THE UNIVERSITY OF CALGARY

ROBUST ESTIMATION OF MULTIVARIATE LOCATION AND SCALE IN THE PRESENCE OF ASYMMETRY

by

© DOUGLAS P. WIENS

A THESIS

SUBMITTED TO THE FACULTY OF GRADUATE STUDIES

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE

OF DOCTOR OF PHILOSOPHY

DEPARTMENT OF MATHEMATICS AND STATISTICS

CALGARY, ALBERTA

APRIL, 1982

To my parents

ABSTRACT

We consider the problem of estimating the location vector and scale matrix of a random vector, when the distribution is only approximately known. Within the ellipse $E_r = \{\underline{x} | (\underline{x}-\underline{\mu})'\Sigma^{-1}(\underline{x}-\underline{\mu}) \leq r^2\}$, the density of the random vector $\underline{x}$ is elliptically symmetric, and arises from a known density through ε-contamination. Outside of E_r the distribution is arbitrary and unknown. The problem is to estimate $\underline{\mu}$ and Σ.

It turns out that the parameter point $(\underline{\mu},\Sigma)$ is not identifiable in the model, and so we estimate instead the identifiable point $\theta = \left(\underline{\mu}, \frac{\Sigma}{\tau(\Sigma)}\right)$, where $\tau(\cdot)$ is a scalar-valued function such as Σ_{11}. Following Huber [3] and Collins [1] we use the method of M-estimators and re-descending influence functions. The estimators are defined as the zeroes of a function of the form $\sum_i \underline{\psi}(\underline{x}_i;\hat{\theta})$. In order that the observations from outside of $E_r(\theta)$ should have no influence on the estimators, $\underline{\psi}$ is chosen to vanish off of $E_r(\hat{\theta})$. We show that $\hat{\theta}$ is a consistent estimator of θ, and that $\sqrt{n}\hat{\theta}$ is asymptotically normally distributed. Confidence regions for linear functions of θ are constructed. In the case $r = \infty$ our results complement those of Huber [5] and Maronna [8], who considered related problems assuming global symmetry.

The associated optimization problem - that of finding functions $\underline{\psi}$ which minimize the maximum asymptotic variance of $\hat{\theta}$ as the distribution of $\underline{x}$ varies - is also considered. It is exhibited as a special case of the problem of minimaxing a general variance functional which typically arises in problems of robust estimation. Under fairly general conditions, this problem is solved for the ε-contamination model.

ACKNOWLEDGEMENTS

I am greatly indebted to Dr. Collins for his helpful advice and suggestions during the preparation of this thesis. I am also grateful to Karen McDermid for her accurate typing of the thesis.

Financial support for the period during which this research was carried out was provided by the University of Calgary, the Province of Alberta Graduate Fellowship fund, and the Izaak Walton Killam Memorial Scholarship fund.

TABLE OF CONTENTS

Page

LIST OF TABLES

0. INTRODUCTION

In this thesis, we apply Huber's theory of robustness to the problem of multivariate estimation of location and scale, when the underlying distribution is only approximately known. It is assumed that within an ellipse of fixed radius, with centre and shape defined by the unknown parameters, the observations are symmetrically distributed, and with probability exceeding 1/2 are generated by some known law. The remaining, "contaminated" observations from within this ellipse are symmetrically distributed, but according to an unknown law. The observations from outside the ellipse may be arbitrarily distributed.

Although there is no particularly compelling reason to believe that contaminated observations should be even locally symmetrically distributed, the assumption of local symmetry is mathematically necessary. Otherwise, there is no natural parameter point which can even be identified, let alone estimated. On the other hand, global symmetry may be unduly restrictive.

Definition of the model

An unobservable random vector (r.vec.) $\underline{y} \in R^m$, $m > 1$, has a partially known density

$$u(\underline{y}) = \begin{cases} (1-\varepsilon)w(\underline{y}) + \varepsilon v(\underline{y}), & \underline{y}'\underline{y} \leq r^2, \\ \text{arbitrary}, & \underline{y}'\underline{y} > r^2. \end{cases} \tag{1}$$

Here, $\varepsilon \in [0,1)$, $r \in (0,\infty]$ and $w(\underline{y})$ are known and fixed. The densities

u, w and v are spherically symmetric in that they depend upon $\underline{y}$ only through $\underline{y}'\underline{y} = |\underline{y}|^2$. One observes n independent realizations of an affine transformation $\Sigma^{\frac{1}{2}}\underline{y}+\underline{\mu}$ of $\underline{y}$, where $\underline{\mu} \in R^m$ and $\Sigma_{m\times m} > 0$ are unknown. For the moment, $\Sigma^{\frac{1}{2}}$ will represent any matrix satisfying $\Sigma^{\frac{1}{2}}\Sigma^{\frac{1}{2}'} = \Sigma$, in a class within which it is unique.

We denote by $E_r(\underline{\mu},\Sigma)$, or simply by E_r where possible, the ellipse $\{\underline{x}|(\underline{x}-\underline{\mu})'\Sigma^{-1}(\underline{x}-\underline{\mu}) \leq r^2\}$, and by D_r the sphere $E_r(\underline{0},I)$. Put

$$
\begin{aligned}
w(\underline{x};\underline{\mu},\Sigma) &= |\Sigma|^{-\frac{1}{2}}w(\Sigma^{-\frac{1}{2}}(\underline{x}-\underline{\mu})), \\
v(\underline{x};\underline{\mu},\Sigma) &= |\Sigma|^{-\frac{1}{2}}v(\Sigma^{-\frac{1}{2}}(\underline{x}-\underline{\mu})).
\end{aligned} \tag{2}
$$

Then the observed r.vec. $\underline{x} = \Sigma^{\frac{1}{2}}\underline{y}+\underline{\mu}$ has density

$$
u(\underline{x};\underline{\mu},\Sigma) = \begin{cases} (1-\varepsilon)w(\underline{x};\underline{\mu},\Sigma) + \varepsilon v(\underline{x};\underline{\mu},\Sigma), & \underline{x} \in E_r(\underline{\mu},\Sigma), \\ \\ \text{arbitrary}, & \underline{x} \notin E_r(\underline{\mu},\Sigma). \end{cases} \tag{3}
$$

Define

$$
\mathcal{U}_{\varepsilon,r} = \{u(\underline{x};\underline{\mu},\Sigma)|\underline{\mu} \in R^m,\ \Sigma_{m\times m} > 0;\ u \text{ is given by (3) for some } v(\underline{x};\underline{\mu},\Sigma) \text{ which is elliptically symmetric throughout } E_r\}.
$$

Our original intention was to estimate $\underline{\mu}$, and Σ itself. As will be shown in Section 1, however, the parameter point $(\underline{\mu},\Sigma)$ is not identifiable in $\mathcal{U}_{\varepsilon,r}$. Although $\underline{\mu}$ is identifiable if Σ is known, Σ is only identifiable up to a scalar multiple, even if $\underline{\mu}$ is known. For this reason, we will instead construct an estimator of $\left(\underline{\mu}, \frac{\Sigma}{\tau(\Sigma)}\right)$, where $\tau(\cdot)$ is some scalar valued, linear function such as Σ_{11} or $tr(\Sigma)$. An immediate consequence of this is that affine invariance is lost. Also, if $r < \infty$, then $\tau(\Sigma)$ must be assumed to be known. The consequences of imperfect knowledge of $\tau(\Sigma)$ are discussed in Section 3.

In Section 1, we derive necessary and sufficient conditions on ε and r under which $\left(\underline{\mu}, \frac{\Sigma}{\tau(\Sigma)}\right)$ is identifiable. Intuitively, the mass of $(1-\varepsilon)w(\underline{x};\underline{\mu},\Sigma)$ within E_r must be slightly greater than 1/2, the limit of 1/2 being approached as $r \to \infty$.

In Section 3, we will define the estimator $\hat{\theta}$ of $\theta = \left(\underline{\mu}, \frac{\Sigma}{\tau(\Sigma)}\right)$ as the Newton-Raphson solution to an equation of the form $\sum_{i=1}^{n} \underline{\psi}(\underline{x}_i;\hat{\theta}) = \underline{0}$. In order that, asymptotically, the asymmetrically distributed observations should have no influence on the estimator, the functions $\underline{\psi}$ will be chosen to vanish off of $E_r(\hat{\theta})$. This poses problems with uniqueness of solutions, which are dealt with in Section 2 by constructing a preliminary, consistent estimator of θ which is used as the starting value of the iterative process. As is shown in Section 3, this implies that $\hat{\theta}$ is consistent for θ. A simple one-step Taylor expansion, and the multivariate CLT, establish asymptotic normality of $\sqrt{n}\hat{\theta}$. The limiting covariance matrix of $\sqrt{n}(\hat{\theta}-\theta)$ is exhibited. This leads to the construction of asymptotic confidence regions for arbitrary linear functions of θ, based on the normal theory. We will as well show that the Newton-Raphson method may be replaced by a much simpler fixed point process which, asymptotically, has the same optimal convergence properties.

The limiting covariance matrix of $\sqrt{n}\hat{\theta}$ is the product of a matrix depending only upon θ, and a scalar functional depending only upon $\underline{\psi}$ and $\dot{u}$. A natural optimization problem is then suggested: find functions $\underline{\psi}$ which are "most robust" in that they minimize, with respect to the ordering by positive definiteness of p.d. matrices, the maximum asymptotic variance of $\sqrt{n}\hat{\theta}$ as u ranges over $\mathcal{U}_{\varepsilon,r}$. We will exhibit this problem as a special case of a more general problem - that of minimaxing a general variance functional which typically arises in connection with robust estimation problems. Under fairly general conditions, this problem is solved in Section 4, for the ε-contamination model. Special applications include the solutions to the optimization problem for $\mathcal{U}_{\varepsilon,r}$. These are given in Section 5.

In the one-dimensional location problem considered by Huber [3], the most robust estimator turned out to be the maximum likelihood estimator for that density which is "least favourable" in the sense that it has minimum Fisher information. For that reason, this estimator is termed an M-estimator. Although a treatment of these problems in a more general framework requires a broader formulation of "information", and the interpretation of the estimator as a maximum likelihood estimator is obscured, we retain the term M-estimator for historical reasons.

Huber [5] and Maronna [8] have considered the problem of estimation of location and scale, under the assumption of global symmetry. Their analyses are essentially carried out in our class $\mathcal{U}_{0,\infty}$. In Section 3 we will discuss the applicability of their results to the class $\mathcal{U}_{\varepsilon,\infty}$, $\varepsilon > 0$. It will be argued that in this case our approaches are significantly different, that their results are not merely limiting cases, as $r \to \infty$, of ours, and that the estimators should be considered as competitors.

Notation

Recall (1) - (3). We define $f, g, h, f_s, g_s, h_s : R \to R$ by

$$u(\underline{y}) = f_s(\underline{y}'\underline{y}) = f(|\underline{y}|),$$

$$w(\underline{y}) = h_s(\underline{y}'\underline{y}) = h(|\underline{y}|), \tag{4}$$

$$v(\underline{y}) = g_s(\underline{y}'\underline{y}) = g(|\underline{y}|).$$

This formalizes the notion of spherical symmetry of u, w and v.

If $\underline{y} \sim u(\underline{y};\underline{0},I)$, then on D_r, $Z = \underline{y}'\underline{y}$ has d.f. $F_m(z)$ and density

$$f_m(z) = \frac{mC_m}{2} z^{m/2\ -1} f_s(z), \tag{5}$$

where $C_m = \dfrac{\pi^{m/2}}{\Gamma(m/2\ +1)}$ is the volume of D_1. We define h_m, H_m, g_m and G_m analogously. In the important special case that $w(\underline{y};\underline{0},I)$ is the m-variate normal density $\phi(\underline{y};\underline{0},I)$, so that $h_s(z) = (2\pi)^{-m/2} e^{-z/2}$, we write $X_m^2(z)$ and $\chi_m^2(z)$ for $H_m(z)$ and $h_m(z) = \dfrac{(z/2)^{m/2\ -1} e^{-z/2}}{2\Gamma(m/2)}$.

Model Assumptions

M1) The r.v. Z has a finite first moment:

$$\int_{R^m} (\underline{y}'\underline{y}) u(\underline{y};\underline{0},I) d\underline{y} = \int_0^\infty z f_m(z) dz < \infty \text{ for all } u \in \mathcal{U}_{\varepsilon,r}.$$

M2) The function $h_s(z)$ is non-increasing on $[0,r^2]$. Until Section 5, we do not assume that $h_s(z)$ is continuous.

The following assumption is made purely because it proves to be mathematically convenient at one point.

M3) For every $u \in U_{\varepsilon,r}$, f_s is non-constant on $[a^2,r^2]$, for every a^2.

1. IDENTIFIABILITY

Definition: Let $m(\cdot)$ be a matrix valued function. The pair $(\underline{\mu}_1, m(\Sigma_1))$ is $(m\text{-})$ *identifiable* in $U_{\varepsilon,r}$ if

$$u_1(\underline{x};\underline{\mu}_1,\Sigma_1) \equiv u_2(\underline{x};\underline{\mu}_2,\Sigma_2) \in U_{\varepsilon,r} \text{ implies } (\underline{\mu}_1, m(\Sigma_1)) = (\underline{\mu}_2, m(\Sigma_2)).$$

The following lemma shows that when $\underline{\mu}_1 = \underline{\mu}_2$, we cannot distinguish Σ_1 from $\Sigma_2 = \alpha\Sigma_1$ for α in a neighborhood of 1, even if $r = \infty$.

LEMMA 1.1 For any $\varepsilon \in (0,1)$, $r \in (0,\infty]$, $\underline{\mu} \in R^m$, $\Sigma > 0$, the pair $(\underline{\mu},\Sigma)$ is not identifiable in $U_{\varepsilon,r}$ if $w(\underline{x};\underline{0},I)$ is continuous on $D_r(\underline{0},I)$.

Proof: Let $0 < \alpha < 1$, and define

$$v_1(\underline{x};\underline{0},I) = \max\left\{\frac{1-\varepsilon}{\varepsilon}(w(\underline{x};\underline{0},\alpha I) - w(\underline{x};\underline{0},I)), 0\right\}, \quad \underline{x} \in D_r(\underline{0},\alpha I);$$

$$v_2(\underline{x};\underline{0},\alpha I) = \max\left\{\frac{1-\varepsilon}{\varepsilon}(w(\underline{x};\underline{0},I) - w(\underline{x};\underline{0},\alpha I)), 0\right\}, \quad \underline{x} \in D_r(\underline{0},\alpha I);$$

$$u_1(\underline{x};\underline{0},I) = (1-\varepsilon)w(\underline{x};\underline{0},I) + \varepsilon v_1(\underline{x};\underline{0},I), \quad \underline{x} \in D_r(\underline{0},I);$$

$$u_2(\underline{x};\underline{0},\alpha I) = \begin{cases} (1-\varepsilon)w(\underline{x};\underline{0},\alpha I) + \varepsilon v_2(\underline{x};\underline{0},\alpha I), & \underline{x} \in D_r(\underline{0},\alpha I); \\ u_1(\underline{x};\underline{0},I), & \underline{x} \notin D_r(\underline{0},\alpha I); \end{cases}$$

and define these functions to be zero elsewhere. Then

$$u_2(\underline{x};\underline{0},\alpha I) = \begin{cases} (1-\varepsilon)\max\{w(\underline{x};\underline{0},I), w(\underline{x},\underline{0},\alpha I)\}, & \underline{x} \in D_r(\underline{0},\alpha I); \\ u_1(\underline{x};\underline{0},I), & \underline{x} \notin D_r(\underline{0},\alpha I); \end{cases}$$

$$\equiv u_1(\underline{x};\underline{0},I).$$

The functions v_1 and v_2 are spherically symmetric on $D_r(\underline{0},I)$ and $D_r(\underline{0},\alpha I)$ respectively. It is easy to see that for α sufficiently close to 1 the functions defined above are, possibly sub-stochastic, densities, and so $u_1 \equiv u_2 \in U_{\varepsilon,r}$. Transforming to $\Sigma^{\frac{1}{2}}\underline{x} + \underline{\mu}$ yields the result. □

If ε is sufficiently small and r is sufficiently large, we can identify the representatives of certain equivalence classes of the parameter space. Let M be the class of $m{\times}m$ real matrices, M^+ the positive definite members of M. Let $\tau\colon M \to R$ be a continuous linear map which commutes with the expectation operator and whose restriction to M^+ is positive; e.g. $\tau(M) = M_{11}$, $\tau(M) = tr(M)$.

LEMMA 1.2 For Σ, Σ_1, $\Sigma_2 > 0$ and $a > 0$:

i) $\Sigma = aI \Rightarrow \dfrac{\Sigma}{\tau(\Sigma)} = \dfrac{I}{\tau(I)}$,

ii) $\dfrac{\Sigma_1}{\tau(\Sigma_1)} = \dfrac{\Sigma_2}{\tau(\Sigma_2)} \Leftrightarrow \dfrac{\Sigma_1^{\frac{1}{2}}}{\sqrt{\tau(\Sigma_1)}} = \dfrac{\Sigma_2^{\frac{1}{2}}}{\sqrt{\tau(\Sigma_2)}}$,

iii) $\Sigma = \Sigma_2^{-\frac{1}{2}}\Sigma_1\Sigma_2^{-\frac{1}{2}\,'}$ implies

$$\frac{\Sigma_1}{\tau(\Sigma_1)} = \frac{\Sigma_2}{\tau(\Sigma_2)} \Leftrightarrow \frac{\Sigma}{\tau(\Sigma)} = \frac{I}{\tau(I)} . \qquad \square$$

The proof is trivial and so is omitted.

By parts ii) and iii), and the fact that $U_{\varepsilon,r}$ is closed under affine transformations, $\left(\underline{\mu},\dfrac{\Sigma}{\tau(\Sigma)}\right)$ is identifiable iff $\left(\underline{\mu},\dfrac{\Sigma^{\frac{1}{2}}}{\sqrt{\tau(\Sigma)}}\right)$ is identifiable iff $\left(\underline{0},\dfrac{I}{\tau(I)}\right)$ is identifiable.

LEMMA 1.3 If Σ^{-1} has eigenvalues $\lambda_1 \geq \ldots \geq \lambda_m > 0$, then
$\max_{E_r} \underline{x}'\underline{x} \geq \underline{\mu}'\underline{\mu} + \frac{r^2}{\lambda_m}$.

Proof: Without loss of generality, $\Sigma^{-1} = \text{diag}(\lambda_1,\ldots,\lambda_m)$, so that
$E_r = \{\underline{x} \mid \sum_i \lambda_i (x_i-\mu_i)^2 \leq r^2\}$.

Since $\underline{x}_0 = \left(\mu_1,\ldots,\mu_{m-1},\mu_m + \frac{r}{\sqrt{\lambda_m}} \cdot \frac{\mu_m}{|\mu_m|}\right)' \in E_r, \left(\frac{0}{|0|} = 1\right)$;

$\max_{E_r} \underline{x}'\underline{x} \geq \underline{x}_0'\underline{x}_0 \geq \underline{\mu}'\underline{\mu} + \frac{r^2}{\lambda_m}$. □

The following result will be used to obtain bounds on ε and r which ensure identifiability of $\left(\underline{\mu},\frac{\Sigma}{\tau(\Sigma)}\right)$.

THEOREM 1.4 Let $D_r = E_r(\underline{0},I)$, $E_r = E_r(\underline{\mu},\Sigma)$. Suppose that $D_r \cap E_r \neq \emptyset$, and that $f:R^m \to R$ is a function with the property that on $D_r \cap E_r$, f is both spherically and elliptically symmetric; i.e. $f(\underline{x}) = f_1(\underline{x}'\underline{x}) = f_2((\underline{x}-\underline{\mu})'\Sigma^{-1}(\underline{x}-\underline{\mu}))$ on $D_r \cap E_r$ for two functions $f_1,f_2 : R \to R$. Then either

i) $\left(\underline{\mu},\frac{\Sigma}{\tau(\Sigma)}\right) = \left(\underline{0},\frac{I}{\tau(I)}\right)$;

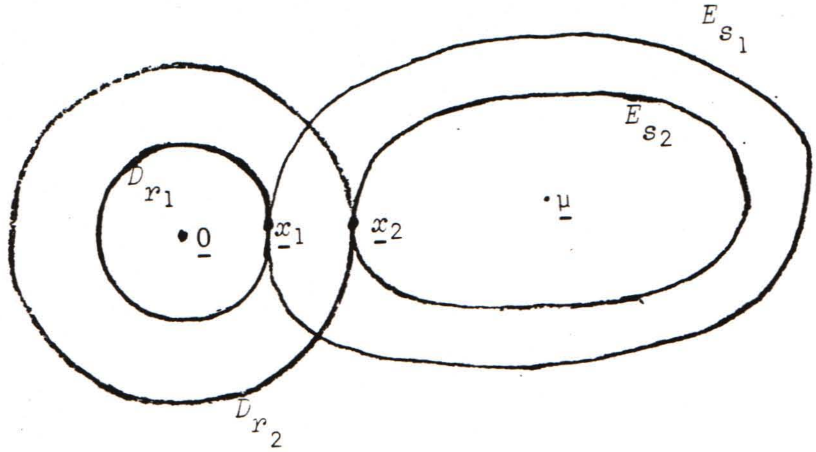
ii) $\underline{\mu} \neq \underline{0}$ and f is constant on $D_r \cap E_r$; or

iii) $\underline{\mu} = \underline{0}$ and f is constant on $D_r \cap E_r \setminus \{\underline{0}\}$.

Proof. We may assume that $f(\underline{x}) = f_1(\underline{x}'\underline{x})$ throughout D_r, and that $f(\underline{x}) = f_2((\underline{x}-\underline{\mu})'\Sigma^{-1}(\underline{x}-\underline{\mu}))$ throughout E_r, since f_1 and f_2 may be extended symmetrically through these regions. Denote by D_r', D_r^0,

E_r' and E_r^0 the boundaries and interiors of D_r and E_r. Let $\underline{x}_1$ and $\underline{x}_2$ be any two points in $D_r \cap E_r$, and suppose that $\underline{\mu} \neq \underline{0}$. We will show that $f(\underline{x}_1) = f(\underline{x}_2)$, by showing that $\underline{x}_1$ and $\underline{x}_2$ lie on intersecting surfaces, along each of which the symmetry of f forces it to be constant. Define $r_i^2 = \underline{x}_i'\underline{x}_i$, $s_i^2 = (\underline{x}_i-\underline{\mu})'\Sigma^{-1}(\underline{x}_i-\underline{\mu})$, and assume that $r_1 < r_2$. Then $f(\underline{x}_i) = f_1(r_i^2) = f_2(s_i^2)$.

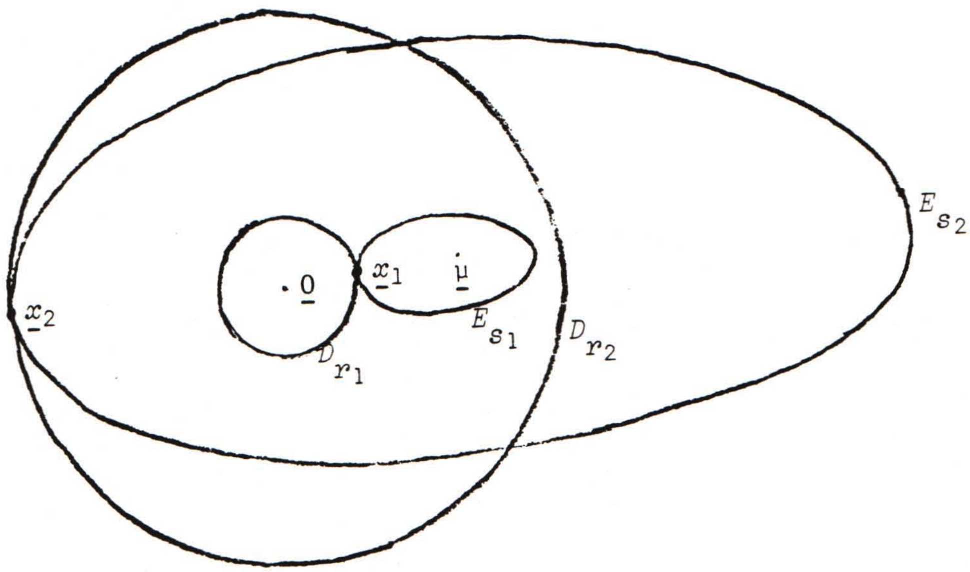
Case i): $0 \leq s_2 < s_1 \leq r$



To show that $f(\underline{x}_1) = f(\underline{x}_2)$, we need only show that there is some $\underline{x}_0 \in D_{r_2}' \cap E_{s_1}'$, since then $f(\underline{x}_1) = f_2(s_1^2) = f_2(s_0^2) = f(\underline{x}_0) = f_1(r_0^2) = f_1(r_2^2) = f(\underline{x}_2)$, using first the elliptical, then the spherical, symmetry of f. To establish the existence of $\underline{x}_0$, it suffices to show that D_{r_2}' intersects both E_{s_1} and $(E_{s_1}^0)^c$ (complementation). But since $\underline{x}_2 \in D_{r_2}' \cap E_{s_2}' \subset D_{r_2}' \cap E_{s_1}$, this will follow from $D_{r_2}' \cap (E_{s_1}^0)^c \neq \emptyset$. Suppose for contradiction that $D_{r_2}' \subset E_{s_1}^0$. Then $D_{r_1} \subset D_{r_2} \subset E_{s_1}^0$, so that $D_{r_1}' \cap E_{s_1}' = \emptyset$, contradicting $\underline{x}_1 \in D_{r_1}' \cap E_{s_1}'$.

