The Annals of Statistics
2000, Vol. 28, No. 4, 1219–1243

ON THE UNIQUENESS OF S-FUNCTIONALS AND M-FUNCTIONALS UNDER NONELLIPTICAL DISTRIBUTIONS

BY KAY S. TATSUOKA AND DAVID E. TYLER[1]

National Institute of Statistical Sciences and Rutgers University

The S-functionals of multivariate location and scatter, including the MVE-functionals, are known to be uniquely defined only at unimodal elliptically symmetric distributions. The goal of this paper is to establish the uniqueness of these functionals under broader classes of symmetric distributions. We also discuss some implications of the uniqueness of the functionals and give examples of striclty unimodal and symmetric distributions for which the MVE-functional is not uniquely defined.

The uniqueness results for the S-functionals are obtained by embedding them within a more general class of functionals which we call the M-functionals with auxiliary scale. The uniqueness results of this paper are then obtained for this class of multivariate functionals. Besides the S-functionals, the class of multivariate M-functionals with auxiliary scale include the constrained M-functionals recently introduced by Kent and Tyler, as well as a new multivariate generalization of Yohai's MM-functionals.

1. Introduction and summary. Most robust high breakdown point estimates of multivariate location and scatter are not explicitly defined. For example, the minimum volume ellipsoid (MVE) estimators and their generalization, the S-estimators, are defined via a minimization subject to some constraint. Consequently, little is known about when the S-estimators are uniquely defined. This question of uniqueness is not a simple one. The bulk of the original paper by Davies (1987), where the multivariate S-estimators are introduced, is devoted to showing that under fairly broad conditions, the functional version of the S-estimators, that is, the S-functionals, are uniquely defined at unimodal elliptically symmetric distributions. This is apparently still the only theoretical result on the uniqueness of the S-estimators or the S-functionals, including the MVE estimators and functionals [see, e.g., Davies (1993)]. As is shown in Davies (1987), and also in Lopuhaä (1989), the importance of establishing the uniqueness of the S-functionals at some distribution is that among other things, once given uniqueness other important properties such as the weak continuity and the influence function of the S-functionals, as well as the strong consistency and the asymptotic normality of the S-estimates, follow fairly readily. Analogous comments hold true for a more flexible variant of the S-functionals of multivariate location and scatter, recently introduced by Kent

Received September 1997; revised November 1999.

[1]Supported in part by NSF Grant IRI-95-30546.

AMS 1991 *subject classifications*. 62G35, 62H05.

Key words and phrases. CM-estimates, elliptical distributions, majorization, M-estimates, minimum volume ellipsoid, MM-estimates, permutation invariance, robustness, Schur-concavity, S-estimates, symmetric exchangeable distributions, unimodal distributions.

and Tyler (1996), called the constrained M-functionals or CM-functionals for short.

The assumption of an elliptically symmetric distribution is often made simply because of its mathematical tractability, but this is a somewhat restrictive assumption and is often unrealistic in practice. This motivates us to study in this paper the uniqueness of the S-functionals, including the MVE-functionals, as well as the uniqueness of the CM-functionals under symmetric distributions other than elliptical distributions.

The uniqueness results for the S-functionals and the CM-functionals are obtained by embedding them within a more general class of functionals which we refer to as the multivariate M-functionals with auxiliary scale. The uniqueness results are then obtained for this general class of functionals. The class of multivariate M-functionals with auxiliary scale represents an adaptation of the M-functionals of general scale for regression introduced by Martin, Yohai and Zamar (1989) to the multivariate setting. It also enables us to introduce a multivariate version of Yohai's (1987) MM-functionals.

Formal definitions and a brief review of the S-functionals and the CM-functionals are given in Section 2.1. In Section 2.2, we define the M-functionals of auxiliary scale and show that the S-functionals, the CM-functionals and our newly defined MM-functionals lie within this class.

The main uniqueness results of the paper are given in Sections 3 and 4. In Section 3.1, we show that for symmetric unimodal distributions as defined by Anderson (1955), the multivariate location M-functionals with auxiliary scale are uniquely defined by the center of symmetry. This is a result that one might anticipate. The main tool used in obtaining this result is Anderson's (1955) classical inequality on the integral of a symmetric unimodal function over a convex set.

By Anderson's definition, the densities of symmetric unimodal distributions in $\Re^p$ have convex contours and include unimodal elliptically symmetric distributions. There are other symmetric distributions in $\Re^p$ of interest, however, which do not have convex contours. For example, if the marginal components of the distribution are independent Cauchy distributions, then their joint distribution in $\Re^p$ does not have convex contours. We therefore discuss a more general definition of symmetric unimodal distributions within Section 3.2, and note that the MVE location functional does not necessarily correspond to the center of symmetry. This leads us to restrict our attention to distributions for which each coordinate slice of the density is symmetric and strictly unimodal. What we mean by this terminology is defined within Section 3.2. For such distributions, or for any affine transformations of such distributions, we show that the location M-functionals with auxiliary scale are again uniquely defined by the center of symmetry.

The joint uniqueness of the location and scatter functionals is considered in Section 4. To obtain joint uniqueness results, we place some additional conditions on the class of distributions considered in Section 3.2. Specifically, we consider the class $\mathscr{P}_p$ of distributions in $\Re^p$ which are invariant under permutations and sign changes of its components and which have densities f such that $f \circ \exp$ is Schur-concave. The assumption that $f \circ \exp$ be Schur-concave is more general than the assumption that f be Schur-concave. The class $\mathscr{P}_p$ includes the unimodal spherically symmetric distributions, as well as a large class of distributions whose components are independent and identically distributed according to some univariate unimodal symmetric distribution, such as the univariate t-distributions. It is well known that the only distributions within the class of spherically symmetric distributions which have independent and identically distributed components are the spherical normal distributions. The class $\mathscr{P}_p$ can thus be viewed as a much broader generalization of the spherical normals than unimodal spherically symmetric distributions. In Section 4, we obtain uniqueness results for both the location and scatter components of any M-functional with auxiliary scale for distributions in $\mathscr{P}_p$, and consequently for affine transformations of distributions in $\mathscr{P}_p$. The location functional is again uniquely defined by the center of symmetry. The scatter functional is a multiple of the identity matrix whenever the distribution is in $\mathscr{P}_p$. For affine transformations of distributions in $\mathscr{P}_p$, the scatter matrix is transformed accordingly. The proofs of these results rely heavily on results on majorization.

In Section 5, we discuss our results and make some concluding remarks about the uniqueness problem. We believe that the conditions given in this paper are close to the best general conditions possible to ensure the uniqueness of the M-functionals with auxiliary scale. We conclude by giving and discussing examples of strictly unimodal symmetric distributions for which the MVE-functionals are not uniquely defined. Technical proofs are given in the Appendix.

2. Classes of functionals.

2.1. *S-functionals and CM-functionals.* Perhaps the most widely known high breakdown point estimator of multivariate location and scatter is the minimum volume ellipsoid (MVE) estimator introduced by Rousseeuw (1986). For n points in $\Re^p$, the MVE is defined to be the minimum volume ellipsoid covering at least half of the data points, or more generally covering at least a proportion of $(1 - \varepsilon)$ of the data for some fixed constant $0 < \varepsilon < 1$. The corresponding MVE estimator of scatter V is taken to be proportional to the positive definite symmetric matrix $\boldsymbol{A}$ associated with the quadratic equation $(\mathbf{X}-\boldsymbol{\mu})'\boldsymbol{A}^{-1}(\mathbf{X}-\boldsymbol{\mu}) = 1$ which characterizes the minimum volume ellipsoid. As with the least median of squares (LMS) estimator in regression, a drawback to the MVE estimator is that it is only $\sqrt[3]{n}$ consistent rather than $\sqrt{n}$ consistent, and so its asymptotic efficiency at the normal model is zero [see, e.g., Nolan (1991) and Davies (1992)]. To correct this, multivariate S-estimators were introduced by Davies (1987) and by Rousseeuw and Leroy (1987). These are a generalization of the MVE estimator in much the same way that the S-estimators of regression introduced by Rousseeuw and Yohai (1984) are a generalization of the LMS estimator of regression.

Let PDF(p) denote the set of all positive definite symmetric matrices of order p and let $x_1, \ldots, x_n$ represent a sample of size n in $\Re^p$. The S-estimators of multivariate location and scatter are defined as any pair $\hat{\boldsymbol{\mu}} \in \Re^p$ and $\widehat{V} \in$ PDF (p), respectively, which minimizes $\det(V)$ subject to the constraint

$$\text{ave}_i\big[\rho\{(\mathbf{x}_i - \boldsymbol{\mu})' V^{-1}(\mathbf{x}_i - \boldsymbol{\mu})\}\big] \le \varepsilon\rho(\infty), \tag{1}$$

where ε is a fixed value between 0 and 1, ave_i denotes the arithmetic average over $i = 1, \ldots, n$, and $\rho(s)$ satisfies Condition 2.1 below.

CONDITION 2.1. For $s \ge 0$, $\rho(s)$ is nondecreasing, $0 = \rho(0) < \rho(\infty) < \infty$, and $\rho(s)$ is continuous from above at zero.

When $\rho(s)$ is zero–one step function, an S-estimator corresponds to an MVE-estimator.

The simultaneous S-functionals of multivariate location and scatter at a distribution F in $\Re^p$ are defined analogously as any pair $\boldsymbol{\mu}(F) \in \Re^p$ and $V(F) \in$ PDF(p), respectively, which minimizes $\det(V)$ subject to the constraint

$$E\big[\rho\{(\mathbf{X} - \boldsymbol{\mu})' V^{-1}(\mathbf{X} - \boldsymbol{\mu})\}\big] \le \varepsilon\rho(\infty), \tag{2}$$

where $\mathbf{X}$ is a p-dimensional random vector having distribution F. If F is the empirical distribution F_n, then (2) reduces to (1).

