*The Annals of Statistics*
1987, Vol. 15, No. 1, 234–251

# A DISTRIBUTION-FREE *M*-ESTIMATOR OF MULTIVARIATE SCATTER[1]

BY DAVID E. TYLER

*Rutgers University*

The existence and uniqueness of a limiting form of a Huber-type $M$-estimator of multivariate scatter is established under certain conditions on the observed sample. These conditions hold with probability one when sampling randomly from a continuous multivariate distribution. The existence of the estimator is proven by showing that it is the limiting point of a specific algorithm. Hence, the proof is constructive. For continuous populations, the estimator of multivariate scatter is shown to be strongly consistent and asymptotically normal. An important property of the estimator is that its asymptotic distribution is distribution-free with respect to the class of continuous elliptically distributed populations. This distribution-free property also holds for the finite sample size distribution when the location parameter is known. In addition, the estimator is the "most robust" estimator of the scatter matrix of an elliptical distribution in the sense of minimizing the maximum asymptotic variance.

**1. Introduction and summary.** For a sample $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n$ from an $m$-variate distribution with known center, say $\mathbf{t}$, Maronna (1976) defines an affine invariant $M$-estimator of multivariate scatter to be the symmetric positive definite matrix $V_n$ that satisfies the equation

$$\operatorname{ave}\{u_2(s_i)(\mathbf{x}_i - \mathbf{t})(\mathbf{x}_i - \mathbf{t})'\} = V_n,$$

where $s_i = (\mathbf{x}_i - \mathbf{t})' V_n^{-1} (\mathbf{x}_i - \mathbf{t})$. The function $u_2$ must satisfy some general conditions given in Section 2 of Maronna's paper. One condition is that for every hyperplane $L$, $P_n(L) < 1 - m/K_2$, where $P_n$ is the empirical distribution measure and $K_2 = \sup_{s>0} s u_2(s)$. As noted by Maronna, since any set of $m$ points in $\mathbb{R}^m$ is contained in some hyperplane, this condition implies $n > m(1 - m/K_2)^{-1}$.

The affine invariant $M$-estimators of scatter are particularly suited for estimating the pseudocovariance or scatter matrix $V$ of an elliptical population, that is, one with density of the form

$$f(\mathbf{x}; \mathbf{t}, V, g) = |V|^{-1/2} g\{(\mathbf{x} - \mathbf{t})' V^{-1} (\mathbf{x} - \mathbf{t})\}, \tag{1.1}$$

where $g$ is some nonnegative function not dependent on $\mathbf{t}$ and $V$. In fact, the definition of the $M$-estimators of scatter given by Maronna (1976) corresponds to a generalization of the maximum likelihood estimators of the scatter matrix of an elliptical population. The family of elliptical distributions provides a multi-

Received May 1984; revised November 1985.
[1]Research supported by NSF grants MCS-8301354 and DMS-8405325.
*AMS* 1980 *subject classifications*. Primary 62H12, 62H10; secondary 62G05.
*Key words and phrases*. Covariance matrix, distribution-free, elliptical distribution, $M$-estimation, multivariate scatter, pseudocovariance matrix, robustness.

variate location-scatter family of distributions that primarily serve as long-tailed alternatives to the multivariate normal model. If the second moments exist, then the parameter $V$ is proportional to the covariance matrix. Properties of elliptical distributions have been studied by Kelker (1970), and a recent discussion of these distributions is given in the text by Muirhead (1982). Within the class of elliptical distributions, the scatter matrix $V$ is not well defined. For a specific distribution, the functional form $g$ in (1.1) can be replaced by $g_1(s) = c^m g(cs)$, where $c$ is a fixed positive scalar. This results in a change in the scatter matrix parameter from $V$ to $cV$. However, parameters of the form $H(V)$ are well defined whenever the function $H$ satisfies the following condition.

CONDITION 1.1. For any symmetric positive definite matrix $V$ and any positive scalar $c$, $H(V) = H(cV)$.

In other words, the directions and relative magnitudes of the axes of the elliptical contours are well defined, but unless further specified, a parameter measuring the spread of the contours is not. This paper focuses on parameters satisfying Condition 1.1.

In this paper, a special case of the affine invariant $M$-estimators of scatter is considered, namely the solution to

$$m \operatorname{ave}\{(\mathbf{x}_i - \mathbf{t})(\mathbf{x}_i - \mathbf{t})'/(\mathbf{x}_i - \mathbf{t})'V_n^{-1}(\mathbf{x}_i - \mathbf{t})\} = V_n, \tag{1.2}$$

which corresponds to choosing $u_2(s) = ms^{-1}$. The existence proof given by Maronna (1976) does not apply to (1.2) since $su_2(s) = m$ and hence $K_2 = m$. Huber (1981), Chapter 8, defines a more general class of affine invariant $M$-estimators of scatter and gives a more general proof of existence. For analogous reasons, his proof also does not apply to (1.2). A constructive proof of the existence of a solution to (1.2) for finite sample sizes is given in Section 2; see Theorem 2.2 and Corollaries 2.2 and 2.3. The solution to (1.2) is not unique, since if $V_n$ is a solution, then $cV_n$ is a solution for any positive scalar $c$. Except for this positive scalar factor, though, the solution is unique; see Corollary 2.1. Thus, the statistic $H(V_n)$ is uniquely defined whenever $H$ satisfies Condition 1.1.

A motivation behind the author's study of the estimator defined by (1.2) stems from observing that it is a limiting form of a Huber-type $M$-estimator of scatter. These estimators are defined by choosing for some fixed $r$, $u_2(s) = \alpha$ if $s \le r^2$ and $u_2(s) = \alpha r^2/s$ for $s > r^2$, where the scaling factor $\alpha$ is defined so that $E\{\chi_m^2 u_2(\chi_m^2)\} = m$. In an earlier paper (Tyler (1983)), it is noted that as $r \to 0$ the asymptotic variance of the Huber-type estimators of scatter when sampling from an elliptical population does not depend on the functional form $g$ in (1.1). In addition, this limiting value corresponds to the strict upper bound for the asymptotic variances of the maximum likelihood estimators of scatter derived when $g$ is assumed known. It is shown in Section 3 that these limiting properties of the asymptotic variance of the Huber-type estimators hold for the asymptotic variance of the estimator defined by (1.2). This is expected since for the Huber-type estimators as $r \to 0$, $u_2(s) \to ms^{-1}$ for $s > 0$, which corresponds to (1.2). Consequently, one can argue that the proposed estimator is the "most robust" estimator for the scatter matrix of an elliptical population; see Remark 3.1. More importantly, for inferential purposes, no nuisance parameters are involved in the asymptotic standard error of the proposed estimator, at least for elliptical populations. This distribution-free property also holds for the finite sample size distribution of the estimator of scatter, see Theorem 3.3. Finally, the strong consistency and asymptotic normality of the proposed estimator of multivariate scatter are established in Theorems 3.1 and 3.2, respectively.

The results of Section 3 assume the center **t** is known. Section 4 treats the case when **t** is not known. It is shown that the asymptotic results of Section 3 are valid when **t** is replaced by an estimate that converges to **t** sufficiently fast and if the population being sampled satisfies some inverse moment conditions. Theorem 4.1 establishes consistency and Theorem 4.2 establishes asymptotic normality.

**2. Existence and uniqueness.** Without loss of generality, it is assumed in this section and in the succeeding section that $\mathbf{t} = \mathbf{0}$. For any proper subspace $S$ of $\mathbb{R}^m$, let $0 < R(S) < m$ represent its rank. For the sample $\{\mathbf{x}_i, 1 \le i \le n\}$, let $P_n$ be the empirical distribution measure and let $n_0$ represent the largest subsample size such that no $q$ vectors from this subsample lie in a $(q-1)$-dimensional subspace. Some conditions on the observed sample used to establish the various results in this section are listed for easy reference.

CONDITION 2.1. The sample $\{\mathbf{x}_i, 1 \le i \le n\}$ contains no values equal to $\mathbf{0}$, and furthermore:

(i) span $\mathbb{R}^m$;
(ii) $n_0 \ge m + 1$;
(iii) $P_n(S) < 1/m$ for any proper subspace $S$;
(iv) $P_n(S) < \min\{1/m, n_0 R(S)/(nm)\}$ for any proper subspace $S$ and $n_0 > m(m-1)$;
(v) $n_0 = n > m(m-1)$.

Note that each condition is implied by any succeeding condition, and that conditions (iii)–(v) imply $n > m(m-1)$. For such $n$, all five conditions hold with probability one when the sample represents a random sample from a continuous distribution in $\mathbb{R}^m$ since $n_0 = n$ and $P_n(S) = R(S)/n$ with probability one.

Define on the set of symmetric positive definite matrices the matrix valued function

$$M(\Gamma) = m\,\mathrm{ave}\{\Gamma^{1/2}\mathbf{x}_i\mathbf{x}_i'\Gamma^{1/2}/\mathbf{x}_i'\Gamma\mathbf{x}_i\}, \tag{2.1}$$

where $A^{1/2}$ refers to the unique symmetric positive semidefinite square root of the symmetric positive semidefinite matrix $A$. The function $M$ is essentially one-to-one in the following sense.

THEOREM 2.1. *If Condition* 2.1(ii) *holds, then for any two symmetric positive definite matrices* $\Gamma_1$ *and* $\Gamma_2$, $M(\Gamma_1) = M(\Gamma_2)$ *if and only if* $\Gamma_1 = c\Gamma_2$ *for some positive scalar* $c$.

