The Annals of Statistics
1989, Vol. 17, No. 4, 1662–1683

ON THE RELATION BETWEEN S-ESTIMATORS AND M-ESTIMATORS OF MULTIVARIATE LOCATION AND COVARIANCE[1]

BY HENDRIK P. LOPUHAÄ

Delft University of Technology

We discuss the relation between S-estimators and M-estimators of multivariate location and covariance. As in the case of the estimation of a multiple regression parameter, S-estimators are shown to satisfy first-order conditions of M-estimators. We show that the influence function IF $(\mathbf{x}; \mathbf{S}, F)$ of S-functionals exists and is the same as that of corresponding M-functionals. Also, we show that S-estimators have a limiting normal distribution which is similar to the limiting normal distribution of M-estimators. Finally, we compare asymptotic variances and breakdown point of both types of estimators.

1. Introduction and preliminaries. Recently Rousseeuw and Yohai (1984) introduced S-estimators in the framework of multiple regression. These estimators were shown to have the same asymptotic properties as corresponding regression M-estimators, and also to have good robustness properties, as their breakdown point (which can be interpreted as the percentage of outliers in the sample that an estimator can handle) was shown to be 50%.

Davies (1987) investigated some properties of S-estimators of multivariate location and covariance. Using a slightly different definition from the one suggested in Rousseeuw and Yohai (1984), he treated existence, consistency, asymptotic normality and breakdown point. However, the close correspondence with multivariate M-estimators, as was found in the case of estimating a regression parameter, remained hidden.

In this paper multivariate S-estimators are related to multivariate M-estimators. First the definition of multivariate M- and S-estimators is discussed and it is shown that S-estimators of multivariate location and covariance satisfy the first-order conditions of multivariate M-estimators.

This will have the consequence that the asymptotic normality results and the expression for the influence function of multivariate S-estimators are the same as those of corresponding multivariate M-estimators.

Finally, we will compare asymptotic variances in relation with the breakdown point for both types of estimators. It turns out that S-estimators can achieve the variances attained by M-estimators, but they have the additional advantage that in high dimensions (at the same level of asymptotic variance) the breakdown point is considerably higher than that of M-estimators.

Received September 1987; revised October 1988.

[1]Research supported by the 'Nederlandse organisatie voor Wetenschappelijk Onderzoek (NWO)' under Grant 10-62-10.

AMS 1980 *subject classifications*. Primary 62F35, 62H12.

Key words and phrases. S-estimators, M-estimators, influence function, asymptotic normality, efficiency.

All proofs have been saved for an Appendix at the end of the paper.

We will denote p-vectors by $\mathbf{t} = (t_1, \dots, t_p)^T$ and $(p \times p)$-matrices by $\mathbf{M} = (m_{ij})$. For any $p \times p$-matrix $\mathbf{M}$, we write $\mathbf{D_M}$ for the diagonal matrix consisting of the diagonal of $\mathbf{M}$; eigenvalues are denoted by $\lambda_p(\mathbf{M}) \le \cdots \le \lambda_1(\mathbf{M})$. The class of positive definite symmetric matrices is written as PDS(p) and by $\Theta = \mathbb{R}^p \times \mathrm{PDS}(p)$ we denote the set of pairs $\theta = (\mathbf{t}, \mathbf{C})$, which can be seen as an open subset of $\mathbb{R}^{p+(1/2)p(p+1)}$.

By $\mathbf{x}_1, \mathbf{x}_2, \dots$ we will mean vectors in $\mathbb{R}^p$ and we will write $X_1, X_2, \dots$ instead if an underlying distribution is assumed.

The Euclidean norm is denoted by $\|\cdot\|$ and, because of the frequent appearance of quadratic forms $(\mathbf{x} - \mathbf{t})^T\mathbf{C}^{-1}(\mathbf{x} - \mathbf{t})$, we will sometimes abbreviate them by $d^2(\mathbf{x}; \mathbf{t}, \mathbf{C})$. Denote by $E(\mathbf{t}, \mathbf{C}, c)$ an ellipsoid $\{\mathbf{x}\colon (\mathbf{x} - \mathbf{t})^T\mathbf{C}^{-1}(\mathbf{x} - \mathbf{t}) \le c^2\}$. Partial derivatives $\partial g(\mathbf{x}, \theta)/\partial\theta$ will sometimes be abbreviated by $g_\theta(\mathbf{x}, \theta)$.

We will focus on the estimation of the parameter $\theta = (\mu, \Sigma)$ which characterizes an elliptical distribution $F_{\mu,\Sigma}$ with a density of the form

$$(\det(\Sigma))^{-1/2} f\left[(\mathbf{x} - \mu)^T\Sigma^{-1}(\mathbf{x} - \mu)\right], \tag{1.1}$$

where $f\colon [0, \infty) \to [0, \infty)$ is a fixed function and $(\mu, \Sigma) \in \Theta$. Expectations with respect to these distributions are denoted by $\mathbb{E}_{\mu,\Sigma}$. The matrix $\mathbf{B}$ represents the unique positive definite symmetric matrix such that $\Sigma = \mathbf{B}\mathbf{B}^T$. Note that it is often easier to write $\mathbb{E}_{0,\mathbf{I}} h(\|X_0\|)$ instead of $\mathbb{E}_{\mu,\Sigma} h[(X_1 - \mu)^T\Sigma^{-1}(X_1 - \mu)]$, where $X_0 = \mathbf{B}^{-1}(X_1 - \mu)$ and h is any real-valued function.

2. Definitions.

2.1. *M-estimators.* M-estimators were originally constructed by Huber (1964) for the estimation of a one-dimensional location parameter. Later Huber (1967) considered a very general framework in which consistency and asymptotic normality were proved under relatively weak conditions.

Maronna (1976) was the first to define M-estimators for multivariate location and covariance. Huber (1981) extended Maronna's definition by defining M-estimates based upon $\mathbf{x}_1, \dots, \mathbf{x}_n \in \mathbb{R}^p$ as solutions of the simultaneous equations:

$$\begin{aligned} &(1/n)\sum_{i=1}^{n} v_1(d_i)(\mathbf{x}_i - \mathbf{t}) = \mathbf{0}, \\ &(1/n)\sum_{i=1}^{n}\left\{v_2(d_i^2)(\mathbf{x}_i - \mathbf{t})(\mathbf{x}_i - \mathbf{t})^T - v_3(d_i)\mathbf{C}\right\} = \mathbf{0}, \end{aligned} \tag{2.1}$$

where $d_i = d(\mathbf{x}_i; \mathbf{t}, \mathbf{C})$ and v_1, v_2 and v_3 are real-valued functions on $[0, \infty)$.

EXAMPLE 2.1 (Huber's Proposal 2). Take $v_3(y) = 1$ and $v_i(y) = \psi_i(y)/y$, for $i = 1, 2$, where $\psi_1(y) = \psi_H(y, k)$ and $\psi_2(y) = \psi_H(y, k^2)$. The function $\psi_H(y, k) = \min\{y, \max\{y, -k\}\}$ is known as Huber's psi-function.

Both existence and uniqueness of solutions of (2.1) were only shown for $v_3 = 1$ [Maronna (1976); Huber (1981)]. For this case Maronna (1976) shows consistency and asymptotic normality by means of Huber's (1967) results.

Maronna (1976) and Huber (1981) consider the breakdown point ε^* and the influence function IF to measure the robustness of these estimators. They both indicate that for solutions of (2.1) the breakdown point is at most $1/(p+1)$. A detailed treatment on the (finite-sample) breakdown behaviour of this type of M-estimators is given in Tyler (1986). One should note that these results assume monotonicity of v_2 and v_3 to be constant. So from the viewpoint of breakdown, M-estimators become more sensitive to outliers in high dimensions. From the viewpoint of the influence function (which describes the effect of *one* outlier on the estimate), M-estimators are robust, as their influence function remains bounded when v_1, v_2 and v_3 in (2.1) are chosen suitably [see Huber (1981)].

2.2. *S-estimators.* Rousseeuw and Yohai (1984) introduced S-estimators in a regression context and defined them as the solution to the problem of minimizing σ subject to

$$\frac{1}{n}\sum_{i=1}^{n}\rho\left(\frac{y_i-\theta^T\mathbf{x}_i}{\sigma}\right)=b_0 \tag{2.2}$$

among all $(\theta,\sigma)\in\mathbb{R}^p\times(0,\infty)$, where $0<b_0<\sup\rho$. The special case $\rho(y)=y^2$ in (2.2) obviously leads to the least squares estimates. In order to obtain more robust estimates and preserve asymptotic normality the function ρ was assumed to satisfy:

(R1) ρ is symmetric, has a continuous derivative ψ and $\rho(0)=0$.
(R2) There exists a finite constant $c_0>0$ such that ρ is strictly increasing on $[0,c_0]$ and constant on $[c_0,\infty)$. (Put $a_0=\sup\rho$.)

A direct generalization to S-estimators of multivariate location and covariance is obtained simply by adjustment of (2.2).

Definition 2.1. Let $\mathbf{x}_1,\mathbf{x}_2,\dots,\mathbf{x}_n\in\mathbb{R}^p$ and let $\rho\colon\mathbb{R}\to[0,\infty)$ satisfy (R1) and (R2). Then the S-estimate of multivariate location and covariance is defined as the solution $\theta_n=(\mathbf{t}_n,\mathbf{C}_n)$ to the problem of minimizing $\det(\mathbf{C})$ subject to

$$\frac{1}{n}\sum_{i=1}^{n}\rho\left[\left\{(\mathbf{x}_i-\mathbf{t})^T\mathbf{C}^{-1}(\mathbf{x}_i-\mathbf{t})\right\}^{1/2}\right]=b_0 \tag{2.3}$$

among all $(\mathbf{t},\mathbf{C})\in\Theta$. Denote this minimization problem by $(\mathscr{P}_n)$.

The constant $0<b_0<a_0$ can be chosen in agreement with an assumed underlying distribution. For instance, if $\mathbf{x}_1,\mathbf{x}_2,\dots,\mathbf{x}_n$ are assumed to be a sample $X_1,X_2,\dots,X_n$ with an underlying elliptical distribution (1.1), then the constant b_0 is generally chosen to be $\mathbb{E}_{0,I}\rho(\|X_0\|)$. In that case the constant c_0

can be chosen such that $0 < b_0/a_0 = r \le (n-p)/2n$, which leads to a (finite-sample) breakdown point $\varepsilon_n^* = \lceil nr \rceil/n$ [see Lopuhaä and Rousseeuw (1987)]. For $r = (n-p)/2n$ one obtains the maximal breakdown point $\lfloor (n-p+1)/2 \rfloor/n$, or asymptotically 0.50. However, the constant c_0 at the same time determines the asymptotic variance and, as we will see in Section 6, it is not possible to achieve small asymptotic variance and 50% breakdown point simultaneously.

It might be worthwhile to mention that S-estimators of location and covariance can also be seen as robustifications of the least squares method. If $b_0 = p$, then using $\rho(y) = y^2$ in (2.3) yields the sample mean and sample covariance as unique solutions of $(\mathscr{P}_n)$ [see for instance Grübel (1988)].

EXAMPLE 2.2 (Tukey's biweight). An example of a rho-function for (2.3) is

$$\rho_B(y, c_0) = \begin{cases} \dfrac{y_2}{2} - \dfrac{y^4}{2c_0^2} + \dfrac{y^6}{6c_0^4}, & \text{for } |y| \le c_0, \\ \dfrac{c_0^2}{6}, & \text{for } |y| \ge c_0. \end{cases}$$

Its derivative, which is a redescending (psi-) function, is known as Tukey's biweight function $\psi_B(y, c_0) = y(1 - (y/c_0)^2)^2 \mathbf{1}_{[-c_0, c_0]}(y)$.

