The Annals of Statistics
1976, Vol. 4, No. 1, 51–67

ROBUST M-ESTIMATORS OF MULTIVARIATE LOCATION AND SCATTER

RICARDO ANTONIO MARONNA

Fundación Bariloche

Let $\mathbf{x}_1, \cdots, \mathbf{x}_n$ be a sample from an m-variate distribution which is spherically symmetric up to an affine transformation. This paper deals with the robust estimation of the location vector $\mathbf{t}$ and scatter matrix $\mathbf{V}$ by means of "M-estimators," defined as solutions of the system: $\sum_i u_1(d_i)(\mathbf{x}_i - \mathbf{t}) = \mathbf{0}$ and $n^{-1} \sum_i u_2(d_i^2)(\mathbf{x}_i - \mathbf{t})(\mathbf{x}_i - \mathbf{t})' = \mathbf{V}$, where $d_i^2 = (\mathbf{x}_i - \mathbf{t})'\mathbf{V}^{-1}(\mathbf{x}_i - \mathbf{t})$.

Existence and uniqueness of solutions of this system are proved under general assumptions about the functions u_1 and u_2. Then the estimators are shown to be consistent and asymptotically normal. The breakdown bound and the influence function are calculated, showing some weaknesses of the estimates for high dimensionality. An algorithm for the numerical calculation of the estimators is described. Finally, numerical values of asymptotic variances, and Monte Carlo small-sample results are exhibited.

1. Introduction. There are several situations in multivariate analysis in which it is desirable to obtain robust affine-invariant estimates of (some substitutes for) the vector of means and of some scalar multiple of the covariance matrix. More precisely, let $\mathbf{x}_1, \cdots, \mathbf{x}_n$ be a sample from an m-variate density f of the form $f(\mathbf{x}) = (\det \mathbf{V})^{-\frac{1}{2}}h[(\mathbf{x} - \mathbf{t})'\mathbf{V}^{-1}(\mathbf{x} - \mathbf{t})]$, where $h(|\mathbf{x}|)$ is a density in R^m ($|\cdot|$ stands for Euclidean norm). Then it is desired to estimate the location vector $\mathbf{t}$ and scatter matrix $\mathbf{V}$, assuming that h is only approximately known.

The most obvious case arises when one is mainly interested in location, and the simultaneous estimation of scatter is simply an auxiliary device to obtain affine-invariant estimates, as in Proposal 2 of Huber (1964). Other situations in which robust estimation of some scalar multiple of the scatter matrix is important in itself are: linear discrimination (Lachenbruch et al., 1973), principal components, and outlier detection (Gnanadesikan and Kettenring, 1972). Besides its mathematical appeal, invariance is a natural requirement when one wants to take into account the linear dependence among the variables, and to represent the situations geometrically.

Among the robust noninvariant procedures considered in the literature, Bickel (1964) and Sen and Puri (1971) treat the coordinatewise application of location estimates based on rank tests (R-estimators), and Gentleman (1965) studies a particular M-estimator of location. Among the affine-invariant estimators, there is a procedure proposed by Tukey—cited by Huber (1972)—called "peeling" which, like univariate trimming, rejects extremal points of the sample; and another one, based on iterative trimming, described in Gnanadesikan and

Received July 1974; revised June 1975.

AMS 1970 *subject classifications.* Primary G2G35; Secondary G2H99.

Key words and phrases. Robust estimation, M-estimators, outlier detection, scatter matrix.

Kettenring (1972), Section 2.2.3. Very little seems to be known about the properties of these two procedures (the one-dimensional version of the second one has been studied by Yohai and Maronna (1975)). Finally, Hampel (1973) suggested an invariant iterative procedure for the estimation of the scatter matrix, which is actually equivalent to an M-estimate.

In this paper we shall study M-estimators of location and scatter, defined as solutions of systems of equations of the form

$$n^{-1} \sum_{i=1}^{n} u_1[\{(\mathbf{x}_i - \mathbf{t})'\mathbf{V}^{-1}(\mathbf{x}_i - \mathbf{t})\}^{\frac{1}{2}}](\mathbf{x}_i - \mathbf{t}) = \mathbf{0}\,, \tag{1.1}$$

$$n^{-1} \sum_{i=1}^{n} u_2[(\mathbf{x}_i - \mathbf{t})'\mathbf{V}^{-1}(\mathbf{x}_i - \mathbf{t})](\mathbf{x}_i - \mathbf{t})(\mathbf{x}_i - \mathbf{t})' = \mathbf{V}\,, \tag{1.2}$$

where u_1 and u_2 are functions satisfying a set of general assumptions which will be stated in Section 2. These estimators are obviously affine-invariant, and include as a particular case the maximum-likelihood estimates for the situation described at the beginning of this section, with the functions $u_1(s) = -s^{-1}\, d[\log h(s)]/ds$ and $u_2(s^2) = u_1(s)$, for $s > 0$.

Under the assumptions mentioned above, the existence and uniqueness of solutions of (1.1)—(1.2) are proved in Sections 3 and 4, and in Sections 5 and 6 the estimates are shown to be consistent and asymptotically normal. Then some robustness measures are calculated in Section 7, and their numerical values, together with those of the asymptotic variances of some particular estimates, are exhibited in Section 8. Finally in Section 9 a procedure for the numerical calculation of the estimators is proposed, and the results of a Monte Carlo experiment on their finite-sample behavior are reported on.

2. Notation and general assumptions. Euclidean norms of vectors and operator norms of matrices will be denoted by $|\cdot|$; $\mathbf{I}$ will be the identity matrix, and $\mathbf{U}'$ the transpose of $\mathbf{U}$. If $\mathbf{M}$ is a positive semidefinite matrix, the squared distance between $\mathbf{x}$ and $\mathbf{y}$ with respect to $\mathbf{M}$ will be denoted by $d^2(\mathbf{x}, \mathbf{y}; \mathbf{M}) = (\mathbf{x} - \mathbf{y})'\mathbf{M}(\mathbf{x} - \mathbf{y})$. The relation "$\mathbf{A} - \mathbf{B}$ is positive semidefinite" will be written as "$\mathbf{A} \geqq \mathbf{B}$".

In general P will denote a probability measure in R^m, and E_P the respective expectation operator, which will be written as E when this causes no ambiguity; $\mathbf{x}$ will always be the "dummy variable" of expectations. The indicator function of the set A will be written $I(\mathbf{x} \in A)$ or $I(A)$. The closed ball with center $\mathbf{0}$ and radius r will be denoted by B_r.

We shall consider solutions $(\mathbf{t}, \mathbf{V})$ of systems of equations of the form

$$E_P u_1(d(\mathbf{x}, \mathbf{t}; \mathbf{V}^{-1}))(\mathbf{x} - \mathbf{t}) = \mathbf{0} \tag{2.1}$$

$$E_P u_2(d^2(\mathbf{x}, \mathbf{t}; \mathbf{V}^{-1}))(\mathbf{x} - \mathbf{t})(\mathbf{x} - \mathbf{t})' = \mathbf{V}\,, \tag{2.2}$$

where the functions $u_1(s)$ and $u_2(s)$ are defined for $s \geqq 0$. If $\{\mathbf{x}_1, \cdots, \mathbf{x}_n\}$ is a sample in R^m of size n, and P is the respective empirical distribution—i.e., the atomic measure such that $P(\{\mathbf{x}_i\}) = n^{-1}$ for $i = 1, \cdots, n$—these equations become (1.1)—(1.2), thus defining the estimators. If P is the underlying distribution, they define the parameters to estimate. Remark that the parameter space is $\Theta = \{(\mathbf{t}, \mathbf{V})\} = \Theta_1 \times \Theta_2$, where $\Theta_1 = R^m$ and Θ_2 is the set of $m \times m$-symmetric, positive definite matrices.

If $\mathbf{S}$ is any matrix such that

$$\mathbf{V} = \mathbf{S}\mathbf{S}' , \tag{2.3}$$

then (2.1)—(2.2) can be written as

$$E_P u_1(|\mathbf{S}^{-1}(\mathbf{x} - \mathbf{t})|)\mathbf{S}^{-1}(\mathbf{x} - \mathbf{t}) = \mathbf{0} \tag{2.4}$$

$$E_P u_2(|\mathbf{S}^{-1}(\mathbf{x} - \mathbf{t})|^2)[\mathbf{S}^{-1}(\mathbf{x} - \mathbf{t})][\mathbf{S}^{-1}(\mathbf{x} - \mathbf{t})]' = \mathbf{I} , \tag{2.5}$$

which determine $\mathbf{t}$ and $\mathbf{V}$ through (2.3). If one wants to determine $\mathbf{S}$ uniquely, he can add to (2.4)—(2.5) the condition that $\mathbf{S} \in \Theta_2$. Either pair of equations will be used, according to mathematical convenience.

To facilitate notation and comparison with the univariate case, define for $s \geqq 0$: $\psi_i(s) = s u_i(s)$ $(i = 1, 2)$. Throughout the paper we shall without further notice assume the following four conditions about these functions:

(A) u_1 and u_2 are nonnegative, nonincreasing, and continuous for $s \geqq 0$.

(B) ψ_1 and ψ_2 are bounded. Let $K_i = \sup_{s \geqq 0} \psi_i(s)$.

(C) ψ_2 is nondecreasing, and is strictly increasing in the interval where $\psi_2 < K_2$.

(D) There exists s_0 such that $\psi_2(s_0^2) > m$, and that $u_1(s) > 0$ for $s \leqq s_0$ (and hence $K_2 > m$).

EXAMPLES. It is easy to verify (A)—(B)—(C)—(D) for two families of estimators to be considered later.