PROOF. If $\Gamma_1 = c\Gamma_2$, then it easily follows that $M(\Gamma_1) = M(\Gamma_2)$. To prove the converse suppose $M(\Gamma_1) = M(\Gamma_2)$ and without loss of generality assume $\Gamma_2 = I$. Let $\gamma_1$ represent the largest eigenvalue of $\Gamma_1$ and $R_1$ its associated eigenprojection. Multiplying both sides of the equation $M(\Gamma_1) = M(\Gamma_2)$ by $R_1$ and then taking the trace gives

$$\gamma_1 m \operatorname{ave}\{\mathbf{x}_i' R_1 \mathbf{x}_i / \mathbf{x}_i' \Gamma \mathbf{x}_i\} = m \operatorname{ave}\{\mathbf{x}_i' R_1 \mathbf{x}_i / \mathbf{x}_i' \mathbf{x}_i\}. \tag{2.2}$$

Since $\mathbf{x}_i'\Gamma\mathbf{x}_i < \gamma_1 \mathbf{x}_i'\mathbf{x}_i$ unless $R_1\mathbf{x}_i = \mathbf{x}_i$, (2.2) holds only if either $R_1\mathbf{x}_i = \mathbf{0}$ or $R_1\mathbf{x}_i = \mathbf{x}_i$ for all $1 \le i \le n$. This contradicts Condition 2.1(ii) unless $R_1 = I$, which implies $\Gamma_1 = \gamma_1 I$. □

If there exists a $\Gamma$ such that $M(\Gamma) = I$, then a solution to (1.2) is obtained by setting $V_n = \Gamma^{-1}$. The uniqueness of the solution is an immediate consequence of Theorem 2.1.

COROLLARY 2.1. *If a symmetric positive definite solution $V_{1,n}$ exists to* (1.2) *and if Condition* 2.1(ii) *holds, then the symmetric positive definite matrix $V_{2,n}$ is also a solution if and only if $V_{1,n} = cV_{2,n}$ for some positive scalar c.*

To construct a solution to the equation $M(\Gamma) = I$, and hence a solution to (1.2), consider some initial symmetric positive definite matrix $\Gamma_1$. Assuming the sample spans $\mathbb{R}^m$, recursively define the sequence of positive definite symmetric matrices

$$\begin{aligned} \Gamma_{k+1} &= \Gamma_k^{1/2} M_k^{-1} \Gamma_k^{1/2} / \operatorname{tr}(\Gamma_k M_k^{-1}) \\ &= \{m \operatorname{ave}(\mathbf{x}_i \mathbf{x}_i' / \mathbf{x}_i' \Gamma_k \mathbf{x}_i)\}^{-1} / \operatorname{tr}(\Gamma_k M_k^{-1}), \end{aligned} \tag{2.3}$$

where $M_k = M(\Gamma_k)$. If the sequence $\Gamma_k$ converges to a nonsingular $\Gamma$, then $M(\Gamma) = I$. Before proving the convergence of the sequence under certain conditions on the sample, some general preliminary results are established.

LEMMA 2.1. *Let $\lambda_{1,k}$ and $\lambda_{m,k}$ represent the largest and smallest eigenvalues of $M_k$, respectively. If Condition* 2.1(i) *holds, then $\lambda_{1,k}$ is a decreasing sequence with $\lambda_{1,k} \to \lambda_1 \ge 1$ and $\lambda_{m,k}$ is an increasing sequence with $\lambda_{m,k} \to \lambda_m \le 1$.*

PROOF. By definition,

$$\begin{aligned} I &= m \operatorname{ave}\{M_k^{-1/2}\Gamma_k^{1/2}\mathbf{x}_i\mathbf{x}_i'\Gamma_k^{1/2}M_k^{-1/2}/\mathbf{x}_i'\Gamma_k\mathbf{x}_i\} \\ &= m \operatorname{ave}\{(\Gamma_{k+1}^{1/2}\mathbf{x}_i\mathbf{x}_i'\Gamma_{k+1}^{1/2}/\mathbf{x}_i'\Gamma_{k+1}\mathbf{x}_i)(\mathbf{x}_i'\Gamma_k^{1/2}M_k^{-1}\Gamma_k^{1/2}\mathbf{x}_i/\mathbf{x}_i'\Gamma_k\mathbf{x}_i)\}. \end{aligned} \tag{2.4}$$

The second equality is obtained by noting that the symmetric positive definite square root of $\Gamma_k^{1/2}M_k^{-1}\Gamma_k^{1/2}$ has the form $Q_k M_k^{-1/2}\Gamma_k^{1/2}$, where $Q_k$ is an orthogonal matrix. By the extremal properties of the largest and smallest eigenvalues, (2.4) implies

$$\lambda_{m,k}^{-1} M_{k+1} \ge I \ge \lambda_{1,k}^{-1} M_{k+1}, \tag{2.5}$$

where the ordering refers to the partial ordering of symmetric positive semidefinite matrices. Since $\operatorname{tr}(M_k) = m$, the left-sided inequality in (2.5) implies $\lambda_{m,k} \le \lambda_{m,k+1} \le 1$ and the right-sided inequality implies $\lambda_{1,k} \ge \lambda_{1,k+1} \ge 1$. Application of the monotone convergence theorem completes the proof. □

THEOREM 2.2. *If Condition* 2.1(iii) *holds, and if for some* $k$, $\lambda_{m,k} > mP_n(S)/R(S)$ *for all proper subspaces* $S$, *then* $\Gamma_k \to \Gamma$, *a symmetric positive definite matrix satisfying* $M(\Gamma) = I$.

PROOF. Since $\operatorname{tr}(\Gamma) = 1$, there exists a subsequence $\Gamma_j \to \Gamma$, a symmetric positive definite or semidefinite matrix. This subsequence can be chosen so that

$$\Gamma_j^{1/2}\mathbf{x}_i/\left(\mathbf{x}_i'\Gamma_j\mathbf{x}_i\right)^{1/2} \to \theta_i$$

whenever $\Gamma\mathbf{x}_i = \mathbf{0}$, and furthermore so that $Q_j \to Q$, where $Q_j$ is defined after (2.4). The subsequences $M_j$ and $M_{j+1}$ then converge, respectively, to

$$M = n^{-1}m\left\{\sum_{i\in w}\Gamma^{1/2}\mathbf{x}_i\mathbf{x}_i'\Gamma^{1/2}/(\mathbf{x}_i'\Gamma\mathbf{x}_i) + \sum_{i\notin w}\theta_i\theta_i'\right\} \tag{2.6}$$

and

$$M_0 = n^{-1}m\left\{\sum_{i\in w}\Gamma_0^{1/2}\mathbf{x}_i\mathbf{x}_i'\Gamma_0^{1/2}/(\mathbf{x}_i'\Gamma_0\mathbf{x}_i) + \sum_{i\notin w}\phi_i\phi_i'\right\}, \tag{2.7}$$

where

$$\Gamma_{j+1} \to \Gamma_0 = \Gamma^{1/2}M^{-1}\Gamma^{1/2}/\operatorname{tr}(M^{-1}\Gamma), \qquad \phi_i = QM^{1/2}\theta_i/\left(\theta_i'M^{-1}\theta_i\right)^{1/2},$$

and the index set $w = \{i \mid \Gamma\mathbf{x}_i \ne 0,\ 1 \le i \le n\}$. The orthogonal matrix $Q$ satisfies the property $(\Gamma^{1/2}M^{-1}\Gamma^{1/2})^{1/2} = QM^{-1/2}\Gamma^{1/2}$. Lemma 2.1 insures $M$ is nonsingular since it implies $\lambda_m > 0$.

By Lemma 2.1, the largest and smallest eigenvalues of both $M$ and $M_0$ are $\lambda_1$ and $\lambda_m$, respectively. Let $P$ and $P_0$ represent the eigenprojections associated with the largest root of $M$ and $M_0$, respectively, and let $r = \operatorname{rank}(P)$ and $r_0 = \operatorname{rank}(P_0)$. Without loss of generality, assume $r_0 \ge r$. If this condition does not hold, the subsequences $(\Gamma_j, \Gamma_{j+1})$ can be replaced by $(\Gamma_{j+1}, \Gamma_{j+2})$ and so on until the condition is met.

Arguments similar to those used in justifying (2.4) can be applied to (2.6) and (2.7) to obtain

$$I = n^{-1}m\left\{\sum_{i\in w}M^{-1/2}\Gamma^{1/2}\mathbf{x}_i\mathbf{x}_i'\Gamma^{1/2}M^{-1/2}/\mathbf{x}_i'\Gamma\mathbf{x}_i + \sum_{i\notin w}M^{-1/2}\theta_i\theta_i'M^{-1/2}\right\} \tag{2.8}$$

and

$$\begin{aligned} I = \lambda_1^{-1}M_0 + n^{-1}m\Big\{&\sum_{i\in w}\left(\Gamma_0^{1/2}\mathbf{x}_i\mathbf{x}_i'\Gamma_0^{1/2}/\mathbf{x}_i'\Gamma_0\mathbf{x}_i\right) \\ &\times\left(\mathbf{x}_i'\Gamma^{1/2}M^{-1}\Gamma^{1/2}\mathbf{x}_i/\mathbf{x}_i'\Gamma\mathbf{x}_i - \lambda_1^{-1}\right)\Big\} \\ &+n^{-1}m\left\{\sum_{i\notin w}\phi_i\phi_i'\left(\theta_i'M^{-1}\theta_i - \lambda_1^{-1}\right)\right\}. \end{aligned} \tag{2.9}$$

Pre- and postmultiply both sides of (2.9) by $P_0$ and recall $P_0^2 = P_0$ and $\lambda_1^{-1}P_0M_0P_0 = P_0$. This gives for $i \in w$ either $P_0QM^{-1/2}\Gamma^{1/2}\mathbf{x}_i = \mathbf{0}$ or $P\Gamma^{1/2}\mathbf{x}_i = \Gamma^{1/2}\mathbf{x}_i$, and for $i \notin w$ either $P_0QM^{-1/2}\theta_i = \mathbf{0}$ or $P\theta_i = \theta_i$. Using this result, premultiplying both sides of (2.8) by $P_0Q$ and postmultiplying by $I - P$ gives $P_0Q(I - P) = 0$ or equivalently $Q'P_0Q = Q'P_0QP$. However, since $r_0 \geq r$, this implies $P = Q'P_0Q$.