Case ii): $0 \leq s_1 < s_2 \leq r$

In this case, $\underline{x}_1 \in E'_{s_1} \cap D'_{r_1}$, $D'_{r_1} \subset D_{r_2}$, and $E'_{s_1} \subset E_{s_2}$, so that $\underline{x}_1 \in E'_{s_1} \cap D_{r_2}$ and $\underline{x}_1 \in D'_{r_1} \cap E_{s_2}$. If $E'_{s_1} \cap (D^0_{r_2})^c \neq \emptyset$, then there is an $\underline{x}_0 \in E'_{s_1} \cap D'_{r_2}$. If $D'_{r_1} \cap (E^0_{s_2})^c \neq \emptyset$, then there is an $\underline{x}_0 \in D'_{r_1} \cap E'_{s_2}$. In either case, we may argue as in Case i) to get $f(\underline{x}_1) = f(\underline{x}_2)$. Suppose then that $E'_{s_1} \subset D^0_{r_2}$, and that $D'_{r_1} \subset E^0_{s_2}$.



The points $\underline{x}_1$ and $\underline{x}_2$ can be connected by intersecting surfaces along which f is constant. We do this by constructing the largest sphere around $\underline{0}$ intersecting E_{s_1}, then the largest ellipse around $\underline{\mu}$ intersecting this sphere, etc. The surfaces so generated must eventually intersect D'_{r_2}, unless $\underline{\mu} = \underline{0}$ and either $\underline{x}_1 = \underline{0}$, or E_{s_1} is in fact a sphere centered at $\underline{0}$.

To this end, define $r_{11}^2 = \max_{E'_{s_1}} \underline{x}'\underline{x}$, and suppose that r_{11}^2 is attained at $\underline{x}_{11}$. Then $E_{s_1} \subset D_{r_{11}}$; and $\underline{x}_1 \in E'_{s_1} \Rightarrow r_1^2 = \underline{x}_1'\underline{x}_1 \leq r_{11}^2$ $\Rightarrow D_{r_1} \subset D_{r_{11}}$, so that $D_{r_1} \cup E_{s_1} \subset D_{r_{11}}$ and $\underline{x}_{11} \in E'_{s_1} \cap D'_{r_{11}}$. Define $s_{12}^2 = \max_{D'_{r_{11}}}((\underline{x}-\underline{\mu})'\Sigma^{-1}(\underline{x}-\underline{\mu}))$, and suppose that s_{12}^2 is attained at $\underline{x}_{12}$. Then $D_{r_{11}} \subset E_{s_{12}}$ and $\underline{x}_{12} \in D'_{r_{11}} \cap E'_{s_{12}}$. Continue, generating the chain $D_{r_1} \cup E_{s_1} \subset D_{r_{11}} \subset E_{s_{12}} \subset D_{r_{13}} \subset \ldots$, with $\underline{x}_1 \in D'_{r_1} \cap E'_{s_1}$, $\underline{x}_{11} \in E'_{s_1} \cap D'_{r_{11}}$, $\underline{x}_{12} \in D'_{r_{11}} \cap E'_{s_{12}}, \ldots$. Suppose that $r_{1,2n+1} \to \infty$ as $n \to \infty$. Then the connected surface

$$\bigcup_{i=0}^{n} \left(D'_{r_{1,2i+1}} \cup E'_{s_{1,2i+2}}\right) \cup E'_{s_1} \cup D'_{r_1} \text{ intersects } D'_{r_2} \text{ for some } n.$$

Since the symmetry of f forces it to be constant everywhere on this surface, we will have $f(\underline{x}_1) = f(\underline{x}_2)$.

To see under what conditions $r_{1,2n+1} \to \infty$, let Σ^{-1} have eigenvalues $\lambda_1 \geq \ldots \geq \lambda_m > 0$. By Lemma 1.3, $r_{11}^2 \geq \underline{\mu}'\underline{\mu} + s_1^2/\lambda_m$. Thus $E_{s_{12}}$ contains a sphere of radius $\left(\underline{\mu}'\underline{\mu} + s_1^2/\lambda_m\right)^{\frac{1}{2}}$, and so $s_{12}/\sqrt{\lambda_1}$ the length of the shortest semi-axis of $E_{s_{12}}$, is at least $\left(\underline{\mu}'\underline{\mu} + s_1^2/\lambda_m\right)^{\frac{1}{2}}$. With $k = \dfrac{\lambda_1}{\lambda_m} \geq 1$, we have $s_{12}^2 \geq \lambda_1\underline{\mu}'\underline{\mu} + ks_1^2$. Again by Lemma 1.3, $r_{13}^2 \geq \underline{\mu}'\underline{\mu} + s_{12}^2/\lambda_m \geq (k+1)\underline{\mu}'\underline{\mu} + ks_1^2/\lambda_m$. Continuing,

$$r_{1,2n+1} \geq \left(\sum_{i=0}^{n} k^i\right)\underline{\mu}'\underline{\mu} + \frac{k^n s_1^2}{\lambda_m} \to \infty, \text{ unless } \underline{\mu} = \underline{0} \text{ and either } k = 1 \text{ or}$$

$s_1^2 = 0$.

If $\underline{\mu} = \underline{0}$ and $k = 1$, then $\Sigma = aI$ for $a = \lambda_1 = \ldots = \lambda_m$, so that $\left(\underline{\mu}, \dfrac{\Sigma}{\tau(\Sigma)}\right) = \left(\underline{0}, \dfrac{I}{\tau(I)}\right)$. If $\underline{\mu} = \underline{0}$ and $s_1^2 = 0$, choose new points $\underline{x}_1, \underline{x}_2 \neq \underline{0}$ and repeat the argument, thus seeing that f is constant on $D_r \cap E_r$ except possibly for a discontinuity at $\underline{0} = \underline{\mu}$. □

The following corollary is immediate.

COROLLARY 1.5 Suppose that $f = f_1$ throughout D_r, that $f = f_2$ throughout E_r, and that $\left(\underline{\mu}, \frac{\Sigma}{\tau(\Sigma)}\right) \neq \left(\underline{0}, \frac{I}{\tau(I)}\right)$. Consider the following possibilities:

i) $E_r \subset D_r$;

ii) $D_r \subset E_r$;

iii) $E_r \not\subset D_r$, $D_r \not\subset E_r$, $\underline{0} \in E_r$;

iv) $E_r \not\subset D_r$, $D_r \not\subset E_r$, $\underline{\mu} \in D_r$;

v) $E_r \not\subset D_r$, $D_r \not\subset E_r$, $\underline{0} \notin E_r$, $\underline{\mu} \notin D_r$.

Corresponding to these cases f is constant, except possibly at $\underline{\mu}$ if $\underline{\mu} = \underline{0}$, on

i) E_r

ii) D_r

iii) D_r

iv) E_r

v) $(D_r \setminus D^0_{r_1}) \cup (E_r \setminus E^0_{r_2})$,

where $r_1^2 = \min\limits_{E_r} \underline{x}'\underline{x}$ and $r_2^2 = \min\limits_{D_r}((\underline{x}-\underline{\mu})'\Sigma^{-1}(\underline{x}-\underline{\mu}))$. □

The pair $\left(\underline{\mu}, \frac{\Sigma}{\tau(\Sigma)}\right)$ is not identifiable in $U_{\varepsilon,r}$ if ε is so large or r so small as to permit the existence of two disjoint ellipses $E_r(\underline{\mu}_i, \Sigma_i)$ supporting $(1-\varepsilon)w(\underline{x};\underline{\mu}_i,\Sigma_i)$. One is then unable to determine which region carries the symmetric mass and which the asymmetric contamination. It is thus necessary that

$$\sum_{i=0,1} \int_{E_r(\underline{\mu}_i,\Sigma_i)} (1-\varepsilon)w(\underline{x};\underline{\mu}_i,\Sigma_i)d\underline{x} = 2(1-\varepsilon)H_m(r^2)$$

exceed unity. Intuitively, the proportion of symmetric, uncontaminated mass must exceed 1/2. This rather appealing bound is not quite sufficient if $r < \infty$, however, as the following example, based upon Corollary 1.5(v), shows.

Let $m = 2$, $2(1-\varepsilon)H_2(r^2) > 1$. Consider two overlapping disks $D_r(\underline{0},I)$ and $D_r(\underline{\mu},I)$, where $\underline{\mu} = (2r-k,0)'$. Write $w(x,y)$ for $w(\underline{x};\underline{0},I)$, and define

$$v_1(x,y) = \frac{1-\varepsilon}{\varepsilon}(w(r-k,0)-w(x,y)) \text{ on } D_r(\underline{0},I)\setminus D_{r-k}(\underline{0},I);$$

$$v_2(x,y) = \frac{1-\varepsilon}{\varepsilon}(w(r-k,0)-w(x-(2r-k),y)) \text{ on } D_r(\underline{\mu},I) \setminus D_{r-k}(\underline{\mu},I);$$

and let v_1 and v_2 be zero elsewhere.

Define

$$u_1(x,y;\underline{0},I) = u_2(x,y;\underline{\mu},I) = \begin{cases} (1-\varepsilon)w(x,y)+\varepsilon v_1(x,y) & \text{on } D_r(\underline{0},I); \\ (1-\varepsilon)w(x-(2r-k),y)+\varepsilon v_2(x,y) & \text{on } D_r(\underline{\mu},I); \\ \text{arbitrary,} & \text{elsewhere.} \end{cases}$$

We have defined u_1 and u_2 in such a way that the outer annuli of the disks, generated by rotating the region of intersection, each carry constant density of $(1-\varepsilon)w(r-k,0)$. The disks within these annuli each support a "bi-variate $(1-\varepsilon)\cdot w(x,y)$." It can be shown that for sufficiently small $k > 0$ the functions defined above are densities, and so $u_1 \equiv u_2 \in \mathcal{U}_{\varepsilon,r}$. It appears that that mass on the union of the overlapping disks, which is added through requiring constancy on the annuli, is for small k less than that mass which is removed through no longer having the region of intersection appear twice in the two disjoint disks.

Another way in which $\left(\underline{\mu},\frac{\Sigma}{\tau(\Sigma)}\right)$ may fail to be identifiable in $\mathcal{U}_{\varepsilon,r}$ is if constant density, necessarily exceeding $(1-\varepsilon)w(\underline{\mu};\underline{\mu},\Sigma)$, is permitted throughout E_r. Although this is precluded by our assumption M3), it is in any event ruled out by $\varepsilon < 1 - [1+h_s(0)\mathrm{Vol}(D_r)-H_m(r^2)]^{-1}$, which requires the contaminating mass to exceed unity in order that u attain constancy on E_r. Since this bound on ε exceeds the previous bound for sufficiently large r, it too is generally insufficient.

These observations prove part i) of Theorem 1.6 below.

For $0 \leqslant r_1 \leqslant r \leqslant \infty$, define $f(r,r_1) = H_m(r_1^2) + \frac{1}{2}h_s(r_1^2)\mathrm{Vol}(D_r \setminus D_{r_1})$. Define $\varepsilon_1^* - \varepsilon_5^*$ by

$$\varepsilon_1^* = \min(\varepsilon_3^*,\varepsilon_4^*), \quad \varepsilon_2^* = \min(\varepsilon_4^*,\varepsilon_5^*),$$

$$(1-\varepsilon_3^*)^{-1} = f(r,r) + \inf_{[0,r]} f(r,r_1) = H_m(r^2) + \inf_{[0,r]} f(r,r_1),$$

$$(1-\varepsilon_4^*)^{-1} = 1 + 2f(r,0) - f(r,r) = 1 + h_s(0)\mathrm{Vol}(D_r) - H_m(r^2),$$

$$(1-\varepsilon_5^*)^{-1} = 2f(r,r) = 2H_m(r^2).$$

<u>THEOREM 1.6</u> In order that $\left(\underline{\mu},\frac{\Sigma}{\tau(\Sigma)}\right)$ be identifiable in $\mathcal{U}_{\varepsilon,r}$:

i) It is necessary, but not sufficient if $r < \infty$, that ε be less than ε_2^*.

ii) It is sufficient that ε be less than ε_1^*.

Also

iii) $\varepsilon_1^* \leqslant \varepsilon_2^* \leqslant \frac{1}{2}$, and $\varepsilon_1^* \to \frac{1}{2}$ as $r \to \infty$.

iv) If $f(r_0,0) + 1 - 2f(r_0,r_0) \geqslant 0$, where r_0 is the root of $4h_s(r_0^2) = h_s(0)$, then $\varepsilon_1^* = \varepsilon_3^*$ for all r.

v) If $w(\underline{x};\underline{\mu},\Sigma) = \phi(\underline{x};\underline{\mu},\Sigma)$, then

a) $$(1-\varepsilon_1^*)^{-1} = (1-\varepsilon_3^*)^{-1} = \begin{cases} X_m^2(r^2) + X_m^2(r_1^2) + X_m^2(r_1^2);\ m > 2, \text{ or } m = 2 \text{ and } r^2 > 2, \\ X_2^2(r^2) + \frac{r^2}{4};\ m = 2,\ r^2 \leq 2, \end{cases}$$

where $r_1 \in (0,r)$ is the sole positive root of

$r_1^m + mr_1^{m-2} - r^m = 0.$

b) For fixed $\varepsilon_1^* \geq 0$, the minimum permissible value of r^2 as $m \to \infty$ is asymptotically equal to $m + \sqrt{2m}\ \Phi^{-1}((2(1-\varepsilon_1^*))^{-1})$.

Proof: We first prove iii). That $\varepsilon_1^* \leq \varepsilon_2^*$ follows from $\varepsilon_3^* \leq \varepsilon_5^*$, which in turn follows from the definitions. Clearly, $\varepsilon_2^* \leq 1/2$. For sufficiently large r, $\varepsilon_1^* = \varepsilon_3^*$, since $\varepsilon_4^* \to 1$ as $r \to \infty$. Thus the second assertion of iii) will follow from $1 \geq \inf_{[0,r]} f(r,r_1) \to 1$ as $r \to \infty$. For this, put $x = \frac{r_1}{r}$, so that

$$\inf_{[0,r]} f(r,r_1) = \inf_{[0,1]} f(r,rx) = \inf_{[0,1]} \{H_m(r^2x^2) + \tfrac{1}{2}C_m r^m(1-x^m)h_s(r^2x^2)\}.$$

Suppose that the limit, as $r \to \infty$, of this last term is less than one. Then there exists $\delta > 0$, and sequences $\{r\}$, $\{x_r\}$; $r \to \infty$, such that

$$H_m(r^2x_r^2) + \tfrac{1}{2}C_m r^m(1-x_r^m)h_s(r^2x_r^2) < 1 - \delta \text{ for all } x_r. \tag{1}$$

Then $H_m(r^2x_r^2) < 1 - \delta$, so that $r^2x_r^2 < H_m^{-1}(1-\delta) = k^2$, say. But then

$$r^m(1-x_r^m)h_s(r^2x_r^2) \geq r^m(1-x_r^m)h_s(k^2)$$

$$> (r^m-k^m)h_s(k^2)$$

$$\to \infty \text{ as } r \to \infty,$$

contradicting (1).

Proof of ii): Let r be large enough that $\varepsilon_1^* > 0$, and let $0 < \varepsilon < \varepsilon_1^*$, so that $\varepsilon < \min(\varepsilon_3^*,\varepsilon_4^*,\varepsilon_5^*)$. By Lemma 1.2 iii), it will suffice to derive a contradiction from the supposition that $\left(\underline{0},\frac{I}{\tau(I)}\right)$ is not identifiable in $\mathcal{U}_{\varepsilon,r}$. Suppose then that there exist $u_2(\underline{x};\underline{\mu},\Sigma) \equiv u_1(\underline{x};\underline{0},I) \in \mathcal{U}_{\varepsilon,r}$, with $\left(\underline{\mu},\frac{\Sigma}{\tau(\Sigma)}\right) \neq \left(\underline{0},\frac{I}{\tau(I)}\right)$. For some densities $v_1(\underline{x};\underline{0},I)$ and $v_2(\underline{x};\underline{\mu},\Sigma)$ on D_r and E_r;

$$u_1(\underline{x};\underline{0},I) = (1-\varepsilon)w(\underline{x};\underline{0},I) + \varepsilon v_1(\underline{x};\underline{0},I) \text{ on } D_r,$$

and

$$u_2(\underline{x};\underline{\mu},\Sigma) = (1-\varepsilon)w(\underline{x};\underline{\mu},\Sigma) + \varepsilon v_2(\underline{x};\underline{\mu},\Sigma) \text{ on } E_r.$$

Denote by $u(\underline{x})$ the common value of $u_1(\underline{x})$ and $u_2(\underline{x})$.

We cannot have $D_r^0 \cap E_r^0 = \emptyset$, since then

$$\int_{R^m} u(\underline{x})d\underline{x} \geq (1-\varepsilon)\int_{D_r} w(\underline{x};\underline{0},I)d\underline{x} + (1-\varepsilon)\int_{E_r} w(\underline{x};\underline{\mu},\Sigma)d\underline{x}$$

$$= 2(1-\varepsilon)H_m(r^2) = \frac{1-\varepsilon}{1-\varepsilon_5^*} \geq \frac{1-\varepsilon}{1-\varepsilon_1^*} > 1.$$

Thus Corollary 1.5 applies. If $u(\underline{x})$ is constant throughout E_r

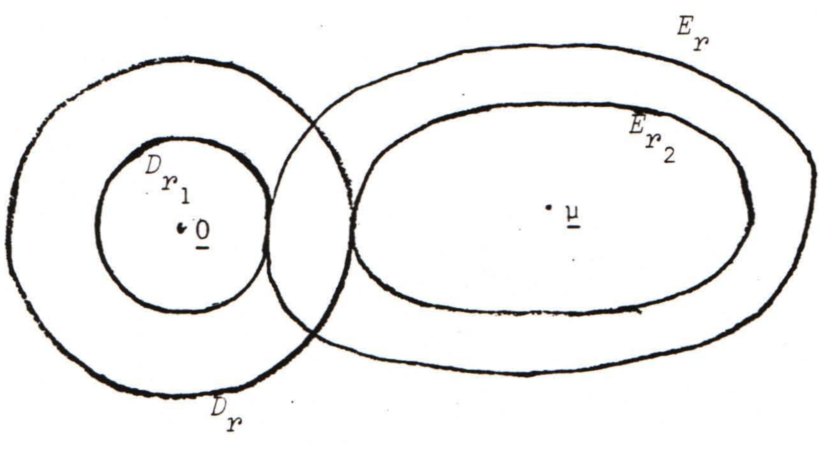
(or $E_r \setminus \{\underline{\mu}\}$), then this constant must exceed $(1-\varepsilon)w(\underline{\mu};\mu,\Sigma) = (1-\varepsilon)|\Sigma|^{-\frac{1}{2}}h_s(0)$. Thus $v_2(\underline{x};\mu,\Sigma) \geq \frac{1-\varepsilon}{\varepsilon}(|\Sigma|^{-\frac{1}{2}}h_s(0)-w(\underline{x};\mu,\Sigma))$, and so

$$\int_{E_r} v_2(\underline{x};\underline{\mu},\Sigma)\,d\underline{x} \geq \frac{1-\varepsilon}{\varepsilon}\left\{h_s(0)\int_{E_r}|\Sigma|^{-\frac{1}{2}}d\underline{x} - \int_{E_r} w(\underline{x};\underline{\mu},\Sigma)\,d\underline{x}\right\}$$

$$= \frac{1-\varepsilon}{\varepsilon}\left\{h_s(0)\mathrm{Vol}(D_r)-H_m(r^2)\right\}$$

$$= \frac{1-\varepsilon}{\varepsilon}\cdot\frac{\varepsilon_4^*}{1-\varepsilon_4^*} \geq \frac{1-\varepsilon}{\varepsilon}\cdot\frac{\varepsilon_1^*}{1-\varepsilon_1^*} > 1.$$

Similarly, $u(\underline{x})$ cannot be constant on $D_r \setminus \{\underline{0}\}$. If $u(\underline{x})$ is constant on $(D_r \setminus D_{r_1}^0) \cup (E_r \setminus E_{r_2}^0)$, then this constant exceeds $(1-\varepsilon)\max(h_s(r_1^2), |\Sigma|^{-\frac{1}{2}}h_s(r_2^2))$.



With $c = \max(h_s(r_1^2),\ |\Sigma|^{-\frac{1}{2}}h_s(r_2^2))$,

$$\int u \geq \int_{D^0_{r_1}} u + \int_{E^0_{r_2}} u + \int_{E_r \setminus E^0_{r_2}} u + \frac{1}{2}\int_{D_r \setminus D^0_{r_1}} u$$

$$\geq (1-\varepsilon)\left[\int_{D^0_{r_1}} w(\underline{x};\underline{0},I)d\underline{x} + \int_{E^0_{r_2}} w(\underline{x};\underline{\mu},\Sigma)d\underline{x} + c\int_{E_r \setminus E^0_{r_2}} 1\, d\underline{x} + \frac{c}{2}\int_{D_r \setminus D^0_{r_1}} 1\, d\underline{x}\right]$$

$$\geq (1-\varepsilon)\left[H_m(r_1^2)+H_m(r_2^2)+h_s(r_2^2)\int_{E_r \setminus E^0_{r_2}} |\Sigma|^{-\frac{1}{2}}d\underline{x} + \frac{h_s(r_1^2)}{2}\int_{D_r \setminus D^0_{r_1}} 1\, d\underline{x}\right]$$

$$= (1-\varepsilon)\left[(H_m(r_2^2)+h_s(r_2^2)\mathrm{Vol}(D_r \setminus D^0_{r_2})) + (H_m(r_1^2)+\tfrac{1}{2}h_s(r_1^2)\mathrm{Vol}(D_r \setminus D^0_{r_1}))\right].$$

The first term in brackets above can be shown to be a non-increasing function of r_2, assuming only that $h_s(\cdot)$ is non-increasing, but not necessarily continuous. The second is $f(r,r_1)$. Thus

$$\int u \geq (1-\varepsilon)\left[H_m(r^2)+f(r,r_1)\right] \geq \frac{1-\varepsilon}{1-\varepsilon_3^*} \geq \frac{1-\varepsilon}{1-\varepsilon_1^*} > 1.$$

These contradictions to $\left(\underline{\mu},\frac{\Sigma}{\tau(\Sigma)}\right) \neq \left(\underline{0},\frac{I}{\tau(I)}\right)$ prove ii).

Proof of iv): That $\varepsilon_1^* = \varepsilon_3^*$ is equivalent to

$$f(r,r) + \inf_{[0,r]} f(r,r_1) \leq 1 + 2f(r,0) - f(r,r).$$ Since

$\inf_{[0,r]} f(r,r_1) \leq f(r,0)$, it is sufficient that $f(r,0) + 1 - 2f(r,r)$ exceed zero. But this latter function is minimized at r_0.