Davies (1987) defines S-estimates similarly only instead of ρ he uses a nonincreasing function $\kappa\colon \mathbb{R}_+ \to [0,1]$ in (2.3). It is related to ρ as $\kappa(y) = 1 - \rho(y^2)/a_0$. If "continuous differentiability of ρ" is weakened to "ρ being left-continuous on $(0, \infty)$ and continuous at 0," if "strict increasing" on $[0, c_0]$ is weakened to "nondecreasing," and if " = " is replaced by " $\le$ " in (2.3), then the two definitions are equivalent. Under these weaker conditions Davies proves existence and consistency of S-estimates, and he obtains asymptotic normality assuming that the function κ has a third continuous derivative.

We will extend existence and consistency of S-estimates to existence and continuity of S-functionals and obtain the influence function, and we will extend the asymptotic normality result by considering S-estimators as a special type of M-estimators and use Huber's (1967) results.

2.3. *Relationship between M- and S-estimators.* A drawback of using κ instead of ρ is that the conjectured correspondence with M-estimators remains hidden. In this section we will show that a solution to the minimization problem $(\mathscr{P}_n)$ also satisfies the first-order M-estimator conditions (2.1).

Let $\boldsymbol{\theta}_n = (\mathbf{t}_n, \mathbf{C}_n)$ be a solution of $(\mathscr{P}_n)$. Then, if by λ_n we denote the corresponding Lagrange multiplier, the pair $(\boldsymbol{\theta}_n, \lambda_n)$ is a zero of all partial derivatives $\partial \mathbf{L}_n/\partial \mathbf{t}$, $\partial \mathbf{L}_n/\partial \mathbf{C}$ and $\partial \mathbf{L}_n/\partial \lambda$, where $\mathbf{L}_n$ is the Lagrangian

$$\mathbf{L}_n(\boldsymbol{\theta}, \lambda) = \log(\det(\mathbf{C})) - \lambda\left\{\frac{1}{n}\sum_{i=1}^{n} \rho\left[\left\{(\mathbf{x}_i - \mathbf{t})^T \mathbf{C}^{-1}(\mathbf{x}_i - \mathbf{t})\right\}^{1/2}\right] - b_0\right\}.$$

This means that besides constraint (2.3), (θ_n, λ_n) satisfies the equations

$$(2.4)\qquad \begin{aligned} &(\lambda/n)\sum_{i=1}^{n} u(d_i)\mathbf{C}^{-1}(\mathbf{x}_i - \mathbf{t}) = \mathbf{0},\\ &2\mathbf{C}^{-1} - \mathbf{D}_{\mathbf{C}^{-1}} + (\lambda/2n)\sum_{i=1}^{n} u(d_i)\big(2\mathbf{V}_i - \mathbf{D}_{\mathbf{V}_i}\big) = \mathbf{0},\end{aligned}$$

where $u(y) = \psi(y)/y$, $d_i = d(\mathbf{x}_i; \mathbf{t}, \mathbf{C})$ and $\mathbf{V}_i = \mathbf{C}^{-1}(\mathbf{x}_i - \mathbf{t})(\mathbf{x}_i - \mathbf{t})^T\mathbf{C}^{-T}$ [for derivatives with respect to symmetric matrices, see Graybill (1983)]. But the second (matrix) equation can be written as

$$(2.5)\qquad \mathbf{I} + \frac{\lambda}{2n}\sum_{i=1}^{n} u(d_i)\mathbf{A}^{-1}(\mathbf{x}_i - \mathbf{t})(\mathbf{x}_i - \mathbf{t})^T\mathbf{A}^{-T} = \mathbf{0},$$

where $\mathbf{A}\mathbf{A}^T = \mathbf{C}$. When we take the trace we get $p + (\lambda/2n)\sum_{i=1}^{n}\psi(d_i)d_i = 0$. Obviously we can solve λ_n from this equation, yielding

$$\lambda_n = -2p\left(\frac{1}{n}\sum_{i=1}^{n}\psi(d_{i,n})d_{i,n}\right)^{-1},$$

where $d_{i,n} = d(\mathbf{x}_i; \mathbf{t}_n, \mathbf{C}_n)$. If we substitute this into (2.5), together with (2.4) and (2.3) we find that any solution θ_n of $(\mathscr{P}_n)$ satisfies the equations

$$(2.6)\qquad \begin{aligned} &(1/n)\sum_{i=1}^{n} u(d_i)(\mathbf{x}_i - \mathbf{t}) = \mathbf{0},\\ &(1/n)\sum_{i=1}^{n}\left\{pu(d_i)(\mathbf{x}_i - \mathbf{t})(\mathbf{x}_i - \mathbf{t})^T - v(d_i)\mathbf{C}\right\} = \mathbf{0},\end{aligned}$$

where $v(y) = \psi(y)y - \rho(y) + b_0$. The term $-\rho(y) + b_0$ is added to $\psi(y)y$ because merely substituting λ_n into (2.5) would give us a system of linear dependent equations for any pair $(\mathbf{t}, \mathbf{C}) \in \Theta$.

So any solution of $(\mathscr{P}_n)$ turns out to be also a solution of (2.6) which obviously are of M-estimator type (2.1). To match the notation used in Huber (1967) we write (2.6) as

$$(2.7)\qquad \frac{1}{n}\sum_{i=1}^{n}\Psi(\mathbf{x}_i, \theta) = \mathbf{0},$$

where $\theta = (\mathbf{t}, \mathbf{C}) \in \Theta$ and $\Psi = (\Psi_1, \Psi_2)$ is the function

$$(2.8)\qquad \begin{aligned} \Psi_1(\mathbf{x}, \theta) &= u(d)(\mathbf{x} - \mathbf{t}),\\ \Psi_2(\mathbf{x}, \theta) &= pu(d)(\mathbf{x} - \mathbf{t})(\mathbf{x} - \mathbf{t})^T - v(d)\mathbf{C},\end{aligned}$$

with $d = d(\mathbf{x}; \mathbf{t}, \mathbf{C})$. We conclude that S-estimators satisfy first-order conditions (2.1) of M-estimators as defined in Huber (1981), or rather (2.7) of the type considered in Huber (1967).

However, recall that S-estimators are originally defined by the minimization problem $(\mathscr{P}_n)$, which is not equivalent to (2.7), and that in any dimension they can still be constructed with high breakdown point. The cause of these differences might lie in the functions $v_2(\cdot)$ and $v_3(\cdot)$ of (2.1) and the functions $u(\cdot)$ and $v(\cdot)$ of (2.6). For instance, Huber (1981) chooses $v_2 > 0$ and $v_3 \geq 0$ to be monotone, and the latter even equal to a constant for proving both existence and uniqueness of solutions of (2.1). The functions $u(y) = \psi(y)/y$ and $v(y) = \psi(y)y - \rho(y) + b_0$ will never satisfy either condition.

One could call any solution of (2.7) an M-estimate. However, M-estimates are generally associated with low breakdown point and with implicit equations (2.1), with v_2 decreasing and v_3 being constant. As this is not the case for S-estimates we tend to consider these estimates to be of a different type.

Although the S-estimate is probably not the only solution of (2.7), it is a solution with high breakdown point. To find it, one must therefore solve $(\mathscr{P}_n)$ and not just (2.7). Nevertheless S-estimators do satisfy (2.7) which has the consequence that their asymptotic behaviour and their influence function are the same as for M-estimators.

3. S-functionals and influence function. For the derivation of the influence function we have to extend Definition 2.1 to a functional formulation. We will identify every distribution function F with its corresponding probability measure P_F on $\mathbb{R}^p$, and for brevity we call this distribution F. Denote by $\mathscr{F}$ the class of all distributions on $\mathbb{R}^p$. The functional analogue of the S-estimator of multivariate location and covariance is defined as follows.

DEFINITION 3.1. Let $\rho: \mathbb{R} \to [0, \infty)$ be a function satisfying (R1) and (R2). Then the S-functional $\mathbf{S}: \mathscr{F} \to \Theta$ is defined as the solution $\mathbf{S}(F) = (\mathbf{t}(F), \mathbf{C}(F))$ to the problem of minimizing $\det(\mathbf{C})$ subject to

$$\int \rho\left[\left\{(\mathbf{y} - \mathbf{t})^T \mathbf{C}^{-1}(\mathbf{y} - \mathbf{t})\right\}^{1/2}\right] dF(\mathbf{y}) = b_0 \tag{3.1}$$

among all $(\mathbf{t}, \mathbf{C}) \in \Theta$, where $0 < b_0 < a_0$. Denote this problem by $(\mathscr{P}_F)$.

Existence of solutions of $(\mathscr{P}_F)$ is ensured if there is not too much mass concentrated at infinity or at arbitrarily thin strips $H(\boldsymbol{\alpha}, l, \delta) = \{\mathbf{y}\colon l \leq \boldsymbol{\alpha}^T\mathbf{y} \leq l + \delta\}$, where $\|\boldsymbol{\alpha}\| = 1$, $\delta \geq 0$ and $l \in \mathbb{R}$. Let $0 < \varepsilon < 1$ and consider the following property for the measure P_F on $\mathbb{R}^p$ corresponding with distribution F:

(C_ε) There exists a compact set $B_\varepsilon \subset \mathbb{R}^p$ such that $P_F(B_\varepsilon) \geq 1 - \varepsilon$, and the value $\delta_\varepsilon = \inf\{\delta\colon P_F(H(\boldsymbol{\alpha}, l, \delta)) \geq \varepsilon, \|\boldsymbol{\alpha}\| = 1, \delta \geq 0, l \in \mathbb{R}\}$ is strictly positive.

THEOREM 3.1. *Let F satisfy property* (C_ε) *for some* $0 < \varepsilon \leq 1 - r$, *where* $r = b_0/a_0$. *Then* $(\mathscr{P}_F)$ *has at least one solution.*

Let G_k, $k \geq 0$, be a sequence of probability distributions on $\mathbb{R}^p$ that converges weakly to F as $k \to \infty$. The following theorem shows when S-functionals are continuous.

THEOREM 3.2. *Let $\mathscr{C}$ be the class of all measurable convex subsets of* $\mathbb{R}^p$ *and suppose that every* $E \in \mathscr{C}$ *is a* P_F*-continuity set, i.e.,* $P_F(\partial E) = 0$. *Suppose that F satisfies* (C_ε) *for some* $0 < \varepsilon < 1 - r$, *and assume that the solution* $\mathbf{S}(F)$ *of* $(\mathscr{P}_F)$ *is unique. Then for k sufficiently large,* $(\mathscr{P}_{G_k})$ *has at least one solution* $\mathbf{S}(G_k)$ *and for any sequence of solutions* $\mathbf{S}(G_k)$, $k \geq 0$, $\lim_{k\to\infty} \mathbf{S}(G_k) = \mathbf{S}(F)$ *holds.*

REMARK 3.1. In the proof of Theorem 3.1 strict monotonicity of ρ on $[0, c_o]$ is not needed and continuity of ρ is not essential. This means that Theorem 3.1 can easily be shown to hold also for S-functionals that correspond to the larger class of S-estimates considered by Davies (1987) (see Section 2.2). With a stronger condition on F, which will ensure $\int \rho(\|\mathbf{y}\|/(1+\eta))\, dF(\mathbf{y})$ to be strictly decreasing at $\eta = 0$, also Theorem 3.2 can be shown to hold for these S-functionals.

