The Annals of Statistics
1992, Vol. 20, No. 1, 398–413

HIGHLY EFFICIENT ESTIMATORS OF MULTIVARIATE LOCATION WITH HIGH BREAKDOWN POINT[1]

BY HENDRIK P. LOPUHAÄ

Delft University of Technology

We propose an affine equivariant estimator of multivariate location that combines a high breakdown point and a bounded influence function with high asymptotic efficiency. This proposal is basically a location M-estimator based on the observations obtained after scaling with an affine equivariant high breakdown covariance estimator. The resulting location estimator will inherit the breakdown point of the initial covariance estimator and within the location-covariance model only the M-estimator will determine the type of influence function and the asymptotic behaviour. We prove consistency and asymptotic normality and obtain the breakdown point and the influence function.

1. Introduction. Consider the standard location-covariance model, that is, one observes p-dimensional $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n$ and assumes these are realizations of independent random vectors $X_1, X_2, \dots, X_n$, with an elliptical distribution $P_{\mu,\Sigma}$ which has a density

$$f_{\mu,\Sigma}(\mathbf{x}) = |\mathbf{B}|^{-1} f\left(\left\|\mathbf{B}^{-1}(\mathbf{x} - \boldsymbol{\mu})\right\|\right), \tag{1.1}$$

where $\boldsymbol{\mu} \in \mathbb{R}^p$ and $\boldsymbol{\Sigma} = \mathbf{B}\mathbf{B}^T$ is a positive definite symmetric $p \times p$-matrix and where $f\colon [0,\infty) \to [0,\infty)$ is a known function.

A well-known estimator for the location parameter $\boldsymbol{\mu}$ is the least squares estimator, defined as the vector $\mathbf{t}_n \in \mathbb{R}^p$ that minimizes $\sum_{i=1}^n \|\mathbf{x}_i - \mathbf{t}\|^2$, which results in the sample mean. In case $P_{\mu,\Sigma}$ is a normal distribution, this estimator corresponds to the maximum likelihood estimator for $\boldsymbol{\mu}$ and is therefore asymptotically efficient at $P_{\mu,\Sigma}$. However, it is not robust at all, as only one single outlier already has a large effect on the estimator.

Hampel (1968) introduced the breakdown point ε^* and the influence function IF to measure the global and local sensitivity of an estimator. Donoho and Huber (1983) proposed a finite-sample version ε_n^* of the breakdown point, which may be interpreted as the minimum fraction of outliers that can make the estimate arbitrarily large. The influence function $\mathrm{IF}(\mathbf{x}; \mathbf{t}, P)$ describes the effect of one single outlier $\mathbf{x}$ on the estimator [see Hampel (1974) and Hampel, Ronchetti, Rousseeuw and Stahel (1986) for a discussion]. The poor robustness of the sample mean, for instance, is illustrated by its breakdown point $\varepsilon_n^* = 1/n$

Received April 1989; revised April 1991.

[1]Research is part of author's Ph.D. thesis and is financially supported by NWO under Grant 10-62-10.

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

Key words and phrases. Multivariate location, high breakdown point, bounded influence, high efficiency.

and unbounded influence function $\mathrm{IF}(\mathbf{x}; \mathbf{t}, P) = \mathbf{x}$ at any distribution P with mean zero.

Only a few multivariate location estimators have been proposed that combine good global and local robustness with another desirable property for multivariate estimators, namely, equivariance under affine transformations of the observations. Unfortunately, all these estimators exhibit relatively poor asymptotic behaviour. The rate of convergence is generally slower than the usual $\sqrt{n}$ rate, the limiting distribution may not be normal or the asymptotic efficiency is disappointingly low.

The proposal in this paper is basically a location M-estimator based upon the observations obtained after scaling with an affine equivariant covariance estimator. The resulting location estimator will be affine equivariant and if we estimate within the location-covariance model, only the M-estimator determines the rate of convergence as well as the asymptotic efficiency, independently of the initial covariance estimator as long as it is consistent. Concerning the robustness properties, the type of influence function at an elliptically contoured distribution is the same as that of the location M-functional independently of the initial covariance functional as long as it is continuous. The breakdown point of the initial covariance estimator is inherited by the resulting location estimator. Using a high breakdown covariance estimator and a suitable highly efficient location M-estimator will provide an affine equivariant multivariate location estimator that combines a high finite-sample breakdown point and a bounded influence function with high asymptotic efficiency. Other work in this direction has been developed by Yohai (1987) and Yohai and Zamar (1988) for regression estimators, and recently by Davies (1990) and Lopuhaä (1990) for simultaneous estimators of location and covariance.

In Section 2 we define the estimator and give sufficient conditions for the corresponding functional to exist. In Section 3 we prove consistency and show that the estimator is asymptotically normal. The robustness of the estimator is treated in Section 4.

2. Definition. Location M-estimators are a well-known robustification of the least squares method. Similar to Huber (1964), one may define an M-estimator of multivariate location as the vector of $\mathbf{m}_n \in \mathbb{R}^p$ that minimizes

$$\sum_{i=1}^{n} \rho(\|\mathbf{x}_i - \mathbf{m}\|). \tag{2.1}$$

Typically, $\rho(y)$ is a symmetric function which is quadratic in the middle and which increases slower than y^2 as $y \to \infty$. An example is the function

$$\rho_{\mathrm{H}}(y; k) = \begin{cases} \frac{1}{2}y^2, & |y| \le k, \\ -\frac{1}{2}k^2 + k|y|, & |y| \ge k, \end{cases} \tag{2.2}$$

whose derivative ψ_{H} is a bounded monotone function known as Huber's ψ-function. In general, an unbounded function ρ in (2.1) which does not increase too fast may lead to location M-estimators with breakdown point

$\lfloor (n+1)/2 \rfloor / n$ [Huber (1984)], with a bounded influence function, and with good asymptotic efficiency relative to the maximum likelihood estimator at several spherically symmetric distributions [Maronna (1976)]. Unfortunately, these location M-estimators are not equivariant with respect to affine transformations of the $\mathbf{x}_i$. Maronna (1976) solves this by defining M-estimators simultaneously for location and covariance, but these estimators become more sensitive to outliers as the dimension p increases: $\varepsilon_n^* \leq 1/(p+1)$, due to breakdown of the covariance M-estimator [Tyler (1986)]. To obtain affine equivariance and to retain the good breakdown properties of $\mathbf{m}_n$, we propose the following alternative.

DEFINITION 2.1. Let $\mathbf{C}_n = \mathbf{C}_n(\mathbf{x}_1, \dots, \mathbf{x}_n)$ be a positive definite symmetric covariance estimator which is affine equivariant; that is, $\mathbf{C}_n(\mathbf{A}\mathbf{x}_1 + \mathbf{v}, \dots, \mathbf{A}\mathbf{x}_n + \mathbf{v}) = \mathbf{A}\mathbf{C}_n(\mathbf{x}_1, \dots, \mathbf{x}_n)\mathbf{A}^T$ for all nonsingular matrices $\mathbf{A}$ and vectors $\mathbf{v}$. Define $\mathbf{t}_n$ as the vector that minimizes

$$(2.3)\quad R_n(\mathbf{t}) = \frac{1}{n}\sum_{i=1}^{n}\left\{\rho\left(\sqrt{(\mathbf{x}_i - \mathbf{t})^T\mathbf{C}_n^{-1}(\mathbf{x}_i - \mathbf{t})}\right) - \rho\left(\sqrt{\mathbf{x}_i^T\mathbf{C}_n^{-1}\mathbf{x}_i}\right)\right\}.$$

The corresponding location functional $\mathbf{t}(\cdot)$ is defined at a distribution P as the vector $\mathbf{t}(P)$ that minimizes the function

$$(2.4)\quad \begin{aligned} R_P(\mathbf{t}) = \int\Big\{&\rho\left(\sqrt{(\mathbf{x}-\mathbf{t})^T\mathbf{C}(P)^{-1}(\mathbf{x}-\mathbf{t})}\right) \\ &-\rho\left(\sqrt{\mathbf{x}^T\mathbf{C}(P)^{-1}\mathbf{x}}\right)\Big\}\,dP(\mathbf{x}), \end{aligned}$$

where $\mathbf{C}(\cdot)$ is the affine equivariant covariance functional corresponding with $\mathbf{C}_n$, that is, $\mathbf{C}(P_{\mathbf{A}X+\mathbf{v}}) = \mathbf{A}\mathbf{C}(P_X)\mathbf{A}^T$, where P_X denotes the distribution of a random vector X.