(i) Multivariate Huber's Proposal 2. Let $\psi(z, k)$ be the well-known "Huber's psi function" defined as $\psi(z, k) = \max(-k, \min(z, k))$. Let k_1 and k_2 be two positive constants, and take $\psi_1(s) = \psi(s, k_1)$ and $\psi_2(s) = \psi(s, k_2^2)/\beta$, where $\beta = E_P \psi(|\mathbf{x}|^2, k_2^2)$, P being the m-variate normal distribution with mean $\mathbf{0}$ and identity covariance matrix. As in Huber (1964), the object of dividing by β is to make $\mathbf{V}$ an asymptotically unbiased estimate of the covariance matrix in the normal case. In his Proposal 2, Huber takes $k_1 = k_2$.

(ii) Maximum likelihood estimator for the Student distribution. The m-variate radial Student distribution with p degrees of freedom has density

$$f_{m,p}(\mathbf{x}) = C(p + |\mathbf{x}|^2)^{-(m+p)/2} ,$$

where C is a constant. The maximum-likelihood estimators are given by $\psi_1(s) = (m + p)s/(p + s^2)$ and $\psi_2(s) = (m + p)s/(p + s)$.

A further assumption relating P to u_2 will be needed:

(E) There exists $a > 0$ such that for every hyperplane H, $P(H) \leqq 1 - m/K_2 - a$.

As will be seen below, this condition, which is equivalent to the hypothesis of the proposition on page 97 of Huber (1964), is essential for the existence, uniqueness and consistency of the estimators. This causes no problem with the underlying distribution, since it is customarily assumed continuous. We may however have trouble with finite samples; since any set of m points in R^m is contained in some hyperplane H, if P is the empirical distribution corresponding to a sample of size n, (E) implies that $m/n \leqq P(H) \leqq 1 - m/K_2$. For not too large samples of high dimensionality this imposes a restriction on the estimators, since K_2 must be chosen larger than $mn/(n - m)$. But, as will be seen in (7.1), a large K_2 may imply a loss of robustness.

3. Existence of solutions. To simplify notation, define

$$\mathbf{F}(\mathbf{t}, \mathbf{V}) = Eu_2(d^2(\mathbf{x}, \mathbf{t}; \mathbf{V}^{-1}))(\mathbf{x} - \mathbf{t})(\mathbf{x} - \mathbf{t})' , \tag{3.1}$$

so that (2.2) becomes

$$\mathbf{F}(\mathbf{t}, \mathbf{V}) = \mathbf{V} . \tag{3.2}$$

For each $\mathbf{t}$, $\mathbf{F}$ is a function of Θ_2 into Θ_2.

Equation (3.2) will be first treated separately. The following simple result is stated for later use.

LEMMA 1. *For each* $A \geqq 0$ *there exists* $r_0 > 0$ *such that, for all* $\mathbf{t} \in B_A$ *and all* $r \geqq r_0$, $\mathbf{F}(\mathbf{t}, r\mathbf{I}) \leqq r\mathbf{I}$.

PROOF. The result follows easily from

$$\lim_{r\to\infty} \sup_{|\mathbf{z}|=1} \sup_{|\mathbf{t}|\leqq A} r^{-1}\mathbf{z}'\mathbf{F}(\mathbf{t}, r\mathbf{I})\mathbf{z} = 0 ,$$

which follows from the boundedness of ψ_2.

THEOREM 1. *Let P satisfy* (E). *Then for each* $\mathbf{t}$ *there exists a unique solution* $\mathbf{V} = \mathbf{V}_P(\mathbf{t})$ *of the equation* $\mathbf{F}(\mathbf{t}, \mathbf{V}) = \mathbf{V}$.

PROOF. There is no loss of generality in supposing $\mathbf{t} = \mathbf{0}$. Throughout this proof $\mathbf{F}(\mathbf{V})$ will stand for $\mathbf{F}(\mathbf{0}, \mathbf{V})$.

If $\mathbf{U}$ is positive definite, so is $\mathbf{F}(\mathbf{U})$; otherwise there would exist $\mathbf{z} \neq \mathbf{0}$ such that $0 = \mathbf{z}'\mathbf{F}(\mathbf{U})\mathbf{z} = Eu_2(\mathbf{z}'\mathbf{U}^{-1}\mathbf{z})(\mathbf{z}'\mathbf{x})^2$, and since u_2 is positive by (C), this would entail $P(\mathbf{z}'\mathbf{x} = 0) = 1$, contradicting (E). Applying now Lemma 1 (with $A = 0$), there is a matrix $\mathbf{V}_0 = r_0\mathbf{I}$ such that $\mathbf{F}(\mathbf{V}_0) \leqq \mathbf{V}_0$. Define recursively $\mathbf{V}_{n+1} = \mathbf{F}(\mathbf{V}_n)$; this is legitimate since by the reasoning above each $\mathbf{V}_n$ is nonsingular.

Since u_2 is nonincreasing, $\mathbf{V}_{n+1} \leqq \mathbf{V}_n$ for all n, so that $\lim_{n\to\infty} \mathbf{V}_n$ exists. This limit, denoted by $\mathbf{V}$, will be the desired solution if it is proved to be nonsingular. To this end, a contradiction will be derived from the supposition that there exists $\mathbf{z} \neq 0$ such that $\mathbf{z}'\mathbf{V}\mathbf{z} = 0$.

For each $b > 0$, there exists by (D) an s_b such that $s \geqq s_b$ implies $\psi_2(s) \geqq K_2 - b$. Define $C_{bn} = \{\mathbf{x} \mid \mathbf{x}'\mathbf{V}_n^{-1}\mathbf{x} < s_b\}$ and $C_b = \bigcap_n C_{bn}$. Obviously $C_{b,n+1} \subseteq C_{b,n}$. Besides, C_b is contained in the hyperplane $H = \{\mathbf{x} \mid \mathbf{z}'\mathbf{x} = 0\}$. In effect let $\mathbf{x} \in C_b$; putting $\mathbf{V}_n = \mathbf{S}_n'\mathbf{S}_n$ we have $(\mathbf{z}'\mathbf{x})^2 = [(\mathbf{S}_n\mathbf{z})'((\mathbf{S}_n')^{-1}\mathbf{x})]^2 \leqq (\mathbf{z}'\mathbf{V}_n\mathbf{z})(\mathbf{x}'\mathbf{V}_n^{-1}\mathbf{x})$

for all n. The last factor is less than s_b, and the first one tends to zero, so that $\mathbf{x} \in H$.

Hence, given $c > 0$ there exists n_{bc} such that $n \geqq n_{bc}$ implies $P(C_{bn}^c) \geqq 1 - P(H) - c$. Besides, since ψ_2 is nondecreasing

$$\mathbf{V}_n \geqq \mathbf{F}(\mathbf{V}_n) \geqq EI(\mathbf{x} \in C_{bn}^c)[(K_2 - b)/(\mathbf{x}'\mathbf{V}_n^{-1}\mathbf{x})]\mathbf{x}\mathbf{x}' .$$

Multiplying this expression by $(\mathbf{S}_n')^{-1}$ on the left and by $\mathbf{S}_n^{-1}$ on the right, and taking the trace, it results for $n \geqq n_{bc}$

$$m \geqq (K_2 - b)P(C_{bn}^c) \geqq (K_2 - b)(1 - P(H) - c) ,$$

and this inequality being valid for all positive b and c, it follows that $m/K_2 \geqq 1 - P(H)$. This contradicts (E), thus proving the existence of a solution.

It may be now assumed—by a change of coordinates—that $\mathbf{I}$ is a solution. To prove uniqueness, it will be shown that in this case, $\mathbf{U} \neq \mathbf{I}$ implies $|\mathbf{F}(\mathbf{U}) - \mathbf{I}| < |\mathbf{U} - \mathbf{I}|$ if $\mathbf{U} \in \Theta_2$. To this end let $a_1 \leqq \cdots \leqq a_m$ and $b_1 \leqq \cdots \leqq b_m$ be respectively the eigenvalues of $\mathbf{U}$ and of $\mathbf{F}(\mathbf{U})$. It will suffice to show that $a_1 < 1$ entails $b_1 > a_1$, and that $a_m > 1$ entails $b_m < a_m$. Only the first implication will be proved, the proof of the second one being completely analogous.

Let $\mathbf{R} = E\psi_2(a_1^{-1}\mathbf{x}'\mathbf{x})(\mathbf{x}'\mathbf{x})^{-1}\mathbf{x}\mathbf{x}'$. Since $\mathbf{x}'\mathbf{U}^{-1}\mathbf{x} \leqq a_1^{-1}\mathbf{x}'\mathbf{x}$, then $\mathbf{F}(\mathbf{U}) \geqq a_1\mathbf{R}$: hence it suffices to prove that the smallest eigenvalue of $\mathbf{R}$ is larger than one, or equivalently, that $\mathbf{z}'\mathbf{R}\mathbf{z} > \mathbf{z}'\mathbf{I}\mathbf{z}$ for all $\mathbf{z} \neq \mathbf{0}$. Since $\mathbf{F}(\mathbf{I}) = \mathbf{I}$, this amounts to proving that

$$0 < \mathbf{z}'(\mathbf{R} - \mathbf{I})\mathbf{z} = E[\psi_2(a_1^{-1}\mathbf{x}'\mathbf{x}) - \psi_2(\mathbf{x}'\mathbf{x})](\mathbf{x}'\mathbf{x})^{-1}(\mathbf{z}'\mathbf{x})^2 .$$

The integrand is nonnegative, since $a_1 < 1$ and ψ_2 is nondecreasing. Suppose that the last member of the inequality vanished for some $\mathbf{z} \neq \mathbf{0}$. By (C), $\psi_2(a_1^{-1}\mathbf{x}'\mathbf{x}) - \psi_2(\mathbf{x}'\mathbf{x}) = 0$ implies $\psi_2(\mathbf{x}'\mathbf{x}) = K_2$ if $\mathbf{x} \neq \mathbf{0}$. Hence $P(\psi_2(\mathbf{x}'\mathbf{x}) = K_2) + P((\mathbf{z}'\mathbf{x}) = 0) = 1$. The second summand is less than $1 - m/K_2$ by (E). Taking the trace in equation $\mathbf{F}(\mathbf{I}) = \mathbf{I}$, it follows that the first one is $\leqq m/K_2$. This contradiction completes the theorem.