The S-estimators are affine equivariant in the sense that if $(\hat{\boldsymbol{\mu}}, \widehat{V})$ are S-estimates for the data $\mathbf{x}_1, \ldots, \mathbf{x}_n$ then $(\mathbf{B}\hat{\boldsymbol{\mu}} + \mathbf{b}, \mathbf{B}\widehat{\boldsymbol{V}}\mathbf{B}')$ are S-estimates for the transformed data $\mathbf{Bx}_1 + \mathbf{b}, \ldots, \mathbf{Bx}_n + \mathbf{b}$, where $\mathbf{B}$ is a nonsingular matrix of order p and $\mathbf{b} \in \Re^p$. The S-functionals are affine equivariant in an analogus fashion; that is, if $F_{\mathbf{Y}}$ represents the distribution of the p-dimensional vector $\mathbf{Y} = \mathbf{BX} + \mathbf{b}$ and $(\boldsymbol{\mu}(F_{\mathbf{X}}), V(F_{\mathbf{X}}))$ are S-functionals at $F_{\mathbf{X}}$ then

$$(\boldsymbol{\mu}(F_{\mathbf{Y}}), \boldsymbol{V}(F_{\mathbf{Y}})) = (\mathbf{B}\boldsymbol{\mu}(F_{\mathbf{X}}) + \mathbf{b}, \mathbf{B}\boldsymbol{V}(F_{\mathbf{X}})\mathbf{B}') \tag{3}$$

corresponds to an S-functional at $F_{\mathbf{Y}}$.

Davies (1987) shows that the multivariate S-estimators have an asymptotic breakdown point which depends on the choice of ε in (2), namely, $\min(\varepsilon, 1 - \varepsilon)$. In addition, he also shows that if $\rho(s)$ is sufficiently smooth, then the S-estimator is $\sqrt{n}$ consistent and asymptotically normal. Lopuhaä (1989), however, notes that the S-estimators tend to become more inefficient at the multivariate normal model as the value of ε increases from 0 to 1/2 and can be quite inefficient for the higher breakdown point S-estimators. In addition, the gross error sensitivities of the higher breakdown point S-estimators at the multivariate normal model tend to be fairly large.

These shortcomings of the S-estimators motivated Kent and Tyler (1996) to introduce a more flexible variant of the S-estimators called the constrained M-estimators or CM-estimators for short. The CM-estimators for multivariate location and scatter are defined as any pair $\hat{\boldsymbol{\mu}} \in \Re^p$ and $\widehat{V} \in$ PDF(p), respectively, which minimizes

$$l_n(\boldsymbol{\mu}, V) = \text{ave}_i\big[\rho\big\{(\mathbf{x}_i - \boldsymbol{\mu})' V^{-1}(\mathbf{x}_i - \boldsymbol{\mu})\big\}\big] + (1/2)\log(\det(V)) \tag{4}$$

subject to the constraint (1), with ε and $\rho(s)$ being defined as in the S-estimators. The CM-functionals of multivariate location and scatter at a distribution F in $\Re^p$ are defined analogously as any pair $\boldsymbol{\mu}(F) \in \Re^p$ and $V(F) \in \mathrm{PDF}(p)$, respectively, which minimizes

$$L(\boldsymbol{\mu}, V) = E\big[\rho\{(\mathbf{X}-\boldsymbol{\mu})'V^{-1}(\mathbf{X}-\boldsymbol{\mu})\}\big] + (1/2)\log(\det(V)), \tag{5}$$

under the constraint (2), where again $\mathbf{X}$ is a p-dimensional random vector having distribution F. If F is taken to be the empirical distribution function then (5) reduces to (4).

Like the S-estimators and functionals, the CM-estimators and functionals are affine equivariant. Kent and Tyler (1996) show that the asymptotic breakdown points of the CM-estimators are analogous to those of the S-estimators, namely $\min(\varepsilon, 1-\varepsilon)$. They also show that, as with the S-estimators, if $\rho(s)$ is sufficiently smooth, then the CM-estimator is $\sqrt{n}$ consistent and asymptotically normal. In addition, unlike the S-estimators, the higher breakdown point CM-estimators can be tuned to have good efficiency and a relatively low gross error sensitivity at the multivariate normal and other models without affecting the breakdown points of the CM-estimators. For details, see Kent and Tyler (1996).

As noted in the introduction, except for the results on the breakdown points, these results concerning consistency, asymptotic normality and gross error sensitivity as well as results on weak continuity, the influence function and maximum bias curves depend on the S-functional or CM-functional being uniquely defined at the distribution of interest; see Davies (1987), Lopuhaä (1989) and Kent and Tyler (1996). Existence of the S-functionals and the CM-functionals have been established under the following mild condition on the underlying distribution [Theorem 3.1 of Kent and Tyler (1996)]. Here P_F refers to the probability measure associated with the distribution function F.

CONDITION 2.2. For all hyperplanes $B \subset \Re^p$, $P_F(B) < 1-\varepsilon$.

Note that if F is absolutely continuous then $P_F(B) = 0$ for any hyperplane B and so Condition 2.2 holds. However, the uniqueness of the S-functionals and the CM-functionals has only been established so far for unimodal elliptically symmetric distributions. We address this problem for much broader classes of symmetric distributions in Sections 3 and 4.

2.2. *M-functionals with auxiliary scale.* Before studying the uniqueness of the multivariate S-functional and the CM-functionals, we note that they can both be embedded within a larger class of functionals which we call the multivariate M-functionals with auxiliary scale. M-estimates with auxilary scale have been previously introduced in the regression setting by Martin, Yohai and Zamar (1989) as a way of studying S-estimates, M-estimates and MM-estimates of regression within a unified framework. They also contain the constrained M-estimates of regression introduced by Mendes and Tyler

(1996). Martin, Yohai and Zamar (1989) use the terminology "M-estimates with general scale," but we feel the term "auxiliary scale" is more descriptive.

For a given "scale" functional $\sigma(F) > 0$, we define the multivariate location and scatter M-functionals with auxiliary scale $\sigma(F)$ to be $\boldsymbol{\mu}(F)$ and $V(F) = \sigma^2(F)\Gamma(F)$, respectively, where $(\boldsymbol{\mu}(F), \Gamma(F))$ minimize

$$E\big[\rho\{(\mathbf{X}-\boldsymbol{\mu})'\Gamma^{-1}(\mathbf{X}-\boldsymbol{\mu})/\sigma^2(F)\}\big] \tag{6}$$

over all $(\boldsymbol{\mu}(F), \Gamma) \in \Theta(p)$, and where

$$\Theta(p) = \{(\boldsymbol{\mu}(F), \Gamma) | \boldsymbol{\mu}(F) \in \Re^p \text{ and } \Gamma \in \text{ PDF}(p) \text{ with } \det(\Gamma) = 1\} \tag{7}$$

If $F = F_n$, the above definition defines multivariate location and scatter M-estimates with auxiliary scale statistic $\sigma(F_n)$.

In the sample univariate case, the above definition reduces to an M-estimate of location with auxiliary scale statistic $\sigma(F_n)$. Location is often viewed as the important statistic in the univariate setting, with scale being viewed as an auxiliary or nuisance parameter. In the multivariate setting, though, the scatter statistic is also of central importance, with primary interest usually being focused on the shape components; that is, functions such that $H(V) = H(\lambda V)$ for any $\lambda > 0$; see Tyler (1983) for more details. These shape components can be identified through $\Gamma(F_n)$ alone and so the scale statistic $\sigma(F_n)$ can still be regarded as an auxiliary statistic.

As in the univariate setting or the regression setting, one can use either a preliminary estimate or a simultaneous estimate of scale. To assure that $(\boldsymbol{\mu}(F), V(F))$ is affine equivariate, the scale statistic or functional must be invariant in the following sense. If $F_{\mathbf{Y}}$ represents the p-dimensional distribution of $\mathbf{Y} = \mathbf{BX}+\mathbf{b}$, where $\mathbf{B}$ is a nonsingular matrix of order p and $\mathbf{b} \in \Re^p$, then

$$\sigma(F_{\mathbf{Y}}) = |\det(\mathbf{B})|^{1/p}\sigma(F_{\mathbf{X}}). \tag{8}$$

Both the S-estimates and the CM-estimates satisfy the definition of M-estimates with auxilary scale and both depend upon a simultaneous estimate of scale. The difference between the S-estimates and the CM-estimates lie in the definition of the auxiliary scale statistic. To be specific, we state the following theorem. The proof of the theorem is fairly straightforward for the CM-functionals. The proof for the S-functionals involves a few additional arguments and is given in the Appendix.

THEOREM 2.1. *For a given ρ-function, suppose*$(\boldsymbol{\mu}(F), V(F))$ *represents either an S-functional or a CM-functional of multivariate location and scatter. If we set* $\sigma(F) = \det(V(F))^{1/(2p)}$, *then* (6) *obtains its minimum over* $\Theta(p)$ *at* $(\boldsymbol{\mu}, \Gamma) = (\boldsymbol{\mu}(F), \sigma^{-2}(F)V(F))$. *That is,* $(\boldsymbol{\mu}(F), V(F))$ *corresponds to a multivariate M-functional with auxiliary scale* $\sigma(F)$.

For the S-functionals, the scale is defined simultaneously as the M-functional of scale corresponding to the largest value of $\sigma(F)$ satisfying the inequality

$$E\big[\rho((\mathbf{X}-\boldsymbol{\mu})'\Gamma^{-1}(\mathbf{X}-\boldsymbol{\mu})/\sigma^2)\big] \le \varepsilon\rho(\infty) \tag{9}$$

for fixed $\boldsymbol{\mu}$ and Γ. This formulation gives an interpretation to the term "S" in S-functionals which is analogous to its interpretation in the regression setting. That is, rather than viewing the S-functional as minimizing the "scale" $\det(V)$, we can view it as choosing $\boldsymbol{\mu}$ and Γ for which the corresponding M-functional of scale is minimized.

For the CM-functionals, the scale is defined simultaneously as the value $\sigma(F)$ which minimizes

$$E\big[\rho((\mathbf{X}-\boldsymbol{\mu})'\Gamma^{-1}(\mathbf{X}-\boldsymbol{\mu})/\sigma^2)\big] + p\log(\sigma) \tag{10}$$

for fixed $\boldsymbol{\mu}$ and Γ, subject to the constraint $\sigma \geq \sigma_S$, where σ_S is the largest value satisfying (9).