REMARK 3.2. A part of the proof of Theorem 3.2 consists of showing that solutions $\mathbf{S}(G_k)$ eventually stay inside a fixed compact set. For the special case $G_h = (1-h)F + h\Delta_{\mathbf{x}}$ (see Definition 3.2) one can show that if $0 < r < \frac{1}{2}$ and if F only satisfies (C_ε) for $\varepsilon = (1-2r)(1-r)^{-1}$, then for any $0 < \alpha < 1$ there exists a compact set $K(\alpha)$ independent of $\mathbf{x}$ such that for all $h \in [0, \alpha r]$ the problem $(\mathscr{P}_{G_k})$ has at least one solution and all solutions are contained in $K(\alpha)$. Condition $0 < r < \frac{1}{2}$ is similar to the condition $0 < r < (n-p)/2n$ which ensures (finite-sample) breakdown point $\varepsilon_n^* = \lceil nr \rceil / n$ (see Section 2.2). The latter means that the S-estimator stays in some fixed compact subset of Θ when the amount of contamination is less than ε_n^*. This is in agreement with the statement above that when the amount of contamination at $\mathbf{x} \in \mathbb{R}^p$ is less than r, the S-functional stays within a compact subset of Θ.

The robustness of the S-estimator can be measured by means of the influence function [see Hampel (1974)]. It is defined in terms of the S-functional in the following manner.

DEFINITION 3.2. Let $\mathbf{S}(\cdot)$ be a vector-valued mapping from a subset of $\mathscr{F}$ into Θ and let F lie in the domain of $\mathbf{S}(\cdot)$. If $\Delta_{\mathbf{x}}$ denotes the atomic probability distribution concentrated in $\mathbf{x} \in \mathbb{R}^p$, then the influence function of $\mathbf{S}(\cdot)$ at F is defined pointwise by

$$\mathrm{IF}(\mathbf{x}; \mathbf{S}, F) = \lim_{h \downarrow 0} \frac{\mathbf{S}((1-h)F + h\Delta_{\mathbf{x}}) - \mathbf{S}(F)}{h}, \tag{3.2}$$

if this limit exists for every $\mathbf{x} \in \mathbb{R}^p$.

If we replace F by the empirical distribution F_{n-1} and h by $1/n$, we realize that the IF measures a weighted alteration of the value of the estimator when one additional observation is added to a large sample of size $n-1$. The importance of the influence function lies in its heuristic interpretation: It describes the effect of an infinitesimal contamination at point $\mathbf{x}$ on the estimate. A bounded influence function is therefore considered to be a good robustness property.

To derive the IF at a distribution F we need to be sure that $\mathbf{S}(\cdot)$ is uniquely defined at $(1-h)F + h\Delta_{\mathbf{x}}$, for all $\mathbf{x} \in \mathbb{R}^p$ at least for small h, and second, that the limit (3.2) exists for all $\mathbf{x} \in \mathbb{R}^p$. Theorem 3.2 ensures that for all $\mathbf{x} \in \mathbb{R}^p$ and h sufficiently small the problem $(\mathscr{P}_{G_h})$, with $G_h = (1-h)F + h\Delta_{\mathbf{x}}$, has at least one solution and that all solutions are continuous.

We conclude that for h sufficiently small there exist solutions $\boldsymbol{\theta}_h = (\mathbf{t}_h, \mathbf{C}_h)$ of $(\mathscr{P}_{G_h})$ and that they all converge to the same limit $(\mathbf{t}(F), \mathbf{C}(F))$ as h tends to 0. Therefore there exists an open neighbourhood N of $\mathbf{S}(F)$ which contains all solutions $\boldsymbol{\theta}_h$ for h sufficiently small.

Remember that (2.7) is obtained from differentiation of the Lagrangian corresponding with problem $(\mathscr{P}_n)$. Similarly one could now differentiate the Lagrangian corresponding with the problem $(\mathscr{P}_F)$. If we restrict to the neighbourhood N, we may interchange the order of differentiation and integration, and similar to (2.7) we obtain the equation

$$\int \boldsymbol{\Psi}(\mathbf{y}, \boldsymbol{\theta})\, dF(\mathbf{y}) = \mathbf{0}, \tag{3.3}$$

where $\boldsymbol{\Psi}(\mathbf{y}, \boldsymbol{\theta})$ is defined in (2.8).

Solutions $(\mathbf{t}_h, \mathbf{C}_h)$ of $(\mathscr{P}_{G_h})$ must be a solution (not necessarily the only one) of (3.3), at least for h sufficiently small. Note that if we had considered a functional $\mathbf{M}$: $\mathscr{F} \to \Theta$, defined as the solution of (3.3), we could have some problems to ensure the uniqueness, and therefore for obtaining the influence function $\mathrm{IF}(\mathbf{x}; \mathbf{S}, F)$ we explicitly consider the solution $\mathbf{S}(F)$ of (3.3). The implicit function theorem, applied to this equation will ensure the uniqueness of $\mathbf{S}(\cdot)$ at G_h for h sufficiently small, and also the existence of $\mathrm{IF}(\mathbf{x}; \mathbf{S}, F)$.

THEOREM 3.3. *Let* ρ: $\mathbb{R} \to [0, \infty)$ *satisfy* (R1) *and* (R2). *Assume that* ρ *has a second derivative* ψ' *and suppose that*

(R3) $\psi'(y)$ *and* $u(y) = \psi(y)/y$ *are bounded and continuous.*

Suppose that the conditions of Theorem 3.2 *hold. Let* $\boldsymbol{\Psi}$ *be defined as in* (2.8) *and let* $\boldsymbol{\lambda}_F(\boldsymbol{\theta}) = \mathbb{E}_F \boldsymbol{\Psi}(X, \boldsymbol{\theta})$. *Suppose that* $\boldsymbol{\lambda}_F(\cdot)$ *has a nonsingular derivative* Λ *at* $\mathbf{S}(F) = (\mathbf{t}(F), \mathbf{C}(F))$. *Then the influence function* $\mathrm{IF}(\mathbf{x}; \mathbf{S}, F)$ *exists and satisfies*

$$\mathrm{IF}(\mathbf{x}; \mathbf{S}, F) = -\Lambda^{-1}\boldsymbol{\Psi}(\mathbf{x}, \mathbf{S}(F)). \tag{3.4}$$

Huber (1981) showed that (3.3) has a unique solution when certain monotonicity conditions on the functions $u(\cdot)$ and $v(\cdot)$ are satisfied. One of these conditions is that the function $v(\cdot)$ is constant. In our case $v(\cdot)$ is certainly not a constant function and so (3.3) may have many solutions. However, it is possible that there is a unique "S-solution" among all solutions of (3.3). For this solution we have derived the IF and naturally the expression is of the same type as for multivariate M-estimators. Note that the properties of the function ρ imply that the influence function $\mathrm{IF}(\mathbf{x}; \mathbf{S}, F)$ is bounded.

4. Asymptotic normality of S-estimators. Let $X_1, X_2, \dots$ be a sequence of independent random vectors $X_i = (X_{i1}, \dots, X_{ip})^T$ with a distribution F on $\mathbb{R}^p$. Suppose that for $n \geq p + 1$ the sample $X_1, \dots, X_n$ is in *general position*, i.e., no $p + 1$ points lie in some $(p - 1)$-dimensional subspace, almost surely.

When F in Definition 3.1 equals the empirical distribution F_n, we get the definition of the S-estimator. Note that F_n satisfies (C_ε) for $\varepsilon = (p + 1)/n$ almost surely, so as a special case of Theorem 3.1 we have that for $n(1 - r) \geq p + 1$ the problem ($\mathscr{P}_n$) has at least one solution almost surely. If F satisfies the conditions of Theorem 3.2 one has consistency: $\boldsymbol{\theta}_n \to (\mathbf{t}(F), \mathbf{C}(F))$ almost surely.

As we have seen in Section 2.3 solutions $\boldsymbol{\theta}_n$ of ($\mathscr{P}_n$) satisfy first-order conditions (2.7) of M-estimators. An immediate consequence is that the asymptotic behaviour of S-estimators is similar to that of M-estimators.

THEOREM 4.1. *Let ρ: $\mathbb{R} \to [0, \infty)$ satisfy* (R1)–(R3) *and suppose that the conditions of Theorem* 3.2 *hold. Let Ψ be defined as in* (2.8) *and let $\lambda_F(\cdot)$ be defined as in Theorem* 3.3. *Suppose that the solution* $\mathbf{S}(F)$ *of* ($\mathscr{P}_F$) *is unique and that $\lambda_F(\cdot)$ has a nonsingular derivative Λ at $\theta_0 = \mathbf{S}(F)$. Let $\boldsymbol{\theta}_n$ be a solution of ($\mathscr{P}_n$). Then $n^{1/2}(\boldsymbol{\theta}_n - \boldsymbol{\theta}_0)$ has a limiting normal distribution with zero mean and covariance matrix $\Lambda^{-1}\mathbf{M}\Lambda^{-T}$, where* $\mathbf{M}$ *stands for the covariance matrix of $\Psi(X_1, \mathbf{S}(F))$.*

REMARK 4.1. One might try to prove asymptotic normality of S-estimators directly from Definition 2.1 and avoid (2.7). A first derivative ψ of ρ in (R1), needed to arrive at (2.7) is then no longer required. At least continuity of ρ seems necessary. This is indicated by the results of Kim and Pollard (1988) on Rousseeuw's (1986) minimum volume ellipsoid estimator, which can be seen as an S-estimator with a discontinuous ρ-function.

5. Elliptical distributions. Consider the case that $F = F_{\mu, \Sigma}$ is elliptical and therefore take $b_0 = \mathbb{E}_{0, I}\rho(\|X_0\|)$ in (2.3) and (3.1). For this choice of F, Huber (1981) obtains the expression for the IF of M-functionals and Maronna (1976) gives a detailed description for the asymptotic covariance matrix of location M-estimators. We compare these results with Theorems 3.3 and 4.1 applied to $F_{\mu, \Sigma}$.

It is not difficult to show that $F_{\mu, \Sigma}$ satisfies property (C_ε) for any $0 < \varepsilon < 1$, so according to Theorem 3.1 at least one solution of ($\mathscr{P}_{F_{\mu,\Sigma}}$) exists. Davies (1987) showed that it is even unique and *Fisher consistent*,

$$\mathbf{S}(F_{\mu, \Sigma}) = (\mu, \Sigma). \tag{5.1}$$

The following corollary gives a detailed description of the limiting normal distribution of $n^{1/2}(\boldsymbol{\theta}_n - \boldsymbol{\theta}_0)$. In particular, the asymptotic covariance of the location S-estimate $\mathbf{t}_n$ as defined in Definition 2.1 is exactly the same as the asymptotic covariance found for the location M-estimate considered by Maronna (1976) if one chooses $v_1(y) = \psi(y)/y$ in (2.1).