LEMMA 2. *If P satisfies* (E), *then the function $\mathbf{V}(\mathbf{t}) = \mathbf{V}_P(\mathbf{t})$ has the following properties*:

(i) *There exists A_1 such that for all $\mathbf{t}$*: $|\mathbf{V}(\mathbf{t})^{-1}| \leqq A_1^2$.

(ii) *There exists A_2 such that for all $\mathbf{t}$*: $\mathbf{t}'\mathbf{V}(\mathbf{t})^{-1}\mathbf{t} \leqq A_2^2$.

(iii) $\limsup_{|\mathbf{t}|\to\infty} |\mathbf{t}|^{-2}|\mathbf{V}(\mathbf{t})| = A_3 < \infty$.

(iv) $\mathbf{V}(\mathbf{t})$ *is a continuous function of R^m into Θ_2 (taking in Θ_2 the topology induced by $R^{m\times m}$).*

PROOF. (i) From (E) it is easy to deduce the existence of $c > 0$ such that

$$\sup_{\mathbf{t}} \sup_{|\mathbf{z}|=1} P\{\mathbf{x} \mid |\mathbf{z}'(\mathbf{x} - \mathbf{t})| \leqq c\} \leqq 1 - m/K_2 - a/2 , \tag{3.3}$$

where a is the constant appearing in (E). There exists $b > 0$ such that $(K_2 - b)(m/K_2 + a/2) > m$; and by (D) there exists $A_1 > 0$ such that $\psi_2[(cA_1)^2] > K_2 - b$. For each $\mathbf{t}$ let $e = e(\mathbf{t})$ be the smallest eigenvalue of $\mathbf{V} = \mathbf{V}(\mathbf{t})$, and

$\mathbf{z} = \mathbf{z}(\mathbf{t})$ be the corresponding eigenvector with unit norm. It will be shown that $1/e \leqq A_1^2$, and this will prove (i) since $1/e = |\mathbf{V}^{-1}|$.

Suppose by the contrary that $1/e > A_1^2$ for some $\mathbf{t}$; taking the trace in (2.5) yields

$$m = E\psi_2(d^2(\mathbf{x}, \mathbf{t}; \mathbf{V}^{-1})) ; \tag{3.4}$$

and since $d^2(\mathbf{x}, \mathbf{t}; \mathbf{V}^{-1}) \geqq e^{-1}|\mathbf{z}'(\mathbf{x} - \mathbf{t})|^2$, and ψ_2 is nondecreasing, it follows that

$$m \geqq P(|\mathbf{z}'(\mathbf{x} - \mathbf{t})| > c)\psi_2[(cA_1)^2] > (m/K_2 + a/2)(K_2 - b) > m .$$

(ii) Let $\mathbf{S} = \mathbf{S}(\mathbf{t})$ be any matrix such that $\mathbf{V}(\mathbf{t}) = \mathbf{S}\mathbf{S}'$, so that (i) is equivalent to $|\mathbf{S}(\mathbf{t})^{-1}| \leqq A_1$, and (3.4) may be written as

$$m = E\psi_2(|\mathbf{S}^{-1}\mathbf{x} - \mathbf{S}^{-1}\mathbf{t}|^2) . \tag{3.4'}$$

Since for each $\mathbf{x}$

$$\lim_{|\mathbf{s}|\to\infty} \inf_{|\mathbf{M}|\leqq A_1} \psi_2(|\mathbf{M}\mathbf{x} - \mathbf{s}|^2) = K_2 ,$$

then the Dominated Convergence Theorem and $K_2 > m$ yield the existence of positive b and A_2 such that $|\mathbf{s}| \geqq A_2$ implies

$$E \inf_{|\mathbf{M}|\leqq A_1} \psi_2(|\mathbf{M}\mathbf{x} - \mathbf{s}|^2) \geqq m + 2b . \tag{3.5}$$

Hence if some $\mathbf{t}$ satisfied $\mathbf{t}'\mathbf{V}(\mathbf{t})^{-1}\mathbf{t} = |\mathbf{S}(\mathbf{t})^{-1}\mathbf{t}|^2 > A_2$, we would obtain a contradiction with (3.4′) by putting in (3.5): $\mathbf{M} = \mathbf{S}^{-1}$ and $\mathbf{s} = \mathbf{S}^{-1}\mathbf{t}$.

(iii) Let now $e = e(\mathbf{t})$ be the largest eigenvalue of $\mathbf{V}(\mathbf{t})(= |\mathbf{V}|)$ and $\mathbf{z} = \mathbf{z}(\mathbf{t})$ the corresponding unit eigenvector, so that for all $\mathbf{w}$: $\mathbf{w}'\mathbf{V}^{-1}\mathbf{w} \geqq e^{-1}|\mathbf{w}|^2$. Applying this in (2.2), since u_2 is nonincreasing it follows that

$$\mathbf{V} \leqq u_2(e^{-1}|\mathbf{x} - \mathbf{t}|^2)(\mathbf{x} - \mathbf{t})(\mathbf{x} - \mathbf{t})' .$$

Multiplying on the left and on the right respectively by $\mathbf{z}'$ and by $\mathbf{z}$, dividing by e, and applying Cauchy–Schwarz, it results that

$$1 = e^{-1}\mathbf{z}'\mathbf{V}\mathbf{z} \leqq E\psi_2(e^{-1}|\mathbf{x} - \mathbf{t}|^2) . \tag{3.6}$$

Now (ii) implies that $A_2^2 \geqq |\mathbf{t}|^2/e(\mathbf{t})$, so that $\liminf_{|\mathbf{t}|\to\infty} e(\mathbf{t}) = \infty$; hence, for (3.6) to hold, it is necessary that $\liminf_{|\mathbf{t}|\to\infty} |\mathbf{t}|^2/e(\mathbf{t}) > 0$, which is equivalent to the thesis.

(iv) When $\mathbf{t}$ ranges on a compact set of R^m, $|\mathbf{V}(\mathbf{t})|$ remains bounded. In effect, given $A > 0$, there exists by Lemma 1 an r such that $\mathbf{F}(\mathbf{t}, r\mathbf{I}) \leqq r\mathbf{I}$ for all $\mathbf{t} \in B_A$; proceeding recursively as at the beginning of the proof of Theorem 1, for all $\mathbf{t} \in B_A$ we obtain that $\mathbf{V}(\mathbf{t}) \leqq r\mathbf{I}$, and a fortiori $|\mathbf{V}(\mathbf{t})| \leqq r$, as asserted.

Besides, (i) entails that $|\mathbf{V}(\mathbf{t})^{-1}|$ is also bounded. Hence the image under $\mathbf{V}(\mathbf{t})$ of a compact of Θ_1 is contained in a compact of Θ_2. Since $\mathbf{F}(\mathbf{t}, \mathbf{V})$ is a continuous function of Θ into Θ_2, and $\mathbf{F}(\mathbf{t}, \mathbf{V}(\mathbf{t})) = \mathbf{I}$ for all $\mathbf{t}$, the continuity of $\mathbf{V}(\mathbf{t})$ results from a simple topological argument. This completes the proof of the theorem.

THEOREM 2. *If P satisfies* (E), *then there exists a solution* $(\mathbf{t}_0, \mathbf{V}_0)$ *of the system* (2.1)—(2.2). *Besides*, $\mathbf{t}_0$ *belongs to the convex hull of the support of* P.

PROOF. In view of Theorem 1, it must be proved that there exists $\mathbf{t}_0$ such that

$$Eu_1[d(\mathbf{x}, \mathbf{t}_0; \mathbf{V}(\mathbf{t}_0)^{-1})](\mathbf{x} - \mathbf{t}_0) = \mathbf{0} , \tag{3.7}$$

which yields $\mathbf{V}_0 = \mathbf{V}_P(\mathbf{t}_0)$. Let $\mathbf{S}(\mathbf{t})$ be the square root of $\mathbf{V}(\mathbf{t})$ in Θ_2 which, by Lemma 2 (iv), is a continuous function of $\mathbf{t}$. To simplify notation write $\mathbf{h}(\mathbf{x}) = u_1(|\mathbf{x}|)\mathbf{x}$, and define the function $\mathbf{g}$ of R^m into R^m:

$$\mathbf{g}(t) = \mathbf{g}_c(t) = \mathbf{t} + c(|\mathbf{t}| + 1)E\mathbf{h}[\mathbf{S}(\mathbf{t})^{-1}(\mathbf{x} - \mathbf{t})]$$

where c is a constant to be later conveniently determined. It is clear that, if $c \neq 0$, solutions of (3.7) coincide with fixed points of $\mathbf{g}$. We shall now prove that for some $c \neq 0$ there exists a closed ball B such that $\mathbf{g}(B) \subseteqq B$. Since $\mathbf{g}$ is obviously continuous, Brower's fixed point theorem (Dunford–Schwarz, 1958) entails the existence of a fixed point of $\mathbf{g}$, as desired.