If we define $\sigma(F)$ as a preliminary scale functional, then this gives a new class of multivariate location and scatter functional which we call the MM-functionals, or in the case $F = F_n$, the MM-estimates. The class of MM-estimates were first introduced by Yohai (1987) in the regression setting. The motivating idea behind their definition is that one can begin with a high breakdown point but inefficient S-estimate as a preliminary regression estimate. One then uses the scale based upon this preliminary estimate along with a better tuned ρ-function to obtain a more efficient M-estimate of regression while maintaining the high breakdown point.

In the multivariate setting, one can obtain a preliminary scale estimate by taking $\sigma(F_n) = \det(V_o(F_n))^{1/(2p)}$, where $V_o(F_n)$ is some preliminary high breakdown point S-estimate of scatter. Once given the scale estimate, a different ρ-function can be used in defining an M-estimate with auxiliary scale $\sigma(F_n)$. Note that if the same ρ-function were used, then the resulting MM-estimate would be the same as the preliminary S-estimate. Properties of the multivariate MM-estimates are similar to those of the multivariate CM-estimates. It is not the intent of this paper to pursue this topic here. The reader can refer to Mendes and Tyler (1996) for a comparison of the CM-estimates and MM-estimates in the regression setting.

A multivariate version of the MM-estimates for location has also been proposed by Lopuhaä (1992). His definition differs in that it treats the entire scatter matrix as a preliminary auxiliary statistic which then involves minimizing (6) over $\boldsymbol{\mu} \in \Re^p$ only.

We conclude this section by noting that it is difficult to obtain simple conditions to ensure in general the existence of the M-functionals with auxiliary scale. The following lemma, however, does assure their existence for any absolutely continuous distribution. Simpler conditions, such as Condition 2.2 for the S and CM-functionals, would depend on the exact definition of $\sigma(F)$.

THEOREM 2.2. *Let ρ satisfy Condition* 2.1.
(a) *For a given $\sigma(F)$, if for all hyperplanes $B \in \Re^p$,*

$$\inf_{\Theta(p)} E\big[\rho\{(\mathbf{X}-\boldsymbol{\mu})'\Gamma^{-1}(\mathbf{X}-\boldsymbol{\mu})/\sigma^2(F)\}\big] < \{1 - P_F(B)\}\rho(\infty),$$

then there exists $(\boldsymbol{\mu}_o, \Gamma_o) \in \Theta(p)$ which minimizes (6).

(b) *For a given* $\sigma(F)$, *if for some hyperplane* $B \in \Re^p$,

$$\inf_{\Theta(p)} E\big[\rho\{(\mathbf{X}-\boldsymbol{\mu})'\Gamma^{-1}(\mathbf{X}-\boldsymbol{\mu})/\sigma^2(F)\}\big] > \{1 - P_F(B)\}\rho(\infty),$$

then there exists no $(\boldsymbol{\mu}_o, \Gamma_o) \in \Theta(p)$ *which minimizes* (6).

3. Uniqueness results for location.

3.1. *Symmetric unimodal distributions.* Some conditions on the distribution are necessary in order to obtain uniqueness results for the M-functionals with auxiliary scale in general. For example, consider a distribution which is a 50–50 mixture of uniform distributions inside the p-dimensional unit spheres centered at plus and minus $\mathbf{c}$, with $\|\mathbf{c}\| \geq 1$. For this distribution, it is easy to see that the minimum volume ellipsoids covering half of the distribution correspond to either of the two unit spheres. Thus, the MVE-functional is not unique in this case and neither of the location components corresponds to the center of symmetry. This example suggests that besides symmetry, some type of unimodality is needed to assure that the location functional gives the center of symmetry.

We begin with the following definition from Anderson (1955).

DEFINITION 3.1. A function f on $\Re^p$ is said to be symmetric and unimodal if:

(a) $f(\mathbf{x}) = f(-\mathbf{x})$ for all $\mathbf{x} \in \Re^p$.

(b) $H_u = \{\mathbf{x}\colon f(\mathbf{x}) \geq u\}$ is convex for all $0 \leq u < \infty$.

If a function f is symmetric and unimodal then it is radially decreasing. That is, for any $\mathbf{x} \in \Re^p$ if $\alpha_1 > \alpha_2 \geq 0$ then $f(\alpha_1\mathbf{x}) \leq f(\alpha_2\mathbf{x})$. If the second inequality is always strict, then f is said to be strictly radially decreasing. We will also say that such functions are strictly unimodal. Unimodal elliptically symmetric distributions satisfy the conditions of Definition 3.1.

Throughout this subsection we assume the random variable $\mathbf{X}$ has a symmetric unimodal probability density function f on $\Re^p$, as defined in Anderson (1955). For such distributions, one would anticipate that any reasonable location functional would give the origin. Such is shown here to be the case for any M-functional of location with auxiliary scale. In addition, if f is strictly unimodal, then it is shown that the location functionals are uniquely defined to be the origin.

To help obtain our results, we recall Anderson's (1955) classic inequality for probabilities of convex symmetric sets when shifted away from the origin.

LEMMA 3.1. *Let* $\mathbf{X}$ *have a symmetric and unimodal pdf, and let h be a symmetric function on* $\Re^p$ *such that* $E_v = \{\mathbf{x} \in \Re^p | h(\mathbf{x}) \leq v\}$ *is convex, where* $v \in \Re$. *Then for any* $\boldsymbol{\mu} \in \Re^p$ *and* $0 \leq \alpha < 1$,

$$P[h(\mathbf{X} + \alpha\boldsymbol{\mu}) \leq v] \geq P[h(\mathbf{X} + \boldsymbol{\mu}) \leq v]. \tag{11}$$

In order to obtain our uniqueness results, we will need strict inequality in (11) for some v. This would allow us to conclude that $E[h(\mathbf{X} + \alpha\boldsymbol{\mu})] < E[h(\mathbf{X} + \boldsymbol{\mu})]$. According to Corollary 1 of Anderson (1955), equality in (11) holds if and only if for every $u, (E_v + \boldsymbol{\mu}) \cap H_u = E_v \cap H_u + \boldsymbol{\mu}$. Note that if $v > \rho(\infty)$ then $E_v = \Re^p$ and so (11) is an equality since both sides are equal to 1. Hence, if $E_v = \Re^p$, then the condition of Corollary 1 of Anderson (1955) reduces to $H_u = H_u + \boldsymbol{\mu}$. This clearly does not hold for typical $\mathbf{X}$. Thus, some additional conditions are needed in order to assure equality in (11), but we do not pursue this line of thought here.

Anderson (1955) also gives the following necessary and sufficient condition for (11) to hold with strict inequality. Let $\mathscr{V}_p[\cdot]$ denote the volume of the enclosed subset of $\Re^p$. Strict inequality holds in (11) if and only if there exists u such that

$$\mathscr{V}_p\big[(E_v + \alpha\boldsymbol{\mu}) \cap H_u\big] > \mathscr{V}_p\big[(E_v + \boldsymbol{\mu}) \cap H_u\big]. \tag{12}$$

Inequality (12) may be difficult to check for general f and E_v. We can use it, however, to obtain a simple sufficient condition for strict inequality to hold in (11) for the special case of interest in this section, that is, when $h(\mathbf{x}) = \rho(\mathbf{x}'V^{-1}\mathbf{x})$ and ρ satisfies Condition 2.1. Here $\supset$ denotes a strict inclusion.

LEMMA 3.2. *Suppose* $\mathbf{X}$ *has a symmetric and unimodal pdf and* $0 \le \alpha < 1$. *If for some* $u > 0$, $(E_v + \alpha\boldsymbol{\mu}) \cap H_u$ *is not contained in a hyperplane and*

$$(E_v + \alpha\boldsymbol{\mu}) \cap H_u \supset \delta((E_v + \boldsymbol{\mu}) \cap H_u) + (1-\delta)((E_v - \boldsymbol{\mu}) \cap H_u), \tag{13}$$

where $E_v = \{\mathbf{x} \in \Re^p : \rho(\mathbf{x}'V^{-1}\mathbf{x}) \le v\}$ *and* $\delta = (1+\alpha)/2$, *then* (11) *holds with strict inequality.*

LEMMA 3.3. *Let* $\mathbf{X}$ *have a symmetric and unimodal pdf* f, *and suppose* ρ *satisfies Condition* 2.1. *For* $(\boldsymbol{\mu}, V) \in \Re^p \times PDS(p)$ *and* $0 \le \alpha < 1$,

$$E\big[\rho\{(\mathbf{X} - \alpha\boldsymbol{\mu})'V^{-1}(\mathbf{X} - \alpha\boldsymbol{\mu})\}\big] \le E\big[\rho\{(\mathbf{X} - \boldsymbol{\mu})'V^{-1}(\mathbf{X} - \boldsymbol{\mu})\}\big]. \tag{14}$$

Further, if f *is symmetric and strictly unimodal and* $\boldsymbol{\mu} \neq 0$, *then*

$$E\big[\rho\{\mathbf{X}'V^{-1}\mathbf{X}\}\big] < E\big[\rho\{(\mathbf{X} - \boldsymbol{\mu})'V^{-1}(\mathbf{X} - \boldsymbol{\mu})\}\big]. \tag{15}$$

REMARK. Lemma 3.3 is analogous to Lemma 4 in Davies (1987). His result, though, only applies to elliptically symmetric distributions.

Lemma 3.3 allows us to conclude that any M-functional of location with auxiliary scale is uniquely given by the origin, or more generally the center of symmetry, whenever the distribution has a symmetric strictly unimodal density.

THEOREM 3.1. *Let F denote the distribution of* $\mathbf{X}$ *with* $\mathbf{X}$ *having a symmetric strictly unimodal pdf, and suppose* ρ *satisfies Condition* 2.1. *For a given* $\sigma(F)$, *there exists* $(\boldsymbol{\mu}(F), \Gamma(F)) \in \Theta(p)$ *which minimizes* (6). *For any such solution* $\boldsymbol{\mu}(F) = 0$.