To describe the asymptotic covariance matrix of $n^{1/2}(\mathbf{C}_n - \Sigma)$ in condensed form, we represent $(p \times p)$-matrices $\mathbf{M}$ by $\mathrm{vec}(\mathbf{M}) = (m_{11}, \ldots, m_{p1}, \ldots, m_{1p}, \ldots, m_{pp})^T$. The operator $\mathrm{vec}(\cdot)$ just stacks the columns of $\mathbf{M}$ on top of each other. Magnus and Neudecker (1979) investigated algebraic properties of this operator in relation with the commutation matrix $\mathbf{K}_{m,n}$. Here we will only use the special case $\mathbf{K}_{p,p}$, which is a $(p^2 \times p^2)$-block matrix with (i, j)-block being equal to Δ_{ji}. The latter is a $(p \times p)$-matrix which is 1 at entry (j, i) and 0 everywhere else. Finally, $\mathbf{M} \otimes \mathbf{N}$ denotes the Kronecker product of the matrices $\mathbf{M}$ and $\mathbf{N}$ which is a $(p^2 \times p^2)$-block matrix with $(p \times p)$-blocks, the (i, j)-block equal to $m_{ij}\mathbf{N}$.

COROLLARY 5.1. *Let ρ: $\mathbb{R} \to [0, \infty)$ satisfy* (R1)–(R3) *and let F be the elliptical distribution with parameter $\theta_0 = (\mu, \Sigma)$. Suppose that*

$$\mathbb{E}_{\mathbf{0},\mathbf{I}}\psi'(\|X_0\|) > 0,$$
$$\mathbb{E}_{\mathbf{0},\mathbf{I}}\left[\psi'(\|X_0\|)\|X_0\|^2 + (p+1)\psi(\|X_0\|)\|X_0\|\right] > 0. \tag{5.2}$$

When $\theta_n = (\mathbf{t}_n, \mathbf{C}_n)$ is a solution of $(\mathscr{P}_n)$, then $n^{1/2}(\theta_n - \theta_0)$ has a limiting normal distribution with zero mean and $\mathbf{t}_n$ and $\mathbf{C}_n$ are asymptotically independent. The covariance matrix of the limiting distribution of $n^{1/2}(\mathbf{t}_n - \mu)$ is given by $(\alpha/\beta^2)\Sigma$, where

$$\alpha = \frac{1}{p}\mathbb{E}_{\mathbf{0},\mathbf{I}}\psi^2(\|X_0\|),$$
$$\beta = \mathbb{E}_{\mathbf{0},\mathbf{I}}\left[\left(1 - \frac{1}{p}\right)u(\|X_0\|) + \frac{1}{p}\psi'(\|X_0\|)\right]. \tag{5.3}$$

The covariance matrix of the limiting distribution of $n^{1/2}(\mathbf{C}_n - \Sigma)$ is given by

$$\sigma_1(\mathbf{I} + \mathbf{K}_{p,p})(\Sigma \otimes \Sigma) + \sigma_2 \,\mathrm{vec}(\Sigma)\mathrm{vec}(\Sigma)^T, \tag{5.4}$$

where

$$\sigma_1 = \frac{p(p+2)\mathbb{E}_{\mathbf{0},\mathbf{I}}\psi^2(\|X_0\|)\|X_0\|^2}{\left\{\mathbb{E}_{\mathbf{0},\mathbf{I}}\left[\psi'(\|X_0\|)\|X_0\|^2 + (p+1)\psi(\|X_0\|)\|X_0\|\right]\right\}^2},$$
$$\sigma_2 = -\frac{2}{p}\sigma_1 + \frac{4\mathbb{E}_{\mathbf{0},\mathbf{I}}\left(\rho(\|X_0\|) - b_0\right)^2}{\left\{\mathbb{E}_{\mathbf{0},\mathbf{I}}\psi(\|X_0\|)\|X_0\|\right\}^2}. \tag{5.5}$$

For the influence function, it is sufficient to give the expression of $\mathrm{IF}(\mathbf{x}; \mathbf{S}, F_{\mathbf{0},\mathbf{I}})$ because affine equivariance of $\mathbf{S}(\cdot)$ yields the general expressions

$$\mathrm{IF}(\mathbf{x}; \mathbf{t}, F_{\mu,\Sigma}) = \mathbf{B}\,\mathrm{IF}(\mathbf{B}^{-1}(\mathbf{x} - \mu); \mathbf{t}, F_{\mathbf{0},\mathbf{I}}),$$
$$\mathrm{IF}(\mathbf{x}; \mathbf{C}, F_{\mu,\Sigma}) = \mathbf{B}\,\mathrm{IF}(\mathbf{B}^{-1}(\mathbf{x} - \mu); \mathbf{C}, F_{\mathbf{0},\mathbf{I}})\mathbf{B}^T. \tag{5.6}$$

We will describe $\mathrm{IF}(\mathbf{x}; \mathbf{S}, F_{0,\mathbf{I}})$ such that it can be compared with the expressions found for M-functionals in Huber (1981).

COROLLARY 5.2. *Let* ρ: $\mathbb{R} \to [0, \infty)$ *satisfy* (R1)–(R3) *and suppose that conditions* (5.2) *hold. Then the influence function* $\mathrm{IF}(\mathbf{x}; \mathbf{S}, F_{0,\mathbf{I}})$ *of the S-functional defined in Definition* 3.2 *exists and it holds that*

$$\mathrm{IF}(\mathbf{x}; \mathbf{t}, F_{0,\mathbf{I}}) = \frac{1}{\beta}\psi(\|\mathbf{x}\|)\frac{\mathbf{x}}{\|\mathbf{x}\|}, \tag{5.7}$$

where β *is defined in* (5.3) *and* $\mathrm{IF}(\mathbf{x}; \mathbf{C}, F_{0,\mathbf{I}})$ *satisfies*

$$\begin{aligned} &\mathrm{IF}(\mathbf{x}; \mathbf{C}, F_{0,\mathbf{I}}) - \frac{1}{p}\operatorname{trace}[\mathrm{IF}(\mathbf{x}; \mathbf{C}, F_{0,\mathbf{I}})]\mathbf{I} \\ &\quad = \frac{1}{\gamma_1} p \psi(\|\mathbf{x}\|)\|\mathbf{x}\|\left(\frac{\mathbf{x}\mathbf{x}^T}{\|\mathbf{x}\|^2} - \frac{1}{p}\mathbf{I}\right) \end{aligned} \tag{5.8}$$

and

$$\frac{1}{p}\operatorname{trace}[\mathrm{IF}(\mathbf{x}; \mathbf{C}, F_{0,\mathbf{I}})] = \frac{2}{\gamma_3}(\rho(\|\mathbf{x}\|) - b_0), \tag{5.9}$$

where γ_1 *is defined in* (A.11) *and* $\gamma_3 = \mathbb{E}_{0,\mathbf{I}}\psi(\|X_0\|)\|X_0\|$.

6. Asymptotic variance in relation to breakdown point. We compute asymptotic variances of the S-estimator defined by Tukey's biweight function $\rho_{\mathbf{B}}(\cdot, c_0)$ of Example 2.2. The variances are computed for different values of p ($= 1, 2$ and 10) and for each p the constant c_0 is given five different values that correspond with the values for $r = 0.1, 0.2$, 0.3, 0.4 and 0.5, by means of the relation

$$\frac{\mathbb{E}_{\Phi}\rho_{\mathbf{B}}(\|X\|, c_0)}{(c_0^2/6)} = r, \tag{6.1}$$

where the expectation is with respect to the standard normal distribution. The values of r are the limiting values of the finite sample breakdown point $\varepsilon_n^* = \lceil nr \rceil / n$. Denote the corresponding S-estimator by $S(r, p)$.

We compare these results with the asymptotic variances of the M-estimator defined by Huber's Proposal 2 of Example 2.1. The different choices of k correspond to "windsorizing proportions" $w = P_{\Phi}\{\|X\| > k\}$ ($= 0.3$, 0.2, 0.1 and 0). Denote the corresponding M-estimator by $H(w, p)$. Note that in all cases $\sup \psi_2 = k^2 > p$, which is needed for the existence of $H(w, p)$. By $H(0, p)$ we mean the limiting case of $H(w, p)$ as $k \to \infty$. Note that $H(0, p)$ is also the limiting case of $S(r, p)$ as $c_0 \to \infty$, namely sample mean and sample covariance.

Maronna (1976) already computed asymptotic variances for the $H(w, p)$-location estimator at the multivariate student and the multivariate normal distribution, and Tyler (1983) computed an index for the asymptotic variance of the $H(w, p)$-covariance estimator also at these distributions as well as at a symmetric contaminated normal distribution with thicker tails.

TABLE 1
Asymptotic variances of $S(r, p)$ and $H(w, p)$ attained at NOR *and* SCN

		$p = 1$		$p = 2$		$p = 10$	
		NOR	SCN	NOR	SCN	NOR	SCN
$S(0.5)$	λ	3.485	4.007	1.725	1.952	1.072	1.191
	η	3.711	4.301	2.656	3.020	1.093	1.215
$S(0.4)$	λ	2.165	2.499	1.356	1.542	1.036	1.152
	η	2.949	3.554	1.736	1.991	1.045	1.163
$S(0.3)$	λ	1.512	1.757	1.157	1.327	1.016	1.133
	η	2.467	3.174	1.298	1.516	1.020	1.140
$S(0.2)$	λ	1.181	1.392	1.055	1.232	1.006	1.139
	η	2.176	3.173	1.096	1.334	1.007	1.174
$S(0.1)$	λ	1.035	1.271	1.011	1.252	1.001	1.250
	η	2.035	3.919	1.018	1.430	1.001	1.527
$H(0.3)$	λ	1.100	1.327	1.048	1.260	1.009	1.185
	η	3.974	4.231	1.256	1.302	1.047	1.066
$H(0.2)$	λ	1.060	1.302	1.029	1.257	1.005	1.190
	η	3.186	3.536	1.171	1.246	1.030	1.060
$H(0.1)$	λ	1.026	1.299	1.012	1.272	1.002	1.203
	η	2.561	3.119	1.087	1.230	1.014	1.070
$H(0)$	λ	1.000	1.800	1.000	1.800	1.000	1.800
	η	2.000	7.333	1.000	2.778	1.000	2.778

We consider the multivariate normal (NOR) distribution $N(\mu, \Sigma)$ and the symmetric contaminated normal (SCN) distribution $0.9N(\mu, \Sigma) + 0.1N(\mu, 9\Sigma)$. Table 1 lists the asymptotic variances. It partly overlaps similar tables in Maronna (1976) and Tyler (1983).

In all cases the location estimator has an asymptotic covariance which is a certain multiple λ of Σ. The expression for λ for $S(r, p)$ is obtained from (5.3), and the expression for λ for $H(w, p)$ is given in Maronna (1976). The values of λ are listed in Table 1.

In all cases the covariance estimator has an asymptotic covariance that is of type (5.4) [Tyler (1982)]. To measure the asymptotic variance of the covariance estimators we distinguish the cases $p = 1$ and $p \geq 2$. If $p = 1$, then (1.1) reduces to $(1/\sigma)f[(x - \mu)^2/\sigma^2]$ and we give the value $\eta = 2\sigma_1 + \sigma_2$ which represents the asymptotic variance of $n^{1/2}(s_n^2 - \sigma^2)$, where s_n^2 denotes the estimate for the scale parameter σ^2 of the underlying distribution.

For $p \geq 2$ we give the value $\eta = \sigma_1$. Tyler (1983) compared values of σ_1 for different covariance M-estimators with simulated values of a Monte Carlo study of robust covariance estimators in Devlin, Gnanadesikan and Kettenring (1981). It turned out that σ_1 suffices as an index for the asymptotic variance of the correlation estimator based upon the robust covariance estimator. The expression for σ_1 for $S(r, p)$ is given in (5.4), and the expression for σ_1 for $H(w, p)$ is given in Tyler (1982).