Note in the first place that $\lim_{|\mathbf{t}|\to\infty} |\mathbf{t}|^{-1}\mathbf{t}'\mathbf{S}(\mathbf{t})^{-1}\mathbf{t} = 0$ by Lemma 2 (ii); and that Lemma 2 (iii) entails

$$\liminf_{|\mathbf{t}|\to\infty} |\mathbf{t}|^{-1}\mathbf{t}'\mathbf{S}(\mathbf{t})^{-1}\mathbf{t} \geqq \liminf_{|\mathbf{t}|\to\infty} |\mathbf{t}||\mathbf{S}(\mathbf{t})|^{-1} = \delta > 0 .$$

Hence

$$\begin{aligned} &\limsup_{|\mathbf{t}|\to\infty} |\mathbf{t}|^{-1}\mathbf{t}'E\mathbf{h}[\mathbf{S}(\mathbf{t})^{-1}(\mathbf{x} - \mathbf{t})] \\ &\qquad < -\delta \limsup_{|\mathbf{t}|\to\infty} Eu_1(|\mathbf{S}(\mathbf{t})^{-1}(\mathbf{x} - \mathbf{t})|) < -\delta p u_1(s_0) , \end{aligned} \tag{3.8}$$

where s_0 is the constant in (D), and $p = 1 - m/\psi_2(s_0^2)$, which by (3.4′) verifies $P(|\mathbf{S}(\mathbf{t})^{-1}(\mathbf{x} - \mathbf{t})| \leqq s_0) \geqq p$ for all $\mathbf{t}$.

Summing up, since $|\mathbf{h}(\mathbf{x})| \leqq K_1$ for all $\mathbf{x}$, then for each c

$$\limsup_{|\mathbf{t}|\to\infty} ||\mathbf{t}|^{-1}\mathbf{t} + cE\mathbf{h}[\mathbf{S}(\mathbf{t})^{-1}(\mathbf{x} - \mathbf{t})]|^2 \leqq 1 + c^2K_1^2 - 2[\delta p u_1(s_0)]c . \tag{3.9}$$

Since $u_1(s_0) > 0$ by (D), taking c positive near 0, the second member of (3.9) becomes < 1; hence there exist $d > 0$ and $c > 0$ such that

$$\limsup_{|\mathbf{t}|\to\infty} ||\mathbf{t}|^{-1}\mathbf{t} + cE\mathbf{h}[\mathbf{S}(\mathbf{t})^{-1}(\mathbf{x} - \mathbf{t})]| \leqq 1 - d . \tag{3.10}$$

Hence there exists A such that $|\mathbf{t}| \leqq A$ implies $|\mathbf{g}(\mathbf{t})| \leqq A$. Otherwise there would exist a sequence $\{\mathbf{t}_n\}$ with $|\mathbf{t}_n| \leqq n$, such that $|\mathbf{g}(\mathbf{t}_n)| > n$ for all n. As $\mathbf{g}$ is continuous, this would imply $\limsup_{n\to\infty} |\mathbf{t}_n| = \infty$, and besides $|\mathbf{t}_n|^{-1}|\mathbf{g}(\mathbf{t}_n)| > 1$, which contradicts (3.10). Hence, for some $A > 0$ we must have $\mathbf{g}(B_A) \subseteqq B_A$, thus completing the proof of the existence.

To prove the second assertion of the theorem, let C be the convex hull of the support of P, and $b = \inf_{\mathbf{x}\in C} d^2(\mathbf{x}, \mathbf{t}_0; \mathbf{V}_0^{-1})$; so that (3.4) entails $m > \psi_2(b)$. If $\mathbf{t}_0 \notin C$, there exist a hyperplane H separating $\mathbf{t}_0$ and C, a vector $\mathbf{z}$ orthogonal to H, and a positive c such that $\mathbf{z}'(\mathbf{x} - \mathbf{t}) \geqq c$ for all $\mathbf{x} \in C$. Multiplication of (2.1) by $\mathbf{z}'$ yields $0 \geqq cEu_1(d(\mathbf{x}, \mathbf{t}_0; \mathbf{V}_0^{-1}))$, and by (D) we have $P(d(\mathbf{x}, \mathbf{t}_0; \mathbf{V}_0^{-1}) > s_0) = 1$, which implies that $\psi_2(b) > m$. This finishes the proof of the theorem.

4. Uniqueness. It is easy to prove uniqueness of solutions for the equations defining the model parameters, i.e. when P is the underlying distribution. Greater generality for the distribution is obtained by stronger restrictions on ψ_1.

THEOREM 3. *Either of the following two hypotheses suffices for the uniqueness of the solution of* (2.1)—(2.2):

(i) *The distribution P has a density $f(\mathbf{x})$ which is a decreasing function of $|\mathbf{x}|$.*

(ii) *The distribution P is symmetric* (i.e., $P(\mathbf{x} \in A) = P(-\mathbf{x} \in A)$ *for all Borel sets A*) *and satisfies* (E); *ψ_1 is nondecreasing, and $m > 1$.*

PROOF. In both cases, remember that $\mathbf{t}_0 = \mathbf{0}$ is a solution of (2.1) for all $\mathbf{V}$, and that for each $\mathbf{t}$, (2.2) has a unique solution $\mathbf{V}(\mathbf{t})$. Hence to prove (i) it suffices to show that for all $\mathbf{t} \neq 0$ and all nonsingular $\mathbf{S}$

$$\mathbf{t}' \int_{R^m} u_1(|\mathbf{S}^{-1}(\mathbf{x} - \mathbf{t})|)(\mathbf{x} - \mathbf{t}) f(\mathbf{x})\, d\mathbf{x} < 0\,. \tag{4.1}$$

Let $A = \{\mathbf{x} \mid \mathbf{t}'(\mathbf{x} - \mathbf{t}) > 0\}$. Splitting the above integral as $\int_A + \int_{A^c}$, and applying in the latter the change of variables $\mathbf{y} = -(\mathbf{x} - \mathbf{t}) + \mathbf{t}$, we see that the expression in (4.1) is equal to

$$\int_A \mathbf{t}'(\mathbf{x} - \mathbf{t}) u_1(|\mathbf{S}^{-1}(\mathbf{x} - \mathbf{t})|)(f(\mathbf{x}) - f(2\mathbf{t} - \mathbf{x}))\, d\mathbf{x};$$

and this is negative, since $|\mathbf{x}| > |2\mathbf{t} - \mathbf{x}|$ for all $\mathbf{x} \in A$.

Since the symmetry of a distribution is conserved under linear transformations, to prove (ii) it suffices to show that $E\mathbf{t}'[u_1(|\mathbf{x} - \mathbf{t}|)(\mathbf{x} - \mathbf{t})] < 0$ for all $\mathbf{t} \neq \mathbf{0}$. The Cauchy–Schwarz inequality and the fact that ψ_1 is nondecreasing, yield $(\mathbf{a} - \mathbf{b})'[u_1(|\mathbf{a}|)\mathbf{a} - u_1(|\mathbf{b}|)\mathbf{b}] \geqq 0$ for all $\mathbf{a}$ and $\mathbf{b}$, with equality holding only if $\mathbf{a}'\mathbf{b} = |\mathbf{a}||\mathbf{b}|$, i.e., if $\mathbf{a}$ and $\mathbf{b}$ are linearly dependent. Hence, taking $\mathbf{a} = \mathbf{x} - \mathbf{t}$ and $\mathbf{b} = \mathbf{x}$, the integrand in the expectation above is negative; and if the expectation vanishes, $\mathbf{x}$ must be linearly dependent with $\mathbf{x} - \mathbf{t}$ with probability one, which contradicts (E) if $m > 1$. This finishes the proof.

When P is the empirical distribution, the fact that P is atomic seems to be of no help in proving the uniqueness for the equations defining the estimators, so that it is necessary to prove it for arbitrary P. Naturally this will require stronger conditions on the ψ_i's. An insight on the type of conditions required may be obtained from the univariate case.

THEOREM 4. *Let $\chi(s) = d\psi_2(s^2)/ds$. The following set of conditions* (F) *is sufficient for the uniqueness of the solution of* (2.1)—(2.2) *if P satisfies* (E) *and $m = 1$:*

(F1) *$\psi_1'(s) \geqq 0$ for all $s \geqq 0$.*

(F2) *Let $s_1 = \sup\{s \mid \psi_1'(s) > 0\}$. Then $\chi(s) = 0$ if $s \geqq s_1$.*

(F3) *There exists $s_2 < s_1$ such that $\psi_2(s_2^2) > m$, and such that $\chi(s)/\psi_1'(s)$ is increasing on $[0, s_2]$ and nondecreasing on $[0, s_1]$.*

The proof is not difficult and follows along the same lines as the uniqueness proof in page 98 of Huber (1964). Details may be found in Maronna (1974).

Assumptions (F2) and (F3) are obviously valid if $\psi_2(s^2) = C(\psi_1(s))^2$, which is the case of Proposal 2. I conjecture that Theorem 4 is also valid for $m > 1$, but I have not been able to obtain a rigorous proof.

5. Consistency. Let P be a distribution in R^m satisfying (E), for which (2.1)—(2.2) have a unique solution $(\mathbf{t}, \mathbf{V})$, and let $\mathbf{x}_1, \cdots, \mathbf{x}_n, \cdots$ be independent variables with common distribution P. For each n let P_n be the empirical distribution corresponding to the sample $\{\mathbf{x}_1, \cdots, \mathbf{x}_n\}$, and $(\mathbf{t}_n, \mathbf{V}_n)$ a solution of (2.1)—(2.2) when the measure is P_n.

THEOREM 5. $\lim_{n\to\infty} (\mathbf{t}_n, \mathbf{V}_n) = (\mathbf{t}, \mathbf{V})$ *with probability one.*

PROOF. The generalization given by Wolfowitz (1954) of the Glivenko-Cantelli theorem establishes that for half-spaces $S \subseteqq R^m$, $\lim_{n\to\infty} P_n(S) = P(S)$ uniformly in S, with probability one. It follows easily that with probability one, P_n satisfies (E) for large n, so that by Theorem 2 of Section 3, the estimators $(\mathbf{t}_n, \mathbf{V}_n)$ indeed exist.