3.2. *Coordinatewise symmetric unimodal distributions.* Although the class of symmetric unimodal distributions is broader than the unimodal elliptical distributions, there are other distributions of interest which are not symmetric and unimodal as defined by Definition 3.1. In particular, consider a distribution in $\Re^p$ whose components are independent and identically distributed. If the marginal distribution is a double exponential centered at zero, then the joint distribution is symmetric and unimodal. However, if the marginal distribution has longer tails such as a Cauchy distribution centered at zero, then the joint distribution is not symmetric and unimodal since the contours of the joint distribution are not convex. A contour plot of the joint density of two independent univariate standard Cauchy distributions is illustrated in Figure 1a.

A simpler and more general definition for a symmetric unimodal or strictly unimodal distribution would be one with a symmetric radially decreasing or a symmetric strictly radially decreasing density, respectively. It is easy, however, to construct such distributions for which the MVE location functional is not unique and does not correspond to the center of symmetry. Consider again the example of a 50–50 mixture of uniform distributions inside the unit spheres centered at plus and minus $\mathbf{c}$. If $\|\mathbf{c}\| = 1$, then the distribution satisfies the more general definition of symmetric unimodal distribution with its center of symmetry at the origin. However, the center of symmetry does not correspond to one of the possible location components of the MVE-functional. This still holds even if one mixes this density with a small enough percentage of a standard normal distribution in $\Re^p$, producing a symmetric strictly radially decreasing density with support $\Re^p$. Thus, some additional restrictions on this class of symmetric unimodal distributions are needed in order to obtain general uniqueness results for the M-functionals with auxiliary scale.

We proceed by assuming in this subsection that the p-dimensional random vector $\mathbf{X} = (X_1, \dots, X_p)'$ possesses a density $f(x_1, \dots, x_i, \dots, x_p)$ which is a symmetric, strictly unimodal univariate function of x_i, for each $i = 1, \dots, p$, when all other coordinates are held fixed. We say each coordinate slice of f is symmetric and strictly unimodal. For this class of distributions, we show not only that any M-functional of location with auxiliary scale is uniquely defined to be $\mathbf{0}$, but also that the scatter functional must be a diagonal matrix. Due to the affine equivariance of the functionals, the uniqueness of the location functional carries over to affine transformations of this class of distributions. Note that any distribution or any affine transformation of a distribution whose density is symmetric and strictly unimodal in each coordinate slice is symmetric about some point and is strictly radially decreasing from the point of symmetry.

For a symmetric positive definite matrix V, express $V = TT'$, where T is a lower triangular matrix and let

$$U_p = T^{-1} = \begin{pmatrix} \mathbf{u}_1' \\ \vdots \\ \mathbf{u}_p' \end{pmatrix} = \begin{bmatrix} u_{11} & 0 & 0 & \cdots & 0 \\ u_{21} & u_{22} & 0 & \cdots & 0 \\ & & \vdots & & \\ u_{p1} & u_{p2} & u_{p3} & \cdots & u_{pp} \end{bmatrix}.$$

The following lemma is the key result needed to obtain the uniqueness results of this subsection.

LEMMA 3.4. *If* $\mathbf{X}$ *have density* f *in* $\Re^p$ *for which each coordinate slice is symmetric and unimodal and* ρ *satisfies Condition* 2.1, *then*

$$E\big[\rho\{(\mathbf{X}-\boldsymbol{\mu})'V^{-1}(\mathbf{X}-\boldsymbol{\mu})\}\big] \geq E\big[\rho(u_{11}^2 X_1^2 + \cdots + u_{pp}^2 X_p^2)\big]. \tag{16}$$

Further, if each coordinate slice of f *is symmetric and stricly unimodal, then the inequality is strict.*

If we let $D_p^{-1} = \operatorname{diag}(u_{11}^2, \ldots, u_{pp}^2)$, then $\det(V) = \det(D_p)$. Thus, it follows directly from its definition that any multivariate location and scatter M-functional with auxiliary scale must be of the form $(\mathbf{0}, \Delta)$ for some diagonal matrix Δ. We summarize this in the following theorem.

THEOREM 3.2. *Suppose* $\mathbf{X}$ *has a density* f *in* $\Re^p$ *for which each coordinate slice is symmetric and strictly unimodal, and* ρ *satisfies Condition* 2.1. *Let* F *denote the distribution of* $\mathbf{X}$. *For a given* $\sigma(F)$, *there exists* $(\boldsymbol{\mu}(F), \Gamma(F)) \in \Theta(p)$ *which minimizes* (6). *For any such solution,* $\boldsymbol{\mu}(F) = 0$ *and* $V(F)$ *is a diagonal matrix.*

The proof of Lemma 3.4 is essentially based upon the recursive application of the following lemma, which can be viewed as a univariate version of Anderson's (1955) inequality on the integral of symmetric unimodal functions over shifted convex sets.

LEMMA 3.5. *Let* $v(s)$, $s \geq 0$, *be nondecreasing in* s, $g: \Re \to \Re$ *be symmetric and nonincreasing in* $|y|$. *Also if*

$$I(c) = \int_{-\infty}^{\infty} v((c+y)^2) g(y)\, dy$$

is integrable for each $c \in \Re$, *then* $I(c) \geq I(0)$. *Further, if either* v *is strictly increasing or* g *is strictly radially decreasing,* $I(c) > I(0)$.

It is possible to unify the results of Sections 3.1 and 3.2 and make a more general statement about the uniqueness of the location functional. For example, it can be shown that if $\mathbf{X} = (X_1, \ldots, X_p)'$ and if for $r \leq p$, $(X_1, \ldots, X_r)$ has a symmetric unimodal distribution with radially strictly decreasing density function, then $(\mu_1, \ldots \mu_r)' = \mathbf{0}$. For the sake of clarity and brevity, we do not give a full treatment here.

4. Joint uniqueness of location and scatter. In Section 3.2, we considered a class of symmetric distributions for which not only are the location M-functionals with auxiliary scale uniquely defined to be the center of symmetry, but also for which the scatter functional must be a diagonal matrix. Under some additional conditions on this class of distributions, we show in this section that the scatter functional is uniquely defined to be a constant times the identity matrix. Thus, we obtain a class of distributions for which the location and scatter M-functionals with auxiliary scale are uniquely defined. Again due to the affine equivariance of the functionals, the simultaneous uniqueness of the location and scatter functionals carries over to affine transformations of this class of distributions.

We begin by introducing some concepts and notations before precisely defining the class of distributions considered in this section. Let $O(p)$ be the group of $p \times p$ orthogonal matrices, and let W_p be the subgroup of $O(p)$ generated by permutation matrices and reflection matrices, that is, diagonal matrices with each diagonal entry being $+1$ or -1. We say a function $f: \Re^p \to \Re$ is W_p-invariant if $f(y) = f(wy)$ for all $y \in \Re^p$ and $w \in W_p$. Examples of W_p-invariant functions include the density of a spherically symmetric random vector, the density of a random vector for which the entries are independent and identically distributed symmetric random variables, or in general the density of a symmetric exchangeable random vector.

A major tool in this section is the concept of majorization, which represents a partial ordering of vectors in $\Re^p$. We recall the definitions of majorization and Schur-concavity as given in Marshall and Olkin (1979). Given a vector $\mathbf{x} = (x_1, \dots, x_p)' \in \Re^p$, let $x_{[1]} \geq \cdots \geq x_{[p]}$ denote the components of $\mathbf{x}$ in decreasing order.

DEFINITION 4.1. For $\mathbf{x}, \mathbf{y} \in \Re^p$,

$$\mathbf{x} \prec \mathbf{y} \text{ if } \begin{cases} \sum_{j=1}^{k} x_{[j]} \leq \sum_{j=1}^{k} y_{[j]}, & \text{for } 1 \leq k \leq p-1, \text{ and} \\ \sum_{j=1}^{p} x_{[j]} = \sum_{j=1}^{p} y_{[j]}. \end{cases}$$

We say $\mathbf{x}$ is majorized by $\mathbf{y}$.

DEFINITION 4.2. A function $f: \Re^p \to \Re$ is said to be Schur-concave provided for every $\mathbf{x}, \mathbf{y} \in \Re^p$ with $\mathbf{x} \prec \mathbf{y}$, $f(\mathbf{x}) \geq f(\mathbf{y})$. The function f is said to be strictly Schur-concave if $\mathbf{x} \prec \mathbf{y}$ implies strict inequality holds unless there exists a permutation w such that $w\mathbf{x} = \mathbf{y}$.

Let $\Re^p_+$ denote the set of positive p-tuples of real numbers. For a W_p-invariant function f in $\Re^p$, f is Schur-concave on $\Re^p_+$ if and only if f is symmetric and unimodal; see Marshall and Olkin (1979). As in Section 3.2 we wish to be able to consider densities that do not necessarily have convex contours and so we need a concept more general than Schur-concavity. For our purposes we formally introduce a multiplicative version of majorization which we call M-majorization. This concept was informally used by Davies (1987). It formalizes the vague notion that the components of a vector $\mathbf{x}$ are "less spread out" than the components of a vector $\mathbf{y}$.

DEFINITION 4.3. For $\mathbf{x},\ \mathbf{y} \in \Re_+^p$,

$$\mathbf{x} \prec_M \mathbf{y} \text{ if } \begin{cases} \prod_{j=1}^{k} x_{[j]} \le \prod_{j=1}^{k} y_{[j]}, & \text{for } 1 \le k \le p-1, \text{ and} \\ \prod_{j=1}^{p} x_{[j]} = \prod_{j=1}^{p} y_{[j]}. \end{cases}$$

We say $\mathbf{x}$ is M-majorized by $\mathbf{y}$.

An alternate definition of M-majorization is the following: for $\mathbf{x}, \mathbf{y} \in \Re_+^p$,

$$\mathbf{x} \prec_M \mathbf{y} \text{ provided } \log\ \mathbf{x} \prec \log\ \mathbf{y},$$

where $\log \mathbf{x} = (\log x_1, \ldots, \log x_p)'$ and $\prec$ refers to the usual notion of majorization. Note there is a correspondence between points $\mathbf{x} \in \Re_+^p$ and ellipses with lengths of semi-major axes $x_1, \ldots, x_p$. Intuitively, $\mathbf{x} \prec_M \mathbf{y}$ means the ellipse corresponding to $\mathbf{x}$ is "less disperse" than the ellipse (with the same volume) corresponding to $\mathbf{y}$. The "least disperse" ellipse among all ellipses with the same volume is the sphere.

