F)\, dF(x) \tag{3.1}$$

for $1 \leqslant i, j \leqslant n$. The above expression certainly holds if the MCD-functional is Frechet differentiable but a formal verification is still open. To our knowledge, a proof of Frechet differentiability has not been given yet in the literature. The normal location-scale model (2.10) with

$$g_\Phi(t) = \frac{1}{(2\pi)^{p/2}} \exp\left(-\frac{t}{2}\right)$$

is without any doubt the most important case. We will denote $F = \Phi$ the multivariate standard normal distribution. Explicit expressions for the asymptotic variances can be worked out:

$$\begin{aligned}\mathrm{ASV}(\Sigma_{ii}, \Phi) = {} & \{b_1(b_1 - pb_2)(1-\alpha)\}^{-2} \\ & \{(1-\alpha)\, b_1^2(\alpha((c_\alpha q_\alpha/p) - 1)^2 - 1) - 2c_3 c_\alpha^2(3(b_1 - pb_2)^2 \\ & + (p+2)\, b_2(2b_1 - pb_2))\} \\ \mathrm{ASV}(\Sigma_{ij}, \Phi) = {} & -\frac{1}{2c_3} \qquad \text{if} \quad i \neq j.\end{aligned}$$

Please note that the above expressions vary both with the dimension p and with the trimming proportion α. The limit case $\alpha = 0$ gives the asymptotic variances of the usual covariance matrix:

$$\mathrm{ASV}(C_{ii}, \Phi) = 2 \qquad \text{and} \qquad \mathrm{ASV}(C_{ij}, \Phi) = 1 \qquad \text{for} \quad i \neq j.$$

These asymptotic variances are used to compute the asymptotic efficiencies of the estimators at a model distribution F, which are defined by the ratios

$$\mathrm{Eff}(\Sigma_{ij}, F) = \frac{1}{\mathrm{ASV}(\Sigma_{ij}, F)\, \mathscr{I}(\Sigma_{ij}, F)} \qquad \forall 1 \leqslant i, \quad j \leqslant p, \tag{3.2}$$

where the inverse of the Fisher Information $\mathscr{I}(\Sigma_{ij}, F)$ is the Cramèr-Rao lower bound. It is well known that the classical covariance matrix attains the lower bound at normal models.

Figure 2 illustrates how the asymptotic efficiencies $\mathrm{Eff}(\Sigma_{ij}, \Phi)$ vary with α and p under the normal model. First of all, note that efficiency decreases as the breakdown point increases, showing a conflict between efficiency and high breakdown. Choosing the highest possible breakdown point ($\alpha = 0.5$) results in a severe loss of efficiency. The suggestion to take $\alpha = 0.25$ as a compromise between efficiency and robustness finds its motivation here: the corresponding estimator can still cope with realistic amounts of contamination in the data, but is much more precise (when no outliers are present) than the usual choice $\alpha = 0.5$. One should not forget that the bias of the MCD with $\alpha = 0.25$ remains bounded for percentages of contamination smaller than 25% but can still become unpractically large. In these cases, the maximal breakdown point MCD may become an alternative. Optimally, a data-driven choice of α should be undertaken, but this possibility has not been explored further.

Figure 2 also reveals that the efficiencies of MCD are increasing with the dimension. Even more, one can formally prove that

$$\lim_{p \to \infty} \mathrm{Eff}(\Sigma_{ij}, \Phi) = 1 - \alpha \qquad \forall 1 \leqslant i, \quad j \leqslant p. \tag{3.3}$$

Result (3.3) can be understood intuitively: every observation $X_i = (X_{i1}, ..., X_{ip})^t$ following the reference distribution Φ lies at a squared distance $X_{i1}^2 + \cdots X_{ip}^2 \approx p$

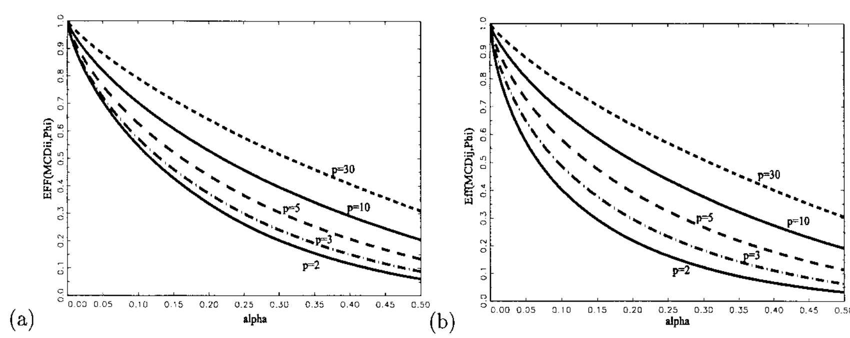

FIG. 2. Asymptotic efficiency of respectively (a) a diagonal element (b) an off-diagonal element of the MCD scatter matrix estimator at the normal model for several values of p and α varying in (0, 0.5).

from the origin, when p is large. (Here we used the strong law of large numbers and independency of the different components of X_i.) This implies that all observations are "equally likely" to constitute the subset of size $n(1-\alpha)$ whose covariance matrix will yield the MCD-scatter estimator. We conclude that, at the reference distribution Φ and for p huge, the MCD is like a usual covariance matrix but computed from a random subsample of size $n(1-\alpha)$ instead of size n, which explains (3.3). The convergence in (3.3) is however rather slow, as is shown by the table below for $\alpha = 0.25$:

p	100	10^3	10^4	10^7	∞
$\mathrm{Eff}(\Sigma_{ii}, \Phi)$	0.6559	0.7210	0.7415	0.7497	0.75
$\mathrm{Eff}(\Sigma_{ij}, \Phi)$	0.6542	0.7209	0.7411	0.7497	0.75

Another class of multivariate elliptically symmetric distributions which is regularly used (e.g. in econometrics), is generated by setting in (2.10)

$$g_{F_\nu}(t) = \frac{\Gamma((p+\nu)/2)}{\Gamma(\nu/2)(\pi\nu)^{p/2}} \frac{1}{(1+(t/\nu))^{(p+\nu)/2}}.$$

The corresponding model distribution is a multivariate Student with ν degrees of freedom. Consistency factors and influence functions can be easily computed, using symbolic computation software. Moreover, asymptotic variances have been computed by (3.1) for the reference distribution F_ν, using numerical integration and a lot of computation time in order to obtain accurate results. The classical covariance estimator (2.12) needs now to be premultiplied with $c_0 = (\nu-2)/\nu$, and has asymptotic variances $\mathrm{ASV}(C_{ii}, F_\nu) = (2\nu-2)/(\nu-4)$ and $\mathrm{ASV}(C_{ij}, F_\nu) = (\nu-2)/(\nu-4)$ for $\nu > 4$ and $i \neq j$. The Fisher Informations at F_ν are equal to $\mathcal{I}(\Sigma_{ii}, F_\nu) = (p+\nu-1)/(2(p+\nu+2))$ and $\mathcal{I}(\Sigma_{ij}, F_\nu) = (p+\nu)/(p+\nu+2)$ for $i \neq j$.

Tables II and III list the asymptotic efficiencies of the elements of the MCD estimator at some Student distributions F_ν with ν degrees of freedom, for several values of p and for $\alpha = 0.25$ and 0.5. Corresponding values for the normal distribution (which is the limit case for $\nu \to \infty$) are also reported. One sees that at these heavier tailed distributions it remains true that the efficiency gain taking $\alpha = 0.25$ instead of the usual $\alpha = 0.5$ is considerable. Moreover, at the Student distribution with 5 degrees of freedom, the MCD (with $\alpha = 0.25$) is even more efficient than the classical covariance matrix estimator.