Since ψ_1 and ψ_2 are continuous and bounded, it is easy to verify that the estimators satisfy conditions (B-1), (B-2′) and (B-3) of Section 3 of Huber (1967), so that according to Huber's Theorem 2, to prove consistency it suffices to show the existence of a compact set $K \subseteqq \Theta$ such that with probability one the sequence $(\mathbf{t}_n, \mathbf{V}_n)$ ultimately stays in K. Unfortunately, Huber's condition (B-4), which would entail the desired result, does not hold for the simultaneous estimation of location and scale, so that a direct proof will be necessary. It will suffice to prove the existence of finite constants A, B, C such that with probability one

$$\text{(5.1)} \qquad \limsup_{n\to\infty} |\mathbf{V}_n^{-1}| \leqq A\,; \qquad \limsup_{n\to\infty} |\mathbf{V}_n| \leqq B\,; \qquad \limsup_{n\to\infty} |\mathbf{t}_n| \leqq C\,.$$

To prove the first inequality in (5.1) note that, in the notation of Section 3, $\mathbf{V}_n = \mathbf{V}_{P_n}(\mathbf{t}_n)$. It is easy to see that the bound A_1 in Lemma 2(i) of Section 3 depends upon the distribution P only through the constant c of (3.3). Now it follows from Wolfowitz's theorem that (3.3) holds also for P_n, i.e., there exists c' such that with probability one:

$$\limsup_{n\to\infty} \sup_{\mathbf{t}} \sup_{|\mathbf{z}|=1} P_n\{\mathbf{x} \mid |\mathbf{z}'(\mathbf{x}-\mathbf{t})| \leqq c'\} \leqq 1 - m/K_2 - a/2\,.$$

Hence proceeding as in Lemma 2(i) we verify the existence of A_1' such that, with probability 1, $\limsup_{n\to\infty} \sup_{\mathbf{t}} |\mathbf{V}_{P_n}(\mathbf{t})^{-1}| \leqq A_1'$, and hence $\limsup_{n\to\infty} |\mathbf{V}_n^{-1}| \leqq A_1'$ a.s., as stated.

In the same manner, one proves that there exists A_2' such that with probability one

$$\text{(5.2)} \qquad \limsup_{n\to\infty} \mathbf{t}_n'\mathbf{V}_n^{-1}\mathbf{t}_n \leqq (A_2')^2\,.$$

In effect, the bound A_2 in Lemma 2(ii) depends on P only through A_1 (which in turn depends only on c) and through the expectation in (3.5). Wolfowitz's theorem entails that for each ε there exists a compact Q such that $\liminf_{n\to\infty} P_n(\mathbf{Q}) \geqq 1 - \varepsilon$ a.s., and hence (3.5) holds also for P_n. That is, there exist A_2' and $b' > 0$ such that $|\mathbf{s}| \geqq A_2'$ implies that with probability one

$$\liminf_{n\to\infty} E_{P_n} \inf_{|\mathbf{M}|\leqq A_1'} \psi_2(|\mathbf{Mx} - \mathbf{s}|^2) \geqq m + 2b'\,,$$

so that again proceeding as in Lemma 2(ii) one proves that

$$\limsup_{n\to\infty} \sup_t \mathbf{t}'\mathbf{V}_{P_n}(\mathbf{t})^{-1}\mathbf{t} \leqq (A_2')^2 \quad \text{a.s.},$$

which implies (5.2).

To prove the theorem it will now suffice to prove the second inequality in (5.1), since it, together with (5.2), entail the third one. Let $\mathbf{S}_n$ be the square root of $\mathbf{V}_n$ in Θ_2, $\mathbf{z}_n$ the unit eigenvector corresponding to the largest eigenvalue of $\mathbf{V}_n$, and $b_n = |\mathbf{V}_n|^{-\frac{1}{2}}$. We shall equivalently prove the existence of $c > 0$ such that $\liminf_{n\to\infty} b_n \geqq c$ a.s. In the rest of the proof E_n will stand for E_{P_n}, and the subscript n will be generally omitted, except from E_n and P_n.

Write for simplicity $\mathbf{w} = \mathbf{S}^{-1}\mathbf{t}$ and $q = q(\mathbf{x}) = |\mathbf{S}^{-1}\mathbf{x} - \mathbf{w}|$. Multiplying (2.5) by $\mathbf{z}'$ and by $\mathbf{z}$, and recalling that $\mathbf{S}\mathbf{z} = \mathbf{z}/b$, we obtain

$$1 = E_n u_2(q^2)(b\mathbf{z}'\mathbf{x} - \mathbf{z}'\mathbf{w})^2 . \tag{5.3}$$

Let $\delta \in (0, 1)$ to be later conveniently determined, and take $\varepsilon < \delta/[K_1 + K_2 + u_2(0)A_2'^2 + u_1(0)A_1']$. Wolfowitz's theorem entails the existence of an r such that $\limsup_{n\to\infty} P_n(B_r^c) < \varepsilon$ a.s., and hence for large n, (5.3) implies that a.s.

$$|1 - E_n I(B_r)u_2(q^2)(b\mathbf{z}'\mathbf{x} - \mathbf{z}'\mathbf{w})^2| \leqq K_2\varepsilon < \delta ; \tag{5.4}$$

and since (5.2) implies $\limsup_n |\mathbf{w}| \leqq A_2'$ a.s.,

$$E_n I(B_r^c)u_2(q^2)(\mathbf{z}'\mathbf{w})^2 \leqq \varepsilon u_2(0)(A_2')^2 < \delta . \tag{5.5}$$

Hence subtracting and adding (5.3) in the modulus below and applying (5.4)—(5.5) it follows that

$$|1 - (\mathbf{z}'\mathbf{w})^2 E_n u_2(q^2)| < 2\delta + u_2(0)br(br + 2A_2') . \tag{5.6}$$

Choose c_0 such that for $b < c_0$ the last summand in (5.6) is less than δ, and $c = \min(c_0, \delta/r)$. A contradiction will be derived from the assumption that $\liminf_{n\to\infty} b < c$ with positive probability. Passing over to a subsequence $\{n'\}$ such that $b_{n'} < c$, (5.6) entails that $\limsup_{n\to\infty} |1 - (\mathbf{z}'\mathbf{w})^2 E_n u_2(q^2)| < 3\delta$ with positive probability, and hence taking $\delta \leqq \frac{1}{6}$ one obtains

$$\tfrac{1}{2} < (\mathbf{z}'\mathbf{w})^2 E_n u_2(q^2) \leqq (\mathbf{z}'\mathbf{w})^2 u_2(0) . \tag{5.7}$$

Hence since u_1 is nonincreasing

$$|(\mathbf{z}'\mathbf{w})E_n u_1(q)| \geqq p u_1(s_0)/(2u_2(0))^{\frac{1}{2}} , \tag{5.8}$$

where s_0 is the constant in (D) and $p = 1 - m/\psi_2(s_0^2)$, which by (3.4′) verifies $P_n(q \leqq s_0) \geqq p$ for all n.

At the same time, multiplying (2.4) by $\mathbf{z}'$ we obtain the equation $a_1 + a_2 - a_3 = 0$, where

$$a_1 = E_n I(B_r^c)u_1(q)(b\mathbf{z}'\mathbf{x} - \mathbf{z}'\mathbf{w}) ,$$
$$a_2 = E_n I(B_r)u_1(q)b\mathbf{z}'\mathbf{x} ,$$
$$a_3 = E_n I(B_r)u_1(q)\mathbf{z}'\mathbf{w} .$$

From their respective definitions it follows that $|a_1| \leqq \varepsilon K_1 < \delta$ and $|a_2| \leqq br < \delta$, so that $|a_3| \leqq 2\delta$. Hence subtracting and adding a_3 in the modulus below, and recalling (5.2) it results that

$$\limsup_{n\to\infty} |E_n u_1(q)\mathbf{z}'\mathbf{w}| \leqq \varepsilon|\mathbf{z}'\mathbf{w}| + |a_3| \leqq 3\delta\,; \tag{5.9}$$

so that taking, for example, δ equal to one-fourth of the right member of (5.8) we obtain a contradiction between (5.9) and (5.8). This finishes the proof.

6. Asymptotic normality. The proof of the asymptotic normality of the estimate $(\mathbf{t}_n, \mathbf{V}_n)$ will be based on a very general result on M-estimators given in Section 4 of Huber (1967). The notation of this section will be chosen to match that of Huber's paper. We shall generically write $\boldsymbol{\theta} = (\mathbf{t}, \mathbf{V})$. Let Θ_3 be the set of $m \times m$-symmetric matrices, and $\Theta^0 = \Theta_1 \times \Theta_3$. The vector space Θ^0 will be normed with $|\boldsymbol{\theta}| = \max(|\mathbf{t}|, |\mathbf{V}|)$. Let $\boldsymbol{\Psi}$ be the function of $R^m \times \Theta$ into Θ: $\boldsymbol{\Psi}(\mathbf{x}, \boldsymbol{\theta}) = (\boldsymbol{\Psi}_1(\mathbf{x}, \boldsymbol{\theta}), \boldsymbol{\Psi}_2(\mathbf{x}, \boldsymbol{\theta}))$ defined by

$$\boldsymbol{\Psi}_1(\mathbf{x}, \boldsymbol{\theta}) = u_1(d(\mathbf{x}, \mathbf{t}; \mathbf{V}^{-1}))(\mathbf{x} - \mathbf{t})$$
$$\boldsymbol{\Psi}_2(\mathbf{x}, \boldsymbol{\theta}) = u_2(d^2(\mathbf{x}, \mathbf{t}; \mathbf{V}^{-1}))(\mathbf{x} - \mathbf{t})(\mathbf{x} - \mathbf{t})'\,.$$

