The Annals of Statistics
1996, Vol. 24, No. 3, 1346–1370

CONSTRAINED *M*-ESTIMATION FOR MULTIVARIATE LOCATION AND SCATTER

BY JOHN T. KENT AND DAVID E. TYLER[1]

University of Leeds and Rutgers University

Consider the problem of estimating the location vector and scatter matrix from a set of multivariate data. Two standard classes of robust estimates are *M*-estimates and *S*-estimates. The *M*-estimates can be tuned to give good local robustness properties, such as good efficiency and a good bound on the influence function at an underlying distribution such as the multivariate normal. However, *M*-estimates suffer from poor breakdown properties in high dimensions. On the other hand, *S*-estimates can be tuned to have good breakdown properties, but when tuned in this way, they tend to suffer from poor local robustness properties. In this paper a hybrid estimate called a constrained *M*-estimate is proposed which combines both good local and good global robustness properties.

1. Introduction and summary. A fundamental problem in multivariate statistics is the development of affine equivariant robust alternatives to the sample mean vector and sample covariance matrix. An early class of robust alternatives was the *M*-estimates of multivariate location and scatter; see Maronna (1976) and Huber (1977). Besides being intuitively appealing, the multivariate *M*-estimates have good local robustness properties; that is, they are not greatly influenced by small perturbations in the data and have reasonably good efficiencies over a broad range of population models. On the other hand, Maronna (1976) and Stahel (1981) have shown that the multivariate *M*-estimates are not globally robust in the sense that they have relatively low breakdown points in high dimensions. Subsequently, high breakdown point affine equivariant multivariate location and scatter statistics have been introduced by Stahel (1981), Donoho (1982), Rousseeuw (1985), Davies (1987) and others. An ironic drawback to many of the high breakdown point statistics though is that they tend to have poor local robustness properties. A discussion of this problem is given in Tyler (1991). There currently exists a need to better understand the interplay between local and global robustness in the multivariate setting and to develop multivariate location and scatter statistics which are both locally and globally robust. To this end we introduce here a new class of multivariate location and scatter statistics which we call constrained *M*-estimates, or *CM*-estimates for short,

Received January 1994; revised May 1995.

[1]Research supported in part by NSF Grant DMS-91-06706 and SERC visiting fellowship at the University of Leeds.

AMS 1991 *subject classifications*. Primary 62F35, 62H12.

Key words and phrases. Breakdown, *M*-estimates, robustness, *S*-estimates.

and show that they not only have high breakdown points but when properly tuned also have good local robustness properties.

Let $X_n = \{\mathbf{x}_1, \dots, \mathbf{x}_n\}$ be a data set in $\mathbb{R}^p$ and let $\mathscr{P}_p$ denote the set of all $p \times p$ positive definite symmetric matrices. For the data set X_n we define the constrained M-estimates of multivariate location $\hat{\mu}(X_n) \in \mathbb{R}^p$ and scatter $\hat{V}(X_n) \in \mathscr{P}_p$ to be any pair which minimizes the objective function

$$(1.1)\quad L(\mu, V; X_n) = \operatorname{ave}\left[\rho\left\{(\mathbf{x}_i - \mu)'V^{-1}(\mathbf{x}_i - \mu)\right\}\right] + \tfrac{1}{2}\log\{\det(V)\}$$

over all $\mu \in \mathbb{R}^p$ and $V \in \mathscr{P}_p$ subject to the constraint

$$(1.2)\qquad \operatorname{ave}\left[\rho\left\{(\mathbf{x}_i - \mu)'V^{-1}(\mathbf{x}_i - \mu)\right\}\right] \leq \varepsilon\rho(\infty),$$

where "ave" stands for the arithmetic average over $i = 1, \dots, n$, ε is a fixed value between 0 and 1 and $\rho(s)$ is a bounded nondecreasing function for $s \geq 0$. In general, the minimization problem which defines the *CM*-estimates may have multiple solutions. The notation $(\hat{\mu}(X_n), \hat{V}(X_n))$ will be used to refer to an arbitrary solution to the minimization problem rather than to the set of all solutions.

Regardless of the dimension p, we show in Section 4 that the breakdown point of the *CM*-estimate is approximately $\min(\varepsilon, 1 - \varepsilon)$, and so for $\varepsilon = \frac{1}{2}$ the breakdown point is approximately $\frac{1}{2}$. Furthermore, the *CM*-estimates are affine equivariant in the following sense. For any nonsingular $p \times p$ matrix A and any $\mathbf{b} \in \mathbb{R}^p$, if $\hat{\mu}(X_n)$ and $\hat{V}(X_n)$ correspond to *CM*-estimates of location and scatter, respectively, for the data set X_n, then

$$(1.3)\quad \hat{\mu}(AX_n + \mathbf{b}) = A\hat{\mu}(X_n) + \mathbf{b} \quad \text{and} \quad \hat{V}(AX_n + \mathbf{b}) = A\hat{V}(X_n)A'$$

correspond to *CM*-estimates of location and scatter, respectively, for the data set $AX_n + b = \{A\mathbf{x}_1 + \mathbf{b}, \dots, A\mathbf{x}_n + \mathbf{b}\}$.

When ρ is differentiable, the *CM*-estimates satisfy the following estimating equation:

$$(1.4)\qquad \mu = \operatorname{ave}\{u(s_i)\mathbf{x}_i\}/\operatorname{ave}\{u(s_i)\},$$

$$(1.5)\qquad V = p\operatorname{ave}\{u(s_i)(\mathbf{x}_i - \mu)(\mathbf{x}_i - \mu)'\}/\operatorname{ave}\{\psi(s_i)\},$$

and either

$$(1.6)\qquad \operatorname{ave}\{\psi(s_i)\} = p$$

or

$$(1.7)\qquad \operatorname{ave}\{\rho(s_i)\} = \varepsilon\rho(\infty),$$

where $s_i = (\mathbf{x}_i - \mu)'V^{-1}(\mathbf{x} - \mu)$, $u(s) = 2\rho'(s)$ and $\psi(s) = su(s)$. Equations (1.4), (1.5) and (1.6) hold whenever strict inequality holds in (1.2) for the *CM*-estimate, and (1.4), (1.5) and (1.7) hold whenever the *CM*-estimate results in (1.2) being an equality. Equations (1.4), (1.5) and (1.6) arise as the critical points of (1.1). Usually, these critical points are expressed as the simultaneous solutions to (1.4) and (1.5) but with the term $\operatorname{ave}\{\psi(s_i)\}$ replaced by p in (1.5), in which case (1.6) holds automatically. To verify this latter claim, just multiply the modified equation by V^{-1} and then take the trace. See, for example, Maronna (1976) or Kent and Tyler (1991). It is more convenient for our purposes here to use the three expressions (1.4), (1.5) and (1.6). Equations (1.4), (1.5) and (1.7) arise as the critical points of (1.1) after introducing a Lagrange multiplier to account for the constraint (1.7).

The estimating equations are useful for establishing and studying local properties of the *CM*-estimates, such as influence functions, asymptotic normality and asymptotic relative efficiencies. All solutions to the estimating equations though need not correspond to *CM*-estimates since we are seeking a global minimum rather than just critical points. Global properties such as existence, uniqueness, consistency and breakdown points rely on the complete definition of the *CM*-estimates.

The paper is organized as follows. The functional version of the *CM*-estimates and conditions for the existence of the *CM*-estimates and *CM*-functionals are given in Section 3. The problem of uniqueness is discussed in Section 4. The finite sample breakdown points of the *CM*-estimates are given in Section 5. In Section 6, the influence functions of the *CM*-functionals are derived and asymptotic normality of the *CM*-estimates is proven under general conditions. The form of the influence functions and the asymptotic variance–covariance matrices of the *CM*-estimates can be expressed in relatively simple forms when the underlying population has a multivariate normal distribution or more generally when it has an elliptically symmetric distribution; see Section 7. Based on the results of Section 7, simple indices for assessing the local robustness of the *CM*-estimates are given in Section 8.

In Section 9, we consider classes of ρ-functions of the form $\rho_c(s) = c\rho(s)$ and propose a procedure for choosing the tuning constant $c > 0$ in order to achieve desirable local robustness properties. The choice of the tuning constant, though, does not affect the breakdown point. We demonstrate the results of the paper by taking a more detailed look at the influence functions and the asymptotic relative efficiencies under the multivariate normal model for the class of biweighted *CM*-estimates.

Proofs are reserved for the Appendix. Throughout the paper, data points, whether viewed as random variables, observations or realizations of random variables, are represented by lowercase boldface letters and always have a subscript, for example, $\mathbf{x}_1$. Whether or not a data point is a random variable is to be understood from its context. Data sets are represented by uppercase italic letters, for example, X_n, and random vectors are represented by uppercase boldface letters, for example, $\mathbf{X}$.

Before presenting the technicalities of the paper, we discuss in Section 2 several ways to heuristically motivate and interpret the *CM*-estimates.

2. Motivation.

Relationship to redescending M-estimates. Our initial motivation arose from first addressing the question of why the multivariate *M*-estimates break down easily. The solutions to the simultaneous equations (1.4), (1.5) and (1.6) correspond to *M*-estimates of multivariate location and scatter. Maronna's

(1976) upper bound of $1/(p + 1)$ for the breakdown point of these multivariate M-estimates presumes that $\psi(s)$ is nondecreasing. Maronna's (1976) result does not apply when $\rho(s)$ is bounded since a bounded ρ-function implies that the function $\psi(s) = 2s\rho'(s)$ must redescend. More generally, Stahel (1981) shows that the multivariate M-estimates cannot have breakdown points greater than $1/p$. What breakdown means in Stahel's (1981) context is that at least one solution to the M-estimating equations can be made to break down when the proportion of contamination in the data exceeds $1/p$, and not necessarily that all solutions can be made to break down; see Tyler (1991) for more details.

It is natural to ask if it is possible to choose a solution to (1.4), (1.5) and (1.6) in such a way so that the resulting statistics have a high breakdown point. Since the solutions to (1.4), (1.5) and (1.6) represent the critical points of (1.1), a natural choice is to choose (μ, V) which minimizes (1.1) unconditionally. This approach is an alternative to the M-estimating equation approach for defining the M-estimates of multivariate location and scatter; see, for example, Huber (1981) and Kent and Tyler (1991). For general ρ, however, it can be shown that this approach gives statistics with breakdown points of at most $1/(p + 1)$. Moreover, when ρ is bounded, the breakdown point is 0 since no values of (μ, V) exist which minimize (1.1) unconditionally. This follows since, as V approaches a singular matrix, $\log|V| \to -\infty$ and so $(1.1) \to -\infty$. This phenomenon is reminiscent of Kiefer and Wolfowitz's (1956) classical observation on the behavior of the likelihood function for a mixture of two normal distributions.

In order to define statistics based on the log-likelihood type objective function $L(\mu, V; X_n)$, it is necessary to bound V away from singularity. Introducing the constraint (1.2) does this in an affine equivariant manner and, as is to be shown, results in high breakdown point statistics which are locally robust when $\rho(s)$ is properly tuned.