From Table 1 we see that the asymptotic variances of $S(r, p)$ for r not too large are of similar magnitude as the asymptotic variances of $H(w, p)$, except at

TABLE 2

Comparison of breakdown points of $S(r, p)$ and $H(w, p)$ at the same level of asymptotic variance attained at NOR *and* SCN

			$S(r,p)$ at NOR		$S(r,p)$ at NOR		$S(r,p)$ at SCN	
p	$H(w,p)$	δ^*	λ	$r(\lambda)$	σ_1	$r(\sigma_1)$	λ	$r(\lambda)$
2	$H(0.1)$	0.217	1.012	0.104	1.087	0.192	1.272	0.256
	$H(0.223)$	0.333	1.033	0.162	1.190	0.256	1.257	0.239
	$H(0.3)$	0.169	1.048	0.189	1.256	0.285	1.260	0.243
10	$H(0.1)$	0.063	1.002	0.124	1.014	0.263	1.203	0.500
	$H(0.2)$	0.074	1.005	0.186	1.030	0.349	1.190	0.497
	$H(0.3)$	0.085	1.009	0.238	1.047	0.406	1.185	0.487
	$H(0.358)$	0.091	1.011	0.258	1.057	0.432	1.183	0.483

the SCN distribution where the $H(w, p)$-covariance estimator has a better performance. In general the asymptotic variance of $S(r, p)$ decreases simultaneously with the breakdown point r. However, in contrast with M-estimators, for every dimension p it is possible to construct an S-estimator with a high breakdown point.

It is interesting to compare the breakdown points of $S(r, p)$ and $H(w, p)$ at the same level of asymptotic variance. Table 2 gives such a comparison.

According to Tyler (1986), when $k^2 > p$, then the limiting value of the breakdown point of $H(w, p)$ equals $\delta^* = \min\{1/k^2, 1 - p/k^2\}$, which is maximal when $k^2 = p + 1$. The values $w = 0.223$ and 0.358 in Table 2 correspond with the values $k^2 = p + 1$.

Given the asymptotic variance λ of the $H(w, p)$-location estimator at the NOR distribution, the constant c_0 of ρ_B $(\cdot, c_0)$ is determined such that the $S(r, p)$-location estimator achieves the same level of λ. With this value of c_0 the breakdown point $r(\lambda)$ is computed by means of (6.1). Next this procedure is repeated given the value σ_1 of the $H(w, p)$-covariance estimator at the NOR distribution, and finally the procedure is repeated given the value λ of the $H(w, p)$-location estimator at the SCN distribution.

We conclude that the S-estimator is able to achieve the asymptotic variances attained by the M-estimator, but in addition it has a breakdown point that becomes considerably higher when the dimension p increases.

APPENDIX

The proofs of existence and consistency of S-estimators in Davies (1987) extend fairly easily to existence and continuity of S-functionals. The following lemma is fundamental.

LEMMA 3.1. *Let* $(\mathbf{t}, \mathbf{C}) \in \Theta$, $0 < m_0 < \infty$, $0 < c < \infty$ *and* $0 < \varepsilon < 1$.

(i) *If F satisfies* (C_ε) *and* $P_F(E(\mathbf{t}, \mathbf{C}, c)) \geq \varepsilon$, *then* $\lambda_p(\mathbf{C}) \geq k_1 > 0$, *where k_1 only depends on ε, F and c.*

(ii) *Assume* $\lambda_p(\mathbf{C}) \geq k_1 > 0$. *If* $\int \rho(\|\mathbf{y}\|/m_0)\, dF(\mathbf{y}) \leq b_0$, *then any solution* $(\mathbf{t}, \mathbf{C})$ *of* $(\mathscr{P}_F)$ *must have* $\lambda_1(\mathbf{C}) \leq k_2 < \infty$, *where* k_2 *only depends on* k_1 *and* m_0.

(iii) *Let* F *satisfy* (C_ε) *and suppose* $P_F(E(\mathbf{t}, \mathbf{C}, c)) \geq \varepsilon$. *If* $\lambda_1(\mathbf{C}) \leq k_2 < \infty$, *then* $(\mathbf{t}, \mathbf{C})$ *is contained in a compact set* $K \subset \Theta$, *which only depends on* ε, F, c *and* k_2.

Proof. Because $E(\mathbf{t}, \mathbf{C}, c)$ is contained in some strip $H(\alpha, l, 2c\sqrt{\lambda_p(\mathbf{C})})$ it follows from (C_ε) that $\lambda_p(\mathbf{C}) \geq (\delta_\varepsilon/c)^2/4 \geq 0$, which proves (i).

The function ρ is continuous and nondecreasing on $[0, \infty)$, so any solution of $(\mathscr{P}_F)$ is also a solution to the same minimization problem with constraint (3.1) replaced by

$$\text{(A.1)} \qquad \int \rho\left[\left\{(\mathbf{y} - \mathbf{t})^T \mathbf{C}^{-1} (\mathbf{y} - \mathbf{t})\right\}^{1/2}\right] dF(y) \leq b_0.$$

As the pair $(\mathbf{0}, m_0^2 \mathbf{I})$ satisfies (A.1) we conclude that any possible solution of $(\mathscr{P}_F)$ must have $\det(\mathbf{C}) \leq m_0^{2p}$. Because $\lambda_p(\mathbf{C}) \geq k_1 > 0$ we find that $\lambda_1(\mathbf{C}) \leq m_0^{2p}/k_1^{p-1} < \infty$ which proves (ii).

Let B_ε be the compact set such that $P_F(B_\varepsilon) \geq 1 - \varepsilon$. Then $\|\mathbf{t} - \mathbf{y}\| \leq c\sqrt{k_2}$ for some $\mathbf{y} \in B_\varepsilon$. Otherwise $E(\mathbf{t}, \mathbf{C}, c)$ would be contained in B_ε^c which would be in contradiction with $P_F(E(\mathbf{t}, \mathbf{C}, c)) \geq \varepsilon$. So $\|\mathbf{t}\|$ is bounded and together with (i) the lemma follows. □

Proof of Theorem 3.1. Let $(\mathbf{t}, \mathbf{C}) \in \Theta$ satisfy constraint (3.1). Then we find

$$\text{(A.2)} \qquad \begin{aligned} P_F(E(\mathbf{t}, \mathbf{C}, c_0)) &\geq 1 - \frac{1}{a_0} \int \rho\left[\left\{(\mathbf{y} - \mathbf{t})^T \mathbf{C}^{-1} (\mathbf{y} - \mathbf{t})\right\}^{1/2}\right] dF(\mathbf{y}) \\ &= 1 - r \geq \varepsilon. \end{aligned}$$

Lemma 3.1(i) implies that $\lambda_p(\mathbf{C}) \geq k_1 > 0$. Because $\lim_{m \to \infty} \int \rho(\|\mathbf{y}\|/m)\, dF(\mathbf{y}) = 0$, there exists an $m_0 > 0$ such that $\int \rho(\|\mathbf{y}\|/m_0)\, dF(\mathbf{y}) \leq b_0$. Lemma 3.1(ii) yields that $\lambda_1(\mathbf{C}) \leq k_2 < \infty$. Finally Lemma 3.1(iii) implies that for solving $(\mathscr{P}_F)$ one may restrict to a compact subset $K \subset \Theta$. As $\det(\mathbf{C})$ is a continuous function of $(\mathbf{t}, \mathbf{C})$ it must attain a minimum on K. □

Lemma 3.2. *Let* G_k, $k \geq 0$, *be a sequence of distributions on* $\mathbb{R}^p$ *that converges weakly to* F *as* $k \to \infty$. *Let* θ_k, $k \geq 0$, *be a sequence in* Θ *such that* $\theta_k \to \theta_L$ *as* $k \to \infty$. *If* $g(\mathbf{y}, \theta) = \rho[\{(\mathbf{y} - \mathbf{t})^T \mathbf{C}^{-1} (\mathbf{y} - \mathbf{t})\}^{1/2}]$, *then*

$$\text{(A.3)} \qquad \lim_{k \to \infty} \int g(\mathbf{y}, \theta_k)\, dG_k(\mathbf{y}) = \int g(\mathbf{y}, \theta_L)\, dF(\mathbf{y}).$$

Proof. Put $g_k(\mathbf{y}) = g(\mathbf{y}, \theta_k)$ and $g_L(\mathbf{y}) = g(\mathbf{y}, \theta_L)$. Then for every sequence $\{\mathbf{y}_k\}$ with $\mathbf{y}_k \to \mathbf{y}$ we have

$$\lim_{k \to \infty} g_k(\mathbf{y}_k) = g_L(\mathbf{y}).$$

Next apply Theorem 5.5 of Billingsley (1968). Let Γ: $[0, \infty) \to [0, \infty)$ be the function $\Gamma(\mathbf{y}) = \mathbf{y}\mathbf{1}_{[0, a_0]}(\mathbf{y}) + a_0 \mathbf{1}_{(a_0, \infty)}(\mathbf{y})$, which is a bounded uniformly continuous function. Then as a consequence of $G_k \Rightarrow F$ we have

$$\lim_{k\to\infty} \int \Gamma(g_k(\mathbf{y}))\, dG_k(\mathbf{y}) = \int \Gamma(g_L(\mathbf{y}))\, dF(\mathbf{y}),$$

which proves (A.3). □

PROOF OF THEOREM 3.2. According to Ranga Rao (1962), Theorem 4.2 we have

$$\sup_{E\in\mathscr{C}} |P_{G_k}(E) - P_F(E)| \to 0 \quad \text{as } k \to \infty. \tag{A.4}$$

Because strips $H(\alpha, l, \delta) \in \mathscr{C}$, (A.4) implies that for k sufficiently large every strip with $P_{G_k}(H(\alpha, l, \delta)) \geq 1 - r$ must also satisfy $P_F(H(\alpha, l, \delta)) \geq \varepsilon$. This means that

$$\inf\{\delta\colon P_{G_k}(H(\alpha, l, \delta)) \geq 1 - r\} \geq \inf\{\delta\colon P_F(H(\alpha, l, \delta)) \geq \varepsilon\} > 0.$$

Next consider the compact set B_ε of (C_ε). We may assume that it belongs to $\mathscr{C}$ and therefore, as $P_F(B_\varepsilon) \geq 1 - \varepsilon > r$, for k sufficiently large $P_{G_k}(B_\varepsilon) \geq r$. We conclude that for k sufficiently large G_k satisfies (C_{1-r}) and according to Theorem 3.1 at least one solution exists.

Denote $\mathbf{S}(G_k) = \boldsymbol{\theta}_k = (\mathbf{t}_k, \mathbf{C}_k)$. Because compact (convex) sets are transformed affinely into compact (convex) sets and because $\mathbf{S}(\cdot)$ is affine equivariant, we may assume that $\mathbf{S}(F) = (\mathbf{0}, \mathbf{I})$. Similar to (A.2) we find $P_{G_k}(E(\mathbf{t}_k, \mathbf{C}_k, c_0)) \geq 1 - r$, such that from (A.4) it follows that for k sufficiently large $P_F(E(\mathbf{t}_k, \mathbf{C}_k, c_0)) \geq \frac{1}{2}(1 - r)$. Lemma 3.1(i) implies that $\lambda_p(\mathbf{C}_k) \geq k_1 > 0$ for k sufficiently large.