Analogously to Schur-concavity, we define M-concavity not only for functions defined on $\Re_+^p$ but also for W_p-invariant functions on $\Re^p$. Note that W_p-invariant functions are determined by their definition on $\Re_+^p$.

DEFINITION 4.4. A function $f\colon \Re_+^p \to \Re$ or a W_p-invariant function $f\colon \Re^p \to \Re$ is said to be M-concave provided for every $\mathbf{x}, \mathbf{y} \in \Re_+^p$ and $\mathbf{x} \prec_M \mathbf{y}$, $f(\mathbf{x}) \ge f(\mathbf{y})$. The function f is said to be strictly M-concave if $\mathbf{x} \prec_M \mathbf{y}$ implies strict inequality holds unless there exists a permutation $w \in W_p$ such that $w\mathbf{x} = \mathbf{y}$.

To obtain the definition of M-convexity just reverse the inequalities. Note that f is M-concave if and only if $f \circ \exp$ is Schur-concave, where $f \circ \exp(\mathbf{x}) = f(e^{x_1}, \ldots, e^{x_p})$. We say a function f is M-concave, instead of M-decreasing, to stay consistent with the classical terminology of Schur. The concept of M-concavity is broader than Schur-concavity in the following sense.

LEMMA 4.1. *Let $f\colon \Re^p \to \Re$ be Schur-concave such that $f|\Re_+^p$ is decreasing in each argument. Then $f|\Re_+^p$ is M-concave. For W_p-invariant functions, if f is Schur-concave and $f|\Re_+^p$ is decreasing in each argument then if is M-concave.*

Examples of Schur-concave functions include a spherically symmetric, radially decreasing probability density function, or more generally W_p-invariant probability density functions with convex contours. The most extreme example of a Schur-concave W_p-invariant probability density is one with contours of the form $\{\mathbf{x} \in \Re_+^p | x_1 + \cdots + x_p = c\}$ $\Re_+^p$. This is obtained by the joint density of i.i.d. double exponential distributions; that is,

$$f(x_1, \ldots, x_p) \propto \exp(-|x_1| - \cdots - |x_p|)$$

is Schur-concave.

An example of an M-concave function that is not Schur-concave is the joint density of i.i.d. t-distributions on $\nu > 0$ degrees of freedom; that is,

$$f(x_1, \ldots, x_p) \propto \prod_{i=1}^{p} \left(1 + \frac{x_i^2}{\nu}\right)^{-(\nu+1)/2}. \tag{17}$$

For this case, f is not Schur-concave since $\{\mathbf{x} \in \Re^p \mid f(\mathbf{x}) \geq c\}$ is not convex. The proof of the M-concavity of f in (17) is given in the Appendix. The most extreme example of an M-concave W_p-*invariant* function is one with contours on $\Re_+^p$ of the form $\{\mathbf{x} \in \Re_+^p | x_1 \cdots x_p = c\}$, which is approached by the joint density of the i.i.d t-distributions as $\nu \to 0$.

From a graphical perspective, we note that a W_p-*invariant* density is M-concave if and only if the contours on $\Re_+^p$ are convex when plotted on logarithmic axes. Figure 1 illustrates this for the case of two independent standard univariate Cauchy distributions.

A fundamental result due to Marshall and Olkin (1979) concerns preservation of Schur-concavity under convolution: if ϕ and g are Schur-concave functions defined on $\Re^p$, then the function Ψ defined on $\Re^p$ by

$$\Psi(\theta) = \int_{\Re^p} g(\theta - \mathbf{x})\phi(\mathbf{x})\, d\mathbf{x}$$

is Schur-concave. The following analogous result for M-concavity is the key result for obtaining the uniqueness of the scatter functionals. This result can be found on page 300 of Marshall and Olkin (1979). Our statement and proof [based on Marshall and Olkin (1979), page 100] uses the terminology of M-majorization and M-concavity.

THEOREM 4.1. *Let* $f: \Re^p \to \Re$ *be* W_p*-invariant and M-concave. Let* $h: \Re^p \to \Re$ *be* W_p*-invariant and M-convex. Then*

$$I(\Lambda) = \int h(\lambda_1 y_1, \ldots, \lambda_p \mathbf{y}_p) f(y)\, d\mathbf{y}$$

is M-convex as a function of $\Lambda = (\lambda_1, \ldots, \lambda_p)'$, *where each* $\lambda_i > 0$. *Further, if either* f *is strictly M-concave or* h *is strictly M-convex*, I *is strictly M-convex.*

In particular, if $\rho: \Re \to \Re$ satisfies Condition 2.1, then since the mapping $\mathbf{y} \to \mathbf{y}'\mathbf{y}$ is M-convex, the map $\mathbf{y} \to \rho(\mathbf{y}'\mathbf{y})$ is M-convex, so Theorem 4.1 applies when h is taken to be $h(\mathbf{y}) = \rho(\mathbf{y}'\mathbf{y})$.

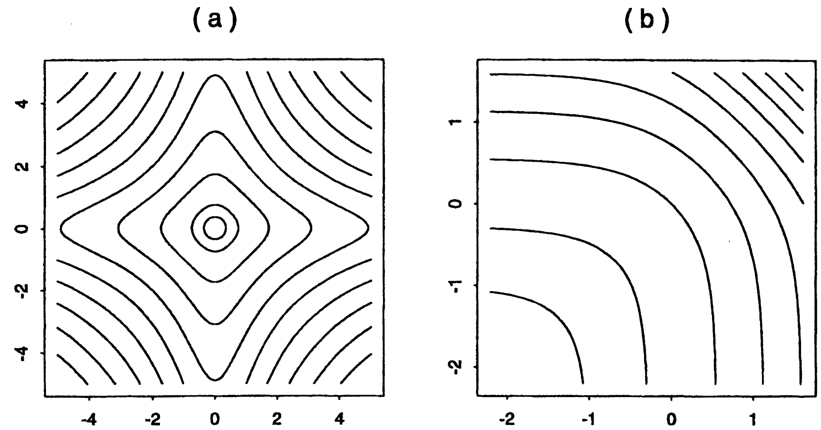

FIG. 1. (a) *The contours for the joint density of two independent standard univariate Cauchy distributions.* (b) *The same contours plotted using a log-log scale for the axes.*

Recall now the class of distributions $\mathscr{P}_p$ defined in the introduction. Using the terminology of the present section, the class $\mathscr{P}_p$ is the class of distributions on $\Re^p$ having densities which are W_p-*invariant* and M-concave. Densities associated with this class have coordinate slices which are symmetric and unimodal, and so Lemma 3.4 of Section 3.2 applies. Using the notation of Section 3.2, and noting that $\kappa^{1/2}\mathbf{1}_p' \prec_M (u_{11}, \dots, u_{pp})'$, where $\kappa^p = \prod_{i=1}^p u_{ii}^2 = \det(V^{-1})$ and where $\mathbf{1}_p$ is the p-dimensional vector with each entry 1, we can put Lemma 3.4 and Theorem 4.1 together to obtain the following key inequality.

LEMMA 4.2. *If ρ satisfies Condition* 2.1 *and the distribution of* $\mathbf{X}$ *is in* $\mathscr{P}_p$, *then*

$$E\big[\rho\{(\mathbf{X}-\boldsymbol{\mu})'V^{-1}(\mathbf{X}-\boldsymbol{\mu})\}\big] \geq E[\rho(\kappa\mathbf{X}'\mathbf{X})].$$

Further, if either ρ is strictly increasing or the density f of $\mathbf{X}$ *is strictly M-concave, then the inequality is strict.*

Since $\det(V) = \det(\kappa^{-1}I_p)$, it follows that for any multivariate M-functional with auxiliary scale the location component must be $\mathbf{0}$ and the scatter component must be proportional to the identity matrix. We summarize this in the following uniqueness theorem.

THEOREM 4.2. *Suppose ρ satisfies Condition* 2.1 *and the distribution F of* $\mathbf{X}$ *is in* $\mathscr{P}_p$. *Further, suppose either ρ is strictly increasing or the density f of* $\mathbf{X}$ *is strictly M-concave. Then, for a given $\sigma(F)$, there exists $(\boldsymbol{\mu}(F), \Gamma(F)) \in \Theta(p)$ which minimizes* (6). *For any such solution,* $\boldsymbol{\mu}(F) = \mathbf{0}$ *and* $V(F) = \sigma^2(F)I_p$.

The uniqueness of any particular multivariate M-functional with auxiliary scale also depends upon the uniqueness of the scale functional $\sigma(F)$. The uniqueness of the S-functionals hold since $\sigma(F)$ is uniquely defined as the largest value of σ which satisfies the inequality

$$E\big[\rho(\mathbf{X}'\mathbf{X}/\sigma^2)\big] \le \varepsilon\rho(\infty). \tag{18}$$

This gives a unique value for $\sigma(F)$ since the left-hand side of (18) is a strictly decreasing function of σ. The uniqueness of the CM-functional depends upon the uniqueness of the solution to the problem of minimizing

$$E\big[\rho(\mathbf{X}'\mathbf{X}/\sigma^2)\big] + p\,\log(\sigma) \tag{19}$$

subject to the constraint $\sigma \ge \sigma_S$, where σ_S is the unique largest value satisfying (18). As discussed in Kent and Tyler (1996), for a specific ρ function and distribution F, this uniqueness problem can in general be checked numerically since (19) is a univariate function. In specific cases it can be verified theoretically. This is discussed in more detail in Kent and Tyler (1997).

We conclude by again noting that by the affine equivariance properties of the M-functionals with auxiliary scale, if $\mathbf{Y} = \mathbf{B}\mathbf{X} + \mathbf{b}$ where $\mathbf{B}$ is a nonsingular matrix of order p and $\mathbf{b} \in \Re^p$, then an M-functional with auxiliary scale under the distribution of $\mathbf{Y}$ must be of the form $(\mathbf{b}, \sigma^2(F)\mathbf{B}\mathbf{B}')$ when the conditions on ρ and $\mathbf{X}$ in Theorem 4.2 are satisfied.