Relationship to MVE- and S-estimates. Another heuristic interpretation of the CM-estimates arises by relating them to S-estimates. Suppose (1.1) is minimized under the equality constraint (1.7) rather than under the inequality constraint (1.2). This is equivalent to minimizing $\det(V)$ subject to (1.7), and this corresponds to the definition of the S-estimates of multivariate location and scatter. These S-estimates have breakdown points approximately equal to ε for $\varepsilon \le \frac{1}{2}$; see Davies (1987) and Lopuhaä (1989). Moreover, they satisfy the joint estimating equations (1.4), (1.5) and (1.7). As with the CM-estimates, the local robustness properties of an S-estimate such as its influence function and its asymptotic efficiency follow from its estimating equations. Lopuhaä (1989) observed that for a fixed ρ-function the asymptotic relative efficiency at the multivariate normal model for the S-estimate severely decreases when the value of ε increases. The reason for this trade-off between the breakdown point and the efficiency of an S-estimate is that larger values of ε tend to force the weight function in the estimating equations to be badly tuned.

Now, if one had information concerning the maximum proportion of bad observations in a data set, then one could tune an S-estimate by choosing ε accordingly. In this way, one could maintain some efficiency while still protecting against breakdown. Generally speaking, such information is contained in the data set itself, and the CM-estimates can be viewed as taking advantage of this information. The realization of a CM-estimate with a breakdown point of approximately $\frac{1}{2}$ corresponds to an S-estimate defined using the same ρ-function but with ε chosen to be equal to the realized value of ave$\{\rho(s_i)\}/\rho(\infty)$ from the CM-estimate. When sampling from a multivariate normal or a 10% symmetrically contaminated normal, this realized value can be substantially less than $\frac{1}{2}$. In this sense, the CM-estimates are adaptive versions of the S-estimates.

It is helpful to relate the above discussion to the MVE-estimates. The MVE-estimates of multivariate location and scatter are, respectively, the center and the orientation matrix associated with the minimum volume ellipsoid covering at least half of the data. They correspond to S-estimates with $\varepsilon = \frac{1}{2}$ and

$$\rho(s) = \{c \text{ for } s \geq s_0;\, 0 \text{ for } s < s_0\}, \tag{2.1}$$

where the value of s_0 simply imposes a scaling factor on the scatter matrix. The constant $c > 0$ has no effect in defining the MVE-estimate since it cancels out in constraint (1.7). The CM-estimates of multivariate location and scatter defined using the ρ-function (2.1) and with $\varepsilon = \frac{1}{2}$ are, respectively, the center and orientation matrix associated with the ellipsoid which minimizes

$$c(1 - \pi) + \tfrac{1}{2}\log(v) \tag{2.2}$$

over all ellipsoids containing at least one-half of the data with π being the proportion of the data within the ellipsoid and v being the volume of the ellipsoid. Here, c now acts as a tuning constant and does not affect the breakdown point. This CM-estimate is an adaptive version of the MVE-estimate in the following sense. If one enlarges the minimum volume ellipsoid containing one-half of the data, then both the volume of the ellipsoid and the proportion of the data within the ellipsoid increase. If nearly one-half of the data is contaminated, then a relatively small increase in the proportion results in a substantial increase in the volume. On the other hand, if the data contains relatively little contamination, then the increase in the volume relative to the increase in the proportion is modest. The CM-estimate discriminates between these two cases by balancing the proportion π and the volume v. In the first case, the CM-estimate and the MVE-estimate will essentially result in the same statistics. In the second case, the CM-estimate will be more stable than the MVE-estimate. In general, the resulting CM-estimate would correspond to an MVE-estimate but not to one covering at least one-half of the data, but rather to one covering at least a proportion π of the data, where π is the value arising from the resulting CM-estimate. The ρ-function (2.1) is not necessarily recommended, but is intended here to serve only as a pedagogical example.

3. Existence of *CM*-estimates and *CM*-functionals. We begin by addressing the question of existence. We assume throughout the paper that $\rho(s)$ satisfies the following condition. Stronger assumptions are needed in later sections.

CONDITION 3.1. *For* $s \geq 0$, $\rho(s)$ *is nondecreasing*, $0 = \rho(0) < \rho(\infty) < \infty$ *and* $\rho(s)$ *is continuous from above at* 0.

Results on existence can be made more general by first introducing the notion of *CM*-functionals. For a random vector $\mathbf{X} \in \mathbb{R}^p$, we define the *CM*-functionals for location $\mu(\mathbf{X})$ and scatter $V(\mathbf{X})$ in a manner analogous to (1.1) and (1.2); that is, $(\mu(\mathbf{X}), V(\mathbf{X}))$ minimizes over all $(\mu, V) \in \mathbb{R}^p \times \mathscr{P}_p$ the objective function

$$L(\mu, V) = E\big[\rho\{(\mathbf{X} - \mu)'V^{-1}(\mathbf{X} - \mu)\}\big] + \tfrac{1}{2}\log|V| \tag{3.1}$$

subject to the constraint

$$E\big[\rho\{(\mathbf{X} - \mu)'V^{-1}(\mathbf{X} - \mu)\}\big] \leq \varepsilon\rho(\infty), \tag{3.2}$$

with $0 < \varepsilon < 1$ being fixed. Estimating equations for the *CM*-functional are analogous to (1.3)–(1.7). More specifically, if ρ is differentiable, then $(\mu(\mathbf{X}), V(\mathbf{X}))$ must satisfy the following equations:

$$\mu = E[u(S)\mathbf{X}]/E[u(S)], \tag{3.3}$$

$$V = pE[u(S)(\mathbf{X} - \mu)(\mathbf{X} - \mu)']/E[\psi(S)], \tag{3.4}$$

and either

$$E[\psi(S)] = p \tag{3.5}$$

or

$$E[\rho(S)] = \varepsilon\rho(\infty), \tag{3.6}$$

where $S = (\mathbf{X} - \mu)'V^{-1}(\mathbf{X} - \mu)$ and u and ψ are as previously defined.

Under the following mild condition on the distribution of $\mathbf{X}$, the existence of the *CM*-functionals can be guaranteed. In addition, although Theorem 3.1 below makes no statement concerning the uniqueness of the *CM*-functionals, it does state that all solutions to the *CM*-criterion must lie in some compact set.

CONDITION 3.2. *For any hyperplane* $B \subset \mathbb{R}^p$, $P(\mathbf{X} \in B) < 1 - \varepsilon$.

THEOREM 3.1. *If Conditions* 3.1 *and* 3.2 *hold*, *then there exists a* $(\mu_0, V_0) \in \mathbb{R}^p \times \mathscr{P}_p$ *which minimizes* $L(\mu, V)$ *subject to the constraint* (3.2). *Furthermore*, *the set of all such* (μ_0, V_0) *is bounded away from* $\partial(\mathbb{R}^p \times \mathscr{P}_p)$, *the boundary set of* $\mathbb{R}^p \times \mathscr{P}_p$.

Condition 3.2 holds for any absolutely continuous distribution in $\mathbb{R}^p$. The results of the theorem also apply to the *CM*-estimates $(\hat{\mu}(X_n), \hat{V}(X_n))$, by taking the distribution of $\mathbf{X}$ to be the empirical distribution of X_n. If X_n is in general position, that is, no $(p+1)$ vectors from X_n lie in a hyperplane in $\mathbb{R}^p$, then Theorem 3.1 holds for the empirical distribution of X_n provided $n > p/(1-\varepsilon)$.

4. The uniqueness problem. Finding general conditions under which the *CM*-estimates or the *CM*-functionals are unique is a far more difficult problem than the existence problem. Establishing the uniqueness of the *CM*-functionals is particularly important since, given uniqueness, the weak continuity of the *CM*-functionals, the strong consistency of the *CM*-estimates and other properties readily follow; see Section 6. Unfortunately, as with Davies' (1987) work on the multivariate *S*-estimates, we are only able to establish the uniqueness of the *CM*-functionals under elliptically symmetric distributions, and we leave the general uniqueness problem open.

We assume throughout the rest of this section that $\mathbf{X}$ has a density in $\mathbb{R}^p$ of the form

$$f(x;t,\Sigma) = |\Sigma|^{-1/2} g\{(x-t)'\Sigma^{-1}(x-t)\} \tag{4.1}$$

for some function $g\colon \mathbb{R}^+ \to \mathbb{R}^+$ and where $(t,\Sigma) \in \mathbb{R}^p \times \mathscr{P}_p$. Properties of elliptically symmetric distribution have been studied in detail by Kelker (1970); see also Muirhead (1982). The parameter t corresponds to the expected value of $\mathbf{X}$ if it exists. If the second moments of $\mathbf{X}$ exist, then the variance–covariance matrix of $\mathbf{X}$ is proportional to Σ. One would anticipate that, for the *CM*-functionals, $\mu(\mathbf{X}) = t$ and $V(\mathbf{X})$ is proportional to Σ. This is shown in Theorem 4.1 below.

CONDITION 4.1.

(a) *$g(s)$ is nonincreasing for $s \ge 0$.*

(b) *There exists an $s_0 > 0$ such that, for all $s_1 < s_0 < s_2$, $\rho(s_1) < \rho(s_0) < \rho(s_2)$ and $g(s_1) > g(s_0) > g(s_2)$.*

THEOREM 4.1. *Assume $\mathbf{X}$ has a density in $\mathbb{R}^p$ of the form* (4.1). *Under Condition* 3.1, *there exists a solution to minimizing* (3.1) *subject to the constraint* (3.2). *Furthermore, if Condition* 4.1 *also holds, then any such solution must be of the form* $(\mu(\mathbf{X}, V(\mathbf{X})) = (t, \lambda\Sigma)$ *for some* $0 < \lambda < \infty$.

Although the results of Theorem 4.1 seem intuitive and natural, the proof is surprisingly delicate. The proof itself is heavily reliant on Davies' (1987) results on the uniqueness of the *S*-functionals of multivariate location and scatter at elliptically symmetric distributions. We refer the reader to Davies' (1987) paper for a better appreciation of the difficulties arising when trying to establish uniqueness results. The condition that g be nonincreasing and that $\rho(s)$ and $g(s)$ have a point of common change (i.e., Condition 4.1) is needed to apply Davies' (1987) results. We note here that Theorem 4.1 does not imply that all solutions to the corresponding estimating equations (3.3)–(3.6) are necessarily of the form $(t, \lambda\Sigma)$.

To complete the discussion of the uniqueness of the *CM*-functionals under elliptically symmetric distributions, we must establish when λ in Theorem 4.1 is unique. Now, any such value of λ is a solution which minimizes

$$l(\lambda; g) = E\{\rho(S/\lambda)\} + \tfrac{1}{2}p \log \lambda \tag{4.2}$$

subject to the constraint $\lambda \geq \lambda_L$, where λ_L is the unique solution to

$$E\{\rho(S/\lambda)\} = \varepsilon\rho(\infty) \tag{4.3}$$

and $S = (\mathbf{X} - t)'\Sigma^{-1}(\mathbf{X} - t)$ has density

$$f(s) = C_p s^{p/2-1} g(s), \tag{4.4}$$

with C_p being a constant dependent on p. Note that since $\rho(s)$ is nondecreasing, $E[\rho(S/\lambda)] \leq \varepsilon\rho(\infty)$ for $\lambda \geq \lambda_L$. Uniqueness of the *CM*-functionals for elliptically symmetric distributions now follows assuming the following condition holds. For a given ρ-function, a given value of ε and a given function g, this condition can in general be checked numerically since $l(\lambda; g)$ is a univariate function, as will be demonstrated in Section 9.