TABLE II

Asymptotic Efficiencies of a Diagonal Element of the MCD Scatter Matrix with $\alpha = 0.25$ and 0.5 and of the Covariance Matrix C at Some Student Distributions for Several Values of p

		$p=2$	$p=3$	$p=5$	$p=10$	$p=30$
MCD	F_5	0.455	0.486	0.528	0.560	0.580
$\alpha = 0.25$	F_8	0.393	0.435	0.496	0.558	0.620
	F_{15}	0.335	0.380	0.455	0.548	0.617
	Φ	0.262	0.300	0.366	0.459	0.577
MCD	F_5	0.139	0.187	0.253	0.325	0.393
$\alpha = 0.5$	F_8	0.109	0.153	0.284	0.307	0.395
	F_{15}	0.086	0.124	0.184	0.275	0.380
	Φ	0.062	0.089	0.134	0.205	0.310
C	F_5	0.375	0.357	0.333	0.304	0.272
	F_8	0.762	0.743	0.714	0.672	0.618
	F_{15}	0.935	0.926	0.912	0.886	0.841
	Φ	1.000	1.000	1.000	1.000	1.000

Remark. Table IV reports the asymptotic efficiencies (only for the diagonal elements) of the MCD estimator at the Cauchy distribution (which is a Student distribution with 1 degree of freedom). The results are less clearcut here: the best choice in high dimensions is $\alpha = 0.5$. Notice that the efficiencies in Table IV decrease with the dimension.

TABLE III

Asymptotic Efficiencies of an Off-diagonal Element of the MCD Scatter Matrix with $\alpha = 0.25$ and 0.5 and of the Covariance Matrix C at Some Student Distributions for Several Values of p

		$p=2$	$p=3$	$p=5$	$p=10$	$p=30$
MCD	F_5	0.284	0.375	0.474	0.564	0.623
$\alpha = 0.25$	F_8	0.241	0.332	0.440	0.551	0.639
	F_{15}	0.206	0.291	0.398	0.521	0.639
	Φ	0.163	0.233	0.324	0.438	0.570
MCD	F_5	0.069	0.124	0.206	0.307	0.403
$\alpha = 0.5$	F_8	0.055	0.103	0.178	0.283	0.397
	F_{15}	0.045	0.085	0.151	0.252	0.380
	Φ	0.033	0.063	0.113	0.191	0.304
C	F_5	0.429	0.417	0.400	0.378	0.352
	F_8	0.800	0.788	0.770	0.741	0.702
	F_{15}	0.946	0.940	0.931	0.914	0.884
	Φ	1.000	1.000	1.000	1.000	1.000

TABLE IV

Asymptotic Efficiencies of a Diagonal Element of the MCD Scatter Matrix Estimator at the Cauchy Distribution for Several Values of p with $\alpha = 0.25$ and 0.5

	$p=2$	$p=3$	$p=5$	$p=10$	$p=30$
$\alpha=0.25$	0.443	0.347	0.279	0.229	0.198
$\alpha=0.5$	0.356	0.322	0.294	0.266	0.247

4. RELATED ROBUST ESTIMATORS OF SCATTER

Since the efficiency of high breakdown methods at normal distributions can be quite low, it is often recommended to compute reweighted versions of them, which maintain the breakdown point of the initial estimators while attaining (hopefully) a better efficiency. Quite recently, Lopuhaä (1997) derived asymptotic properties of reweighted multivariate estimators of location and scatter. Denote (T^0, Σ^0) initial robust estimators of multivariate location and scatter. For a sample $\{x_1, ..., x_n\}$ one step reweighted estimators are computed as

$$T^1 = \frac{\sum_{i=1}^n w_i x_i}{\sum_{i=1}^n w_i} \qquad \text{and} \qquad \Sigma^1 = c_1 \frac{\sum_{i=1}^n w_i (x_i - T^1)(x_i - T^1)^t}{\sum_{i=1}^n w_i}, \tag{4.1}$$

where the weights are computed from the initial estimators by

$$w_i = w((X_i - T^0)^t (\Sigma^0)^{-1} (X_i - T^0))$$

with $w\colon [0, \infty[\to \mathbb{R}$ a suitable weight function.

A simple, but common choice is to take

$$w(t) = \mathrm{I}_{[0, q_\delta]}(t) \qquad \text{with } q_\delta = G^{-1}(1-\delta), \tag{4.2}$$

where $G(u) = P_F(X^t X \leqslant u)$. At the p-dimensional normal model we have $q_\delta = \chi^2_{p, 1-\delta}$. The consistency factor c_1 is given by

$$c_1 = (1-\delta) \left\{ \frac{\pi^{p/2}}{\Gamma((p/2)+1)} \int_0^{\sqrt{q_\delta}} r^{p+1} g(r^2)\, dr \right\}^{-1}.$$

The influence function of Σ^1 at the model distribution F for the weight function (4.2) follows from Lopuhaä (1997),

$$IF(x, \Sigma^1, F) = \frac{d_2 + 2d_3}{d_2} \left(IF(x, \Sigma^0, F) + \frac{1}{2} \operatorname{trace}(IF(x, \Sigma^0, F))\, I \right)$$

$$+ \frac{1}{d_2} I(x^t x \leqslant q_\delta)\, x x^t - I,$$

where the constant d_2 equals $(1-\delta)\, c_1^{-1}$ and the expression for d_3 is the same as for c_3 (given in Theorem 1) but with α replaced by δ.

Taking the MCD estimator for (T^0, Σ^0) and $\delta = 0.025$ in (4.2) is advocated and used by (Rousseeuw and Van Driessen, 1999) and we denote the resulting estimator by MCD^1. Figure 3 gives the influence functions for a diagonal and an off-diagonal element of the MCD^1 scatter estimator in the bivariate case for $\alpha = 0.25$. We observe two jumps: one due to the discontinuity of the weight function and the other due to the discontinuity in the influence function of the initial MCD estimator. Note that the influence of extreme outliers on MCD^1 is smaller than on the ordinary MCD.

One could also iterate the process and define MCD^2 as a reweighted estimator using MCD^1 as starting estimator in (4.1). The general belief is that reiterating could increase efficiency, but at the cost of a higher bias (Rousseeuw and Croux, 1994).

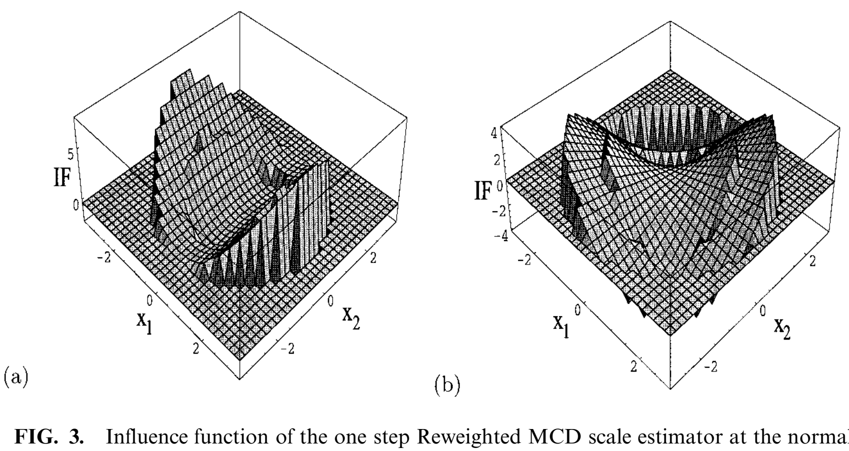

FIG. 3. Influence function of the one step Reweighted MCD scale estimator at the normal model, with $p = 2$, and $\alpha = 0.25$, (a) for the first diagonal element of the scatter matrix (b) for an off-diagonal element of the scatter matrix.

As a competitor for the MCD estimator, S-estimators will be considered. Recall that the S-functional $(t(G), S(G))$ is defined as the minimizer of $\det(S)$ subject to

$$\int \rho\left(\sqrt{(x-t)^t\, S^{-1}(x-t)}\right) dF(x) = b_0$$

among all (t, S), with $t \in \mathbb{R}^p$ and $S \in \mathrm{PD}(p)$. The function ρ: $[0, \infty[\rightarrow [0, +\infty[$ should be bounded, increasing and sufficiently smooth. A standard choice is Tuckey's biweight function:

$$\rho_c(y) = \min\left(\frac{y^2}{2} - \frac{y^4}{2c^2} + \frac{y^6}{6c^4}, \frac{c^2}{6}\right).$$

To obtain Fisher consistency at the model distribution F take $b_0 = E_F[\rho_c(\|x\|)]$. In order to attain a breakdown point of α, one needs to select the constant c as the solution of the equation $b_0 = \alpha\rho_c(\infty)$. The reweighted version of an S-estimator which will be considered, S^1, is obtained using the biweight S as initial estimator in (4.1). Asymptotics of S-estimators were derived by Davies (1987) and an expression for the influence function can be found in Lopuhaä (1989). Figure 4 pictures $IF(x, \mathrm{S}_{11}, \Phi)$ and $IF(x, \mathrm{S}_{12}, \Phi)$, for the 25% breakdown biweight S-estimator. Comparison with Figure 1 reveals that the IF for the S-estimator can be considered as a smoothed version of the MCD's influence function.