The results of the previous paragraph imply for $i \in w$ either $P\Gamma^{1/2}\mathbf{x}_i = \mathbf{0}$ or $P\Gamma^{1/2}\mathbf{x}_i = \Gamma^{1/2}\mathbf{x}_i$. This contradicts Condition 2.1(ii) unless $P\Gamma^{1/2} = 0$ or $P\Gamma^{1/2} = \Gamma^{1/2}$. If $P\Gamma^{1/2} = 0$, then after noting that the results of the previous paragraph also imply for $i \in w$ either $P\theta_i = \mathbf{0}$ or $P\theta_i = \theta_i$, it follows from (2.6) that $\lambda_1 P = n^{-1}m\Sigma_{i \in v}\theta_i\theta_i'$ where the index set

$$v = \{i | \Gamma\mathbf{x}_i = \mathbf{0},\ P\theta_i = \theta_i,\ 1 \leq i \leq n\}.$$

Applying the trace to the last equation and recalling that $\theta_i'\theta_i = 1$ gives $\lambda_1 r \leq n^{-1}mn_\Gamma$, where $n_\Gamma$ represents the number of sample vectors in the null space of $\Gamma$. However, by Condition 2.1(iii) this implies $\lambda_1 \leq mn_\Gamma/(rn) < 1$, a contradiction. Thus, $P\Gamma^{1/2} = \Gamma^{1/2}$ and so

$$(2.10) \qquad P = n^{-1}m\lambda_1^{-1}\sum_{i \in w}\Gamma^{1/2}\mathbf{x}_i\mathbf{x}_i'\Gamma^{1/2}/\mathbf{x}_i'\Gamma\mathbf{x}_i + n^{-1}m\,\lambda_1^{-1}\sum_{i \in w}\theta_i\theta_i'.$$

The ranks of the two summations in (2.6) are additive, and hence the ranks of the two summations in (2.10) are additive. This implies that the two terms on the right-hand side of (2.10) are both orthogonal projections. By Condition 2.1(i), the first term has rank equal to rank($\Gamma$). Applying the trace to the first term gives

$$(2.11) \qquad \lambda_1 = n^{-1}m(n - n_\Gamma)/\text{rank}(\Gamma).$$

Applying the trace to the last term of (2.10) gives $\lambda_1 < 1$. This implies $v$ is empty and thus by Condition 2.1(i), $r = \text{rank}(\Gamma)$. The inequality $m = \text{tr}(M) \geq r\lambda_1 + (m - r)\lambda_m$ then gives $\lambda_m \leq mn_\Gamma/\{n(m - r)\}$.

The last inequality violates the condition on $\lambda_{m,k}$ unless $M_k \to I$, since by Theorem 3.2, $\lambda_{m,k}$ increases to $\lambda_m$. Since $M_k \to I$, it follows from (3.3) that the sequence $\Gamma_k \to \Gamma$. By (2.11), $\Gamma$ must be nonsingular. Finally, by the continuity of the function $M(\cdot)$ when the argument is nonsingular, $M(\Gamma) = I$. □

As a consequence of Theorems 2.1 and 2.2, in order to prove the existence of a "unique" solution to (1.2) and to construct the solution via (2.3), it is sufficient to show the existence of an initial symmetric positive definite matrix $\Gamma_1$ such that the largest root of $M(\Gamma_1)$ satisfies the condition stated in Theorem 2.2. Two examples are given in the following corollaries.

COROLLARY 2.2. *If Condition* 2.1(v) *holds, then for the initial value* $\Gamma_1 = \{\Sigma_{i=1}^n \mathbf{x}_i\mathbf{x}_i'\}^{-1}$, *the sequence* $\Gamma_k \to \Gamma$, *a symmetric positive definite matrix satisfying the equation* $M(\Gamma) = I$.

PROOF. Since Condition 2.1(v) implies 2.1(iii), it is sufficient to show $\lambda_{m,k} > m/n$. Let $\mathbf{w}_i = \Gamma_1^{1/2}\mathbf{x}_i$ and $\alpha = \max\{\mathbf{w}_i'\mathbf{w}_i,\ 1 \leq i \leq n\}$. Since $\Sigma_{i=1}^n\mathbf{w}_i\mathbf{w}_i' = I$, it follows that $0 < \alpha < 1$ and $M_1 \geq \{m/(n\alpha)\} I$. Thus, $\lambda_{m,1} \geq m/(n\alpha)$. However, $\alpha = 1$ only if there exists a $j$ such that $\mathbf{w}_i'\mathbf{w}_j = 0$ for $i \neq j$. This contradicts Condition 2.1(v) and hence $\lambda_{m,k} > m/n$. □

Let $\{\mathbf{y}_i, 1 \leq i \leq n_0\}$ represent a subsample of size $n_0$ such that any subset of size $m$ from this subsample spans $\mathbb{R}^m$. If Condition 2.1(iv) holds, then by the preceding corollary there exists a positive definite matrix $\Gamma_1$ such that

$$n_0^{-1} m \sum_{i=1}^{n_0} \Gamma_1^{1/2} \mathbf{y}_i \mathbf{y}_i' \Gamma_1^{1/2} / \mathbf{y}_i' \Gamma_1 \mathbf{y}_i = I. \tag{2.12}$$

Corollary 2.3. *If Condition* 2.1(iv) *holds and* $\Gamma_1$ *satisfies* (2.12), *then the sequence* $\Gamma_k \to \Gamma$, *a symmetric positive definite matrix satisfying the equation* $M(\Gamma) = I$.

Proof. From the definition of $\Gamma_1$, it follows that $M_1 \geq (n_0/n)I$. Thus, $\lambda_{m,1} \geq n_0/n > mP_n(S)/R(S)$ for any proper subspace $S$. □

Remark 2.1. Condition 2.1 demands that no sample vector $\mathbf{x}$ be equal to the center $\mathbf{t}$ since for such cases the left-hand side of (1.2) is undefined. A practical modification in such cases is obtained by disregarding these sample vectors in the average. Such sample vectors contain no directional information, that is, information on functions of the scatter matrix that satisfy Condition 1.1.

**3. Distribution theory: Location known.** Let $V_n$ represent the solution to (1.2) that is standardized so that $\operatorname{tr}(V_n) = m$, and where it is still assumed without loss of generality that $\mathbf{t} = \mathbf{0}$.

By the results of the previous section, for $n > m(m-1)$, $V_n$ exists and is unique on a set with probability one when the random sample is drawn from a continuous distribution in $\mathbb{R}^m$. For a continuous population, say represented by an $m$-dimensional random vector $\mathbf{X}$ possessing a density in $\mathbb{R}^m$, let $V$ be the symmetric positive definite solution to

$$mE(\mathbf{X}\mathbf{X}'/\mathbf{X}'V^{-1}\mathbf{X}) = V, \tag{3.1}$$

which is standardized so that $\operatorname{tr}(V) = m$. Application of Theorems 2.1 and 2.2 to a continuous population rather than a sample insures that $V$ exists and is unique. If the population is elliptical, then $V$ corresponds to the pseudocovariance or scatter matrix parameter.

The strong consistency of $V_n$ as an estimate of $V$ does not follow readily from the general theory for $M$-estimators. In particular, Condition (B-4) as given by Huber ((1981), page 131) does not hold. The proof of the following consistency theorem is similar to the classical proof of the maximum likelihood estimators.

Theorem 3.1. *If the sample* $\{\mathbf{x}_i, i = 1, 2, \dots\}$ *represents a random sample from a continuous distribution, then* $V_n \to V$ *almost surely.*

PROOF. Let $M_n(\Gamma)$ and $M(\Gamma)$ be defined by (2.1) when the average refers to the sample and population, respectively, and let $h_n(\Gamma) = \text{tr}\{M_n(\Gamma)^2\}$ and $h(\Gamma) = \text{tr}\{M(\Gamma)^2\}$ for symmetric positive definite matrices $\Gamma$ of order $m$. Since $\text{tr}\{M_n(\Gamma)\} = \text{tr}\{M(\Gamma)\} = m$, it follows that $m \le h_n(\Gamma) \le m^2$ and $m \le h(\Gamma) \le m^2$. Furthermore, $h_n(\Gamma) = m$ or $h(\Gamma) = m$ if and only if $M_n(\Gamma) = I$ or $M(\Gamma) = I$, respectively. Let $C$ be any compact set such that for all $\Gamma$ in $C$, $\text{tr}(\Gamma^{-1}) = m$ and $\text{tr}(\Gamma) < K < \infty$, where $K > m$. It is to be shown that

$$\sup_{\Gamma \in C} |h_n(\Gamma) - h(\Gamma)| \to 0 \quad \text{almost surely.} \tag{3.2}$$

It is also to be shown that under Condition 2.1(ii),

(3.3) $\Gamma_n$ is a critical value of $h_n$ if and only if $M_n(\Gamma_n) = I$.