CONDITION 4.2. *There exists a unique* $0 < \lambda_0 < \infty$ *which minimizes* (4.2) *over* $\lambda \geq \lambda_L$.

COROLLARY 4.1. *Under Conditions* 3.1, 4.1 *and* 4.2, *there exists a unique solution to minimizing* (3.1) *subject to the constraint* (3.2). *Specifically*, $(\mu(\mathbf{X}), V(\mathbf{X})) = (t, \lambda_0\Sigma)$ *with* λ_0 *defined as in Condition* 4.2.

If a distribution is close to an elliptically symmetric distribution, then the weak continuity of the *CM*-functionals given by Theorem 6.1 ensures that even if the *CM*-functional is not unique at the distribution, then at least all possible solutions to the *CM* criterion are within some small neighborhood of each other. Likewise, if we have a random sample from an elliptically symmetric distribution, then the strong consistency of the *CM*-estimates, which is established in Section 6, ensures that all possible solutions to the *CM* criterion must lie within some shrinking compact set as the sample size goes to ∞. The uniqueness of the *CM*-estimates themselves is not necessary in establishing strong consistency, asymptotic normality or in deriving its finite sample breakdown point.

5. Finite sample breakdown point. In this section, we derive the breakdown point of the *CM*-estimates, using the concept of the finite sample replacement breakdown point introduced by Donoho and Huber (1983). For the multivariate location–scatter problem, this concept can be defined as follows: suppose $m \leq n$ arbitrary data points $Y = \{\mathbf{y}_1, \ldots, \mathbf{y}_m\}$ replace m data points from the original data $X_n = \{\mathbf{x}_1, \ldots, \mathbf{x}_n\}$ producing a corrupted sample Z which consists of a fraction of $\varepsilon_m = m/n$ bad values. For a given ε_m, the statistic $(\hat{\mu}(\cdot), V(\cdot))$ is said to break down at X_n under ε_m-contamination if for *at least one solution* $(\hat{\mu}(Z), \hat{V}(Z))$ $\|\hat{\mu}(Z)\|$ can be made arbitrarily large, the largest eigenvalue of $\hat{V}(Z)$ can be made arbitrarily large or the smallest eigenvalue of $\hat{V}(Z)$ can be made arbitrarily close to 0, for varying choices of Y and of the m replaced data points from X_n. The finite sample replacement breakdown point of $(\hat{\mu}(\cdot), \hat{V}(\cdot))$ at X_n is then defined to be $\varepsilon^*(X_n)$, the minimum of all ε_m causing breakdown. Note that the definition of breakdown here does not depend on $(\hat{\mu}(\cdot), \hat{V}(\cdot))$ being uniquely defined.

For X_n in general position, Davies (1987) gives a strict upper bound for the finite sample replacement breakdown point $\varepsilon^*(X_n)$ for any affine equivariant location and scatter statistics, namely

$$\varepsilon^*(X_n) \le \lceil (n-p)/2 \rceil / n, \tag{5.1}$$

where $\lceil k \rceil$ represents the smallest integer greater than or equal to k if positive and 0 otherwise. By choosing $\varepsilon = (n-p)/(2n)$ in (1.2), the *CM*-estimate $(\hat{\mu}(\cdot), \hat{V}(\cdot))$ obtains this upper bound, as is seen from the following general result.

THEOREM 5.1. *If Condition* 3.1 *holds, then, for* X_n *in general position,* $0 < \varepsilon < 1$, *and* $n > p/(1-\varepsilon)$,

$$\varepsilon^*(X_n) = \min\{\lceil n\varepsilon \rceil / n, \lceil n(1-\varepsilon) - p \rceil / n\}.$$

6. Consistency, influence functions and asymptotic normality. Unless otherwise stated, we assume throughout this section that the p-dimensional random vector $\mathbf{X}$ has an absolutely continuous distribution in $\mathbb{R}^p$. The existence of the *CM*-functional $(\mu(\mathbf{X}), V(\mathbf{X}))$ is assured by Theorem 3.1. The results of this section also are dependent on the *CM*-functional being uniquely defined. Although we have only been able to establish the uniqueness of the *CM*-functional at elliptically symmetric distributions, rather than assuming $\mathbf{X}$ has an elliptically symmetric distribution, we will assume throughout this section that they are unique at $\mathbf{X}$. We make this general uniqueness assumption here since, given uniqueness, most other important properties follow. One such property is the weak continuity of the *CM*-functionals, which we establish in Theorem 6.1 below. We also need some conditions on the ρ-function which are stronger than Condition 3.1.

CONDITION 6.1. *The CM-functional* $(\mu(\mathbf{X}), V(\mathbf{X}))$ *is uniquely defined at* $\mathbf{X}$.

CONDITION 6.2. *For* $s \ge 0$, $\rho(s)$ *is continuous, nondecreasing when* $0 = \rho(0) < \rho(\infty) < \infty$ *and strictly increasing when* $0 < \rho(s) < \rho(\infty)$.

THEOREM 6.1. *Under Conditions* 6.1 *and* 6.2, *if* $\mathbf{X}_k \to \mathbf{X}$ *in distribution, then* $(\mu(\mathbf{X}_k), V(\mathbf{X}_k))$ *exists for large* k *and* $(\mu(\mathbf{X}_k), V(\mathbf{X}_k)) \to (\mu(\mathbf{X}), V(\mathbf{X}))$ *for any* $(\mu(\mathbf{X}_k), V(\mathbf{X}_k))$ *satisfying the definition of the CM-functional at* $\mathbf{X}_k$.

The condition that $\rho(s)$ is strictly increasing when $0 < \rho(s) < \rho(\infty)$ plays a role in the proof of Theorem 6.1 only for certain distributions on $\mathbf{X}$. Also, the presumption that $\mathbf{X}$ is absolutely continuous can essentially be relaxed to Condition 3.2. These subtle technicalities are discussed in the Appendix; see Remark 10.1.

Strong consistency of the *CM*-estimates follows readily from Theorem 6.1 since the empirical distribution function converges to the underlying distribution function almost surely. Thus, if $\mathbf{x}_1, \ldots, \mathbf{x}_n$ are i.i.d. $\mathbf{X}$, then

$$(6.1) \qquad \left(\hat{\mu}(X_n), \hat{V}(X_n) \right) \to \left(\mu(\mathbf{X}), V(\mathbf{X}) \right) \quad \text{almost surely}$$

for any sequence $(\hat{\mu}(X_n), \hat{V}(X_n))$ satisfying the definition of the *CM*-estimates.

Given weak continuity and consistency, local properties of the *CM*-estimates such as the influence functions and the asymptotic distributions follow from the estimating equations (3.3)–(3.6). To obtain general expressions for the influence functions and the asymptotic distributions, we note that (μ, V) consists of $m = p + \frac{1}{2}p(p+1)$ parameters and so let $\theta \in \mathbb{R}^m$ represent an m-dimensional vector parameterization of (μ, V) with μ being the first p components of $\boldsymbol{\theta}$ and with the upper triangular part of V being the remaining $\frac{1}{2}p(p+1)$ components. Analogously, the *CM*-functionals $(\mu(\mathbf{X}), V(\mathbf{X}))$ can be represented by $\boldsymbol{\theta}(\mathbf{X}) \in \mathbb{R}^m$. Equations (3.3), (3.4) and (3.5) together, or (3.3), (3.4) and (3.6) together, consist of m implicit equations for $\boldsymbol{\theta}$ since (3.4) itself contains only $\frac{1}{2}p(p+1) - 1$ implicit equations. This observation concerning (3.4) follows from the symmetry of V and by noting that multiplying both sides of (3.4) by V^{-1} and then taking the trace gives $p = p$. Thus, there exists a function $\Psi_1\colon \mathbb{R}^p \times \mathbb{R}^m \to \mathbb{R}^m$ such that (3.3), (3.4) and (3.5) together are equivalent to

$$(6.2) \qquad E\left[\Psi_1(\mathbf{X}, \boldsymbol{\theta}) \right] = \mathbf{0},$$

and there exists a function $\Psi_2\colon \mathbb{R}^p \times \mathbb{R}^k \to \mathbb{R}^k$ such that (3.3), (3.4) and (3.6) together are equivalent to

$$(6.3) \qquad E\left[\Psi_2(\mathbf{X}, \boldsymbol{\theta}) \right] = \mathbf{0},$$

with the first $(k-1)$ entries of Ψ_1 and Ψ_2 being the same.

The local properties of the *CM*-estimates are now obtained by using standard expansion arguments on the estimating equations. To do this, we need to ensure that the expansions are valid and so hereafter we assume the following two conditions hold.

CONDITION 6.3. *The function $\rho(s)$ has a continuous second derivative, and $u(s)$, $\psi(s)$ and $\psi'(s)$ are bounded.*

CONDITION 6.4. *For the CM-functional $(\mu(\mathbf{X}), V(\mathbf{X}))$,* (3.5) *and* (3.6) *do not both hold at* $\mathbf{X}$.

Another way of phrasing Condition 6.4 is that if the global minimum of (3.1), subject to the constraint (3.2), occurs on the boundary of the constraint, then it is assumed not to be a critical point of (3.1). This property carries over to distributions near the distribution of $\mathbf{X}$.

THEOREM 6.2. *Under Conditions* 6.1, 6.2 *and* 6.4, *if* $\mathbf{X}_k \to \mathbf{X}$ *in distribution, then* (3.5) *and* (3.6) *cannot both hold for* $(\mu(\mathbf{X}_k), V(\mathbf{X}_k))$ *for large enough* k. *Furthermore, for large* k, *if* (3.5) *holds for* $\mathbf{X}$, *then* (3.5) *holds for* $\mathbf{X}_k$ *and if* (3.6) *holds for* $\mathbf{X}$, *then* (3.6) *holds for* $\mathbf{X}_k$.

The importance of Theorem 6.2 is that it allows us to treat (6.2) and (6.3) as two separate cases when studying the local properties of the *CM*-estimates. For convenience, we treat both cases together by defining $\Psi = \Psi_1$ if (3.5) holds and $\Psi = \Psi_2$ if (3.6) holds.

The influence function of the functional $\boldsymbol{\theta}(\mathbf{X})$, the $\mathbb{R}^m$ representation of $(\mu(\mathbf{X}), V(\mathbf{X}))$, at $\mathbf{X}$ is defined by

$$\mathrm{IF}(\mathbf{x}; \boldsymbol{\theta}(\mathbf{X})) = \lim_{\varepsilon \to 0^+} \{\boldsymbol{\theta}(\mathbf{X}_\varepsilon) - \boldsymbol{\theta}(\mathbf{X})\}/\varepsilon, \tag{6.4}$$

provided the limit exists, where $\mathbf{X} \sim F$ and $\mathbf{X}_\varepsilon \sim F_\varepsilon = (1 - \varepsilon)F + \varepsilon\delta_x$ with δ_x being the distribution function of the constant x.