Throughout the paper we will assume that $\rho: \mathbb{R} \to [0,\infty)$ satisfies:

(R) ρ is symmetric, $\rho(0) = 0$ and $\rho(y) \to \infty$ as $y \to \infty$. The functions $\psi = \rho'$ and $u(y) = \psi(y)/y$ are continuous, $\psi \geq 0$ on $[0,\infty)$ and there exists a $y_0 > 0$ such that ψ is nondecreasing on $(0, y_0)$ and nonincreasing on (y_0, ∞).

Conditions (R) are sufficient for $\mathbf{m}_n$ to have a high breakdown point and they will guarantee that $\mathbf{t}_n$ inherits the breakdown point of $\mathbf{C}_n$. They are somewhat weaker than in Huber (1984) to include the function ρ_{H} of (2.2). These conditions are also sufficient for the existence of at least one vector $\mathbf{t}(P)$ that minimizes $R_P(\mathbf{t})$.

There are several candidates for a robust covariance estimator. Examples of covariance estimators with a high breakdown point are the Stahel–Donoho [Stahel (1981), Donoho (1982)] estimator, the minimum volume ellipsoid (MVE) estimator and the minimum covariance determinant (MCD) estimator [Rousseeuw (1985)] or smoothed versions of the MVE estimator called S-estimators [Davies (1987), Lopuhaä (1989)]. And of course there are alternative choices of the covariance estimator, such as a distribution-free M-estimator

[Tyler (1987)], which have a lower breakdown point but which may have other robustness properties that are better. For instance, Yohai and Maronna (1990) have shown that despite its high breakdown point the maximum asymptotic bias of the MVE can be much larger than that of the distribution-free M-estimator for amounts of contamination smaller than $1/p$.

Denote by $\lambda_p(\mathbf{A}) \leq \cdots \leq \lambda_1(\mathbf{A})$ the eigenvalues of a positive definite symmetric matrix $\mathbf{A}$ and recall the property

$$\frac{\|\mathbf{v}\|^2}{\lambda_1(\mathbf{A})} \leq \mathbf{v}^T\mathbf{A}^{-1}\mathbf{v} \leq \frac{\|\mathbf{v}\|^2}{\lambda_p(\mathbf{A})} \tag{2.5}$$

for such matrices. Let

$$M(\mathbf{t}) = \sup_{\mathbf{x}} |\rho(\|\mathbf{x} + \mathbf{t}\|) - \rho(\|\mathbf{x}\|)|. \tag{2.6}$$

LEMMA 2.1. *The difference* $\eta(\mathbf{t}) = M(\mathbf{t}) - \rho(\|\mathbf{t}\|)$ *is bounded*: $0 \leq \eta(\mathbf{t}) \leq y_0\psi(y_0)$.

PROOF. By symmetry, we may rotate the vectors $\mathbf{x}$ and $\mathbf{x} + \mathbf{t}$ in (2.6) and consider them as multiples of the vector $\mathbf{t}$. Since ρ is increasing on $[0, \infty)$, we may then write $M(\mathbf{t}) = \sup_{\alpha \geq 0}\{\rho((1 + \alpha)\|\mathbf{t}\|) - \rho(\alpha\|\mathbf{t}\|)\}$. Clearly $\eta(\mathbf{t}) \geq 0$. For $\mathbf{t} \neq \mathbf{0}$ fixed, the conditions on the function ψ imply that the function $\alpha \mapsto \rho((1 + \alpha)\|\mathbf{t}\|) - \rho(\alpha\|\mathbf{t}\|)$ attains its maximum at some α^*, where $0 \leq \alpha^* \leq y_0/\|\mathbf{t}\|$. By the mean value theorem, it follows that $\eta(\mathbf{t}) = \rho((1 + \alpha^*)\|\mathbf{t}\|) - \rho(\|\mathbf{t}\|) - \rho(\alpha^*\|\mathbf{t}\|) \leq \psi(y_0)\alpha^*\|\mathbf{t}\| \leq \psi(y_0)y_0$. □

THEOREM 2.1. *Suppose* ρ: $\mathbb{R} \to [0, \infty)$ *satisfies* (R).

(i) *There is at least one vector* $\mathbf{t}(P)$ *that minimizes* $R_P(\mathbf{t})$. *When* ρ *is also strictly convex, then* $\mathbf{t}(P)$ *is uniquely defined.*

(ii) *When* P *is an elliptical distribution with parameters* $\boldsymbol{\mu}$ *and* $\boldsymbol{\Sigma}$ *and if* $\mathbf{C}(P) = \boldsymbol{\Sigma}$, *then* $R_P(\mathbf{t}) \geq R_P(\boldsymbol{\mu})$. *When* f *in* (1.1) *is strictly decreasing, then* $R_P(\mathbf{t})$ *is minimized uniquely by* $\mathbf{t}(P) = \boldsymbol{\mu}$.

PROOF. (i) Denote by λ_1 and λ_p the largest and smallest eigenvalue of $\mathbf{C}(P)$ and let $L = \sup \psi < \infty$. Note that $R_P(\mathbf{t}) \to \infty$ as $\|\mathbf{t}\| \to \infty$. This means that there exists a constant $M > 0$ such that

$$R_P(\mathbf{t}) > 0 = R_P(\mathbf{0}) \quad \text{for all } \|\mathbf{t}\| > M. \tag{2.7}$$

Therefore, for minimizing $R_P(\mathbf{t})$ we may restrict to the set $K = \{\mathbf{t} \in \mathbb{R}^p\colon \|\mathbf{t}\| \leq M\}$. By Lemma 2.1 and dominated convergence it follows that $R_P(\mathbf{t})$ is continuous and therefore it must attain at least one minimum on the compact set K. It is easily seen that strict convexity of ρ implies strict convexity of R_P, which rules out more than one minimum.

(ii) Because $\mathbf{C}(\cdot)$ is affine equivariant, we may assume that $(\boldsymbol{\mu}, \boldsymbol{\Sigma}) = (\mathbf{0}, \mathbf{I})$. For $s \geq 0$, define $\rho^{-1}(s) = \inf\{y \geq 0\colon \rho(y) \geq s\}$ and for $r > 0$, let

$B(\mathbf{t}, r) = \{\mathbf{x}: \|\mathbf{x} - \mathbf{t}\| \le r\}$. Let $\mathbf{t} \neq \mathbf{0}$, then by Fubini's theorem we have that

$$\begin{aligned} R_P(\mathbf{t}) &= \int\int \left(\{0 \le s \le \rho(\|\mathbf{x} - \mathbf{t}\|)\} - \{0 \le s \le \rho(\|\mathbf{x}\|)\}\right) f(\|\mathbf{x}\|)\, ds\, d\mathbf{x} \\ &= \int \left(\int_{B(\mathbf{0}, \rho^{-1}(s))} f(\|\mathbf{x}\|)\, d\mathbf{x} - \int_{B(\mathbf{t}, \rho^{-1}(s))} f(\|\mathbf{x}\|)\, d\mathbf{x} \right) ds. \end{aligned} \tag{2.8}$$

It follows from Anderson's theorem [see, for instance, Tong (1980)] that for every $r > 0$,

$$\int_{B(\mathbf{0}, r)} f(\|\mathbf{x}\|)\, d\mathbf{x} \ge \int_{B(\mathbf{0}, r) + \mathbf{t}} f(\|\mathbf{x}\|)\, d\mathbf{x} \tag{2.9}$$

with equality if and only if $[(B(\mathbf{0}, r) + \mathbf{t}) \cap D_u] = [(B(\mathbf{0}, r) \cap D_u) + \mathbf{t}]$ for every level set $D_u = \{\mathbf{x}: f(\|\mathbf{x}\|) \ge u\}$, $u \ge 0$. This immediately gives $R_P(\mathbf{t}) \ge R_P(\mathbf{0})$.