Proof of v): If $w(\underline{x};\underline{\mu},\Sigma) = \phi(\underline{x};\mu,\Sigma)$, then $h_s(z) = (2\pi)^{-m/2}e^{-z/2}$, so that $r_0^2 = 4\ln 2 \approx 2.8$. An integration by parts gives

$$f(r_0,0) + 1 - 2f(r_0,r_0) = 1 - 2X^2_{m+2}(r_0^2) \geq 0$$

for $m \geq 2$, so that $\varepsilon_1^* = \varepsilon_3^*$ for all r. From

$$f(r,r_1) = X^2_m(r_1^2) + \left\{m2^{\frac{m}{2}}\Gamma\left(\frac{m}{2}\right)\right\}^{-1}(r^m - r_1^m)e^{-r_1^2/2}$$

we get

$$f_2(r,r_1) = \left(m2^{\frac{m}{2}}\Gamma\left(\frac{m}{2}\right)\right)^{-1} r_1 e^{-r_1^2/2}(r_1^m + mr_1^{m-2} - r^m).$$

Thus for $m = 2$ and $r^2 \leq 2$, $\inf\limits_{[0,r]} f(r,r_1) = f(r,0) = \frac{r^2}{4}$. For $m > 2$, or $m = 2$ and $r^2 > 2$, the extreme values of $f(r,r_1)$ in $(0,r)$ are attained at zeroes of $p(r_1) = r_1^m + mr_1^{m-2} - r^m$. Since $p(0) < 0$, $p(r) > 0$, and $p'(r_1) > 0$; p has only one positive zero, $f(r,r_1)$ has only one positive critical point, and this point r_1 is in $(0,r)$. Since $f_2(r,0^+) < 0$ and $f_2(r,r) > 0$, this point provides a minimum of $f(r,r_1)$. At this point, $r^m - r_1^m = mr_1^{m-2}$, so that

$$f(r,r_1) = X^2_m(r_1^2) + \left(2^{\frac{m}{2}}\Gamma\left(\frac{m}{2}\right)\right)^{-1} r_1^{m-2}e^{-r_1^2/2} = X^2_m(r_1^2) + X^2_m(r_1^2).$$

This proves a).

For b), we use the fact that a χ^2_m $r.v$ is asymptotically normally distributed, with mean m and variance $2m$. From this, it follows that the minimum permissible value of r^2 for fixed $\varepsilon_2^* = \varepsilon_5^*$ is asymptotically equal to $m + \sqrt{2m}\Phi^{-1}((2(1-\varepsilon_2^*))^{-1})$. But if $r^2 \gtrsim m$ as $m \to \infty$, then $p(x) < 0$ for $\frac{x}{r} < 2^{-1/m-2}$. Thus the zero r_1 of p satisfies $\frac{r_1}{r} \geq 2^{-1/m-2}$, and so $r_1 \sim r$ as $m \to \infty$. But this implies that $\varepsilon_1^* \sim \varepsilon_2^*$ as $m,r \to \infty$ in the manner described. This proves b). □

The case $r = \infty$ is sufficiently interesting that we state the simpler result for $\mathcal{U}_{\varepsilon,\infty}$ as a corollary.

<u>COROLLARY 1.7</u> The pair $(\underline{\mu},\Sigma)$ is not identifiable in $\mathcal{U}_{\varepsilon,\infty}$ if $w(\underline{x},\underline{0},I)$ is continuous. The pair $\left(\underline{\mu},\frac{\Sigma}{\tau(\Sigma)}\right)$ is identifiable iff $\varepsilon < 1/2$. □

Remarks:

1) It is easy to construct an example - $m = 2$, $w(\underline{x};\underline{0},I) = \pi^{-1}\cos(\underline{x}'\underline{x})$, $\underline{x}'\underline{x} \leq r^2 \leq \frac{\pi}{2}$ - in which ε_4^* may be smaller than both ε_3^* and ε_5^*, or lie between them, depending upon the value of r.

2) For the normal case $w(\underline{x};\underline{\mu},\Sigma) = \phi(\underline{x};\underline{\mu},\Sigma)$, the asymptotic values given in iii) and v) of Theorem 1.6 are approached rapidly. Table I below 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. $\int_{E_r}(1-\varepsilon_1^*)\phi(\underline{x};\underline{\mu},\Sigma)d\underline{x}$. Table II gives the maximum permissible proportion ε_1^* of symmetric contamination for specified r and m.

TABLE I

ε_1^* \ m	2	3	4	5	10	20	30	$m\to\infty$
0	1.306 (.574)	1.614 (.543)	1.892 (.534)	2.137 (.529)	3.089 (.518)	4.420 (.513)	5.434 (.510)	$m^{\frac{1}{2}}$ (.500)
.01	1.314 (.572)	1.622 (.542)	1.900 (.533)	2.145 (.528)	3.098 (.518)	4.429 (.513)	5.443 (.510)	$(m+.018\sqrt{m})^{\frac{1}{2}}$ (.500)
.05	1.349 (.568)	1.659 (.540)	1.937 (.531)	2.183 (.527)	3.136 (.518)	4.466 (.512)	5.481 (.510)	$(m+.093\sqrt{m})^{\frac{1}{2}}$ (.500)
.10	1.397 (.561)	1.710 (.537)	1.989 (.529)	2.234 (.525)	3.188 (.516)	4.519 (.511)	5.533 (.509)	$(m+.198\sqrt{m})^{\frac{1}{2}}$ (.500)

TABLE II

m \ r	1.5	2	3	4	5	6	∞
2	.190	.405	.493	.500	.500	.500	.500
3	-	.294	.483	.499	.500	.500	.500
4	-	.110	.463	.498	.500	.500	.500
5	-	-	.432	.496	.500	.500	.500
10	-	-	-	.303	.497	.500	.500
20	-	-	-	-	.367	.492	.500
30	-	-	-	-	-	.363	.500

2. CONSISTENT ESTIMATORS OF $\underline{\mu}$ AND $\frac{\Sigma}{\tau(\Sigma)}$.

In Section 3, we will define estimators of $\left(\underline{\mu},\frac{\Sigma}{\tau(\Sigma)}\right)$ as solutions to certain non-linear equations of the form $\sum_{i=1}^{n} \underline{\psi}\left(\underline{x}_i;\hat{\underline{\mu}},\frac{\hat{\Sigma}}{\tau(\hat{\Sigma})}\right) = \underline{0}$. In order that, asymptotically, the asymmetrically distributed observations should have no influence on the estimators, the functions $\underline{\psi}$ will be chosen to vanish off of $E_r\left(\hat{\underline{\mu}},\frac{\hat{\Sigma}}{\tau(\hat{\Sigma})}\right)$. As a result, the solutions are not unique if $r < \infty$ - any pair $\left(\hat{\underline{\mu}},\frac{\hat{\Sigma}}{\tau(\hat{\Sigma})}\right)$ which is such that the associated ellipse excludes all of the observations will constitute a "solution". However, if the starting value of the iterative process used to solve the equations is itself a consistent estimator of the parameters, and if the true parameters are asymptotic solutions, then the solution to which the process converges will be consistent. Choices of $\underline{\psi}$ can then be made to minimize the maximum asymptotic variance of the estimators, over the class of symmetric contaminating distributions.

In this section we construct an initial estimator of $\left(\underline{\mu},\frac{\Sigma}{\tau(\Sigma)}\right)$ and show that it is consistent. Before beginning the construction, we require some elementary facts about truncated spherical distributions.

If $\underline{y} \sim w(\underline{y};\underline{0},I)$, then the conditional density of $\underline{x} = \Sigma^{\frac{1}{2}}\underline{y} + \underline{\mu}$, given that $\underline{x} \in E_r(\underline{\mu},\Sigma)$, is

$$w(\underline{x}|E_r) = \begin{cases} \dfrac{w(x;\mu,\Sigma)}{H_m(r^2)}, & \underline{x} \in E_r \\ 0\ , & \underline{x} \notin E_r. \end{cases}$$

We say that $\underline{x} \sim w(\underline{x}|E_r)$ if $\underline{x} \in E_r$ and has conditional density $w(\underline{x}|E_r)$, and that $\underline{x} \sim \phi(\underline{x}|E_r)$ if $w(\underline{x};\underline{\mu},\Sigma) = \phi(\underline{x};\underline{\mu},\Sigma)$.

LEMMA 2.1 If $\underline{x} \sim w(\underline{x}|E_r)$, then

i) $\underline{y} = \Sigma^{-\frac{1}{2}}(\underline{x}-\underline{\mu}) \sim w(\underline{y}|D_r)$

ii) $E[\underline{x}] = \underline{\mu}$, $E[(\underline{x}-\underline{\mu})(\underline{x}-\underline{\mu})'] = a\Sigma$, where

$$a = \int_0^{r^2} \frac{zh_m(z)}{mH_m(r^2)}\, dz.$$

If the columns of $X_{m\times n} = \|\underline{x}_1 \ \ldots \ \underline{x}_n\|$ are independently distributed as $w(\underline{x}|E_r)$, if $\bar{\underline{x}} = \dfrac{X\underline{1}}{n}$, $V_1 = \dfrac{XX' - n\bar{\underline{x}}\bar{\underline{x}}'}{n-1}$, and $V = a^{-1}V_1$, then

iii) $E[\bar{\underline{x}}] = \underline{\mu}$, $E[V] = \Sigma$.

If $\underline{x} \sim \phi(\underline{x}|E_r)$ then $a = \chi^2_{m+2}(r^2) \ / \ \chi^2_m(r^2)$. □

A rough description of the method of estimation is as follows. Given a sample of size n from $u(\underline{x};\mu,\Sigma) \in \mathcal{U}_{\varepsilon,r}$, we form all subsets of size l_n. The sequence $\{l_n\}$ grows sufficiently slowly that at least some of the subsets will consist entirely of observations from $w(\underline{x}|E_r)$, with arbitrarily high probability, as $n \to \infty$. Each subset is randomly partitioned into two parts. From one part we form the mean vector $\bar{\underline{x}}$ and covariance matrix V, adjusted for unbiasedness at the ideal distribution, as in Lemma 2.1. Then if $\underline{x}$ is a member of the other part, it is independent of $(\bar{\underline{x}}, V)$.

If $\underline{x}^n$, $\bar{\underline{x}}^n$, V^n are members of, or estimates formed from, subsets all of whose members are from $w(\underline{x}|E_r)$, then $(\bar{\underline{x}}^n, V^n)$ converges in probability to $(\underline{\mu}, \Sigma)$, and $\underline{y}^n = (V^n)^{-\frac{1}{2}}(\underline{x}^n - \bar{\underline{x}}^n)$ converges weakly to $\underline{y} \sim w(\underline{y}|D_r)$, as $n \to \infty$. This implies that the corresponding empirical distribution functions (e.d.f.s) converge uniformly to the d.f. $W(\underline{y}|D_r)$ of $\underline{y}$. Thus the minimum distance, with respect to the sup norm, between $W(\underline{y}|D_r)$ and the e.d.f.s. of the $\binom{n}{l_n}$ sub-samples $\{\underline{y}_i^n\}$, tends to zero. The estimators of $\underline{\mu}$ and $\frac{\Sigma}{\tau(\Sigma)}$ are the $\bar{\underline{x}}$ and $\frac{V}{\tau(V)}$ corresponding to the e.d.f. which minimizes this distance. In Theorem 2.3 we show that these estimators are consistent.

We shall require the following version of the Glivenko-Cantelli Theorem.

<u>LEMMA 2.2</u> If

i) $W(\underline{x})$ is an absolutely continuous d.f. on R^m,

ii) $\{G_n(\underline{x})\}$ is a sequence of d.f.s converging pointwise to $W(\underline{x})$,

iii) $\{m_n\}$ is a sequence of positive integers satisfying $\sum_n m_n^{-2} < \infty$,

iv) $\{F_n(\underline{x})\}$ is a sequence of e.d.f.s of independent samples, of size m_n, from G_n; then

$$\sup_{C \in \mathcal{C}} |F_n(C) - W(C)| \xrightarrow{a.e.} 0,$$

where $\mathcal{C}$ is the set of measureable convex subsets of R^m.

Proof: Consider $F_n(\underline{x})$ as the e.d.f. based on m_n independent realizations $\underline{X}_1(\omega),\ldots,\underline{X}_{m_n}(\omega)$; where the $\underline{X}_i$ are r.vec.s on a probability space $(\Omega,\mathcal{F},P)$, with d.f. G_n. Let C be a convex, measureable subset of R^m, and define m_n independent r.v.s $\xi_j(C,\omega)$ by

$$\xi_j(C,\omega) = \begin{cases} 1 & \text{if } \underline{X}_j(\omega) \in C, \\ 0 & \text{otherwise.} \end{cases}$$

Then $m_n F_n(C,\omega) = \sum_{j=1}^{m_n} \xi_j(C,\omega) \sim b(m_n,\delta_n)$, where $\delta_n = G_n(C) \to \delta = W(C)$.

Let $\varepsilon > 0$ be arbitrary, and let N be large enough that $|\delta_n - \delta| < \varepsilon/2$ for $n \geq N$. Then

$$\lim_{N\to\infty} P(|F_n(C,\omega)-\delta| > \varepsilon \text{ for some } n \geq N)$$

$$\leq \lim_{N\to\infty} P(|F_n(C,\omega)-\delta_n| > \varepsilon/2) \text{ for some } n \geq N)$$

$$\leq \lim_{N\to\infty} \sum_{n=N}^{\infty} P(|F_n(C,\omega)-\delta_n| > \varepsilon/2)$$

$$\leq \lim_{N\to\infty} \sum_{n=N}^{\infty} \frac{E[(F_n(C,\cdot)-\delta_n)^4]}{(\varepsilon/2)^4}$$

$$= \frac{16}{\varepsilon^4} \lim_{N\to\infty} \sum_{n=N}^{\infty} \left[\frac{3m_n^2\delta_n^2(1-\delta_n)^2 + m_n\delta_n(1-\delta_n)(1-6\delta_n(1-\delta_n))}{m_n^4} \right]$$

$$= 0,$$

by iii). The third inequality is Chebyshev's Inequality. We have shown that $F_n(C,\omega) \xrightarrow{a.e.} \delta = W(C)$, i.e. for *a.e.* ω, $F_n(\cdot,\omega) \xrightarrow{D} W(\cdot)$. The uniformity then follows from Theorem 4.2 of R.R. Rao, [10]. □

Now let $\underline{x}_1,\ldots,\underline{x}_n$ be a random sample from $u \in U_{\varepsilon,r}$, with $\varepsilon < \varepsilon_1^*$. For any such u,

$$p = P(\underline{x}_i \sim w(\underline{x} \mid E_r)) = \int_{E_r} (1-\varepsilon)w(\underline{x};\underline{\mu},\Sigma)d\underline{x}$$

$$= (1-\varepsilon)H_m(r^2) = \frac{(1-\varepsilon)}{2(1-\varepsilon_5^*)} \geq \frac{(1-\varepsilon)}{2(1-\varepsilon_1^*)} > \frac{1}{2} .$$

Define the r.v. L_n by L_n = "Number of members of the sample from $w(\underline{x}|E_r)$". Then $L_n \sim b(n,p)$. Let $\{l_n\}$ be a sequence of positive integers satisfying

i) $\frac{l_n}{n} \uparrow p$ as $n \to \infty$ (implying $\sum_n l_n^{-2} < \infty$),

ii) $\sum_n n^2(l_n - np)^{-4} < \infty$;

e.g. $l_n = [np - n^\delta p]$, where $3/4 < \delta < 1$. Put $a_n = \binom{n}{l_n}$, and let $\{T_1^n,\ldots,T_{a_n}^n\}$ be the set of all subsets of $\{\underline{x}_1,\ldots,\underline{x}_n\}$ of size l_n. Let $m_n = l_n - o(n)$. Then $\frac{m_n}{n} \to p$ and $\sum m_n^{-2} < \infty$. For each i, randomly partition T_i^n as $D_i^n \cup E_i^n$, where $|D_i^n| = m_n$ and $|E_i^n| = l_n - m_n$. From the elements of E_i^n, calculate the mean vector $\underline{\bar{x}}_i^n$ and adjusted covariance matrix V_i^n. Note that the set $T_i^n = \{ <\underline{\bar{x}}_i^n, V_i^n>, \underline{x}_{ij}^n \mid \underline{x}_{ij}^n \in D_i^n, 1 \leq j \leq m_n\}$ is totally independent, and that its distribution does not depend upon i.

Let $n \to \infty$, and consider the infinite array, with independence between the rows,

$$\begin{array}{c} \vdots \\ T_1^n, \ldots T_{a_n}^n \\ T_1^{n+1}, \ldots, T_{a_{n+1}}^{n+1} \\ \vdots \end{array}$$

Put $B_n = \{1,2,\ldots,a_n\}$, $B = \overset{\infty}{\underset{n=1}{X}} B_n$. For $\underline{b} \in B$, i.e. for $\underline{b} = \langle b_1,\ldots,b_n,\ldots\rangle$, where $1 \leq b_n \leq a_n$, define $T_{\underline{b}} = \langle T_{b_1}^1,\ldots,T_{b_n}^n,\ldots\rangle$. Then $\{T_{\underline{b}} | \underline{b} \in B\}$ is the set of all sequences through the array, with one set chosen from each row. We claim that

$$1 = \lim_{n\to\infty} P \text{ (There exists a sequence } T_{\underline{\beta}} \text{ with the property that for all } k \geq n,\ T_{\beta_k}^k \text{ consists entirely of vectors from } w(\underline{x}|E_r)\text{).}$$

This claim is equivalent to

$$0 = \lim_{n\to\infty} P \text{ (There exists a row } k \text{ in the array, with } k \geq n, \text{ no member of which contains only vectors from } w(\underline{x}|E_r))$$

$$= \lim_{n\to\infty} \sum_{k=n}^{\infty} P(L_k < l_k).$$

This follows easily from ii) and Chebyshev's Inequality.

If $T_{\underline{\beta}}$ is any such sequence, then the corresponding sequence $\{\underline{\bar{x}}_{\beta_n}^n, V_{\beta_n}^n\}$ converges in probability to $(\underline{\mu},\Sigma)$, so that if $\{\underline{x}_{\beta_n}^n\}$ is a sequence of independent r.vec.s, all of which come from $w(\underline{x}|E_r)$ from some point onwards, then

$$(V^n_{\beta_n})^{-\frac{1}{2}}(\underline{x}^n_{\beta_n} - \bar{\underline{x}}^n_{\beta_n}) \xrightarrow{D} \underline{y} \sim w(\underline{y}|D_r). \qquad (1)$$

In particular, if $\underline{x}^n_{\beta_n} \in D^n_{\beta_n}$, then (1) holds. Now form "standardized" set $S^n_{i\cdot}$ from the $T^n_{i\cdot}$. Define $S^n_{i\cdot} = \left\{\underline{y}^n_{i,1},\ldots,\underline{y}^n_{i,m_n}\right\}$, $1 \le i \le a_n$, by

$$\underline{y}^n_{i,j} = \begin{cases} \underline{0} & \text{if } V^n_{i\cdot} \not> 0 \\ (V^n_{i\cdot})^{-\frac{1}{2}}(\underline{x}^n_{ij\cdot} - \bar{\underline{x}}^n_{i\cdot}) & \text{if } V^n_{i\cdot} > 0;\ \underline{x}^n_{ij\cdot} \in D^n_{i\cdot},\quad 1 \le j \le m_n. \end{cases}$$

The members of the $S^n_{i\cdot}$ are *i.i.d.*, and their distribution does not depend upon i. By (1),

$$S_{\underline{\beta}} \xrightarrow{D} \underline{y} \sim w(\underline{y}|D_r), \qquad (2)$$

in the sense that the sequence $<\underline{y}'_{\beta_1,1},\ldots,\underline{y}'_{\beta_1,m_1},\ldots,\underline{y}^n_{\beta_n,1},\ldots,\underline{y}^n_{\beta_n,m_n},\ldots>$ converges weakly to $\underline{y}$. Let G_n be the d.f. of $S^n_{i\cdot}$, and let $F^n_{i\cdot}$ be the e.d.f. Define

$$X_{n,i} = \sup_{C \in \mathcal{C}_1} |F_{n,i}(C) - W(C|D_r)|,$$

$$X_{n,\alpha} = \min_{i \in \beta_n} X_{n,i},$$

where $\mathcal{C}_1$ is any fixed subset of $\mathcal{C}$. By (2) and Lemma 2.2, $X_{n,\beta_n} \xrightarrow{\text{a.e.}} 0$, so that $X_{n,\alpha} \xrightarrow{a.e.} 0$. Let $\bar{\underline{x}}^n_\alpha$, V^n_α be the mean vector and covariance matrix corresponding to $X_{n,\alpha}$.

<u>THEOREM 2.3</u> If $\varepsilon < \varepsilon_1^*$, then

$$\left(\bar{\underline{x}}^n_\alpha, \frac{V^n_\alpha}{\tau(V^n_\alpha)}\right) \xrightarrow{P} \left(\underline{\mu}, \frac{\Sigma}{\tau(\Sigma)}\right).$$

<u>Proof</u>: We first show that

$$S^n_\alpha \xrightarrow{D} \underline{y} \sim w(\underline{y}|D_r). \tag{3}$$

Let $G_{n,\alpha}$ be the d.f. of the members of S^n_α. For fixed $\underline{y}$, put $G_{n,\alpha}(\underline{y}) = \delta_n$, $W(\underline{y}|D_r) = \delta$. Then (3) will follow from

$$\delta_n \longrightarrow \delta. \tag{4}$$

But $X_{n,\alpha} \xrightarrow{a.e.} 0$ implies that $F_{n,\alpha}(\underline{x}) \xrightarrow{a.e.} \delta$. Since $m_n F_{n,\alpha}(\underline{x}) \sim b(m_n, \delta_n)$ the second moments of the $F_{n,\alpha}(\underline{x})$ are uniformly bounded in n. In the presence of this condition, convergence a.e. implies convergence in L_1: $E[F_{n,\alpha}(\underline{x})] \to \delta$, which is (4).

By the *WLLN*, $\bar{\underline{x}}^n_\alpha$ and V^n_α converge in probability to their expectations, say $\underline{\mu}_\alpha$ and Σ_α, which exist by assumption (model assumption M1)). Since $\tau(\cdot)$ is continuous, $\left(\bar{\underline{x}}^n_\alpha, \frac{V^n_\alpha}{\tau(V^n_\alpha)}\right) \xrightarrow{P} \left(\underline{\mu}_\alpha, \frac{\Sigma_\alpha}{\tau(\Sigma_\alpha)}\right)$.

The result will then follow from

$$\left(\underline{\mu}_\alpha, \frac{\Sigma_\alpha}{\tau(\Sigma_\alpha)}\right) = \left(\underline{\mu}, \frac{\Sigma}{\tau(\Sigma)}\right). \qquad (5)$$

Recall that D_α^n and S_α^n are related by $\underline{x}_{\alpha j}^n = (V_\alpha^n)^{\frac{1}{2}} \underline{y}_{\alpha j}^n + \underline{\bar{x}}_\alpha^n$, for $\underline{x}_{\alpha j}^n \in D_\alpha^n$, $\underline{y}_{\alpha j}^n \in S_\alpha^n$. Since $\underline{y}_{\alpha j}^n \xrightarrow{D} \underline{y}$ by (3), and $(\underline{\bar{x}}_\alpha^n, V_\alpha^n) \xrightarrow{P} (\underline{\mu}_\alpha, \Sigma_\alpha)$; $D_\alpha^n \xrightarrow{D} \underline{x}_\alpha \sim w(\underline{x}_\alpha | E_r(\underline{\mu}_\alpha, \Sigma_\alpha))$. For a randomly chosen $\underline{x} \in \{\underline{x}_1, \ldots, \underline{x}_n\}$,

$$\lim_{n\to\infty} P(\underline{x} \in D_\alpha^n) = \lim_{n\to\infty} \frac{m_n}{n} = p,$$

$$\lim_{n\to\infty} P(\underline{x} \in E_\alpha^n) = \lim_{n\to\infty} \frac{l_n - m_n}{n} = 0,$$

$$\text{and } \lim_{n\to\infty} P(\underline{x} \notin T_\alpha^n) = 1-p.$$

Thus the limiting distribution of $\underline{x}$ is simultaneously $U(\underline{x};\underline{\mu},\Sigma)$ and

$$pW(\underline{x}_\alpha | E_r(\underline{\mu}_\alpha, \Sigma_\alpha)) + (1-p) G_\alpha(\underline{x}), \qquad (6)$$

for some $G_\alpha(\underline{x})$ *whose restriction to* $E_r(\underline{\mu},\Sigma)$ *is symmetric.* Furthermore, G_α must be differentiable. W.l.o.g., we now assume that $(\underline{\mu},\Sigma) = (\underline{0},I)$, and write D_r for $D_r(\underline{0},I)$, E_r for $E_r(\underline{\mu}_\alpha,\Sigma_\alpha)$. Equating $u(\underline{x};\underline{0},I)$ to the derivative of the term at (6), and inserting the definitions of p and $w(\cdot|E_r)$, gives

$$u(\underline{x};\underline{0},I) = ((1-\varepsilon) w(\underline{x};\underline{\mu}_\alpha,\Sigma_\alpha)) 1_{E_r} + (1-p) g_\alpha(\underline{x})$$

for some density g_α which is symmetric within D_r. But by definition,

$$u(\underline{x};\underline{0},I) = ((1-\varepsilon) w(\underline{x};\underline{0},I) + \varepsilon v(\underline{x};\underline{0},I)) 1_{D_r} + \eta(\underline{x}) 1_{D_r^c}$$

for some symmetric v and arbitrary η. We cannot have $D_r \cap E_r = \emptyset$, else $\int u \geq \int_{D_r} u + \int_{E_r} u > 1$. Thus Theorem 1.4 applies. If $\left(\underline{\mu}_\alpha, \frac{\Sigma_\alpha}{\tau(\Sigma_\alpha)}\right) \neq \left(\underline{0}, \frac{I}{\tau(I)}\right)$, and if $E_r \subset D_r$, or $D_r \subset E_r$, then the required constancy of u on the inner region violates $\varepsilon < \varepsilon_1^* \leq \varepsilon_4^*$. Constancy of u on an outer annulus of D_r violates assumption M3). This leaves only $\left(\underline{\mu}_\alpha, \frac{\Sigma_\alpha}{\tau(\Sigma_\alpha)}\right) = \left(\underline{0}, \frac{I}{\tau(I)}\right)$, which is (5). □

3. CONSISTENT, ASYMPTOTICALLY NORMALLY DISTRIBUTED M-ESTIMATORS

In this section we derive M-estimators of $\underline{\mu}$ and that multiple of Σ which has $\tau(\Sigma) = 1$. We assume that the observations come from $u(\underline{x};\underline{\mu},\Sigma) \in U'_{\varepsilon,r} = \{u \in U_{\varepsilon,r} | \tau(\Sigma)=1\}$. This is equivalent to assuming that the observations come from a general $u \in U_{\varepsilon,r_0}$, but that the region of known symmetry is $\{\underline{x} | (\underline{x}-\underline{\mu})'\Sigma^{-1}(\underline{x}-\underline{\mu}) \leq r_0^2 = r^2/\tau(\Sigma)\}$. This is a serious drawback to the model, if $r < \infty$, but is forced by the identifiability problem. Even if, as is assumed, r is known exactly, the user must still substitute an estimate of $\tau(\Sigma)$ in $r^2/\tau(\Sigma)$. An incorrect estimate yields an estimator of Σ which either excludes symmetrically distributed observations, hence has reduced efficiency; or includes possibly asymmetrically distributed observations. If $r = \infty$ then $\tau(\Sigma)$ need not be known.