THEOREM 6.3. *Let* $\lambda(\boldsymbol{\theta}) = E[\Psi(\mathbf{X}; \boldsymbol{\theta})]$ *and suppose* $\lambda(\boldsymbol{\theta})$ *has a nonsingular derivative* $\Lambda = \partial\lambda(\boldsymbol{\theta})/\partial\boldsymbol{\theta}$ *at* $\boldsymbol{\theta}_0 = \boldsymbol{\theta}(\mathbf{X})$. *Under Conditions* 6.1 *through* 6.4, *the influence function of* $\boldsymbol{\theta}(\mathbf{X})$ *exists and is given by* $IF(\mathbf{x}; \boldsymbol{\theta}(\mathbf{X})) = -\Lambda^{-1}\Psi(\mathbf{x}; \boldsymbol{\theta}_0)$.

The proof of Theorem 6.3 involves standard expansion arguments, as does the proof of the following theorem concerning the asymptotic normality of the *CM*-estimates. Whenever (3.6) holds, the influence function given in Theorem 6.3 and the asymptotic distribution given in Theorem 6.4 below are the same as the multivariate *S*-estimates defined using the same ρ-function and the same value of ε; see Lopuhaä (1989).

THEOREM 6.4. *Under Conditions* 6.1, 6.2, 6.3 *and* 6.4, *if* $\mathbf{x}_1, \dots, \mathbf{x}_n$ *are i.i.d.* $\mathbf{X}$, *then* $\sqrt{n}\{\hat{\theta}(X_n) - \theta(\mathbf{X})\}$ *converges in distribution to a multivariate normal with mean* $\mathbf{0}$ *and variance–covariance matrix* $\Lambda^{-1}M\Lambda'^{-1}$, *where* M *is the variance–covariance matrix of* $\Psi(\mathbf{X}; \theta_0)$ *and* Λ *is defined as in Theorem* 6.3.

The form of the influence function in Theorem 6.3 and the form of the asymptotic normal distribution in Theorem 6.4 simplify whenever the distribution of $\mathbf{X}$ is symmetric about some vector μ_0. For this case, $\mu(\mathbf{X}) = \mu_0$ and the matrix Λ is block diagonal in the sense that when the first p parameters of $\boldsymbol{\theta}$ are taken to correspond to μ, the upper right $p \times (m - p)$ entries of Λ and the lower right $(m - p) \times p$ entries of Λ are 0. The matrix M is similarly block diagonal, and so $\sqrt{n}\{\hat{\mu}(X_n) - \mu(\mathbf{X})\}$ and $\sqrt{n}\{\hat{V}(X_n) - V(\mathbf{X})\}$ are asymptotically independent. The influence function given in Theorem 6.3 and the asymptotic variance–covariance matrix given in Theorem 6.4 simplify further whenever $\mathbf{X}$ has an elliptically symmetric distribution.

7. Behavior under elliptical distributions. In this section, we assume $\mathbf{X}$ has a multivariate normal or some other elliptically symmetric density in $\mathbb{R}^p$; see (4.1). We then know by Theorem 4.1 that under general conditions the *CM*-functional $(\mu(\mathbf{X}), V(\mathbf{X}))$ is uniquely defined. Further, we note that Condition 6.3, which is assumed in establishing the asymptotic normality of the *CM*-estimates and in deriving the influence functions of the *CM*-functionals, can be expressed in the simpler form given by Condition 7.1 below. Here, λ_0, λ_L and $l(\lambda; g)$ are defined as in Section 4. As with Condition 4.2, for a given ρ-function, a given value of ε and a given function g, this condition can in general be checked numerically.

CONDITION 7.1. *If* $\lambda_0 = \lambda_L$, *then* λ_L *is neither a local minimum nor an inflection point of* $l(\lambda; g)$.

Now, since $\mathrm{IF}(x; \theta(\mathbf{X}))$ is block diagonal, we can treat the influence function for $\mu(\mathbf{X})$ and $V(\mathbf{X})$ separately. Using standard invariance arguments [see, e.g., Lopuhaä (1989) or Huber (1981)], we have

$$\mathrm{IF}(x; t(\mathbf{X})) = \lambda_0^{1/2} h_1(s/\lambda_0)\Sigma^{1/2} z, \tag{7.1}$$

$$\mathrm{IF}(x; V(\mathbf{X})) = \lambda_0 h_2(s/\lambda_0)\Sigma^{1/2}(zz' - p^{-1}I)\Sigma^{1/2} + \lambda_0 h_3(s/\lambda_0)\Sigma, \tag{7.2}$$

where $s = (x-t)'\Sigma^{-1}(x-t)$, $z = \Sigma^{-1/2}(x-t)/\sqrt{s}$ and h_1, h_2 and h_3 are scalar-valued functions dependent on ρ and g. Specifically,

$$h_1(w) = w^{1/2}u(w)/\beta_1, \tag{7.3}$$

$$h_2(w) = (p+1)\psi(w)/\beta_2, \tag{7.4}$$

$$h_3(w) = \begin{cases} 2(\rho(w) - \varepsilon\rho(\infty))/\beta_3, & \text{if } \lambda_0 = \lambda_L, \\ 2(\psi(w)/p - 1)/(\beta_2 - \beta_3), & \text{if } \lambda_0 > \lambda_L, \end{cases} \tag{7.5}$$

where

$$\beta_1 = E\left[u\left(\frac{S}{\lambda_0}\right) + \frac{2}{p}\frac{S}{\lambda_0}u'\left(\frac{S}{\lambda_0}\right)\right], \tag{7.6}$$

$$\beta_2 = E\left[\psi\left(\frac{S}{\lambda_0}\right) + \frac{2}{p}\frac{S}{\lambda_0}\psi'\left(\frac{S}{\lambda_0}\right)\right], \tag{7.7}$$

$$\beta_3 = E\left[\psi\left(\frac{S}{\lambda_0}\right)\right]. \tag{7.8}$$

In this section we are treating $V \in \mathscr{P}_p$ as a p^2 parameter [which therefore includes $\frac{1}{2}p(p-1)$ duplications], and so (7.2) is a $p \times p$ symmetric matrix.

For the asymptotic distribution of the *CM*-estimates, we know that $\sqrt{n}\{\hat{\mu}(X_n) - \mu(\mathbf{X})\}$ and $\sqrt{n}\{\hat{V}(X_n) - V(\mathbf{X})\}$ are asymptotically independent.

For the location estimate, we have by Theorem 6.3 and the equivariance properties of $\hat{\mu}(X_n)$ and $\mathbf{X}$ that

$$\sqrt{n}\,\{\hat{\mu}(X_n) - \mu(\mathbf{X})\} \to \text{Normal}_p(\mathbf{0}, \alpha\Sigma), \tag{7.9}$$

where

$$\alpha = E\big[u^2(S_1/\lambda_0)S\big]/\big(p\beta_1^2\big). \tag{7.10}$$

For the scatter estimate, we have

$$\text{vec}\Big\{\sqrt{n}\,\big(\hat{V}(X_n) - V(\mathbf{X})\big)\Big\} \to \text{Normal}_{p^2}(0, \boldsymbol{\Gamma}), \tag{7.11}$$

where

$$\Gamma = \lambda_0^2\sigma_1\big(I + K_{p,p}(\Sigma \otimes \Sigma)\big) + \lambda_0^2\sigma_2\ \text{vec}(\Sigma)\text{vec}(\Sigma)', \tag{7.12}$$

with

$$\sigma_1 = \left(1 + \frac{2}{p}\right)E\left[\psi^2\left(\frac{S}{\lambda_0}\right)\right]\bigg/\beta_2^2, \tag{7.13}$$

$$\sigma_2 = E\left[h_3^2\left(\frac{S}{\lambda_0}\right)\right] - \frac{2}{p}\sigma_1. \tag{7.14}$$

The notation in (7.11) and (7.12) is standard notation which arises when describing the asymptotic distribution of random matrices. For a $p_1 \times p_2$ matrix A, vec(A) represents the $p_1 p_2 \times 1$ vector formed by stacking the columns of A. The Kronecker product $\otimes$ between a $p_1 \times p_2$ matrix A and a $p_3 \times p_4$ matrix B is the $p_1 p_3 \times p_2 p_4$ matrix $A \otimes B = \{a_{ij}B\}$ with i ranging over rows of matrices and j ranging over columns of matrices. The commutation matrix $K_{p,p}$ is the $p^2 \times p^2$ matrix $K_{p,p} = \Sigma e_i e_j' \otimes e_j e_i'$ with the sum being over $i, j = 1, 2, \dots, p$ and where $\mathbf{e}_i \in \mathbb{R}^p$ has a 1 in the ith position and 0's otherwise; see, for example, Muirhead (1982), Lopuhaä (1989) or Tyler (1983).

8. Indices of local robustness. The behavior of the *CM*-estimates obviously depends on the choice of the ρ-function. For a given ρ-function, one way of assessing the local robustness of the resulting *CM*-estimate is to consider the asymptotic relative efficiency of the estimate over a range of possible distributions. A simpler and perhaps more appealing way of assessing the local robustness of the estimate is to consider the behavior of its influence function along with its relative asymptotic efficiency at a specific distribution, in particular the multivariate normal distribution or some other elliptically symmetric distribution. We use the latter approach here in developing simple indices for accessing the local robustness of the *CM*-estimates.

Hereafter we assume the data $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n$ represents a random sample from an elliptically symmetric distribution. Consider first the location component. Although the asymptotic variance–covariance matrix of $\hat{\mu}(\mathbf{X}_n)$ is matrix valued, we note from (7.9) that the asymptotic relative efficiency of the location component for varying choices of ρ can be compared by simply comparing the corresponding values of the scalar α given in (7.10). Comparing the influence functions of the location component of *CM*-functions for varying choices of ρ is more complicated since the influence functions are vector-valued functions which differ by more than a simple scalar multiple. We consider instead the most common measure of local robustness based on the influence function, namely the gross error sensitivity (GES). In general, the GES is defined to be $\sup_{\mathbf{x} \in \mathbb{R}^p} \|\mathrm{IF}(\mathbf{x}; \cdot)\|$ for some norm $\|\cdot\|$. For the location functional we note from (7.1) that regardless of the choice of the norm the GES is proportional to the scalar

$$(8.1) \qquad G_1 = (\lambda_0)^{1/2} K_1/\beta_1,$$

where $K_1 = \sup_{s>0} s^{1/2} u(s)$. If the norm is taken to be the p-dimensional Euclidean norm, then (8.1) is equal to the gross error sensitivity when $\Sigma = I$.