Without loss of generality, assume $V = I$ and hence $M(I) = I$. Let $C$ be chosen so that it satisfies the aforementioned conditions and contains $I$ in its interior. Since $h(I) = m < h(\Gamma)$ for $\Gamma \in C$, $\Gamma \ne I$, and $h(\Gamma)$ is continuous, statement (3.2) implies with probability one that for large enough $n$, $h_n(I)$ is less than $h_n(\Gamma)$ for any value on the boundary of $C$. Hence, with probability one, $h_n$ contains a local minimum in $C$ for large enough $n$. By (3.3), this implies with probability one that $V_n^{-1}$ is eventually in $C$. Since $C$ can be chosen to be arbitrarily small, it follows that $V_n^{-1} \to I$ almost surely and hence $V_n \to I$ almost surely.

PROOF OF STATEMENT (3.2). For $\mathbf{x} \in \mathbb{R}^m$, $\mathbf{x} \ne 0$, define

$$G(\mathbf{x}, \Gamma) = \Gamma^{1/2}\mathbf{x}\mathbf{x}'\Gamma^{1/2}/\mathbf{x}'\Gamma\mathbf{x}.$$

It can be shown that the function $G$ is an equicontinuous function of $\Gamma$ on $C$ in the sense that for all $\varepsilon \to 0$ there exist $\delta_\varepsilon < 0$, which is not dependent on $\mathbf{x} \ne 0$, $\Gamma_0 \in C$, nor $\Gamma \in C$, such that if $\|\Gamma - \Gamma_0\| < \delta_\varepsilon$, then $\|G(\mathbf{x}, \Gamma) - G(\mathbf{x}, \Gamma_0)\| < \varepsilon$. Since $C$ is compact, for any $\delta_\varepsilon > 0$ there exist a finite partitioning of $C$, say $\{C_{\varepsilon, k}\}$, such that $\|\Gamma_1 - \Gamma_2\| < \delta_\varepsilon$ for $\Gamma_1 \in C_{\varepsilon, k}$ and $\Gamma_2 \in C_{\varepsilon, k}$. Choose one element from each of the sets $C_{\varepsilon, k}$, say $\Gamma_{\varepsilon, k}$, and label the set $C_\varepsilon$. Since $P(\mathbf{X} = \mathbf{0}) = 0$, it follows from the equicontinuity of $G$ and the strong law of large numbers that

$$\begin{aligned} &\sup_{\Gamma \in C} \|M_n(\Gamma) - M(\Gamma)\| \\ &\quad \le \max_k \sup_{\Gamma \in C_{\varepsilon, k}} \{\|M_n(\Gamma) - M_n(\Gamma_{\varepsilon, k})\| + \|M(\Gamma) - M(\Gamma_{\varepsilon, k})\|\} \\ &\qquad + \max_k \|M_n(\Gamma_{\varepsilon, k}) - M(\Gamma_{\varepsilon, k}) \\ &\quad \le 2m\varepsilon + \max_k \|M_n(\Gamma_{\varepsilon, k}) - M(\Gamma_{\varepsilon, k})\| \to 2m\varepsilon \end{aligned} \tag{3.4}$$

almost surely. Since $\varepsilon$ can be chosen to be arbitrarily small, it follows that $\sup_{\Gamma \in C}\|M_n(\Gamma) - M(\Gamma)\| \to 0$ almost surely, and hence (3.2) holds. □

PROOF OF STATEMENT (3.3). The critical points of $h_n$ can be found by using the expansion

$$h_n(\Gamma + \dot{\Gamma}) = h_n(\Gamma) + 2m\,\text{tr}\big[\{B(\Gamma)\Gamma B(\Gamma) - C(\Gamma)\}\dot{\Gamma}\big] + O(\|\dot{\Gamma}\|^2), \tag{3.5}$$

where

$$B(\Gamma) = \text{ave}\{\mathbf{x}_i\mathbf{x}_i'/\mathbf{x}_i'\Gamma\mathbf{x}_i\}$$

and

$$C(\Gamma) = \text{ave}\left[\left\{\mathbf{x}_i'\Gamma B(\Gamma)\Gamma\mathbf{x}_i/(\mathbf{x}_i'\Gamma\mathbf{x}_i)^2\right\}\mathbf{x}_i\mathbf{x}_i'\right].$$

Thus, $\Gamma_n$ is a critical point of $h_n$ if and only if $B(\Gamma_n)\Gamma_n B(\Gamma_n) = C(\Gamma_n)$. Multiplying both sides of this last equation by $\{B(\Gamma_n)\}^{-1}$, which by Condition 2.1(ii) exists, and then taking the trace gives

$$1 = \text{ave}\left[\theta_i' M_n(\Gamma_n)\theta_i \times \theta_i'\{M_n(\Gamma_n)\}^{-1}\theta_i\right], \tag{3.6}$$

where

$$\theta_i = \Gamma_n^{1/2}\mathbf{x}_i/\mathbf{x}_i'\Gamma_n\mathbf{x}_i.$$

However, by Kantorovich's inequality, equation (3.6) holds only if $\theta_i$ is an eigenvector of $M(\Gamma_n)$, $1 \le i \le n$. This contradicts Condition 2.1(ii) unless $M_n(\Gamma_n) = I$. □

Rather than using the general theory for $M$-estimators, the asymptotic normality of $V_n$ can be proven directly. The most cumbersome part involves specifying the form of the asymptotic covariance of $V_n$. For this reason, the asymptotic distribution of $V_{0,n} = mV_n/\text{tr}(V^{-1}V_n)$ rather than $V_n$ is given in the next theorem. Note that $V_{0,n}$ represents the unique symmetric positive definite solution to (1.2), which is standardized so that $\text{tr}(V^{-1}V_{0,n}) = m$. For any well defined function of the scatter parameter, that is functions satisfying Condition 1.1, $H(V_{0,n}) = H(V_n)$.

The following conventional notation is used. For a symmetric positive semidefinite matrix $A$, let $A^+$ represent its unique Moore–Penrose generalized inverse. If $B$ is a $b \times t$ matrix, then $\text{vec}(B)$ is the transformation of $B$ into the $bt$-dimensional vector formed by stacking the columns of $B$. If $B$ is $b \times t$ and $C$ is $c \times u$, then the Kronecker product of $B$ and $C$ is the $bc \times tu$ partitioned matrix $B \otimes C = [b_{ij}C]$. The commutation matrix is the $ab \times ab$ matrix

$$K_{a,b} = \sum_{i=1}^{a}\sum_{j=1}^{b} J_{ij} \otimes J_{ij}',$$

where $J_{ij}$ is an $a \times b$ matrix with one in the $(i, j)$ position and zeroes elsewhere. A good overview of the algebraic properties of the vec transformation, the Kronecker product and the commutation matrix is given by Magnus and Neudecker (1979). These properties and the properties of the Moore–Penrose generalized inverse are to be used without further reference. For the sake of brevity, many algebraic manipulations used hereafter are not explicitly stated since most are fairly straightforward but cumbersome.

Let $M_n(\Gamma) = m\,\text{ave}\{\Gamma^{1/2}\mathbf{x}_i\mathbf{x}_i'\Gamma^{1/2}/\mathbf{x}_i'\Gamma\mathbf{x}_i\}$ where the average is over $1 \le i \le n$. For random samples from continuous populations in $\mathbb{R}^m$, the central limit theorem implies $n^{1/2}\{M_n(V^{-1}) - I\} \to Z$ in distribution, where $\text{vec}(Z)$ is multivariate normal with mean zero and covariance matrix $\Sigma$, which depends on the fourth moment of $\theta = V^{-1/2}\mathbf{X}/(\mathbf{X}'V^{-1}\mathbf{X})^{1/2}$. Specifically,

$$\Sigma = m\left[\Sigma_0 - m^{-1}\operatorname{vec}(I)\{\operatorname{vec}(I)\}'\right], \quad \text{where } \Sigma_0 = mE(\boldsymbol{\theta}\boldsymbol{\theta}' \otimes \boldsymbol{\theta}\boldsymbol{\theta}').$$

Since $\boldsymbol{\theta}$ has a continuous distribution on the unit $m$-sphere, $\operatorname{Var}(\theta' B\boldsymbol{\theta}) = 0$ if and only if $B = \lambda I$ or $B + B' = 0$. This implies $\operatorname{vec}(B)$ is in the null space of $\Sigma$ if and only if $B$ satisfies the aforementioned property. Hence, $\operatorname{rank}(\Sigma) = \frac{1}{2}m(m-1) - 1$ and the orthogonal projection onto the range of $\Sigma$ is

$$Q = \tfrac{1}{2}(I + K_{m,m}) - m^{-1}\operatorname{vec}(I)\{\operatorname{vec}(I)\}'.$$

THEOREM 3.2. *If the sample* $\{\mathbf{x}_i, 1 \le i \le n\}$ *represents a random sample from a continuous distribution in* $\mathbb{R}^m$, *then* $n^{1/2}(V_{0,n} - V) \to N$ *in distribution, where* $\operatorname{vec}(N)$ *is multivariate normal with mean zero and variance–covariance matrix*

$$\Sigma_N(V) = 4m(V^{1/2} \otimes V^{1/2})\{(I + K_{m,m})(I - \Sigma_0)\}^{+} \times \Sigma_0\{(I + K_{m,m})(I - \Sigma_0)\}^{+}(V^{1/2} \otimes V^{1/2}),$$

*where* $\Sigma_0$ *is previously defined. The rank of* $\Sigma_N(V)$ *is* $\frac{1}{2}m(m-1) - 1$.