5. Some comments on uniqueness. It was noted in the introduction that many properties of the S-estimates and CM-estimates under a given distribution follow once the uniqueness of the corresponding S-functionals and CM-functionals is established. The results of this paper show that these properties hold not only for unimodal elliptical distributions but also for much broader classes of unimodal symmetric distributions. This does not imply that the associated estimating equations have a unique solution, but only that the one which optimizes the S or CM criterion is unique.

The uniqueness results can be extended to the estimates themselves. That is, for large enough random samples from a distribution for which an S-functional or a CM-functional is uniquely defined, the corresponding S-estimate or CM-estimate will be uniquely defined; see Kent and Tyler (1996) for further discussion. In practice, though, we conjecture that the S-estimates or CM-estimates themselves are unique with probability 1 when random sampling from an absolutely continuous distribution, even one for which the corresponding S- or CM-functional is not uniquely defined. Davies (1992) shows this to be true for the MVE-estimates. It is the uniqueness of the functional, though, which is important in establishing results such as consistency.

The nonuniqueness of a functional under some distributions need not be viewed as an undesirable property, but rather simply as a descriptive property. Knowing when a functional is uniquely defined and when it is not uniquely defined can give some insight into the nature of the functional. Recall the example from Section 3 consisting of a 50–50 mixture of uniform distributions within unit spheres with centers plus and minus $\mathbf{c}$. One may argue that having two solutions is desirable here, one with location minus $\mathbf{c}$ and the other with location plus $\mathbf{c}$ since this exposes the structure of the distribution. Any uniquely defined affine equivariate functional would give the location as the origin, which may mask the two clusters. Application of this argument to the special case $\mathbf{c} = \mathbf{1}/\sqrt{p}$, for which $\|\mathbf{c}\| = 1$, and its variation as discussed in Section 3.2, supports the seemingly unorthodox statement:

> It is not necessarily desirable that a location functional corresponds to the center of symmetry for a distribution with a symmetric, exchangeable and strictly radially decreasing density.

Of course, for a random sample from such distributions, the MVE-estimate will have a unique solution. The MVE-estimate, though, would tend to be near one of the two MVE-functionals and the cluster structure could then easily be picked up in a residual analysis.

Davies (1993) gives another example of when the MVE-functional is not uniquely defined. He considers a 40–15–15–15–15 mixture of uniform, distributions within spheres of radius r in $\Re^2$ with centers $\mathbf{0}, (1,0)', (0,1)', (-1,0)'$ and $(0,-1)'$, respectively. He notes that for small enough r, any minimum volume ellipsoid will concentrate on the component at the origin and two diametrically opposed components of the remaining four. This results in two solutions for the MVE-functional, both with location $\mathbf{0}$ but with different scatters. Davies (1993) uses this example to illustrate that S-functionals of scatter are not necessarily uniquely defined. He views this as a fundamental difficulty of S-functionals and so argues that it may not be reasonable to apply the S-functionals to a design matrix for the purpose of downweighting leverage points in GM-estimates of regressions. One can again make the alternative argument, though, that uniqueness is not necessarily desirable here since the nonunique solution would help uncover the curious nature of the design space. One may then wish to consider separate regressions on the different clusters.

The above examples may appear somewhat pathological, as is probably the case with any distribution that produces some nonunique M-functional with auxiliary scale. These examples, though, give some insight into the nature of the uniqueness problem. One can see that trying to impose uniqueness on the functionals by defining them to be some affine equivariate average of all possible solutions is not necessarily desirable and does not make the corresponding estimates consistent. In the first example, any affine equivariate average of the MVE-functionals would give the location as the origin, but the MVE-estimates would still be near either plus or minus $\mathbf{c}$.

These examples are also indicative of the nature of redescending high breakdown point estimates. Although in practice one would seldom deal with such balanced clusters, clusters do arise in data, possibly due to some unmeasured factor. Redescending high breakdown point estimates tend to concentrate on the majority cluster or collection of clusters, while ignoring the minority. The nature of the data can then be detected through a residual analysis. Most examples which demonstrate the utility of high breakdown point estimates involve a cluster or clusters of data. For example, the classic Herzsprung–Russell star data set considered on page 261 in Rousseeuw and Leroy (1987) contains a cluster of four outliers corresponding to giant stars, which represents an unmeasured factor in the analysis.

We conclude by considering a scatter analog of the first example. This interesting example provides some insight into the relationship between M-concavity and scatter. Consider a 50–50 mixture of two bivariate normal distributions with means **0** and variance-covariance matrices

$$\begin{pmatrix} 1 & 0 \\ 0 & \gamma^2 \end{pmatrix} \quad \text{and} \quad \begin{pmatrix} \gamma^2 & 0 \\ 0 & 1 \end{pmatrix}, \tag{20}$$

respectively, for some $0 < \gamma < 1$. This distribution has a density which is symmetric and strictly unimodal in each variable, and so we know by Theorem 3.2 that any M-functional with auxiliary scale gives a unique location **0**, with the scatter functional being a diagonal matrix. The density is also exchangeable, but it is neither concave nor M-concave. Figure 2 illustrates this for the case $\gamma^2 = 0.25$. One can verify this in general by noting if $\gamma < \beta < 1$, then $f(a, a) < f(a\beta, a/\beta)$ for large enough $a > 0$ even though $(a, a) \prec_M (a\beta, a/\beta)$. Consequently, Theorem 4.2 does not apply to this example.

Whether or not the scatter functional is proportional to the identity will depend upon the specific scale functional $\sigma(F)$ and the specific ρ-function being used, as well as on the value of γ in the distribution. If $\gamma = 1$, then Theorem 4.2 applies. The scatter functional will also be uniquely defined and proportional to the identity for values of γ sufficiently close to 1, provided the scale functional $\sigma(F)$ is uniquely defined and continuous in some neighborhood of the standard normal. For small enough γ, however, the minimum volume ellipsoid covering half of the distribution will not be a circle. This

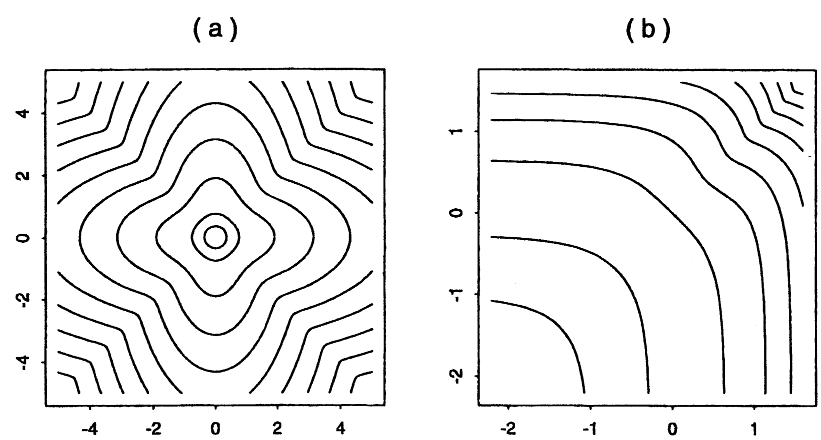

FIG. 2. (a) *The contours of the density of a* 50–50 *mixture of bivariate normal distributions. The means of the normal distributions are both zero, and the variance-covariance matrices are given in* (20) *with* $\gamma^2 = 0.25$. (b) *The same contours plotted using a log-log scale for the axes.*

implies the MVE scatter functional cannot be proportional to the identity and so by symmetry there must be at least two distinct diagonal solutions for the scatter functional.

In this last example, one can verify that the minimum volume ellipsoid is not a circle for small enough γ by noting that the volume of the circle about zero covering half the distribution converges to πr^2 as $\gamma \to 0$, where $P(|Z| < r) = 0.5$ and with Z having a standard normal distribution. On the other hand, an ellipsoid centered at zero with axes corresponding to the coordinate axes and with half-lengths $r_1 = r\kappa$ and $r_2 = r/\kappa$, respectively, for some $\kappa > 1$ also has volume πr^2. As $\gamma \to 0$, though, its limiting probability coverage is $0.5\ P(|Z| < r_1) + 0.5P(|Z| < r_2)$, which can be shown to always be greater than 0.5. Hence for small enough γ the minimum volume ellipsoid cannot be a circle. Some numerical calculations suggests that this holds for at least $\gamma < 0.38$. Using a similar argument one can show that that for small enough γ the minimum volume ellipsoid covering a proportion τ of the distribution will not be a circle for $\tau < P(|Z| < 1) = 0.68269$.

APPENDIX

PROOF OF THEOREM 2.1. Let $\Gamma(F) = V(F)/\sigma^2(F)$ and suppose $(\boldsymbol{\mu}(F), \Gamma(F))$ does not minimize (6) over $\Theta(p)$. This implies there exists $(\boldsymbol{\mu}_o, \Gamma_o)$ with $\det(\Gamma_o) = 1$ such that $(\boldsymbol{\mu}_o, V_o = \sigma^2(F)\Gamma_o)$ satisfies the S-functional constraint (2) with strict inequality. Furthermore, $\det(V_o) = \sigma^{2p}(F) = \det(V(F))$. We can then find a $\sigma_o < \sigma(F)$ such that $(\boldsymbol{\mu}_o, V_o^* = \sigma_o^2\Gamma_o)$ still satisfies the constraint (2). In addition, we have $\det(V_o^*) = \sigma_o^{2p} < \sigma^{2p}(F) = \det(V(F))$, a contradiction. □

PROOF OF THEOREM 2.2. Part (a) can be justified by showing that if $(\boldsymbol{\mu}, \Gamma)$ approaches the boundary of $\Theta(p)$ then for any $\delta > 0$, (6) is eventually greater than $\inf_B(1-P_F(B))\rho(\infty)-\delta$, where the infimum is taken over all hyperplanes $B \subset \Re^p$. Let $\lambda_1(\Gamma) \geq \cdots \geq \lambda_p(\Gamma)$ denote the eigenvalues of Γ. We first note that if $\boldsymbol{\gamma}_1(\Gamma)$ is bounded above, then (6) goes to $\rho(\infty)$ as $\|\boldsymbol{\mu}\| \to \infty$.