The parameter spaces are $\Theta_1 = R^m$, $\Theta_2 = \{V_{m\times m} | V>0\}$, $\Theta = \Theta_1 \times \Theta_2$. The members of Θ_2 will be thought of variously as matrices, or as the vectors in $R^{m(m+1)/2}$ consisting of the functionally independent elements of these matrices. Put $\theta_0 = (\underline{\mu},\Sigma)$, $\theta_1 = \left(\underline{\bar{x}}^n_\alpha, \frac{V^n_\alpha}{\tau(V^n_\alpha)}\right)$, from Theorem 2.3.

We will derive a system of equations, and define an estimator $\hat{\theta}$ as the Newton-Raphson solution to these equations, with θ_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. From this, asymptotic confidence regions for linear functions of θ_0 are constructed. We will also show that $\hat{\theta}$ may be obtained as the limit of a simple fixed point iteration process which, asymptotically, has the same optimal convergence properties as the Newton-Raphson method.

In the case $r = \infty$, it will be shown that under fairly mild restrictions, the solution to the equations is asymptotically unique, and so the construction of θ_1 is "asymptotically unnecessary."

Throughout this section, $\Sigma^{\frac{1}{2}}$ is assumed to be upper triangular with positive diagonal elements.

Derivation of the equations

If $\underline{x} \sim u(\underline{x};\theta_0) \in U'_{\varepsilon,r}$, then θ_0 is characterized uniquely by

i) $E[a_0((\underline{x}-\underline{\mu})'\Sigma^{-1}(\underline{x}-\underline{\mu}))(\underline{x}-\underline{\mu})] = \underline{0}$ for all functions $a_0(\cdot)$ vanishing off of $[0,r^2]$;

ii) $E[a_2((\underline{x}-\underline{\mu})\Sigma^{-1}(\underline{x}-\underline{\mu}))\Sigma^{-\frac{1}{2}}(\underline{x}-\underline{\mu})(\underline{x}-\underline{\mu})'\Sigma^{-\frac{1}{2}'}] = kI_m$ for all $a_2(\cdot)$ vanishing off of $[0,r^2]$ and $k = k(a_2,v)$;

iii) $\tau(\Sigma) = 1$.

Pre- and post-multiplying by $\Sigma^{\frac{1}{2}}$ and $\Sigma^{\frac{1}{2}'}$ in ii), then applying τ to both sides (recalling that τ commutes with the expectation operator), and using iii) gives $k = E[\tau(a_2(\cdot)(\underline{x}-\underline{\mu})(\underline{x}-\underline{\mu})')]$. It follows that ii) and iii) are together equivalent to

iv) $E[a_2(\cdot)\{\Sigma^{-\frac{1}{2}}(\underline{x}-\underline{\mu})(\underline{x}-\underline{\mu})'\Sigma^{-\frac{1}{2}'} - \tau((\underline{x}-\underline{\mu})(\underline{x}-\underline{\mu})')I\}] = 0_{m\times m}$.

Let $a_0(z)$, $a_2(z)$ be continuous, piecewise smooth functions vanishing off of $[0,r^2]$. For any $\theta = (\underline{t},V) \in \Theta$ and sample values $\underline{x}_1,\ldots,\underline{x}_n$; define

$$\underline{a} = (\ldots, a_0((\underline{x}_i-\underline{t})'V^{-1}(\underline{x}_i-\underline{t})),\ldots)' : n \times 1, \tag{1}$$

$$A = \mathrm{diag}(\ldots, a_2((\underline{x}_i-\underline{t})'V^{-1}(\underline{x}_i-\underline{t})),\ldots) : n \times n \tag{2}$$

$$X = \|\underline{x}_1 \ldots \underline{x}_n\| : n \times n \qquad T = \|\underline{t},\ldots,\underline{t}\| : n \times n. \tag{3}$$

In analogy with i) and iv), we seek the appropriate zero of $F_n\colon \Theta \to R^{m(m+3)/2}$ defined by

$$F_n(\theta) = n^{-1}((X-T)\underline{a}, [V^{-\frac{1}{2}}(X-T)A(X-T)'V^{-\frac{1}{2}'} - \tau((X-T)A(X-T)')I]).$$

Of course, the zeroes of F_n coincide with those of

$$G_n(\theta) = n^{-1}((X-T)\underline{a}, [(X-T)A(X-T)' - \tau((X-T)A(X-T)')V]). \tag{4}$$

The first function is better suited to the calculations which follow. The second is the one to which the fixed point process referred to above is applied.

The estimator of θ_0 is

$$\hat{\theta} = \begin{cases} \text{The Newton-Raphson solution } \theta^* \text{ to } F_n(\theta) = \underline{0}, \\ \text{starting with } \theta_1, \text{ if the iteration process} \\ \text{converges;} \\ \\ \theta_1, \text{ otherwise.} \end{cases} \tag{5}$$

Consistency of $\hat{\theta}$

Put $\underline{y}_i = V^{-\frac{1}{2}}(\underline{x}_i-\underline{t})$, and define

$$\underline{\psi}_0(\underline{x}_i;\theta) = a_0(\underline{y}_i'\underline{y}_i)V^{\frac{1}{2}}\underline{y}_i,$$

$$\underline{\psi}_2(\underline{x}_i;\theta) = a_2(\underline{y}_i'\underline{y}_i)[\underline{y}_i\underline{y}_i'-\tau(V^{\frac{1}{2}}\underline{y}_i\underline{y}_i'V^{\frac{1}{2}'})I],$$

$$\underline{\psi}(\underline{x}_i;\theta) = (\underline{\psi}_0(\underline{x}_i;\theta),\underline{\psi}_2(\underline{x}_i;\theta)).$$

Then $F_n(\theta) = n^{-1}\sum_i \underline{\psi}(\underline{x}_i;\theta) \xrightarrow{P} E[\underline{\psi}(\underline{x};\theta)]$, and so $F_n(\theta_0) \xrightarrow{P} \underline{0}$ by i) and iv) above. We show that

C1) $\lim_{n\to\infty} P(\hat{\theta}=\theta^*) = 1$,

C2) $\hat{\theta} \xrightarrow{P} \theta_0$,

C3) $nF_n(\hat{\theta}) = \sum_i \underline{\psi}(\underline{x}_i;\hat{\theta}) \xrightarrow{P} \underline{0}$.

Of course, C1) implies C3). Since a_0, a_2 are piecewise smooth, we have

i) Around each $\theta \in \Theta$ is an open neighborhood within which $\left(\frac{\partial F_n}{\partial \theta}\right)$ exists, and is continuous, with probability one.

In Lemma 3.4 below, we give conditions ensuring that

ii) $\lim_{n\to\infty} \left(\frac{\partial F_n}{\partial\theta}\right)_{\theta_0} = E\left[\left(\frac{\partial\underline{\psi}}{\partial\theta}\right)_{\theta_0}\right]$ is non-singular.

Points i) and ii), together with $F_n(\theta_0) \xrightarrow{P} \underline{0}$, imply that θ_0 is asymptotically a point of attraction of the iteration process:

iii) There exists an open neighborhood S of θ_0 such that starting anywhere in S the iterates remain in Θ and converge to θ_0.

Then iii) and the consistency of θ_1 establish C1) and C2).

Asymptotic Normality of $\hat{\theta}$

Put $p = \frac{m(m+1)}{2}$, and denote by $f_1,\ldots,f_{m+p}$ the components of F_n. By the mean value theorem, there exist $t_1,\ldots,t_{m+p} \in (0,1)$ such that

$$\sqrt{n}(F_n(\hat{\theta})-F_n(\theta_0)) = B(\hat{\theta},\theta_0)(\sqrt{n}(\hat{\theta}-\theta_0)), \qquad (6)$$

where

$$B(\hat{\theta},\theta_0) = \|\ldots,(\text{grad } f_i(\theta_0+t_i(\hat{\theta}-\theta_0))',\ldots\|' \;:\; (m+p) \times (m+p)$$

$$\xrightarrow{P} E\left[\left(\frac{\partial \underline{\psi}}{\partial \theta}\right)_{\theta_0}\right].$$

The first term on the left of (6) tends to $\underline{0}$ in probability by C3), and the second is

$$n^{-\frac{1}{2}} \sum_i \underline{\psi}(\underline{x}_i;\theta_0) \xrightarrow{D} N_{m+p}(\underline{0},\text{cov}[\underline{\psi}]_{\theta_0}).$$

Thus the right-hand-side of (6) is asymptotically normally distributed as well, and so

$$\sqrt{n}(\hat{\theta}-\theta_0) \xrightarrow{D} N_{m+p}\left(\underline{0}, \left(E\left[\left(\frac{\partial \underline{\psi}}{\partial \theta}\right)_{\theta_0}\right]\right)^{-1} (\text{cov}[\underline{\psi}]_{\theta_0}) \left(E\left[\left(\frac{\partial \underline{\psi}}{\partial \theta}\right)_{\theta_0}\right]\right)^{-1'}\right).$$

The proof of Lemma 3.4 will then complete the proof of consistency and asymptotic normality of $\hat{\theta}$. This result is stated formally as Theorem 3.5 below, where the limiting covariance matrix of $\sqrt{n}\hat{\theta}$ is given explicitly. It is of rank one less than its order, due to the condition $\tau(\Sigma) = 1$. In Lemma 3.6, this singularity is removed.

We first introduce a system of zero-one matrices which will be used to determine $\text{cov}[\underline{\psi}]_{\theta_0}$ (Lemma 3.2), and to exhibit certain Jacobian matrices (Lemma 3.1) required in the calculation of $E\left[\left(\frac{\partial\underline{\psi}}{\partial\theta}\right)_{\theta_0}\right]$ (Lemma 3.3).

Matrix notation

For the purpose of calculating Jacobian matrices, we enumerate the elements of symmetric and upper triangular matrices row-by-row on and above the main diagonal; viz.,

$$V = \{v_{11},\ldots,v_{1m};v_{22},\ldots,v_{2m};\ldots;v_{mm}\}.$$

If $A_{m\times m}$, $B_{m\times m}$ and C_i are matrices, then

$$A \otimes B = (a_{ij}B) \;:\; m^2 \times m^2;$$

$$\bigoplus_{i=1}^{m} C_i = \text{diag}(C_1,\ldots,C_m), \text{ so that } I_m \otimes B = \bigoplus_{i=1}^{m} B;$$

$$\text{vec } A = (a_{11},\ldots,a_{1m};a_{21},\ldots,a_{2m};\ldots;a_{m1},\ldots,a_{mm})' \;:\; m^2 \times 1$$

$$\underset{s}{\text{vec}}\, A = (a_{11},\ldots,a_{1m};a_{22},\ldots,a_{2m};\ldots;a_{mm})' \;:\; \frac{m(m+1)}{2} \times 1.$$

Thus if A and B are symmetric or upper triangular, the $(i,j)^{\text{th}}$ element of $\left(\frac{\partial A}{\partial B}\right) \;:\; \frac{m(m+1)}{2} \times \frac{m(m+1)}{2}$ is the partial derivative of the i^{th} element of $\underset{s}{\text{vec}}\, A$ with respect to the j^{th} element of $\underset{s}{\text{vec}}\, B$.

Extensive use is made of the following zero-one matrices.

$$\underline{f}_i = (0,\ldots,0,\underset{\substack{\uparrow \\ i}}{1},0,\ldots,0)' : m \times 1; \qquad \underline{f} = \begin{pmatrix} \underline{f}_1 \\ \cdot \\ \cdot \\ \cdot \\ \underline{f}_m \end{pmatrix} = \text{vec } I: m^2 \times 1;$$

$$J_i = \|\underline{f}_i \cdots \underline{f}_m\| = \overset{m-i+1}{\begin{pmatrix} O \\ \hline I \end{pmatrix}} \begin{matrix} i-1 \\ m-i+1 \end{matrix} : m \times m-i+1;$$

$$J = \bigoplus_{i=1}^{m} J_i = \frac{\partial \text{ vec } A}{\partial \underset{s}{\text{vec}} A} : m^2 \times \frac{m(m+1)}{2};$$

$$\underline{e}_i = J_i' \underline{f}_i = (1,0,\ldots,0)' : m-i+1 \times 1; \quad \underline{e} = \begin{pmatrix} \underline{e}_1 \\ \cdot \\ \cdot \\ \cdot \\ \underline{e}_m \end{pmatrix} = \underset{s}{\text{vec}} \; I: \frac{m(m+1)}{2} \times 1.$$

Then $J_i \underline{e}_i = \underline{f}_i$ and $J\underline{e} = \underline{f}$.

$$F = (\underline{f}_j \underline{f}_i') = \frac{\partial \text{ vec } A'}{\partial \text{ vec } A} : m^2 \times m^2; \quad F_d = \bigoplus_{i=1}^{m} \underline{f}_i \underline{f}_i' \ .$$

The $(i,j)^{\text{th}}$ block of F has 1 in the $(j,i)^{\text{th}}$ spot, zeroes elsewhere.

$$P = \overset{m(m+1)/2}{\begin{pmatrix} P_1 \\ \hline P_2 \end{pmatrix}} \begin{matrix} m \\ m(m-1)/2 \end{matrix}, \text{ where } P_1 = \|\underline{f}_1 \underline{e}_1' | \cdots | \underline{f}_m \underline{e}_m'\|, \text{ is the permutation}$$

matrix defined by its action

$$P(\underset{s}{\text{vec}} \; A) = (a_{11}, a_{22}, \ldots, a_{mm}; a_{12}, a_{13}, \ldots, a_{1m}; a_{23}, \ldots, a_{2m}; \ldots; a_{m-1,m})'$$

for any A: $m \times m$. Thus, e.g. $P\underline{e} = \begin{pmatrix} \underline{1} \\ \underline{0} \end{pmatrix} \begin{matrix} m \\ m(m-1)/2 \end{matrix}$.

The most frequently employed properties of these matrices are:

Z1) $J'(\text{vec } A) = \text{vec}_s A$;

Z2) $F(\text{vec } A) = \text{vec } A'$;

Z3) $F(A \otimes B)F = B \otimes A$;

Z4) $\text{vec } A = (I \otimes A')\underline{f} = (A \otimes I)\underline{f}$, $\text{vec } AB' = (A \otimes B)\underline{f}$;

Z5) $\underline{f}'(\text{vec } A) = tr\ A$;

Z6) $FF_d = F_d$, $F\underline{f} = F_d\underline{f} = \underline{f}$;

Z7) $P_1'P_1 = I - P_2'P_2 = \bigoplus_{i=1}^{m} \underline{e}_i\underline{e}_i'$, the diagonal matrix with $\underline{e}$ as its diagonal.

LEMMA 3.1

i) If V is symmetric and non-singular, then

$$\left(\frac{\partial V^{-1}}{\partial V}\right) = -J'(V^{-1} \otimes V^{-1})(I+F-F_d)J, \qquad (7)$$

and so

$$\left(\frac{\partial V^{-1}}{\partial V}\right)^{-1} = \left(\frac{\partial V}{\partial V^{-1}}\right) = -J'(V \otimes V)(I+F-F_d)J.$$

ii) If W is upper triangular, then

$$\left(\frac{\partial W^{-1}}{\partial W}\right) = -J'(W^{-1} \otimes W^{-1'})J,$$

$$\left(\frac{\partial W'W}{\partial W}\right) = J'(I+F)(W' \otimes I)J,$$

$$\left(\frac{\partial WW'}{\partial W}\right) = J'(I+F)(I \otimes W)J.$$

iii) If $V > 0$, $V^{-\frac{1}{2}}$ is an upper triangular matrix satisfying

$$(V^{-\frac{1}{2}})'V^{-\frac{1}{2}} = V^{-1}, \text{ and } \Delta = \left(\frac{\partial V^{-\frac{1}{2}}}{\partial V}\right), \text{ then}$$

$$\Delta^{-1} = -J'(I+F)(V^{\frac{1}{2}}\otimes V)J.$$

iv) If A is upper triangular, then

$$\left(\frac{\partial A\underline{y}}{\partial A}\right) = \|\underline{f}_1\underline{y}'|\ldots|\underline{f}_m\underline{y}'\|J = (I_m\otimes\underline{y}')J.$$

v) $$\left(\frac{\partial \underline{y}\underline{y}'}{\partial \underline{y}}\right) = J'((\underline{y}\otimes I_m)+(I_m\otimes\underline{y})).$$

vi) $$\left(\frac{\partial\tau(\underline{x}\underline{x}')}{\partial\underline{x}}\right) = (\tau(\underline{f}_1\underline{x}'+\underline{x}\underline{f}_1'),\ldots,\tau(\underline{f}_m\underline{x}'+\underline{x}\underline{f}_m')).$$

Proof: We prove i), iii) and vi), the others are left to the reader.

i): For any parameter ϕ,

$$0 = \frac{\partial}{\partial\phi}I = \frac{\partial}{\partial\phi}VV^{-1} = \left(\frac{\partial V}{\partial\phi}\right)V^{-1} + V\left(\frac{\partial V^{-1}}{\partial\phi}\right),$$

so that

$$\left(\frac{\partial V^{-1}}{\partial\phi}\right) = -V^{-1}\left(\frac{\partial V}{\partial\phi}\right)V^{-1}.$$

Putting $\phi = v_{jj}$, then $\phi = v_{jk}$ $(j<k)$ gives

$$\left(\frac{\partial V^{-1}}{\partial v_{jj}}\right) = -V^{-1}(\underline{f}_j\underline{f}_j')V^{-1}, \quad \left(\frac{\partial V^{-1}}{\partial v_{jk}}\right) = -V^{-1}(\underline{f}_j\underline{f}_k'+\underline{f}_k\underline{f}_j')V^{-1}.$$

Thus, representing V^{-1} as (v^{il}),

$$\frac{\partial v^{il}}{\partial v_{jj}} = \left(\frac{\partial V^{-1}}{\partial v_{jj}}\right)_{i,l} = \underline{f}_i'\left(\frac{\partial V^{-1}}{\partial v_{jj}}\right)\underline{f}_l = -v^{ij}v^{lj},$$

$$\frac{\partial v^{il}}{\partial v_{jk}} = \left(\frac{\partial V^{-1}}{\partial v_{jk}}\right)_{i,l} = \underline{f}_i'\left(\frac{\partial V^{-1}}{\partial v_{jk}}\right)\underline{f}_l = -(v^{ij}v^{lk}+v^{ik}v^{lj}).$$

Upon expansion, this is seen to agree with (7).

iii): $\Delta^{-1} = \left(\frac{\partial V}{\partial V^{-\frac{1}{2}}}\right) = \left(\frac{\partial V}{\partial V^{-1}}\right)\left(\frac{\partial V^{-1}}{\partial V^{-\frac{1}{2}}}\right) = \left(\frac{\partial V}{\partial V^{-1}}\right)\left(\frac{\partial W'W}{\partial W}\right)_{W = V^{-\frac{1}{2}}}$

$= -J'(V\otimes V)(I+F-F_d)J\cdot J'(I+F)(V^{-\frac{1}{2}'}\otimes I)J$ by i) and ii),

$= -J'(V\otimes V)(I+F)(V^{-\frac{1}{2}'}\otimes I)J$, using $(I+F-F_d)JJ'(I+F) = I+F$,

$= -J'(I+F)(V\otimes V)(V^{-\frac{1}{2}'}\otimes I)J$ by Z3)

$= -J'(I+F)(V^{\frac{1}{2}}\otimes V)J.$

vi): This follows from $\underline{x}\underline{x}' = \sum_{i,j} x_i x_j \underline{f}_i \underline{f}_j'$, and the linearity of τ. □

For $\underline{x} \sim u(\underline{x};\underline{\mu},\Sigma) \in \mathcal{U}_{\varepsilon,r}$, put $\underline{y} = \Sigma^{-\frac{1}{2}}(\underline{x}-\underline{\mu})$. Define

$\alpha = E[2y_1^2 a_0'(\underline{y}'\underline{y})+a_0(\underline{y}'\underline{y})]$, $\alpha_1 = E[y_1^2 a_0^2(\underline{y}'\underline{y})]$,

$\gamma = E[y_1^2 y_2^2 a_2'(\underline{y}'\underline{y})]$, $\gamma_1 = E[y_1^2 y_2^2 a_2^2(\underline{y}'\underline{y})]$,

$\delta = E[y_1^4 a_2'(\underline{y}'\underline{y})]$, $\delta_1 = E[y_1^4 a_2^2(\underline{y}'\underline{y})]$,

$\beta = E[y_1^2 a_2(\underline{y}'\underline{y})]$.

A property of spherically symmetric functions is that $\delta = 3\gamma$ and $\delta_1 = 3\gamma_1$. Represent $\Sigma^{\frac{1}{2}}$ by its columns as $\Sigma^{\frac{1}{2}} = \|\underline{\lambda}_1,\ldots\underline{\lambda}_m\|$, and in partitioned form as

$$\Sigma^{\frac{1}{2}} = \begin{pmatrix} \overset{i-1}{R_{i1}} & \overset{m-i+1}{R_{i2}} \\ 0 & R_{i3} \end{pmatrix}\begin{matrix} i-1 \\ \\ m-i+1 \end{matrix} .$$

Put $R = \bigoplus_{i=1}^{m} R_{i3} = J'(I\otimes\Sigma^{\frac{1}{2}})J$. An identity which will prove useful, and which follows from the triangularity of $\Sigma^{\frac{1}{2}}$, is

$$JR' = (I\otimes\Sigma^{\frac{1}{2}'})J. \tag{8}$$

Define $m\times m$ matrices S and T by

$$S_{ij} = \tau(\underline{f}_i\underline{\lambda}_j'+\underline{\lambda}_j\underline{f}_i'), \quad T_{ij} = \tau(\underline{\lambda}_i\underline{\lambda}_j'+\underline{\lambda}_j\underline{\lambda}_i').$$

The vectors $\underline{\tau}_d{:}m\times 1$ and $\underline{\tau}_0{:}\frac{m(m-1)}{2}\times 1$ of diagonal and off-diagonal elements of T are defined by

$$P(\underset{s}{\text{vec}}\ T) = \begin{pmatrix} \underline{\tau}_d \\ \underline{\tau}_0 \end{pmatrix}.$$

Define $\frac{m(m+1)}{2} \times \frac{m(m+1)}{2}$ matrices

$$H = (\beta+2\gamma)(I+P_1'P_1), \quad \Delta_0 = \left(\frac{\partial V^{-\frac{1}{2}}}{\partial V}\right)_{V=\Sigma},$$

$$G_1 = J'(I\otimes S)JJ'(\Sigma^{\frac{1}{2}}\otimes I)J, \quad G_2 = J'(I\otimes T)J,$$

$$G = \beta G_1 - (\beta+2\gamma)G_2, \quad D_1 = 2I_m \oplus I_{m(m-1)/2}$$

$$D = P_1'P_1 + 2P_2'P_2 = \text{diag}(1,2,\ldots,2;1,2,\ldots,2;\ldots;1,2;1)$$

$$U = \begin{pmatrix} I_m - \tfrac{1}{2}\underline{1}\underline{\tau}_d' & -\underline{1}\underline{\tau}_0' \\ 0 & I \end{pmatrix} \begin{matrix} m \\ m(m-1)/2 \end{matrix} .$$

<u>LEMMA 3.2</u> The covariance matrix of $\underline{\psi}$ at θ_0 is

$$\text{cov}[\underline{\psi}]_{\theta_0} = \alpha_1 \Sigma \oplus \gamma_1 P' U D_1 U' P,$$

and has rank one less than its order.