When f is strictly decreasing, for every $r > 0$, we can find a $u > 0$ such that $D_u = B(\mathbf{0}, r)$. As $\mathbf{t} \neq \mathbf{0}$, it follows that $[(B(\mathbf{0}, r) + \mathbf{t}) \cap B(\mathbf{0}, r)] \neq B(\mathbf{0}, r) + \mathbf{t}$. We conclude that inequality (2.9) is strict for $r > 0$ and hence from (2.8) we have that $R_P(\mathbf{t}) > 0 = R_P(\mathbf{0})$. □

As a consequence of Theorem 2.1(i), under the conditions (R), there exists at least one vector $\mathbf{t}_n = \mathbf{t}(P_n)$ that minimizes $R_n(\mathbf{t})$, where P_n is the empirical distribution. Although it may not be uniquely defined, the robustness properties and the asymptotic behaviour of each $\mathbf{t}_n$ will be the same. Since $\mathbf{C}_n$ is affine equivariant, $\mathbf{t}_n$ will be affine equivariant in the following sense. Denote by $V(\mathbf{x}_1, \dots, \mathbf{x}_n)$ the set of vectors that minimize $R_n(\mathbf{t})$, then $V(\mathbf{A}\mathbf{x}_1 + \mathbf{v}, \dots, \mathbf{A}\mathbf{x}_n + \mathbf{v}) = \mathbf{A}V(\mathbf{x}_1, \dots, \mathbf{x}_n) + \mathbf{v}$ for all nonsingular matrices $\mathbf{A}$ and vectors $\mathbf{v}$, where $\mathbf{A}V + \mathbf{v}$ denotes the set $\{\mathbf{A}\mathbf{x} + \mathbf{v}: \mathbf{x} \in V\}$.

That $\mathbf{t}_n$ inherits the breakdown point of $\mathbf{C}_n$ is suggested by the following argument. Suppose that we start with a covariance estimator $\mathbf{C}_n = \mathbf{A}_n\mathbf{A}_n^T$ and suppose that $\mathbf{C}_n$ does not break down. Since for unbounded functions ρ, the breakdown point of the unscaled M-estimator $\mathbf{m}_n$ at any collection is $\lfloor (n+1)/2 \rfloor / n$, independent of the structure of the collection, we can expect the M-estimate $\tilde{\mathbf{m}}_n = \mathbf{m}_n(\mathbf{A}_n^{-1}\mathbf{x}_1, \dots, \mathbf{A}_n^{-1}\mathbf{x}_n)$ to stay bounded and hence $\mathbf{t}_n$, which is nothing else but $\mathbf{A}_n\tilde{\mathbf{m}}_n$, will stay bounded. This argument will be made rigorous in Section 4.

For $\mathbf{C}_n$ one may use any affine equivariant covariance estimator. A choice for the function ρ in (2.3) may be the function ρ_{H} of (2.2). The location M-estimator defined with ρ_{H} turned out to be Huber's (1964) robust minimax solution. It has good asymptotic efficiency relative to the maximum likelihood estimator at several spherically symmetric distributions [Maronna (1976)] and a bounded influence function as well. The idea is that $\mathbf{t}_n$ will inherit the affine equivariance and the breakdown point from $\mathbf{C}_n$, and that the asymptotic behaviour of $\mathbf{t}_n$ will be similar to that of the corresponding M-estimator. The latter will be shown in the next section.

3. Asymptotic normality. Let $X_1, X_2, \ldots$ be a sequence of independent identically distributed random vectors $X_i = (X_{i1} \cdots X_{ip})^T$ with a distribution P on $\mathbb{R}^p$. Denote by P_n the empirical distribution corresponding with the sample $X_1, \ldots, X_n$. To prove consistency for $\mathbf{t}_n$, we will need that the initial covariance estimator $\mathbf{C}_n$ in (2.3) is consistent for the value of its corresponding functional $\mathbf{C}(\cdot)$ at P, that is,

$$\lim_{n\to\infty} \mathbf{C}_n = \mathbf{C}(P) \tag{3.1}$$

with probability 1. Examples of such estimators are S-estimators, including the MVE estimator [Davies (1987), Lopuhaä (1989)], the MCD estimator [Butler and Jhun (1990)] and M-estimators, including the distribution-free M-estimator [Tyler (1987)].

To prove consistency, we will apply a uniform strong law for empirical processes $(P_n - P)\phi$, indexed by functions ϕ in a class $\mathscr{F}$, as is given in Pollard (1984). This involves concepts like polynomial discrimination and permissibility for which we refer to Pollard (1984). By the *envelope* F of $\mathscr{F}$ is meant a function F for which $|\phi| \le F$ for every $\phi \in \mathscr{F}$. For the sake of brevity, we will sometimes write $Pg(\cdot) = \int g(\mathbf{x})\, dP(\mathbf{x})$ or simply Pg. Let $\Theta = \mathbb{R}^p \times \mathrm{PDS}(p)$, where $\mathrm{PDS}(p)$ denotes the class of all positive definite symmetric matrices.

THEOREM 3.1 (Consistency). *Let* $\rho: \mathbb{R} \to [0,\infty)$ *satisfy conditions* (R). *Suppose that* $\mathbf{t}(P)$ *is uniquely defined. When* $\mathbf{C}_n$ *satisfies* (3.1), *then* $\lim_{n\to\infty} \mathbf{t}_n = \mathbf{t}(P)$ *with probability* 1.

PROOF. For $\boldsymbol{\theta} = (\mathbf{t}, \mathbf{C}) \in \Theta$, write $h_1(\mathbf{x}, \boldsymbol{\theta}) = \rho(\sqrt{(\mathbf{x}-\mathbf{t})^T\mathbf{C}^{-1}(\mathbf{x}-\mathbf{t})}\,)$, $h_2(\mathbf{x}, \boldsymbol{\theta}) = \rho(\sqrt{\mathbf{x}^T\mathbf{C}^{-1}\mathbf{x}}\,)$, $h(\mathbf{x}, \boldsymbol{\theta}) = h_1(\mathbf{x}, \boldsymbol{\theta}) - h_2(\mathbf{x}, \boldsymbol{\theta})$, $H(\boldsymbol{\theta}) = Ph(\cdot\,; \boldsymbol{\theta})$ and $H_n(\boldsymbol{\theta}) = P_n h(\cdot\,; \boldsymbol{\theta})$. Because $\mathbf{C}_n$ is affine equivariant, we can restrict ourselves to the case $(\mathbf{t}(P), \mathbf{C}(P)) = (\mathbf{0}, \mathbf{I})$. In that case $\mathbf{C}_n \to \mathbf{I}$, so that with probability 1, all eigenvalues of $\mathbf{C}_n$ are between $1/4$ and 4, say, for n sufficiently large. By law of large numbers we have that for every $\mathbf{t}$,

$$R_n(\mathbf{t}) \to R_P(\mathbf{t})$$

with probability 1 as $n \to \infty$. Choose $M > 0$ as in (2.7). Then it must hold that eventually $\|\mathbf{t}_n\| \le M$ with probability 1, since otherwise $R_n(\mathbf{t}_n) > 0 = R_n(\mathbf{0})$.