Let P be the underlying distribution and $\boldsymbol{\lambda}(\boldsymbol{\theta}) = (\boldsymbol{\lambda}_1(\boldsymbol{\theta}), \boldsymbol{\lambda}_2(\boldsymbol{\theta})) = E_P \boldsymbol{\Psi}(\mathbf{x}, \boldsymbol{\theta})$, so that the "true" parameter $\boldsymbol{\theta}_0$ to be estimated is defined by $\boldsymbol{\lambda}(\boldsymbol{\theta}_0) = \mathbf{0}$, and if P_n is the empirical distribution, the estimator $\boldsymbol{\theta}_n{}^* = (\mathbf{t}_n, \mathbf{V}_n)$ is defined by $E_{P_n} \boldsymbol{\Psi}(\mathbf{x}, \boldsymbol{\theta}^*) = \mathbf{0}$. For $j = 1, 2$ define

$$U_j(\mathbf{x}, \boldsymbol{\theta}, \delta) = \sup_{|\boldsymbol{\theta}_1 - \boldsymbol{\theta}| < \delta} |\boldsymbol{\Psi}_j(\mathbf{x}, \boldsymbol{\theta}_1) - \boldsymbol{\Psi}_j(\mathbf{x}, \boldsymbol{\theta})|\,.$$

According to Huber's Theorem 3 and its corollary, if there exist positive numbers b, c and δ_0 such that $EU_j(\mathbf{x}, \boldsymbol{\theta}, \delta) \leqq b\delta$ and $EU_j{}^2(\mathbf{x}, \boldsymbol{\theta}, \delta) \leqq c\delta$ for $|\boldsymbol{\theta} - \boldsymbol{\theta}_0| + \delta \leqq \delta_0$ $(j = 1, 2)$, and if the derivative $(D\boldsymbol{\lambda})_{\boldsymbol{\theta}_0}$ is nonsingular, then the distribution of $n^{\frac{1}{2}}(\boldsymbol{\theta}_n - \boldsymbol{\theta}_0)$ tends to a normal law with zero mean, and covariance matrix $(D\boldsymbol{\lambda})_{\boldsymbol{\theta}_0}^{-1}\mathbf{C}((D\boldsymbol{\lambda})_{\boldsymbol{\theta}_0}^{-1})'$, where $\mathbf{C}$ is the covariance matrix of the variable $\boldsymbol{\Psi}(\mathbf{x}, \boldsymbol{\theta}_0)$. We shall assume that P is of radial type, i.e., is obtained from a radial distribution by an affine transformation; although the results are probably true for symmetric distributions.

THEOREM 6. *If the functions $s\psi_j{}'(s)$ are bounded $(j = 1, 2)$, and P is a distribution of radial type such that* (2.1)—(2.2) *have a unique solution $\boldsymbol{\theta}_0 = (\mathbf{t}_0, \mathbf{V}_0)$ and such that $E_P \psi_1{}'(d(\mathbf{x}, \mathbf{t}_0; \mathbf{V}_0{}^{-1})) > 0$, then $n^{\frac{1}{2}}(\boldsymbol{\theta}_n - \boldsymbol{\theta}_0)$ has a limit normal distribution with zero means, $\mathbf{t}_n$ and $\mathbf{V}_n$ are asymptotically independent, and the covariance matrix of the asymptotic distribution of $\mathbf{t}_n$ is given by $(a/b^2)\mathbf{V}_0$, where*

$$a = m^{-1}E\psi_1{}^2(d(\mathbf{x}, \mathbf{t}_0; \mathbf{V}_0{}^{-1})$$
$$b = E[u_1(d(\mathbf{x}, \mathbf{t}; \mathbf{V}_0{}^{-1})(1 - m^{-1}) + \psi_1{}'(d(\mathbf{x}, \mathbf{t}_0; \mathbf{V}_0{}^{-1})m^{-1}]\,.$$

PROOF. We must prove that the hypotheses of Huber's theorem hold. It will be first shown that if K is a compact in Θ, then there exists $b > 0$ such that $U_j(\mathbf{x}, \boldsymbol{\theta}, \delta) \leqq b\delta$ $(j = 1, 2)$ for all $\boldsymbol{\theta} \in K$, all $\mathbf{x} \in R^m$ and δ sufficiently small; this will obviously imply the desired conditions on U_1 and U_2. We begin by calculating the derivatives of $\boldsymbol{\Psi}_1$ and $\boldsymbol{\Psi}_2$ at some $\boldsymbol{\theta} = (\mathbf{t}, \mathbf{V})$, applied to a generic element $(\mathbf{h}, \mathbf{M}) \in \Theta^\circ$. Application of the chain rule yields (writing for simplicity $d = d(\mathbf{x}, \mathbf{t}; \mathbf{V}^{-1})$)

$$(6.1)\qquad (D\mathbf{\Psi}_1)_\theta(\mathbf{h}, \mathbf{M}) = -u_1(d)\mathbf{h} - u_1'(d)/(2d)[2\mathbf{h}'\mathbf{V}^{-1}(\mathbf{x}-\mathbf{t}) + (\mathbf{x}-\mathbf{t})'\mathbf{V}^{-1}\mathbf{M}\mathbf{V}^{-1}(\mathbf{x}-\mathbf{t})](\mathbf{x}-\mathbf{t}),$$

$$(6.2)\qquad (D\mathbf{\Psi}_2)_\theta(\mathbf{h}, \mathbf{M}) = -u_2(d^2)[\mathbf{h}(\mathbf{x}-\mathbf{t})' + (\mathbf{x}-\mathbf{t})\mathbf{h}'] - \mathbf{M} - u_2'(d^2)[2\mathbf{h}'\mathbf{V}^{-1}(\mathbf{x}-\mathbf{t}) + (\mathbf{x}-\mathbf{t})'\mathbf{V}^{-1}\mathbf{M}\mathbf{V}^{-1}(\mathbf{x}-\mathbf{t})](\mathbf{x}-\mathbf{t})(\mathbf{x}-\mathbf{t})'.$$

Since the functions $s\psi_j'(s)$ and $\psi_j(s)/s$ are bounded, so are the $\psi_j'(s)$ $(j = 1, 2)$. Recalling that $d^{-1}|\mathbf{x} - \mathbf{t}| \leqq |\mathbf{V}|^{\frac{1}{2}}$, it is easy to verify that there exist constants A and B depending only upon $|\mathbf{V}|$ and $|\mathbf{V}^{-1}|$, such that for all $(\mathbf{h}, \mathbf{M})$, $|(D\mathbf{\Psi}_j)_\theta(\mathbf{h}, \mathbf{M})| \leqq A|\mathbf{h}| + B|\mathbf{M}|$ $(j = 1, 2)$; and the Mean Value Theorem entails the desired result on the U_j's.

It may be now assumed without loss of generality that P is spherically symmetric and $\boldsymbol{\theta}_0 = (\mathbf{0}, \mathbf{I})$. From (6.1)—(6.2) and from $u_j'(s) = s^{-1}(\psi_j'(s) - u_j(s))$ it follows that

$$(6.3)\qquad E(D\mathbf{\Psi}_1)_{\theta_0}(\mathbf{h}, \mathbf{M}) = -[Eu_1(|\mathbf{x}|)(\mathbf{h} - \mathbf{h}'\mathbf{x}|\mathbf{x}|^{-2}\mathbf{x}) + E\psi_1'(|\mathbf{x}|)\mathbf{h}'\mathbf{x}|\mathbf{x}|^{-2}\mathbf{x},$$

$$(6.4)\qquad E(D\mathbf{\Psi}_2)_{\theta_0}(\mathbf{h}, \mathbf{M}) = -[Eu_2'(|\mathbf{x}|^2)\mathbf{x}'\mathbf{M}\mathbf{x}\,\mathbf{x}\mathbf{x}' + \mathbf{M}].$$

It will be shown that $(D\boldsymbol{\lambda})_{\theta_0}$ is nonsingular. It follows from (6.3)—(6.4) that the "partial derivatives" $\partial\boldsymbol{\lambda}_1/\partial\mathbf{V}$ and $\partial\boldsymbol{\lambda}_2/\partial\mathbf{t}$ vanish at $\boldsymbol{\theta}_0$, so that it suffices to verify that the operators $\mathbf{D}_{1t} = (\partial\boldsymbol{\lambda}_1/\partial\mathbf{t})_{\theta_0}$ and $\mathbf{D}_{2V} = (\partial\boldsymbol{\lambda}_2/\partial\mathbf{V})_{\theta_0}$ are nonsingular.

Since P is radial, $|\mathbf{x}|$ and $\mathbf{z} = \mathbf{x}/|\mathbf{x}|$ are independent, and the latter is distributed uniformly on the unit spherical surface. Multiplying (6.3) by $\mathbf{h}'$, and taking into account that $E(\mathbf{h}'\mathbf{z})^2 = |\mathbf{h}|^2/m$ for all $\mathbf{h}$, it follows that

$$\mathbf{h}'\mathbf{D}_{1t}\mathbf{h} = -[(1 - m^{-1})Eu_1(|\mathbf{x}|) + m^{-1}E\psi_1'(|\mathbf{x}|)|\mathbf{h}|^2];$$

and this is negative for all $\mathbf{h} \neq \mathbf{0}$, thus proving that $\mathbf{D}_{1t}$ is nonsingular.

Now put $\mathbf{W} = -\mathbf{D}_{2V}\mathbf{M} = A\mathbf{U} + \mathbf{M}$, where according to (6.4), $A = Eu_2'(|\mathbf{x}|^2)|\mathbf{x}|^4$ and $\mathbf{U} = E\mathbf{z}'\mathbf{M}\mathbf{z}\,\mathbf{z}\mathbf{z}'$. It must be proved that $\mathbf{W} = \mathbf{0}$ only if $\mathbf{M} = 0$. Application of (3.4) and (D) yield

$$(6.5)\qquad A = E\psi_2'(|\mathbf{x}|^2)|\mathbf{x}|^2 - E\psi_2(|\mathbf{x}|^2) = c - m,$$

where c is positive.