<u>Proof</u>: We have $U\begin{pmatrix}\underline{1}\\ \underline{0}\end{pmatrix} = \underline{1} - \tfrac{1}{2}\underline{1}\underline{\tau}_d'\underline{1} = \underline{0}$, since

$$\tfrac{1}{2}\underline{\tau}_d'\underline{1} = \tfrac{1}{2}\sum_i T_{ii} = \tfrac{1}{2}\tau\left(\sum_i 2\underline{\lambda}_i \underline{\lambda}_i'\right) = \tau(\Sigma^{\frac{1}{2}}\Sigma^{\frac{1}{2}'}) = \tau(\Sigma) = 1.$$

Since the only linear constraint on $\underline{\psi}$ is that imposed by $\tau(\Sigma) = 1$, the second statement of the lemma follows.

With $\underline{y} = \Sigma^{-\frac{1}{2}}(\underline{x}-\underline{\mu})$, put

$$\underline{z} = a_0(\underline{y}'\underline{y})\Sigma^{\frac{1}{2}}\underline{y}, \quad Z_1 = a_2(\underline{y}'\underline{y})(\underline{y}\underline{y}')$$

$$Z_2 = a_2(\underline{y}'\underline{y})\tau(\Sigma^{\frac{1}{2}}\underline{y}\underline{y}'\Sigma^{\frac{1}{2}'})I, \quad Z = Z_1 - Z_2.$$

Then

$$\text{cov}[\underline{\psi}]_{\theta_0} = \begin{pmatrix} E[\underline{z}\underline{z}'] & E[\underline{z}(\underset{s}{\text{vec}}\ Z)'] \\ E[(\underset{s}{\text{vec}}\ Z)\underline{z}'] & E[(\underset{s}{\text{vec}}\ Z)(\underset{s}{\text{vec}}\ Z)'] \end{pmatrix} .$$

Since a_0 and a_2 vanish off of $[0,r^2]$, the distribution of $\underline{y}$ is symmetric on the range over which the expectations are taken. Thus

$E[\underset{-}{z}(\underset{s}{\text{vec}}\ Z_1)'] = 0_{m\times m(m+1)/2}$, and

$$E[\underset{-}{z}(\underset{s}{\text{vec}}\ Z_2)'] = E[a_0(\underline{y}'\underline{y})\Sigma^{\frac{1}{2}}\underline{y}\cdot a_2(\underline{y}'\underline{y})\tau(\Sigma^{\frac{1}{2}}\underline{y}\underline{y}'\Sigma^{\frac{1}{2}'})]\underline{e}'$$

$$= \Sigma^{\frac{1}{2}}E[a_0(\underline{y}'\underline{y})a_2(\underline{y}'\underline{y})\underline{y}\tau(\Sigma^{\frac{1}{2}}\underline{y}\underline{y}'\Sigma^{\frac{1}{2}'})]\underline{e}'.$$

The i^{th} component of the vector of expectations is

$$E[a_0(\underline{y}'\underline{y})a_2(\underline{y}'\underline{y})y_i\tau(\Sigma^{\frac{1}{2}}\underline{y}\underline{y}'\Sigma^{\frac{1}{2}'})]$$

$$= \tau(E[a_0(\underline{y}'\underline{y})a_2(\underline{y}'\underline{y})y_i\Sigma^{\frac{1}{2}}\underline{y}\underline{y}'\Sigma^{\frac{1}{2}'}])$$

$$= \tau(\Sigma^{\frac{1}{2}}E[a_0(\underline{y}'\underline{y})a_2(\underline{y}'\underline{y})y_i\underline{y}\underline{y}']\Sigma^{\frac{1}{2}'})$$

$$= \tau(0_{m\times m}) = 0.$$

Thus $E[\underset{-}{z}(\underset{s}{\text{vec}}\ Z)'] = 0_{m\times m(m+1)/2}$. Clearly, $E[\underline{z}\underline{z}'] = \alpha_1\Sigma$, and so

$$\text{cov}[\underline{\psi}]_{\theta_0} = \alpha_1\Sigma \oplus E[(\underset{s}{\text{vec}}\ Z)(\underset{s}{\text{vec}}\ Z)'].$$

We have, by properties Z1) and Z4),

$$E[(\underset{s}{\text{vec}}\ Z)(\underset{s}{\text{vec}}\ Z)'] = J'E[(I\otimes Z)\underline{f}\underline{f}'(I\otimes Z)]J$$

$$= J'\{E[A_1]-E[A_2]-E[A_2]'+E[A_3]\}J,$$

where

$$A_1 = (I\otimes Z_1)\underline{f}\underline{f}'(I\otimes Z_1),\ A_2 = (I\otimes Z_1)\underline{f}\underline{f}'(I\otimes Z_2),\ A_3 = (I\otimes Z_2)\underline{f}\underline{f}'(I\otimes Z_2).$$

i) $E[A_1] = E[(I\otimes a_2(\underline{y}'\underline{y})\underline{y}\underline{y}')\underline{f}\underline{f}'(I\otimes a_2(\underline{y}'\underline{y})\underline{y}\underline{y}')]$

$= E[a_2^2(\underline{y}'\underline{y})(I\otimes \underline{y}\underline{y}')\underline{f}\underline{f}'(I\otimes \underline{y}\underline{y}')]$

$= E[a_2^2(\underline{y}'\underline{y})(\text{vec }\underline{y}\underline{y}')(\text{vec }\underline{y}\underline{y}')']$.

The $(k,l)^{\text{th}}$ element of the $(i,j)^{\text{th}}$ block of the matrix of expectations is

$$E[a_2^2(\underline{y}'\underline{y})y_iy_ky_jy_l] = \begin{cases} \delta_1 = 3\gamma_1, & k=l=i=j; \\ \gamma_1, & k=l\neq i=j,\ k=i\neq l=j,\ k=j\neq l=i; \\ 0, & \text{otherwise}; \end{cases}$$

so that the $(i,j)^{\text{th}}$ block is

$$\gamma_1(\underline{f}_i\underline{f}_j'+\underline{f}_j\underline{f}_i') + \begin{cases} 0, & i\neq j; \\ \gamma_1 I, & i=j; \end{cases} \tag{9}$$

and so $E[A_1] = \gamma_1(\underline{f}\underline{f}'+F+I)$.

ii) $E[A_2] = E[(I\otimes a_2(\underline{y}'\underline{y})\underline{y}\underline{y}')\underline{f}\underline{f}'(I\otimes a_2(\underline{y}'\underline{y})\tau(\Sigma^{\frac{1}{2}}\underline{y}\underline{y}'\Sigma^{\frac{1}{2}'})I)]$

$= E[a_2^2(\underline{y}'\underline{y})\tau(\Sigma^{\frac{1}{2}}\underline{y}\underline{y}'\Sigma^{\frac{1}{2}'})(I\otimes \underline{y}\underline{y}')\underline{f}\underline{f}'(I\otimes I)]$

$= I\otimes E[a_2^2(\underline{y}'\underline{y})\tau(\Sigma^{\frac{1}{2}}\underline{y}\underline{y}'\Sigma^{\frac{1}{2}'})\underline{y}\underline{y}']\underline{f}\underline{f}'$.

The matrix of expectations has $(i,j)^{\text{th}}$ element

$$E[a_2^2(\underline{y}'\underline{y})\tau(\Sigma^{\frac{1}{2}}\underline{yy}'\Sigma^{\frac{1}{2}'})y_i y_j] = \tau(\Sigma^{\frac{1}{2}}E[a_2^2(\underline{y}'\underline{y})\underline{yy}'y_i y_j]\Sigma^{\frac{1}{2}'}). \quad (10)$$

The last matrix of expectations coincides with (9) above, so that (10) is

$$\gamma_1\tau(\Sigma^{\frac{1}{2}}\underline{f}_i\underline{f}_j'\Sigma^{\frac{1}{2}'}+\Sigma^{\frac{1}{2}}\underline{f}_j\underline{f}_i'\Sigma^{\frac{1}{2}'}) + \begin{cases}\tau(0), & i\neq j;\\ \gamma_1\tau(\Sigma), & i=j;\end{cases}$$

$$= \gamma_1\tau(\underline{\lambda}_i\underline{\lambda}_j'+\underline{\lambda}_j\underline{\lambda}_i') + \begin{cases}0, & i\neq j;\\ \gamma_1, & i=j.\end{cases}$$

Thus $E[a_2^2(\underline{y}'\underline{y})\tau(\Sigma^{\frac{1}{2}}\underline{yy}'\Sigma^{\frac{1}{2}'})\underline{yy}'] = \gamma_1(I+T)$, and

$$E[A_2] = \gamma_1(I\otimes(I+T))\underline{ff}' = \gamma_1(\underline{ff}'+(\text{vec } T)\underline{f}').$$

iii) $E[A_3] = E[(I\otimes a_2(\underline{y}'\underline{y})\tau(\Sigma^{\frac{1}{2}}\underline{yy}'\Sigma^{\frac{1}{2}'})I)\,\underline{ff}'(I\otimes a_2(\underline{y}'\underline{y})\tau(\Sigma^{\frac{1}{2}}\underline{yy}'\Sigma^{\frac{1}{2}'})I]$

$$= E[a_2^2(\underline{y}'\underline{y})\tau^2(\Sigma^{\frac{1}{2}}\underline{yy}'\Sigma^{\frac{1}{2}'})]\underline{ff}',$$

where the expectation is

$$E[a_2^2(\underline{y}'\underline{y})\tau(\Sigma^{\frac{1}{2}}\tau(\Sigma^{\frac{1}{2}}\underline{yy}'\Sigma^{\frac{1}{2}'})\underline{yy}'\Sigma^{\frac{1}{2}'})]$$

$$= \tau(\Sigma^{\frac{1}{2}}E[a_2^2(\underline{y}'\underline{y})\tau(\Sigma^{\frac{1}{2}}\underline{yy}'\Sigma^{\frac{1}{2}'})\underline{yy}']\Sigma^{\frac{1}{2}'})$$

$$= \gamma_1\tau(\Sigma^{\frac{1}{2}}(I+T)\Sigma^{\frac{1}{2}'}) \text{ as above,}$$

$$= \gamma_1(1+\tau(\Sigma^{\frac{1}{2}}T\Sigma^{\frac{1}{2}'})).$$

Expanding $\tau(\Sigma^{\frac{1}{2}}T\Sigma^{\frac{1}{2}'})$ gives

$$E[A_3] = \gamma_1(1+\underline{\tau}_0'\underline{\tau}_0+\tfrac{1}{2}\underline{\tau}_d'\underline{\tau}_d)\underline{f}\underline{f}'.$$

Summing up, and using $J'FJ = \bigoplus_{i=1}^{m} \underline{e}_i\underline{e}_i' = P_1'P_1$, we get

$$E[(\underset{s}{\text{vec}}\ Z)(\underset{s}{\text{vec}}\ Z)'] = \gamma_1[I+P_1'P_1+(\underline{\tau}_0'\underline{\tau}_0+\tfrac{1}{2}\underline{\tau}_d'\underline{\tau}_d)\underline{e}\underline{e}'-(\underset{s}{\text{vec}}\ T)\underline{e}'-\underline{e}(\underset{s}{\text{vec}}\ T)'].$$

The result then follows from

$$PE[(\underset{s}{\text{vec}}\ Z)(\underset{s}{\text{vec}}\ Z)']P' = \gamma_1 UD_1U'. \qquad \square$$

<u>LEMMA 3.3</u> The expectation of the Jacobian matrix of $\underline{\psi}$ at θ_0 is

$$E\left[\left(\frac{\partial\underline{\psi}}{\partial\theta}\right)_{\theta_0}\right] = -\alpha I_m \oplus (H+\underline{e}\underline{e}'G)R'\Delta_0.$$

<u>Proof</u>: Make the following transformations:

$\theta = (\underline{t},V) \to (\underline{t},W)$, where $W = V^{-\frac{1}{2}}$ is upper triangular, and $W'W = V^{-1}$;

$(\underline{t},W) \to (\underline{y},W)$, where $\underline{y} = W(\underline{x}-\underline{t})$;

$(\underline{y},W) \to (\underline{z},Z)$, where $\underline{z} = a_0(\underline{y}'\underline{y})W^{-1}\underline{y} = \underline{\psi}_0(\underline{x};\theta)$,

and $Z = a_2(\underline{y}'\underline{y})[\underline{y}\underline{y}'-\tau(W^{-1}\underline{y}\underline{y}'W^{-1'})I] = \underline{\psi}_2(\underline{x};\theta)$.

Then $\left(\dfrac{\partial\psi}{\partial\theta}\right) = \left(\dfrac{\partial \underline{z},Z}{\partial \underline{y},W}\right)\left(\dfrac{\partial \underline{y},W}{\partial \underline{t},W}\right)\left(\dfrac{\partial \underline{t},W}{\partial \underline{t},V}\right).$

Clearly, $\left(\dfrac{\partial \underline{t},W}{\partial \underline{t},V}\right)_{\theta_0} = I_m \oplus \Delta_0$. From Lemma 3.1 iv) and (8),

$$\left(\frac{\partial \underline{y},W}{\partial \underline{t},W}\right)_{\theta_0} = \left(\begin{array}{c|c} -W & (I_m \otimes (\underline{x}-\underline{t})')J \\ \hline 0 & I \end{array}\right)_{\theta_0} = \left(\begin{array}{c|c} -\Sigma^{-\frac{1}{2}} & (I_m \otimes \underline{y}')JR' \\ \hline 0 & I \end{array}\right).$$

Calculation of $\left(\dfrac{\partial \underline{z},Z}{\partial \underline{y},W}\right)_{\theta_0}$:

i) $$\left(\frac{\partial \underline{z}}{\partial \underline{y}}\right)_{\theta_0} = W^{-1}\left(\underline{y}\left(\frac{\partial a_0(\underline{y}'\underline{y})}{\partial \underline{y}}\right) + a_0(\underline{y}'\underline{y})\left(\frac{\partial \underline{y}}{\partial \underline{y}}\right)\right)_{\theta_0}$$

$$= W^{-1}(\underline{y}(2a_0'(\underline{y}'\underline{y})\underline{y}') + a_0(\underline{y}'\underline{y})I)_{\theta_0}$$

$$= \Sigma^{\frac{1}{2}}(2a_0'(\underline{y}'\underline{y})\underline{y}\underline{y}' + a_0(\underline{y}'\underline{y})I).$$

ii) $$\left(\frac{\partial \underline{z}}{\partial W}\right)_{\theta_0} = \left(\frac{\partial W^{-1}a_0(\underline{y}'\underline{y})\underline{y}}{\partial W^{-1}}\right)\left(\frac{\partial W^{-1}}{\partial W}\right)_{\theta_0}$$

$$= a_0(\underline{y}'\underline{y})(I_m \otimes \underline{y}')J\left(\frac{\partial W^{-1}}{\partial W}\right)_{\theta_0}.$$

iii) $$\left(\frac{\partial Z}{\partial W}\right)_{\theta_0} = -a_2(\underline{y}'\underline{y})\underline{e}\left(\frac{\partial \tau(W^{-1}\underline{y}\underline{y}'W^{-1'})}{\partial W^{-1}\underline{y}}\right)_{\theta_0}\left(\frac{\partial W^{-1}\underline{y}}{\partial W^{-1}}\right)\left(\frac{\partial W^{-1}}{\partial W}\right)_{\theta_0}$$

$$= -a_2(\underline{y}'\underline{y})\underline{e}\left(\frac{\partial \tau(\underline{x}\underline{x}')}{\partial \underline{x}}\right)_{\underline{x}=\Sigma^{\frac{1}{2}}\underline{y}}(I_m \otimes \underline{y}')J\left(\frac{\partial W^{-1}}{\partial W}\right)_{\theta_0}.$$

iv) $$\left(\frac{\partial Z}{\partial \underline{y}}\right)_{\theta_0} = \left(\frac{\partial \alpha_2(\underline{y}'\underline{y})\underline{y}\underline{y}'}{\partial \underline{y}}\right) - \underline{e}\left(\frac{\partial \alpha_2(\underline{y}'\underline{y})\tau(W^{-1}\underline{y}\underline{y}'W^{-1'})}{\partial \underline{y}}\right)_{\theta_0}$$

$$= \left[(\mathrm{vec}_s\ \underline{y}\underline{y}')\left(\frac{\partial \alpha_2(\underline{y}'\underline{y})}{\partial \underline{y}}\right) + \alpha_2(\underline{y}'\underline{y})\left(\frac{\partial \underline{y}\underline{y}'}{\partial \underline{y}}\right)\right]$$

$$-\underline{e}\left[\tau(W^{-1}\underline{y}\underline{y}'W^{-1'})\left(\frac{\partial \alpha_2(\underline{y}'\underline{y})}{\partial \underline{y}}\right) + \alpha_2(\underline{y}'\underline{y})\left(\frac{\partial \tau(\underline{x}\underline{x}')}{\partial \underline{x}}\right)_{\underline{x}=W^{-1}\underline{y}}\left(\frac{\partial W^{-1}\underline{y}}{\partial \underline{y}}\right)\right]_{\theta_0}$$

$$= \left[2\alpha_2'(\underline{y}'\underline{y})(\mathrm{vec}_s\ \underline{y}\underline{y}')\underline{y}' + \alpha_2(\underline{y}'\underline{y})\left(\frac{\partial \underline{y}\underline{y}'}{\partial \underline{y}}\right)\right]$$

$$-\underline{e}\left[2\alpha_2'(\underline{y}'\underline{y})\tau(\Sigma^{\frac{1}{2}}\underline{y}\underline{y}'\Sigma^{\frac{1}{2}'})\underline{y}' + \alpha_2(\underline{y}'\underline{y})\left(\frac{\partial \tau(\underline{x}\underline{x}')}{\partial \underline{x}}\right)_{\underline{x}=\Sigma^{\frac{1}{2}}\underline{y}}\Sigma^{\frac{1}{2}}\right].$$

We have

$$E\left[\left(\frac{\partial \underline{\psi}}{\partial \theta}\right)_{\theta_0}\right] = E\left[\left(\frac{\partial \underline{z}, Z}{\partial \underline{y}, W}\right)\left(\frac{\partial \underline{y}, W}{\partial \underline{t}, W}\right)\right]_{\theta_0}\left(\frac{\partial \underline{t}, W}{\partial \underline{t}, V}\right)_{\theta_0}$$

$$= \begin{pmatrix} -E\left[\left(\frac{\partial \underline{z}}{\partial \underline{y}}\right)_{\theta_0}\right]\Sigma^{-\frac{1}{2}} & E\left[\left(\frac{\partial \underline{z}}{\partial \underline{y}}\right)_{\theta_0}(I_m \otimes \underline{y}')\right]JR' + E\left[\left(\frac{\partial \underline{z}}{\partial W}\right)_{\theta_0}\right] \\ -E\left[\left(\frac{\partial Z}{\partial \underline{y}}\right)_{\theta_0}\right]\Sigma^{-\frac{1}{2}} & E\left[\left(\frac{\partial Z}{\partial \underline{y}}\right)_{\theta_0}(I_m \otimes \underline{y}')\right]JR' + E\left[\left(\frac{\partial Z}{\partial W}\right)_{\theta_0}\right] \end{pmatrix} \cdot (I_m \oplus \Delta_0).$$

As in Lemma 3.2, the two off-diagonal blocks are zero, and

$E\left[\left(\frac{\partial \underline{z}}{\partial \underline{y}}\right)_{\theta_0}\right] = \alpha\Sigma^{\frac{1}{2}}$, so that

$$E\left[\left(\frac{\partial\psi}{\partial\theta}\right)_{\theta_0}\right] = -\alpha I_m \oplus \left\{E\left[\left(\frac{\partial Z}{\partial \underline{y}}\right)_{\theta_0}(I_m \otimes \underline{y}')\right]JR' + E\left[\left(\frac{\partial Z}{\partial W}\right)_{\theta_0}\right]\right\}\Delta_0. \qquad (11)$$

From iii) above,

$$E\left[\left(\frac{\partial Z}{\partial W}\right)_{\theta_0}\right] = -\underline{e}E\left[a_2(\underline{y}'\underline{y})\left(\frac{\partial\tau(\underline{x}\underline{x}')}{\partial\underline{x}}\right)_{\underline{x}=\Sigma^{\frac{1}{2}}\underline{y}}(I_m \otimes \underline{y}')\right]J\left(\frac{\partial W^{-1}}{\partial W}\right)_{\theta_0}.$$

The vector of expectations is $1 \times m^2$. The j^{th} term in the i^{th} block $(1 \leq i, j \leq m)$ is, from Lemma 3.1 vi)

$$E[a_2(\underline{y}'\underline{y})\tau(\underline{f}_i\underline{y}'\Sigma^{\frac{1}{2}'} + \Sigma^{\frac{1}{2}}\underline{y}\underline{f}_i')y_j]$$

$$= \tau(\underline{f}_i E[a_2(\underline{y}'\underline{y})\underline{y}'y_j]\Sigma^{\frac{1}{2}'} + \Sigma^{\frac{1}{2}}E[a_2(\underline{y}'\underline{y})\underline{y}y_j]\underline{f}_i')$$

$$= \tau(\beta\underline{f}_i\underline{f}_j'\Sigma^{\frac{1}{2}'} + \beta\Sigma^{\frac{1}{2}}\underline{f}_j\underline{f}_i')$$

$$= \beta\tau(\underline{f}_i\underline{\lambda}_j' + \underline{\lambda}_j\underline{f}_i')$$

$$= \beta S_{ij}.$$

Thus the vector of expectations is $(\text{vec } S)'$, and so

$$E\left[\left(\frac{\partial Z}{\partial W}\right)_{\theta_0}\right] = -\beta\underline{e}(\text{vec } S)'J\left(\frac{\partial W^{-1}}{\partial W}\right)_{\theta_0}$$

$$= -\beta\underline{e}((I \otimes S')\underline{f})'JJ'(\Sigma^{\frac{1}{2}} \otimes \Sigma^{\frac{1}{2}'})J \text{ by Z4) and Lemma 3.1 ii)},$$

$$= -\beta\underline{e}\underline{f}'(I \otimes S)JJ'(\Sigma^{\frac{1}{2}} \otimes I)(I \otimes \Sigma^{\frac{1}{2}'})J$$

$$= -\beta\underline{e}\underline{f}'(I \otimes S)JJ'(\Sigma^{\frac{1}{2}} \otimes I)JR' \text{ by (8)}$$

$$= -\beta\underline{e}\underline{e}'G_1R'. \qquad (12)$$

Calculations very similar to those in Lemma 3.2 give

$$E\left[\left(\frac{\partial Z}{\partial \underline{y}}\right)_{\theta_0}(I_m \otimes \underline{y}')\right] JR'$$

$$= (\beta+2\gamma)(J'(I+F) - \underline{e}(\text{vec } T)')JR'$$

$$= (\beta+2\gamma)((I+P_1'P_1) - \underline{e}(\text{vec } T)'J)R'$$

$$= (H-(\beta+2\gamma)\underline{e}\underline{e}'J'(I\otimes T)J)R' \text{ by Z4),}$$

$$= (H-(\beta+2\gamma)\underline{e}\underline{e}'G_2)R'. \qquad (13)$$

Substituting (12) and (13) into (11) gives

$$E\left[\left(\frac{\partial \underline{\psi}}{\partial \theta}\right)_{\theta_0}\right] = -\alpha I_m \oplus (H+\underline{e}\underline{e}'G)R'\Delta_0. \qquad \square$$

It is not difficult to show that $\underline{e}'G_1\underline{e} = \underline{e}'G_2\underline{e} = 2\tau(\Sigma) = 2$, $H\underline{e} = 2(\beta+2\gamma)\underline{e}$, and $\underline{e}'G_1 = \underline{e}'G_2 = (\underset{s}{\text{vec}}\ T)'$. From Lemma 3.2 iii) and (8),

$$\Delta_0^{-1} = -J'(I+F)(\Sigma^{\frac{1}{2}}\otimes\Sigma)J$$

$$= -J'(I+F)(\Sigma^{\frac{1}{2}}\otimes\Sigma^{\frac{1}{2}})(I\otimes\Sigma^{\frac{1}{2}'})J$$

$$= -J'(I+F)(\Sigma^{\frac{1}{2}}\otimes\Sigma^{\frac{1}{2}})JR',$$

so that $(R'\Delta_0)^{-1} = -J'(I+F)(\Sigma^{\frac{1}{2}}\otimes\Sigma^{\frac{1}{2}})J$. The following lemma may then be verified directly.