Let $T = \{\boldsymbol{\theta}: \|\mathbf{t}\| \le M, 1/4 \le \lambda_p(\mathbf{C}) \le \lambda_1(\mathbf{C}) \le 4\}$. Let $\varepsilon > 0$ and choose $K > 0$ such that $P\{\|X\| > K\} < \varepsilon$. Consider the classes of functions $\mathscr{H}_{i,K} = \{h_i(\mathbf{x}, \boldsymbol{\theta})\{\|\mathbf{x}\| \le K\}: \boldsymbol{\theta} \in T\}$ for $i = 1, 2$. It is easily seen [see, for instance, Lemma 22 in Nolan and Pollard (1987)] that the class of graphs of functions $h_1(\mathbf{x}, \boldsymbol{\theta})$ has polynomial discrimination [Pollard (1984), page 17]. The graph of a function in $\mathscr{H}_{1,K}$ is the intersection of the graph of $h_1(\mathbf{x}, \boldsymbol{\theta})$ and the cylinder $\{(\mathbf{x}, s): \|\mathbf{x}\| \le K, s \in \mathbb{R}\}$. Since the class of cylinders has polynomial discrimination, it follows from Lemma II.15 in Pollard (1984) that the class of graphs of functions in $\mathscr{H}_{1,K}$ has polynomial discrimination. Similarly, the class of graphs of functions in $\mathscr{H}_{2,K}$ also has polynomial discrimination. Moreover, since the functions are restricted to $\|\mathbf{x}\| \le K$, both classes have bounded envelopes. Finally, since T can be seen as a subset of $\mathbb{R}^{p+(1/2)p(p+1)}$, $\mathscr{F}$ is permissible in the sense of Pollard [(1984), Appendix C]. Then, according to Theorem II.24 and Lemma II.25 in Pollard (1984),

$$\begin{aligned} &\sup_{\theta \in T} \left|(P_n - P) h_1(\cdot, \boldsymbol{\theta})\{\|\mathbf{x}\| \le K\}\right| \to 0, \\ &\sup_{\theta \in T} \left|(P_n - P) h_2(\cdot, \boldsymbol{\theta})\{\|\mathbf{x}\| \le K\}\right| \to 0, \end{aligned} \tag{3.2}$$

with probability 1.

Next, consider the difference $(h_1(\mathbf{x}, \boldsymbol{\theta}) - h_2(\mathbf{x}, \boldsymbol{\theta}))\{\|\mathbf{x}\| > K\}$. Then by Lemma 2.1, for $\boldsymbol{\theta} \in T$ this difference is bounded by some positive constant a, where a only depends on ρ and M. Hence

$$\begin{aligned} &\sup_{\theta \in T} \left|(P_n - P)(h_1(\mathbf{x}, \boldsymbol{\theta}) - h_2(\mathbf{x}, \boldsymbol{\theta}))\{\|\mathbf{x}\| > K\}\right| \\ &\qquad \le a(P_n + P)\{\|\mathbf{x}\| > K\}. \end{aligned} \tag{3.3}$$

With (3.2) and (3.3), we find that

$$\begin{aligned} \gamma_n &= \sup_{\theta \in T} |H_n(\boldsymbol{\theta}) - H(\boldsymbol{\theta})| \\ &\le \sup_{\theta \in T} \left|(P_n - P) h_1(\mathbf{x}, \boldsymbol{\theta})\{\|\mathbf{x}\| \le K\}\right| \\ &\quad + \sup_{\theta \in T} \left|(P_n - P) h_2(\mathbf{x}, \boldsymbol{\theta})\{\|\mathbf{x}\| \le K\}\right| + a(P_n + P)\{\|\mathbf{x}\| > K\} \\ &\to 2aP\{\|X\| > K\} < 2a\varepsilon \end{aligned}$$

with probability 1. As this holds for every $\varepsilon > 0$, we conclude that

$$\sup_{\theta \in T} |H_n(\boldsymbol{\theta}) - H(\boldsymbol{\theta})| \to 0 \tag{3.4}$$

with probability 1.

Because $R_P(\mathbf{t})$ has a unique minimum at $\mathbf{t}(P) = \mathbf{0}$, for all $\delta > 0$ there exist $0 < \alpha < 1$ and $\beta > 0$ such that

$$\inf_{\|\mathbf{t}\| > \delta} \int \left\{ \rho\left(\frac{\|\mathbf{x} - \mathbf{t}\|}{1 + \alpha} \right) - \rho\left(\frac{\|\mathbf{x}\|}{1 - \alpha} \right) \right\} dP(\mathbf{x}) > \beta. \tag{3.5}$$

Let n be sufficiently large such that $\gamma_n \le \beta$ and $1 - \alpha \le \lambda_p(\mathbf{C}_n) \le \lambda_1(\mathbf{C}_n) \le 1 + \alpha$. Then with (3.4) and (3.5), we have

$$\inf_{\|\mathbf{t}\| > \delta} H_n(\mathbf{t}, \mathbf{C}_n) > H_n(\mathbf{0}, \mathbf{C}_n).$$

As $\mathbf{t}_n$ minimizes $H_n(\mathbf{t}, \mathbf{C}_n)$, it follows that $\|\mathbf{t}_n\| \le \delta$ for n sufficiently large with probability 1. We conclude that $\mathbf{t}_n \to \mathbf{0}$ with probability 1. □

To obtain the limiting distribution of $\mathbf{t}_n$, consider the function $R_P(\mathbf{t})$ in (2.4). Because ψ is bounded, $R_P(\mathbf{t})$ has derivative

$$-\mathbf{C}(P)^{-1}\int u\Big(\sqrt{(\mathbf{x}-\mathbf{t})^T\mathbf{C}(P)^{-1}(\mathbf{x}-\mathbf{t})}\Big)(\mathbf{x}-\mathbf{t})\,dP(\mathbf{x}),$$

where $u(y) = \psi(y)/y$. Hence $(\mathbf{t}(P), \mathbf{C}(P))$ will always be a zero of the function

$$G(\boldsymbol{\theta}) = Pg(\cdot\,;\boldsymbol{\theta}), \tag{3.6}$$

where for $\boldsymbol{\theta} = (\mathbf{t}, \mathbf{C})$,

$$g(\mathbf{x};\boldsymbol{\theta}) = u\Big(\sqrt{(\mathbf{x}-\mathbf{t})^T\mathbf{C}^{-1}(\mathbf{x}-\mathbf{t})}\Big)(\mathbf{x}-\mathbf{t}). \tag{3.7}$$

A vector $\mathbf{v} \in \mathbb{R}^p$ is called a *point of symmetry of* P, if $P(\mathbf{v}+A) = P(\mathbf{v}-A)$ for all P-measurable sets $A \subset \mathbb{R}^p$, where for $\lambda \in \mathbb{R}$ and $\mathbf{v} \in \mathbb{R}^p$, $\mathbf{v} + \lambda A$ denotes the set $\{\mathbf{v} + \lambda\mathbf{x}\colon \mathbf{x} \in A\}$. If $\boldsymbol{\mu}$ is a point of symmetry of P, it has the property that

$$G(\boldsymbol{\mu}, \mathbf{C}) = \mathbf{0} \tag{3.8}$$

for all nonsingular matrices $\mathbf{C}$.

We will use the following tightness property from Pollard (1984). It is a combination of the approximation lemma (page 27), Lemma II.36 (page 34) and the equicontinuity lemma (page 150).

LEMMA 3.1. *Let $\mathscr{F}$ be a permissible class of real-valued functions with envelope F and suppose that $0 < PF^2 < \infty$. If the class of graphs of functions in $\mathscr{F}$ has polynomial discrimination, then for each $\eta > 0$ and $\varepsilon > 0$, there exists a $\delta > 0$ for which*

$$\limsup_{n\to\infty} \mathbb{P}\Big\{\sup_{\phi_1,\phi_2\in[\delta]} \big|\sqrt{n}\,(P_n - P)(\phi_1 - \phi_2)\big| > \eta\Big\} < \varepsilon,$$

where $[\delta] = \{(\phi_1, \phi_2)\colon \phi_1, \phi_2 \in \mathscr{F}$ and $P(\phi_1 - \phi_2)^2 \le \delta^2\}$.

To apply Lemma 3.1 we will need that the function $u(y) = \psi(y)/y$ in (3.7) is of bounded variation. This holds for instance for the function $u(y)$ that corresponds with the function ρ_{H} of (2.2).