PROOF. Applying the identity

$$\left(\mathbf{x}_i' V_{0,n}^{-1}\mathbf{x}_i\right)^{-1} = \left(\mathbf{x}_i' V^{-1}\mathbf{x}_i\right)^{-1} - \mathbf{x}_i'\left(V_{0,n}^{-1} - V^{-1}\right)\mathbf{x}_i/\left\{\mathbf{x}_i'V^{-1}\mathbf{x}_i \times \mathbf{x}_i'V_{0,n}^{-1}\mathbf{x}_i\right\}$$

to the equation $M_n(V_{0,n}^{-1}) = I$ gives

$$A_n = M_n(V^{-1}) - m\operatorname{ave}\left\{\boldsymbol{\theta}_i\boldsymbol{\theta}_i'\left(A_n^{-1} - I\right)\boldsymbol{\theta}_i\boldsymbol{\theta}_i'/\boldsymbol{\theta}_i'A_n^{-1}\boldsymbol{\theta}_i\right\},$$

where

$$A_n = V^{-1/2}V_{0,n}V^{-1/2} \quad \text{and} \quad \boldsymbol{\theta}_i = V^{-1/2}\mathbf{x}_i/\left(\mathbf{x}_i'V^{-1}\mathbf{x}_i\right)^{1/2}.$$

This equation can be restated as

$$\left\{I - \Sigma_{1,n}\left(A_n^{-1/2} \otimes A_n^{-1/2}\right)\right\}\operatorname{vec}(A_n - I) = \operatorname{vec}\left\{M_n(V^{-1}) - I\right\}, \tag{3.7}$$

where

$$\Sigma_{1,n} = m\operatorname{ave}\left\{\boldsymbol{\theta}_i\boldsymbol{\theta}_i' \otimes \boldsymbol{\theta}_i\boldsymbol{\theta}_i'/\boldsymbol{\theta}_i'A_n^{-1}\boldsymbol{\theta}_i\right\}.$$

The term $\Sigma_{1,n}$ can be bounded above and below by $\lambda_m^{-1}(A_n)\Sigma_{0,n}$ and $\lambda_1^{-1}(A_n)\Sigma_{0,n}$, respectively, where $\lambda_1(A_n)$ and $\lambda_m(A_n)$ are the largest and smallest roots of $A_n$, respectively, and $\Sigma_{0,n} = m\operatorname{ave}\{\boldsymbol{\theta}_i\boldsymbol{\theta}_i' \otimes \boldsymbol{\theta}_i\boldsymbol{\theta}_i'\}$. Since $A_n \to I$ almost surely and by the large numbers $\Sigma_{0,n} \to \Sigma_0$ almost surely, it follows that $\Sigma_{1,n} \to \Sigma_0$ and hence

$$(I - \Sigma_0)\left\{n^{1/2}\operatorname{vec}(A_n - I)\right\} \to \operatorname{vec}(Z) \tag{3.8}$$

in distribution. The range of $(\mathrm{I} - \Sigma_0)$ contains the range of the orthogonal projection

$$Q = \tfrac{1}{2}(I + K_{m,m}) - m^{-1}\operatorname{vec}(I)\{\operatorname{vec}(I)\}',$$

and furthermore

$$Q\operatorname{vec}(A_n - I) = \operatorname{vec}(A_n - I) \quad \text{and} \quad Q\operatorname{vec}(Z) = \operatorname{vec}(Z).$$

This implies

$$n^{1/2}\text{vec}(A_n - I) \to \{Q(I - \Sigma_0)Q\}^{+}\text{vec}(Z)$$

in distribution, or equivalently

(3.9) $n^{1/2}\text{vec}(V_{0,n}) \to \text{vec}(N) = (V^{1/2} \otimes V^{1/2})\{Q(I - \Sigma_0)Q\}^{+}\text{vec}(Z)$

in distribution. This gives

$$\Sigma_n(V) = (V^{1/2} \otimes V^{1/2})\{Q(I - \Sigma_0)Q\}^{+}\Sigma\{Q(I - \Sigma_0)Q\}^{+}(V^{1/2} \otimes V^{1/2}),$$

which reduces to the form given in the statement of the theorem after noting that

$$\{Q(I - \Sigma_0)Q\}^{+} = \{Q(I - \Sigma_0)Q\}^{+}Q, \quad Q\Sigma Q = m\Sigma_0$$

since $Q$ vec$(I) = \mathbf{0}$, and

$$Q(I - \Sigma_0)Q = \tfrac{1}{2}(I + K_{m,m})(I - \Sigma_0)$$

since $(I - \Sigma_0)\text{vec}(I) = \mathbf{0}$. □

If the continuous distribution being sampled has an elliptical distribution, then $\boldsymbol{\theta}$ is uniformly distributed on the unit $m$-sphere. The distribution of $\boldsymbol{\theta}$ is the same as $P\boldsymbol{\theta}$ for any orthogonal $P$ and $\theta_i^2$ has a $\beta\{\frac{1}{2}, \frac{1}{2}(m-1)\}$ distribution. These two properties imply after some calculations that

$$\Sigma_0 = mE(\boldsymbol{\theta\theta}' \otimes \boldsymbol{\theta\theta}') = (m+2)^{-1}[I + K_{m,m} + \{\text{vec}(I)\}\{\text{vec}(I)\}'].$$

The form of $\Sigma_n$ in Theorem 4.2 then simplifies to

(3.10) $\Sigma_n = (1 + 2/m)[(I + K_{m,m})(V \otimes V) + 2m^{-1}\text{vec}(V)\{\text{vec}(V)\}'].$

Note that the asymptotic distribution is not dependent on the particular elliptical distribution being sampled, that is, it does not depend upon the "nuisance functional parameter" $g$ in (1.1). This is not a consequence of the asymptotic theory.

THEOREM 3.3. *If the sample* $\{\mathbf{x}_i, 1 \le i \le n\}$ *represents a random sample from a population with an elliptical contoured density* $f(\mathbf{0}, V; g)$ (*see* (1.1)), *then for* $n > m(m+1)$, *the distribution of* $V_{0,n}$ *does not depend on the function* $g$. *Furthermore the distribution of* $Z_n = V^{-1/2}V_{0,n}V^{-1/2}$ *is invariant under the transformation* $Z \to P'ZP$ *for any orthogonal* $P$.

PROOF. With probability one, the random matrix $Z_n$ is defined implicitly as the unique symmetric positive definite solution to the equations

$$m\,\text{ave}\{Z_n^{-1/2}\boldsymbol{\theta}_i\boldsymbol{\theta}_i'Z_n^{-1/2}/\boldsymbol{\theta}_i'Z_n^{-1}\boldsymbol{\theta}_i\} = I \quad \text{and} \quad \text{tr}(Z_n) = m,$$

where $\boldsymbol{\theta}_i = V^{-1/2}\mathbf{x}_i/\mathbf{x}_i'V^{-1}\mathbf{x}_i$. The theorem then follows after recalling that $\boldsymbol{\theta}_i$ is uniformly distributed on the unit $m$-sphere, irregardless of the functional form $g$. □

Although the distribution of $V_{0,n}$ does not depend upon the specific elliptical distribution from which the sample is taken, the exact distribution theory appears to be formidable. In any event, Theorem 3.3 has important implications if one wishes to simulate the distribution of $V_{0,n}$ or some function of $V_{0,n}$.

REMARK 3.1. The estimate $V_n$ can be viewed as the "most robust" estimate of scatter for elliptical distributions in the following minimax sense. Suppose the function $g$ is known, and let $\hat{V}_g$ represent the maximum likelihood estimate of $V$. Under the usual regularity conditions, the limiting distribution of $\text{vec}\{n^{1/2}(\hat{V}_{0,g} - V_0)\}$, where $V_0 = mV/\text{tr}(V)$ and $\hat{V}_{0,g} = m\hat{V}_g/\text{tr}(V_0^{-1}\hat{V}_g)$, is multivariate normal with mean $\mathbf{0}$. The asymptotic covariance matrix can be shown to be $\sigma_{1,g}[(I + K_{m,m})(V_0 \otimes V_0) - 2m^{-1}\text{vec}(V_0)\{\text{vec}(V_0)\}']$, where $\sigma_{1,g}$ is a scalar dependent on $g$ but not on $V$. The scalar $\sigma_{1,g}$ is strictly bounded above by $(1 + 2/m)$; see Tyler (1983), Theorem 1 and Section 4 for a more detailed discussion. Consider a real-valued continuously differentiable function $H(V)$ satisfying Condition 1.1. It follows from (3.10) and the bound on $\sigma_{1,g}$ that over the class of continuous elliptical distribution satisfying the aforementioned regularity conditions, the maximum asymptotic variance of $H(V_n)$ is less than the maximum asymptotic variance of any other consistent and uniformly asymptotically normal estimator of $H(V)$.