Consider now the scatter component. The asymptotic variance–covariance matrix of the scatter component, given by (7.12), depends on the ρ-function through two scalar quantities $\lambda_0^2 \sigma_1$ and $\lambda_0^2 \sigma_2$. It is helpful here to separate the scatter component into a shape component and a scale component. By a shape component, we mean any function of $V(\mathbf{X})$ which is invariant under a positive scalar multiple, such as $V(\mathbf{X})/\mathrm{trace}\{V(\mathbf{X})\}$, and by a scale component we mean any function of $V(\mathbf{X})$ which is equivariant under a positive scalar multiple, such as $\mathrm{trace}\{V(\mathbf{X})\}$. In practice, one is often interested in scatter only through its shape component. For example, the ratio of principal component roots, principal component vectors, correlations, canonical vectors and correlations, multiple correlation coefficients, multivariate linear regression coefficients, and the ratio of variances are all shape components. Moreover, the scale component is essentially an ill-defined nuisance parameter. For elliptically symmetric distributions, the scale component for Σ is confounded with the function g in (4.1). Thus, we choose here to focus only on the shape of the scatter component, for which we have the following result concerning their asymptotic distributions. For $p \geq 2$, if $H\colon \mathscr{P}_p \to \mathbb{R}^q$ is continuously differentiable and $H(V) = H(\lambda V)$ for any $\lambda > 0$, then

$$(8.2) \qquad \sqrt{n}\left\{H\left[\hat{V}(X_n)\right] - H\left[V(\mathbf{X})\right]\right\} \to_d \mathrm{Normal}_p(\mathbf{0}, \sigma_1 \Gamma_H(\Sigma)),$$

where $\Gamma_H(\cdot)$ is a specific function on $\mathscr{P}_p$ which is not dependent on ρ, g, $\mathbf{t}$ or Σ, and for which $\Gamma_H(\Sigma) = \Gamma_H(\lambda\Sigma)$ for all $\Sigma \in \mathscr{P}_p$ and $\lambda > 0$; see Tyler (1983) for more detail. We note from (8.2) that the asymptotic relative efficiency of *CM*-estimates of shape for varying choices of ρ can be compared by simply comparing the corresponding values of σ_1. For $p = 1$, the concept of shape is obviously meaningless.

Comparisons of gross error sensitivity for the shape of the scatter component can also be reduced to a scalar comparison. If the function H is defined as in (8.2), then we show in the Appendix that

$$(8.3) \qquad \mathrm{IF}(\mathbf{x}; H(V(\mathbf{X}))) = h_2(s/\lambda_0)\gamma_H(\Sigma)(\mathbf{z} \otimes \mathbf{z}),$$

where $\gamma_H(\cdot)$ is a $q \times p^2$ matrix-valued function dependent on H but not on ρ, g, $\mathbf{t}$ or Σ, with the property $\gamma_H(\lambda\Sigma) = \gamma_H(\Sigma)$ for any $\lambda > 0$. Thus, regardless of the norm chosen for the influence function, we see that the GES for the shape functional $H(V(\mathbf{X}))$ is proportional to the scalar

$$G_2 = K_2/\beta_2, \tag{8.4}$$

where $K_2 = \sup_{s>0} \psi(s)$. For example, the influence function of $V(\mathbf{X})/\text{trace}\{V(\mathbf{X})\}$ is

$$h_2(s/\lambda_0)\left[\Sigma^{1/2}\mathbf{z}\mathbf{z}'\Sigma^{1/2} - \{\mathbf{z}'\Sigma\mathbf{z}/\text{trace}(\Sigma)\}\Sigma\right]/\text{trace}(\Sigma). \tag{8.5}$$

If we use the p^2-dimensional Euclidean norm in defining the GES of $V(\mathbf{X})/\text{trace}\{V(\mathbf{X})\}$, then when Σ is proportional to I the GES equals $a_p G_2$, where $a_p = (1 + 2/p)(1 - 1/p)^{1/2}$.

Summarizing, our method for judging the local robustness of the *CM*-estimates associated with different ρ-functions is to consider the corresponding values of α, G_1, σ_1 and G_2 at some specific elliptically symmetric distribution of interest. We note though that the scalars α, G_1, σ_1 and G_2 are dependent on the underlying elliptically symmetric distribution only through the function g in (4.1) and not through $\mathbf{t}$ or Σ. Thus, for a given g and ρ, the scalars α and σ_1 serve as indices for the asymptotic efficiency of the location and shape components, respectively, of the associated *CM*-estimate for the entire location–scatter elliptically symmetric family given by (4.1). An analogous statement holds for G_1 and G_2 as indices of gross error sensitivity. A good ρ-function is one for which α, G_1, σ_1 and G_2 are as small as possible.

9. Tuning the *CM*-estimates. The broad problem of choosing a ρ-function from the class of all possible ρ-functions is not addressed here. Instead, for a given function ρ, we introduce a multiplicative tuning constant $c > 0$ in order to generate and study the class of ρ-functions defined by

$$\mathscr{C}_\rho = \{\rho_c(s) \mid \rho_c(s) = c\rho(s),\ c > 0\}. \tag{9.1}$$

From Theorem 5.1, we note that the breakdown point of the *CM*-estimate associated with $\rho_c(s)$ does not depend on the tuning constant c. The value of c though does affect the influence functions and the asymptotic distributions of the *CM*-estimates, and hence the local robustness indices α, G_1, σ_1 and G_2.

For the multivariate normal model or some other elliptical model of interest, one can plot as a function of c the indices of α, G_1, σ_1 and G_2 in order to determine a desirable value for c. A plot of the indices of α, G_1, σ_1 and G_2 can be simplified by noting that they depend on the tuning constant c only through the value of λ_0, and, moreover, the lower bound λ_L for λ_0 does not depend on c. Thus, we suggest plotting α, G_1, σ_1 and G_2 as a function of λ_0 rather than as a function of c. Based on these plots, one can determine a desirable value for λ_0 and then find the value of c, if any, which produces this value of λ_0. Since ρ is increasing, it can be shown that λ_0 is a nondecreasing function of c. For $\lambda_0 > \lambda_L$, the only possible value of c is

$$c = p/E[\psi(S/\lambda_0)], \tag{9.2}$$

where, as before, $\psi(s) = 2s\rho'(s)$ and S has density (4.4). This follows since if $\lambda_0 > \lambda_L$, then λ_0 must be a critical point of

$$l_c(\lambda; g) = E[\rho_c(S/\lambda)] + \tfrac{1}{2}p \log(\lambda), \tag{9.3}$$

and the critical points of $l_c(\lambda; g)$ must satisfy $E[\psi_c(S/\lambda)] = p$, where $\psi_c(s) = 2s\rho_c'(s) = c\psi(s)$. After obtaining c from (9.2), it is necessary to check that the desired value of λ_0 does indeed minimize $l_c(\lambda; g)$ subject to the constraint $\lambda \geq \lambda_L$. To obtain $\lambda_0 = \lambda_L$, one can choose any $c \leq p/E[\psi(S/\lambda_L)]$.

The formulas for the influence functions and the asymptotic distributions of the *CM*-estimates depend on the value of ε only through the lower bound λ_L. Since ρ is nondecreasing, λ_L is a nonincreasing function of ε. For a fixed value of λ_0, the influence function and the asymptotic distribution of the scatter component have a different form depending on whether $\lambda_0 > \lambda_L$ or $\lambda_0 = \lambda_L$. However, we note from (7.1), (7.9), (8.2) and (8.3) that this is not the case for the location component or the shape part of the scatter component. Thus, if for a given $\varepsilon < \frac{1}{2}$ we determine a desirable value of $\lambda_0 \geq \lambda_L$, then one can always choose $\varepsilon = \frac{1}{2}$ instead and still use the desirable value of λ_0. In this way we obtain a breakdown point of approximately $\frac{1}{2}$ in large samples without affecting the local robustness properties of the *CM*-estimates of location and shape.

In contrast, Lopuhaä (1989) observes that the breakdown point and the local robustness properties of the multivariate *S*-estimates are heavily related. Recall that the influence functions and the asymptotic distribution of the *CM*-estimates when $\lambda_0 = \lambda_L$ correspond to those of the *S*-estimates. The relationship between the large sample breakdown point of the *S*-estimate and the value of λ_0 is given by

$$\varepsilon = E[\rho(S/\lambda_0)]/\rho(\infty). \tag{9.4}$$

The *S*-estimates can be tuned by either λ_0 or ε, but once one is specified, the other is determined.

To illustrate these ideas, we assume a multivariate normal model and consider the biweighted *CM*-estimates, which we defined by choosing $\varepsilon = \frac{1}{2}$ and

$$\rho_c(s) = c\begin{cases} 3s - 3s^2 + s^3, & s \leq 1, \\ 1, & s > 1, \end{cases} \tag{9.5}$$

with $c > 0$. These ρ-functions are also studied by Lopuhaä (1989) in connection with the *S*-estimates of multivariate location and scatter. If we normalize the scatter matrix so that it is consistent for the covariance matrix of the multivariate normal distribution, that is, define $\Sigma = V/\lambda_0$, then the weight function $u_c(s_i) = 2\rho_c'(s_i)$ in (1.4) can be replaced by $u(r_i; \sigma_0)$, where $r_i = \{(\mathbf{x}_i - \mu)'\Sigma^{-1}(\mathbf{x}_i - \mu)\}^{1/2}$, $\sigma_0 = \lambda_0^{1/2}$ and

$$u(r; \sigma_0) = \left(1 - (r/\sigma_0)^2\right) I(|r| < \sigma_0). \tag{9.6}$$

The weight function $u(r; \sigma_0)$ corresponds to Tukey's biweight function with tuning constant σ_0.

The values of α, G_1, σ_1 and G_2 can be expressed in terms of ratios of linear combinations of cumulative chi-square probabilities with varying degrees of freedom. These expressions are fairly easy to derive but are not very tractable. We give them in the Appendix; see (A.6)–(A.9). It is better for our purposes here to use graphs. Graphs of α, G_1, σ_1 and G_2 as functions of λ_0 for $p = 2$, 5 and 10 are given in Figures 2, 3 and 4, respectively, of the Appendix. The values of λ_L for $p = 2$, 5 and 10 are 7.08, 21.64 and 45.91, respectively. The values of α, G_1, σ_1 and G_2 at λ_L coincide with those of the biweighted S-estimates of Lopuhaä (1989) which has a limiting breakdown point of $\frac{1}{2}$. We again emphasize that for the S-estimates with $\varepsilon = \frac{1}{2}$, one has no control of the biweight tuning constant in (9.6). The limiting values of α, G_1, σ_1 and G_2 as $\lambda_0 \to \infty$ correspond to those of the sample mean vector and sample covariance matrix, namely $\alpha = 1$, $G_1 = \infty$, $\sigma_1 = 1$ and $G_2 = \infty$.

For $p = 2$, we note from the graphs that it is possible to substantially improve upon both the asymptotic relative efficiencies and the gross error sensitivities of the biweighted S-estimate by using an appropriately chosen value for λ_0. For example, by choosing $\lambda_0 = 17$, we obtain values of $\alpha = 1.130$, $G_1 = 1.927$, $\sigma_1 = 1.243$ and $G_2 = 1.369$ for the CM-estimate compared to values of $\alpha = 1.725$, $G_1 = 2.390$, $\sigma_1 = 2.646$ and $G_2 = 1.673$ for the S-estimate. In order to obtain the value $\lambda_0 = 17$, we use (9.2) to determine that we must choose $c = 4.627$. A graph of $l_c(\lambda; g)$ for $c = 4.627$ is given in Figure 1. From the graph we can verify that Conditions 4.2 and 7.1 hold and that λ_0 does equal 17. From (9.4) we note that a biweighted S-estimate having the same values of α, G_1, σ_1 and G_2 as the CM-estimate with $c = 4.627$ has a limiting breakdown point of 0.280. Similar observations arise from considering the graphs for $p = 5$.