Some conclusions about the comparison of the different influence functions can now be given. The influence function of the MCD estimator looks like that of the usual covariance matrix at the center of the distribution.

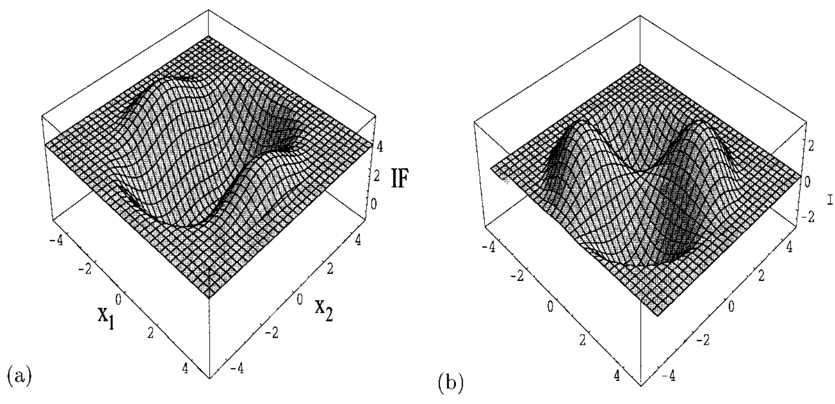

FIG. 4. Influence function for the 25% breakdown biweight S-estimator at the normal model, with $p = 2$, and $\alpha = 0.25$, (a) for the first diagonal element of the scatter matrix (b) for an off-diagonal element of the scatter matrix.

However, points further away (which can be considered as outliers) are downweighted by MCD. The same holds for the influence function of MCD^1 but there the downweighting happens for points somewhat further away from the origin. Moreover, two circles of discontinuity instead of one appear in the graph of $IF(x, \mathrm{MCD}^1, \Phi)$. The big advantage of the S-estimator is that its influence function is very smooth, also downweights outliers and resembles the influence function of the covariance matrix at the center of the distribution. One could conclude that, with respect to influence functions, an S-estimator is to be preferred.

A comparative study of the asymptotic efficiencies of the elements of the scatter matrix estimators MCD^1, MCD^2, S, and S^1 has been done, of which the results are reported in Tables V and VI. As before, the two cases $\alpha = 25\,\%$ and 50 % breakdown point are considered.

TABLE V

Asymptotic Efficiency of a Diagonal Element of the Reweighted MCD and S Scatter Matrices with $\alpha = 0.25$ and 0.5 and of the Covariance Matrix C at Several Student Distributions.

α			$p=2$	$p=3$	$p=5$	$p=10$	$p=30$
0.25	Φ	MCD^1	0.599	0.680	0.753	0.836	0.901
		MCD^2	0.573	0.656	0.727	0.806	0.872
		S	0.899	0.941	0.968	0.990	0.997
		S^1	0.678	0.745	0.793	0.853	0.905
		C	1.000	1.000	1.000	1.000	1.000
	F_5	MCD^1	0.760	0.743	0.698	0.655	0.577
		S	0.899	0.876	0.835	0.765	0.691
		C	0.375	0.357	0.333	0.304	0.272
	F_8	MCD^1	0.739	0.778	0.778	0.776	0.725
		S	0.930	0.942	0.923	0.884	0.819
		C	0.762	0.743	0.714	0.672	0.618
	F_{15}	MCD^1	0.695	0.760	0.809	0.843	0.821
		S	0.932	0.962	0.980	0.965	0.906
		C	0.935	0.926	0.912	0.886	0.841
0.5	Φ	MCD^1	0.455	0.595	0.720	0.820	0.896
		MCD^2	0.572	0.651	0.731	0.808	0.877
		S	0.502	0.647	0.803	0.920	0.973
		S^1	0.633	0.706	0.782	0.849	0.909
	F_5	MCD^1	0.702	0.737	0.709	0.668	0.586
		S	0.639	0.718	0.778	0.796	0.783
	F_8	MCD^1	0.651	0.749	0.780	0.786	0.730
		S	0.603	0.712	0.810	0.859	0.872
	F_{15}	MCD^1	0.579	0.706	0.799	0.844	0.825
		S	0.564	0.694	0.828	0.925	0.933

TABLE VI

Asymptotic Efficiency of an Off-diagonal Element of the Reweighted MCD and S Scatter Matrices with $\alpha = 0.25$ and 0.5 and of the Covariance Matrix C at Several Student Distributions

α			$p=2$	$p=3$	$p=5$	$p=10$	$p=30$
0.25	Φ	MCD^1	0.637	0.736	0.814	0.878	0.928
		MCD^2	0.710	0.773	0.829	0.881	0.929
		S	0.850	0.924	0.967	0.988	0.997
		S^1	0.749	0.804	0.849	0.891	0.932
		C	1.000	1.000	1.000	1.000	1.000
	F_5	MCD^1	0.813	0.822	0.807	0.772	0.709
		S	0.883	0.929	0.936	0.905	0.851
		C	0.429	0.417	0.400	0.378	0.352
	F_8	MCD^1	0.794	0.856	0.873	0.866	0.809
		S	0.888	0.951	0.975	0.958	0.911
		C	0.800	0.788	0.770	0.741	0.702
	F_{15}	MCD^1	0.737	0.819	0.884	0.912	0.892
		S	0.882	0.954	0.990	0.991	0.961
		C	0.946	0.940	0.931	0.914	0.884
0.5	Φ	MCD^1	0.401	0.618	0.783	0.873	0.934
		MCD^2	0.684	0.765	0.836	0.884	0.934
		S	0.377	0.579	0.778	0.915	0.979
		S^1	0.682	0.765	0.842	0.890	0.934
	F_5	MCD^1	0.683	0.794	0.811	0.779	0.712
		S	0.466	0.637	0.785	0.879	0.910
	F_8	MCD^1	0.619	0.804	0.871	0.873	0.812
		S	0.440	0.626	0.798	0.913	0.956
	F_{15}	MCD^1	0.531	0.742	0.871	0.915	0.894
		S	0.414	0.611	0.800	0.933	0.985

First of all, note that there is no real efficiency gain using MCD^2 instead of MCD^1, at least for $\alpha = 0.25$. In the 50% breakdown case, where the efficiency of the initial estimator MCD is very low (cf. Fig. 2), two times reweighting may be an option. A comparison with Tables II and III shows that a considerable asymptotic efficiency gain is obtained by reweighting the MCD-estimator. Also at the heavier tailed Student distributions, MCD^1 is more precise than ordinary MCD and even achieves a better efficiency than the classical estimator for small values of ν. As a preliminary conclusion, one can say that the one step reweighted 25% breakdown MCD seems to be the best out of the class of MCD-based estimators.

On the other hand, the 25% breakdown S-estimator outperforms all the others in the normal case (except the classical estimator) and under the Student distributions. Using the reweighted S-estimator for efficiency reasons has not much sense, unless for $\alpha = 50\%$ and p small. The ordinary S-estimator combines a high efficiency with a high breakdown point, yielding a very appealing estimator. Rocke (1996) noticed that the efficiency of *S*-estimators tends to 100% when the dimension tends to infinity, but argued that S-estimators are in fact not so robust since their *asymptotic rejection probability* is extremely small in high dimensions. One may not forget that a positive breakdown point is not a guarantee for robustness, since the corresponding bias may become extremely large, but still remain bounded.