According to Lemma 3.2 for any $\eta > -1$, we have

$$\int \rho(\|\mathbf{y}\|/(1+\eta))\, dG_k(\mathbf{y}) \to \int \rho(\|\mathbf{y}\|/(1+\eta))\, dF(\mathbf{y}),$$

as $k \to \infty$. As the limit is strictly decreasing at $\eta = 0$ we see that for any $\eta > 0$,

$$\int \rho\left(\frac{\|\mathbf{y}\|}{1+\eta}\right) dG_k(\mathbf{y}) \leq \int \rho(\|\mathbf{y}\|)\, dF(\mathbf{y}) = b_0$$

for k sufficiently large. Then similar to the proof of Lemma 3.1(ii) we find that $\det(\mathbf{C}_k) \leq (1+\eta)^{2p}$ eventually. Because $\eta > 0$ may be taken arbitrarily small, we conclude that

$$\limsup_{k\to\infty} \det(\mathbf{C}_k) \leq 1 \tag{A.5}$$

and we find that $\lambda_1(\mathbf{C}_k) \leq 4^p/k_1^{p-1}$ eventually. With Lemma 3.1(iii) we see that there exists a compact set K such that for k sufficiently large the sequence $\{\boldsymbol{\theta}_k\} \subset K$.

Consider a convergent subsequence $\{\boldsymbol{\theta}_{k_j}\}$ with $\boldsymbol{\theta}_{k_j} \to \boldsymbol{\theta}_L = (\mathbf{t}_L, \mathbf{C}_L)$. With Lemma 3.2 we find that

$$\int \rho[d(\mathbf{y}; \mathbf{t}_L, \mathbf{C}_L)]\, dF(\mathbf{y}) = \lim_{j\to\infty} \int \rho\left[d(\mathbf{y}; \mathbf{t}_{k_j}, \mathbf{C}_{k_j})\right] dG_{k_j}(\mathbf{y}) = b_0.$$

So $\boldsymbol{\theta}_L$ turns out to satisfy constraint (3.1) of $(\mathscr{P}_F)$ which has solution $(\mathbf{0}, \mathbf{I})$. This means that $\det(\mathbf{C}_L) \geq 1$. Next, (A.5) yields $\det(\mathbf{C}_L) = 1$. But then uniqueness of $(\mathbf{0}, \mathbf{I})$ implies $\theta_L = (\mathbf{0}, \mathbf{I})$. As $\{\theta_k\}$ eventually stays in a compact set we must have $\lim_{k\to\infty} \theta_k = (\mathbf{0}, \mathbf{I})$. □

To prove Theorem 3.3 we will use the following version of the implicit function theorem.

THEOREM 3.4. *Let $\mathscr{M}$ be a metric space, $(h_0, \theta_0) \in \Omega \subset \mathscr{M} \times \mathbb{R}^k$, Ω open. When $\mathbf{W}$: $\mathscr{M} \times \mathbb{R}^k \to \mathbb{R}^k$, with $\mathbf{W}(h_0, \theta_0) = \mathbf{0}$ is such that*

1. *$\mathbf{W}$ is continuous on Ω,*
2. *$\partial \mathbf{W}/\partial \theta$ is continuous on Ω,*
3. *$\partial \mathbf{W}/\partial \theta$ is nonsingular at (h_0, θ_0),*

then there exists a neighbourhood $B_1 \times B_2$ of (h_0, θ_0) on which a function $\theta(\cdot)$: $B_1 \to B_2$ exists such that $\mathbf{W}(h, \theta(h)) = \mathbf{0}$. Moreover it holds that:

4. *If $(\tilde{h}, \tilde{\theta}) \in B_1 \times B_2$ with $\mathbf{W}(\tilde{h}, \tilde{\theta}) = \mathbf{0}$, then $\tilde{\theta} = \theta(\tilde{h})$.*
5. *$\theta(\cdot)$ is continuous on B_1.*

Solutions θ_h of $(\mathscr{P}_{G_h})$ eventually are contained in an open neighbourhood N of $\mathbf{S}(F)$, and they satisfy (3.3) or equivalently the pairs (h, θ_h) are a zero of the function $\mathbf{W}(\cdot; \mathbf{x})$: $[0,1] \times \Theta \to \Theta$, defined as

$$\text{(A.6)} \qquad \begin{aligned} \mathbf{W}(h, \theta; \mathbf{x}) &= \int \Psi(\mathbf{y}, \theta)\, dG_h(\mathbf{y}) \\ &= (1-h)\int \Psi(\mathbf{y}, \theta)\, dF(\mathbf{y}) + h\Psi(\mathbf{x}, \theta), \end{aligned}$$

where $\Psi = (\Psi_1, \Psi_2)$ is defined in (2.8). Theorem 3.4 will be applied to the function $\mathbf{W}(h, \theta; \mathbf{x})$ considered on the open subset $\Omega = [0,1] \times N$ of $[0,1] \times \mathbb{R}^{p+(1/2)p(p+1)}$. In this setup Theorem 3.4(4) will ensure the uniqueness of $\mathbf{S}(\cdot)$ at G_h, sufficiently close to F, but first we note that the existence of limit (3.2) is also implied by conditions 1–3 above.

COROLLARY 3.5. *Let $\mathbf{x} \in \mathbb{R}^p$ and let $\mathbf{W}(h, \theta; \mathbf{x})$, as defined in (A.6), satisfy conditions 1–3 of Theorem 3.4 in the setup described above. Let $\theta_0 = (\mathbf{t}(F), \mathbf{C}(F))$. When $\mathbf{W}(\cdot, \theta_0; \mathbf{x})$ has a right derivative*

$$\frac{\partial \mathbf{W}}{\partial h}(0, \theta_0; \mathbf{x}) = \lim_{h \downarrow 0} \frac{\mathbf{W}(h, \theta_0; \mathbf{x}) - \mathbf{W}(0, \theta_0)}{h}$$

at $h = 0$, then also the function $\theta(\cdot; \mathbf{x})$ has a right derivative at $h = 0$, with

$$\text{(A.7)} \qquad \frac{\partial \theta}{\partial h}(0; \mathbf{x}) = -\left[\frac{\partial \mathbf{W}}{\partial \theta}(0, \theta_0)\right]^{-1} \frac{\partial \mathbf{W}}{\partial h}(0, \theta_0; \mathbf{x}).$$

PROOF. When $\mathscr{M}$ would be a Banach space it is a known consequence of the implicit function theorem that when $\mathbf{W}(\cdot; \mathbf{x})$ is continuously differentiable on Ω

and satisfies (3), the function $\theta(\cdot)$ is continuously differentiable on $\mathbf{B}_1$ [see Dieudonné (1960)]. We are not dealing with a Banach space $\mathscr{M}$, but we are also not interested in the total (or Fréchet) derivative of $\theta(\cdot;\mathbf{x})$ but only in the right derivative of $\theta(\cdot;\mathbf{x})$ at $h=0$. It is not difficult to prove (A.7) similar to Theorem 10.2.1 in Dieudonné (1960). □

LEMMA 3.3. *Let* $\rho: \mathbb{R} \to [0,\infty)$ *satisfy* (R1)–(R3) *and consider the function* $\Psi(\mathbf{x},\theta)$ *of* (2.8). *Then*:

(i) Ψ *is bounded and continuous on* $\mathbb{R}^p \times \Theta$.

(ii) $\partial\Psi/\partial\theta$ *is continuous on* $\mathbb{R}^p \times \Theta$ *and is bounded by a constant which depends only upon* $\|\mathbf{C}\|$ *and* $\|\mathbf{C}^{-1}\|$.

PROOF. Continuity of Ψ is obvious and boundedness of the functions $u(y)y$, $u(y)y^2$ and $v(y)$ proves (i).

For (ii) compute the derivative $\partial\Psi/\partial\theta$:

$$\frac{\partial\Psi_1}{\partial\mathbf{t}} = -\left(\frac{u'(d)}{d}\mathbf{C}^{-1}(\mathbf{x}-\mathbf{t})(\mathbf{x}-\mathbf{t})^T + u(d)\mathbf{I}\right),$$

$$\frac{\partial\Psi_{1,j}}{\partial\mathbf{C}} = -\frac{u'(d)}{2d}(x_j - t_j)(2\mathbf{V} - \mathbf{D}_\mathbf{V}),$$

$$\frac{\partial\Psi_{2,ij}}{\partial\mathbf{t}} = -p\frac{u'(d)}{d}(x_i - t_i)(x_j - t_j)\mathbf{C}^{-1}(\mathbf{x}-\mathbf{t})$$
$$+ pu(d)\frac{\partial(x_i - t_i)(x_j - t_j)}{\partial\mathbf{t}} + \frac{v'(d)}{d}c_{ij}\mathbf{C}^{-1}(\mathbf{x}-\mathbf{t}),$$

$$\frac{\partial\Psi_2}{\partial c_{ij}} = -p\frac{u'(d)}{2d}(2\mathbf{V} - \mathbf{D}_\mathbf{V})_{ij}(\mathbf{x}-\mathbf{t})(\mathbf{x}-\mathbf{t})^T$$
$$+ \frac{v'(d)}{2d}(2\mathbf{V} - \mathbf{D}_\mathbf{V})_{ij}\mathbf{C} - v(d)\frac{\partial\mathbf{C}}{\partial c_{ij}},$$

where $d = d(\mathbf{x};\mathbf{t},\mathbf{C})$ and $\mathbf{V} = \mathbf{C}^{-1}(\mathbf{x}-\mathbf{t})(\mathbf{x}-\mathbf{t})^T\mathbf{C}^{-T}$.

Because $\|\mathbf{x}-\mathbf{t}\|^2/d^2 \le \|\mathbf{C}\|$ and (R3), the second statement (ii) follows. □

PROOF OF THEOREM 3.3. Apply Theorem 3.4 and its Corollary 3.3 to the function $\mathbf{W}(\cdot;\mathbf{x})$ of (A.6) considered on $[0,1]\times N$. Let $\theta_0 = (\mathbf{t}(F),\mathbf{C}(F))$ be the unique solution of $(\mathscr{P}_F)$ such that $\mathbf{W}(0,\theta_0) = \mathbf{0}$.

According to Lemma 3.3(i) the function $\Psi(\mathbf{y},\theta)$ is bounded and continuous, so we conclude from (A.6) that also $\mathbf{W}(h,\theta;\mathbf{x})$ is continuous on Ω for all $\mathbf{x}\in\mathbb{R}^p$. Lemma 3.3(ii) implies that $\partial\Psi/\partial\theta$ is bounded and continuous on $\mathbb{R}^p\times N$, so

$$\frac{\partial\mathbf{W}}{\partial\theta}(h,\theta;\mathbf{x}) = (1-h)\int\frac{\partial\Psi}{\partial\theta}(\mathbf{y},\theta)\,dF(\mathbf{y}) + h\frac{\partial\Psi}{\partial\theta}(\mathbf{x},\theta)$$

is also a continuous function on Ω. Finally, we have that

$$\frac{\partial \mathbf{W}}{\partial \boldsymbol{\theta}}(0, \boldsymbol{\theta}_0) = \int \frac{\partial \boldsymbol{\Psi}}{\partial \boldsymbol{\theta}}(\mathbf{y}, \boldsymbol{\theta}_0)\, dF(\mathbf{y}) = \Lambda, \tag{A.8}$$

which is nonsingular. So conditions 1–3 of Theorem 3.4 hold.