THEOREM 3.2. *Let $\rho\colon \mathbb{R} \to [0, \infty)$ satisfy condition* (R) *and suppose that the function $u(y) = \psi(y)/y$ is of bounded variation. Let $\mathbf{t}(P)$ be uniquely defined and suppose that it is a point of symmetry of P. Let G in* (3.6) *have a partial derivative $\partial G/\partial\mathbf{t}$ that is continuous at $\boldsymbol{\theta}_0 = (\mathbf{t}(P), \mathbf{C}(P))$ and suppose that $\boldsymbol{\Lambda} = (\partial G/\partial\mathbf{t})(\boldsymbol{\theta}_0)$ is nonsingular. When $\mathbf{C}_n$ satisfies* (3.1), *then $\sqrt{n}\,(\mathbf{t}_n - \mathbf{t}(P))$ has a limiting normal distribution with zero mean and covariance matrix $\boldsymbol{\Lambda}^{-1}\mathbf{M}\boldsymbol{\Lambda}^{-T}$, where $\mathbf{M}$ is the covariance matrix of $g(X_1, \boldsymbol{\theta}_0)$, with g defined in* (3.7).

PROOF. Write $\boldsymbol{\theta}_n = (\mathbf{t}_n, \mathbf{C}_n)$. We first show that

$$\left|\sqrt{n}\,(P_n - P)(g(\cdot,\boldsymbol{\theta}_n) - g(\cdot,\boldsymbol{\theta}_0))\right| \to 0$$

in probability. Write $k(\mathbf{x}, \boldsymbol{\theta}) = u(\sqrt{(\mathbf{x}-\mathbf{t})^T\mathbf{C}^{-1}(\mathbf{x}-\mathbf{t})}\,)$ and consider the jth component $g_j(\mathbf{x}, \boldsymbol{\theta}) = k(\mathbf{x}, \boldsymbol{\theta})(x_j - t_j)$ of the function $g(\mathbf{x}, \boldsymbol{\theta})$. Write $\mathbf{t}(P) = \boldsymbol{\mu} = (\mu_1 \cdots \mu_p)^T$ and $\mathbf{t}_n = (t_{n1} \cdots t_{np})^T$. Decompose $g_j(\mathbf{x}, \boldsymbol{\theta}_n) - g_j(\mathbf{x}, \boldsymbol{\theta}_0)$ as follows:

$$\begin{aligned}(3.9)\qquad &\big(k(\mathbf{x},\boldsymbol{\theta}_n)x_j - k(\mathbf{x},\boldsymbol{\theta}_0)x_j\big) - t_{nj}\big(k(\mathbf{x},\boldsymbol{\theta}_n) - k(\mathbf{x},\boldsymbol{\theta}_0)\big)\\ &\quad + (\mu_j - t_{nj})k(\mathbf{x},\boldsymbol{\theta}_0).\end{aligned}$$

Consider the second difference in (3.9). Because the function $u(y)$ is of bounded variation, it follows from Lemma 22 of Nolan and Pollard (1987) that the class of graphs of the functions $\mathscr{F} = \{k(\mathbf{x}, \boldsymbol{\theta}): \boldsymbol{\theta} \in \Theta\}$ has polynomial discrimination and a bounded envelope. Since Θ can be seen as a subset of $\mathbb{R}^{p+(1/2)p(p+1)}$, $\mathscr{F}$ is permissible in the sense of Pollard (1984), so that Lemma 3.1 applies. Because $\boldsymbol{\theta}_n \to \boldsymbol{\theta}_0$, for each $\delta > 0$, the functions $k(\mathbf{x}, \boldsymbol{\theta}_n)$ and $k(\mathbf{x}, \boldsymbol{\theta}_0)$ are in the class $[\delta]$ of Lemma 3.1 for n sufficiently large. This means that if we integrate (3.9) with respect to $(P_n - P)$, the second difference is $o_P(1/\sqrt{n}\,)$.

Choose $M > 0$ as in (2.7) and define the set T as in the proof of Theorem 3.1. It is easily seen that for $j = 1, \dots, p$ the class of graphs of the functions $\mathscr{F}_j = \{k(\mathbf{x}, \boldsymbol{\theta})x_j: \boldsymbol{\theta} \in T\}$ has polynomial discrimination. Furthermore, since

$$k(\mathbf{x},\boldsymbol{\theta})x_j = k(\mathbf{x},\boldsymbol{\theta})(x_j - t_j) + k(\mathbf{x},\boldsymbol{\theta})t_j,$$

$\mathscr{F}_j$ has a bounded envelope. Since $\mathscr{F}_j$ is permissible for the same reason as $\mathscr{F}$ is, we conclude that the first difference in (3.9) will also be $o_P(1/\sqrt{n}\,)$.

Finally, the last term in (3.9) is $o_P(1/\sqrt{n}\,)$, because $(P_n - P)k(\cdot, \boldsymbol{\theta}_0)$ is $O_P(1/\sqrt{n}\,)$ according to the central limit theorem and $\mathbf{t}_n \to \boldsymbol{\mu}$. It follows that $(P_n - P)(g_j(\cdot, \boldsymbol{\theta}_n) - g_j(\cdot, \boldsymbol{\theta}_0)) = o_P(1/\sqrt{n}\,)$. Since this holds for every $j = 1, \dots, p$, we conclude that

$$(3.10)\qquad (P_n - P)(g(\cdot,\boldsymbol{\theta}_n) - g(\cdot,\boldsymbol{\theta}_0)) = o_P(1/\sqrt{n}\,).$$

Because $\partial G/\partial \mathbf{t}$ is continuous at $\boldsymbol{\theta}_0$, we have that

$$(3.11)\qquad G(\mathbf{t},\mathbf{C}) = G(\boldsymbol{\mu},\mathbf{C}) + \frac{\partial G}{\partial \mathbf{t}}(\boldsymbol{\mu},\mathbf{C})(\mathbf{t}-\boldsymbol{\mu}) + (\mathbf{t}-\boldsymbol{\mu})r(\boldsymbol{\theta}),$$

where $r(\boldsymbol{\theta}) \to 0$ as $\boldsymbol{\theta} \to \boldsymbol{\theta}_0$. Since $\mathbf{t}_n$ minimizes the function $R_n(\mathbf{t})$, the pair $\boldsymbol{\theta}_n = (\mathbf{t}_n, \mathbf{C}_n)$ is a zero of the function $P_n g(\cdot; \theta)$. Hence, together with (3.10), it follows that

$$\begin{aligned}\mathbf{0} &= P_n g(\cdot,\boldsymbol{\theta}_n)\\ &= Pg(\cdot,\boldsymbol{\theta}_n) + (P_n - P)g(\cdot,\boldsymbol{\theta}_0) + (P_n - P)(g(\cdot,\boldsymbol{\theta}_n) - g(\cdot,\boldsymbol{\theta}_0))\\ &= Pg(\cdot,\boldsymbol{\theta}_n) + (P_n - P)g(\cdot,\boldsymbol{\theta}_0) + o_P(1/\sqrt{n}\,).\end{aligned}$$

Then use expansion (3.11) for $Pg(\cdot, \boldsymbol{\theta}_n)$ together with property (3.8). This gives

$$\mathbf{0} = \frac{\partial G}{\partial \mathbf{t}}(\boldsymbol{\mu}, \mathbf{C}_n)(\mathbf{t}_n - \boldsymbol{\mu}) + (\mathbf{t}_n - \boldsymbol{\mu})r(\boldsymbol{\theta}_n) + (P_n - P)g(\cdot, \boldsymbol{\theta}_0) + o_P\left(\frac{1}{\sqrt{n}}\right).$$

Because $\partial G/\partial \mathbf{t}$ is continuous at $\boldsymbol{\theta}_0$ and since $r(\boldsymbol{\theta}_n) = o_P(1)$, this reduces to

$$\mathbf{0} = (\boldsymbol{\Lambda} + o_P(1))(\mathbf{t}_n - \boldsymbol{\mu}) + (P_n - P)g(\cdot, \boldsymbol{\theta}_0) + o_P(1/\sqrt{n}). \tag{3.12}$$

According to the central limit theorem $(P_n - P)g(\cdot, \boldsymbol{\theta}_0) = O_P(1/\sqrt{n})$, and since $\boldsymbol{\Lambda}$ is nonsingular, it follows that $\mathbf{t}_n - \boldsymbol{\mu} = O_P(1/\sqrt{n})$. When we insert this in (3.12), we find that

$$\mathbf{0} = \boldsymbol{\Lambda}(\mathbf{t}_n - \boldsymbol{\mu}) + (P_n - P)g(\cdot, \boldsymbol{\theta}_0) + o_P(1/\sqrt{n}).$$