5. FINITE SAMPLE EFFICIENCIES

The results given in the preceeding sections are of an asymptotic nature. In this section, finite-sample efficiencies for the MCD, S and their reweighted versions are obtained by means of a simulation study. Both estimators are defined as minimizers of a certain criterion under an additional constraint, implying that it is not so obvious to compute them in practice. Fortunately, algorithms were proposed which give approximations to the actual value of the estimator. For computing the MCD the FAST-MCD algorithm of Rousseeuw and Van Driessen (1999) was used, while S-estimators are based on the SURREAL algorithm of Ruppert (1992). Both algorithms start from an initial mean and covariance matrix obtained from a $(p+1)$ *subset* of observations, which is iterated towards better approximations using Newton-steps (for S) or the so-called *C-steps* which are used in the FAST-MCD algorithm. This choice for the starting values guarantees that the computed version of the estimator shares the robustness properties of the theoretical counterpart. Implementation of the algorithms was done in GAUSS and in both cases 500 different starting values were considered for each computation of the estimator. To illustrate the fastness of the algorithms: at a (50×5) normal data set it took about one second to compute the MCD and half a second for the S-estimator on a Pentium 200Mhz. One may certainly say that computational feasibility is no longer an obstacle for the use of high breakdown methods (at least not for small p). Furthermore, note that reweighting the estimators comes at almost no additional computational cost.

For $m = 5000$ samples of sizes $n = 50$ and 200, observations were generated from a $\mathrm{N}(0, \mathrm{I}_p)$ with $p = 2, 3, 5$ or 10. Denote $\hat{\Sigma}^k_{ij}$ the element (i, j) of the estimator obtained from the kth sample, with $1 \leqslant k \leqslant m$. The accuracy of a diagonal element is measured by the *standardized variance*

$$\mathrm{StVar}(\hat{\Sigma}_{ii}) = \frac{n \operatorname{var}_m(\hat{\Sigma}_{ii})}{[\operatorname{ave}_m(\hat{\Sigma}_{ii})]^2}, \tag{5.1}$$

where $\text{ave}_m(\hat{\Sigma}_{ii})$ and $\text{var}_m(\hat{\Sigma}_{ii})$ are the average and variance computed from the sequence of m replicates $\hat{\Sigma}^k_{ij}$. (Use of measure (5.1) is motivated by Bickel and Lehmann, page 1142, 1976). For an off-diagonal element the mean squared error (MSE) is used to measure the deviation from the true value,

$$\text{MSE}(\hat{\Sigma}_{ij}) = \frac{n}{m} \sum_{k=1}^{m} (\hat{\Sigma}^k_{ij} - I_{ij})^2, \tag{5.2}$$

with I_{ij} of course equal to zero for $i \neq j$.

The simulation results are summarized in Tables VII and VIII. Since 2 is the lower bound for (5.1), the reported efficiencies in Table VII equal $2/\text{ave}_p \text{StVar}(\hat{\Sigma}_{ii})$. Efficiencies for the off-diagonal elements are obtained in a similar way, now with 1 as lower bound. The standard error of any value reported in the tables equals approximately 2% of the value.

First off all, note that the finite-sample efficiencies of the MCD converge well to the asymptotic ones which are listed under $n = \infty$ in the tables. There is some discrepancy for $p = 10$, which enforces the idea that convergence to the asymptotic distribution is slower in higher dimensions. Also, it might happen that for $p = 10$, 500 starting values are not enough to ensure that the actual estimate resembles the exact one. For the one step reweighted MCD there is quite serious loss of efficiency at finite samples, certainly for $p = 5$ and $p = 10$, but it still dominates the ordinary MCD. We can repeat the conclusions of Section 4 here: (a) reweighted MCD with $\alpha = 0.25$ yields very reasonable efficiencies (b) 25% breakdown

TABLE VII

Finite-Sample Efficiencies of Diagonal Elements of the MCD, S, and Reweighted MCD and S Scatter Matrix Estimators at the Normal Distribution

	$p=2$			$p=3$			$p=5$			$p=10$		
$n =$	50	200	∞	50	200	∞	50	200	∞	50	200	∞
$\alpha = 0.25$												
MCD	0.299	0.275	0.262	0.358	0.315	0.300	0.424	0.386	0.366	0.528	0.484	0.577
MCD^1	0.575	0.629	0.599	0.604	0.701	0.680	0.587	0.749	0.753	0.543	0.774	0.836
S	0.877	0.873	0.899	0.917	0.935	0.941	0.955	0.957	0.968	0.970	0.996	0.990
S^1	0.736	0.728	0.678	0.783	0.788	0.745	0.842	0.828	0.793	0.903	0.895	0.853
$\alpha = 0.5$												
MCD	0.105	0.071	0.062	0.150	0.103	0.089	0.220	0.157	0.134	0.330	0.247	0.205
MCD^1	0.376	0.423	0.455	0.373	0.497	0.595	0.355	0.575	0.720	0.350	0.628	0.820
S	0.435	0.468	0.502	0.582	0.633	0.647	0.748	0.782	0.803	0.869	0.920	0.920
S^1	0.597	0.647	0.633	0.662	0.733	0.706	0.758	0.799	0.782	0.834	0.875	0.849

TABLE VIII

Finite-Sample Efficiencies of Off-Diagonal Elements of the MCD, S, and Reweighted MCD and S Scatter Matrix Estimators at the Normal Distribution

	$p=2$			$p=3$			$p=5$			$p=10$		
$n=$	50	200	∞	50	200	∞	50	200	∞	50	200	∞
$\alpha=0.25$												
MCD	0.191	0.169	0.163	0.252	0.235	0.233	0.359	0.359	0.324	0.462	0.450	0.438
MCD^1	0.643	0.649	0.637	0.678	0.714	0.736	0.726	0.823	0.814	0.692	0.820	0.878
S	0.917	0.820	0.850	0.930	0.893	0.924	0.976	1.020	0.967	1.019	1.000	0.988
S^1	0.792	0.725	0.749	0.781	0.769	0.804	0.846	0.879	0.849	0.941	0.909	0.891
$\alpha=0.5$												
MCD	0.046	0.036	0.033	0.079	0.064	0.063	0.140	0.123	0.113	0.221	0.205	0.191
MCD^1	0.440	0.407	0.401	0.472	0.500	0.618	0.515	0.671	0.783	0.502	0.714	0.873
S	0.332	0.357	0.377	0.513	0.532	0.579	0.711	0.810	0.778	0.924	0.909	0.915
S^1	0.697	0.667	0.682	0.730	0.735	0.767	0.803	0.850	0.842	0.904	0.905	0.890

S-estimators have the highest efficiencies among the considered estimators: almost always above 90% (c) reweighting an S-estimator (except for p small and $\alpha=0.5$) leads to a small loss in precision.

Since the MCD and S estimators are meant to be robust estimators of multivariate location and scatter, we compared their behavior at contaminated normal distributions. The generated samples contain now 20% of outliers equal to $k\,\vec{e_1}$ where $\vec{e_1}$ is the first unit vector, while the remaining 80% observations are again normally distributed. The contamination will affect the whole scatter-matrix, and the median squared error as well as the 0.9 quantile of the squared errors (reported between parenthesis) have been computed for several values of k. Using quantiles of the squared errors instead of the mean squared error was suggested by a referee who argued that MSEs may be misleading under this type of contamination, but the obtained results appeared to be comparable.

In Tables IX and X results for two representative cases, $\hat{\Sigma}_{11}$ and $\hat{\Sigma}_{12}$, for $k=100$ and $k\approx\sqrt{q_\delta}$ (with $\delta=0.025$) are reported for the 25% and 50% breakdown point MCD, MCD^1, S and S^1 estimators. Note that $k=100$ gives samples containing far away outliers which are easily detected by the robust estimators. The MCD estimator is however more sensible to intermediate outliers, which are much harder to detect. Some simulation experiments indicated that $k\approx\sqrt{q_\delta}$ is close to the value resulting in the maximal squared errors one can expect for this type of contamination. One sees that the reweighted MCD remains much more precise than its initial estimator even under severe contamination and that it also outperforms the

TABLE IX

Median Squared Error and 0.9 Quantile of the Squared Errors (between Parentheses) of Element (1,1) of the MCD, S, and Reweighted MCD and S Scatter Matrix Estimators at a 20% Contaminated Normal Distribution

