The Canadian Journal of Statistics
Vol. 14, No. 2, 1986, pages 161–176
La Revue Canadienne de Statistique

Robust *M*-estimators of multivariate location and scatter in the presence of asymmetry

D.P. WIENS* and Z. ZHENG†

Dalhousie University and Peking University

Key words and phrases: Robust estimation, location vectors, scatter matrices, *M*-estimators, asymmetry.
AMS 1980 subject classifications: Primary 62H12; secondary 62G05, 62G20.

ABSTRACT

Robust estimation of location vectors and scatter matrices is studied under the assumption that the unknown error distribution is spherically symmetric in a central region and completely unknown in the tail region. A precise formulation of the model is given, an analysis of the identifiable parameters in the model is presented, and consistent initial estimators of the identifiable parameters are constructed. Consistent and asymptotically normal *M*-estimators are constructed (solved iteratively beginning with the initial estimates) based on "influence functions" which vanish outside specified compact sets. Finally *M*-estimators which are asymptotically minimax (in the sense of Huber) are derived.

RÉSUMÉ

Cet article concerne l'estimation robuste de paramètres de position et de matrices de dispersion dans les situations où la loi des erreurs est totalement inconnue sauf pour la présence d'une symétrie sphérique à l'intérieur d'une région centrale. On formule le modèle de façon précise et on en analyse les paramètres identifiables pour lesquels on construit des estimateurs initiaux convergents. A partir de ces estimations initiales, un processus itératif nous permet de déduire des *M*-estimateurs convergents et asymptotiquement normaux. Ceux-ci sont fondés sur des "fonctions d'influence" qui s'annulent en dehors de certains ensembles compacts. Enfin, on obtient des *M*-estimateurs qui sont asymptotiquement minimax au sens de Huber.

1. INTRODUCTION AND SUMMARY

Huber (1964) developed a theory of robust estimation of a location parameter using *M*-estimators, which was later extended to a theory of robust regression (Huber 1973) and robust estimation of multivariate location and scatter (Huber 1977, Marona 1976). Collins (1976) modified the theory of Huber (1964) to show that when the error distribution in the model is assumed to be symmetric in the centre, but unknown and asymmetric in the tails, then robust estimates of location can be obtained using *M*-estimators with "redescending" influence curves. Such estimators were first considered by F. Hampel—see Andrews *et al.* (1972). The same sort of modification of Huber's theory was successfully extended

*Research supported by the National Sciences and Engineering Research Council of Canada under Grant No. A8603.

†Research completed while visiting the Department of Statistics, University of California, Berkeley.

to the case of robust regression by Collins, Sheahan, and Zheng (1981), and to estimation of multivariate location, with scale known, by Collins (1982).

In this paper we carry out this type of modification to the theory of robust estimation of unknown location vectors and scatter matrices. That is, we use redescending M-estimators to obtain robust estimates of location and scatter in the presence of asymmetry in the tails of the error distribution. In Section 2 the model is presented. It assumes that the distribution is elliptically symmetric inside a given ellipsoid $\{(\mathbf{x} - \boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\mathbf{x} - \boldsymbol{\mu}) \leq r^2\}$, $r \leq \infty$, and arbitrary outside. Inside the ellipsoid, the distribution arises through ϵ-contamination of a known law. If $r = \infty$, the model then coincides with that of the location–scatter-estimation problem of Huber (1977). There is an important difference, however, in our treatment of the problem, which carries over to this limiting case. For all r, the parameters are unidentifiable if $\epsilon > 0$. Huber (1977) and Maronna (1976) circumvent this difficulty by defining the parameters in terms of the limiting values of the estimators themselves. In contrast, we estimate $\boldsymbol{\mu}$ and an appropriately standardized scatter matrix, e.g. $\boldsymbol{\Sigma}/\mathrm{tr}(\boldsymbol{\Sigma})$, $\boldsymbol{\Sigma}/\boldsymbol{\Sigma}_{11}$, which *are* identifiable provided that ϵ is less than an explicitly given bound $\epsilon(r)$, increasing to $\frac{1}{2}$ as $r \to \infty$.

Section 3 then presents the construction of consistent estimators of the identifiable parameters. In Section 4, M-estimators are constructed, defined as the Newton-Raphson solutions to an appropriate system of equations. The corresponding influence functions are chosen to redescend, so as to trim to zero the influence of observations from the arbitrary tail area of the distribution. The estimators of Section 3 are used as the starting values of the iteration. We address some of the computational difficulties associated with the construction of the initial estimator, by giving conditions under which, if $r = \infty$, the solutions to the equations are asymptotically unique. As well, we give a much simpler scoring algorithm which has, asymptotically, the same convergence properties as the Newton-Raphson process.

Our estimators are shown to be consistent and asymptotically normal, and their asymptotic covariance matrix is exhibited. In Section 5, Huber's minimax variance criterion for robustness is then applied to the class of consistent and asymptotically normal M-estimators. We solve for particular estimators which are optimal, subject to the side condition that the influence functions have compact support.

2. THE MODEL AND IDENTIFIABLE PARAMETERS

The model is that an unobservable random vector $\mathbf{Y} \in \mathbb{R}^m$, $m > 1$, has a partially known density which is a convex combination $(1 - \epsilon)w_s(\mathbf{y}) + \epsilon v_s(\mathbf{y})$ of densities w_s, v_s within the sphere $\mathbf{y}'\mathbf{y} \leq r^2$, and is arbitrary off of this sphere. The multivariate densities w_s, v_s are spherically symmetric: $w_s(\mathbf{y}) = w(|y|)$, $v_s(\mathbf{y}) = v(|\mathbf{y}|)$ for functions $w, v : [0, r] \to [0, \infty)$. Put $u_s = (1 - \epsilon)w_s + \epsilon v_s$, $u = (1 - \epsilon)w + \epsilon v$, so that within the sphere, $u_s(\mathbf{y}) = u(|\mathbf{y}|)$ is the density of $\mathbf{Y}$. We assume that $\epsilon \in [0, 1)$, $r \in [0, \infty]$, and w_s are known and fixed, whereas v_s is unknown and free to vary over the class of all spherically symmetric densities which place mass ≤ 1 within the sphere of radius r.

One observes n independent realizations of an affine transformation $\mathbf{X} = \boldsymbol{\Sigma}^{\frac{1}{2}}\mathbf{Y} + \boldsymbol{\mu}$ of $\mathbf{Y}$. The problem is to estimate $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma} = \boldsymbol{\Sigma}^{\frac{1}{2}}\boldsymbol{\Sigma}^{\frac{1}{2}\prime}$. Within the ellipsoid $E_r(\boldsymbol{\mu}, \boldsymbol{\Sigma}) = \{\mathbf{x} \mid (\mathbf{x} - \boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\mathbf{x} - \boldsymbol{\mu}) \leq r^2\}$, $\mathbf{X}$ then has an elliptically symmetric density $u_e(\mathbf{x}; \boldsymbol{\mu}, \boldsymbol{\Sigma}) = (1 - \epsilon)w_e(\mathbf{x}; \boldsymbol{\mu}, \boldsymbol{\Sigma}) + \epsilon v_e(\mathbf{x}; \boldsymbol{\mu}, \boldsymbol{\Sigma})$, where $w_e(\mathbf{x}; \boldsymbol{\mu}, \boldsymbol{\Sigma}) = |\boldsymbol{\Sigma}|^{-\frac{1}{2}} w_s(\boldsymbol{\Sigma}^{-\frac{1}{2}}(\mathbf{x} - \boldsymbol{\mu}))$, $v_e(\mathbf{x}; \boldsymbol{\mu}, \boldsymbol{\Sigma}) = |\boldsymbol{\Sigma}|^{-\frac{1}{2}} v_s(\boldsymbol{\Sigma}^{-\frac{1}{2}}(\mathbf{x} - \boldsymbol{\mu}))$.

The parameter space is $\Theta = \{\mathbf{t}, \mathbf{V} \mid \mathbf{t} \in \mathbb{R}^m, \mathbf{V}$ a positive definite $m \times m$ matrix$\}$, with typical member $\theta = (\mathbf{t}, \mathbf{V})$; and $\theta_0 = (\boldsymbol{\mu}, \boldsymbol{\Sigma})$. Denote by $\mathcal{U}_{\epsilon, r}$ the set of all densities $u_\theta(\mathbf{x})$

whose restrictions to $E_r(\mathbf{t}, \mathbf{V})$ are of the form $u_e(\mathbf{x}; \mathbf{t}, \mathbf{V})$ for some $\theta \in \Theta$.

In several places, our analysis depends only on the distribution of the norm of $\mathbf{Y} = \mathbf{\Sigma}^{-\frac{1}{2}}(\mathbf{X} - \boldsymbol{\mu}) \sim u_{\theta_0}(\mathbf{y})$. Put $\nu(y) = my^{m-1}\pi^{m/2}/\Gamma(m/2 + 1)$. Then, within $[0, r]$, the density of $|\mathbf{Y}|$ is given by $f(y) = u(y)\nu(y)$; if $\epsilon = 0$ it is $h(y) = w(y)\nu(y)$. We use F and H for the corresponding distribution functions.

Throughout this paper, we assume:

(C1) Each $u_\theta(\mathbf{x}) \in \mathcal{U}_{\epsilon, r}$ is continuously differentiable, and the corresponding $u(|\mathbf{x}|)$ is nonconstant on $[a, r]$, for each $a < r$.

(C2) The function $w(|\mathbf{x}|)$ is nonincreasing on $[0, r]$.

In Section 5, additional assumptions will be made about w.

In the model described above, the scale parameter $\mathbf{\Sigma}$ may fail to be identifiable in $U_{\epsilon, r}$, even if $r = \infty$; and for $r < \infty$ the location parameter $\boldsymbol{\mu}$ may also be unidentifiable. The following examples illustrate these remarks.