**4. Estimating scatter when location is unknown.** The simultaneous affine invariant $M$-estimators of multivariate location and scatter are defined by Maronna (1976) to be the vector $\mathbf{t}_n$ and the symmetric positive definite matrix $V_n$, respectively, which satisfy the equations

$$\text{ave}\{u_1(d_i)(\mathbf{x}_i - \mathbf{t}_n)\} = \mathbf{0}$$

and

$$\text{ave}\{u_2(s_i)(\mathbf{x}_i - \mathbf{t}_n)(\mathbf{x}_i - \mathbf{t}_n)'\} = V_n,$$

where

$$d_i = s_i^{1/2} = \{(\mathbf{x}_i - \mathbf{t}_n)'V_n^{-1}(\mathbf{x}_i - \mathbf{t}_n)\}^{1/2}.$$

Sufficient conditions on $u_1, u_2$ and the empirical distribution are given by Maronna (1976) and Huber (1981) to insure the existence of a joint solution $(\mathbf{t}_n, V_n)$. The proof of joint existence depends on showing that for a fixed $\mathbf{t}_n$ the solution to the second equation, say $V_n(\mathbf{t}_n)$, is a continuous function of $\mathbf{t}_n$. As far as the author is aware, the uniqueness of the joint solution is still an open question.

The corresponding limit to (1.2) of the Huber-type estimator for $u_1$ is $u_1(d) = d^{-1}$. This choice of $u_1$ possesses the property that for a fixed $V_n$ the solution $\mathbf{t}_n$ to the equation $\text{ave}\{u_1(d_i)(\mathbf{x}_i - \mathbf{t}_n)\} = \mathbf{0}$ is invariant under multiplication of $V_n$ by a positive scalar. This property is necessary if the equation is used in conjunction with (1.2). Unfortunately, although it is possible to show that the solution to (1.2) is a continuous function of $\mathbf{t}$ for $\mathbf{t} \neq \mathbf{x}_i$, $1 \leq i \leq n$, the discontinuities at the sample vectors are not removable. These discontinuities prevent the direct application of Brouwer's fixed point theorem, as in Maronna's (1976) proof, in proving the joint existence when $u_1(d) = d^{-1}$ and $u_2(s) = ms^{-1}$. The author has not resolved this difficulty.

Alternatively, rather than simultaneously estimating location and scatter, an estimate of scatter can be obtained by replacing the location parameter $\mathbf{t}$ in (1.2) by some estimate say $\mathbf{t}_n$. For example, $\mathbf{t}_n$ could be taken to be the sample mean, the componentwise median, or some other componentwise estimate of location. One drawback to componentwise estimates of location, except for the sample mean, is that they are not affine equivariant, and so the resulting estimate of scatter will not be affine equivariant. If affine equivariance is desired, rather than using the sample mean vector, $\mathbf{t}_n$ can be taken to be one of the affine equivariance $M$-estimates of location previously discussed. In the following discussion, $\mathbf{t}_n$ need not be affine equivariant.

If the estimate $\mathbf{t}_n$ is equal to some sample observation and $\mathbf{t}_n$ replaces $\mathbf{t}$ in (1.2), then a solution to (1.2) for $V_n$ does not exist. This poses no problem when using the sample mean since such an event has zero probability when sampling from a continuous distribution. However, for some other estimates of location, e.g., the componentwise median, such an event has positive probability. Rather than restrict the choice of the location estimate, define the estimate of scatter $V_n(\mathbf{t}_n)$ to be the symmetric positive definite matrix satisfying $\operatorname{tr}\{V_n(\mathbf{t}_n)\} = m$ and

$$(4.1)\quad m \operatorname{ave}^*\{(\mathbf{x}_i - \mathbf{t}_n)(\mathbf{x}_i - \mathbf{t}_n)'/(\mathbf{x}_i - \mathbf{t}_n)'V_n^{-1}(\mathbf{t}_n)(\mathbf{x}_i - \mathbf{t})\} = V_n(\mathbf{t}_n),$$

where ave* refers to the average over $1 \le i \le n$ for which $\mathbf{x}_i \neq \mathbf{t}_n$. A heuristic justification for this estimate is given in Remark 2.1. Hereafter, it is assumed that $\{\mathbf{x}_i,\ i = 1, 2, \dots, \}$ represents a random sample of a continuous random variable $\mathbf{X}$ of dimension $m$. This guarantees the existence and uniqueness of $V_n(\mathbf{t}_n)$ when $n > m(2m - 1)$.

LEMMA 4.1. *If $n > m(2m-1)$, then with probability one, $V_n(\mathbf{t})$ exists for all $\mathbf{t} \in \mathbb{R}^m$ and is unique for each $\mathbf{t}$.*

PROOF. With probability one, no $q + 1$ points lie in any $(q - 1)$-dimensional plane. Also, with probability one, no $m + 1$ planes generated by $m + 1$ distinct subsets of size $m$ from $\{\mathbf{x}_i, 1 \le i \le n\}$ contain a common point. Without loss of generality, consider $\mathbf{t} = \mathbf{0}$. The lemma follows after noting that on the intersection of the two aforementioned events $n_0 > n - m^2$ and $P_n(S) \le \{R(S) + 1\}/n$, and so Condition 2.1(iv) holds. □

Without loss of generality, it is still assumed that $\mathbf{t} = \mathbf{0}$ and hence $V_n(\mathbf{t}_n)$ estimates the parameter $V$ defined by (3.1). The consistency and asymptotic normality of $V_n(\mathbf{t}_n)$ are established in the following two theorems, provided $t_n$ converges to $\mathbf{t} = \mathbf{0}$ at an appropriate rate. It is interesting to note that in order to establish consistency and asymptotic normality, the population distribution cannot be too heavily concentrated about the center. Intuitively, such a condition seems necessary since small changes in $\mathbf{t}_n$ would result in large changes in the direction of $\mathbf{x} - \mathbf{t}_n$ for $\mathbf{x}$ near $\mathbf{t}_n$.

THEOREM 4.1. *Assume there exists a sequence of random variables $T_n$ such that for some $\alpha \geq 0$, $n^\alpha \mathbf{t}_n'\mathbf{t}_n \leq T_n$ and $T_n$ converges almost surely or in distribution, with convergence to* $\mathbf{0}$ *when $\alpha = 0$. If $E\{\|X\|^{-2/(2+\alpha)}\} < \infty$ for some norm then $V_n(\mathbf{t}_n) \to V$ almost surely or in probability, respectively.*

PROOF. Let $M(\Gamma)$, $M_n(\Gamma)$, $h(\Gamma)$ and $h_n(\Gamma)$ be defined as in the proof of Theorem 3.1, and define

$$(4.2)\quad M_n(\Gamma, \mathbf{t}_n) = m \operatorname{ave}^*\{\Gamma^{1/2}(\mathbf{x}_i - \mathbf{t}_n)(\mathbf{x}_i - \mathbf{t}_n)'\Gamma^{1/2}/(\mathbf{x}_i - \mathbf{t}_n)'\Gamma(\mathbf{x}_i - \mathbf{t}_n)\}$$

and $h_n(\Gamma, \mathbf{t}_n) = \operatorname{tr}\{M_n(\Gamma, \mathbf{t}_n)^2\}$. Statement (3.3) is valid if $h_n(\Gamma)$ and $M_n(\Gamma_n)$ are replaced by $h_n(\Gamma, \mathbf{t}_n)$ and $M_n(\Gamma_n, \mathbf{t}_n)$, respectively. As with Theorem 3.1, Theorem 4.1 follows if it can be shown that $\sup_{\Gamma \in C}|h_n(\Gamma, \mathbf{t}_n) - h(\Gamma)| \to 0$ almost surely or in probability, depending on the condition on $\mathbf{t}_n$. The set $C$ is defined as is (3.2). The almost sure case is considered here. The in probability case is analogous. By (3.2) it is sufficient to show $\sup_{\Gamma \in C}\|h_n(\Gamma, \mathbf{t}_n) - h_n(\Gamma)\| \to 0$ or

$$(4.3)\qquad \sup_{\Gamma \in C}\|M_n(\Gamma, \mathbf{t}_n) - M_n(\Gamma)\| \to 0$$

almost surely. Since norms are equivalent in finite dimensional spaces, let the norm in (4.3) be defined by $\|A\|^2 = \operatorname{tr}(A'A)$ for any $m \times m$ matrix $A$.