Consider now the case when $\lambda_1(\Gamma) \to \infty$. Since $\det(\Gamma) = 1$, this implies $\lambda_p(\Gamma) \to 0$. Let $r \geq 2$ be the smallest value for which $\boldsymbol{\gamma}_r(\Gamma) \to 0$, and let $Q_r(\Gamma)$ denote the orthogonal projection onto the space spanned by the eigenvectors of Γ associated with the roots $\lambda_r(\Gamma), \ldots, \lambda_p(\Gamma)$. We then have

$$(6) \geq E\big[\rho\{\lambda_r^{-1}(\Gamma)(\mathbf{X}-\boldsymbol{\mu})' Q_r(\Gamma)(\mathbf{X}-\boldsymbol{\mu})/\sigma^2(F)\}\big]. \tag{21}$$

If $\|Q_r(\Gamma)\boldsymbol{\mu}\| \to \infty$, then the right-hand side of (21) goes to $\rho(\infty)$.

To complete the proof of part (a) we only need to consider the case for which $\|Q_r(\Gamma)\boldsymbol{\mu}\|$ is bounded above. By compactness, we can assume without loss of generality that $Q_r(\Gamma) \to Q_r$, an orthogonal projection matrix of rank $(p-r+1)$, and that $Q_r(\Gamma)\boldsymbol{\mu} \to Q_r\boldsymbol{\mu}_o$. For this case the right-hand side of (21) goes to $\rho(\infty)(1 - P_F(B_r))$, where $B_r = \{\mathbf{x} \in \Re^p | Q_r(\mathbf{x}-\boldsymbol{\mu}_o) = 0\}$. Since B_r is contained in some hyperplane of R^p, part (a) follows.

To prove part (b), let $B = \{\mathbf{x} \in \Re^p | Q(\mathbf{x} - \boldsymbol{\mu}) = 0\}$ denote the hyperplane with largest probability mass $P_F(B)$, with Q being an orthogonal projection matrix of rank $p-1$. Consider a sequence $\Gamma = \lambda_1(I - Q) + \lambda_p Q$ with $\lambda_1 \to \infty$ and $\lambda_p = \lambda_1^{-1/(p-1)} \to 0$. For such Γ and $\boldsymbol{\mu}$, (6)$\to \rho(\infty)(1 - P_F(B))$. Hence (6) cannot have a minimum in the interior of $\Theta(p)$. □

PROOF OF COROLLARY 3.2. This proof follows closely the proof of Theorem 1 in Anderson (1955). The proof of Anderson's theorem rests on two inequalities. Set $\delta = (1+\alpha)/2$. Note $\delta\boldsymbol{\mu} + (1-\delta)(-\boldsymbol{\mu}) = \alpha\boldsymbol{\mu}$. Thus

$$E_v + \alpha\boldsymbol{\mu} = \delta(E_v + \boldsymbol{\mu}) + (1-\delta)(E_v - \boldsymbol{\mu});$$

hence

$$\begin{aligned}(E_v + \alpha\boldsymbol{\mu}) \cap H_u &= \{\delta(E_v + \boldsymbol{\mu}) + (1-\delta)(E_v - \boldsymbol{\mu})\} \cap H_u \\ &\supseteq \delta(E_v + \boldsymbol{\mu}) \cap H_u + (1-\delta)(E_v - \boldsymbol{\mu}) \cap H_u.\end{aligned}$$

Taking the volume of the first and last sets yields the first inequality of Anderson's proof. The second inequality (the Brunn–Minkowski theorem) states that

$$\begin{aligned}&\mathscr{V}_p\{\delta(E_v + \boldsymbol{\mu}) \cap H_u + (1-\delta)(E_v - \boldsymbol{\mu}) \cap H_{\boldsymbol{\mu}}\} \\ &\qquad \geq \mathscr{V}_p\{(E_v + \boldsymbol{\mu}) \cap H_u\}.\end{aligned}$$

Since $(E_v + \alpha\boldsymbol{\mu}) \cap H_u$ is not contained in a hyperplane, $(E_v + \alpha\boldsymbol{\mu}) \cap H_u$ has nonzero volume [Bonneson and Fenchel (1987)]. Thus a strict inclusion of the convex sets in the first inequality is enough to show the volumes differ [Bonnesen and Fenchel (1987), page 42]. Using (12), that is enough to show strict inequality in (11). □

PROOF OF LEMMA 3.3. We verify the strict inequality (15). It suffices to find $u > 0$ and $v > 0$ such that the conditions of Corollary 3.2 hold. Set $\mathbf{e} = t\boldsymbol{\mu} \in \partial E_v$ and $u = f(\mathbf{e} + \alpha\boldsymbol{\mu})$, where $v > 0$ is chosen so that H_u contains an open set whose closure contains $\mathbf{e} + \alpha\boldsymbol{\mu}$. For large enough v such an H_u exists since $f > 0$ and f has convex contours. Then $\mathbf{e} + \alpha\boldsymbol{\mu} \in (E_v + \alpha\boldsymbol{\mu}) \cap H_u$, and the latter set has nonzero volume.

We claim $\mathbf{e} + \alpha\boldsymbol{\mu} \notin \delta(E_v + \boldsymbol{\mu}) \cap H_u + (1-\delta)(E_v - \boldsymbol{\mu}) \cap H_u$. Otherwise we can find $\mathbf{e}_1, \mathbf{e}_2 \in E_v$ such that $\mathbf{e}_1 + \boldsymbol{\mu} \in H_u$, $\mathbf{e}_2 - \boldsymbol{\mu} \in H_u$ and

$$\begin{aligned}\mathbf{e} + \alpha\boldsymbol{\mu} &= \delta(\mathbf{e}_1 + \boldsymbol{\mu}) + (1-\delta)(\mathbf{e}_2 - \boldsymbol{\mu}) \\ &= \delta\mathbf{e}_1 + (1-\delta)\mathbf{e}_2 + \alpha\boldsymbol{\mu}.\end{aligned}$$

However, $\mathbf{e} \in \partial E_v$ and E_v is an ellipse implies $\mathbf{e} = \mathbf{e}_1 = \mathbf{e}_2$. Then $\mathbf{e} + \boldsymbol{\mu} = (t+1)\boldsymbol{\mu} \in H_u$, but this contradicts the assumption that f is strictly radially decreasing ($f((t+1)\boldsymbol{\mu}) < f((t+\alpha)\boldsymbol{\mu}) = u$). □

PROOF OF LEMMA 3.4. Recall there is a one-to-one correspondence between lower traingular matrices U_p with $u_{11}, \dots, u_{pp} > 0$ and positive-definite p-dimensional matrices. We therefore assume hereafter that each $u_{ii} > 0$. Write $\boldsymbol{\mu} = (\mu_1, \dots, \mu_p)'$. We first claim

$$\int \rho(\|U_P(\mathbf{y}-\boldsymbol{\mu})\|^2) f(\mathbf{y}) d\mathbf{y} \geq \int \rho(\|U(\mathbf{y}-\boldsymbol{\mu}_{p-1})\|^2) f(\mathbf{y}) d\mathbf{y}, \tag{22}$$

where

$$U = \begin{pmatrix} U_{p-1} & \mathbf{0}_{p-1} \\ \mathbf{0}'_{p-1} & u_{pp} \end{pmatrix}$$

and $\boldsymbol{\mu}_{p-1} = (\mu_1, \dots, \mu_{p-1}, 0)'$.

To see this we first express U_p as a partitioned matrix,

$$U_p = \begin{pmatrix} U_{p-1} & \mathbf{0}_{p-1} \\ \mathbf{u}_p^1 & \end{pmatrix},$$

where $\mathbf{0}_{p-1}$ is the $p-1$-vector of zeroes. We will make use of the identity

$$(U_p(\mathbf{y}-\boldsymbol{\mu}))'(U_p(\mathbf{y}-\boldsymbol{\mu})) = s_{p-1} + (\mathbf{u}'_p(\mathbf{y}-\boldsymbol{\mu}))^2,$$

where $s_{p-1} = \|U_{p-1}(y_1-\mu_1, \dots, y_{p-1}-\mu_{p-1})'\|^2$. Then

$$\begin{aligned} \int \rho(\|U_p(\mathbf{y}-\boldsymbol{\mu})\|^2) f(\mathbf{y})\, d\mathbf{y} &= \int_{y_1,\dots,y_{p-1}} \int_{y_p} \rho(s_{p-1} + u_{pp}^2(c_{p-1} + y_p - \mu_p)^2) \\ &\qquad \times f(y_1, \dots, y_{p-1}, y_p)\, dy_p\, dy_1 \cdots dy_{p-1} \\ &\geq \int_{y_1,\dots,y_{p-1}} \int_{y_p} \rho(s_{p-1} + u_{pp}^2 y_p^2) f(\mathbf{y})\, d\mathbf{y}, \end{aligned}$$

where $c_{p-1} = (u_{p1}/u_{pp})(y_1-\mu_1) + \cdots + (u_{p,\,p-1}/u_{pp})(y_{p-1}-\mu_{p-1})$. We check the last inequality. Since $f(y_1, \dots, y_{p-1}, y_p)$ is symmetric and unimodal in y_p for $y_1, \dots, y_{p-1}$ fixed, we can apply Lemma 3.5 to the inner integral to verify the last inequality and hence (22). Note if strict unimodality of f holds we have a strict inequality in the two inner integrals for almost all $y_1, \dots, y_{p-1}$ in $\Re^{p-1}$ (the set of all $(y_1, \dots, y_{p-1})$ satisfying $c_{p-1} - \mu_p = 0$ is a hyperplane of $\Re^{p-1}$). It follows that strict inequality holds in (22).