EXAMPLE 2.1. Suppose first that $m = 1$. Define

$$u_e(y) = \max\{(1 - \epsilon)w_e(y; 0, 1), (1 - \epsilon)w_e(y; 0, \sigma^2)\}I\{y^2 \le r^2\},$$

where $I(A)$ is the indicator of the event A, and σ^2 is less than 1, but sufficiently close to 1 that the mass of u_e within $[-r, r]$ does not exceed unity. Assume further that r is sufficiently large that $w_e(y; 0, 1) > w_e(y; 0, \sigma^2)$ on (k, r) for some $k < r\sigma$. We may take $r = \infty$. Now $u_e(y)$ may be extended arbitrarily for $y^2 > r^2$; the resulting $u_\theta(y)$ is a member of $U_{\epsilon, r}$ under two distinct parametrizations:

(1) As $(1 - \epsilon)w_e(y; 0, 1)I\{y^2 \le r^2\}$, contaminated symetrically in $[-k, k]$ and possibly asymmetrically in $(-\infty, -r) \cup (r, \infty)$.

(2) As $(1 - \epsilon)w_e(y; 0, \sigma^2)I\{y^2 \le r^2\sigma^2\}$, contaminated symmetrically in $(-r\sigma, -k) \cup (k, r\sigma)$ and possibly asymmetrically in $(-\infty, -r\sigma) \cup (r\sigma, \infty)$.

Now rotate $u_e(y)$ around the coordinate axes, and make a transformation to $\mathbf{x} = \mathbf{\Sigma}^{\frac{1}{2}}\mathbf{y} + \boldsymbol{\mu}$, thus obtaining an m-dimensional density $u_{\theta_1}(\mathbf{x}) \equiv u_{\theta_2}(\mathbf{x}) \in \mathcal{U}_{\epsilon, r}$, with $\theta_1 = (\boldsymbol{\mu}, \mathbf{\Sigma})$, $\theta_2 = (\boldsymbol{\mu}, \sigma^2\mathbf{\Sigma})$. Note that $\mathbf{\Sigma}$ *is* identifiable up to a scalar factor.

EXAMPLE 2.2.

(i) Suppose now that $r < \infty$, and that ϵ is so large, or r so small, that there may exist two disjoint ellipsoids $E_j = E_r(\boldsymbol{\mu}_j, \mathbf{\Sigma}_j)$, supporting $(1 - \epsilon)w_e(\mathbf{x}; \boldsymbol{\mu}_j, \mathbf{\Sigma}_j)$, such that the total mass thus supported does not exceed unity. Then neither $\boldsymbol{\mu}$ nor $\mathbf{\Sigma}$ is identifiable—we have $u_e(\mathbf{x}) = (1 - \epsilon)w_e(\mathbf{x}; \boldsymbol{\mu}_1, \mathbf{\Sigma}_1)I(\mathbf{x} \in E_1) + (1 - \epsilon)w_e(\mathbf{x}; \boldsymbol{\mu}_2, \mathbf{\Sigma}_2)I(\mathbf{x} \in E_2) \in U_{\epsilon, r}$ and are unable to determine whether $\theta = (\boldsymbol{\mu}_1, \mathbf{\Sigma}_1)$ or $\theta = (\boldsymbol{\mu}_2, \mathbf{\Sigma}_2)$. The total mass on $E_1 \cup E_2$ is $2(1 - \epsilon)H(r)$, and so identifiability requires $\epsilon < \epsilon_5^*$, where

$$(1 - \epsilon_5^*)^{-1} = 2H(r). \tag{2.1}$$

(ii) Suppose instead that $E_1 \cap E_2 \neq \emptyset$. Elliptically revolve the region of intersection around $\boldsymbol{\mu}_1$, and around $\boldsymbol{\mu}_2$, thus generating two annuli $A_j = \{\mathbf{x} \mid r_j^2 \le (\mathbf{x} - \boldsymbol{\mu}_j)\mathbf{\Sigma}_j^{-1}(\mathbf{x} - \boldsymbol{\mu}_j) \le r^2\}$, $j = 1, 2$. (If, say, $\boldsymbol{\mu}_2 \in E_1$, then $r_2 = 0$ and $A_2 = E_2$.) Define

$$\begin{aligned} u_\theta(\mathbf{x}) = \max\{&(1 - \epsilon)w(\mathbf{x}; \boldsymbol{\mu}_1, \mathbf{\Sigma}_1), (1 - \epsilon)w(\mathbf{x}; \boldsymbol{\mu}_2, \mathbf{\Sigma}_2), \\ &(1 - \epsilon)|\mathbf{\Sigma}_1|^{-\frac{1}{2}}w(r_1), (1 - \epsilon)|\mathbf{\Sigma}_2|^{-\frac{1}{2}}w(r_2)\} \qquad \text{on} \quad E_1 \cup E_2, \end{aligned}$$

arbitrary elsewhere. Then the restriction of $u_\theta(\mathbf{x})$ to each of E_1, E_2 is elliptically symmetric, and we lose identifiability of both components of θ if such a $u_\theta(\mathbf{x})$ is a proper

TABLE 1: Lower bounds on r, and on the proportion of uncontaminated mass, which are sufficient for identifiability of the parameters.

ϵ_1^*	$m = 2$	3	4	5	10	20	30	$m \to \infty$
0	1.306	1.614	1.892	2.137	3.089	4.420	5.434	$m^{\frac{1}{2}}$
	(0.574)	(0.543)	(0.534)	(0.529)	(0.518)	(0.513)	(0.510)	(0.500)
0.01	1.314	1.622	1.900	2.145	3.098	4.429	5.443	$(m + 0.018\sqrt{m})^{\frac{1}{2}}$
	(0.572)	(0.542)	(0.533)	(0.528)	(0.518)	(0.513)	(0.510)	(0.500)
0.05	1.349	1.659	1.937	2.183	3.136	4.466	5.481	$(m + 0.093\sqrt{m})^{\frac{1}{2}}$
	(0.568)	(0.540)	(0.531)	(0.527)	(0.518)	(0.512)	(0.510)	(0.500)
0.10	1.397	1.710	1.989	2.234	3.188	4.519	5.533	$(m + 0.198\sqrt{m})^{\frac{1}{2}}$
	(0.561)	(0.537)	(0.529)	(0.525)	(0.516)	(0.511)	(0.509)	(0.500)

density. Note that u_0 is constant on the annuli. This requires the contaminating density v_e to have substantial mass, which can be determined in terms of ϵ and r. If $\boldsymbol{\mu}_1 \in E_2$ or $\boldsymbol{\mu}_2 \in E_1$, this mass exceeds unity for $\epsilon < \epsilon_4^*$; otherwise a sufficient bound is $\epsilon < \epsilon_3^*$, where

$$(1 - \epsilon_4^*)^{-1} = 1 - H(r) + \frac{r\nu(r)}{m} w(0), \tag{2.2}$$

$$(1 - \epsilon_3^*)^{-1} = H(r) + \inf_{r_1 \in [0,r]} \left(H(r_1) + \frac{w(r_1)\{r\nu(r) - r_1\nu(r_1)\}}{2m} \right) \tag{2.3}$$

See Wiens (1982) for details.

Motivated by Example 2.1, we propose estimating $(\boldsymbol{\mu}, \boldsymbol{\Sigma}/\tau(\boldsymbol{\Sigma}))$, where $\tau(\boldsymbol{\Sigma})$ is any continuous, linear real-valued function which commutes with the expectation operator and whose restriction to the class of positive definite matrices is positive-valued. Examples are $\operatorname{tr}\boldsymbol{\Sigma}$ and $\boldsymbol{\Sigma}_{11}$. We note that this approach, with $\tau(\cdot) = \operatorname{tr}(\cdot)$, was also adopted by Bickel (1982) in his treatment of adaptive estimation of the parameters of the multivariate elliptical model, with $r = \infty$. If $r = \infty$, this in fact settles the identifiability problem. If ϵ is sufficiently small that the possibilities of Example 2.2 are ruled out, then $(\boldsymbol{\mu}, \boldsymbol{\Sigma}/\tau(\boldsymbol{\Sigma}))$ is identifiable for $r < \infty$ as well, in the sense that

$$u(\mathbf{x}; \boldsymbol{\mu}_1, \boldsymbol{\Sigma}_1) \equiv u(\mathbf{x}; \boldsymbol{\mu}_2, \boldsymbol{\Sigma}_2) \in \mathcal{U}_{\epsilon,r} \Rightarrow \left(\boldsymbol{\mu}_1 \frac{\boldsymbol{\Sigma}_1}{\tau(\boldsymbol{\Sigma}_1)}\right) = \left(\boldsymbol{\mu}_2 \frac{\boldsymbol{\Sigma}_2}{\tau(\boldsymbol{\Sigma}_2)}\right). \tag{2.4}$$

More precisely, we have

THEOREM 2.1. *Let* $\epsilon_1^* = \min(\epsilon_3^*, \epsilon_4^*)$, $\epsilon_2^* = \min(\epsilon_4^*, \epsilon_5^*)$, *where* $\epsilon_3^* - \epsilon_5^*$ *are defined by* (2.1)–(2.3). *In order that* (2.4) *hold, for* $r < \infty$:

(i) *It is necessary, but not sufficient, that* ϵ *be less than* ϵ_2^*.
(ii) *It is sufficient that* ϵ *be less than* ϵ_1^*.
(iii) $\epsilon_1^* \le \epsilon_2^* \le \frac{1}{2}$, *and* $\epsilon_1^* \to \frac{1}{2}$ *as* $r \to \infty$.

Proof: See Wiens and Zheng (1985).

Note that, as a referee has pointed out, if *each* $u(|\mathbf{y}|)$ may be assumed to be a strictly decreasing function of $|\mathbf{y}|$ in $(0, r)$, then Example 2.2(ii) is ruled out and the larger bound ϵ_5^* is sufficient for identifiability.

In the important special case that $w_s(y)$ is the m-variate normal density $(2\pi)^{-m/2} \exp(-\frac{1}{2}\mathbf{y}'\mathbf{y})$, so that $H(y^2)$ is the χ_m^2 distribution function, Table 1 gives the minimum permissible values of r for specified proportions ϵ_1^* of symmetric contamination. The figures in parentheses are the minimum permissible amounts of uncontaminated mass, i.e., $\int_{E_r} (1 - \epsilon_1^*)\phi(\mathbf{x}; \boldsymbol{\mu}, \boldsymbol{\Sigma})\, d\mathbf{x}$.