		$k=100$				$k\approx\sqrt{q_\delta}$			
		$p=2$		$p=3$		$p=2$		$p=3$	
α		$n=50$	$n=200$	$n=50$	$n=200$	$n=50$	$n=200$	$n=50$	$n=200$
0.25	MCD	15.15	58.40	8.09	30.88	315.43	1265.97	378.75	1480.03
		(64.0)	(137.0)	(42.9)	(82.8)	(540.8)	(1660.4)	(582.8)	(1841.1)
	MCD^1	1.50	2.00	1.41	1.66	62.23	248.60	106.42	405.79
		(9.4)	(12.6)	(8.1)	(9.8)	(109.1)	(332.2)	(171.0)	(517.9)
	S	611.60	2642.38	564.38	2459.76	64.63	268.39	108.62	447.09
		(1225.3)	(3800.6)	(1146.9)	(3533.5)	(113.8)	(360.7)	(170.4)	(562.2)
	S^1	1.45	2.39	1.33	1.92	60.11	233.40	97.42	380.03
		(9.8)	(14.5)	(8.8)	(11.9)	(102.4)	(312.3)	(153.2)	(480.5)
0.5	MCD	10.99	26.34	9.51	18.82	1541.11	6266.32	1392.39	5281.43
		(138.6)	(186.4)	(91.7)	(121.3)	(2763.5)	(8574.7)	(2254.7)	(6888.2)
	MCD^1	1.78	1.75	1.91	1.57	73.45	305.99	144.66	519.21
		(9.87)	(11.07)	(10.48)	(9.28)	(145.93)	(420.31)	(247.74)	(676.94)
	S	15.14	68.75	11.98	54.85	270.21	1013.36	358.03	1300.84
		(67.5)	(154.7)	(52.8)	(120.2)	(616.8)	(1478.7)	(643.5)	(1691.2)
	S^1	1.53	2.04	1.40	1.80	62.83	244.80	103.84	393.80
		(9.2)	(12.8)	(8.5)	(10.6)	(108.6)	(329.8)	(165.3)	(501.9)

S-estimator in the case of extreme outliers. Although the percentage of outliers is close to the breakdown point of the MCD^1 estimator with $\alpha=0.25$, the median squared error of the latter is comparable to that of the maximal breakdown point MCD^1. It is surprising to note that the MCD estimator performs better with $\alpha=0.25$ than with $\alpha=0.5$. The S-estimator with $\alpha=0.25$, which was the most efficient in the uncontaminated case, becomes very unstable at the 20% contamination level. On the other hand, the reweighted S-estimator is much more robust but does not outperform the MCD^1 estimator. The (small) loss in efficiency paid for by reweighting the S-estimator, is apparently compensated by a gain in robustness. This in contrast with the MCD, where the reweighted version performs better both in presence and absence of contamination.

To conclude the simulation study, one could say that reweighted S and MCD with 25% breakdown are the most appealing among all considered estimators. Comparing MCD^1 with S^1 is slightly favorable for the latter due to its better efficiency at the normal model, especially at finite samples

TABLE X

Median Squared Error and 0.9 Quantile of the Squared Errors (between Parentheses) of Element (1,2) of the MCD, S, and Reweighted MCD and S Scatter Matrix Estimators at a 20% Contaminated Normal Distribution

.

		$k=100$				$k \approx \sqrt{q_\delta}$			
		$p=2$		$p=3$		$p=2$		$p=3$	
α		$n=50$	$n=200$	$n=50$	$n=200$	$n=50$	$n=200$	$n=50$	$n=200$
0.25	MCD	2.22	2.55	1.68	1.66	3.56	4.28	2.50	2.72
		(13.3)	(14.8)	(10.0)	(10.2)	(21.1)	(25.4)	(14.8)	(15.9)
	MCD^1	0.73	0.73	0.68	0.67	0.79	0.83	0.80	0.71
		(4.3)	(4.4)	(4.2)	(4.1)	(4.9)	(4.8)	(5.0)	(4.3)
	S	11.52	12.14	11.24	11.77	0.55	0.57	0.56	0.56
		(71.9)	(72.2)	(68.7)	(71.4)	(3.3)	(3.2)	(3.3)	(3.4)
	S^1	0.67	0.68	0.65	0.67	0.58	0.59	0.62	0.61
		(4.2)	(4.1)	(4.1)	(4.0)	(3.7)	(3.5)	(3.8)	(3.7)
0.5	MCD	10.02	13.17	5.60	7.04	26.90	54.99	12.83	18.80
		(60.1)	(76.8)	(34.3)	(42.0)	(139.3)	(289.5)	(77.6)	(118.2
	MCD^1	0.86	0.85	0.88	0.77	2.28	4.07	1.96	1.98
		(5.2)	(5.1)	(5.2)	(4.7)	(12.3)	(20.7)	(11.6)	(12.2)
	S	2.40	2.39	1.68	1.64	2.34	2.31	1.60	1.44
		(14.8)	(13.9)	(10.7)	(10.1)	(18.0)	(13.7)	(11.1)	(9.1)
	S^1	0.73	0.73	0.70	0.69	0.80	0.72	0.71	0.69
		(4.4)	(4.2)	(4.1)	(4.1)	(4.9)	(4.2)	(4.5)	(3.9)

in higher dimensions. Let us not forget, however, the very intuitive finite-sample definition of the MCD, which could make MCD^1 more attractive to non-specialists in the field.

6. CONCLUSION

The development and availability of fast algorithms (Hawkins, 1994; Rousseeuw and Van Driessen, 1999) for computing the minimum covariance determinant (MCD) has brought renewed interest to this estimator. Asymptotic properties were given in Butler, Davies and Jhun (1993), but the asymptotic variance of the MCD-scatter part remained unknown. In this paper, we worked out the influence function of the MCD scatter matrix estimator and used it to evaluate the asymptotic efficiency of this robust estimator. The efficiencies of other robust estimators, whose influence function has already been derived in the literature, i.e., the

S-estimator and reweighted estimators, have also been evaluated numerically.

The main conclusion is that the Gaussian efficiency of the maximal breakdown MCD estimator is rather poor, but is already substantially higher for the 25% breakdown MCD and increases even more after reweighting. The efficiency of the reweighted MCD with $\alpha = 0.25$ seems to be acceptable: almost always above 60% in the Gaussian case, even for finite samples.

With respect to efficiency, 25% breakdown S-estimators are very attractive and outperform the reweighted MCD estimators. Davies (1992b) and Lopuhaä (1991) proposed alternatives to S-estimators with a higher efficiency, but these improvements seem only to be worthwhile in the 50% breakdown point case with p small. In spite of their positive breakdown property, the bias of S-estimators can be considerably high. Yohai and Maronna (1990) showed this by means of the *Maxbias curve*, which is a powerful tool to quantify the robustness of an estimator. It is therefore not sufficient to consider only breakdown point and efficiency of robust estimators, also Maxbias curves should be computed. This has been done for the MCD-estimator in the univariate case (Croux and Haesbroeck, 1999), but the multivariate case seems to be rather hard to handle.

7. APPENDIX

Proof of Proposition 1. Call $G = F_{\varepsilon, x}$, $t_\varepsilon = T_{(A, y)}(G)$, $\Sigma_\varepsilon = \Sigma_{(A, y)}(G)$ and $d_\varepsilon^2(z) = (z - t_\varepsilon)^t \, \Sigma_\varepsilon^{-1} (z - t_\varepsilon)$. Define for every $s > 0$

$$\mathscr{E}_s = \{ z \in \mathbb{R}^p \mid d_\varepsilon^2(z) < s \},$$

denote

$$D^2(G) = \sup \{ s > 0 \mid P_G(\mathscr{E}_s) \leqslant 1 - \alpha \},$$

and take $\mathscr{E} = \mathscr{E}_{D^2(G)}$. Since $\lim_{s \to \infty} P_G(\mathscr{E}_s) > 1 - \alpha$ and $\lim_{s \to 0} P_G(\mathscr{E}_s) \leqslant \varepsilon < 1 - \alpha$, $\mathscr{E}$ is a well defined, bounded and non degenerate open ellipsoid with $1 - \alpha - \varepsilon \leqslant P_G(\mathscr{E}) \leqslant 1 - \alpha$. We conclude that $\mathscr{E} \in \mathscr{D}_G(\alpha)$ or $(\mathscr{E}, x) \in \widetilde{\mathscr{D}}_G(\alpha)$, since x is the only atom of the distribution G.