Now we calculate $\mathbf{U}$. Let $a_1, \cdots, a_m$ be the eigenvalues of $\mathbf{M}$, and $\mathbf{e}_1, \cdots, \mathbf{e}_m$ the respective unit eigenvectors. Inserting the decomposition $\mathbf{M} = \sum_j \mathbf{a}_j\mathbf{e}_j\mathbf{e}_j'$ in the definition of $\mathbf{U}$, the spherical symmetry of $\mathbf{z}$ implies that $\mathbf{e}_j'\mathbf{U}\mathbf{e}_k = 0$ if $i \neq k$; this in turn implies that the $\mathbf{e}_j$'s are the eigenvectors of $\mathbf{U}$. To calculate the respective eigenvalues u_j, recall that for all j the variable $(\mathbf{z}'\mathbf{e}_j)^2$ has a beta distribution with parameters $(m - 1)/2$ and $\frac{1}{2}$, and hence $E(\mathbf{z}'\mathbf{e}_j)^4 = 3/[m(m + 2)]$; this implies that $u_j = (\sum_{i=1}^m a_i + 2a_j)/[m(m + 2)]$. Hence $\mathbf{W}$ has eigenvectors $\mathbf{e}_j$ with respective eigenvalues

$$(6.6)\qquad w_j = [(c - m)\textstyle\sum_{i=1}^m a_i + (m^2 + 2c)a_j]/[m(m + 2)].$$

We must prove that if the w_j's are all zero, so are the a_j's. If $c = m$, this fact is obvious by (6.6). If $c \neq m$, the system of linear equations in the a_j's given by $w_j = 0$ $(j = 1, \cdots, m)$ is equivalent to $\sum_{i=1}^{m} a_i + Ca_j = 0$ $(j = 1, \cdots, m)$, with $C = (m^2 + 2c)/(c - m)$. The matrix of this system has eigenvalues C (with multiplicity $m - 1$) and $C + m$. As both are obviously nonnull, the system has a unique solution, proving that $D\boldsymbol{\lambda}$ is nonsingular at $\boldsymbol{\theta}_0$, and hence that Huber's conditions for asymptotic normality hold.

If $z_1, \cdots, z_m$ are the coordinates of the vector $\mathbf{z}$, the symmetry of the distribution implies that $Ez_i z_j z_k = 0$ for all i, j, k, and hence that the covariances between each component of $\boldsymbol{\Psi}_1$ and each element of $\boldsymbol{\Psi}_2$ are all null. Then the covariance matrix of the asymptotic distribution of $(\mathbf{t}_n, \mathbf{V}_n)$ can be represented in block form as

$\mathbf{D}_{1t}^{-1}\mathbf{A}(\mathbf{D}_{1t}^{-1})'$	$\mathbf{0}$
$\mathbf{0}$	$\mathbf{D}_{2V}^{-1}\mathbf{B}(\mathbf{D}_{2V}^{-1})'$

where $\mathbf{A}$ and $\mathbf{B}$ are respectively the covariance matrices of $\boldsymbol{\Psi}_1(\mathbf{x}, \boldsymbol{\theta}_0)$ and of $\boldsymbol{\Psi}_2(\mathbf{x}, \boldsymbol{\theta}_0)$. This proves asymptotic independence, and application of (6.3) yields the last assertion, so the proof is complete.

7. Robustness measures. We shall briefly consider two indicators of robustness introduced by Hampel (1968 and 1971): the influence function and the breakdown bound. The distribution P will be assumed to be spherically symmetric, so that with the notation of the former section $\boldsymbol{\theta}_0 = (\mathbf{0}, s^2\mathbf{I})$ for some $s > 0$.

A standard argument (Andrews et al., 1972) yields the influence function of the location estimator $\mathbf{t}$:

$$IF(\mathbf{x}; \mathbf{t}, P) = -\mathbf{D}_{1t}^{-1}\boldsymbol{\Psi}_1(\mathbf{x}, \boldsymbol{\theta}_0) = b^{-1}u_1(s^{-1}|\mathbf{x}|)\mathbf{x} ,$$

where b and $\mathbf{D}_{1t}$ are defined in Theorem 6. From this we calculate easily the "gross error sensitivity" of $\mathbf{t}$: GES $(\mathbf{t}, P) = \sup_{\mathbf{x}} |IF(\mathbf{x}; \mathbf{t}, P)| = K_1 s/b$. The influence function of $\mathbf{V}$ is obtained likewise, but the resulting expression is not easy to handle.

To calculate the breakdown bound δ^* of $(\mathbf{t}, \mathbf{V})$, operating with distributions of the form $Q = (1 - \varepsilon)P + \varepsilon\delta_{\mathbf{x}}$, and letting $\mathbf{x} \to \infty$, one obtains (see Maronna, 1974):

$$\delta^* \leqq \min(1/K_2, 1 - m/K_2) . \tag{7.1}$$

It is natural to conjecture that there is actually equality above. In fact the values of the right member computed for Huber's Proposal 2 in the univariate case, coincide with the numerical values given by Hampel (1971); but unfortunately Hampel does not give an explicit proof of his results.

In any case, a consequence of (7.1) is that if one wants to take K_2 large enough to ensure (E) as explained in Section 2, he may damage δ^*. Moreover, since

$K_2 > m$, (7.1) implies that for *all* M-estimators, $\delta^* \leqq (m + 1)^{-1}$, so that for high dimensionality there may be difficulties which will require further thought. At the same time, operating with measures of the form $Q = (1 - \varepsilon)P + \varepsilon H_r$, where H_r is the uniform distribution on the spherical surface of radius r, and letting $r \to \infty$, it follows that ε must be larger than m/K_2 in order for the estimate to break down. This fact, together with the reasonable deficiencies given in the next section, suggest that it is extreme departure from the radial type, rather than longtailedness, that may cause the estimates to fail when m is large.

It is not difficult to prove that, if P is the m-variate spherical normal distribution, $m^{-\frac{1}{2}}$ GES $(\mathbf{t}, P) \geqq (m/2)^{-\frac{1}{2}}\Gamma(m/2)/\Gamma[(m - 1)/2]$ for all estimators; and the right-hand member approaches 1 rapidly for large m. This points out another intrinsic (lesser) drawback of M-estimators for high dimensionality.

8. Some asymptotic values. In this section we shall consider numerically the asymptotic behavior of the two families of estimators described in Section 2:

(i) Multivariate Huber's Proposal 2. For simplicity only the case $k_1 = k_2$ is considered, although other possibilities might be better. To allow some comparison between different dimensions, the family is parameterized with the number $p = P(|\mathbf{x}| > k_1)$, where P is the unit spherical normal; i.e., p is the "Winsorization proportion." The estimator is denoted by $H(p, m)$ or $H(p)$. The values chosen are $p = 0.5$, 0.3 and 0.2, which for $m = 1$ correspond approximately to $k_1 = 0.7$, 1.0 and 1.2, which are three of the values considered in Andrews et al. (1972). We include also the limit cases $p = 1$—i.e. the "median"—with some positive k_2, and $p = 0$—i.e., sample means and covariances.

(ii) Maximum likelihood estimator for the Student distribution with g degrees of freedom, to be denoted by MLST (g, m) or MLST (g) for the values $g = 1$, 2, 3 and 5.

The distributions chosen were the unit radial normal, and the radial Student distribution with g degrees of feedom, denoted by ST (g, m) or ST (g).

Since the location estimators in consideration will have an asymptotically normal distribution with mean $\mathbf{0}$ and covariance matrix $c\mathbf{I}$, it will be only necessary to report the constant c, which will be called the "variance" of the estimator. Table 1 exhibits their deficiencies, obtained by dividing the "variance" by that of the corresponding maximum likelihood estimate. No asymptotic parameters are given for the scatter matrix.

In addition we give for δ^* the value (7.1), and the sensitivity measure SENS $=$ GES$^2/m$, where GES is computed for the unit normal.

The variances may be obtained from the deficiencies by recalling that the variance of MLST (g, m) for ST (g, m) is $1 + 2/(g + m)$. The value of δ^* for $H(1.0)$ is omitted since it depends on k_2.

9. Results for finite samples. Some parameters of the distributions of the estimators for finite samples in the spherically symmetric case were computed by the Monte Carlo method. The algorithm used for the numerical solution of

TABLE 1

Asymptotic deficiencies and robustness of MLST(g, m) *and* $H(p, m)$.