3. CONSISTENT INITIAL ESTIMATORS

In the model given in Section 2, it was seen that the parameters $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}/\tau(\boldsymbol{\Sigma})$ are identifiable when the known value of ϵ is sufficiently small ($\epsilon < \epsilon_1^*$). Section 4 will consider M-estimators of the identifiable parameters based on "influence functions" which vanish outside a compact set. As in the one-dimensional location problem (Collins 1976), solutions to such equations are not unique. To obtain consistent M-estimators of the parameters, one must take solutions close to initial estimators of the parameters which are themselves consistent. In this section consistent initial estimators of $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}/\tau(\boldsymbol{\Sigma})$ are constructed, to be used as starting points for the iteratively defined M-estimators of Section 4.

First note that by the definition of ϵ_1^*, there is a number $\eta > 0$ such that

$$\int_{|\mathbf{y}| \le r} u_s(\mathbf{y})\,d\mathbf{y} > \tfrac{1}{2} + \eta \tag{3.1}$$

for all $u \in \mathcal{U}_{\epsilon,r}$. Hence for each $u \in \mathcal{U}_{\epsilon,r}$, there is a constant $\lambda(u) < 1$ such that

$$\int_{|\mathbf{y}| \le \lambda(u) r} u_s(\mathbf{y})\,d\mathbf{y} = \tfrac{1}{2} + \eta. \tag{3.2}$$

Let $D = \{\mathbf{x} : \mathbf{x}'\mathbf{x} \le r^2\}$, and let $\{D_1^{(n)}, D_2^{(n)}, \ldots, D_{k_n}^{(n)}\}$ be any sequence of partitions of the set D which satisfies the following three conditions:

(i) $k_n/\sqrt{n} \to 1$ as $n \to \infty$;
(ii) $d(D_i^{(n)}) \le \sqrt{m}\,2\{L(D)/\sqrt{n}\}^{1/m}$, where $d(D_i^{(n)})$ denotes the diameter of $D_i^{(n)}$ and $L(D)$ denotes the Lebesque measure of D; and
(iii) if the set $D_i^{(n)}$ does not contain any point which belongs to the boundary of D, then it has the following form:

$$D_i^{(n)} = \left\{(x_1, \ldots, x_m) : \ell_j \left(\frac{L(D)}{\sqrt{n}}\right)^{1/m} < x_j < \ell_{j+1}\left(\frac{L(D)}{\sqrt{n}}\right)^{1/m}, j = 1, \ldots, m\right\},$$

where the ℓ_j's are fixed scalars.

The main idea of the construction is as follows. For each $\theta \in \Theta$, we can empirically estimate the restriction to D of the density of the transformed observations $g_\theta(\mathbf{x}_i) = V^{-1/2}(\mathbf{x}_i - \mathbf{t})$. If $\theta = \theta_0$, this density is spherically symmetric about $\mathbf{0}$. We select the value of θ which, using a criterion of average squared deviation, is closest to yielding a spherically symmetric density estimate on D. Specifically, we begin by defining

$$u_n(\mathbf{x}, \theta, \mathbf{X}_1, \ldots, \mathbf{X}_n) = \frac{1}{nL(D_i^{(n)})} \sum_{j=1}^{n} I\{g_\theta(\mathbf{X}_j) \in D_i^{(n)}\} \quad \text{for} \quad x \in D_i^{(n)}, \quad i = 1, 2, \ldots, k_n. \tag{3.3}$$

The estimator $\tilde{\theta}^{(n)}$ is defined by

$$\begin{aligned}
&\sup_{Q \in \mathcal{Q}} \int_D [u_n(\mathbf{x}, \tilde{\theta}^{(n)}, \mathbf{X}_1, \ldots, \mathbf{X}_n) - u_n(Q\mathbf{x}, \tilde{\theta}^{(n)}, \mathbf{X}_1, \ldots, \mathbf{X}_n)]^2\,d\mathbf{x} \\
&\quad + \frac{1}{n}\left| \sum_{j=1}^{n} I\{|g_{\tilde{\theta}^{(n)}}(\mathbf{X}_j)| \le r\} - (\tfrac{1}{2} + \eta)\right| \\
&\qquad = \inf_{\theta \in \Theta} \left\{ \sup_{Q \in \mathcal{Q}} \int_D [u_n(\mathbf{x}, \theta, \mathbf{X}_1, \ldots, \mathbf{X}_n) - u_n(Q\mathbf{x}, \theta, \mathbf{X}_1, \ldots, \mathbf{X}_n]^2\,d\mathbf{x} \right. \\
&\qquad\qquad \left. + \frac{1}{n}\left| \sum_{j=1}^{n} I\{|g_\theta(\mathbf{X}_j)| \le r\} - (\tfrac{1}{2} + \eta)\right| \right\},
\end{aligned} \tag{3.4}$$

where $\mathfrak{Q}$ denotes the set of all orthogonal matrices.

THEOREM 3.1. *Let* $\mathbf{X}_1, \ldots, \mathbf{X}_n$ *be a random sample from a distribution with density* $u_{\theta_0}(x) \in \mathcal{U}_{\epsilon,r}$, *with* $\epsilon < \epsilon_1^*$. *Then the estimator* $\tilde{\theta}^{(n)}$ *given by* (3.4) *is well defined and converges in probability to* $(\boldsymbol{\mu}, \lambda(u_{\theta_0})\boldsymbol{\Sigma})$, *where* $\lambda(u_{\theta_0}) < 1$ *is the scalar uniquely determined by Equation* (3.2).

The proof of this theorem is long and technical, and so the reader is referred to Zheng (1980) for the details. We remark that the second term on each side of (3.4) is included to preclude the possibility of the first coordinate of $\tilde{\theta}^{(n)}$ converging to some point *outside* of $E_r(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, which may have a neighbourhood (of necessarily small total mass) in which the density is elliptically symmetric.

An immediate corollary of Theorem 3.1 is that a consistent estimator of the identifiable parameter $(\boldsymbol{\mu}, \boldsymbol{\Sigma}/\tau(\boldsymbol{\Sigma}))$ is obtained as follows. Write $\tilde{\theta}^{(n)} = (\hat{\boldsymbol{\mu}}, \hat{\boldsymbol{\Sigma}})$, and then define

$$\hat{\theta}_1 = \left(\hat{\boldsymbol{\mu}}\frac{\hat{\boldsymbol{\Sigma}}}{\tau(\hat{\boldsymbol{\Sigma}})}\right). \tag{3.5}$$

In practice, very large sample sizes are needed to obtain good initial estimates $\hat{\theta}_1$ because of the use of empirical density estimation. For an alternative construction of a consistent initial estimator of $(\boldsymbol{\mu}, \boldsymbol{\Sigma}/\tau(\boldsymbol{\Sigma}))$ which avoids density estimation, but which still requires quite large sample sizes for practical success, see Section 2 of Wiens (1982).

4. THE *M*-ESTIMATORS

In this section we derive *M*-estimators of $\boldsymbol{\mu}$ and that multiple of $\boldsymbol{\Sigma}$ which has $\tau(\boldsymbol{\Sigma}) = 1$. Since $\boldsymbol{\Sigma}$ is not identifiable, the region within which the observations are symmetrically distributed can also not be known, if $r < \infty$. We thus assume that $\tau(\boldsymbol{\Sigma})$ is known, and may then assume, by applying the methods of this section to $\mathbf{x}_i/\sqrt{\tau(\boldsymbol{\Sigma})}$, that $\tau(\boldsymbol{\Sigma}) = 1$. If $r = \infty$, then $\tau(\boldsymbol{\Sigma})$ need not be known, and we estimate $\boldsymbol{\Sigma}/\tau(\boldsymbol{\Sigma})$.

We will derive a system of equations, and define an estimator $\hat{\theta}$ as the Newton-Raphson solution to these equations, with $\hat{\theta}_1$ as starting value. We show that $\hat{\theta}$ is a consistent estimator of θ_0, and that $\sqrt{n}(\hat{\theta} - \theta_0)$ is asymptotically normally distributed. In the case $r = \infty$, it will be shown that under fairly mild restrictions the solution to the equations is asymptotically unique. Comparisons are made with the estimators of Maronna (1976) and Huber (1977).

4.1. *Derivation of the Equations.*

If $\mathbf{X}$ has density $u_{\theta_0}(\mathbf{x})$, with $\tau(\boldsymbol{\Sigma}) = 1$, then

(i) for all functions $a_0(\cdot)$ vanishing off of $[0, r^2]$,

$$\mathscr{E}\{a_0((\mathbf{x} - \boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\mathbf{x} - \boldsymbol{\mu}))(\mathbf{x} - \boldsymbol{\mu})\} = 0;$$

(ii) for all such $a(\cdot)$,

$$\mathscr{E}\{a_1((\mathbf{x} - \boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\mathbf{x} - \boldsymbol{\mu}))\boldsymbol{\Sigma}^{-1/2}(\mathbf{x} - \boldsymbol{\mu})(\mathbf{x} - \boldsymbol{\mu})'\boldsymbol{\Sigma}^{-\frac{1}{2}\prime}\} = k\mathbf{I}_m$$

for some $k = k(a_1, v)$;

(iii) $\tau(\boldsymbol{\Sigma}) = 1$.

Pre- and postmultiplying by $\boldsymbol{\Sigma}^{\frac{1}{2}}$ and $\boldsymbol{\Sigma}^{\frac{1}{2}\prime}$ in (ii), then applying τ to both sides (recalling that τ commutes with the expectation operator), and using (iii) gives $k = \mathscr{E}\{\tau(a_1(\cdot)(\mathbf{x} - \boldsymbol{\mu})(\mathbf{x} - \boldsymbol{\mu})')\}$. It follows that (ii) and (iii) are together equivalent to

(iv) $\mathscr{E}[a_1((\mathbf{x}-\boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu}))\{(\mathbf{x}-\boldsymbol{\mu})(\mathbf{x}-\boldsymbol{\mu})' - \tau((\mathbf{x}-\boldsymbol{\mu})(\mathbf{x}-\boldsymbol{\mu})')\boldsymbol{\Sigma}\}] = \mathbf{0}_{m\times m}$.