Consider the two following probability measures $v_{(A, y)}$ and $v_{(\mathscr{E}, x)}$ defined for each measurable set B by

$$v_{(A, y)}(B) = (P_G(A \cap B) + \delta I(y \in B))/(1 - \alpha),$$

with $\delta = 1 - \alpha - P_G(A)$ and

$$\nu_{(\mathscr{E}, x)}(B) = (P_G(\mathscr{E} \cap B) + \tilde{\delta} I(x \in B))/(1-\alpha),$$

where $\tilde{\delta} = 1 - \alpha - P_G(\mathscr{E})$. By definition of $\mathscr{E}$ one has that $\tilde{\delta} = 0$, as long as the atom x does not lie at the border of $\mathscr{E}$. Therefore

$$\tilde{\delta}\,(d_\varepsilon^2(x) - D^2(G)) = 0. \tag{7.1}$$

It is sufficient to show that

$$E_{\nu_{(\mathscr{E}, x)}}[d_\varepsilon^2(z)] \leqslant E_{\nu_{(A, y)}}[d_\varepsilon^2(z)]. \tag{7.2}$$

Indeed, if the above equation holds, we know that there exists a $0 < c \leqslant 1$ for which

$$E_{\nu_{(\mathscr{E}, x)}}[(z - t_\varepsilon)^t (c\Sigma_\varepsilon)^{-1} (z - t_\varepsilon)] = p, \tag{7.3}$$

since the RHS of (7.2), like the sum of the diagonal elements of the hat matrix, equals p. Since (A, y) provides an MCD solution we have that $\det(c\Sigma_\varepsilon) \leqslant \det(\Sigma_\varepsilon) \leqslant \det(\Sigma_{(\mathscr{E}, x)}(G))$, which in combination with (7.3) contradicts the result of Grübel (1988) unless $t_\varepsilon = T_{(\mathscr{E}, x)}(G)$, and $c\Sigma_\varepsilon = \Sigma_{(\mathscr{E}, x)}(G)$. Then also c should equal 1, which will end the proof.

To prove (7.2), write

$$E_{\nu_{(\mathscr{E}, x)}}[d_\varepsilon^2(z)] = \frac{1}{1-\alpha} \left\{ \int_{\mathscr{E} \cap A} d_\varepsilon^2(z)\, dG(z) \right.$$
$$\left. + \int_{\mathscr{E} \setminus (A \cup \{y\})} d_\varepsilon^2(z)\, dG(z) + d_\varepsilon^2(y)\, P_G(\mathscr{E} \cap \{y\}) + \tilde{\delta}\, d_\varepsilon^2(x) \right\}. \tag{7.4}$$

Now

$$\begin{aligned}\int_{\mathscr{E} \setminus (A \cup \{y\})} d_\varepsilon^2(z)\, dG(z) &\leqslant D^2(G)\, P_G(\mathscr{E} \setminus (A \cup \{y\}))\\ &= D^2(G)(1 - \alpha - \tilde{\delta} - P_G(\mathscr{E} \cap A) - P_G(\mathscr{E} \cap \{y\}))\\ &= D^2(G)(P_G(A \setminus \mathscr{E}) + \delta - \tilde{\delta} - P_G(\mathscr{E} \cap \{y\}))\\ &\leqslant \int_{A \setminus \mathscr{E}} d_\varepsilon^2(z)\, dG(z) + D^2(G)(\delta - \tilde{\delta} - P_G(\mathscr{E} \cap \{y\})).\end{aligned} \tag{7.5}$$

Combining (7.4) and (7.5) with (7.1), yields

$$
\begin{aligned}
E_{\nu(\mathscr{E},x)}[d_\varepsilon^2(z)] &\leqslant \frac{1}{1-\alpha}\left\{\int_A d_\varepsilon^2(z)\,dG(z)\right.\\
&\qquad \left.+D^2(G)(\delta-P_G(\mathscr{E}\cap\{y\}))+d_\varepsilon^2(y)\,P_G(\mathscr{E}\cap\{y\})\right\}\\
&= E_{\nu(A,y)}[d_\varepsilon^2(z)]+\frac{1}{1-\alpha}\{(D^2(G)-d_\varepsilon^2(y))(\delta-P_G(\mathscr{E}\cap\{y\}))\}.
\end{aligned}
\tag{7.6}
$$

The second term of (7.5) is always negative, since $D^2(G)\leqslant d_\varepsilon^2(y)$ for $y\notin\mathscr{E}$ and $D^2(G)\geqslant d_\varepsilon^2(y)$, but $\delta\leqslant P_G(\{y\})$ for $y\in\mathscr{E}$. Therefore, (7.2) holds. ▮

Proof of Theorem 1: For ease of notation, denote $F_{\varepsilon,x_0}=(1-\varepsilon)\,F+\varepsilon\Delta_{x_0}$, where x_0 is an arbitrary point, $\Sigma(F_{\varepsilon,x_0})=\Sigma_\varepsilon$ and $T(F_{\varepsilon,x_0})=t_\varepsilon$. Proposition 1 guarantees that the MCD scatter estimator at F_{ε,x_0} satisfies

$$
\begin{aligned}
\Sigma_\varepsilon &= c_\alpha\left\{\frac{1}{1-\alpha}\int_{A(F_{\varepsilon,x_0})} xx^t\,dF_{\varepsilon,x_0}(x)-t_\varepsilon t_\varepsilon^t\right\}\\
&= c_\alpha\left\{\frac{1-\varepsilon}{1-\alpha}\int_{A(F_{\varepsilon,x_0})} xx^t\,dF(x)+\frac{\varepsilon}{1-\alpha}\,\mathrm{I}(x_0\in A(F_{\varepsilon,x_0}))\,x_0x_0^t-t_\varepsilon t_\varepsilon^t\right\},
\end{aligned}
$$

where $A(F_{\varepsilon,x_0})=\{x\in\mathbb{R}^p\colon (x-t_\varepsilon)^t\,\Sigma_\varepsilon^{-1}(x-t_\varepsilon)<q_\alpha(\varepsilon)\}$, for a certain positive number $q_\alpha(\varepsilon)$. (As long as $\|x_0\|^2\neq q_\alpha$ one may suppose that x_0 does not belong to the border of $A(F_{\varepsilon,x_0})$ for ε small, cf. Remark 4, Section 2.) By definition, $IF(x_0,\Sigma,F)=(\partial\Sigma_\varepsilon/\partial\varepsilon)_{|\varepsilon=0}$. Let us differentiate Σ_ε w.r.t. ε and set $\varepsilon=0$,

$$
\begin{aligned}
\frac{\partial\Sigma_\varepsilon}{\partial\varepsilon}_{|\varepsilon=0} &= c_\alpha\left\{-\frac{1}{1-\alpha}\int_{A(F)} xx^t\,dF(x)+\frac{1}{1-\alpha}\frac{\partial}{\partial\varepsilon}\int_{A(F_{\varepsilon,x_0})} xx^t\,dF(x)_{|\varepsilon=0}\right.\\
&\qquad\left.+\frac{1}{1-\alpha}\,\mathrm{I}(x_0\in A(F))\,x_0x_0^t\right\}.
\end{aligned}
\tag{7.7}
$$

Using the Fisher consistency of $\Sigma(F)$ and the fact that $A(F)=\{x\in\mathbb{R}^p : x^tx\leqslant q_\alpha\}$, we get

$$
\begin{aligned}
IF(x_0,\Sigma,F) &= -I+\frac{c_\alpha}{1-\alpha}\frac{\partial}{\partial\varepsilon}\int_{A(F_{\varepsilon,x_0})} xx^t\,dF(x)_{|\varepsilon=0}\\
&\qquad+\frac{c_\alpha}{1-\alpha}\,\mathrm{I}(\|x_0\|^2\leqslant q_\alpha)\,x_0x_0^t.
\end{aligned}
\tag{7.8}
$$

In order to compute the second term on the right hand side of (7.8), we transform the integration variable x to $y=\Sigma_{\varepsilon}^{-1/2}(x-t_{\varepsilon})$. The integration domain becomes now a ball with center at the origin and radius $\sqrt{q_\alpha(\varepsilon)}$. When more convenient, we will express y in polar coordinates $y=r\ e(\theta)$ with $r\in[0, \sqrt{q_\alpha(\varepsilon)}]$, $e(\theta)\in S^{p-1}$ and $\theta=(\theta_1, ..., \theta_{p-1})\in\Theta=[0, \pi[\times\cdots[0, \pi[\times[0, 2\pi[$. The integral in (7.8), denoted by $I(\varepsilon)$, becomes

$$I(\varepsilon)=\det(\Sigma_\varepsilon^{1/2})\int_0^{\sqrt{q_\alpha(\varepsilon)}}\int_\Theta J(\theta, r)(r\Sigma_\varepsilon^{1/2}e(\theta)+t_\varepsilon)(r\Sigma_\varepsilon^{1/2}e(\theta)+t_\varepsilon)^t$$

$$g((r\Sigma_\varepsilon^{1/2}e(\theta)+t_\varepsilon)^t(r\Sigma_\varepsilon^{1/2}e(\theta)+t_\varepsilon))\,dr\,d\theta, \qquad (7.9)$$

where $J(\theta, r)$ is the Jacobian of the transformation into polar coordinates.