For $p = 10$, we again note that an appropriately chosen biweighted CM-estimate of location improves upon both the gross error sensitivity and the asymptotic efficiency of the biweighted S-estimate of location. However, the CM-estimates cannot improve upon the gross error sensitivity for shape, that is, upon G_2, of the S-estimate. To improve upon the asymptotic efficiency of the S-estimate of shape, a trade-off with gross error sensitivity must be made. For example, by choosing $\lambda_0 = 60$, we obtain values of $\alpha = 1.037$,

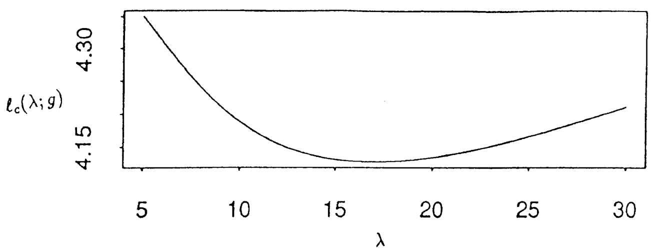

FIG. 1. *Plot of* (9.3) *at the $p = 2$ dimensional standard normal distribution for the biweighted CM-estimate with $c = 4.627$.*

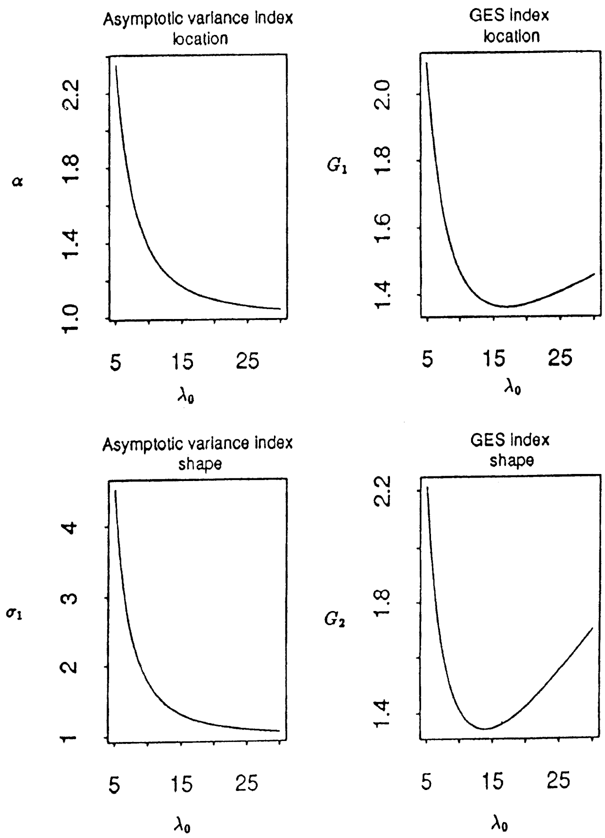

FIG. 2. *Indices of local robustness biweighted CM-estimates*: $p = 2$.

$G_1 = 3.428$, $\sigma_1 = 1.048$ and $G_2 = 1.244$ for the *CM*-estimate compared to values of $\alpha = 1.072$, $G_1 = 3.482$, $\sigma_1 = 1.093$ and $G_2 = 1.142$ for the *S*-estimate. The value $\lambda_0 = 60$ corresponds to choosing $c = 15.464$. Again, a graph of $l_c(\lambda; g)$ for $c = 15.464$ can be made to verify that Conditions 4.2 and 7.1 hold and that λ_0 does equal 60. Also, from (9.4) we note that a biweighted *S*-estimate having the same values of α, G_1, σ_1 and G_2 as the *CM*-estimate with $c = 15.464$ would have a limiting breakdown point of 0.408.

If we let $\lambda_0 = \lambda_L/\tau$ with $0 < \tau \leq 1$ and then let $p \to \infty$, we get $\alpha \to 1$, $G_1/p^{1/2} \to G_1^\infty = 0.64(\tau\gamma_0)^{-1/2}(1 - \tau\gamma_0)^{-2}/5^{1/2}$, $\sigma_1 \to 1$ and $G_2 \to G_2^\infty = (4/27)(\tau\gamma_0)^{-1}(1 - \tau\gamma_0)^{-2}$, where $\gamma_0 \simeq 0.2063$. The constant γ_0^{-1} is the limiting value of λ_L/p. These limiting results are derived in the Appendix. The minimum value of G_1^∞ equals 1, and this occurs at $\tau = (5\gamma_0)^{-1} = 0.9695$. When $r = 1$, $G_1^\infty = 1.0003$. The minimum value of G_2^∞ is 1.140 which occurs when $\tau = 1$.

It was noted in Section 7 that G_1 represents the gross error sensitivity of the location vector whenever the p-dimensional Euclidean norm is used in

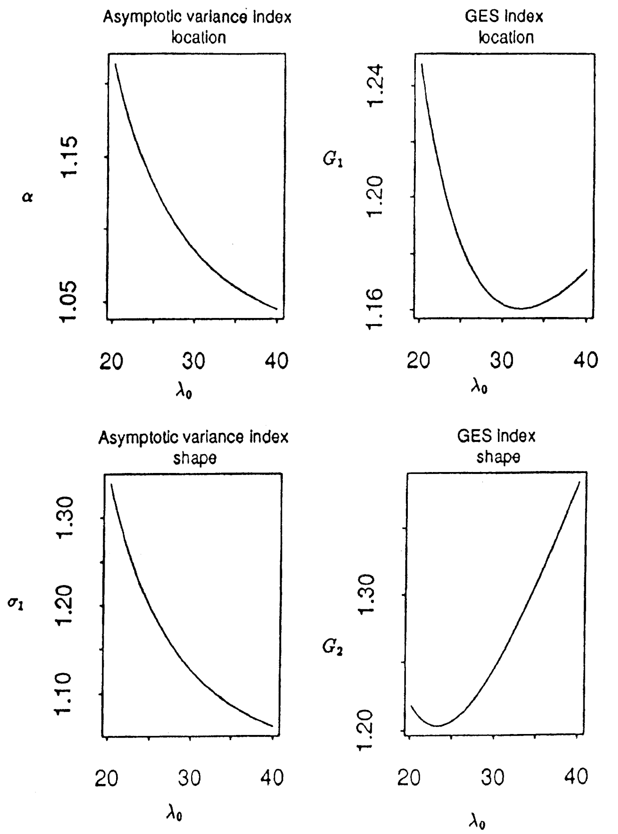

FIG. 3. *Indices of local robustness biweighted CM-estimates*: $p = 5$.

defining the GES. It is interesting to compare the GES of the *CM*-estimate of location to the GES of an estimate formed by using a univariate location estimate on each of the p variables. The latter is not generally affine equivariant. At the multivariate standard normal model, the componentwise location estimate has GES $= p^{1/2}g_1$, where g_1 represents the gross error sensitivity of the univariate location estimate at the standard normal distribution. For the univariate median $g_1 = (\pi/2)^{1/2} \simeq 1.2533$, and this is known to be the smallest possible value for g_1 among all translation equivariant estimates of univariate location; see, for example, Hampel, Ronchetti, Rousseeuw and Stahel (1986), Section 2.5c. Curiously, for large p, we note that biweighted *CM*-estimates of location can have gross error sensitivities which are smaller than the componentwise median. This is true even for $p = 4$. If we choose $\lambda_0 = 27$ for $p = 4$, then $G_1/p^{1/2} = 1.197$.

For large p, the biweighted *CM*-estimates do not improve upon the biweighted *S*-estimates in terms of the GES for shape (i.e., G_2). Some computations show that this is true for $p \geq 6$. If we do choose $\lambda_0 > \lambda_L$, though, a

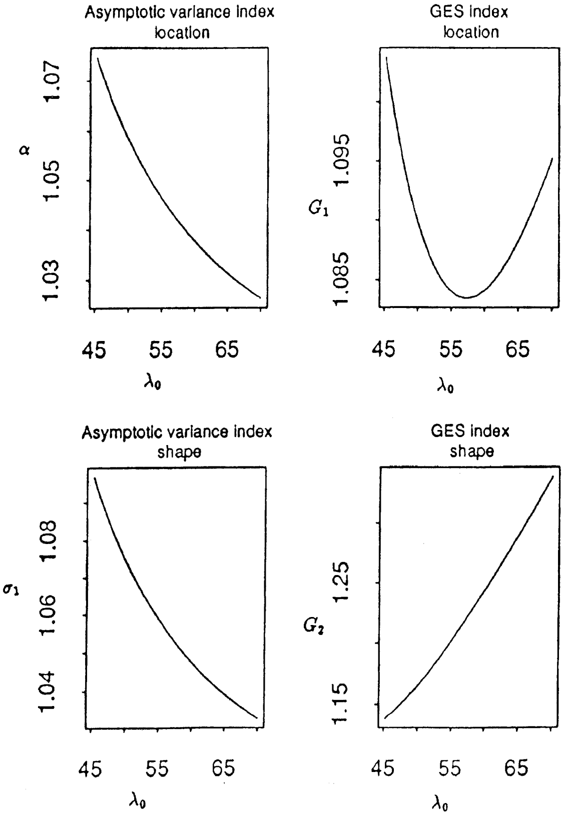

FIG. 4. *Indices of local robustness biweighted CM-estimates*: $p = 10$.

modest increase in G_2 may be a worthwhile sacrifice in order to obtain a smaller value of σ_1. A small value of σ_1 may be especially important when considering problems involving simultaneous inference since there are $\frac{1}{2}p(p+1) - 1$ functionally independent shape parameters.

APPENDIX

Proof of Theorem 3.1. First note that if $(\mu, V) \in \mathbb{R}^p \times \mathscr{P}_p$, then $L(\mu, V) < \infty$. Also, $(\mu, \lambda V) \in \mathbb{R}^p \times \mathscr{P}_p$ satisfies (3.2) for large enough λ. Thus, we only need to show that if $(\mu_k, V_k) \to \partial(\mathbb{R}^p \times \mathscr{P}_p)$, then either $L(\mu_k, V_k) \to \infty$ or (3.2) is not met for large k.