Let $a_0(z)$, $a_1(z)$ be continuous, piecewise continuously differentiable functions vanishing off of $[0, r^2]$. For $\theta = (\mathbf{t}, \mathbf{V}) \in \Theta$ and sample values $\mathbf{x}_1, \ldots, \mathbf{x}_n$ put $\mathbf{y}_i = V^{-\frac{1}{2}}(\mathbf{x}_i - \mathbf{t})$. The estimator $\hat{\theta}$ is defined to be the Newton-Raphson solution θ^* to

$$n^{-1}\sum_i a_0(\mathbf{y}_i'\mathbf{y}_i)(\mathbf{x}_i - \mathbf{t}) = \mathbf{0}_{m\times 1}, \tag{4.1}$$

$$n^{-1}\sum_i a_1(\mathbf{y}_i'\mathbf{y}_i)\{(\mathbf{x}_i - \mathbf{t})(\mathbf{x}_i - \mathbf{t})' - \tau((\mathbf{x}_i - \mathbf{t})(\mathbf{x}_i - \mathbf{t})')\mathbf{V}\} = \mathbf{0}_{m\times m}, \tag{4.2}$$

starting with $\hat{\theta}_1$ if the iteration process converges; $\hat{\theta}_1$ otherwise.

Recall that vec $\mathbf{A}$ is the column vector formed by stacking the columns of a matrix $\mathbf{A}$, in their natural order. For $\mathbf{A}$ symmetric, denote by $\underset{s}{\mathrm{vec}}\,\mathbf{A}$ the subvector of vec $\mathbf{A}$ containing only those elements of $\mathbf{A}$ on or below the main diagonal. For any $\theta \in \Theta$, put

$$\boldsymbol{\theta} = (\mathbf{t}', (\underset{s}{\mathrm{vec}}\, V)')', \qquad \boldsymbol{\psi}_0(\mathbf{x}_i; \boldsymbol{\theta}) = a_0(\mathbf{y}_i'\mathbf{y}_i)(\mathbf{x}_i - \mathbf{t}),$$
$$\boldsymbol{\psi}_1(\mathbf{x}_i; \boldsymbol{\theta}) = a_1(\mathbf{y}_i'\mathbf{y}_i)\{\underset{s}{\mathrm{vec}}\,(\mathbf{x}_i - \mathbf{t})(\mathbf{x}_i - \mathbf{t})' - \tau((\mathbf{x}_i - \mathbf{t})(\mathbf{x}_i - \mathbf{t})')\,\underset{s}{\mathrm{vec}}\,\mathbf{V}\},$$
$$\boldsymbol{\psi} = (\boldsymbol{\psi}_0', \boldsymbol{\psi}_1')', \qquad F_n(\boldsymbol{\theta}) = n^{-1}\sum_{i=1}^{n}\boldsymbol{\psi}(\mathbf{x}_i; \boldsymbol{\theta}).$$

Then (4.1), (4.2) can be written more succintly as $F_n(\hat{\boldsymbol{\theta}}) = \mathbf{0}$.

4.2. Consistency and Asymptotic Normality.

The consistency of $\hat{\boldsymbol{\theta}}_1$ easily implies consistency of $\hat{\boldsymbol{\theta}}$, and that $\lim_{n\to\infty} P(\hat{\boldsymbol{\theta}} = \boldsymbol{\theta}^*) = 1$, as long as $\mathscr{E}(\partial\boldsymbol{\psi}/\partial\boldsymbol{\theta})_{\boldsymbol{\theta}_0}$ is nonsingular. From this,

$$nF_n(\hat{\boldsymbol{\theta}}) \xrightarrow{P} \mathbf{0}.$$

The mean-value and central limit theorems then imply that

$$\sqrt{n}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) \xrightarrow{D} \mathbf{N}_p(\mathbf{0}, \mathscr{E}[\Phi\Phi']),$$

where

$$\Phi = -\left\{\mathscr{E}\left(\frac{\partial\boldsymbol{\psi}}{\partial\boldsymbol{\theta}}\right)_{\boldsymbol{\theta}_0}\right\}^{-1}\boldsymbol{\psi}\Big|_{\boldsymbol{\theta}_0}$$

is the influence function. The details are given in Wiens (1982).

Theorem 4.1 below gives conditions under which $\mathscr{E}(\partial\boldsymbol{\psi}/\partial\boldsymbol{\theta})_{\boldsymbol{\theta}_0}$ is nonsingular. With $\mathbf{y} = \boldsymbol{\Sigma}^{-\frac{1}{2}}(\mathbf{x} - \boldsymbol{\mu})$, define

$$\alpha = \mathscr{E}[2y_1^2 a_0'(\mathbf{y}'\mathbf{y}) + a_0(\mathbf{y}'\mathbf{y})], \qquad \alpha_1 = \mathscr{E}[y_1^2 a_0^2(\mathbf{y}'\mathbf{y})],$$
$$\beta = \mathscr{E}[y_1^2 a_1(\mathbf{y}'\mathbf{y})], \qquad \gamma = \mathscr{E}[y_1^2 y_2^2 a_1'(\mathbf{y}'\mathbf{y}), \qquad \gamma_1 = \mathscr{E}[y_1^2 y_2^2 a_1^2(\mathbf{y}'\mathbf{y})].$$

Let $\boldsymbol{\delta}_{ij}$ be the $m \times m$ matrix with 1 in the $(i,j)^{\text{th}}$ position, zeros elsewhere, and define $\mathbf{T}: m \times m$ by $T_{ij} = \tau(\boldsymbol{\delta}_{ij} + \boldsymbol{\delta}_{ji})$.

THEOREM 4.1. *Necessary and sufficient for the nonsingularity of* $\mathscr{E}(\partial\boldsymbol{\psi}/\partial\boldsymbol{\theta})_{\boldsymbol{\theta}_0}$ *are* (i) $\alpha \neq 0$, (ii) $\beta \neq 0$, (iii) $\beta + 2\gamma \neq 0$. *If* (i)–(iii) *hold, then* $\sqrt{n}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)$ *is asymptotically normally distributed with mean* $\mathbf{0}$. *The asymptotic covariance matrix is of the form* $\mathrm{diag}\,(\mathbf{A}_{m\times m}, \mathbf{B}_{(m(m+1)/2)\times(m(m+1)/2)})$, *so that the location and scale components of the estimator are asymptotically independent. Here,* $\mathbf{A} = (\alpha_1/\alpha^2)\boldsymbol{\Sigma}$, *and* $\mathbf{B}$ *is described by*

$$\lim_{n\to\infty} Cov[\sqrt{n}\hat{\sigma}_{ik}, \sqrt{n}\hat{\sigma}_{jl}] = \frac{\gamma_1}{(\beta + 2\gamma)^2}\Big\{\sigma_{ij}\sigma_{kl} + \sigma_{jk}\sigma_{il} + \tau(\mathbf{\Sigma T \Sigma})\sigma_{ik}\sigma_{jl} - \tau(\boldsymbol{\sigma}_i\boldsymbol{\sigma}_k' + \boldsymbol{\sigma}_k\boldsymbol{\sigma}_i')\sigma_{jl} - \tau(\boldsymbol{\sigma}_j\boldsymbol{\sigma}_l' + \boldsymbol{\sigma}_l\boldsymbol{\sigma}_j')\sigma_{ik}\Big\},$$

where $\mathbf{\Sigma} = (\sigma_{ij}) = \|\boldsymbol{\sigma}_1, \ldots, \boldsymbol{\sigma}_m\|$.

The proof of Theorem 4.1 is given in the Appendix. See Tyler (1982) for a similar analysis of the covariance structures of the estimators of Maronna (1976) and Huber (1977).

REMARKS.

(1) The proof of Theorem 4.1 also shows that the influence function is

$$\Phi = \operatorname{diag}(\alpha^{-1}\mathbf{I}_m, (\beta + 2\gamma)^{-1}\mathbf{I}_{m(m+1)/2})\boldsymbol{\psi}|_{\boldsymbol{\theta}_0}. \tag{4.3}$$

The method of scoring, whereby the $(k + 1)$th iterate is given by

$$\boldsymbol{\theta}_{k+1} = \boldsymbol{\theta}_k - \left\{\mathscr{E}\left(\frac{\partial\boldsymbol{\psi}}{\partial\boldsymbol{\theta}}\right)_{\boldsymbol{\theta}_0}\right\}^{-1}\Bigg|_{\boldsymbol{\theta}_0=\boldsymbol{\theta}_k} F_n(\boldsymbol{\theta}_k),$$

is less cumbersome than the Newton-Raphson method and has, asymptotically, the same superlinear convergence properties (Ortega and Rheinboldt 1970). Using (4.3), this iteration method becomes

$$\boldsymbol{\theta}_{k+1} = \boldsymbol{\theta}_k + \operatorname{diag}(\alpha_k^{-1}I, (\beta_k + 2\gamma_k)^{-1}I)F_n(\boldsymbol{\theta}_k),$$

where α_k, β_k, γ_k are the method-of-moments estimates of α, β, γ based on $\boldsymbol{\theta}_k$.