Because $\boldsymbol{\theta}_0$ is a zero of (3.6), it follows that

$$\sqrt{n}(\mathbf{t}_n - \boldsymbol{\mu}) = -\boldsymbol{\Lambda}^{-1}\frac{1}{\sqrt{n}}\sum_{i=1}^{n} g(X_i, \boldsymbol{\theta}_0) + o_P(1).$$

As $g(\mathbf{x}, \boldsymbol{\theta}_0)$ is bounded, the theorem follows after applying the central limit theorem. □

Note that the rate of convergence of $\mathbf{C}_n$ is irrelevant as long as $\mathbf{t}(P)$ is uniquely defined and is a point of symmetry of P. Let P be elliptically contoured with parameters $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma} = \mathbf{B}\mathbf{B}^T$. According to Theorem 2.1(ii), $\mathbf{t}(P) = \boldsymbol{\mu}$ and it is clear that $\boldsymbol{\mu}$ is a point of symmetry of P. When the function ψ in Theorem 3.2 has a continuous bounded derivative, $\boldsymbol{\Lambda}$ reduces to

$$\boldsymbol{\Lambda} = -\beta \mathbf{I}, \tag{3.13}$$

where

$$\beta = \int \left[\left(1 - \frac{1}{p}\right)u(\|\mathbf{x}\|) + \frac{1}{p}\psi'(\|\mathbf{x}\|)\right] f(\|\mathbf{x}\|)\, d\mathbf{x}. \tag{3.14}$$

When ψ is not differentiable, such as for instance Huber's ψ-function ψ_{H}, the matrix $\boldsymbol{\Lambda}$ may still be of type (3.13) under suitable conditions on the function f. The matrix $\mathbf{M}$ reduces to a multiple $\alpha\boldsymbol{\Sigma}$, where

$$\alpha = \frac{1}{p}\int \psi^2(\|\mathbf{x}\|) f(\|\mathbf{x}\|)\, d\mathbf{x}.$$

When $\boldsymbol{\Lambda}$ is of type (3.13), the limiting covariance of $\sqrt{n}(\mathbf{t}_n - \boldsymbol{\mu})$ reduces to $\gamma\boldsymbol{\Sigma}$. The scalar $\gamma = \alpha/\beta^2$ is independent of $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ and suffices as an index for the asymptotic efficiency. Note that the limiting distribution is the same as that of the corresponding affine equivariant location M-estimator considered in Maronna (1976); in particular, when P is spherically symmetric, the limiting distribution is the same as that of the location M-estimator defined by minimizing (2.1) with the same function ρ. If we use the function $\rho_{\mathrm{H}}(y; k)$ in (2.3), the asymptotic efficiency relative to the maximum likelihood estimator can be read from Table 1 in Maronna (1976). It is reasonable at the multivariate normal, as well as at several multivariate student distributions, for moderate values of k.

4. Robustness. The global behaviour of an estimator under large perturbations of a given situation may be described by the breakdown point, a measure of global sensitivity introduced by Hampel (1968). Donoho and Huber (1983) give the following finite sample version. Let $\mathbf{X} = (\mathbf{x}_1, \dots, \mathbf{x}_n)$ be a collection of n points in $\mathbb{R}^p$ and let $\mathbf{t}_n(\mathbf{X})$ be some location estimator based upon $\mathbf{X}$. The *breakdown point* of a location estimator $\mathbf{t}_n$ at a collection $\mathbf{X}$ is defined as the smallest fraction m/n of outliers for which the estimator can be made arbitrarily large:

$$\varepsilon^*(\mathbf{t}_n, \mathbf{X}) = \min_{1 \le m \le n} \left\{ \frac{m}{n} : \sup_{Y_m} \| \mathbf{t}_n(\mathbf{X}) - \mathbf{t}_n(\mathbf{Y}_m) \| = \infty \right\}, \tag{4.1}$$

where the supremum in (4.1) is taken over all possible corrupted collections of n points $\mathbf{Y}_m = (\mathbf{y}_1, \dots, \mathbf{y}_m, \mathbf{x}_{i_{m+1}}, \dots, \mathbf{x}_{i_n})$ that can be obtained from $\mathbf{X}$ by replacing any m points $\mathbf{x}_{i_1}, \dots, \mathbf{x}_{i_m}$ by arbitrary values $\mathbf{y}_1, \dots, \mathbf{y}_m$. Similarly, the breakdown point of a covariance estimator $\mathbf{C}_n$ at a collection $\mathbf{X}$ is defined as the smallest fraction m/n of outliers that can either take the largest eigenvalue $\lambda_1(\mathbf{C}_n)$ over all bounds, or take the smallest eigenvalue $\lambda_p(\mathbf{C}_n)$ arbitrarily close to zero:

$$\varepsilon^*(\mathbf{C}_n, \mathbf{X}) = \min_{1 \le m \le n} \left\{ \frac{m}{n} : \sup_{Y_m} D(\mathbf{C}_n(\mathbf{X}), \mathbf{C}_n(\mathbf{Y}_m)) = \infty \right\},$$

where $D(\mathbf{A}, \mathbf{B}) = \max\{|\lambda_1(\mathbf{A}) + \lambda_1(\mathbf{B})|, |\lambda_p(\mathbf{A})^{-1} - \lambda_p(\mathbf{B})^{-1}|\}$ and where the supremum is taken over the same corrupted collections as in (4.1). This definition is similar to the one given in Donoho (1982).

The breakdown behaviour of a location M-estimator depends on whether one uses a bounded or unbounded function ρ in (2.1) [see Huber (1984)]. When we use a bounded function ρ, the breakdown point $\varepsilon^*(\mathbf{m}_n, \mathbf{X})$ will depend on the actual structure of the collection $\mathbf{X}$. If we consider the special case of a function ρ that is constant outside an interval $[-c, c]$, this is easy to understand. Namely, if the width $2c$ of such a function ρ is small compared to the distances between the $\mathbf{x}_i$, for instance if the $\mathbf{x}_i$ are at least $2c$ apart, then replacing only one point already forces breakdown of the location M-estimator. On the other hand, if the width $2c$ of ρ is large compared to the distances between the $\mathbf{x}_i$, for instance, if all $\mathbf{x}_i$ are the same, then one needs to replace at least half of the observations to make the M-estimator break down. When ρ satisfies conditions (R), the breakdown point of a location M-estimator will be independent of $\mathbf{X}$ and attains the maximal value possible for translation equivariant location estimators: $\varepsilon^*(\mathbf{m}_n, \mathbf{X}) = \lfloor (n+1)/2 \rfloor / n$. This property is basically the reason why the scaled location M-estimator $\mathbf{t}_n$ will have a breakdown point that is at least equal to that of $\mathbf{C}_n$.