Thus it suffices to show

$$\int \rho(\|U(\mathbf{y}-\boldsymbol{\mu}_{p-1})\|^2) f(\mathbf{y})\, d\mathbf{y} \geq \int \rho(u_{11}^2 y_1^2 + \cdots + u_{pp}^2 y_p^2) f(\mathbf{y})\, d\mathbf{y}.$$

Note for $p = 2$ we can apply (22). Assume $p > 2$ and the lemma holds for dimension $p-1$. Then

$$\int \rho(\|U(y-\boldsymbol{\mu}_{p-1})\|^2) f(\mathbf{y})\, d\mathbf{y} = \int_{y_p} \int_{y_1,\dots,y_{p-1}} \rho(s_{p-1} + u_{pp}^2 y_p^2) f(\mathbf{y})\, d\mathbf{y}.$$

The proof is completed by induction on the inner integral. □

PROOF OF LEMMA 3.5. We take $c \geq 0$ as the two cases are similar. Note $(c+y)^2 \leq y^2$ if and only if $y \leq -c/2$ if and only if $g(y) \leq g(y+c)$. Thus for all $y \in \Re$, and since v is nondecreasing,

$$\left[v((c+y)^2) - v(y^2)\right]\left[g(y+c) - g(y)\right] \leq 0.$$

Note for $c > 0$ the inequality is strict for each y if either v or g is strictly monotone. Integrating,

$$\int v((c+y)^2) g(y+c)\,dy + \int v(y^2) g(y)\,dy$$
$$\leq \int v((c+y)^2) g(y)\,dy + \int v(y^2) g(y+c)\,dy.$$

The first two integrals equal $\int v(y^2) g(y)\,dy$. On substituting $-y' = c + y$ in the fourth integral, and using the symmetry of g about 0,

$$\int v(y^2) g(y)\,dy \leq \int v((y+c)^2) g(y)\,dy,$$

verifying the claim. □

PROOF OF LEMMA 4.1. Define $\phi(\mathbf{x}) = f(\exp(\mathbf{x}))$. From Marshall and Olkin (1979) recall ϕ is Schur-concave iff $-\phi$ is Schur-convex. Since $-f$ is Schur-convex and increasing in each argument, and exp is convex, it follows from page 63 of Marshall and Olkin that ϕ is Schur-concave. □

Before proceeding to the proof of M-concavity of f in (17), we recall a standard lemma from majorization theory [Marshall and Olkin (1979), page 58] which is useful in verifying M-concavity of functions.

LEMMA A.1. *A function* $\phi\colon \Re_+^p \longrightarrow \Re$ *is M-concave provided* $\phi(\mathbf{x}) \geq \phi(\mathbf{y})$ *whenever* $\mathbf{x}, \mathbf{y} \in \Re_+^p$, $\mathbf{x} \prec_M \mathbf{y}$ *and* $\mathbf{x}, \mathbf{y}$ *differ in at most two components.*

Recall the kth elementary symmetric functions $S_\kappa(\mathbf{x})$ are defined as

$$S_0(\mathbf{x}) \equiv 1,$$
$$S_1(\mathbf{x}) \equiv \sum_{i=1}^{p} x_i,$$
$$S_2(\mathbf{x}) \equiv \sum_{i_1 < i_2} x_{i_1} x_{i_2},$$
$$\vdots$$
$$S_p(\mathbf{x}) \equiv \sum_{i_1 < \cdots < i_p} x_{i_1} \cdots x_{i_p}.$$

In the next lemma we illustrate how one verifies a given function is M-concave. The result also follows from Schur-convexity of S_κ [Marshall and Olkin (1979)] and Lemma A.1.

LEMMA A.2. *$S_k(\mathbf{x})$ is M-convex on $\Re_+^p$. If $k < p$, S_k is strictly M-convex.*

PROOF. By Lemma A.1 it suffices to prove $S_k(\mathbf{x}) \le S_k(\mathbf{y})$ when $\mathbf{x} \prec_M \mathbf{y}$ and $\mathbf{x}, \mathbf{y}$ differ in only two components. Write $\mathbf{x} = (x_1, \dots, x_p)'$, $\mathbf{y} = (y_1, \dots, y_p)'$. From the definition of M-majorization we may take $y_1 > y_2$, $0 < a < b$ such that $ab = 1$, $x_1 = ay_1 \ge by_2 = x_2$ and $x_i = y_i$ for $3 \le i \le p$. Write

$$\begin{aligned} S_k(\mathbf{y}) = &\sum_{2<i_3<\cdots<i_k} y_1 y_2 y_{i_3} \cdots y_{i_k} \\ &+ \sum_{2<i_2<\cdots<i_k} y_1 y_{i_2} \cdots y_{i_k} \\ &+ \sum_{2<i_2<\cdots<i_k} y_2 y_{i_2} \cdots y_{i_k} \\ &+ \sum_{2<i_1<\cdots<i_k} y_{i_1} \cdots y_{i_k}. \end{aligned}$$

Note for $k = p$ we have $S_k(\mathbf{x}) = S_k(\mathbf{y})$. For $k < p$ since $0 < y_2 \le b_{y_2} \le ay_1 \le y_1$,

$$S_k(\mathbf{x}) - S_k(\mathbf{y}) = \sum_{2<i_2<\cdots<i_k} (ay_1 + by_2 - (y_1 + y_2)) y_{i_2} \cdots y_{i_k} < 0. \qquad \square$$

PROOF OF (17). From the definition of M-majorization we have for $\mathbf{x}, \mathbf{y} \in \Re_+^p$,

$$(x_1, \dots, x_p)' \prec_M (y_1, \dots, y_p)' \quad \Leftrightarrow \quad (x_1^2, \dots x_p^2)' \prec_M (y_1^2, \dots, y_p^2)'.$$

Note also that we can write

$$f(\mathbf{x}) \propto 1/(1 + S_1(x_1^2/\nu, \dots, x_p^2/\nu) + \cdots + S_p(x_1^2/\nu, \dots, x_p^2/\nu))^{(\nu+1)/2}.$$

The result is then immediate from Lemma A.2. $\square$

PROOF OF THEOREM 4.1. By Lemma A.1 it suffices to prove the theorem for dimension 2. Let $\theta = (\theta_1, \theta_2)'$, $\xi = (\xi_1, \xi_2)'$ and suppose $(\theta_1, \theta_2)' \prec_M (\xi_1, \xi_2)'$. Then

$$0 < \xi_2 < \theta_2 < 1 < \theta_1 < \xi_1,$$

where we may assume $\theta_1\theta_2 = 1 = \xi_1\xi_2$.

Then the substitutions $y_1 \to \theta_1 y_1$ and $y_2 \to \xi_2 y_2$ yield

$$\begin{aligned} I(\theta) - I(\xi) &= \int \big[h(\theta_1 y_1, \theta_2 y_2) - h(\xi_1 y_1, \xi_2 y_2)\big] f(\mathbf{y})\, d\mathbf{y} \\ &= 4/(\theta_1 \xi_2) \int_{y_1,\, y_2 \ge 0} [h(y_1, (\theta_2/\xi_2) y_2) - h((\xi_1/\theta_1) y_1, y_2)] \\ &\qquad \times f(y_1/\theta_{1,\, y2}/\xi_2)\, d\mathbf{y} \end{aligned}$$

$$= 4/(\theta_1\xi_2)\int_{y_1\geq y_2}[h(y_1,(\theta_2/\xi_2)y_2) - h((\xi_1/\theta_1)y_1, y_2)]$$
$$\times (f(y_1/\theta_1, y_2/\xi_2) - f(y_1/\xi_2, y_2/\theta_1))\,d\mathbf{y},$$

where the last integral was obtained by the change of variables $y_1 \to y_2$ and $y_2 \to y_1$ on the set $\{y \in \Re^2\colon y_2 \geq y_1\}$, and using the facts that f and h are permutation-invariant and $\theta_2/\xi_2 = \xi_1/\theta_1$.

Since $y_1 \geq y_2$ it follows that

$$(y_1, (\xi_1/\theta_1)y_2)' \prec_M ((\xi_1/\theta_1)y_1, y_2)'$$

and

$$(y_1/\theta_1, y_2/\xi_2)' \prec_M (y_1/\xi_2, y_2/\theta_1)'$$

Since h is M-convex and f is M-concave, $I(\theta) - I(\xi) \leq 0$ and so I is M-convex. □

REFERENCES

ANDERSON, T. W. (1955). The integral of a symmetric unimodal function over a symmetric convex set and some probability inequalities. *Proc. Amer. Math. Soc.* **6** 170–176.

BONNESON, T. and FENCHEL, W. (1987). *Theory of Convex Bodies*. BCS Assoc., Moscow, Idaho.

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

DAVIES, P. L. (1992). The asymptotics of Rousseeuw's minimum volume ellipsoid estimator. *Ann. Statist.* **20** 1828–1843.

DAVIES, P. L. (1993). Aspects of robust linear regression. *Ann. Statist.* **21** 1843–1899.

KENT, J. and TYLER, D. E. (1996). Constrained M-estimation for multivariate location and scatter. *Ann. Statist.* **24** 1346–1370.

KENT, J. and TYLER, D. E. (1997). Regularity and uniqueness for M-estimates of scale. Technical report, Univ. Leeds, England.

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

LOPUHAÄ, H. (1992). Highly efficient estimators of multivariate location with high breakdown point. *Ann. Statist.* **20** 398–413.

MARSHALL, A. W. and OLKIN, I. (1979). *Inequalities: Theory of Majorization and Its Applications.* Academic Press, New York.

MARTIN, R. D., YOHAI, V. J. and ZAMAR, R. H. (1989). Min–max bias robust regression. *Ann. Statist.* **17** 1608–1630.

MENDES, B. and TYLER, D. E. (1996). Constrained M-estimation for regression. *Robust Statistics, Data Analysis, and Computer Intensive Methods. Lecture Notes in Statist.* **109** 229–320. Springer, New York.

NOLAN, D. (1991). The excess-mass ellipsoid. *J. Multivariate Anal.* **39** 348–371.

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

ROUSSEEUW, P. J. and LEROY, A. M. (1987). *Robust Regression and Outlier Detection.* Wiley, New York.

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

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

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

NATIONAL INSTITUTE OF STATISTICAL SCIENCES
19 T.W. ALEXANDER DRIVE
P.O. BOX 14006
RESEARCH TRIANGLE PARK, NORTH CAROLINA 27709-4006
E-MAIL: kaysuoka@stat.rutgers.edu

DEPARTMENT OF STATISTICS
RUTGERS UNIVERSITY
110 FRELINGHUYSEN ROAD
PISCATAWAY, NEW JERSEY 08854-8019
E-MAIL: dtyler@rci.rutgers.edu