(2) From McCulloch (1982), or from Magnus and Neudecker (1979), one sees that the matrix $\mathbf{C}$, defined at (6.5) in the proof of Theorem 4.1, is the covariance matrix of $n^{-\frac{1}{2}}\underset{s}{\operatorname{vec}}\,\mathbf{W}$, where $\mathbf{W}$ is an $m \times m$ Wishart matrix $\sim W_m(\mathbf{\Sigma}, n)$. From this, it follows that if the scale components of $\hat{\boldsymbol{\theta}}$ are arranged as an $m \times m$ matrix $\hat{\mathbf{\Sigma}}$, and if $\mathbf{U} \sim W_m(\mathbf{I}, n)$, then

$$\sqrt{n}(\mathbf{\Sigma}^{-\frac{1}{2}}\hat{\mathbf{\Sigma}}\mathbf{\Sigma}^{-\frac{1}{2}\prime} - \mathbf{I}_m) \quad \text{and} \quad \frac{\gamma_1^{\frac{1}{2}}}{\sqrt{n}(\beta + 2\gamma)}\{\mathbf{U} - \tau(\Sigma^{\frac{1}{2}}\mathbf{U}\Sigma^{\frac{1}{2}\prime})\mathbf{I}_m\}$$

have the same asymptotic distribution. This suggests the possible use of the distribution of the latter to approximate the former, in small samples. For instance, if $\mathbf{\Sigma}^{\frac{1}{2}}$ is lower triangular, and $\tau(\mathbf{\Sigma}) = \Sigma_{11}$, then the latter matrix is $\{\gamma_1^{\frac{1}{2}}/\sqrt{n}(\beta + 2\gamma)\}(\mathbf{U} - u_{11}\mathbf{I}_m)$.

(3) In Section 5, we consider the problem of choosing functions a_0, a_1 so as to minimize the maximum (with respect to the natural ordering by positive-definiteness) of the asymptotic covariance matrix of $\sqrt{n}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)$ as u varies over $U_{\epsilon,r}$. From Theorem 4.1, it is clear that this reduces to independently minimaxing the scalar funtions α_1/α^2 and $\gamma_1/(\beta + 2\gamma)^2$. We can put this reduced problem into a more unified form.

Recall that on $[0, r]$, $|\mathbf{\Sigma}^{-\frac{1}{2}}(\mathbf{X} - \boldsymbol{\mu})|$ has density $f = uv$. With respect to this density we have

$$\frac{\alpha_1}{\alpha^2} = V(\psi_0, u) = \frac{m\mathscr{E}\{\psi_0^2(X)\}}{\left\{\mathscr{E}\left((m-1)\frac{\psi_0(X)}{X} + \psi_0'(X)\right)\right\}^2},$$

$$\frac{\gamma_1}{(\beta + 2\gamma)^2} = V(\psi_1, u) = \frac{m(m+2)\mathscr{E}\{X^2\psi_1^2(X)\}}{[\mathscr{E}\{(m+1)X\psi_1(X) + X^2\psi_1'(X)\}]^2}, \tag{4.4}$$

where

$$\psi_0(x) = xa_0(x^2), \qquad \psi_1(x) = xa_1(x^2). \tag{4.5}$$

If we may integrate by parts, the variance functionals (4.4) become

$$V(\psi_j, u) = \frac{\int_0^r \psi_j^2(x)u(x)\eta_j(x)\nu(x)\,dx}{2\left[\int_0^r \psi_j(x)u'(x)\eta_j(x)\nu(x)\,dx\right]^2}, \tag{4.6}$$

where $\eta_j(x) = m^{-1}\{2x^2/(m+2)\}^j$, $j = 0, 1$. This is the form that will be used in Section 5 below.

4.3. *The Case* $r = \infty$.

If $r = \infty$, then $\tau(\boldsymbol{\Sigma})$ need not be known and $\hat{\boldsymbol{\theta}}$ is consistent for $\boldsymbol{\Sigma}/\tau(\boldsymbol{\Sigma})$, for all distributions in $U_{\epsilon,\infty}$. In this case, the problem of completely spurious solutions to $F_n(\boldsymbol{\theta}) = \mathbf{0}$ is no longer present. A partial result on asymptotic uniqueness of solutions is given below. Its proof entails only simple modifications to those of Theorems 1 and 3 of Maronna (1976), and so is omitted. The details are given in Wiens (1982).

THEOREM 4.2. *If* $r = \infty$, *then under the following conditions the zero* θ_0 *of* $\mathscr{E}[\psi(\mathbf{X};\boldsymbol{\theta}]$ *is unique:*

(a) *The function* $\psi_1(x)/x$ *is nonincreasing and* $x\psi_1(x)$ *is nondecreasing; and either*
(b1) *the function* $\psi_0(x)$ *is nondecreasing, or*
(b2) *each* $u(|x|)$ *is a strictly decreasing function of* $|x|$.

The approaches taken by Maronna (1976) and Huber (1977) in estimating multivariate location and scale are somewhat different than ours. The major differences are that their estimators are invariant under arbitrary affine transformations, but that in neither case does it appear to be the intent to construct estimators of $\boldsymbol{\mu}$ and a specified multiple of $\boldsymbol{\Sigma}$ which are globally consistent throughout $U_{\epsilon,\infty}$ if $\epsilon > 0$.

In both cases, the estimator $\hat{\boldsymbol{\theta}}$ is the zero of $n^{-1}\sum_i\psi(\mathbf{x}_i;\boldsymbol{\theta})$, where $\mathbf{y}_i = \mathbf{V}^{-\frac{1}{2}}(\mathbf{x}_i - \mathbf{t})$, $\psi = (a_0(\mathbf{y}_i'\mathbf{y}_i)\mathbf{y}_i, a_1(\mathbf{y}_i'\mathbf{y}_i)\mathbf{y}_i\mathbf{y}_i' - a_2(\mathbf{y}_i'\mathbf{y}_i)\mathbf{I})$, and a_0, a_1, a_2 are sufficiently smooth functions. Maronna takes $a_2 \equiv 1$. Under conditions similar to those in Theorem 4.2, Maronna shows that unique solutions $\hat{\boldsymbol{\theta}}$ exist and are asymptotically normally distributed around the zero $\boldsymbol{\theta}^*$ of $\mathscr{E}[\boldsymbol{\psi}(\mathbf{X};\boldsymbol{\theta})]$. He conjectures that a certain set of conditions is sufficient to ensure finite-sample uniqueness. Huber states similar asymptotic results in the more general case.

It is a bit unclear what these estimators really would estimate in $U_{\epsilon,\infty}$. Suppose that $\boldsymbol{\theta}^*$ is to be of the form $(\boldsymbol{\mu}, \boldsymbol{\Sigma}/k^2)$, where k is some scalar. The relationship $\mathscr{E}[\boldsymbol{\psi}(\mathbf{X};\boldsymbol{\theta}^*)] = \mathbf{0}$ is then seen to be equivalent to

$$\mathscr{E}[|k\mathbf{y}|^2a_1(|k\mathbf{y}|^2) - ma_2(|k\mathbf{y}|^2) = 0. \tag{4.7}$$

If $a_1 \equiv 1$ and $za_1(z)$ is nondecreasing, as in Maronna's development, it seems clear that for any distribution in $U_{\epsilon,\infty}$ there exists a scalar k, depending upon the underlying distribution, satisfying (4.7). But global consistency throughout $U_{\epsilon,\infty}$, for the same multiple of $\boldsymbol{\Sigma}$, forces the choice $a_2(z) = m^{-1}za_1(z)$. Note that then any scalar multiple of $\boldsymbol{\Sigma}$ is a solution. See Tyler (1985) for a treatment of this case, with location assumed known.

In Huber's consideration of $U_{\epsilon,\infty}$ (his "F"), with w the normal density and $(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ assumed to be $(\mathbf{0}, \mathbf{I})$, he states that, with $a_2 \equiv 1$, the optimum choice of a_1 has minimax properties with respect to the variance of $\hat{\boldsymbol{\theta}}$ for "that subset of F for which $(\hat{\boldsymbol{\theta}})$ is a consistent estimator of the identity matrix." The problem of identifiability is not considered.

It can be seen [although it is not explicitly stated in Huber (1977)] that the choice $a_2 \equiv 1$ is optimum, in the minimax sense, under the side condition $\mathscr{E}[\boldsymbol{\psi}(\mathbf{X}; \boldsymbol{\theta}^*)] = \mathbf{0}$. With this choice, the minimax problems considered by Huber are equivalent to minimaxing the functionals $V(\psi_j, u)$, defined at (4.6), with $r = \infty$. We shall thus concentrate, in Section 5, on solutions to these problems for finite r.

5. THE MINIMAX SOLUTIONS

We now consider the problem of determining functions a_0, a_1 which minimize the maximum (ordered by positive definiteness) asymptotic covariance matrix of $\sqrt{n}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)$. As in Remark 3 of Section 4, this is equivalent to determining pairs (ψ_{*j}, u_{*j}), $j = 0, 1$, such that

$$\sup_{\mathscr{U}} V(\psi_{*j}, u) \leq \sup_{\mathscr{U}} V(\psi, u) \qquad \text{for all} \quad \psi \in \Psi, \tag{5.1}$$

where $\mathscr{U} = \{u(|\mathbf{x}|) \,|\, u(\mathbf{x}) \in \mathscr{U}_{\epsilon,r}\}$ and Ψ is the set of all continuous, piecewise continuously differentiable functions on $[0, \infty)$ vanishing off of $[0, r]$. We then put $a_j(x) = \psi_j(\sqrt{x})/\sqrt{x}$.

This problem motivated the work in Section 4 of Collins and Wiens (1985), to which the reader is referred for complete details. Here, we proceed in a more heuristic fashion. We temporarily drop the subscript j, in order to treat the two minimax problems simultaneously.

Similar to the treatment of similar problems in Huber (1964, 1981), one first finds $u_* \in \mathscr{U}$ minimizing the Fisher information there, and then chooses $\psi_* = -u_*'/u_*$ to attain maximum efficiency at u_*. For our purposes, a definition of the Fisher information equivalent to the classical one is

$$I(u; \eta) = 2\int_0^r \left(\frac{u'(x)}{u(x)}\right)^2 u(x)\sigma(x)\,\mathrm{d}x,$$

where $\sigma(x) = \nu(x)\eta(x)$. By Lemma 6 of Huber (1964), I is a convex functional of u. A standard variational argument then shows that I is minimized by u_* if

$$0 \leq \frac{\mathrm{d}}{\mathrm{d}t} I((1 - t)u_* + tu; \eta)_{t=0} \qquad \text{for all} \quad u \in \mathscr{U}.$$

Carrying out the differentiation, and then integrating by parts, yields the equivalent condition

$$\int_0^r J(\psi_*)(x)\nu(x)(u - u_*)(x)\,\mathrm{d}x + \lim_{x \downarrow 0} \psi_*(x)\sigma(x)(u - u_*)(x) \geq 0 \tag{5.2}$$

where $J(\psi)(x) = \{2\psi'(x) - \psi^2(x) + 2\psi(x)\sigma'(x)/\sigma(x)\}\eta(x)$.