The event $\mathbf{x}_i = \mathbf{t}_n$ for at most one sample vector and $\mathbf{x}_i \neq \mathbf{0}$ for $1 \leq i \leq n$, has probability one. Therefore, notation is restricted to such events. Let $G(\mathbf{x}, \Gamma)$ be defined as in the proof of statement (3.2), and note $\|G(\mathbf{x}, \Gamma)\| = 1$ for $\mathbf{x} \neq \mathbf{0}$ and so $\|G(\mathbf{x} - \mathbf{t}_n, \Gamma) - G(\mathbf{x}, \Gamma)\| \leq 2$ for $\mathbf{x} \neq 0$ and $\mathbf{x} \neq \mathbf{t}_n$. For $\mathbf{x}_i \in \mathscr{A}_n(\gamma) = \{\mathbf{x} | (\mathbf{x} - \mathbf{t}_n)'(\mathbf{x} - \mathbf{t}_n) \geq \gamma \mathbf{x}'\mathbf{x}\}$, where $0 < \gamma < 1$,

$$\begin{aligned}
&\|G(\mathbf{x}_i - \mathbf{t}_n, \Gamma) - G(\mathbf{x}_i, \Gamma)\|^2 \\
&\quad = 2\{(\mathbf{t}_n'\Gamma\mathbf{t}_n)(\mathbf{x}_i'\Gamma\mathbf{x}_i) - (\mathbf{t}_n'\Gamma\mathbf{x}_i)^2\}/\{\mathbf{x}_i'\Gamma\mathbf{x}_i(\mathbf{x}_i - \mathbf{t}_n)'\Gamma(\mathbf{x}_i - \mathbf{t}_n)\} \\
(4.4)\qquad &\quad \leq 2(\mathbf{t}_n'\Gamma\mathbf{t}_n)/(\mathbf{x}_i - \mathbf{t}_n)'\Gamma(\mathbf{x}_i - \mathbf{t}_n) \\
&\quad \leq 2\lambda(\Gamma)\mathbf{t}_n'\mathbf{t}_n/(\mathbf{x}_n - \mathbf{t}_n)'(\mathbf{x}_i - \mathbf{t}_n) \\
&\quad \leq 2\lambda\gamma^{-1}\mathbf{t}_n'\mathbf{t}_n/\mathbf{x}_i'\mathbf{x}_i,
\end{aligned}$$

where $\lambda(\Gamma)$ is the ratio of the largest to the smallest root of $\Gamma$, and $\lambda = \max_{\Gamma \in C}\lambda(\Gamma) < \infty$ since $C$ is compact. The preceding results imply for $\Gamma \in C$,

$$\begin{aligned}
&\|M_n(\Gamma, \mathbf{t}_n) - M_n(\Gamma)\| \\
&\quad \leq m \operatorname{ave}^*\{\|G(\mathbf{x}_i - \mathbf{t}_n, \Gamma) - G(\mathbf{x}_i, \Gamma)\|\} + 2n^{-1}m \\
(4.5)\qquad &\quad \leq mn(n-1)^{-1}\Big[P_n\{\mathscr{A}_n^c(\gamma)\} + (2\lambda\gamma^{-1})^{1/2}(\mathbf{t}_n'\mathbf{t}_n)^{1/2} \\
&\qquad\qquad \times \operatorname{ave}\big\{(\mathbf{x}_i'\mathbf{x}_i)^{-1/2}\big\}\Big] - 2n^{-1}m,
\end{aligned}$$

where $P_n$ represents the empirical probability measure of $\{\mathbf{x}_i, 1 \leq i \leq n\}$. Since

$\mathbf{t}_n \to \mathbf{0}$ almost surely, $P_n\{\mathscr{A}_n(\gamma)\} \to 1$ almost surely. From Marcinkiewicz's SLLN (see, e.g., Loève (1963), result 16.A.4.4), $n^{-\alpha/2}\text{ave}\{(\mathbf{x}_i'\mathbf{x}_i)^{-1/2}\} \to 0$ almost surely since $E\{\|\mathbf{X}\|^{-2/(2+\alpha)}\} < \infty$. Thus by the conditions on $\mathbf{t}_n$ in the theorem, the right-hand side of (4.5) goes to zero almost surely. Statement (4.3) follows since the bound in (4.5) is uniform for $\Gamma \in C$. □

As an example, suppose $\mathbf{t}_n$ is taken to be the sample mean $\bar{\mathbf{x}}$. By Marcinkiewicz's SLLN for $\alpha < 1$, $n^{\alpha/2}\bar{\mathbf{x}} \to 0$ almost surely if $E\{\|\mathbf{X}\|^{2/(2-\alpha)}\} < \infty$. In addition, $n^{1/2}\bar{\mathbf{x}}$ converges in distribution if $E\{\|\mathbf{X}\|^2\} < \infty$. Thus, if $E\{\|\mathbf{X}\|^{2/(2-\alpha)}\} < \infty$ and $E\{\|\mathbf{X}\|^{-2/(2+\alpha)}\} < \infty$ hold for $\alpha = 1$, then $V_n(\bar{\mathbf{x}})$ is weakly consistent, and if they hold for some $0 \le \alpha \le 1$, then $V_n(\bar{\mathbf{x}})$ is strongly consistent.

As another example, suppose $\mathbf{t}_n$ is taken to be the sample componentwise median. If each population componentwise median is unique and has positive density at the median, then $n^{1/2}\mathbf{t}_n$ converges in distribution and by the LIL for medians, $n^{\alpha/2}\mathbf{t}_n \to \mathbf{0}$ almost surely for any $\alpha < 1$. Under the aforementioned conditions on the marginal distributions, $V_n(\mathbf{t}_n)$ is weakly consistent when $E\{\|\mathbf{X}\|^{-2/3}\} < \infty$ and strongly consistent when $E\{\|\mathbf{X}\|^{-2/(2+\alpha)}\} < \infty$ for some $0 \le \alpha < 1$.

Turning now to the question of asymptotic normality, let $V_{0,n}(\mathbf{t}_n) = mV_n(\mathbf{t})/\text{tr}\{V^{-1}V_n(\mathbf{t}_n)\}$. Also define $\mathbf{c} = E(S^{-1/2}\boldsymbol{\theta})$ and $C = E(S^{-1/2}\boldsymbol{\theta} \otimes \boldsymbol{\theta}\boldsymbol{\theta}')$, where again $\boldsymbol{\theta} = V^{-1/2}\mathbf{X}/S^{1/2}$ with $S = \mathbf{X}'V^{-1}\mathbf{X}$.

THEOREM 4.2. *Assume there exists a sequence of random variables $T_n$ such that $n\mathbf{t}_n'\mathbf{t}_n \le T_n$ and $T_n$ converges in distribution. If $E(\|X\|^{-3/2}) < \infty$ for some norm, $\mathbf{c} = \mathbf{0}$ and $C = 0$, then $n^{1/2}\{V_{0,n}(\mathbf{t}_n) - V\} \to N$ in distribution, where the distribution of $N$ is same as in Theorem* 3.2.

PROOF. Let $M_n(\Gamma)$, $M_n(\Gamma, \mathbf{t}_n)$ and $G(\mathbf{x}, \Gamma)$ be defined in the proof of Theorem 4.1, and let $\Sigma_{1,n}$, $\Sigma_{0,n}$ and $\boldsymbol{\theta}_i$ be defined as in the proof of Theorem 3.2. Also, let

$$A_n(\mathbf{t}_n) = V^{-1/2}V_{0,n}(\mathbf{t}_n)V^{-1/2},$$

$$\boldsymbol{\theta}_i(\mathbf{t}_n) = V^{-1/2}(\mathbf{x}_i - \mathbf{t}_n)/\left\{(\mathbf{x}_i - \mathbf{t}_n)'V^{-1}(\mathbf{x}_i - \mathbf{t}_n)\right\}^{1/2}$$

for

$$\mathbf{x}_i \neq \mathbf{t}_n,\ \Sigma_{1,n}(\mathbf{t}_n) = m\,\text{ave}^*\left\{\boldsymbol{\theta}_i(\mathbf{t}_n)\boldsymbol{\theta}_i(\mathbf{t}_n)' \otimes \boldsymbol{\theta}_i(\mathbf{t}_n)\boldsymbol{\theta}_i(\mathbf{t}_n^*)'/\theta_i(\mathbf{t}_n)'A_n(\mathbf{t}_n)^{-1}\boldsymbol{\theta}_i(\mathbf{t}_n)\right\}$$

and

$$\Sigma_{0,n}(\mathbf{t}_n) = m\,\text{ave}^*\{\boldsymbol{\theta}_i(\mathbf{t}_n)\boldsymbol{\theta}_i(\mathbf{t}_n)' \otimes \boldsymbol{\theta}_i(\mathbf{t}_n)\boldsymbol{\theta}_i(\mathbf{t}_n)'\}.$$

Statement (3.7) holds if $\Sigma_{1,n}$, $A_n$ and $M_n(V^{-1})$ are replaced by $\Sigma_{1,n}(\mathbf{t}_n)$, $A_n(\mathbf{t}_n)$ and $M_n(V^{-1}, \mathbf{t}_n)$, respectively. It is to be shown that $\Sigma_{1,n}(\mathbf{t}_n) \to \Sigma_0$ in probability and $n^{1/2}\{M_n(V^{-1}, \mathbf{t}_n) - \mathrm{I}\} \to Z$ in distribution. This implies (3.8) holds whenever $A_n$ is replaced by $A_n(\mathbf{t}_n)$. The remainder of the proof of the theorem is then identical to the proof of Theorem 3.2.