<u>LEMMA 3.4</u> Assume that none of α, β, $\beta+2\gamma$ is zero. Then $\left(E\left[\left(\frac{\partial \underline{\psi}}{\partial \theta}\right)_{\theta_0}\right]\right)^{-1}$ exists and equals

$$-\alpha^{-1}I_m \oplus (R'\Delta_0)^{-1}(I + \frac{\gamma}{\beta}\underline{e}(\operatorname{vec}_s T)')H^{-1},$$

where $H^{-1} = \frac{1}{2(\beta+2\gamma)}(2I-P_1'P_1)$ and $(R'\Delta_0)^{-1} = -J'(I+F)(\Sigma^{\frac{1}{2}}\otimes\Sigma^{\frac{1}{2}})J.$ □

Evaluating $\left(E\left[\left(\frac{\partial\underline{\psi}}{\partial\theta}\right)_{\theta_0}\right]\right)^{-1}(\operatorname{cov}[\underline{\psi}]_{\theta_0})\left(E\left[\left(\frac{\partial\underline{\psi}}{\partial\theta}\right)_{\theta_0}\right]\right)^{-1'}$ gives

<u>THEOREM 3.5</u> Under the conditions of Lemma 3.4, the estimator $\hat{\theta}$, defined at (5), is a consistent estimator of θ_0. Furthermore,

$$\sqrt{n}(\hat{\theta}-\theta_0) \xrightarrow{D} N_{m(m+3)/2}(\underline{0}, \frac{\alpha_1}{\alpha^2}\mathbf{\Sigma} \oplus \frac{\gamma_1}{2(\beta+2\gamma)^2}\Lambda\Lambda'),$$

where $\Lambda = [J'(\Sigma^{\frac{1}{2}}\otimes\Sigma^{\frac{1}{2}})(I+F)J][I-\tfrac{1}{2}\underline{e}(\operatorname{vec}_s T)']D^{\frac{1}{2}}$

and $\Lambda\Lambda' = J'(I+F)(\Sigma\otimes\Sigma)(I+F)J + 2\tau(\Sigma^{\frac{1}{2}}T\Sigma^{\frac{1}{2}'})(\operatorname{vec}_s \Sigma)(\operatorname{vec}_s \Sigma)'$

$$-2(\operatorname{vec}_s \Sigma)(\operatorname{vec}_s \Sigma^{\frac{1}{2}}T\Sigma^{\frac{1}{2}'})' - 2(\operatorname{vec}_s \Sigma^{\frac{1}{2}}T\Sigma^{\frac{1}{2}'})(\operatorname{vec}_s \Sigma)'.$$ □

Remarks:

1) With $\Sigma = \|\underline{\sigma}_1,\ldots,\underline{\sigma}_m\|$, we have $(\Sigma^{\frac{1}{2}}T\Sigma^{\frac{1}{2}'})_{ij} = \tau(\underline{\sigma}_i\underline{\sigma}_j'+\underline{\sigma}_j\underline{\sigma}_i')$, so that $\Lambda\Lambda'$ is independent of the choice of $\Sigma^{\frac{1}{2}}$.

2) The location and scale components of $\hat{\theta}$ are asymptotically uncorrelated, hence asymptotically independent. Upon expansion

$$\lim_{n\to\infty}\sqrt{n}\operatorname{cov}(\hat{\sigma}_{ik},\hat{\sigma}_{jl}) = \frac{\gamma_1}{2(\beta+2\gamma)^2}\{2(\sigma_{ij}\sigma_{kl} + \sigma_{jk}\sigma_{il})$$

$$+ 2\tau(\Sigma^{\frac{1}{2}}T\Sigma^{\frac{1}{2}'})\sigma_{ik}\sigma_{jl} - 2\tau(\underline{\sigma}_i\underline{\sigma}_k'+\underline{\sigma}_k\underline{\sigma}_i')\sigma_{jl} - 2\tau(\underline{\sigma}_j\underline{\sigma}_l'+\underline{\sigma}_l\underline{\sigma}_j')\sigma_{ik}\}.$$

We will now remove the singularity

in Λ. Essentially, this involves removing from θ_0 one of the diagonal elements of Σ, and making the corresponding changes in $\hat{\theta}$ and $\Lambda\Lambda'$. By a re-enumeration if necessary, we may assume that it is Σ_{11} which is to be removed.

Put $p = \frac{m(m+1)}{2}$, and let ϕ_0 and $\hat{\phi}$ be the $p+m-1\times 1$ vectors which result from removing the $(m+1)^{\text{th}}$ elements from each of θ_0 and $\hat{\theta}$. Define K: $p-1 \times p$ by $K = \|\underline{0}|I_{p-1}\|$. Then, e.g., $\phi_0 = (I_m \oplus K)\theta_0$. Since $\text{cov}[\underline{\psi}]_{\theta_0}$ has rank $m+p-1$, Λ has rank $p-1$. It is easy to see that $\Lambda\underline{e} = \underline{0}$, so that, partitioning Λ as $\Lambda = \|\underline{c}|\Lambda_1\|$, where $\Lambda_1 = \Lambda K'$ is $p\times p-1$ with rank $p-1$, we have $\underline{0} = \Lambda\underline{e} = \underline{c} + \Lambda_1 K\underline{e}$. Solving for $\underline{c}$ then gives

$$\Lambda\Lambda' = \Lambda_1(I_{p-1}+K\underline{e}\underline{e}'K')\Lambda_1' = \Lambda K'(I_{p-1}+K\underline{e}\underline{e}'K')K\Lambda'.$$

A version of $(I_{p-1}+K\underline{e}\underline{e}'K')^{\frac{1}{2}}$ is $\left(I_{p-1} + \frac{K\underline{e}\underline{e}'K'}{1+\sqrt{m}}\right)$. This gives

$$K\Lambda\Lambda'K' = \left[K\Lambda K'\left(I_{p-1} + \frac{K\underline{e}\underline{e}'K'}{1+\sqrt{m}}\right)\right]\left[\left(I_{p-1} + \frac{K\underline{e}\underline{e}'K'}{1+\sqrt{m}}\right)K\Lambda'K'\right].$$

We then have

<u>LEMMA 3.6</u> Under the conditions of Lemma 3.4, the estimator $\hat{\phi}$ is consistent for ϕ_0, and

$$\sqrt{n}(\hat{\phi}-\phi_0) \xrightarrow{D} N_{p+m-1}\left(\underline{0}, \frac{\alpha_1}{\alpha^2}\,\Sigma \oplus \frac{\gamma_1}{2(\beta+2\gamma)^2}\,BB'\right),$$

where $B = K\Lambda K'\left(I_{p-1} + \frac{K\underline{e}\underline{e}'K'}{1+\sqrt{m}}\right)$: $p-1 \times p-1$ is non-singular. □

As preparation for the next sections, it is convenient to reparameterize the expectations in Lemma 3.6 in terms of $|\underline{y}| = X$. Recall from (0.4-5) that on $D_{\sqrt{r}}(\underline{0},I)$, X has density $mC_m x^{m-1} f(x)$. For p,q even integers, we find

$$E[y_1^p y_2^q a(\underline{y}'\underline{y})] = \frac{\Gamma\left(\frac{m}{2}\right)\Gamma\left(\frac{p+1}{2}\right)\Gamma\left(\frac{q+1}{2}\right)}{\pi\Gamma\left(\frac{m+p+q}{2}\right)} E[X^{p+q} a(X^2)],$$

so that, defining

$$a_1(\underline{y}'\underline{y}) = (\underline{y}'\underline{y}) a_2(\underline{y}'\underline{y}),$$

$$V(a_1,u) = \frac{\gamma_1}{2(\beta+2\gamma)^2}, \quad V(a_0,u) = \frac{\alpha_1}{\alpha^2};$$

we have

$$V(a_i,u) = \left(\frac{m+2}{2}\right)^i \frac{E[mX^{2-2i} a_i(X^2)]}{(E[ma_i(X^2)+2X^2 a_i'(X^2)])^2} \quad . \tag{14}$$

In Section 5, we will consider the problem of choosing functions a_i which minimize, with respect to the natural ordering of positive definite matrices, the maximum asymptotic variance of estimators $\hat{\theta}$ defined by (5). The significance of Theorem 3.5 and Lemma 3.6 is that this is now a problem of independently optimizing the scalar functionals $V(a_i,u)$.

If f is differentiable, then (14) becomes

$$V(a_i,u) = \left(\frac{m+2}{2}\right)^i \frac{m\int_0^r x^{2-2i} a_i^2(x^2) mC_m x^{m-1} f(x)dx}{\left[\int_0^r x a_i(x^2) mC_m x^{m-1} f'(x)dx\right]^2} .$$

With

$$\psi_i(x) = x^{1-2i} a_i(x^2),$$

$$\eta_i(x) = \left(\frac{2}{m+2}\right)^i \frac{x^{2i}}{m},$$

$$\nu(x) = \frac{mC_m}{2}|x|^{m-1},$$

it becomes

$$V(a_i,u) = \frac{\int_{-r}^r \psi_i^2(x) f(x) \eta_i(x) \nu(x) dx}{\left[\int_{-r}^r \psi_i(x) f'(x) \eta_i(x) \nu(x) dx\right]^2} . \qquad (15)$$

This latter version is the one with which we shall work in the following sections.

Confidence Regions

Lemma 3.6 allows us to construct asymptotic confidence regions for linear functions of ϕ_0, based on the normal theory. Let M: $q\times(m+p-1)$ be a matrix of constants of rank q. Put $C = V(a_0,u)\Sigma \oplus V(a_1,u)BB'$, and let $\hat{C}$ be the estimate of C using $\hat{\theta}$. The consistency of $\hat{\theta}$ implies that $\hat{C} \xrightarrow{P} C$. This, together with $\sqrt{n}(M\hat{\phi}-M\phi_0) \xrightarrow{D} N_q(\underline{0},MCM')$ implies that

$$n(M\hat{\phi}-M\phi_0)'(M\hat{C}M')^{-1}(M\hat{\phi}-M\phi_0) \xrightarrow{D} \chi_q^2.$$

Thus an asymptotic 100α% confidence region for $M\phi_0$ is the ellipse $E_t(M\hat{\phi}, n^{-1}M\hat{C}M')$, where t is determined from $\chi^2_q(t^2) = \alpha$. Choices of a_0 and a_1 which are minimax with respect to the asymptotic variance of $\hat{\phi}$ then have the property of minimizing the maximum asymptotic volume of such regions.

Numerical computations

We have defined $\hat{\theta}$ as the root Θ^*, obtained by the Newton-Raphson method, of $F_n(\theta) = \underline{0}$; i.e.

$$\Theta_1 = \left(\underline{\bar{x}}^n_\alpha, \frac{V^\alpha_n}{\tau(V^\alpha_n)}\right),$$

$$\Theta_{k+1} = \Theta_k - \left(\frac{\partial F_n(\theta)}{\partial\theta}\right)^{-1}_{\theta=\Theta_k} F_n(\Theta_k),$$

$$\hat{\theta} = \lim_{k\to\infty} \Theta_k.$$

Consider the modified iteration process

$$\Theta_{k+1} = \Theta_k - [A(\Theta_k)]^{-1}F_n(\Theta_k), \tag{16}$$

where $A(\Theta_k) = E\left[\left(\frac{\partial\underline{\psi}}{\partial\theta}\right)_{\theta_0}\right]_{\theta_0=\Theta_k}$; i.e. $[A(\Theta_k)]^{-1}$ is the matrix in the statement of Lemma 3.4, with the parameters estimated by Θ_k. Since

$$[A(\Theta^*)]^{-1}\left(\frac{\partial F_n}{\partial\theta}\right)_{\Theta^*} = \left(E\left[\left(\frac{\partial\underline{\psi}}{\partial\theta}\right)_{\Theta^*}\right]\right)^{-1}\left(\frac{\partial F_n}{\partial\theta}\right)_{\Theta^*} \xrightarrow{P} I,$$

the spectral radius

$$\rho\left(I-[A(\theta^*)]^{-1}\left(\frac{\partial F_n}{\partial\theta}\right)_{\theta^*}\right) \xrightarrow{P} 0.$$

Then by 10.2.1 of Ortega and Rheinboldt [9], the iteration process (16) has, asymptotically, the same super-linear convergence properties as the Newton-Raphson method. Since the Jacobian matrices no longer need to be calculated and inverted, there is an obvious reduction in computational complexity. In fact, upon expanding (16), it becomes a fixed point process with easily calculated weights. Let X, T_k, $\underline{a}_k$, A_k be as at (1) - (3), but with $\theta = (\underline{t},V)$ replaced by $\theta_k = (\underline{t}_k,V_k)$. Think of $G_n(\theta_k)$, from (4), as a vector in $R^{m(m+3)/2}$, and put

$$r^2_{i,k} = (\underline{x}_i-\underline{t}_k)'V_k^{-1}(\underline{x}_i-\underline{t}_k),$$

$$b_k = \frac{1}{m}\sum_i[2r^2_{i,k}a_0'(r^2_{i,k})+ma_0(r^2_{i,k})],$$

$$c_k = \frac{1}{m(m+2)}\sum_i[2r^2_{i,k}a_1'(r^2_{i,k})+ma_1(r^2_{i,k})],$$

$$D_k = b_k^{-1}I_m \oplus c_k^{-1}I_{m(m+1)/2}.$$

Then (16) becomes

$$\theta_1 = \left(\underline{x}^n_\alpha, \frac{V^n_\alpha}{\tau(V^n_\alpha)}\right),$$

$$\theta_{k+1} = \theta_k + D_kG_n(\theta_k), \qquad (17)$$

$$\hat{\theta} = \lim_{k\to\infty}\theta_k.$$

Robustness measures

From (17) we easily obtain the influence function $IC(\underline{x};\theta_0) = -\left(E\left[\left(\frac{\partial\underline{\psi}}{\partial\theta}\right)_{\theta_0}\right]\right)^{-1}\underline{\psi}(\underline{x};\theta_0)$, it is $IC(\underline{x};\theta_0) = \left(\frac{a_0((\underline{x}-\underline{\mu})'\Sigma^{-1}(\underline{x}-\underline{\mu}))}{\alpha}(\underline{x}-\underline{\mu}),\ \frac{a_1((\underline{x}-\underline{\mu})'\Sigma^{-1}(\underline{x}-\underline{\mu}))}{\beta+2\gamma}[(\underline{x}-\underline{\mu})(\underline{x}-\underline{\mu})'-\tau((\underline{x}-\underline{\mu})(\underline{x}-\underline{\mu})')\Sigma]\right)$. From this, we find that the "gross error sensitivity"

$$GES = \sup_{\underline{x}}|IC(\underline{x};\theta_0)|$$

is, for location, $GES_0 = \frac{\lambda_1^{1/2}}{|\alpha|}\sup_z za_0(z^2)$, where λ_1 is the largest eigenvalue of Σ. For scale the expression is not so simple.

The case $r = \infty$

If $r = \infty$, then $\tau(\Sigma)$ need not be known and $\hat{\theta}$ is consistent for that multiple of Σ which has $\tau(\Sigma) = 1$. In this case, the problem of completely spurious solutions to $F_n(\theta) = \underline{0}$, as discussed at the beginning of Section 2, is no longer present. The construction of the starting value $\theta = \left(\underline{\bar{x}}^n_\alpha, \frac{V^n_\alpha}{\tau(V^n_\alpha)}\right)$ is then unnecessary if the zero of $F_n(\theta)$ is unique.

A partial result in this direction is given below. Recall that $F_n(\theta) \xrightarrow{P} E[\underline{\psi}(\underline{x};\theta)]$.

<u>THEOREM 3.7</u> If $r = \infty$, then under the following conditions the zero θ_0 of $E[\underline{\psi}(\underline{x};\theta)]$ is unique:

a) The function $a_2(z)$ is non-increasing, and $a_1(z) = za_2(z)$ is non-decreasing; and either

b1) The function $a_3(z) = za_0(z^2)$ is non-decreasing; or

b2) For every $u \in U_{\varepsilon,\infty}$, $u(\underline{x};\underline{0},I) = f(|\underline{x}|)$ is a strictly decreasing function of $|\underline{x}|$.

Proof: The proof is a modification of the proofs of Theorems 1) and 3) of Maronna [8]. Put

$$Q_1(\underline{\mu},\Sigma) = E[a_0((\underline{x}-\underline{\mu})'\Sigma^{-1}(\underline{x}-\underline{\mu}))(\underline{x}-\underline{\mu})]$$

$$Q_2(\underline{\mu},\Sigma) = \frac{E[a_2((\underline{x}-\underline{\mu})'\Sigma^{-1}(\underline{x}-\underline{\mu}))(\underline{x}-\underline{\mu})(\underline{x}-\underline{\mu})']}{\tau(E[a_2((\underline{x}-\underline{\mu})'\Sigma^{-1}(\underline{x}-\underline{\mu}))(\underline{x}-\underline{\mu})(\underline{x}-\underline{\mu})'])} .$$

For any $\theta = (\underline{\nu},\Lambda)$, $E[\underline{\psi}(\underline{x};\theta)] = \underline{0}$ iff $(Q_1(\underline{\nu},\Lambda),Q_2(\underline{\nu},\Lambda)) = (\underline{0},\Lambda)$. We may assume that $\underline{\mu} = \underline{0}$, in which case we must show that $(Q_1(\underline{\nu},\Lambda),Q_2(\underline{\nu},\Lambda)) = (\underline{0},\Lambda)$ implies $(\underline{\nu},\Lambda) = (\underline{0},\Sigma)$. This follows from

i) If $\underline{\nu} \neq \underline{0}$, then for any $\Lambda > 0$, $Q_1(\underline{\nu},\Lambda) \neq \underline{0}$.

ii) If $Q_2(\underline{0},\Lambda) = \Lambda$, then $\Lambda = \Sigma$.

We first show that b1) and b2) each imply i).

Put $\underline{y} = \Sigma^{-\frac{1}{2}}\underline{x}$, $\underline{\lambda} = \Sigma^{-\frac{1}{2}}\underline{\nu}$, $S = \Sigma^{-\frac{1}{2}}\Lambda\Sigma^{-\frac{1}{2}'}$. Then $\underline{y} \sim u(\underline{y};\underline{0},I)$ and i) is equivalent to

iii) If $\underline{\lambda} \neq \underline{0}$ then for any $S > 0$, $Q_1(\underline{\lambda},S) \neq \underline{0}$.

To prove iii), it suffices to show that

$$\underline{\lambda}'Q_1(\underline{\lambda},S) = \int_{R^m} a_0((\underline{y}-\underline{\lambda})'S^{-1}(\underline{y}-\underline{\lambda}))\underline{\lambda}'(\underline{y}-\underline{\lambda})f(|\underline{y}|)d\underline{y} < 0,$$

for all $\underline{\lambda} \neq \underline{0}$. Let $A = \{\underline{y}|\underline{\lambda}'(\underline{y}-\underline{\lambda})>0\}$. Write the integral as $\int_A + \int_{A^c}$, and apply the change of variables $\underline{z} = -(\underline{y}-\underline{\lambda}) + \underline{\lambda}$ in the second integral to get

$$\underline{\lambda}'Q_1(\underline{\lambda},S) = \int_A a_0((\underline{y}-\underline{\lambda})'S^{-1}(\underline{y}-\underline{\lambda}))\underline{\lambda}'(\underline{y}-\underline{\lambda})(f(|\underline{y}|)-f(|2\underline{\lambda}-\underline{y}|))d\underline{y},$$

which is negative if b2) holds since $|\underline{y}| > |2\underline{\lambda}-\underline{y}|$ for $\underline{y} \in A$.

To see that b1) also implies i), put $\underline{y} = \Lambda^{-\frac{1}{2}}\underline{x}$, $\underline{\lambda} = \Lambda^{-\frac{1}{2}}\underline{\nu}$, $S = \Lambda^{-\frac{1}{2}}\Sigma\Lambda^{-\frac{1}{2}'}$, so that $\underline{y} \sim u(\underline{y};\underline{0},S)$. Then i) is equivalent to

iv) If $\underline{\lambda} \neq \underline{0}$, then $Q_1(\underline{\lambda},I) \neq \underline{0}$.

It then suffices to show that

$$\underline{\lambda}'Q_1(\underline{\lambda},I) = \underline{\lambda}'E\left[a_3(|\underline{y}-\underline{\lambda}|)\frac{(\underline{y}-\underline{\lambda})}{|\underline{y}-\underline{\lambda}|}\right] < 0 \text{ for all } \underline{\lambda} \neq \underline{0}.$$

The Cauchy-Schwartz inequality and the fact that a_3 is non-decreasing yields

$$(\underline{a}-\underline{b})'\left(a_3(|\underline{a}|)\frac{\underline{a}}{|\underline{a}|} - a_3(|\underline{b}|)\frac{\underline{b}}{|\underline{b}|}\right) \geq 0 \text{ for all } \underline{a},\underline{b},$$

with equality iff $\underline{a}$ is a multiple of $\underline{b}$. With $\underline{a} = \underline{y}-\underline{\lambda}$ and $\underline{b} = \underline{y}$ we then have

$$\underline{\lambda}'Q_1(\underline{\lambda},I) \leq \underline{\lambda}'E\left[a_3(|\underline{y}|)\frac{\underline{y}}{|\underline{y}|}\right] = 0,$$

with equality iff $\underline{y}$ is a multiple of $\underline{y}-\underline{\lambda}$ with probability one. Thus b1) implies i).

We now show that a) implies ii). It suffices to show that a) implies

v) If $\underline{x} \sim u(\underline{x};\underline{0},\Sigma) \in U_{\varepsilon,\infty}$ and $E[a_2(\underline{x}'\Lambda^{-1}\underline{x})\underline{x}\underline{x}'] = k\Lambda$ for some $k > 0$, then $\Lambda = l\Sigma$ for some $l > 0$.

For, if $Q_2(\underline{0},\Lambda) = \Lambda$, then $\tau(\Lambda) = 1$ and $E[a_2(\underline{x}'\Lambda^{-1}\underline{x})\underline{xx}'] = k\Lambda$ for $k = \tau(E[a_2(\underline{x}'\Lambda^{-1}\underline{x})\underline{xx}'])$. Then v) and $\tau(\Sigma) = 1$ imply $\Lambda = \Sigma$.

Let Γ be an orthogonal matrix satisfying $\Gamma\Sigma^{-\frac{1}{2}}\Lambda\Sigma^{-\frac{1}{2}'}\Gamma' = D = \text{diag}(\lambda_1,\ldots,\lambda_m)$, where $\lambda_1 \geq \ldots \geq \lambda_m > 0$ are the eigenvalues of $\Sigma^{-\frac{1}{2}}\Lambda\Sigma^{-\frac{1}{2}'}$. With $\underline{z} = \Gamma\Sigma^{-\frac{1}{2}}\underline{x}$, v) is equivalent to

vi) If $\underline{z} \sim u(\underline{z};\underline{0},I) \in \mathcal{U}_{\varepsilon,\infty}$, and $E[a_2(\underline{z}'D^{-1}\underline{z})\underline{zz}'] = kD$, then $D = lI$ for some l; i.e. $\lambda_1 = \ldots = \lambda_m$.