Write $u = (1 - \epsilon)w + \epsilon v$, $u_* = (1 - \epsilon)w + \epsilon v_*$. It then follows from (5.2) that $J(\psi_*)$ must attain a constant minimum value $-\lambda < 0$ on the support of v_*, and that v_* must place mass 1 within $[0, r]$. It is then also necessary that, with $\zeta(x) = -w'(x)/w(x)$, the region $\{x \mid J(\zeta)(x) < -\lambda\}$ be contained in the support of v_*. This latter condition then determines that support.

We assume:

(A1) $I(w; \eta) < \infty$.
(A2) $w(x) > 0$ on $[0, r]$, and $\zeta(x) = -w'(x)/w(x)$ is continuously differentiable and positive on $(0, r)$, and bounded on $[0, r]$.
(A3) Either (i) $\zeta(x)$ is nondecreasing on $(0, r)$, or (ii) $J(\zeta)(x)$ is nonincreasing on that subset of $(0, r)$ on which it is negative.
(A4) $\epsilon < \epsilon_4^*$.

Let $\xi(x; \lambda)$ be the solution to $J(\xi)(x) \equiv -\lambda$ passing through $(r, 0)$, and let $k(x; \lambda)$ satisfy $-k'(x; \lambda)/k(x, \lambda) = \xi(x; \lambda)$. Define

$$
\begin{aligned}
u(x; \lambda) &= \left\{(1 - \epsilon)w(x), (1 - \epsilon)\frac{w(a_\lambda)}{k(a_\lambda; \lambda)}k(x; \lambda), (1 - \epsilon)w(x)\right\}, \\
\psi(x; \lambda) &= \{\zeta(x), \xi(x; \lambda), 0\}
\end{aligned}
\tag{5.3}
$$

on $[0, a_\lambda], [a_\lambda, r], [r, \infty)$ respectively. Assumptions (A1)−(A4) imply the existence of a unique pair $\bar{\lambda} > 0$, $a_{\bar{\lambda}} > 0$ satisfying

$$
\int_{a_{\bar{\lambda}}}^{r} v(x; \bar{\lambda}) \nu(x)\, dx = 1, \tag{5.4i}
$$

$$
\zeta(a_{\bar{\lambda}}) = \xi(a_{\bar{\lambda}}; \bar{\lambda}) \tag{5.4ii}
$$

and such that $\{x \mid J(\zeta)(x) < -\bar{\lambda}\} \subseteq [a_{\bar{\lambda}}, r]$. Put $\psi_*(x) = \psi(x; \bar{\lambda})$, $u_*(x) = u(x; \bar{\lambda})$. Then in (5.2) the limit is nonnegative, since $v_*(0) = 0$. The integral is

$$
\int_0^{a_{\bar{\lambda}}} J(\zeta)(x) \nu(x) v(x)\, dx - \bar{\lambda}\left(\int_{a_{\bar{\lambda}}}^{r} \nu(x) v(x)\, dx - 1\right)
$$

$$
\geq -\bar{\lambda}\left(\int_0^{r} \nu(x) v(x)\, dx - 1\right) \geq 0.
$$

Thus (5.2) holds, and it then follows in a manner very similar to that in Theorem 2 of Huber (1964) that (ψ_*, u_*) is a saddle-point solution to the minimax problem, in that

$$
V(\psi_*, u) \leq \frac{1}{I(u_*; \eta)} = V(\psi_*, u_*) \leq V(\psi, u_*) \tag{5.5}
$$

for all $u \in \mathcal{U}$, $\psi \in \Psi$.

Before applying this theory to the $V(\psi_j, u)$ at (4.6), we first check the conditions of Theorem 4.1. That $\beta = m^{-1}\mathscr{E}[X\psi_{*,1}(X)] > 0$ is clear, since $\psi_{*,1} > 0$ on $(0, r)$. The requirements that α and $\beta + 2\gamma$ be bounded away from zero are ensured by (5.1). Thus Theorem 4.1 applies, and the minimax solutions are given by (5.3) and (5.4). With

$$
\delta(a) = \frac{\epsilon/(1 - \epsilon) + H(r) - H(a)}{h(a)},
$$

(5.4) becomes

$$\xi(a, \lambda) = \zeta(a), \qquad \int_a^r \left(\frac{x}{a}\right)^{m-1} \frac{k(x;\lambda)}{k(a;\lambda)} \mathrm{d}x = \delta(a). \tag{5.6}$$

5.1. *Case I: Location Estimation* ($\eta_0(x) = m^{-1}$).

For notational convenience we work with the equation $J(\xi) \equiv -4\lambda^2/m$, i.e.,

$$2\xi' - \xi^2 + \frac{2(m-1)}{x}\xi \equiv -4\lambda^2. \tag{5.7}$$

Putting $\xi = -2y'/y = -k'/k$, where $k = y^2$, gives

$$y'' + \frac{m-1}{x}y' -, \lambda^2 y \equiv 0,$$

which in turn transforms into Bessel's equation—see Watson (1966). Compare also (10.7) of Huber (1981). The solution to (5.7) satisfying $\xi(r, \lambda) = 0$ is then obtained in terms of the modified Bessel functions I and K:

$$\xi(x, \lambda) = \frac{2\lambda(\omega K_{m/2}(\lambda x) - I_{m/2}(\lambda x))}{\omega K_{(m-2)/2}(\lambda x) + I_{(m-2)/2}(\lambda x)}, \tag{5.8}$$

$$k(x, \lambda) = x^{2-m}(\omega K_{(m-2)/2}(\lambda x) + I_{(m-2)/2}(\lambda x))^2, \tag{5.9}$$

$$\omega = \frac{I_{m/2}(\lambda r)}{K_{m/2}(\lambda r)}. \tag{5.10}$$

Now (5.5) holds, with $\psi_{*,0}(x)$, $u_{*,0}(x)$ given by (5.3), (5.6), (5.8)–(5.10).

In the families of m-variate densities

$$P_1 = \{w \,|\, w_s(\mathbf{y}) = w(|\mathbf{y}|) = \mathrm{const}(\exp(-|c\mathbf{y}|^{2\ell}), c, \ell > 0\},$$

$$P_2 = \{w \,|\, w_s(\mathbf{y}) = w(|\mathbf{y}|) = \mathrm{const}\left(1 + \frac{|\mathbf{y}|^2}{c}\right)^{-(m+\ell)/2}, c, \ell > 0\},$$

assumptions (A.1)–(A.3) hold for P_1 if $\ell \geq \frac{1}{2}$, and for P_2 if $\ell \leq m - 4$ or if $\ell > m - 4$ and $r^2 < c\{1 + 4m/(\ell - m + 4)\}$.

For odd m, the modified Bessel functions above have closed-form expansions. If $m = 3$, then

$$\xi(x, \lambda) = \frac{2}{x}\left[\frac{(\lambda^2 xr - 1)\tanh(\lambda r - \lambda x) + (\lambda r - \lambda x)}{\lambda r - \tanh(\lambda r - \lambda x)}\right],$$

$$k(x, \lambda) = \frac{2e^{2\lambda r}}{\pi}\left[\frac{\lambda r \cosh(\lambda r - \lambda x) - \sinh(\lambda r - \lambda x)}{\lambda^2 x(r + 1)}\right]^2.$$

In particular, if $w(\mathbf{y}) = (2\pi)^{-\frac{3}{2}} \exp(-\frac{1}{2}\mathbf{y}'\mathbf{y})$, so that $\zeta(x) = x$, then (5.6) becomes

$$\tanh(\lambda r - \lambda a) = \frac{\lambda\{(a^2 - 2)r + 2a\}}{(a^2 - 2) + 2\lambda^2 ra},$$

$$\delta(a) = \frac{\lambda^2 r^3\{1 - \tanh^2(\lambda r - \lambda a)\}}{\{\lambda r - \tanh(\lambda r - \lambda a)\}^2} + \frac{a(a^2 - 2 - 4\lambda^2)}{4\lambda^2}.$$

Incidentally, this answers a conjecture of Collins (1982), who considered the problem of multivariate estimation of location, with scale known and with $w(\mathbf{y})$ the normal density.

It was conjectured there that $(\psi_{*,0}, u_{*,0})$ would be of the above form with, in the case $m = 3$, the constants determined from six equations and inequalities.

5.2. Case II; Scatter Estimation ($\eta_1(x) = 2x^2/m(m+2)$)

Here, we work with the equation $J(\xi) \equiv -2m\lambda/(m+2)$, i.e.,

$$2x^2\xi' - x^2\xi^2 + 2(m+1)x\xi \equiv -m^2\lambda.$$

Put $z_1(x) = x\xi - m$ to get $z_1^2 - 2xz_1' \equiv m^2(1+\lambda)$. A particular solution is $z_1(x) = m\sqrt{1+\lambda}$. The general solution is obtained by setting $z_1(x) = m\sqrt{1+\lambda} + \{1/z(x)\}$ and solving the resulting equation $z' + (m\sqrt{1+\lambda}/x)\,z = -1/2x$ with the integrating factor $x^{m\sqrt{1+\lambda}}$. Unravelling these transformations and choosing the solution through $(r, 0)$ gives

$$\xi(x,\lambda) = \frac{(R+1)m}{x}\,\frac{1-\left(\dfrac{x}{r}\right)^{mR}}{1+\dfrac{R+1}{R-1}\left(\dfrac{x}{r}\right)^{mR}}, \qquad \text{where } R = \sqrt{1+\lambda}, \tag{5.11}$$

whence

$$k(x,\lambda) = \frac{\left\{1+\dfrac{R+1}{R-1}\left(\dfrac{x}{r}\right)^{mR}\right\}^2}{\left(\dfrac{x}{r}\right)^{m(R+1)}}. \tag{5.12}$$

The pair $\psi_{*,1}(x)$, $u_{*,1}(x)$ given by (5.3), (5.6), (5.11), (5.12) then possesses the saddle-point property (5.5).