Let h be sufficiently small such that $(\mathscr{P}_{G_h})$ has at least one solution. Suppose that $\boldsymbol{\theta}_{h,1}$ and $\boldsymbol{\theta}_{h,2}$ would be two solutions of $(\mathscr{P}_{G_h})$. Then according to Theorem 3.2 both $(h, \boldsymbol{\theta}_{h,1})$ and $(h, \boldsymbol{\theta}_{h,2})$ are contained in the neighbourhood $B_1 \times B_2$ mentioned in Theorem 3.4 for h sufficiently small, and they are both a zero of $\mathbf{W}(\cdot\,; \mathbf{x})$. We may therefore conclude that $\boldsymbol{\theta}_{h,1} = \boldsymbol{\theta}_{h,2} = \boldsymbol{\theta}(h; \mathbf{x})$. For h sufficiently small the functional $\mathbf{S}(\cdot)$ is thus uniquely defined as

$$\mathbf{S}(G_h) = \mathbf{S}\big((1-h)F + h\Delta_{\mathbf{x}}\big) = \boldsymbol{\theta}(h; \mathbf{x}).$$

Corollary 3.5 implies the existence of $\mathrm{IF}(\mathbf{x}; \mathbf{S}, F)$ and the expression can be obtained from (A.7). As $\int \boldsymbol{\Psi}(\mathbf{y}, \boldsymbol{\theta}_0)\, dF(\mathbf{y}) = \mathbf{W}(0, \boldsymbol{\theta}_0) = \mathbf{0}$, we find that at $(0, \boldsymbol{\theta}_0)$ the derivative $\partial \mathbf{W}/\partial h = \boldsymbol{\Psi}(\mathbf{x}, \boldsymbol{\theta}_0)$ and the theorem follows. □

PROOF OF THEOREM 4.1. Put

$$U(\mathbf{x}; \boldsymbol{\theta}, \delta) = \sup_{\|\boldsymbol{\tau} - \boldsymbol{\theta}\| \le \delta} \|\boldsymbol{\Psi}(\mathbf{x}, \boldsymbol{\tau}) - \boldsymbol{\Psi}(\mathbf{x}, \boldsymbol{\theta})\|.$$

According to Huber's (1967) Theorem 3 and its corollary it is sufficient to prove the following conditions:

1. There exists a $\boldsymbol{\theta}_0 \in \Theta$ such that $\boldsymbol{\lambda}_F(\boldsymbol{\theta}_0) = \mathbf{0}$.
2. There exist strictly positive constants b, c and d_0 such that (i) $\mathbb{E}_F U(X_1; \boldsymbol{\theta}, \delta) \le b\delta$ for $\|\boldsymbol{\theta} - \boldsymbol{\theta}_0\| + \delta \le d_0$ and (ii) $\mathbb{E}_F U^2(X_1; \boldsymbol{\theta}, \delta) \le c\delta$ for $\|\boldsymbol{\theta} - \boldsymbol{\theta}_0\| + \delta \le d_0$.
3. $\mathbb{E}_F \|\boldsymbol{\Psi}(X_1, \boldsymbol{\theta}_0)\|^2$ is finite.

According to Theorem 3.1 a solution $\mathbf{S}(F)$ of $(\mathscr{P}_F)$ exists and it must therefore satisfy (3.3). In other words $\boldsymbol{\theta}_0 = (\mathbf{t}(F), \mathbf{C}(F))$ is a zero of $\boldsymbol{\lambda}_F(\boldsymbol{\theta})$ which proves 1. Lemma 3.3(i) yields condition 3.

Let K be any compact subset of Θ which contains $\boldsymbol{\theta}_0$. We will show that for all $\boldsymbol{\theta} \in K^0$ and δ sufficiently small, there exists a constant $b > 0$ such that

$$U(\mathbf{x}; \boldsymbol{\theta}, \delta) \le b\delta. \tag{A.9}$$

This obviously yields condition 2. Let $\boldsymbol{\theta} = (\mathbf{t}, \mathbf{C}) \in K$. So both $\|\mathbf{C}\|$ and $\|\mathbf{C}^{-1}\|$ are bounded away from 0 and ∞.

Let δ be sufficiently small such that the ball $B_\delta(\boldsymbol{\theta}) \subset K$. Then the mean value theorem together with Lemma 3.3(ii) yield that there exists some constant $b > 0$ such that for $\boldsymbol{\tau} \in B_\delta(\boldsymbol{\theta})$ we have $\|\boldsymbol{\Psi}(\mathbf{x}, \boldsymbol{\tau}) - \boldsymbol{\Psi}(\mathbf{x}, \boldsymbol{\theta})\| \le b\|\boldsymbol{\theta} - \boldsymbol{\tau}\| \le b\delta$. This proves (A.9) and the theorem follows. □

Before proving Corollaries 5.1 and 5.2 we state three minor lemmas. The first one states a property of elliptical distributions.

LEMMA 5.1. *Let X_0 have an elliptical distribution F with parameter $(\mathbf{0}, \mathbf{I})$. Then $U = X_0/\|X_0\|$ is independent of $\|X_0\|$, has mean zero and covariance* *matrix* $(1/p)\mathbf{I}$. *Furthermore* $\mathbb{E}_{\mathbf{0},\mathbf{I}}UU^TU = 0$ *and* $\mathbb{E}_{\mathbf{0},\mathbf{I}}\,\mathrm{vec}(UU^T)\mathrm{vec}(UU^T) = \sigma_1(\mathbf{I} + \mathbf{K}_{p,p}) + \sigma_2\,\mathrm{vec}(\mathbf{I})\mathrm{vec}(\mathbf{I})^T$, *where* $\sigma_1 = \sigma_2 = (p(p+2))^{-1}$.

PROOF OF LEMMA 5.1. To show independence of U and $\|X_0\|$ it is sufficient to show that $\|X_0\|$ and $(U_1,\dots,U_{p-1})^T$ are independent. This can be proven immediately by performing the coordinate transformation $Y_i = U_i$, for $i = 1,2,\dots,p-1$ and $Y_p = \|X_0\|$, and computing the simultaneous density of $(Y_1, Y_2,\dots,Y_p)^T$. The other results can be obtained by using spherical coordinates in a suitable manner. □

LEMMA 5.2. *Let* $\mathbf{Z}$ *be a random* $p \times p$*-matrix which has zero mean and covariance matrix* $\mathbb{E}\,\mathrm{vec}(\mathbf{Z})\mathrm{vec}(\mathbf{Z})^T = \sigma_1(\mathbf{I} + \mathbf{K}_{p,p}) + \sigma_2\,\mathrm{vec}(\mathbf{I})\mathrm{vec}(\mathbf{I})^T$. *Suppose that* $\mathbf{B}\mathbf{B}^T = \Sigma$, *then* $\mathbf{B}\mathbf{Z}\mathbf{B}^T$ *has zero mean and covariance matrix* $\sigma_1(\mathbf{I} + \mathbf{K}_{p,p})(\Sigma \otimes \Sigma) + \sigma_2\,\mathrm{vec}(\Sigma)\mathrm{vec}(\Sigma)^T$.

PROOF. We use two identities from Magnus and Neudecker (1979). First $\mathrm{vec}(\mathbf{ABC}) = (\mathbf{C}^T \otimes \mathbf{A})\mathrm{vec}(\mathbf{B})$, which implies that $\mathrm{vec}(\mathbf{BZB}^T) = (\mathbf{B} \otimes \mathbf{B})\mathrm{vec}(\mathbf{Z})$ and also that $\mathrm{vec}(\Sigma) = (\mathbf{B} \otimes \mathbf{B})\mathrm{vec}(\mathbf{I})$. The second identity $\mathbf{K}_{m,n}(\mathbf{A} \otimes \mathbf{B}) = (\mathbf{B} \otimes \mathbf{A})\mathbf{K}_{m,n}$ yields $(\mathbf{B} \otimes \mathbf{B})\mathbf{K}_{p,p}(\mathbf{B} \otimes \mathbf{B})^T = \mathbf{K}_{p,p}(\mathbf{B} \otimes \mathbf{B})(\mathbf{B} \otimes \mathbf{B})^T$. As it is not difficult to see that $(\mathbf{B} \otimes \mathbf{B})(\mathbf{B} \otimes \mathbf{B})^T = \Sigma \otimes \Sigma$ the lemma follows. □

LEMMA 5.3. *Let* $\mathbf{1}$ *denote the* $(p \times p)$*-matrix with all entries equal to* 1. *For* $a, b, c, d \in \mathbb{R}$ *it holds that*:

(i) *If* $a \neq 0$ *and* $a + pb \neq 0$, *then* $(a\mathbf{I} + b\mathbf{1})^{-1} = (1/a)\mathbf{I} + (b/(a(a+pb)))\mathbf{1}$.
(ii) $(c\mathbf{I} + d\mathbf{1})(a\mathbf{I} + b\mathbf{1})(c\mathbf{I} + d\mathbf{1}) = c^2a\mathbf{I} + (cad + ad(c+pd) + b(c+pd)^2)\mathbf{1}$.

PROOF. Straightforward. □

PROOF OF COROLLARY 5.1. Affine equivariance of $\mathbf{t}_n$ and $\mathbf{C}_n$ and Lemma 5.2 imply that we may restrict to $\theta_0 = (\mathbf{0}, \mathbf{I})$. Obviously the conditions of Theorem 3.2 hold for elliptical distributions. Therefore in order to apply Theorem 4.1 we are left with showing that $\Lambda_{F_{0,I}}(\cdot)$ has a nonsingular derivative at θ_0. To show this and to derive (5.4) we first consider the symmetric $p \times p$-matrix $\mathbf{C}$ as $\frac{1}{2}p(p+1)$-vector $(c_{11},\dots,c_{pp},c_{12},\dots,c_{p-1,p})^T$ consisting of the upper-right triangle elements of the matrix $\mathbf{C}$.

According to (A.8), $\Lambda = \mathbb{E}_{\mathbf{0},\mathbf{I}}\Psi_\theta(X_1, \theta_0)$. With Lemma 5.1 it follows from the expressions found for $\partial\Psi_1/\partial\theta$ and $\partial\Psi_2/\partial\theta$ in the proof of Lemma 3.3 that Λ is of block form

$$\Lambda = \begin{pmatrix} \Delta_{\mathbf{t}} & \\ & \Delta_{\mathbf{C}} \end{pmatrix}, \tag{A.10}$$

where $\Delta_{\mathbf{t}} = \mathbb{E}_{\mathbf{0},\mathbf{I}}(\Psi_1)_{\mathbf{t}}(X_1, \theta_0)$ and $\Delta_{\mathbf{C}} = \mathbb{E}_{\mathbf{0},\mathbf{I}}(\Psi_2)_{\mathbf{C}}(X_1, \theta_0)$.

Using Lemma 5.1 again we see that $\Delta_{\mathbf{t}} = -\beta\mathbf{I}$ and is therefore nonsingular. The matrix $\Delta_{\mathbf{C}}$ is a $(\frac{1}{2}p(p+1) \times \frac{1}{2}p(p+1))$-matrix which consists of two nonzero block matrices on the main diagonal. The upper-left is a $(p \times p)$-matrix $\Delta_{C,1} = -\gamma_1 \mathbf{I} + \gamma_2 \mathbf{1}$ and the lower-right matrix is a diagonal $(\frac{1}{2}p(p-1) \times \frac{1}{2}p(p-1))$-matrix $\Delta_{C,2} = -\gamma_1 \mathbf{I}$, where

$$\gamma_1 = \frac{\mathbb{E}_{0,\mathbf{I}}\left[\psi'(\|X_1\|)\|X_1\|^2 + (p+1)\psi(\|X_1\|)\|X_1\|\right]}{p+2},$$

(A.11)

$$\gamma_2 = \frac{\mathbb{E}_{0,\mathbf{I}}\left[2\psi'(\|X_1\|)\|X_1\|^2 + p\psi(\|X_1\|)\|X_1\|\right]}{2p(p+2)}.$$

As the matrix $\Delta_{C,1}$ has determinant $(-\gamma_1 + p\gamma_2)(-\gamma_1)^{p-1}$, it follows from (5.2) that Δ_C is also nonsingular and hence Theorem 3.1 applies.