By matrix differentiation and since $\Sigma_0=I$, we obtain

$$\frac{\partial \det(\Sigma_\varepsilon^{1/2})}{\partial\varepsilon}\Big|_{\varepsilon=0}=\frac{1}{2}\frac{\partial \det(\Sigma_\varepsilon)}{\partial\varepsilon}\Big|_{\varepsilon=0}=\frac{1}{2}\,\mathrm{trace}\,(IF(x_0, \Sigma, F)). \qquad (7.10)$$

Applying Leibniz' formula to (7.9) and using (7.10) results in

$$\frac{\partial I(\varepsilon)}{\partial\varepsilon}\Big|_{\varepsilon=0}=\frac{1}{2}\,\mathrm{trace}(IF(x_0, \Sigma, F))\,H(q_\alpha)\,I+\frac{\partial\sqrt{q_\alpha(\varepsilon)}}{\partial\varepsilon}\Big|_{\varepsilon=0}\,q_\alpha g(q_\alpha)\,d_1 I$$

$$+\int_{\|y\|^2\leqslant q_\alpha}\frac{\partial}{\partial\varepsilon}((\Sigma_\varepsilon^{1/2}y+t_\varepsilon)(\Sigma_\varepsilon^{1/2}y+t_\varepsilon)^t g((\Sigma_\varepsilon^{1/2}y+t_\varepsilon)^t(\Sigma_\varepsilon^{1/2}y+t_\varepsilon)))\Big|_{\varepsilon=0}\,dy, \qquad (7.11)$$

with $d_1=\int_\Theta J(\theta, \sqrt{q_\alpha})\,e_1^2(\theta)\,d\theta=1/p\int_\Theta J(\theta, \sqrt{q_\alpha})\,d\theta$ and $H(q_\alpha)=\int_{\|y\|^2\leqslant q_\alpha} y_1^2 g(y^ty)\,dy=(c_\alpha/(1-\alpha))^{-1}$.

The derivative in the third term of (7.11) is given by

$$\frac{\partial}{\partial\varepsilon}((\Sigma_\varepsilon^{1/2}y+t_\varepsilon)(\Sigma_\varepsilon^{1/2}y+t_\varepsilon)^t\,g((\Sigma_\varepsilon^{1/2}y+t_\varepsilon)^t(\Sigma_\varepsilon^{1/2}y+t_\varepsilon)))\Big|_{\varepsilon=0}$$

$$=\frac{1}{2}\{IF(x_0, \Sigma, F)\,yy^t+yy^t\,IF(x_0, \Sigma, F)$$

$$+2IF(x_0, T, F)\,y^t+2yIF(x_0, T, F)^t\}\,g(y^ty)$$

$$+yy^tg'(y^ty)\{y^tIF(x_0, \Sigma, F)\,y+2y^tIF(x_0, T, F)\}. \qquad (7.12)$$

Note that since $\int_{\|y\|^2\leqslant q_\alpha} yg(y^ty)\,dy$ and $\int_{\|y\|^2\leqslant q_\alpha} yy^tg'(y^ty)\,y\,dy$ are zero, the terms in (7.12) including $IF(x_0, T, F)$ give a zero contribution to the integral in (7.11).

Equation (7.11) still depends on $((\partial\sqrt{q_\alpha(\varepsilon)})/\partial\varepsilon)_{|\varepsilon=0}$ which needs to be computed. By definition of $A(F_{\varepsilon, x_0})$,

$$1-\alpha=\int_{A(F_{\varepsilon, x_0})} dF_{\varepsilon, x_0}(x)=(1-\varepsilon)\int_{A(F_{\varepsilon, x_0})} dF(x)+\varepsilon \mathrm{I}(x_0\in A(F_{\varepsilon, x_0})). \quad (7.13)$$

Differentiating both sides w.r.t. ε yields

$$0=-\int_{A(F)} dF(x)+\frac{\partial}{\partial\varepsilon}\int_{A(F_{\varepsilon, x_0})} dF(x)_{|_{\varepsilon=0}}+\mathrm{I}(\|x_0\|^2\leqslant q_\alpha). \quad (7.14)$$

In the same way as before, one can easily verify that

$$\begin{aligned}\frac{\partial}{\partial\varepsilon}&\int_{A(F_{\varepsilon, x_0})} dF(x)_{|_{\varepsilon=0}}\\ &=\frac{1-\alpha}{2}\,\mathrm{trace}(IF(x_0, \Sigma, F))+\frac{\partial\sqrt{q_\alpha(\varepsilon)}}{\partial\varepsilon}_{|_{\varepsilon=0}}\, g(q_\alpha)\int_\Theta J(\theta, \sqrt{q_\alpha})\,d\theta\\ &\quad+\int_{\|y\|^2\leqslant q_\alpha} g'(y^ty)\,y^t\,IF(x_0, \Sigma, F)\,y\,dy. \end{aligned} \quad (7.15)$$

The last term in (7.15) equals, using the symmetry of F, $c_2\,\mathrm{trace}(IF(x_0, \Sigma, F))$, with $c_2=\int_{\|y\|^2\leqslant q_\alpha} y_1^2 g'(y^ty)\,dy$.

By substituting (7.15) in (7.14), the above equation leads to

$$\frac{\partial\sqrt{q_\alpha(\varepsilon)}}{\partial\varepsilon}_{|_{\varepsilon=0}}=\frac{1-\alpha-\mathrm{I}(\|x_0\|^2\leqslant q_\alpha)-\mathrm{trace}\,(IF(x_0, \Sigma, F))(c_2+(1-\alpha)/2)}{g(q_\alpha)\,pd_1}. \quad (7.16)$$

Inserting (7.16) and (7.12) in (7.11) gives

$$\begin{aligned}IF(x_0, \Sigma, F)=&\\ &-I+\frac{c_\alpha}{1-\alpha}\mathrm{I}(\|x_0\|^2\leqslant q_\alpha)\,x_0x_0^t+\frac{1}{2}\,\mathrm{trace}(IF(x_0, \Sigma, F))\,I\\ &+\frac{c_\alpha}{1-\alpha}\frac{q_\alpha}{p}\left((1-\alpha)-\mathrm{I}(\|x_0\|^2\leqslant q_\alpha)-\mathrm{trace}(IF(x_0, \Sigma, F))\left(c_2+\frac{1-\alpha}{2}\right)\right)I\\ &+\frac{c_\alpha}{2(1-\alpha)}\int_{\|y\|^2\leqslant q_\alpha}(IF(x_0, \Sigma, F)\,yy^t+yy^t\,IF(x_0, \Sigma, F))\,g(y^ty)\,dy\\ &+\frac{c_\alpha}{1-\alpha}\int_{\|y\|^2\leqslant q_\alpha} yy^tg'(y^ty)\,y^t\,IF(x_0, \Sigma, F)\,y\,dy. \end{aligned} \quad (7.17)$$

In order to give element wise expressions for the influence function, we note that

$$\frac{1}{2}\sum_{k=1}^{p}\left\{IF(x_0, \Sigma_{ik}, F)\int_{\|y\|^2\leqslant q_\alpha} y_k y_j g(y^t y)\,dy\right.$$

$$\left. + IF(x_0, \Sigma_{kj}, F)\int_{\|y\|^2\leqslant q_\alpha} y_i y_k g(y^t y)\,dy\right\}$$

$$= H(q_\alpha)\, IF(x_0, \Sigma_{ij}, F)$$

for every $1\leqslant i, j\leqslant p$ and

$$\sum_{k=1}^{p}\sum_{l=1}^{p} IF(x_0, \Sigma_{kl}, F)\int_{\|y\|^2\leqslant q_\alpha} y_i y_j y_k y_l g'(y^t y)\,dy$$

$$= \begin{cases} c_3 \operatorname{trace}(IF(x_0, \Sigma, F)) + (c_4 - c_3)\, IF(x_0, \Sigma_{ii}, F), & i = j \\ 2\, c_3\, IF(x_0, \Sigma_{ij}, F), & i \neq j, \end{cases}$$

where $c_3 = \int_{\|y\|^2\leqslant q_\alpha} y_i^2 y_j^2 g'(y^t y)\,dy$ and $c_4 = \int_{\|y\|^2\leqslant q_\alpha} y_i^4 g'(y^t y)\,dy$.