6. APPENDIX: PROOF OF THEOREM 4.1

It is convenient to first recall some notions useful in matrix differentiation, as given in Henderson and Searle (1979) and Wiens (1985). There exists a unique matrix $\mathbf{G}: m^2 \times m(m+1)/2$ such that $\mathbf{G}\operatorname{vec}_s \mathbf{A} = \operatorname{vec}\mathbf{A}$ for all symmetric $\mathbf{A}: m \times m$, and the Moore-Penrose inverse $\mathbf{H} = (\mathbf{G}'\mathbf{G})^{-1}\mathbf{G}'$ has $\mathbf{H}\operatorname{vec}\mathbf{A} = \operatorname{vec}_s \mathbf{A}$. The $mn \times mn$ permutation matrix $\mathbf{I}_{(m,n)}$ is defined by its action $\mathbf{I}_{(m,n)}\operatorname{vec}A = \operatorname{vec}A'$ for all $\mathbf{A}: n \times m$, and has the property that $\mathbf{H}\mathbf{I}_{(m,m)} = \mathbf{H}$. If $\mathbf{A}$, $\mathbf{B}$ are $m \times m$ symmetric, then the matrix of partial derivatives of the functionally independent elements of $\mathbf{A}$ with respect to those of $\mathbf{B}$ is

$$\left(\frac{\partial \operatorname{vec}_s \mathbf{A}}{\partial \operatorname{vec}_s \mathbf{B}}\right) = \mathbf{H}\left(\frac{\partial \operatorname{vec}\mathbf{A}}{\partial \operatorname{vec}\mathbf{B}}\right)\mathbf{G},$$

where the ordinary Jacobian matrix appears on the right, calculated ignoring functional relationships. For arbitrary $\mathbf{A}: m \times m$ we have the identities

$$(\mathbf{A}\otimes\mathbf{A})\mathbf{G} = \mathbf{G}\mathbf{H}(\mathbf{A}\otimes\mathbf{A})\mathbf{G}, \qquad \mathbf{H}(\mathbf{A}\otimes\mathbf{A}) = \mathbf{H}(\mathbf{A}\otimes\mathbf{A})\mathbf{G}\mathbf{H} \tag{6.1}$$

(see Henderson and Searle 1979), implying

$$\{\mathbf{H}(\mathbf{A}\otimes\mathbf{A})\mathbf{G}\}^{-1} = \mathbf{H}(\mathbf{A}^{-1}\otimes\mathbf{A}^{-1})\mathbf{G}. \tag{6.2}$$

Here, $\otimes$ represents the Kronecker product. We denote by $\mathbf{A}\oplus\mathbf{B}$ the direct sum $\operatorname{diag}(\mathbf{A},\mathbf{B})$.

With $p = m + m(m+1)/2$, $q = m^2 + m$, put $\mathbf{H}_p = \mathbf{I}_m \oplus \mathbf{H} : p \times q$, $\mathbf{G}_p = \mathbf{I}_m \oplus \mathbf{G} : q \times p$, $\boldsymbol{\Omega} = \mathbf{V}^{\frac{1}{2}} \oplus (\mathbf{V}^{\frac{1}{2}} \otimes \mathbf{V}^{\frac{1}{2}}) : q \times q$. Define $\mathbf{S} : m \times m$ by $S_{ij} = \tau(\mathbf{V}^{\frac{1}{2}}\delta_{ij}\mathbf{V}^{\frac{1}{2}\prime})$, so that

$$\tau(\mathbf{V}^{\frac{1}{2}}\mathbf{y}\mathbf{y}'\mathbf{V}^{\frac{1}{2}\prime}) = \sum_{i,j} y_i y_j S_{ij} = \mathbf{y}'\mathbf{S}\mathbf{y} = (\operatorname{vec}\mathbf{S})'(\mathbf{y} \otimes \mathbf{y}).$$

Then $\boldsymbol{\psi}$ factors as $\boldsymbol{\psi} = \mathbf{H}_p \mathbf{M}\mathbf{v}$, where

$$\mathbf{M} = \boldsymbol{\Omega}[\mathbf{I}_m \oplus \{\mathbf{I}_{m^2} - (\operatorname{vec}\mathbf{I}_m)(\operatorname{vec}\mathbf{S})'\}] : q \times q,$$
$$\mathbf{v} = (a_0(\mathbf{y}'\mathbf{y})\mathbf{y}', a_1(\mathbf{y}'\mathbf{y})(\mathbf{y}' \otimes \mathbf{y}'))' : q \times 1.$$

Thus

$$\left(\frac{\partial \boldsymbol{\psi}}{\partial \boldsymbol{\theta}}\right) = \mathbf{H}_p \left(\frac{\partial \mathbf{M}\mathbf{v}}{\partial \mathbf{t}} \,\middle|\, \frac{\partial \mathbf{M}\mathbf{v}}{\partial \mathbf{V}}\right) \mathbf{G}_p.$$

A straightforward application of matrix differentiation techniques, as outlined neatly in Nel (1980), then gives

$$\left(\frac{\partial \boldsymbol{\psi}}{\partial \boldsymbol{\theta}}\right) = \mathbf{H}_p \left[-\boldsymbol{\Omega} \begin{pmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{A}_{21} & \mathbf{A}_{22} \end{pmatrix} \boldsymbol{\Omega}^{-1} + \begin{pmatrix} \mathbf{B}_{11} & \mathbf{0} \\ \mathbf{B}_{12} & \mathbf{B}_{22} \end{pmatrix} \right] \mathbf{G}_p,$$

where

$$\mathbf{A}_{11} = 2a_0'(\mathbf{y}'\mathbf{y})(\mathbf{y} \otimes \mathbf{y}') : m \times m, \qquad A_{12} = a_0'(\mathbf{y}'\mathbf{y})(\mathbf{y} \otimes \mathbf{v}' \otimes \mathbf{y}') : m \times m^2,$$
$$\mathbf{A}_{21} = 2a_1'(\mathbf{y}'\mathbf{y})[\mathbf{I} - (\operatorname{vec}\mathbf{I}_m)(\operatorname{vec}\mathbf{S})'](\mathbf{y} \otimes \mathbf{y} \otimes \mathbf{y}') : m^2 \times m,$$
$$\mathbf{A}_{22} = a_1'(\mathbf{y}'\mathbf{y})[I - (\operatorname{vec}\mathbf{I}_m)(\operatorname{vec}\mathbf{S})'](\mathbf{v} \otimes \mathbf{y} \otimes \mathbf{y}' \otimes \mathbf{y}') : m^2 \times m^2,$$
$$\mathbf{B}_{11} = -a_0(\mathbf{y}'\mathbf{y})\mathbf{I}_m, \qquad \mathbf{B}_{22} = -a_1(\mathbf{y}'\mathbf{y})(\operatorname{vec}\mathbf{S})'(\mathbf{y} \otimes \mathbf{y})I_{m^2},$$
$$\mathbf{B}_{12} = a_1(\mathbf{y}'\mathbf{y})[(\operatorname{vec} V)\mathbf{y}'\mathbf{V}^{\frac{1}{2}\prime}\mathbf{T} - (\mathbf{I} + \mathbf{I}_{(m,m)})(\mathbf{V}^{\frac{1}{2}}\mathbf{y} \otimes \mathbf{I}_m)] : m^2 \times m.$$

By symmetry, A_{12}, A_{21}, and B_{12} have zero expectations at $\boldsymbol{\theta} = \theta_0$, and $\mathscr{E}(\mathbf{A}_{11})_{\theta_0} = \mathscr{E}[2y_1^2 a_0'(\mathbf{y}'\mathbf{y})]\mathbf{I}_m$. Put $\mathbf{S}_0 = \mathbf{S}|_{\mathbf{V}=\boldsymbol{\Sigma}}$, $\boldsymbol{\Omega}_0 = \boldsymbol{\Omega}|_{\mathbf{V}=\boldsymbol{\Sigma}}$. Then

$$\mathscr{E}[\mathbf{B}_{22}]_{\theta_0} = -\beta(\operatorname{vec}\mathbf{S}_0)'(\operatorname{vec}\mathbf{I}_m)\mathbf{I}_{m^2} = -\beta(\operatorname{tr} S_0)\mathbf{I}_{m^2} = -\beta\,\tau(\boldsymbol{\Sigma})\mathbf{I}_{m^2} = -\beta\mathbf{I}_{m^2}.$$

For any random vector $\mathbf{Y}$ with a spherically symmetric density, $\mathbf{Y}/|\mathbf{Y}|$ is distributed independently of $|\mathbf{Y}|$, and the former is uniformly distributed over the surface of the unit sphere. It follows that $\mathscr{E}\{Y_1^4\phi(\mathbf{Y}'\mathbf{Y})\} = 3\mathscr{E}\{Y_1^2Y_2^2\phi(\mathbf{Y}'\mathbf{Y})\}$ for all functions ϕ for which the expectations are defined. This fact allows a simple generalization of Theorem 4.1 of Magnus and Neudecker (1979), giving

$$\mathscr{E}\{a_1'(\mathbf{y}'\mathbf{y})(\mathbf{y} \otimes \mathbf{y} \otimes \mathbf{y}' \otimes \mathbf{y}')\} = \gamma\{\mathbf{I} + \mathbf{I}_{(m,m)} + (\operatorname{vec}\mathbf{I}_m)(\operatorname{vec}\mathbf{I}_m)'\} \tag{6.3}$$

so that $\mathscr{E}\mathbf{A}_{22} = \gamma[\mathbf{I} + \mathbf{I}_{(m,m)} - (\operatorname{vec} I_m)\{\operatorname{vec}(S_0 + S_0')\}']$.
Collecting terms and applying (6.1), (6.2) gives

$$\begin{aligned}\mathscr{E}\left(\frac{\partial \boldsymbol{\psi}}{\partial \boldsymbol{\theta}}\right)_{\theta_0} &= -\mathbf{H}_p\boldsymbol{\Omega}_0[\alpha\mathbf{I}_m \oplus \{\beta\mathbf{I}_{m^2} + \mathscr{E}(\mathbf{A}_{22})_{\theta_0}\}]\boldsymbol{\Omega}_0^{-1}\mathbf{G}_p \\ &= -(\mathbf{H}_p\boldsymbol{\Omega}_0\mathbf{G}_p)[\alpha\mathbf{I}_m \oplus \mathbf{H}\{\beta\mathbf{I}_{m^2} + \mathscr{E}(\mathbf{A}_{22})_{\theta_0}\}\mathbf{G}](\mathbf{H}_p\boldsymbol{\Omega}_0^{-1}\mathbf{G}_p) \\ &= -(\mathbf{H}_p\boldsymbol{\Omega}_0\mathbf{G}_p)(\alpha\mathbf{I}_m \oplus [(\beta + 2\gamma)\mathbf{I}_{m(m+1)/2} \\ &\qquad - \gamma\mathbf{H}(\operatorname{vec}\mathbf{I}_m)\{\operatorname{vec}(\mathbf{S}_0 + \mathbf{S}_0')\}'\mathbf{G}])(\mathbf{H}_p\boldsymbol{\Omega}_0^{-1}\mathbf{G}_p).\end{aligned}$$