Put
$$R = E\left[a_2\left(\frac{\underline{z}'\underline{z}}{\lambda_m}\right)\frac{\underline{zz}'}{\lambda_m}\right] = E\left[a_1\left(\frac{\underline{z}'\underline{z}}{\lambda_m}\right)\frac{\underline{zz}'}{\underline{z}'\underline{z}}\right].$$

Note that $\frac{\underline{z}'\underline{z}}{\lambda_1} \leq \underline{z}'D^{-1}\underline{z} \leq \frac{\underline{z}'\underline{z}}{\lambda_m}$, and that the inequalities are strict with probability one unless $\lambda_1 = \ldots = \lambda_m$.

In any event, $\lambda_m R \leq kD$ since a_2 is non-increasing. Suppose that it is not the case that $\lambda_1 = \ldots = \lambda_m$. We claim that then $\lambda_1 R > kD$. Equivalently, $\underline{t}'\lambda_1 R\underline{t} > \underline{t}'kD\underline{t}$ for any $\underline{t} \neq \underline{0}$. But

$$\underline{t}'kD\underline{t} = E[a_2(\underline{z}'D^{-1}\underline{z})(\underline{t}'\underline{z})^2] = E\left[\frac{a_1(\underline{z}'D^{-1}\underline{z})(\underline{t}'\underline{z})^2}{(\underline{z}'D^{-1}\underline{z})}\right].$$

The integrand is strictly less than $\lambda_1 a_1\left(\frac{\underline{z}'\underline{z}}{\lambda_m}\right)\frac{(\underline{t}'\underline{z})^2}{(\underline{z}'\underline{z})}$ with probability one, and so

$$\underline{t}'kD\underline{t} < \lambda_1 E\left[a_1\left(\frac{\underline{z}'\underline{z}}{\lambda_m}\right)\frac{(\underline{t}'\underline{z})^2}{(\underline{z}'\underline{z})}\right] = \underline{t}'\lambda_1 R\underline{t}.$$

Thus $\lambda_m R \leq kD < \lambda_1 R$. In particular,

$$\lambda_m R_{mm} = \lambda_m \underline{f}_m' R \underline{f}_m \leq k \underline{f}_m' D \underline{f}_m = k\lambda_m,$$

and

$$k\lambda_1 = k\underline{f}_1' D \underline{f}_1 < \lambda_1 \underline{f}_1' R \underline{f}_1 = \lambda_1 R_{11};$$

i.e. $R_{mm} < R_{11}$. This contradiction completes the proof. □

Maronna [8] and Huber [5] have also considered the problem of estimation of multivariate location and scale in the case $r = \infty$. Their approaches are somewhat different than ours. The major difference is that in neither case does it appear to be the intent to construct an estimator of $\underline{\mu}$ and a specified multiple of Σ which is globally consistent throughout $U_{\varepsilon,\infty}$, if $\varepsilon > 0$.

In both cases, the estimator $\hat{\theta}$ is, in our notation, the zero of

$$H_n(\theta) = n^{-1}(V^{-\frac{1}{2}}(X-T)\underline{a}, [V^{-\frac{1}{2}}(X-T)A(X-T)'V^{-\frac{1}{2}'} - \sum_{i=1}^{n} a_4((\underline{x}_i-\underline{t})'V^{-1}(\underline{x}_i-\underline{t}))I]),$$

where $a_4(\cdot)$ is some sufficiently smooth function. Maronna takes $a_4 \equiv 1$. With

$$\theta = (\underline{t}, V),\ \underline{y}_i = V^{-\frac{1}{2}}(\underline{x}_i-\underline{t}),\ \underline{\psi}_0(\underline{x}_i;\theta) = a_0(\underline{y}_i'\underline{y}_i)\underline{y}_i,$$

$$\underline{\psi}_1(\underline{x}_i;\theta) = a_2(\underline{y}_i'\underline{y}_i)\underline{y}_i\underline{y}_i' - a_4(\underline{y}_i'\underline{y}_i)I,\ \underline{\psi} = (\underline{\psi}_0, \underline{\psi}_1);$$

we have

$$H_n(\theta) = n^{-1} \sum_i \underline{\psi}(\underline{x}_i;\theta) \xrightarrow{P} E[\underline{\psi}(\underline{x};\theta)].$$

Under conditions similar to those in Theorem 3.7, which is essentially Maronna's result, Maronna shows that unique solutions $\hat{\theta}$ to $E[\underline{\psi}(\underline{x};\theta)] = \underline{0}$ exist, and are asymptotically normally distributed around the zero θ^* of $E[\underline{\psi}(\underline{x};\theta)]$. He conjectures that a certain set of conditions is sufficient to ensure finite sample uniqueness. Huber states, without proof, similar asymptotic results in the more general case.

It is a bit unclear what these estimators really would estimate in $U_{\varepsilon,\infty}$. Suppose that θ^* is to be of the form $(\underline{\mu},k\Sigma)$, where k is some unspecified scalar. We may assume that $(\underline{\mu},\Sigma) = (\underline{0},I)$, in which case the relationship $E[\underline{\psi}(\underline{x};\theta^*)] = \underline{0}$ becomes

$$(\underline{0},O_{m\times m}) = (E[\underline{\psi}_0(\underline{x};\theta^*)],E[\underline{\psi}_1(\underline{x};\theta^*)])$$

$$= (E[k^{-\frac{1}{2}}a_0(k^{-1}\underline{x}'\underline{x})\underline{x}],E[k^{-1}a_2(k^{-1}\underline{x}'\underline{x})\underline{x}\underline{x}'-a_4(k^{-1}\underline{x}'\underline{x})I])$$

$$= (\underline{0},E[(mk)^{-1}\underline{x}'\underline{x}a_2(k^{-1}\underline{x}'\underline{x})-a_4(k^{-1}\underline{x}'\underline{x})]I).$$

With $\underline{x}'\underline{x} = Z \sim f_m(z)$, and $a_1(z) = za_2(z)$, we then have that $\hat{\theta}$ is consistent for $(\underline{\mu},k\Sigma)$ iff

$$E[a_1(k^{-1}Z)-ma_4(k^{-1}Z)] = 0. \tag{18}$$

If $a_4 \equiv 1$ and a_1 is non-decreasing, as in Maronna's development, it seems clear that for any $f_m(z)$ there exists $k = k(f_m) > 0$ satisfying (18). Also, one can set $k = 1$ and then determine a_4 so that (18) is satisfied for a particular, known f_m - i.e. $\hat{\theta}$ can be made consistent for $(\underline{\mu},\Sigma)$ in $U_{0,\infty}$, where there is no problem with identifiability. But global consistency throughout $\mathcal{U}_{\varepsilon,\infty}$, for the same multiple of Σ, or any multiple if a_4 has a more general form, seems unrealistic.

Indeed, in Huber's consideration of $\mathcal{U}_{\varepsilon,\infty}$ (his "$\mathcal{F}$"), with $w(\underline{x};\underline{\mu},\Sigma) = \phi(\underline{x};\underline{\mu},\Sigma)$ and $(\underline{\mu},\Sigma)$ assumed to be $(\underline{0},I)$, he sets $a_4 \equiv 1$ and states that the optimum choice of a_1 has minimax properties with respect to the variance of $\hat{\theta}$ "for that subset of $\mathcal{F}$ for which $(\hat{\theta})$ is a consistent estimator of the identity matrix". The problem of identifiability is not considered.

4. ON MINIMIZING THE MAXIMUM ASYMPTOTIC VARIANCE OF ESTIMATORS IN THE ε-CONTAMINATION MODEL*

In the preceding section, we showed that the location and scale components of the estimator $\sqrt{n}\hat{\theta}$ are asymptotically independent, with asymptotic covariance structures $V(a_i,u)\cdot M_i$; $i = 0$ (location), $i = 1$ (scale). Here, M_i is some matrix depending only upon the unknown parameters, and $V(a_i,u)$ is a scalar functional (see 3.14).

In Section 5 we shall give the solutions to the problems of finding functions a_0,a_1 which minimize the maximum asymptotic variance of $\sqrt{n}\hat{\theta}$ over $\mathcal{U}_{\varepsilon,r}$. This is done by exhibiting functions (a_i^*,u_i^*) satisfying the "saddlepoint property"

$$V(a_i^*,u) \leq V(a_i^*,u_i^*) \leq V(a_i,u_i^*)$$

for all continuous, piecewise smooth functions $a_i(\underline{y}'\underline{y})$ vanishing off of $[0,r^2]$ and all $u(\underline{y};\underline{0},I) \in \mathcal{U}_{\varepsilon,r}$.

In this section we give some results on the minimax problem in the general ε-contamination model. Special applications of these results include the solutions to the problems outlined above. We also apply the results to the analogous problem for univariate estimation of location in some, hitherto uninvestigated, cases in which the underlying ideal density is not strongly, or even weakly, unimodal.

We consider functionals of the form

---
*See also the Appendix.

$$V(\psi,f) = \frac{\int_{-\infty}^{\infty} \psi^2(x)f(x)\eta(x)\nu(x)dx}{\left[\int_{-\infty}^{\infty} \psi(x)f'(x)\eta(x)\nu(x)dx\right]^2}.$$

Notation and Definitions

i) $\psi \in \Psi \subseteq \{\psi: R\setminus\{0\} \to R \mid \psi$ is skew-symmetric, continuous, and piecewise smooth$\}$

$\Psi_r = \{\psi\in\Psi \mid \psi(x)=0 \text{ for } |x|\geq r\}$.

ii) $\eta,\nu: R \to [0,\infty)$ are symmetric, and are positive and smooth on $(0,\infty)$.

$\sigma(x) = \eta(x)\nu(x)$.

iii) G is that subset of the set of symmetric, non-negative, absolutely continuous functions on R satisfying

$$\int_{-\infty}^{\infty} g(x)\nu(x)dx \leq 1 \text{ for all } g \in G.$$

The closure, with respect to the vague topology, of the set

$$G_\nu = \{G_\nu(t) = \int_{-\infty}^{t} g(x)\nu(x)dx \mid g\in G\}$$

contains those distributions with all mass at $\pm k$, for all $k \in [-\infty,\infty]$. In particular, $\overline{G}_\nu$ contains the distribution $G_\nu(t) \equiv 1/2$, with all mass at $\pm\infty$. Statistically this corresponds to rejecting a proportion ε of a data set upon a first screening.

iv) $f \in F = \{f | f=(1-\varepsilon)h+\varepsilon g;\ h,g \in G,\ \varepsilon \in [0,1],\ \varepsilon$ and h known and fixed$\}$.

As at (3.11) of Huber [4], the set $F_\nu = \{(1-\varepsilon)H_\nu+\varepsilon G_\nu | G_\nu \in G_\nu\}$ is convex and vaguely compact.

Example 1: $\eta(x) = \nu(x) = \sigma(x) \equiv 1$ gives the standard univariate location estimation functional, with G the set of symmetric pdf's on R. In the case $h(x) = (2\pi)^{-\frac{1}{2}}\exp(-x^2/2)$, see Huber [6] for Ψ, Collins [1] for Ψ_r.

Example 2: For m a fixed positive integer, and $g \in G$, put $g^m(x) = mC_m|x|^{m-1}g(x)$. Assume that $\int_0^\infty g^m(x)dx \leq 1$ for all $g \in G$. Then $\{g(|\underline{x}|) | g \in G, \underline{x} \in R^m\}$ is the set of symmetric, possibly sub-stochastic densities on R^m, and $\int_{-\infty}^\infty g(x)\nu(x)dx \leq 1$ for all $g \in G$, if $\nu(x) = \frac{mC_m}{2}|x|^{m-1}$. With $\psi_i(x) = x^{1-2i}a_i(x^2) \in \Psi_r$ and $\eta_i(x) = \left(\frac{2}{m+2}\right)^i \frac{x^{2i}}{m}$, we then have $V(\psi_i, f) = V(a_i, u)$, as at (3.15). If $\psi \in \Psi$ then, apart from differences in parameterization, the functionals $V(a_i, u)$ coincide with those of Huber [5].

For $f \in F$, define $I(f) = \int_{-\infty}^\infty (f'/f)^2 f\sigma dx$. If $\nu = \eta = \sigma \equiv 1$, then $I(f)$ is the Fisher information of f. Assume that $I(h) < \infty$. Then the results of Huber [3] go through essentially unchanged. For convenience of presentation we state these results as a lemma.

<u>LEMMA 4.1</u> (Huber) There is a unique $f_0 \in F$ minimizing $I(f)$ over F. Necessary and sufficient for f_0 to minimize $I(f)$ is that

$$\int_{-\infty}^\infty (2(\psi_0\sigma)' - \psi_0^2\sigma)\, l\, dx \geq 0 \qquad (1)$$

for all $l = f-f_0$, $f \in F$, where $\psi_0 = -\frac{f_0'}{f_0}$. If (1) holds and if $\psi_0 \in \Psi$, then (ψ_0,f_0) possesses the saddlepoint property:

$$V(\psi_0,f) \leq \frac{1}{I(f_0)} = V(\psi_0,f_0) \leq V(\psi,f_0) \qquad (2)$$

for all $\psi \in \Psi$ and $f \in F$.

Conversely, if (2) is satisfied by $(\psi_0,f_0) = \left(-\frac{f_0'}{f_0},f_0\right) \in \Psi\times F$, then f_0 minimizes $I(f)$ over F. □

Condition (1) is more conveniently written as

$$\int_0^\infty [2\psi_0'(x)-\psi_0^2(x)+2\psi_0(x)\frac{\sigma'}{\sigma}(x)]\eta(x)\nu(x)(g-g_0)(x)dx \geq 0 \qquad (3)$$

for all $g \in G$, where $f_0 = (1-\varepsilon)h+\varepsilon g_0$.

Define

$$J(\psi)(=J(\psi(x))) = (2\psi'-\psi^2+2\psi\frac{\sigma'}{\sigma})(x)\eta(x),$$

$$\zeta(x) = -\frac{h'}{h}(x).$$

Then (3) becomes $\int_0^\infty J(\psi_0)\nu(g-g_0)dx \geq 0$. The essential ingredient of Theorem 4.2 below is that ψ_0 is pieced together using ζ, where $J(\zeta)$ is relatively large, and another function ξ satisfying $J(\xi) = \text{const.}$, where $J(\zeta)$ is small.

Assume: A1) ζ is smooth,

A2) $\inf_{(0,\infty)} J(\zeta) < 0$.

If A2) fails, put $g_0 \equiv 0$, $\psi_0 \equiv \zeta$, $f_0 \equiv (1-\varepsilon)h$. Then (3) is satisfied and we are through.

<u>THEOREM 4.2</u> In order that $(\psi_0,f_0) = \left(-\frac{f_0'}{f_0},f_0\right) \in \Psi\times F$ satisfy (2), it is necessary and sufficient that there exist $\lambda \geq 0$, and a set

$$B_\lambda = \bigcup_{j=1}^{N(\lambda)} B_{\lambda,j} \subseteq (0,\infty),$$

where $N(\lambda) \leq \infty$ and the $B_{\lambda,j}$ are non-overlapping open intervals, such that:

i) $$\psi_0(x) = -\psi_0(-x) = \begin{cases} \zeta(x), \ x\in B_\lambda^c = (0,\infty)\backslash B_\lambda, \\ \xi(x;\omega_j,\lambda), \ x \in B_{\lambda,j}, \end{cases}$$

where $J(\xi(x;\omega_j,\lambda)) \equiv -\lambda$ for any fixed $\omega_j \in R$;

ii) $$f_0(x) = f_0(-x) = \begin{cases} (1-\varepsilon)h(x), & x \in B_\lambda^c, \\ (1-\varepsilon)\left(\sup\limits_{B_{\lambda,j}} \dfrac{h(x)}{k(x;\omega_j,\lambda)}\right)k(x;\omega_j,\lambda), & x \in B_{\lambda,j}, \end{cases}$$

where each k satisfies $\xi = -\frac{k'}{k}$ and $\sup\limits_{B_{\lambda,j}} \frac{h}{k}(x)$ is attained at each non-zero, finite endpoint of $B_{\lambda,j}$;

iii) $\int_{B_\lambda} \nu g_0 dx \leq \frac{1}{2}$, with equality if $\lambda > 0$, where $g_0 = \dfrac{f_0-(1-\varepsilon)h}{\varepsilon}$;

iv) $A_\lambda = \{x\in(0,\infty)\,|\,J(\zeta)<-\lambda\} \subseteq B_\lambda$;

v) $\psi_0 \in \Psi$;

vi) $B_{\lambda,j} \cap A_\lambda \neq \emptyset$ for all j.

Proof:

Sufficiency: By i), ii) and v), $\psi_0 = -\dfrac{f_0'}{f_0} \in \Psi$. In particular, ψ_0 is continuous. By the second part of ii), f_0 is continuous, hence so is f_0'. Then (ψ_0, f_0) satisfies (2) as long as (3) holds. Write this as

$$\int_0^\infty J(\psi_0)\nu(g-g_0)dx \geq 0 \text{ for all } g \in G. \tag{4}$$

Using i) and ii), then i) and iv), then iii), the integral is

$$2\left[\int_{B_\lambda} J(\xi)\nu(g-g_0)dx + \int_{B_\lambda^c} J(\zeta)\nu g dx\right] \geq -2\lambda\left[\int_{B_\lambda} \nu(g-g_0)dx + \int_{B_\lambda^c}\nu g dx\right]$$

$$= 2\lambda\left[\int_{B_\lambda} \nu g_0 dx - \int_0^\infty \nu g dx\right] \geq 0.$$

Thus i) - v) alone are sufficient.

Necessity: Let $(\psi_0, f_0) = \left(-\dfrac{f_0'}{f_0}, f_0\right) \in \Psi\times F$ satisfy (2). Then $f_0 = (1-\varepsilon)h+\varepsilon g_0$ for some $g \in G$, and (4) above holds. Define $B = \{x\in(0,\infty)\,|\,g_0(x)>0\}$, $\frac{\lambda}{2} = -\int_B J(\psi_0)\nu g_0 dx$. Note that $B \neq \emptyset$, since otherwise $f_0 \equiv (1-\varepsilon)h$, $\psi_0 \equiv \zeta$, and then (4) becomes $\int J(\zeta)\nu g dx \geq 0$ for all $g \in G$, contradicting A2). From (4),

$$0 \leq \int_B J(\psi_0)\nu(g-g_0)dx + \int_{B^c} J(\psi_0)\nu g dx = \frac{\lambda}{2} + \int_0^\infty J(\psi_0)\nu g dx, \tag{5}$$

so that

$$J(\psi_0) \geq -\lambda \text{ on } (0,\infty). \tag{6}$$

If $\lambda < 0$, then (5) is contradicted by, say, $g = \frac{1}{2}g_0$. Thus $\lambda \geq 0$, so that (6) and the definition of λ imply that $J(\psi_0) \equiv -\lambda$ on $\mathcal{B}$, whence $\int_B \nu g_0 dx = \frac{1}{2}$ if $\lambda > 0$.

Define $B_\lambda = B$, and represent B_λ as $B_\lambda = \bigcup_{j=1}^{N(\lambda)} B_{\lambda,j}$, where $N(\lambda) \leq \infty$ and the $B_{\lambda,j}$ are non-overlapping open intervals. On $B_{\lambda,j}$, ψ_0 satisfies the differential equation $J(\psi_0) \equiv -\lambda$. Since this equation has a solution passing through any pre-specified point, on $B_{\lambda,j}$ ψ_0 must have the form specified in i).

By the definition of B_λ, $f_0 = (1-\varepsilon)h$ on B_λ^c. On $B_{\lambda,j}$, $\psi_0 = \xi(x;\omega_j,\lambda)$, so that $f_0(x) = \alpha_j k(x;\omega_j,\lambda)$, where k is any function satisfying $-\frac{k'}{k} = \xi$, and $\alpha_j \in R$. Since $f_0 \geq (1-\varepsilon)h$ on $\overline{B}_{\lambda,j}$, $\alpha_j \geq (1-\varepsilon)\sup_{B_{\lambda,j}} \frac{h}{k}(x)$. The continuity of f_0 then forces equality in this last relation, and forces the sup to be attained at each non-zero, finite endpoint of $B_{\lambda,j}$. Thus f_0 has the specified form on $B_{\lambda,j}$.

Condition iv) is trivially satisfied if $B_\lambda^c = \emptyset$. If $B_\lambda^c \neq \emptyset$ and $\lambda > 0$, write (4) as

$$0 \leq \int_{B_\lambda} (-\lambda)\nu(g-g_0)dx + \int_{B_\lambda^c} J(\zeta)\nu g dx = \lambda\left(\frac{1}{2} - \int_{B_\lambda} \nu g dx\right) + \int_{B_\lambda^c} J(\zeta)\nu g dx.$$

This implies that $\inf_{B_\lambda^c} J(\zeta) \geq -\lambda$, so that $A_\lambda \subseteq B_\lambda$. This follows in a similar manner if $\lambda = 0$.

Thus conditions i) - v) are necessary as well as sufficient. To see that vi) is also necessary, extend F to F', G to G' by admitting piecewise absolutely continuous functions g. By Huber's Theorem 4 [3], there is a unique $f^* \in F'$ minimizing $I(f)$ over F'. By assumption, f_0 satisfies (3) for $g \in G$, and hence for $g \in G'$ as well since G is dense in G'. Thus $f_0 = f^*$. Suppose that vi) fails. Then there exists some $B_{\lambda,j}$, say $B_{\lambda,1}$, with $J(\zeta) \geq -\lambda$ on $B_{\lambda,1}$. We will obtain a contradiction by constructing a member f_1 of F', $f_1 \neq f_0$, such that $\int J(\psi_1)\nu(g-g_1) \geq 0$ for $g \in G'$, where $\psi_1 = -\frac{f_1'}{f_1} \in \Psi$ and $f_1 = (1-\varepsilon)h+\varepsilon g_1$. We assume that $N(\lambda) > 1$ - the method of proof is easily adapted to the exceptional case. Put

$$a = 1 + \frac{\varepsilon\int_{B_{\lambda,1}} \nu g_0 \, dx}{\int_{B_\lambda \setminus B_{\lambda,1}} \nu f_0 \, dx},$$

$$g_1(x) = \begin{cases} 0, & x \in B_{\lambda,1} \cup B_\lambda^c, \\ g_0(x) + \frac{a-1}{\varepsilon} f_0(x), & x \in B_\lambda \setminus B_{\lambda,1}, \end{cases}$$

$$f_1(x) = \begin{cases} (1-\varepsilon)h(x), & x \in B_{\lambda,1} \cup B_\lambda^c, \\ af_0, & x \in B_\lambda \setminus B_{\lambda,1}. \end{cases}$$

Then $f_1 \in F'$, $\psi_1 \in \Psi$, $\int_{B_\lambda \setminus B_{\lambda,1}} \nu g_1 dx = \int_{B_\lambda} \nu g_0 dx$, and for any $g \in G'$,

$$\int J(\psi_1)\nu(g-g_1)dx = \int_{B_{\lambda,1} \cup B_\lambda^c} J(\zeta)\nu g dx + \int_{B_\lambda \setminus B_{\lambda,1}} J(\psi_1)\nu(g-g_1)dx$$

$$\geq -\lambda \int_{B_{\lambda,1} \cup B_\lambda^c} \nu g dx - \lambda \int_{B_\lambda \setminus B_{\lambda,1}} \nu(g-g_1)dx$$

$$= \lambda \left[\int_{B_\lambda \setminus B_{\lambda,1}} \nu g_1 dx - \int_0^\infty \nu g dx \right] \geq 0,$$

implying $f_1 = f^*$. This contradiction completes the proof. □

Theorem 4.2 remains valid for Ψ_r, with minor modifications. Note that neither A2) nor vi) are used in the proof of sufficiency. In fact, the manner in which the failure of A2) is circumvented can lead out of the class Ψ_r, and the proof of necessity of vi) is likewise invalid if $B_{\lambda,1}$ is the right-most interval of $B_\lambda \cap (0,r)$.

When we later consider this class we will not make use of A2), and will verify directly that conditions i) - v) are satisfied.

We have

COROLLARY 4.3 Make the following changes in the statement of Theorem 4.2. Define $I_r(f) = \int_{-r}^{r} (f'/f)^2 f\sigma dx$, replace $(0,\infty)$ by $(0,r)$ and Ψ by Ψ_r throughout, and replace vi) by

vi)': $B_{\lambda,j} \cap A_\lambda \neq \emptyset$ for all $B_{\lambda,j}$ with $r \notin \overline{B}_{\lambda,j}$.

Then under assumption A1), conditions i) - v) are sufficient. Under A1) and A2), i) - vi)' are necessary. □

The conditions of Theorem 4.2 are strong enough to determine the pair (ψ_0, f_0) in two classes of special cases - for $\psi \in \Psi$, if the region A_λ which requires contamination is a single interval; and for $\psi \in \Psi_r$, if h is assumed unimodal.