LEMMA 4.1. *Let Q and H be probability measures and $0 \le \varepsilon \le 1$. Define*

$$Q_{\varepsilon,H} = (1-\varepsilon)Q + \varepsilon H.$$

Suppose that Q has a finite first moment and suppose that there exist k_1 and k_2 such that $0 < k_1^2 \le \inf_H \lambda_p(\mathbf{C}(Q_{\varepsilon,H})) \le \sup_H \lambda_1(\mathbf{C}(Q_{\varepsilon,H})) \le k_2^2 < \infty$. When $\int \rho(\|\mathbf{y}\|/k_1)\,dQ(\mathbf{y}) < \infty$, then there exists a constant K independent of H such that

$$R_{Q_{\varepsilon,H}}(\mathbf{t}) \ge (1-2\varepsilon)\rho\left(\frac{\|\mathbf{t}\|}{k_2}\right) - K.$$

PROOF. Write $\mathbf{C}_{\varepsilon,Q,H} = \mathbf{C}(Q_{\varepsilon,H})$. We have that

$$\begin{aligned}\int\Big\{\rho\Big(\sqrt{(\mathbf{y}-\mathbf{t})^T\mathbf{C}_{\varepsilon,Q,H}^{-1}(\mathbf{y}-\mathbf{t})}\Big) - \rho\Big(\sqrt{\mathbf{y}^T\mathbf{C}_{\varepsilon,Q,H}^{-1}\mathbf{y}}\Big)\Big\}\,dQ(\mathbf{y})\\ = \rho\Big(\sqrt{\mathbf{t}^T\mathbf{C}_{\varepsilon,Q,H}^{-1}\mathbf{t}}\Big)\\ + \int\Big\{\rho\Big(\sqrt{(\mathbf{y}-\mathbf{t})^T\mathbf{C}_{\varepsilon,Q,H}^{-1}(\mathbf{y}-\mathbf{t})}\Big) - \rho\Big(\sqrt{\mathbf{t}^T\mathbf{C}_{\varepsilon,Q,H}^{-1}\mathbf{t}}\Big)\Big\}\,dQ(\mathbf{y})\\ - \int\rho\Big(\sqrt{\mathbf{y}^T\mathbf{C}_{\varepsilon,Q,H}^{-1}\mathbf{y}}\Big)\,dQ(\mathbf{y}).\end{aligned}$$

By property (2.5), the third term on the right-hand side is bounded from below by $-\int\rho(\|\mathbf{y}\|/k_1)\,dQ(\mathbf{y})$. By Lemma 2.1 and (2.5), the second term on the right-hand side is bounded from below by $-y_0\psi(y_0) - \int\rho(\|\mathbf{y}\|/k_1)\,dQ(\mathbf{y})$. Similarly we have that

$$\int\Big\{\rho\Big(\sqrt{(\mathbf{y}-\mathbf{t})^T\mathbf{C}_{\varepsilon,Q,H}^{-1}(\mathbf{y}-\mathbf{t})}\Big) - \rho\Big(\sqrt{\mathbf{y}^T\mathbf{C}_{\varepsilon,Q,H}^{-1}\mathbf{y}}\Big)\Big\}\,dQ(\mathbf{y})$$

is bounded from below by $-y_0\psi(y_0) - \rho(\sqrt{\mathbf{t}^T\mathbf{C}_{\varepsilon,Q,H}^{-1}\mathbf{t}})$. We conclude that

$$R_{Q_{\varepsilon,H}}(\mathbf{t}) \ge (1-2\varepsilon)\rho\Big(\sqrt{\mathbf{t}^T\mathbf{C}_{\varepsilon,Q,H}^{-1}\mathbf{t}}\Big) - K \ge (1-2\varepsilon)\rho\left(\frac{\|\mathbf{t}\|}{k_2}\right) - K,$$

where $K = y_0\psi(y_0) + 2\int\rho(\|\mathbf{y}\|/k_1)\,dQ(\mathbf{y})$. □

THEOREM 4.1. *Let $\mathbf{X} = (\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n)$ be a collection of n points in $\mathbb{R}^p$. Let $\rho\colon \mathbb{R} \to [0,\infty)$ satisfy conditions (R) and let $\mathbf{C}_n$ be an affine equivariant positive definite symmetric covariance estimator. Let $\mathbf{t}_n$ be defined in Definition 2.1. Then $\varepsilon^*(\mathbf{t}_n, \mathbf{X}) \ge \varepsilon^*(\mathbf{C}_n, \mathbf{X})$.*

PROOF. If we replace at most $m \le n\varepsilon^*(\mathbf{C}_n, \mathbf{X}) - 1$ points, then $\mathbf{C}_n$ does not break down and we must show that $\|\mathbf{t}_n(\mathbf{Y}_m)\|$ stays bounded. Since $\mathbf{C}_n$ is affine equivariant, $\varepsilon^*(\mathbf{C}_n, \mathbf{X})$ is at most $\lfloor (n-p+1)/2 \rfloor / n$ [see Davies (1987)] and therefore $2m \le n-1$. Apply Lemma 4.1 with $\varepsilon = m/n$, Q the empirical measure corresponding with $\mathbf{x}_{i_{m+1}}, \dots, \mathbf{x}_{i_n}$ and H the empirical measure corresponding with $\mathbf{y}_1, \dots, \mathbf{y}_m$. Denote by $R_n^*(\mathbf{t})$ the function R_n corresponding to the corrupted sample $\mathbf{Y}_m$. Since $\rho(y) \to \infty$ as $y \to \infty$, $R_n^*(\mathbf{t})$ is bounded away from 0 for $\|\mathbf{t}\|$ sufficiently large, independent of $\mathbf{y}_1, \dots, \mathbf{y}_m$. Because $R_n^*(\mathbf{0}) = 0$ and since $\mathbf{t}_n(\mathbf{Y}_m)$ minimizes $R_n^*(\mathbf{t})$, $\mathbf{t}_n(\mathbf{Y}_m)$ must be within a bounded neighbourhood of $\mathbf{0}$, independent of the choice of $\mathbf{y}_1, \dots, \mathbf{y}_m$. □

REMARK. Define the maximum asymptotic bias of $\mathbf{t}$ by

$$B(\varepsilon) = \sup_H \|\mathbf{t}((1-\varepsilon)P + \varepsilon H) - \mathbf{t}(P)\|.$$

Another consequence of Lemma 4.1 is that if for $0 \le \varepsilon < 1/2$ the maximum asymptotic bias of the covariance functional is finite, in the sense that the smallest and largest eigenvalue are bounded away from 0 and ∞, respectively, then also $B(\varepsilon) < \infty$. This can be shown by a similar argument as in the proof of Theorem 4.1. Also note that this implies that the location estimator inherits the gross-error breakdown point $\varepsilon^* = \inf\{\varepsilon\colon B(\varepsilon) < \infty\}$ from the covariance estimator, where the supremum in $B(\varepsilon)$ runs over all probability measures that assign mass 1 to some $\mathbf{x} \in \mathbb{R}^p$.

Whereas the breakdown point measures the global sensitivity, the local sensitivity may be described by the influence function introduced by Hampel (1968, 1974). Let $\mathbf{t}(\cdot)$ be the location functional corresponding with $\mathbf{t}_n$. If $\delta_{\mathbf{x}}$ denotes the probability distribution that assigns mass 1 to $\mathbf{x} \in \mathbb{R}^p$, then the *influence function* of $\mathbf{t}(\cdot)$ at P is defined pointwise as

$$\mathrm{IF}(\mathbf{x}; \mathbf{t}, P) = \lim_{\varepsilon \downarrow 0} \frac{\mathbf{t}((1-\varepsilon)P + \varepsilon\delta_{\mathbf{x}}) - \mathbf{t}(P)}{\varepsilon}$$

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

To obtain IF($\mathbf{x}; \mathbf{t}, P$), we will need that the initial covariance functional $\mathbf{C}(\cdot)$ in (2.4) is continuous at P, that is,

$$\lim_{\varepsilon \downarrow 0} \mathbf{C}((1-\varepsilon)P + \varepsilon\delta_{\mathbf{x}}) = \mathbf{C}(P). \tag{4.2}$$

Examples of such functionals are S-functionals, including the functional corresponding with the MVE estimator [Lopuhaä (1989)].

THEOREM 4.2. *Let $\rho\colon \mathbb{R} \to [0, \infty)$ satisfy conditions (R). Let $\mathbf{t}(P)$ be uniquely defined and suppose that it is a point of symmetry of P. Suppose that the function G defined in* (3.6) *satisfies the conditions of Theorem* 3.2. *When $\mathbf{C}(\cdot)$ satisfies* (4.2), *then for $\mathbf{x} \in \mathbb{R}^p$, it holds that*

$$\mathrm{IF}(\mathbf{x}; \mathbf{t}, P) = -\Lambda^{-1} g(\mathbf{x}; \mathbf{t}(P), \mathbf{C}(P)),$$

where g is defined in (3.7).

PROOF. Write $P_{\varepsilon,\mathbf{x}} = (1-\varepsilon)P + \varepsilon\delta_{\mathbf{x}}$. First show that $\mathbf{t}(P_{\varepsilon,\mathbf{x}}) \to \mathbf{t}(P)$. This is almost a copy of the proof of Theorem 3.1 if we read $P_{\varepsilon,\mathbf{x}}$ instead of P_n.