The constants c_2, c_3, and c_4 can be rewritten in the form given in the statement of the theorem, using polar coordinates. From (7.17) the influence function for the off-diagonal elements is immediately obtained,

$$IF(x_0, \Sigma_{ij}, F) = \frac{c_\alpha x_{0i} x_{0j}}{1-\alpha}\, \mathrm{I}(\|x_0\|^2 \leqslant q_\alpha) + \frac{c_\alpha}{1-\alpha}(2\, c_3 + H(q_\alpha))\, IF(x_0, \Sigma_{ij}, F),$$

which leads to (2.12), due to the definition of H. For the diagonal elements, we get

$$IF(x_0, \Sigma_{jj}, F) = -1 + \frac{c_\alpha x_{0j}^2}{1-\alpha}\, \mathrm{I}(\|x_0\|^2 \leqslant q_\alpha) + \frac{1}{2}\operatorname{trace}(IF(x_0, \Sigma, F))$$

$$+ \frac{c_\alpha}{1-\alpha}\frac{q_\alpha}{p}\left\{(1-\alpha) - \mathrm{I}(\|x_0\|^2 \leqslant q_\alpha) - \operatorname{trace}(IF(x_0, \Sigma, F))\left(c_2 + \frac{1-\alpha}{2}\right)\right\}$$

$$+ \frac{c_\alpha}{1-\alpha}\{c_4 - c_3 + H(q_\alpha)\}\, IF(x_0, \Sigma_{jj}, F) + \frac{c_\alpha c_3}{1-\alpha}\operatorname{trace}(IF(x_0, \Sigma, F)). \tag{7.18}$$

Using the constants b_1 and b_2 defined in Section 2, (7.18) can be written as

$$b_1\, IF(x_0, \Sigma_{jj}, F) - b_2 \operatorname{trace}(IF(x_0, \Sigma, F))$$
$$= -1 + \frac{c_\alpha x_{0j}^2}{1-\alpha} \mathbf{I}(\|x_0\|^2 \leqslant q_\alpha) + \frac{c_\alpha}{1-\alpha} \frac{q_\alpha}{p} \{1-\alpha-\mathbf{I}(\|x_0\|^2 \leqslant q_\alpha)\}. \tag{7.19}$$

Taking the sum of the diagonal terms given in (7.19) yields the trace of the influence function

$$\operatorname{trace}(IF(x_0, \Sigma, F))$$
$$= (b_1 - pb_2)^{-1} \{(-c_\alpha \|x_0\|^2/(1-\alpha))\, \mathbf{I}(\|x_0\|^2 \leqslant q_\alpha)$$
$$+ p\{(c_\alpha/(1-\alpha))(q_\alpha/p)(1-\alpha-\mathbf{I}(\|x_0\|^2 \leqslant q_\alpha)) - 1\}\}. \tag{7.20}$$

Combining (7.19) and (7.20) gives (2.12). ▮

ACKNOWLEGMENTS

We thank the two referees and the editor for helpful suggestions and remarks.

REFERENCES

1. P. J. Bickel and E. L. Lehmann, Descriptive statistics for nonparametric models III, Dispersion, *Ann. Statist.* **4** (1976), 1139–1159.
2. R. W. Butler, Nonparametric interval and point prediction using data trimmed by a Grubbs type outlier rule, *Ann. Statist.* **10** (1982), 197–204.
3. R. W. Butler, P. L. Davies, and M. Jhun, Asymptotics for the minimum covariance determinant estimator, *Ann. Statist.* **21** (1993), 1385–1400.
4. C. Croux and P. J. Rousseeuw, A class of high-breakdown scale estimators based on subranges, *Comm. Statist. Theory Methods* **21** (1992), 1935–1951.
5. C. Croux and G. Haesbroeck, "Maxbias Curves of Robust Scale Estimators Based on Subranges," Technical Report, Université Libre de Bruxelles, 1999, http://www.sig.egss.ulg.ac.be/Haesbroeck/.
6. P. L. Davies, Asymptotic behavior of S-estimators of multivariate location parameters and dispersion matrices, *Ann. Statist.* **15** (1987), 1269–1292.
7. P. L. Davies, The asymptotics of Rousseeuw's minimum volume ellipsoid, *Ann. Statist.* **20** (1992a), 1828–1843.
8. P. L. Davies, An efficient Fréchet Differentiable high breakdown multivariate location and dispersion estimator, *J. Multivariate Anal.* **40** (1992b), 311–327.
9. R. Grübel, A minimal characterization of the covariance matrix, *Metrika* **35** (1988), 49–52.

10. F. R. Hampel, E. M. Ronchetti, P. J. Rousseeuw, and W. A. Stahel, "Robust Statistics: The Approach Based on Influence Functions," Wiley, New York, 1996.
11. D. M. Hawkins, The feasible solution algorithm for the minimum covariance determinant estimator in multivariate data, *Comput. Statist. Data Anal.* **17** (1994), 197–210.
12. X. He and G. Wang, Cross-checking using the minimum volume ellipsoid estimator, *Statistica Sinica* **6** (1996), 367–374.
13. O. Hössjer and C. Croux, Generalizing univariate signed rank statistics for testing and estimating a multivariate location parameter, *Nonparametric Statist.* **4** (1995), 293–308.
14. H. P. Lopuhaä, On the relation between S-estimators and M-estimators of multivariate location and covariance, *Ann. Statist.* **17** (1989), 1662–1683.
15. H. P. Lopuhaä, Multivariate τ-estimators for location and scatter, *Canad. J. Statist.* **19** (1991), 307–321.
16. H. P. Lopuhaä, Highly efficient estimators of multivariate location with high breakdown point, *Ann. Statist.* **19** (1992), 229–248.
17. H. P. Lopuhaä, Asymptotics of reweighted estimators of multivariate location and scatter, preprint, TU Delft, 1997.
18. H. P. Lopuhaä and P. J. Rousseeuw, Breakdown properties of affine equivariant estimators of multivariate location and covariance matrices, *Ann. Statist.* **19** (1991), 229–248.
19. R. A. Maronna and Yohai, Robust estimation of multivariate location and scatter, *in* "Encyclopedia of Statistical Sciences Update" (S. Kotz, C. Read, and D. Banks, Eds.), Vol. 2, pp. 589–596, Wiley, New York, 1998.
20. D. M. Rocke, Robustness properties of S-estimators of multivariate location and shape in high dimensions, *Ann. Statist.* **24** (1996), 1327–1345.
21. P. J. Rousseeuw, Multivariate estimation with high breakdown point, *in* "Mathematical Statistics and Applications" (W. Grossmann, G. Pflug, I. Vincze, and W. Wertz, Eds.), Vol. B, pp. 283–297, Reidel, Dordrecht, 1985.
22. P. J. Rousseeuw and A. M. Leroy, "Robust Regression and Outlier Detection," Wiley, New York, 1987.
23. P. J. Rousseeuw and C. Croux, The bias of k-step M-estimators, *Statist. Probab. Lett.* **20** (1994), 411–420.
24. P. J. Rousseeuw and K. Van Driessen, A fast algorithm for the minimum covariance determinant estimator, *Technometrics*, to appear.
25. D. Ruppert, Computing S-estimators for regression and multivariate location/dispersion, *J. Comput. Graph. Statist.* **1** (1992), 253–270.
26. V. J. Yohai and R. A. Maronna, The maximum bias of robust covariances, *Comm. Statist. A* **19** (1990), 3924–3933.