Let $\lambda_{1,k}$ and $\lambda_{p,k}$ represent the largest and smallest eigenvalue of V_k, respectively. If condition (3.2) is met for (μ_k, V_k), then, for any s, $\varepsilon\rho(\infty) \geq \rho(s)P[(\mathbf{X} - \mu_k)'V_k^{-1}(\mathbf{X} - \mu_k) \geq s]$, which implies for any $\delta > 0$ there exists an s large enough such that

$$\text{Prob}\big[(\mathbf{X} - \mu_k)'V_k^{-1}(\mathbf{X} - \mu_k) < s\big] \geq 1 - \varepsilon - \delta. \tag{A.1}$$

This, in turn, implies for any $\delta > 0$ there exists an s such that

$$\text{(A.2)} \qquad \text{Prob}\Big[\{\mathbf{a}_k'(\mathbf{X} - \mu_k)\}^2 < \lambda_{p,k} s\Big] \ge 1 - \varepsilon - \delta,$$

where $\mathbf{a}_k$ is an eigenvector of V_k associated with $\lambda_{p,k}$ and is normalized so that $\mathbf{a}_k'\mathbf{a}_k = 1$. If $\lambda_{p,k} \to 0$, then, by the condition on $\mathbf{X}$ in the theorem, (A.2) cannot hold for large k. Thus, we may assume $\lambda_{p,k} > \lambda_p > 0$. If $\lambda_{1,k} \to \infty$, then $L(\mu_k, V_k) \to \infty$ since ρ is bounded and $\lambda_{p,k}$ must be bounded away from 0. Thus, we may assume $\lambda_{1,k} < \lambda_1 < 0$. Finally, we note that if $\|\mu_k\| \to \infty$, then (A.1) cannot be met. □

Proof of Theorem 4.1. Theorem 3.1 ensures the existence of a solution $(\mu(\mathbf{X}), V(\mathbf{X}))$. Define $\varepsilon_0 = E[\rho\{(\mathbf{X} - \mu(\mathbf{X}))'V(\mathbf{X})^{-1}(\mathbf{X} - \mu(\mathbf{X}))\}]/\rho(\infty)$, and note that $(\mu(\mathbf{X}), V(\mathbf{X}))$ must minimize $|V|$ subject to the constraint $E[\rho\{(\mathbf{X} - \boldsymbol{\mu})'V^{-1}(\mathbf{X} - \boldsymbol{\mu})\}] = \varepsilon_0\, \rho(\infty)$. However, application of Theorem 1 of Davies (1987) for multivariate S-estimators implies that the solution to this last problem is unique and must be $(\mu(\mathbf{X}), V(\mathbf{X})) = (\boldsymbol{\mu}, \lambda\Sigma)$ for some fixed $\lambda > 0$. This completes the proof. In applying Davies' (1987) results, it is helpful to note that the function $\kappa(s)$ used by Davies (1987) is equivalent to $1 - \rho(s)/\rho(\infty)$. Also, Davies (1987) normalizes the function g in (4.1) so that $(\mu(\mathbf{X}), V(\mathbf{X})) = (\mu, \Sigma)$. This normalization implicitly depends on ε_0. □

Proof of Theorem 5.1. (i) *Upper bound.* Suppose $m \ge \lceil n\varepsilon \rceil$ and consider a sequence $Y_k = \{\mathbf{y}_{1,k}, \dots, \mathbf{y}_{m,k}\}$ such that $\|\mathbf{y}_{j,k}\| \to \infty$ as $k \to \infty$ for $j = 1, \dots, m$. Define $Z_k = Y_k \cup \{\mathbf{x}_{m+1}, \dots, \mathbf{x}_n\}$ and let $(\mu_k, V_k) = (\hat{\mu}(Z_k), \hat{V}(Z_k))$. Again, let $\lambda_{1,k}$ and $\lambda_{p,k}$ represent the largest and smallest eigenvalues of V_k, respectively. If $(\hat{\mu}(\cdot), \hat{V}(\cdot))$ does not break down, then there exist constants a, v_1, v_2 such that $\|\mu_k\| < a < \infty$ and $0 < v_1 < \lambda_{1,k} \le \lambda_{p,k} < v_2 < \infty$. Thus, $(\mathbf{y}_{j,k} - \mu_k)'V_k^{-1}(\mathbf{y}_{j,k} - \mu_k) \to \infty$ for $j = 1, 2, \dots, m$ and so, for large k, $\text{ave}\{\rho[(\mathbf{z}_{i,k} - \mu_k)'V_k^{-1}(\mathbf{z}_{i,k} - \mu_k)]\} > m\rho(\infty)/n \ge \varepsilon\rho(\infty)$. This contradicts (1.2) and so $\varepsilon^*(X_n) \le \lceil n\varepsilon \rceil/n$.

Suppose now that $m \ge \lceil n(1 - \varepsilon) - p \rceil$ and suppose all m elements of Y equal $\mathbf{x}_1$. Since $(\hat{\mu}(\cdot), \hat{V}(\cdot))$ is affine equivariant, we can assume $\mathbf{x}_1 = \mathbf{0}$, $\mathbf{x}_{p+1} = \mathbf{e}_1$ and, when $p \ge 2$, $\mathbf{x}_j = \mathbf{e}_j$ for $j = 2, \dots, p$, where $\mathbf{e}_j$ is the vector with 1 in the jth position and 0's elsewhere. Define $Z = Y \cup \{x_1, \dots, x_{n-m}\}$ and let $s_0 > 0$ such that $\rho(s_0) < \rho(\infty)$. Such an s_0 exists since $\rho(s)$ is continuous at 0 and $\rho(0) < \rho(\infty)$. Consider now the sequence $V_k = s_0^{-1}e_1e_1' + \lambda_k^{-1}(I - e_1e_1')$ with $\lambda_k \to 0$. Thus, $\mathbf{y}_i'V_k^{-1}\mathbf{y}_i = \mathbf{x}_1'V_k^{-1}\mathbf{x}_1 = 0$, $\mathbf{x}_{p+1}'V_k^{-1}\mathbf{x}_{p+1} = s_0$ and, when $p \ge 2$, $\mathbf{x}_j^{-1}V_k^{-1}\mathbf{x}_j = \lambda_k$ for $j = 2, \dots, p$. If $n - m \ge p + 1$, we then have $\text{ave}\{\rho\mathbf{z}_i'V_k^{-1}\mathbf{z}_i)\} \le (n - m - p - 1)\rho(\infty)/n + \rho(s_0)/n + (p - 1)\lambda_k/n \le \varepsilon\rho(\infty) - \{\rho(\infty) - \rho(s_0)\}/n + (p - 1)\lambda_k/n$. If $n - m \le p$, then $\text{ave}\{\rho(\mathbf{z}_i'V_k^{-1}\mathbf{z}_i)\} \le (p - 1)\lambda_k/n$. In either case, this implies that eventually, for large enough k, the constraint (1.2) holds for $(\mathbf{0}, V_k)$. However, for any fixed (μ, V), $L(\mu, V; Z) > -\infty$ whereas $L(\mathbf{0}, V_k; Z) \to -\infty$ and so $\varepsilon(\mathbf{X}_n) \le \lceil n(1 - \varepsilon) - p \rceil/n$.

(ii) *Lower bound.* Suppose $m < \min\{\lceil n\varepsilon \rceil, \lceil n(1-\varepsilon) - p \rceil\}$ and consider any ε_m-contaminated samples Z. We first note that Theorem 3.1 ensures the existence of $(\hat{\mu}(Z), \hat{V}(Z))$. Let $\lambda_1(Z)$ and $\lambda_p(Z)$ represent the largest and smallest eigenvalues of $\hat{V}(Z)$, respectively. Our goal is to show that, for all possible ε_m-contaminated samples Z, $\lambda_1(Z)$ is bounded above, $\lambda_p(Z)$ is bounded below and $\|\hat{\mu}(Z)\|$ is bounded above. This then implies $\varepsilon^*(X) \geq \min\{\lceil n\varepsilon \rceil / n, \lceil n(1-\varepsilon) - p \rceil / n\}$.

The constraint (1.2) implies that there exists $s_u < \infty$ not dependent on Z such that $\{\mathbf{z} - \hat{\mu}(Z)\}'\hat{V}(Z)^{-1}\{\mathbf{z} - \hat{\mu}(Z)\} < s_u$ for at least $\lceil n(1-\varepsilon) \rceil$ values in Z. Since $m < \lceil n(1-\varepsilon) - p \rceil$, at least $(p+1)$ of such values must be from X_n. Since X_n is in general position, this implies there exists a $\lambda_L > 0$ such that $\lambda_p(Z) > \lambda_L$. Also, if there exists a $\lambda_u < \infty$ such that $\lambda_1(Z) < \lambda_u$, then there must exist a $\mu_u < \infty$ such that $\|\hat{\mu}(Z)\| < \mu_u$. So it remains only to show that $\lambda_1(Z)$ is bounded above.

To show this, note that $L\{\hat{\mu}(Z), \hat{V}(Z); Z\}$ is not bounded above if $\lambda_1(Z)$ is not since ρ is bounded and we have already shown that $\lambda_p(Z) > \lambda_L > 0$. However, since $m < \lceil n\varepsilon \rceil$, constraint (1.2) is met for $(\mu, V) = (0, \lambda I)$ and any Z for large enough λ, or specifically for λ such that $\sup_{\mathbf{x} \in X} \rho(\mathbf{x}'\mathbf{x}/\lambda) \leq n(\varepsilon - m/n)\rho(\infty)/(n-m)$. Moreover, $L(\mathbf{0}, \lambda I; Z)$ is bounded above over Z for any fixed λ. Thus, there must exist a $\lambda_u < \infty$ such that $\lambda_1(Z) < \lambda_u$. □

LEMMA A.1. *Let $\mathbf{Y}_k$ be a sequence of random vectors in $\mathbb{R}^p$ such that $\mathbf{Y}_k \to \mathbf{Y}$ in distribution. Also, suppose $(\mu_k, V_k) \in \mathbb{R}^p \times \mathscr{P}_p$ with $(\mu_k, V_k) \to (\mu, V) \in \mathbb{R}^p \times \mathscr{P}_p$. If G: $\mathbb{R} \to \mathbb{R}$ is bounded and continuous, then $E[G\{(\mathbf{Y}_k - \mu_k)'V_k^{-1}(\mathbf{Y}_k - \mu_k)\}] \to E[G\{(\mathbf{Y} - \mu)'V^{-1}(\mathbf{Y} - \mu)\}]$.*

PROOF. Let $d_k^2 = (\mathbf{Y}_k - \mu_k)'V_k^{-1}(\mathbf{Y}_k - \mu_k)$ and $d^2 = (\mathbf{Y} - \mu)'V^{-1}(\mathbf{Y} - \mu)$. We then have $d_k^2 \to d^2$ in distribution. This implies $E[G(d_k^2)] \to E[G(d^2)]$ for any bounded continuous G. □

PROOF OF THEOREM 6.1. Since $\mathbf{X}$ is continuous, the conditions of Theorem 3.1 hold for $\mathbf{X}_k$ for large enough k, say $k \geq K$. Furthermore, there must exist some compact set $\mathscr{C}$ such that $(\mu_k, V_k) \in \mathscr{C}$ for $k \geq K$, since, if not, then there exists a subsequence of (μ_k, V_k) diverging to $\partial(\mathbb{R}^p \times \mathscr{P}_p)$. An argument analogous to the proof of Theorem 3.1 can be made to show that this last statement leads to a contradiction. Thus, to complete the proof of Theorem 6.1, it is sufficient to show that if $(\mu_k, V_k) \to (\mu_0, V_0) \in \mathbb{R}^p \times \mathscr{P}_p$, then $(\mu_0, V_0) = (\mu(\mathbf{X}), V(\mathbf{X}))$.

Let $L_k(\mu, V)$ be the objective function defined by applying (3.1) to $\mathbf{X}_k$ rather than to $\mathbf{X}$. Also, let $c_k(\mu, V) = E[\rho\{(\mathbf{X}_k - \mu)'V^{-1}(\mathbf{X}_k - \mu)\}]/\rho(\infty)$, and define $c(\mu, V)$ accordingly for $\mathbf{X}$. Thus, (μ_k, V_k) minimizes $L_k(\mu, V)$ subject to $c_k(\mu, V) \leq \varepsilon$, and $(\mu(\mathbf{X}), V(\mathbf{X}))$ minimizes $L(\mu, V)$ subject to $c(\mu, V) \leq \varepsilon$.