Instead of (3.4), we now have that

$$\gamma_{\varepsilon,\mathbf{x}} = \sup_{\theta \in T} |P_{\varepsilon,\mathbf{x}} h(\cdot,\boldsymbol{\theta}) - Ph(\cdot,\boldsymbol{\theta})| \le 2\varepsilon y_0 \psi_0 + 2\varepsilon\rho(2M) \to 0,$$

where h, T and M are chosen as in the proof of Theorem 3.1. One can then show $\mathbf{t}(P_{\varepsilon,\mathbf{x}}) \to \mathbf{t}(P)$ completely similar to the proof of Theorem 3.1.

The expression for IF($\mathbf{x}$; $\mathbf{t}$, P) can now be obtained similar to the proof of Theorem 3.2. Write $\boldsymbol{\theta}_{\varepsilon,\mathbf{x}} = (\mathbf{t}_{\varepsilon,\mathbf{x}}, \mathbf{C}_{\varepsilon,\mathbf{x}}) = (\mathbf{t}(P_{\varepsilon,\mathbf{x}}), \mathbf{C}(P_{\varepsilon,\mathbf{x}}))$. Then $\boldsymbol{\theta}_{\varepsilon,\mathbf{x}}$ is a zero of $P_{\varepsilon,\mathbf{x}} g(\cdot, \mathbf{t}, \mathbf{C}_{\varepsilon,\mathbf{x}})$, where g is defined in (3.7). Together with (3.11) and (3.8) it follows that

$$\begin{aligned} \mathbf{0} &= (1-\varepsilon)G(\mathbf{t}_{\varepsilon,\mathbf{x}}, \mathbf{C}_{\varepsilon,\mathbf{x}}) + \varepsilon g(\mathbf{x};\boldsymbol{\theta}_{\varepsilon,\mathbf{x}}) \\ &= (1-\varepsilon)\left\{\frac{\partial G}{\partial t}(\boldsymbol{\mu}, \mathbf{C}_{\varepsilon,\mathbf{x}})(\mathbf{t}_{\varepsilon,\mathbf{x}} - \boldsymbol{\mu}) + (\mathbf{t}_{\varepsilon,\mathbf{x}} - \boldsymbol{\mu}) r(\boldsymbol{\theta}_{\varepsilon,\mathbf{x}})\right\} \qquad (4.3) \\ &\quad + \varepsilon g(\mathbf{x};\boldsymbol{\theta}_{\varepsilon,\mathbf{x}}), \end{aligned}$$

where $\boldsymbol{\mu} = \mathbf{t}(P)$. Because $\partial G/\partial \mathbf{t}$ and g are continuous at $\boldsymbol{\theta}_0$, we obtain

$$\mathbf{0} = (1-\varepsilon)(\boldsymbol{\Lambda} + o(1))(\mathbf{t}_{\varepsilon,\mathbf{x}} - \boldsymbol{\mu}) + O(\varepsilon)$$

as $\varepsilon \downarrow 0$. Because $\boldsymbol{\Lambda}$ is nonsingular, we conclude that $\mathbf{t}_{\varepsilon,\mathbf{x}} - \boldsymbol{\mu} = O(\varepsilon)$ as $\varepsilon \downarrow 0$. When we insert this into (4.3) and use that g is continuous, (4.3) reduces to

$$\frac{\mathbf{t}(P_{\varepsilon,\mathbf{x}}) - \boldsymbol{\mu}}{\varepsilon} = -\boldsymbol{\Lambda}^{-1} g(\mathbf{x};\boldsymbol{\theta}_0) + o(1)$$

as $\varepsilon \downarrow 0$, which completes the proof. □

Note that IF($\mathbf{x}$; $\mathbf{t}$, P) is bounded and that the rate of convergence in (4.5) is irrelevant as long as $\mathbf{t}(P)$ is uniquely defined and is a point of symmetry of P. At elliptically contoured distributions the influence function is the same as that of the corresponding affine equivariant location M-estimator considered in Maronna (1976). In particular, when P is spherically symmetric, the influence function is the same as that of the location M-estimator defined by minimizing (2.1) with the same function ρ:

$$\mathrm{IF}(\mathbf{x};\mathbf{t},P) = \frac{\psi(\|\mathbf{x}\|)}{\beta\|\mathbf{x}\|}\mathbf{x},$$

where β is defined in (3.14). This influence function is *weakly* redescending, that is, $\|\mathrm{IF}\|$ is nonincreasing for $\|\mathbf{x}\| \to \infty$ and nonzero for $\mathbf{x} \neq \mathbf{0}$. A natural question is whether one can also use functions ρ in Definition 2.1 which will lead to *strongly* redescending influence functions, that is, IF($\mathbf{x}$; $\mathbf{t}$, P) $= \mathbf{0}$ for $\mathbf{x}$ outside a certain region. A function $\rho(y)$ that is constant for $|y| \ge c$, say, would correspond with a derivative ψ that is zero for $|y| \ge c$ and hence with an influence function (at spherically symmetric distributions) that is zero for $\|\mathbf{x}\| \ge c$. When we use such a function ρ, the breakdown behaviour of the location M-estimator $\mathbf{m}_n$ will depend on the actual structure of the collection at which it is computed. This suggests that when we use a bounded function ρ

in Definition 2.1, we cannot hope for a breakdown point of $\mathbf{t}_n$ that will be independent of the breakdown behaviour of $\mathbf{C}_n$. Nevertheless, it is possible to use a biweight type ρ-function. The difference is, however, that one can no longer use *any* affine equivariant covariance estimator in Definition 2.1, but only a covariance S-estimator defined by a different ρ-function, and that in order to obtain a result like Theorem 4.1, the two ρ-functions must be related to each other. In that case, one can obtain an affine equivariant location estimator with a high breakdown point, a *strongly* redescending influence function and high asymptotic efficiency relative to the sample mean. This method can be seen as the multivariate version of Yohai's regression *MM*-estimators [Yohai (1987)] and is treated in detail in Lopuhaä (1990).

Acknowledgments. I would like to thank two anonymous referees for their suggestions and remarks. Especially the second referee, who pointed out that several moment conditions in the earlier versions of Theorems 2.1, 3.1 and 3.2 could be removed.

REFERENCES

BUTLER, R. W. and JHUN, M. (1990). Asymptotics for the minimum covariance determinant estimator. Preprint, Colorado State Univ. and Korea Univ.

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

DAVIES, P. L. (1990). Improving S-estimators by means of k-step M-estimators. Technical Report, GHS, Essen.

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

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

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

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

HAMPEL, F. R., RONCHETTI, E. M., ROUSSEEUW, P. J. and STAHEL, W. A. (1986). *Robust Statistics: The Approach Based on Influence Functions.* Wiley, New York.

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

HUBER, P. J. (1984). Finite sample breakdown point of M- and P-estimators. *Ann. Statist.* **12** 119–126.

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

LOPUHAÄ, H. P. (1990). Estimation of location and covariance with high breakdown point. Ph.D. thesis, Delft Univ. of Technology.

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

NOLAN, D. and POLLARD, D. (1987). U-processes: Rates of convergence. *Ann. Statist.* **15** 780–799.

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

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

TONG, Y. L. (1980). *Probability Inequalities in Multivariate Distributions.* Academic, New York.

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

TYLER, D. E. (1987). A distribution-free *M*-estimator of multivariate scatter. *Ann. Statist.* **15** 234–251.

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

YOHAI, V. J. and MARONNA, R. A. (1990). The maximum bias of robust covariances. *Comm. Statist. Theory Methods* **19** 3925–3933.

YOHAI, V. J. and ZAMAR, R. (1988). High breakdown-point of estimates of regression by means of the minimization of an efficient scale. *J. Amer. Statist. Assoc.* **83** 406–413.

DEPARTMENT OF MATHEMATICS
DELFT UNIVERSITY OF TECHNOLOGY
MEKELWEG 4
2628 CD DELFT
THE NETHERLANDS