	ST(1)	ST(2)	ST(3)	ST(5)	NORMAL	SENS	δ^*
$m = 2$							
MLST(1)	1.000	1.054	1.116	1.206	1.537	2.278	0.333
MLST(2)	1.083	1.000	1.015	1.060	1.279	1.639	0.250
MLST(3)	1.228	1.018	1.000	1.017	1.179	1.533	0.200
MLST(5)	1.563	1.094	1.022	1.000	1.097	1.609	0.143
$H(1.0)$	1.200	1.081	1.071	1.094	1.273	1.273	—
$H(0.5)$	1.460	1.102	1.031	1.005	1.093	1.515	0.361
$H(0.3)$	1.826	1.208	1.081	1.018	1.048	1.802	0.291
$H(0.2)$	2.161	1.300	1.129	1.037	1.029	2.070	0.248
$H(0.0)$	∞	∞	2.143	1.296	1.000	∞	0
$m = 4$							
MLST(1)	1.000	1.030	1.066	1.122	1.377	2.423	0.200
MLST(2)	1.047	1.000	1.010	1.042	1.239	1.589	0.167
MLST(3)	1.136	1.013	1.000	1.013	1.170	1.362	0.143
MLST(5)	1.350	1.069	1.017	1.000	1.103	1.261	0.111
$H(1.0)$	1.270	1.081	1.037	1.023	1.132	1.132	—
$H(0.5)$	1.650	1.199	1.088	1.023	1.045	1.336	0.195
$H(0.3)$	2.018	1.321	1.156	1.055	1.023	1.547	0.165
$H(0.2)$	2.335	1.417	1.210	1.081	1.014	1.735	0.146
$H(0.0)$	∞	∞	1.667	1.364	1.000	∞	0
$m = 6$							
MLST(1)	1.000	1.019	1.043	1.081	1.287	2.792	0.143
MLST(2)	1.032	1.000	1.007	1.031	1.203	1.734	0.125
MLST(3)	1.094	1.009	1.000	1.010	1.155	1.413	0.111
MLST(5)	1.249	1.053	1.013	1.000	1.101	1.210	0.090
$H(1.0)$	1.327	1.107	1.047	1.015	1.086	1.086	—
$H(0.5)$	1.728	1.252	1.126	1.044	1.030	1.264	0.136
$H(0.3)$	2.078	1.375	1.198	1.081	1.015	1.440	0.117
$H(0.2)$	2.369	1.469	1.252	1.109	1.009	1.591	0.106
$H(0.0)$	∞	∞	2.545	1.410	1.000	∞	0
$m = 10$							
MLST(1)	1.000	1.010	1.022	1.043	1.189	3.679	0.090
MLST(2)	1.018	1.000	1.004	1.018	1.151	2.145	0.083
MLST(3)	1.055	1.006	1.000	1.006	1.125	1.652	0.077
MLST(5)	1.152	1.034	1.009	1.000	1.091	1.289	0.006
$H(1.0)$	1.397	1.147	1.073	1.024	1.051	1.051	—
$H(0.5)$	1.787	1.305	1.168	1.073	1.018	1.198	0.084
$H(0.3)$	2.099	1.422	1.240	1.112	1.009	1.335	0.075
$H(0.2)$	2.352	1.509	1.292	1.140	1.005	1.451	0.069
$H(0.0)$	∞	∞	2.600	1.471	1.000	∞	0

(2.1)—(2.2) was inspired by the procedure used in Andrews et al. (1972, page 17) for MLST (1, 1). Taking $\mathbf{F}(\mathbf{t}, \mathbf{V})$ as in (3.1) define

$$\mathbf{G}(\mathbf{t}, \mathbf{V}) = \mathbf{F}(\mathbf{t}, \mathbf{V})\mathbf{V}^{-1}\mathbf{F}(\mathbf{t}, \mathbf{V})$$

and

$$\mathbf{g}(\mathbf{t}, \mathbf{V}) = Eu_1(d(\mathbf{x}, \mathbf{t}; \mathbf{V}^{-1}))\mathbf{x}/Eu_1(d(\mathbf{x}, \mathbf{t}; \mathbf{V}^{-1}))\,,$$

so that (**t**, **V**) is a solution of (2.1)—(2.2) if and only if $\mathbf{G}(\mathbf{t}, \mathbf{V}) = \mathbf{V}$ and $\mathbf{g}(\mathbf{t}, \mathbf{V}) = \mathbf{t}$. Start from an initial approximation $(\mathbf{t}_0, \mathbf{V}_0)$—I chose sample means and covariances—and recursively compute $\mathbf{t}_{n+1} = \mathbf{g}(\mathbf{t}_n, \mathbf{V}_n)$ and $\mathbf{V}_{n+1} = \mathbf{G}(\mathbf{t}_n, \mathbf{V}_n)$; if the limit exists, it must be a solution. I have not been able to prove the convergence of the procedure, but its empirical behavior is good, at least for low dimensionality. The procedure using **F** instead of **G** is instead much slower, especially for samples from long-tailed distributions. However, a much more rapid algorithm will be necessary for simulating in higher dimensionality, probably accompanied by a better election of the starting values. Here, due to machine-time limitations, only the bivariate case could be treated.

Recall that, if the underlying distribution is radial, so is the distribution of $\mathbf{t}_n$, and moreover, $n|\mathbf{t}_n|^2$ ought to be approximately a multiple of a χ^2 with m degrees of freedom, so that $\mathbf{t}_n$ can be well represented by some scale parameters of $n|\mathbf{t}_n|^2$. The parameters chosen were (see Andrews et al., 1972):

(i) The "variance": $v = En|\mathbf{t}_n|^2/m$.

(ii) The "α-pseudovariances": $\mathrm{PV} = \xi_\alpha/\chi_\alpha^2$, where ξ_α and χ_α^2 are respectively the upper α-percent points of $n|\mathbf{t}_n|^2$ and of the χ^2 distribution with m degrees of freedom. Here we report only $\mathrm{PV}_{0.05}$.

(iii) The "index of nonnormality": $\mathrm{INN} = \mathrm{PV}_{.05}/\mathrm{PV}_{.25}$.

(iv) A measure of the "deformation" of $\mathbf{V}_n$ with respect to the "correct" form (in this case, radial) $\mathrm{DEF}(\mathbf{V}) = (\det \mathbf{V})^{1/m}/(\mathrm{tr}\ \mathbf{V}/m)$. We computed the (lower) α-percent points of $\mathrm{DEF}(\mathbf{V}_n)$, DEF_α, for $\alpha = 0.25$ and 0.50.

The distributions used were Normal and ST(3). Since they are of the form "Normal/Independent," a large gain in precision was achieved for $\mathbf{t}_n$ by exploiting their peculiar conditional independence properties as explained in Andrews

TABLE 2

Behavior of estimators for finite samples

	v	PV	INN	DEF_{50}	DEF_{25}
Normal, $n = 10$					
$H(0.0)$	1.0(0)	1.0(0)	1.0	0.83	0.68
$H(0.3)$	1.051(3)	1.050(3)	1.0	0.80	0.64
$H(0.5)$	1.103(6)	1.100(5)	1.0	0.77	0.59
ST(3), $n = 10$					
$H(0.0)$	2.78(11)	2.90(7)	1.18	0.74	0.56
$H(0.3)$	1.663(14)	1.753(21)	1.09	0.78	0.62
$H(0.5)$	1.572(12)	1.651(18)	1.08	0.76	0.57
Normal, $n = 20$					
$H(0.0)$	1.0(0)	1.0(0)	1.0	0.93	0.87
$H(0.3)$	1.048(3)	1.049(3)	1.0	0.91	0.85
$H(0.5)$	1.094(5)	1.094(4)	1.0	0.90	0.82
ST(3), $n = 20$					
$H(0.0)$	2.92(9)	3.06(9)	1.28	0.84	0.71
$H(0.3)$	1.609(14)	1.647(17)	1.04	0.89	0.80
$H(0.5)$	1.528(11)	1.564(15)	1.04	0.89	0.79

et al. (1972), which applies almost literally to the multivariate case. Unfortunately, it does not seem possible to apply the same ideas to $\mathbf{V}_n$, and hence the percent points of DEF had to be computed "naïvely" from its empirical distribution function.

Table 2 shows the results for the estimators $H(0.0)$ (sample means and covariances), $H(0.30)$ and $H(0.50)$, and the sample sizes $n = 10$ and 20, giving in parentheses after each value its estimated standard deviation in units of the last digit. The number of replications was 400 for the Normal, 1500 for ST(3) with $n = 20$, and 2000 for ST(3) with $n = 10$.

Acknowledgment. This paper is a revised version of the author's doctoral thesis. The author is indebted to his thesis director, Dr. Victor J. Yohai, for having introduced him into the problems of robustness.

REFERENCES

ANDREWS, D. F., BICKEL, P. J., HAMPEL, F. R., HUBER, P. J., ROGERS, W. H., and TUKEY, J. W. (1972). *Robust Estimates of Location: Survey and Advances.* Princeton Univ. Press.

BICKEL, P. J. (1964). On some alternative estimates for shift in the p-variate one sample problem. *Ann. Math. Statist.* **35** 1079–1090.

DUNFORD, N. and SCHWARTZ, J. (1958). *Linear Operators.* Interscience, New York.

GENTLEMAN, W. M. (1965). Robust estimation of multivariate location by minimizing pth power deviations. Ph. D. dissertation, Princeton Univ.

GNANADESIKAN, R. and KETTENRING, J. R. (1972). Robust estimates, residuals, and outlier detection with multiresponse data. *Biometrics* **28** 81–124.

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

HAMPEL, F. R. (1971). A general qualitative definition of robustness. *Ann. Math. Statist.* **42** 1887–1896.

HAMPEL, F. R. (1973). Robust estimation: A condensed partial survey. *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete* **27** 87–104.

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. Prob.* **1** 221–233. Univ. of California Press.

HUBER, P. J. (1972). Robust statistics: A review. *Ann. Math. Statist.* **43** 1041–1067.

LACHENBRUCH, P. A., SNEERINGER, C., and REVO, L. T. (1973). Robustness of the linear and quadratic discriminant function to certain types of nonnormality. *Comm. Statist.* **1** 39–56.

MARONNA, R. A. (1974). Estimación robusta de locación y dispersión multivariadas. Doctoral dissertation, Universidad de Buenos Aires.

SEN, P. K. and PURI, M. L. (1971). *Nonparametric Methods in Multivariate Analysis.* Wiley, New York.

WOLFOWITZ, J. (1954). Generalization of the theorem of Glivenko–Cantelli. *Ann. Math. Statist.* **25** 131–138.

YOHAI, V. J. and MARONNA, R. A. (1975). Location estimates based on linear combinations of modified order statistics. *Notas de Mathemática* No. 31, Univ. de la Plata.

DEPARTAMENTO DE MATEMÁTICA
FUNDACIÓN BARILOCHE
C. C. 138
S. C. DE BARILOCHE, ARGENTINA