Applying the identity $|\mathbf{I} - \mathbf{x}\mathbf{y}'| = 1 - \mathbf{y}'\mathbf{x}$ shows that, if and only if (i)–(iii) of Theorem 4.1 hold, the inverse exists and equals

$$-(\mathbf{H}_p\boldsymbol{\Omega}_0\mathbf{G}_p)\Big\{\alpha^{-1}\mathbf{I}_m \oplus (\beta + 2\gamma)^{-1}$$
$$\times\Big(\mathbf{I} + \frac{\gamma}{\beta}\mathbf{H}(\operatorname{vec}\mathbf{I}_m)\{\operatorname{vec}(\mathbf{S}_0 + \mathbf{S}_0')\}'\mathbf{G}\Big)\Big\}(\mathbf{H}_p\boldsymbol{\Omega}_0^{-1}\mathbf{G}_p)$$
$$= -\Big[\alpha^{-1}\mathbf{I}_m \oplus (\beta + 2\gamma)^{-1}\Big(\mathbf{I} + \frac{\gamma}{\beta}(\operatorname{vec}_s \boldsymbol{\Sigma})\{\operatorname{vec}(\mathbf{S}_0 + \mathbf{S}_0')\}'(\boldsymbol{\Sigma}^{-\frac{1}{2}} \otimes \boldsymbol{\Sigma}^{-\frac{1}{2}})\mathbf{G}\Big)\Big].$$

Now (4.3) follows, since

$$\{\operatorname{vec}(\mathbf{S}_0 + \mathbf{S}_0')\}'(\boldsymbol{\Sigma}^{-\frac{1}{2}} \otimes \boldsymbol{\Sigma}^{-\frac{1}{2}})G\boldsymbol{\psi}_2|_{\boldsymbol{\theta}_0}$$
$$= \{\operatorname{vec}(\mathbf{S}_0 + \mathbf{S}_0')\}'\{\mathbf{I}_{m^2} - (\operatorname{vec}\mathbf{I}_m)(\operatorname{vec}\mathbf{S}_0)'\}a_1(\mathbf{y}'\mathbf{y})(\mathbf{y} \otimes \mathbf{y})$$
$$= a_1(\mathbf{y}'\mathbf{y})\{\operatorname{vec}(\mathbf{S}_0' - \mathbf{S}_0)\}'(\mathbf{y} \otimes \mathbf{y}) = a_1(\mathbf{y}'\mathbf{y})(\mathbf{y}'\mathbf{S}_0'\mathbf{y} - \mathbf{y}'\mathbf{S}_0\mathbf{y}) = 0.$$

The asymptotic covariance matrix of $\sqrt{n}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)$ is thus

$$\mathscr{E}(\boldsymbol{\Phi}\boldsymbol{\Phi}') = \{\alpha^{-1}I \oplus (\beta + 2\gamma)^{-1}I\}\mathscr{E}(\boldsymbol{\psi}\boldsymbol{\psi}')_{\boldsymbol{\theta}_0}\{\alpha^{-1}I \oplus (\beta + 2\gamma)^{-1}I\}. \tag{6.4}$$

With $\mathbf{M}_0, \mathbf{M}_1$ defined by $\mathbf{M}|_{\mathbf{V}=\boldsymbol{\Sigma}} = \mathbf{M}_0 = \boldsymbol{\Sigma}^{\frac{1}{2}} \oplus \mathbf{M}_1$,

$$\mathscr{E}(\boldsymbol{\psi}\boldsymbol{\psi}')_{\boldsymbol{\theta}_0} = \mathbf{H}_p\mathbf{M}_0\mathscr{E}(\mathbf{v}\mathbf{v}')_{\boldsymbol{\theta}_0}\mathbf{M}_0'\mathbf{H}_p'$$
$$= \alpha_1\boldsymbol{\Sigma} \oplus \gamma_1\mathbf{H}\mathbf{M}_1\{\mathbf{I} + \mathbf{I}_{(m,m)} + (\operatorname{vec}\mathbf{I}_m)(\operatorname{vec}\mathbf{I}_m)'\}\mathbf{M}_1'\mathbf{H}',$$

by symmetry and the analogue of (6.3). With $\boldsymbol{\Lambda} = I - \frac{1}{2}\mathbf{H}(\operatorname{vec}\boldsymbol{\Sigma})(\operatorname{vec}\mathbf{T})'\mathbf{G}$ and $\mathbf{C} = 2\mathbf{H}(\boldsymbol{\Sigma} \otimes \boldsymbol{\Sigma})\mathbf{H}'$, this becomes

$$\mathscr{E}(\boldsymbol{\psi}\boldsymbol{\psi}')_{\boldsymbol{\theta}_0} = \alpha_1\boldsymbol{\Sigma} \oplus \gamma_1\boldsymbol{\Lambda}\mathbf{C}\boldsymbol{\Lambda}'. \tag{6.5}$$

Combining (6.5), (6.4) gives

$$\mathscr{E}(\boldsymbol{\Phi}\boldsymbol{\Phi}') = \frac{\alpha_1}{\alpha^2}\boldsymbol{\Sigma} \oplus \frac{\gamma_1}{(\beta + 2\gamma)^2}\boldsymbol{\Lambda}\mathbf{C}\boldsymbol{\Lambda}'.$$

Evaluating the individual elements of $\{\gamma_1/(\beta + 2\gamma)^2\}\boldsymbol{\Lambda}\mathbf{C}\boldsymbol{\Lambda}' = \mathbf{B}$ then completes the proof.

ACKNOWLEDGEMENTS

We gratefully acknowledge the guidance and support of Professor J.R. Collins, University of Calgary, who supervised the graduate work of the first author; and Professor P.J. Bickel, at whose invitation the second author visited the University of California, Berkeley.

REFERENCES

Andrews, D.F.; Bickel, P.J.; Hampel, F.R.; Huber, P.J.; Rogers, W.H., and Tukey, J.W. (1972). *Robust Estimates of Location*. Princeton Univ. Press, Princeton, N.J.

Bickel, P.J. (1982). On adaptive estimation. *Ann. Statist.*, 10, 647–671.

Collins, J.R. (1976). Robust estimation of a location parameter in the presence of asymmetry. *Ann. Statist.*, 4, 68–85.

Collins, J.R. (1982). Robust *M*-estimators of location vectors. *J. Multivariate Anal.*, 12, 480–492.

Collins, J.R.; Sheahan, J.N., and Zheng, Z. (1985). Robust estimation in the linear model with asymmetric error distributions. *J. Multivariate Anal.*, to appear.

Collins, J.R., and Wiens, D.P. (1985). Minimax variance *M*-estimators in ϵ-contamination models. *Ann. Statist.*, 13, 1078–1096.

Henderson, H.V., and Searle, S.R. (1979). Vec and Vech operators for matrices, with some uses in Jacobians and multivariate statistics. *Canad. J. Statist.*, 7, 65–81.

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

Huber, P.J. (1973). Robust regression: Asymptotics, conjectures and Monte Carlo. *Ann. Statist.*, 1, 799–821.

Huber, P.J. (1977). Robust covariances. *Statistical Decision Theory and Related Topics, II* (S.S. Gupta and D.S. Moore, eds.), Academic Press, New York.

Huber, P.J. (1981). *Robust Statistics.* Wiley.

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

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

McCulloch, C.E. (1982). Symmetric matrix derivatives with applications. *J. Amer. Statist. Assoc.*, 77, 679–682.

Nel, D.G. (1980). On matrix differentiation in statistics. *South African Statist. J.*, 14, 137–193.

Ortega, J.M., and Rheinboldt, W.C. (1970). *Iterative Solution of Non-linear Equations in Several Variables.* Academic Press, New York.

Tyler, D.E. (1982). Radial estimates and the test for sphericity. *Biometrika*, 69, 429–436.

Tyler, D.E. (1985). A distribution-free *M*-estimator of multivariate scatter. Unpublished manuscript.

Watson, G.N. (1966). *A Treatise on the Theory of Bessel Functions.* University Press, Cambridge.

Wiens, D.P. (1982). Robust estimation of multivariate location and scale in the presence of asymmetry. Ph.D. Thesis, Univ. of Calgary.

Wiens, D.P. (1985). On some pattern reduction matrices which appear in statistics. *Linear Algebra Appl.*, 67, 233–258.

Wiens, D.P., and Zheng, Z. (1985). Robust estimators of multivariate location and scatter in the presence of asymmetry. *Technical Report* DALTRS-85-5, Dalhousie University.

Zheng, Z. (1980). *M*-estimators of multivariate location and scatter. Unpublished technical report, University of California, Berkeley.

Received 25 August 1984
Revised 2 January 1986
Accepted 20 January 1986

Department of Mathematics, Statistics and Computing Science
Dalhousie University
Halifax, Nova Scotia B3H 3J5

Department of Mathematics
Peking University
Beijing, People's Republic of China