To obtain the expressions for the asymptotic covariance matrices first compute the covariance matrix $\mathbf{M}$ of $(\Psi_1, \Psi_2)^T$, with the symmetric matrix Ψ_2 considered as a $\frac{1}{2}p(p+1)$-vector. Lemma 5.1 implies that $\mathbf{M}$ is also a block matrix,

$$\mathbf{M} = \begin{array}{|c|c|} \hline \mathbf{M}_1 & \\ \hline & \mathbf{M}_2 \\ \hline \end{array},$$

where $\mathbf{M}_1 = \mathbb{E}_{0,\mathbf{I}} u^2(\|X_1\|) X_1 X_1^T = \alpha \mathbf{I}$ and $\mathbf{M}_2 = \mathbb{E}_{0,\mathbf{I}} \Psi_2(X_1, \theta_0)\Psi_2(X_1, \theta_0)^T$. The matrix Λ^{-1} is of the same structure as $\mathbf{M}$, which immediately implies that $\mathbf{t}_n$ and $\mathbf{C}_n$ are asymptotically independent and that $\mathbf{t}_n$ has asymptotic covariance matrix

$$\Delta_{\mathbf{t}}^{-1}\mathbf{M}_1\Delta_{\mathbf{t}}^{-T} = \frac{\alpha}{\beta^2}\mathbf{I}.$$

To describe the $(\frac{1}{2}p(p+1) \times \frac{1}{2}p(p+1))$-matrix $\Delta_{\mathbf{C}}^{-1}\mathbf{M}_2\Delta_{\mathbf{C}}^{-1}$, consider the covariance matrix $\mathbf{M}_2$ of $\Psi_2(X_1, \theta_0)$. Because

$$\Psi_{2,ij}(\mathbf{x}, \theta_0) = p\psi(\|\mathbf{x}\|)\|\mathbf{x}\|\frac{x_i x_j}{\|\mathbf{x}\|^2} - v(\|\mathbf{x}\|)\delta_{ij},$$

Lemma 5.1 implies that $\mathbf{M}_2$ is of the same structure as $\Delta_{\mathbf{C}}$. It also consists of two nonzero block matrices on the main diagonal. The upper left is the $(p \times p)$-covariance matrix $\mathbf{M}_{2,1}$ of the diagonal elements $\Psi_{2,ii}(X_1, \theta_0)$: $\mathbf{M}_{2,1} = \delta_1 \mathbf{I} + \delta_2 \mathbf{1}$, and the lower right matrix is a diagonal $(\frac{1}{2}p(p-1) \times \frac{1}{2}p(p-1))$-covariance matrix $\mathbf{M}_{2,2}$ of the off-diagonal elements $\Psi_{2,ij}(X_1, \theta_0)$ $(1 \le i \le j \le p)$: $\mathbf{M}_{2,2} = \frac{1}{2}\delta_1 \mathbf{I}$, where $\delta_1 = 2p(p+2)^{-1}\mathbb{E}_{0,\mathbf{I}}\psi^2(\|X_1\|)\|X_1\|^2$ and $\delta_2 = -\delta_1/p + \mathbb{E}_{0,I}(\rho(\|X_1\|) - b_0)^2$. Therefore $\Delta_{\mathbf{C}}^{-1}$ and $\mathbf{M}_2$ are of the same structure and hence $\Delta_{\mathbf{C}}^{-1}\mathbf{M}_2\Delta_{\mathbf{C}}^{-1}$ is. It follows immediately that the lower-right matrix is a $(\frac{1}{2}p(p-1) \times \frac{1}{2}p(p-1))$-diagonal matrix with diagonal element

$$\text{(A.12)} \quad \sigma_1 = \frac{\delta_1}{2\gamma_1^2} = \frac{p(p+2)\mathbb{E}_{0,\mathbf{I}}\psi^2(\|X_1\|)\|X_1\|^2}{\left\{\mathbb{E}_{0,\mathbf{I}}\left[\psi'(\|X_1\|)\|X_1\|^2 + (p+1)\psi(\|X_1\|)\|X_1\|\right]\right\}^2}.$$

The upper-left matrix is the $(p \times p)$-matrix $\Delta_{\mathbf{C},1}^{-1}\mathbf{M}_{2,1}\Delta_{\mathbf{C},1}^{-T}$. Using Lemma 5.2, easy but tedious computations show that this equals $2\sigma_1 \mathbf{I} + \sigma_2 \mathbf{1}$, with σ_2 as in

(5.5).

The expressions found for $\Delta_{C,1}^{-1}M_{2,1}\Delta_{C,1}^{-T}$ and $\Delta_{C,2}^{-1}M_{2,2}\Delta_{C,2}^{-T}$ tell us that $n^{1/2}(\mathbf{C}_n - \mathbf{I})$ converges in distribution to a symmetric random matrix $\mathbf{Z}$ of which the off-diagonal elements are uncorrelated with each other and uncorrelated with the diagonal elements, of which each off-diagonal element has variance σ_1, and of which the diagonal elements all have variance $2\sigma_1 + \sigma_2$ with the covariance between any two diagonal elements being σ_2. In other words, $\mathbb{E}\,\mathrm{vec}(\mathbf{Z})\mathrm{vec}(\mathbf{Z})^T = \sigma_1(\mathbf{I} + \mathbf{K}_{p,p}) + \sigma_2\,\mathrm{vec}(\mathbf{I})\mathrm{vec}(\mathbf{I})^T$, which proves the corollary for the case $\theta_0 = (\mathbf{0}, \mathbf{I})$. Lemma 5.2 then implies the general form (5.4). □

PROOF OF COROLLARY 5.2. The conditions of Theorem 3.2 hold for elliptical distributions. According to (5.1), $\mathbf{S}(F_{\mathbf{0},\mathbf{I}}) = (\mathbf{0}, \mathbf{I})$ and from the proof of Corollary 5.1 we have that $\lambda_{F_{\mathbf{0},\mathbf{I}}}(\cdot)$ has nonsingular derivative Λ of (A.10) at $(\mathbf{0}, \mathbf{I})$. Therefore Theorem 3.3 applies, which means that $\mathrm{IF}(\mathbf{x}; \mathbf{S}, F_{\mathbf{0},\mathbf{I}})$ exists and its expression can be obtained from (3.4). As Λ consists of the two block matrices $\Delta_{\mathbf{t}}$ and $\Delta_{\mathbf{C}}$ on the main diagonal, $\mathrm{IF}(\mathbf{x}; \mathbf{t}, F_{\mathbf{0},\mathbf{I}})$ and $\mathrm{IF}(\mathbf{x}; \mathbf{C}, F_{\mathbf{0},\mathbf{I}})$ can be treated separately.

Equations (3.4) and (2.8) give

$$\mathrm{IF}(\mathbf{x}; \mathbf{t}, F_{\mathbf{0},\mathbf{I}}) = -\Delta_{\mathbf{t}}^{-1}\Psi_1(\mathbf{x}, (\mathbf{0}, \mathbf{I})) = \frac{1}{\beta}u(\|\mathbf{x}\|)\mathbf{x},$$

which proves (5.7). Let us denote by $\mathbf{IF} = (\mathrm{IF}_{11}, \dots, \mathrm{IF}_{pp}, \mathrm{IF}_{12}, \dots, \mathrm{IF}_{p-1,p})^T$ the influence function $\mathrm{IF}(\mathbf{x}; \mathbf{C}, F_{\mathbf{0},\mathbf{I}})$ of the covariance estimator. Then (3.4) and (2.8), together with the expression found for $\Delta_{\mathbf{C}}$ in the proof of Corollary 5.1, yield

$$-\gamma_1 \mathrm{IF}_{ii} + \gamma_2\,\mathrm{trace}(\mathbf{IF}) = -pu(\|\mathbf{x}\|)x_i^2 + v(\|\mathbf{x}\|), \tag{A.13}$$

$$-\gamma_1 \mathrm{IF}_{ij} = -pu(\|\mathbf{x}\|)x_i x_j, \tag{A.14}$$

where γ_1 and γ_2 are as in (A.11). Summation of (A.13) over $i = 1, 2, \dots, p$ gives

$$\mathrm{trace}(\mathbf{IF}) = \frac{-p\psi(\|\mathbf{x}\|)\|\mathbf{x}\| + pv(\|\mathbf{x}\|)}{-\gamma_1 + p\gamma_2}. \tag{A.15}$$

From (A.11) we have $-\gamma_1 + p\gamma_2 = -\frac{1}{2}\mathbb{E}_{\mathbf{0},\mathbf{I}}\psi(\|\mathbf{X}_0\|)\|\mathbf{X}_0\|$, and when we put in $v(y) = \psi(y)y - \rho(y) + b_0$, we find (5.9).

Finally substitute (A.15) into (A.13). Together with (A.14) this proves (5.8). □

Acknowledgments. I thank Peter Rousseeuw for stimulating discussions and helpful remarks. I also thank the two referees whose comments have led to a considerable improvement of the initial version of this paper.

REFERENCES

BILLINGSLEY, P. (1968). *Convergence of Probability Measures*. Wiley, New York.

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

DEVLIN, S. J., GNANADESIKAN, R. and KETTENRING, J. R. (1981). Robust estimation of dispersion matrices and principal components. *J. Amer. Statist. Assoc.* **76** 354–362.

DIEUDONNÉ, J. (1960). *Foundations of Modern Analysis*. Academic, New York.

GRAYBILL, F. A. (1983). *Matrices with Applications in Statistics*, 2nd ed. Wadsworth, Belmont, Calif.

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

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

HUBER, P. J. (1964). Robust estimation of a location parameter. *Ann. Math. Statist.* **35** 73–101.

HUBER, P. J. (1967). The behavior of maximum likelihood estimates under nonstandard conditions. *Proc. Fifth Berkeley Symp. Math. Statist. Probab.* **1** 221–233. Univ. California Press.

HUBER, P. J. (1981). *Robust Statistics*. Wiley, New York.

KIM, J. and POLLARD, D. (1988). Cube root asymptotics. *Ann. Statist.* To appear.

LOPUHAÄ, H. P. and ROUSSEEUW, P. J. (1987). Breakdown properties of affine equivariant estimators of multivariate location and covariance matrices. Report No. 14, Faculty of Mathematics and Informatics, Delft Univ. Technology.

MAGNUS, J. R. and NEUDECKER, H. (1979). The commutation matrix: Some properties and applications. *Ann. Statist.* **7** 381–394.

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

RANGA RAO, R. (1962). Relations between weak and uniform convergence of measures with applications. *Ann. Math. Statist.* **33** 659–680.

ROUSSEEUW, P. J. (1986). Multivariate estimation with high breakdown point. In *Mathematical Statistics and Applications* (W. Grossmann, G. Pflug, I. Vincze and W. Wertz, eds.) 283–297. Reidel, Dordrecht.

ROUSSEEUW, P. J. and YOHAI, V. J. (1984). Robust regression by means of *S*-estimators. *Robust and Nonlinear Time Series Analysis. Lecture Notes in Statist.* **26** 256–272. Springer, New York.

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

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

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

FACULTY OF MATHEMATICS AND INFORMATICS
DELFT UNIVERSITY OF TECHNOLOGY
JULIANALAAN 132
2628 BL DELFT
THE NETHERLANDS