Unless stated otherwise, all convergence statements for the remainder of the proof are understood to be in distribution. To prove $\Sigma_{1,n}(\mathbf{t}_n) \to \Sigma_0$, as in the proof Theorem 3.2, since $A_n(\mathbf{t}_n) \to I$, it is sufficient to show $\Sigma_{0,n}(\mathbf{t}_n) \to \Sigma_0$, and since $\Sigma_{0,n} \to \Sigma_0$, it is sufficient to show $\|\Sigma_{0,n}(\mathbf{t}_n) - \Sigma_{0,n}\| \to 0$. Using the identities, $\mathrm{tr}(A_1 \otimes A_2) = \mathrm{tr}(A_1)\mathrm{tr}(A_2)$ and $(A_1 \otimes A_2)(B_1 \otimes B_2) = A_1B_1 \otimes A_2B_2$ when the matrices are conformable, one can note $\|\mathbf{b}_1\mathbf{b}_1' \otimes \mathbf{b}_1\mathbf{b}_1'\| = 1$ if $\mathbf{b}_1'\mathbf{b}_1 = 1$. Also

$$\|\mathbf{b}_1\mathbf{b}_1' \otimes \mathbf{b}_1\mathbf{b}_1' - \mathbf{b}_2\mathbf{b}_2' \otimes \mathbf{b}_2\mathbf{b}_2'\| = 2^{1/2}\left(1 - (\mathbf{b}_1'\mathbf{b}_2)^4\right)^{1/2} \le 2\left(1 - (\mathbf{b}_1'\mathbf{b}_2)^2\right)^{1/2}$$

if $\mathbf{b}_1'\mathbf{b}_1 = \mathbf{b}_2'\mathbf{b}_2 = 1$, where again $\|A\|^2 = \mathrm{tr}(A'A)$ for any $m \times m$ matrix $A$. Using these results, one obtains

$$\begin{aligned} &\|\Sigma_{0,n}(\mathbf{t}_n) - \Sigma_{0,n}\| \\ &\quad \le 2m\,\mathrm{ave}^*\left\{\|G(\mathbf{x}_i - \mathbf{t}_n, V^{-1}) - G(\mathbf{x}_i, V^{-1})\|\right\} + 2n^{-1}m \end{aligned} \tag{4.6}$$

on the event $\mathbf{x}_i = \mathbf{t}_n$ for at most one sample vector and $\mathbf{x}_i \ne \mathbf{0}$ for $1 \le i \le n$, which without loss of generality is assumed to hold. As in (4.5), the right-hand side of (4.6) goes to zero.

To show $n^{1/2}\{M_n(V^{-1}, \mathbf{t}_n) - I\} \to Z$, it is sufficient to show

$$n^{1/2}\left\{M_n\left(V^{-1}, \mathbf{t}_n\right) - M_n\left(V^{-1}\right)\right\} \to 0.$$

Using the expansion

$$\mathbf{x}'\mathbf{x}/(\mathbf{x} - \mathbf{t})'(\mathbf{x} - \mathbf{t}) = 1 + 2\mathbf{t}'\mathbf{x}/\mathbf{x}'\mathbf{x} + e(\mathbf{x}, \mathbf{t})/(\mathbf{x} - \mathbf{t})'(\mathbf{x} - \mathbf{t}),$$

where

$$e(\mathbf{x}, \mathbf{t}) = \left\{4(\mathbf{t}'\mathbf{x})^2 - (\mathbf{t}'\mathbf{t})(\mathbf{x}'\mathbf{x}) - 2\mathbf{t}'\mathbf{t}\mathbf{t}'\mathbf{x}\right\}/\mathbf{x}'\mathbf{x},$$

for $\mathbf{x}_i \ne \mathbf{t}_n$ one obtains

$$\theta_i(\mathbf{t}_n)\theta_i(\mathbf{t}_n)' = \theta_i\theta_i' + R_i + E_{1i} + E_{2i},$$

where

$$R_i = \left\{2c_i\theta_i\theta_i' - V^{-1/2}(\mathbf{x}_i\mathbf{t}' + \mathbf{t}_n\mathbf{x}_i')V^{-1/2}\right\}/s_i$$

with $c_i = \mathbf{x}_i'V^{-1}\mathbf{t}_n$ and $s_i = \mathbf{x}_i'V^{-1}\mathbf{x}_i$,

$$E_{1i} = e\left(V^{-1/2}\mathbf{x}_i, V^{-1/2}\mathbf{t}_n\right)\theta_i(\mathbf{t}_n)\theta_i(\mathbf{t}_n)'$$

and

$$E_{2i} = V^{-1/2}\left[(1 + 2c_i/s_i)\mathbf{t}_n\mathbf{t}_n' - (2c_i/s_i)(\mathbf{t}_n\mathbf{x}_i' + \mathbf{x}_i\mathbf{t}_n')\right]V^{-1/2}/s_i.$$

Note that if $\mathbf{x}_i = \mathbf{t}_n$ then $R_i = 0$, $e(V^{-1/2}\mathbf{x}_i, V^{-1/2}\mathbf{t}_n) = 1$ and $E_{2i} = -\theta_i\theta_i'$. Let $a_n = 1$ if $\mathbf{x}_i \ne \mathbf{t}_n$ for $1 \le i \le n$ and $a_n = n/(n-1)$ if $\mathbf{x}_i = \mathbf{t}_n$ for one such $i$. The results of this paragraph imply

$$\begin{aligned} M_n\left(V^{-1}, \mathbf{t}_n\right) &= a_n\left\{M_n\left(V^{-1}\right) + m\,\mathrm{ave}(R_i) + m\,\mathrm{ave}(E_{2i})\right\} \\ &\quad + m\,\mathrm{ave}^*(E_{1i}). \end{aligned} \tag{4.7}$$

After some algebraic manipulations, application of Marcinkiewicz's SLLN gives

$n^{1/2}\text{ave}(\|E_{2i}\|) \to 0$ almost surely and

$$n^{1/2}\text{ave}^*(\|E_{1i}\|) \le n^{1/2}\text{ave}^*\{|e(V^{-1/2}\mathbf{x}_i, V^{-1/2}\mathbf{t}_n)|\}$$
$$\le a_n n^{1/2}\text{ave}\{|e(V^{-1/2}\mathbf{x}_i, V^{-1/2}\mathbf{t}_n)|\} \to 0$$

almost surely. Algebraic manipulations using the vec operator give $\text{vec}(R_i) = \{2(\theta_i \otimes \theta_i\theta_i') - (I + K_{m,m})(\theta_i \otimes I)\}s_i^{-1/2}V^{-1/2}\mathbf{t}_n$, which implies

$$(4.8)\quad \|\text{vec}\{n^{1/2}\text{ave}(R_i)\} - \{2C - (I + K_{m,m})(\mathbf{c} \otimes I)\}V^{-1/2}(n^{1/2}\mathbf{t}_n)\| \to 0.$$

Thus, if $\mathbf{c} = 0$ and $C = 0$, then $n^{1/2}\|M_n(V^{-1}, \mathbf{t}_n) - M_n(V^{-1})\| \to 0$. This completes the proof. □

If $\mathbf{c} \neq \mathbf{0}$ or $C \neq 0$, then Theorem 4.2 is still valid if $n\mathbf{t}_n'\mathbf{t}_n \to 0$ in probability. More realistically, suppose $\mathbf{c} \neq 0$ or $C \neq 0$ but

$$[n^{1/2}\mathbf{t}_n, n^{1/2}\{M_n(V^{-1}) - I\}] \to (\mathbf{Y}, Z)$$

in distribution, where $(\mathbf{Y}, Z)$ is jointly multivariate normal with mean zero, $\Sigma_Y(V) = E(\mathbf{Y}\mathbf{Y}')$, $\Sigma_{ZY}(V) = E\{\text{vec}(Z)\mathbf{Y}'\}$ and the covariance matrix for $\text{vec}(Z)$ is given prior to Theorem 3.2. It follows from (4.7) and (4.8) that Theorem 4.2 still holds for this case but with $\text{vec}(N)$ being multivariate normal, mean zero and covariance matrix

$$(4.9)\quad A(V)\{m\Sigma_0 + \Sigma_{ZY}B(V)' + B(V)\Sigma_{ZY}' + B(V)\Sigma_Y B(V)'\}A(V)',$$

where

$$A(V) = 2(V^{1/2} \otimes V^{1/2})\{(I + K_{m,m})(I - \Sigma_0)\}^+$$

and

$$B(V) = (I + K_{m,m})\{C - (\mathbf{c} \otimes I)\}V^{-1/2}.$$

The form of (4.9) is obtained by using arguments similar to those after (3.8) and by noting $Q\{2C - (I + K_{m,m})(\mathbf{c} \otimes I)\}V^{-1/2} = B(V)$ with $Q$ defined prior to Theorem 3.2.

For distributions that are symmetric, i.e., $\mathscr{L}(\mathbf{X}) = \mathscr{L}(-\mathbf{X})$, $\mathbf{c} = 0$ and $C = 0$ and so Theorem 4.2 holds provided the other conditions of the theorem are met. In particular, for elliptical distributions, the asymptotic distribution of Theorem 4.2 does not depend on the particular functional $g$ in (1.1), at least for those $g$ under which $\mathbf{t}_n$ satisfies the conditions of the theorem and for which $E(\|\mathbf{X}\|^{-3/2}) < \infty$. Thus, if $\mathbf{t}_n$ is the sample median, then Remark 3.1 applies over the class of continuous elliptical distributions for which $E(\|\mathbf{X}\|^{-3/2}) < \infty$. Finally, for elliptical distributions, the condition $E(\|\mathbf{X}\|^{\beta}) < \infty$ is equivalent to

$$(4.10)\qquad \int_0^{\infty} s^{(\beta+m-2)/2}g(s)\,ds < \infty.$$

For normal populations, (4.10) is satisfied for $m > -\beta$, and so Theorem 4.2 applies to normal populations.

## REFERENCES

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

KELKER, D. (1970). Distribution theory of spherical distributions and a location-scale parameter generalization. *Sankhyā Ser. A* **32** 419–430.

LOÈVE, M. (1963). *Probability Theory*, 3rd ed. Van Nostrand Reinhold, New York.

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

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

MUIRHEAD, R. J. (1982). *Aspects of Multivariate Statistical Theory*. Wiley, New York.

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

DEPARTMENT OF STATISTICS
RUTGERS UNIVERSITY
NEW BRUNSWICK, NEW JERSEY 08903