Suppose now that $(\mu_k, V_k) \to (\mu_0, V_0) \neq (\mu(\mathbf{X}), V(\mathbf{X}))$. By the presumed uniqueness of $(\mu(\mathbf{X}), V(\mathbf{X}))$, either $L(\mu_0, V_0) > L(\mu(\mathbf{X}), V(\mathbf{X}))$ or $c(\mu_0, V_0) > \varepsilon$.

The latter is not possible since $\varepsilon \geq c_k(\mu_k, V_k) \to c(\mu_0, V_0)$. The limit follows from Lemma A.1. Thus, we must have

$$L(\mu_0, V_0) > L(\mu(\mathbf{X}), V(\mathbf{X})). \tag{A.3}$$

We divide the remainder of the proof into two cases.

CASE I [$c(\mu(\mathbf{X}), V(\mathbf{X})) < \varepsilon$]. By Lemma A.1, we have $c_k(\mu(\mathbf{X}), V(\mathbf{X})) \to c(\mu(\mathbf{X}), V(\mathbf{X})) < \varepsilon$, $L_k(\mu(\mathbf{X}), V(\mathbf{X})) \to L(\mu(\mathbf{X}), V(\mathbf{X}))$ and $L_k(\mu_k, V_k) \to L(\mu_0, V_0)$. These limits, together with (A.3), imply $c_k(\mu(\mathbf{X}), V(\mathbf{X})) < \varepsilon$ and $L_k(\mu(\mathbf{X}), V(\mathbf{X})) < L_k(\mu_0, V_0)$ for large enough k. This contradicts the definition of (μ_k, V_k) and so we must have $(\mu_0, V_0) = (\mu(\mathbf{X}), V(\mathbf{X}))$.

CASE II [$c(\mu, \mathbf{X}), V(\mathbf{X})) = \varepsilon$]. Since the distribution of $\mathbf{X}$ is absolutely continuous in $\mathbb{R}^p$, the distribution of $D^2 = \{\mathbf{X} - \mu(\mathbf{X})\}' V(\mathbf{X})^{-1} \{\mathbf{X} - \mu(\mathbf{X})\}$ is absolutely continuous in $\mathbb{R}$. Now $\rho(D^2) = \rho(\infty)$ w.p.1 is not possible since $c(\mu(\mathbf{X}), V(\mathbf{X})) = \varepsilon < 1$. Thus, by Condition 6.2, for any $\eta > 0$, $c(\mu(\mathbf{X}), (1 + \eta)V(\mathbf{X})) < \varepsilon$. By (A.3), $\eta > 0$ can be chosen so that $L(\mu(\mathbf{X}), (1 + \eta)V(\mathbf{X})) < L(\mu_0, V_0)$. The proof then proceeds as in Case I. □

REMARK A.1. From the proof of Theorem 6.1 we note for the case $c(\mu(\mathbf{X}), V(\mathbf{X})) < \varepsilon$ that the conditions of Theorem 6.1 can be relaxed. In particular, we do not need to assume that $\rho(s)$ is strictly increasing. Also, the assumption that the distribution of $\mathbf{X}$ is absolutely continuous can be replaced by Condition 3.2. Note that if Condition 3.2 holds for $\mathbf{X}$, then it holds for $\mathbf{X}_k$ for large enough k. This follows since, for any hyperplane $B \subset \mathbb{R}^p$, $\limsup P(X_k \in B) \leq P(X \in B)$.

The conditions of Theorem 6.1 can also be relaxed for the case $c(\mu(\mathbf{X}), V(\mathbf{X})) = \varepsilon$. For this case, the assumption that $\rho(s)$ is strictly increasing and the assumption that $\mathbf{X}$ is absolutely continuous can be replaced by Condition 3.2 and $P[D^2 \in A_p] \neq 0$, where $A_p = \{s > 0 \mid \rho(s_0) < \rho(s) < \rho(s_1)$ for $s_0 < s < s_1\}$. This last assumption simply states that the support of D^2 is not mutually exclusive of the set on which ρ is strictly increasing.

PROOF OF THEOREM 6.2. By Theorem 6.1, we know that $(\mu(\mathbf{X}_k), V(\mathbf{X}_k)) \to (\mu(\mathbf{X}), V(\mathbf{X}))$. Now, if (3.5) holds for $\mathbf{X}$, then (3.6) does not hold for $\mathbf{X}$. Application of Lemma A.1 then implies that (3.6) cannot hold for $\mathbf{X}_k$ for large enough k, and hence (3.5) must hold. An identical argument shows that if (3.6) holds for $\mathbf{X}$, then (3.6) but not (3.5) holds for $\mathbf{X}_k$ for large enough k. □

PROOFS OF THEOREMS 6.3 AND 6.4. These proofs mimic the proofs of Theorems 3.3 and 4.1 given by Lopuhaä (1989). □

PROOF OF (8.2). For vectors $\mathbf{a}$ and $\mathbf{b}$, let $d\mathbf{a}/d\mathbf{b}$ represent the matrix of partial derivatives $\{\partial a_i / \partial b_j\}$ with i varying over rows and j varying over columns. Taking into account the redundancy caused by the symmetry of V, the first two terms of the Taylor series expansion of V about V_0 are given by

$$H(V) = H(V_0) + \{H'(V_0)\}\mathrm{vec}\{V - V_0\} + o(V - V_0), \tag{A.4}$$

where $H'(V) = \frac{1}{2}\{dH(V)/d\,\mathrm{vec}(V)\}(I + J_p)$ with $J_p = \sum_{i=1}^{p}\mathbf{e}_i\mathbf{e}_i' \otimes \mathbf{e}_i\mathbf{e}_i'$. Using (A.4), it follows that

$$\text{(A.5)} \qquad \mathrm{IF}(\mathbf{x}; H(V(\mathbf{X}))) = \{H'(V(\mathbf{X}))\}\mathrm{vec}\{\mathrm{IF}(\mathbf{x}; V(\mathbf{X}))\}.$$

Now, since $H(\lambda V) = H(V)$, $\lambda H'(\lambda V) = H'(V)$ and hence $\lambda_0 H'(V(\mathbf{X})) = H(\Sigma)$. Also, from the proof of Theorem 1 in Tyler (1983), we know that $\{H'(V)\}\mathrm{vec}(V) = 0$. Applying these results to (7.2) and using some standard matrix algebra, we obtain

$$\text{(A.6)} \qquad \mathrm{IF}(\mathbf{x}; H(V(\mathbf{X}))) = h_2(s/\lambda_0)\{H'(\Sigma)\}(\Sigma^{1/2} \otimes \Sigma^{1/2})(\mathbf{z} \otimes \mathbf{z}).$$

The proof is complete by observing $\gamma_H(\Sigma) = \{H'(\Sigma)\}(\Sigma^{1/2} \otimes \Sigma^{1/2})$ and so $\gamma_H(\lambda\Sigma) = \gamma_H(\Sigma)$. □

Derivations for the biweighted CM-estimates. Expressions for α, G_1, σ_1 and G_2 as functions of λ_0 follow from the identity

$$e_k(\lambda) = E\left[(S/\lambda)^k I_{S \le \lambda}\right] = 2^k\Gamma\left(\tfrac{1}{2}p + k\right)\mathrm{Prob}\left[\chi^2_{p+2k} \le \lambda\right]/\left\{\lambda^k\Gamma\left(\tfrac{1}{2}p\right)\right\},$$

where S has a χ^2_p distribution. We then have

$$\text{(A.7)} \qquad \alpha = p^{-1}\lambda_0\{e_5(\lambda_0) - 4e_4(\lambda_0) + 6e_3(\lambda_0) - 4e_2(\lambda_0) + e_1(\lambda_0)\}/b_1^2,$$

$$\text{(A.8)} \qquad G_1 = (96/125)\sqrt{5}\,\lambda_0^{1/2}/b,$$

$$\text{(A.9)} \qquad \sigma_1 = (1 + 2/p)\{e_6(\lambda_0) - 4e_5(\lambda_0) + 6e_4(\lambda_0) - 4e_2(\lambda_0)\}/b_2^2,$$

$$\text{(A.10)} \qquad G_2 = (8/9)(p + 2)/b_2,$$

where $b_1 = (1 + 4/p)e_2(\lambda_0) - 2(1 + 2/p)e_1(\lambda_0) + e_0(\lambda_0)$ and $b_2 = e_6(\lambda_0) - 4e_5(\lambda_0) + 6e_4(\lambda_0) - 4e_3(\lambda_0) + e_2(\lambda_0)$. Furthermore,

$$\text{(A.11)} \quad E[\rho(S/\lambda)]/\rho(\infty) = 3e_1(\lambda) - 3e_2(\lambda) - 3e_3(\lambda) + 1 - e_0(\lambda).$$

Consider now $p \to \infty$. Since $\chi^2_{p+2k}/p \to 1$ in probability, it follows that $e_k(p/\gamma) \to \gamma^k$ for any $0 < \gamma < 1$. Using (A.11), this implies $\lambda_L/p \to \gamma_0^{-1}$, where γ_0 is the unique root of the equation $3\gamma - 3\gamma^2 + \gamma^3 = 1/2$, specifically $\gamma_0 \simeq 0.2063$. If we set $\lambda_0 = \lambda_L/\tau$ with $0 < \tau \le 1$ and let $p \to \infty$, then $\lambda_0/p \to (\tau\gamma_0)^{-1}$ and $e_k(\lambda_0) \to (\gamma_0\tau)^{-1}$. From this we get the limiting results for α, $G_1/p^{1/2}$, σ_1 and G_2 given in Section 9.

REFERENCES

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

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

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

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

HUBER, P. J. (1977). Robust covariances. In *Statistical Decision Theory and Related Topics* (S. S. Guta and D. S. Moore, eds.) 165–191. Academic Press, New York.

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.

KENT, J. T. and TYLER, D. E. (1991). Redescending *M*-estimates of multivariate location and scatter. *Ann. Statist.* **19** 2102–2119.

KIEFER, J. and WOLFOWITZ, J. (1956). Consistency of the maximum likelihood estimator in the presence of infinitely many incidental parameters. *Ann. Math. Statist.* **27** 887–906.

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

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

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

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

STAHEL, W. A. (1981). Breakdown of covariance estimators. Technical Report 31, Pachgruppe für Statistik, ETH, Zurich.

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

TYLER, D. E. (1991). Some issues in the robust estimation of multivariate location and scatter. In *Directions in Robust Statistics and Diagnostics* **2** (W. Stahel and S. Weisberg, eds.) 327–336. Springer, New York.

DEPARTMENT OF STATISTICS
UNIVERSITY OF LEEDS
LEEDS LS2 9JT
ENGLAND

DEPARTMENT OF STATISTICS
RUTGERS UNIVERSITY
PISCATAWAY, NEW JERSEY 08855
E-MAIL: dtyler@stat.rutgers.edu