<u>*Case* I: $\psi \in \Psi$, A_λ *connected.*</u>

In order to be able to use the existence of the optimal pair, we extend the class Ψ, if necessary, to $\Psi' = \{-\frac{f'}{f}; f \in F\}$. By Lemma 4.1, there exists $f_0 \in F$ minimizing $I(f)$, and by the above extension, $(\psi_0, f_0) \in \Psi' \times F$. Then (ψ_0, f_0) must satisfy the conditions of Theorem 4.2.

Suppose that A_λ is a single interval (c_λ, d_λ), necessarily non-empty by condition vi). Then by iv) and vi), B_λ must be a single interval $(a_\lambda, b_\lambda) \supseteq (c_\lambda, d_\lambda)$. If $a_\lambda \neq 0$ $(b_\lambda \neq \infty)$, it must be the largest (smallest) zero of $\xi - \zeta$ to the left (right) of $c_\lambda (d_\lambda)$, since extending the support of g_0 beyond this point would contradict vi).

The problem is then reduced to a numerical one of determining two constants (ω, λ). We have:

<u>THEOREM 4.4</u> Suppose that A1) and A2) hold, and that for all $\lambda \geq 0$, A_λ is a single interval (c_λ, d_λ). Then there exists a unique pair $(\omega, \bar{\lambda})$, where $\omega \in [-\infty, \infty]$ and $0 \leq \bar{\lambda} \leq \inf\{\lambda | A_\lambda = \emptyset\}$, such that the pair (ψ_0, f_0), defined below, satisfies the given conditions.

$$\text{i)}\quad \psi_0(x) = -\psi_0(-x) = \begin{cases} \zeta(x), & x \in (0,a_{\bar\lambda}]\cup[b_{\bar\lambda},\infty), \\ \xi(x;\omega,\bar\lambda), & x \in [a_{\bar\lambda},b_{\bar\lambda}]; \end{cases}$$

where $J(\xi) \equiv -\bar\lambda$,

$$a_{\bar\lambda} = \sup\{x \le c_{\bar\lambda} \mid (\xi-\zeta)(x)=0\},\ (\sup\emptyset = 0),$$

$$b_{\bar\lambda} = \inf\{x \ge d_{\bar\lambda} \mid (\xi-\zeta)(x)=0\},\ (\inf\emptyset = \infty).$$

$$\text{ii)}\quad f_0(x) = f_0(-x) = \begin{cases} (1-\varepsilon)h(x), & x \in (0,a_{\bar\lambda}]\cup[b_{\bar\lambda},\infty), \\ (1-\varepsilon)sk(x), & x \in [a_{\bar\lambda},b_{\bar\lambda}]; \end{cases}$$

where $\xi = -\dfrac{k'}{k}$ and

$$s = \sup_{[a_{\bar\lambda},b_{\bar\lambda}]} \frac{h}{k}(x) = \begin{cases} \dfrac{h}{k}(a_{\bar\lambda}) & \text{if } b_{\bar\lambda} = \infty, \\ \dfrac{h}{k}(b_{\bar\lambda}) & \text{if } a_{\bar\lambda} = 0, \\ \dfrac{h}{k}(a_{\bar\lambda}) = \dfrac{h}{k}(b_{\bar\lambda}) & \text{if } 0 < a_{\bar\lambda} < b_{\bar\lambda} < \infty. \end{cases}$$

$$\text{iii)}\quad \frac{1-\varepsilon}{\varepsilon}\int_{a_{\bar\lambda}}^{b_{\bar\lambda}} \nu(x)[sk(x)-h(x)]\,dx \le \frac{1}{2}, \text{ with equality if } \bar\lambda > 0.$$

If $\psi_0 \in \Psi$, then (ψ_0,f_0) possesses the saddlepoint property:

$$V(\psi_0,f) \le \frac{1}{I(f_0)} = V(\psi_0,f_0) \le V(\psi,f_0)$$

for all $\psi \in \Psi$ and $f \in \mathrm{F}$. □

Remarks:

1) The condition that $a_{\bar\lambda}$ and $b_{\bar\lambda}$ simultaneously maximize $\frac{h}{k}$, if $0 < a_{\bar\lambda} < b_{\bar\lambda} < \infty$, appears to be rather stringent. However, the existence of such a pair is assured by the existence of (ψ_0, f_0). If $a_{\bar\lambda} = 0$ or $b_{\bar\lambda} = \infty$, then the condition is quite natural. It follows from

$$\xi - \zeta = -\frac{k'}{k} + \frac{h'}{h} = \left(\frac{h}{k}\right)'\frac{k}{h}$$

that the zeroes of $\xi - \zeta$ are critical points of $\frac{h}{k}$, and are local maxima if $\xi - \zeta$ is decreasing through zero at these points. Writing A_λ as

$$A_\lambda = \{x \in (0,\infty) \,|\, J(\xi) - J(\zeta) = \frac{2\sigma'}{\sigma}(\xi-\zeta) + 2(\xi-\zeta)' - (\xi^2-\zeta^2) > 0\}, \qquad (7)$$

we see that these local maxima are precisely the zeroes of $\xi - \zeta$ in the interior of A_λ^c. Note also that if $0 < a_{\bar\lambda} < b_{\bar\lambda} < \infty$, then by (7), $\xi - \zeta$ must have a third zero inside $(c_{\bar\lambda}, d_{\bar\lambda})$.

2) It follows from $J(\xi) \equiv -\lambda$ that ξ is decreasing at any zero. It thus changes sign at most once, from positive to negative. Since ψ_0 cannot be negative (f_0 increasing) on an unbounded interval, we see that in the cases where $b_{\bar\lambda} = \infty$, ξ satisfies $\xi > 0$ on $[a_{\bar\lambda}, \infty)$. In particular, if $b_{\bar\lambda} = \infty$ and $a_{\bar\lambda}$ is a zero of $\xi - \zeta$, $a_{\bar\lambda}$ must be a point of decrease of h $(\zeta(a_{\bar\lambda}) > 0)$, even if h is not unimodal. Similar considerations apply to the cases $0 = a_{\bar\lambda} < b_{\bar\lambda} < \infty$ and $0 < a_{\bar\lambda} < b_{\bar\lambda} < \infty$.

3) We have not found any non-pathological instances in which A_λ is <u>not</u> a single interval for all $\lambda \geq 0$ - but see Example 5 below. Huber, in his consideration of multivariate estimation of scale [5], constructs

an f_0 which, in certain instances, places contaminating mass on two disjoint intervals. However, the value of λ in these cases is negative. The pair (ψ_0, f_0) can then not be optimal, by Theorem 4.2, if G_ν contains certain sub-stochastic distributions. The details of this example, and what we believe to be the correct solution, are given in Section 5.

4) The solutions to $J(\xi) \equiv -\lambda$ are often rather unwieldy special functions. Although the determination of the minimax variance

$$V(\psi_0, f_0) = (I(f_0))^{-1} = \left(\int \psi_0^2 f_0 \sigma dx\right)^{-1}$$

would seem to require the integration of these functions, this is not the case. Integrating over $J(\xi) \equiv -\lambda$ allows one to eliminate the terms involving ξ from $I(f_0)$, leaving only an integral which is often either elementary, or whose value is obtainable from the tables of the d.f. H_ν, once the constants are determined.

With $A = (0, a_{\bar\lambda}] \cup [b_{\bar\lambda}, \infty)$, and $I_A(h) = \int_A \left(\frac{h'}{h}\right)^2 h\sigma dx$, we find

$$I(f_0) = 2(1-\varepsilon)\{I_A(h) - \bar\lambda[H_\nu(b_{\bar\lambda}) - H_\nu(a_{\bar\lambda})] + 2s[(k'\sigma)(b_{\bar\lambda}) - (k'\sigma)(a_{\bar\lambda})]\} - \bar\lambda\varepsilon.$$

If $a_{\bar\lambda}$ and $b_{\bar\lambda}$ are zeroes of $\xi - \zeta$, then $(sk'\sigma)(a_{\bar\lambda}) = (h'\sigma)(a_{\bar\lambda})$ and $(sk'\sigma)(b_{\bar\lambda}) = (h'\sigma)(a_{\bar\lambda})$, so that in these cases the dependence on k is eliminated as well.

Example 3: Consider the univariate location estimation problem $(\nu=\eta=\sigma\equiv 1)$, and assume that either

i) ζ is increasing (h strongly unimodal), or

ii) $J(\zeta)$ is decreasing where it is negative.

It is not difficult to see, by considerations as in Remarks 1) and 2), that under these conditions B_λ must be a single half-infinite interval. The solution to $J(\xi) = 2\xi' - \xi^2 \equiv -\lambda$ has several forms, depending upon the desired monotonicity properties:

$$\xi(x;\omega,\lambda) = \begin{cases} \sqrt{\lambda}\ \tanh\left[\frac{-\sqrt{\lambda}}{2}(x-\omega)\right], & \text{decreasing,} \\ \sqrt{\lambda}, & \\ \sqrt{\lambda}\ \coth\left[\frac{-\sqrt{\lambda}}{2}(x-\omega)\right], & \text{increasing.} \end{cases}$$

The "tanh" solution is eventually negative, and the "coth" solution is either always negative or has a pole, depending upon the value of ω. This leaves only the constant solution.

The optimal pair (ψ_0, f_0) is given by

$$\psi_0(x) = -\psi_0(-x) = \begin{cases} \zeta(x), & 0 < x \lesssim a, \\ \\ \zeta(a), & x > a; \end{cases}$$

$$f_0(x) = f_0(-x) = \begin{cases} (1-\varepsilon)h(x), & 0 < x \lesssim a, \\ \\ (1-\varepsilon)h(a)e^{-\zeta(a)(x-a)}, & x > a; \end{cases}$$

where a is determined from

$$\frac{\varepsilon}{2(1-\varepsilon)} = \frac{h(a)}{\zeta(a)} - 1 + H(a).$$

Minimum information is

$$I(f_0) = 2(1-\varepsilon)\left\{\int_0^a \left(\frac{h'}{h}\right)^2 h\, dx - h'(a)\right\}.$$

In the next example, neither of the above conditions is met.

Example 4: Put $\nu = \eta = \sigma = 1$. Let h be the Cauchy density $h(x) = (\pi(1+x^2))^{-1}$. The qualitative results are the same for any t-density. We find

$$\zeta(x) = \frac{2x}{1+x^2}, \quad J(\zeta) = \frac{4(1-2x^2)}{(1+x^2)^2},$$

$$A_\lambda = \begin{cases} (c_\lambda, d_\lambda), & \lambda < 4/3, \\ \emptyset, & \lambda \geq 4/3; \end{cases}$$

where $c_\lambda^2 = \frac{4-\lambda-2\sqrt{4-3\lambda}}{\lambda}$, $d_\lambda^2 = \frac{4-\lambda+2\sqrt{4-3\lambda}}{\lambda}$. The constant and "coth" solutions to $J(\xi) \equiv -\lambda$ violate $A_\lambda \subseteq B_\lambda$, leaving only the "tanh" solution. The possibility that $\lambda = 0$ is likewise untenable. For $\lambda > 0$, $\xi - \zeta$ has three zeroes a_λ, e_λ, b_λ; with $a_\lambda (>0)$, $b_\lambda (<\infty) \in A_\lambda^c$ and $e_\lambda \in A_\lambda$. Thus, for $\lambda \in (0,4/3)$ and $e_\lambda^2 \in \left(\frac{4-\lambda-2\sqrt{4-3\lambda}}{\lambda}, \frac{4-\lambda+2\sqrt{4-3\lambda}}{\lambda}\right)$, set $\omega = e_\lambda + \frac{2}{\sqrt{\lambda}} \tanh^{-1}\left(\frac{2e_\lambda}{\sqrt{\lambda}(1+e_\lambda^2)}\right)$, so that $\xi(e_\lambda;\lambda,\omega) = \zeta(e_\lambda)$. Let a_λ and b_λ be the other two zeroes of $\xi - \zeta$. Put $k(x;\omega,\lambda) = \cosh^2\left(\frac{-\sqrt{\lambda}}{2}(x-\omega)\right)$

so that $-\frac{k'}{k} = \xi$. Then by Theorem 4.4 there exists a unique pair $(\bar{\lambda}, e_{\bar{\lambda}})$ in the indicated region satisfying

$$s = \frac{h}{k}(a_{\bar{\lambda}}) = \frac{h}{k}(b_{\bar{\lambda}}) = \sup_{(a_{\bar{\lambda}}, b_{\bar{\lambda}})} \frac{h}{k}(x),$$

$$\frac{1-\varepsilon}{\varepsilon} \int_{a_{\bar{\lambda}}}^{b_{\bar{\lambda}}} [sk(x)-h(x)]dx = \frac{1}{2};$$

and the optimal pair is given by

$$\psi_0(x) = -\psi_0(-x) = \begin{cases} \dfrac{2x}{1+x^2}, & x \notin [a_{\bar{\lambda}}, b_{\bar{\lambda}}], \\ \sqrt{\bar{\lambda}} \tanh\left(\dfrac{-\sqrt{\bar{\lambda}}}{2}(x-\omega)\right), & x \in [a_{\bar{\lambda}}, b_{\bar{\lambda}}]; \end{cases}$$

$$f_0(x) = f_0(-x) = \begin{cases} \dfrac{1-\varepsilon}{\pi(1+x^2)}, & x \notin [a_{\bar{\lambda}}, b_{\bar{\lambda}}], \\ (1-\varepsilon)s \cdot \cosh^2\left(\dfrac{-\sqrt{\bar{\lambda}}}{2}(x-\omega)\right), & x \in [a_{\bar{\lambda}}, b_{\bar{\lambda}}]. \end{cases}$$

Upon performing some elementary calculations, we find that the constants $(\lambda, \omega, s, a, b)$ may be obtained as follows. Define λ and s as functions of a and b by

$$\lambda = \frac{4}{(b^2-a^2)}\left[\left(\frac{a}{1+a^2}\right)^2 (1+b^2) - \left(\frac{b}{1+b^2}\right)^2 (1+a^2)\right],$$

$$s = \frac{\lambda(1+a^2)^2 - 4a^2}{\pi\lambda(1+a^2)^3}.$$

Determine a and b from

$$\exp(\sqrt{\lambda}(b-a)) = \left(\frac{\sqrt{\lambda}(1+a^2)+2a}{\sqrt{\lambda}(1+b^2)+2b}\right)\left(\frac{\sqrt{\lambda}(1+b^2)-2b}{\sqrt{\lambda}(1+a^2)-2a}\right),$$

$$\frac{\varepsilon}{2(1-\varepsilon)} = \frac{2a}{\pi\lambda(1+a^2)^2} - \frac{2b}{\pi\lambda(1+b^2)^2} + \frac{s(b-a)}{2} - \frac{(\tan^{-1}(b)-\tan^{-1}(a))}{\pi}.$$

The value of ω may be calculated directly from

$$\omega = a + 2/\sqrt{\lambda}\ \tanh^{-1}\left(\frac{2a}{\sqrt{\lambda}(1+a^2)}\right).$$

Minimum information is

$$I(f_0) = I(h) - \frac{1-\varepsilon}{\pi}\left[\frac{b(b^2+7)}{(1+b^2)^2} - \frac{a(a^2+7)}{(1+a^2)^2}\right] - (1+2\lambda)\left[\frac{\varepsilon}{2} + \frac{1-\varepsilon}{\pi}(\tan^{-1}(b)-\tan^{-1}(a))\right],$$

where $I(h) = 2/\sqrt{\pi}$.

The next example illustrates several points.

i) Saddlepoint solutions to the minimax problem need not exist in $\Psi_r \times F$.

ii) The solutions in $\Psi \times F$ may lead to the use of ψ's which are <u>negative</u> for small, positive x.

iii) Even if A_λ is not a single interval for all λ, the least favourable density f_0 may place all of its contaminating mass on one interval, for certain values of ε.

Example 5: Put $\nu = \eta = \sigma \equiv 1$. Let $h(x) = \dfrac{x^4 e^{-|x|}}{48}$, $-\infty < x < \infty$. For $x > 0$, $\zeta(x) = 1 - \dfrac{4}{x}$ and $J(\zeta) = \dfrac{-x^2+8x-8}{x^2}$.

a) We show that, at least for $r < 4$, saddlepoint solutions do not exist in $\Psi_r \times F$. If $r < 4$, then $\zeta(x) < 0$ on $(0,r]$ and for all $\lambda \geq 0$, A_λ is an interval of the form $(0,d_\lambda)$. The optimal $\psi_0 \in \Psi_r$, if such exists, must then satisfy $J(\psi_0) \equiv -\lambda$ both near the origin, to have $A_\lambda \subseteq B_\lambda$, and in a neighborhood of r, to have $\psi_0(r) = 0$. Continuity considerations then dictate that

$$\psi_0(x) = \sqrt{\lambda}\ \tanh\left(\frac{-\sqrt{\lambda}}{2}(x-r)\right),\ \text{throughout } (0,r].$$

Then

$$f_0(x) = (1-\varepsilon)h(x)+\varepsilon g_0(x) = (1-\varepsilon)s\cdot\cosh^2\left(\frac{-\sqrt{\lambda}}{2}(x-r)\right),\ x \in (0,r],$$

for some s and g_0. The requirement that g_0 be a non-negative, proper density becomes

$$h(r) \leq s = \frac{\dfrac{\varepsilon}{1-\varepsilon} + \displaystyle\int_{-r}^{r} h(x)\,dx}{r + \dfrac{\sinh\ r\sqrt{\lambda}}{\sqrt{\lambda}}}.$$

The right hand member above is less than $\dfrac{\dfrac{\varepsilon}{1-\varepsilon} + \displaystyle\int_{-r}^{r} h(x)\,dx}{2r}$ for all $\lambda \geq 0$, and so it is necessary to have

$$\frac{\varepsilon}{1-\varepsilon} \geq \int_{-r}^{r} (h(r)-h(x))\,dx. \tag{8}$$

But these are precisely the values of ε which permit the existence of an $f_1 \in F$ with $I_r(f_1) = 0$.

Define

$$g_1(x) = \frac{1-\varepsilon}{\varepsilon}\left[\int_0^r \frac{h(x)}{r}dx - h(x)\right] + \frac{1}{2r} .$$

Then $\int_{-r}^{r} g_1(x)dx = 1$, $g_1(x)$ is non-negative on $[-r,r]$ iff (8) holds, and

$$f_1(x) = (1-\varepsilon)h(x)+\varepsilon g_1(x) \equiv \frac{(1-\varepsilon)\int_{-r}^{r} h(x)dx + \varepsilon}{2r}, \; x \in [-r,r],$$

has

$$I_r(f_1) = 0 \leq I_r(f_0) = \frac{\lambda(\sinh\ r\sqrt{\lambda} - r\sqrt{\lambda})}{\sinh\ r\sqrt{\lambda} + r\sqrt{\lambda}} \cdot \int_{-r}^{r} f_0(x)dx,$$

with equality iff $\lambda = 0$. Conversely, if (8) fails then constant density on $[-r,r]$ requires the contaminating mass to exceed unity.

Thus if ε is small enough that (8) fails, no $\psi \in \Psi_r$ can meet the optimality criteria; and if (8) holds, the optimal $(\psi_1,f_1) \in \Psi_r\times F$ is the trivial pair above, with $V(\psi_1,f_1) = \infty$.

b) The same problem is solvable in $\Psi\times F$ if Ψ contains certain members which are negative for small, positive values of x. If ε is less than (approximately) .025, and $a \in (0,2)$ is related to ε by

$$\frac{\varepsilon}{2(1-\varepsilon)} = \frac{h(a)}{\zeta(a)}\left[e^{a\zeta(a)}-1\right] + \frac{1}{2} - H(a),$$

then the solution is given by

$$\psi_0(x) = -\psi_0(-x) = \begin{cases} \zeta(a), & 0<x\leq a, \\ \zeta(x), & x\geq a; \end{cases}$$

$$f_0(x) = f_0(-x) = \begin{cases} (1-\varepsilon)h(a)e^{|\zeta(a)|(x-a)}, & 0<x\leq a, \\ (1-\varepsilon)h(x), & x\geq a. \end{cases}$$

For larger values of ε, it appears that ψ_0 is also constant, but positive, for large x.

Case II: $\psi \in \Psi_r$, h unimodal.

The results for Ψ_r are less complete than those for Ψ. In view of Example 5, we cannot posit the existence of an optimal $\psi_0 \in \Psi_r$ and then infer its properties from its existence. The side condition $\psi(r) = 0$, implying $r \in B_\lambda$, is not dictated by any optimality principle, unless as well $r \in A_\lambda$. We will thus restrict our attention to densities for which the condition $A_\lambda \subseteq B_\lambda$ forces $r \in B_\lambda$.

We assume

B1) ζ is smooth and positive on $(0,r)$, and bounded on $[0,r]$.

Let $\xi(x;\omega,\lambda) = \xi(x,\lambda)$ be the solution to

$$J(\xi) = (2\xi'-\xi^2+2\frac{\sigma'}{\sigma}\xi)(x)\eta(x) \equiv -\lambda,\ 0 \leq \lambda \leq \infty,$$

passing through $(r,0)$. Assume that σ and η are such that the following condition is satisfied.

B2) For fixed $\lambda > 0$, $\xi(x,\lambda)$ is strictly decreasing in $x \in (0,r)$.

The following lemma gives an easily checked condition ensuring that B2) is satisfied, and gives some further properties of $\xi(x,\lambda)$.

LEMMA 4.5 For fixed $x \in (0,r)$, $\xi(x,\lambda)$ is a continuously differentiable function of λ, with

i) $\xi(x,\lambda) > 0 \quad (\lambda>0)$,

ii) $\frac{d}{d\lambda}\xi(x,\lambda) > 0$,

iii) $\lim_{\lambda\to 0} \xi(x,\lambda) = 0$,

iv) $\lim_{\lambda\to\infty} \xi(x,\lambda) = \infty$.

Furthermore, assumption B2) is implied by the following condition.

B2') For fixed $\lambda > 0$, $\frac{\sigma'}{\sigma}(x) + \left(\left(\frac{\sigma'}{\sigma}(x)\right)^2 + \frac{\lambda}{\eta(x)}\right)^{\frac{1}{2}}$ is strictly decreasing in $x \in (0,r)$.

In particular, if $\frac{\sigma'}{\sigma}(x)$ is non-negative and decreasing, and $\eta(x)$ is non-decreasing, then $\xi(x,\lambda)$ is strictly decreasing in x for all $\lambda > 0$.

Proof: That ξ is a continuously differentiable function of λ is a standard result in the theory of differential equations. Since ξ is decreasing at any zero, i) is immediate. For ii), let $0 < \lambda_1 < \lambda_2 < \infty$. We claim that $\xi(x,\lambda_1) < \xi(x,\lambda_2)$ for all $x \in (0,r)$.

Since

$$\xi'(r,\lambda_2) = -\frac{\lambda_2}{2\eta(r)} < -\frac{\lambda_1}{2\eta(r)} = \xi'(r,\lambda_1) < 0,$$

there exists a maximal interval (r_0,r) on which $\xi(x,\lambda_1) < \xi(x,\lambda_2)$. Suppose that $r_0 > 0$. Then $\xi(r_0,\lambda_1) = \xi(r_0,\lambda_2)$ and $\xi'(r_0,\lambda_2) < \xi'(r_0,\lambda_1)$. Substituting these relations into $J(\xi(r_0,\lambda_1)) - J(\xi(r_0,\lambda_2)) = -\lambda_1 + \lambda_2 > 0$ yields the desired contradiction, and so $r_0 = 0$.

Now let $\lambda \to 0$. Since ξ is continuous in λ, so is ξ', hence $\xi(x,0)$ satisfies $J(\xi) \equiv 0$. Put $\xi = -2\frac{v'}{v}$, to get $(v'\sigma)' \equiv 0$. This gives

$$\xi(x,0) = \frac{2}{\sigma(x)\left\{c - \int^{x} (\sigma(x))^{-1} dx\right\}},$$

for some $c \in [-\infty,\infty]$. But the only function of this form satisfying $\xi(r,0) = 0$ is $\xi(x,0) \equiv 0$. This proves iii).

For iv), let $0 < x_0 < x_1 \leq r$, and integrate across $J(\xi) \equiv -\lambda$ to get

$$\xi(x_1,\lambda) - \xi(x_0,\lambda) = \int_{x_0}^{x_1} \xi'(x,\lambda)dx = \int_{x_0}^{x_1} [\tfrac{1}{2}\xi^2(x,\lambda) - \frac{\sigma'}{\sigma}(x)\xi(x,\lambda)]dx - \frac{\lambda}{2}\int_{x_0}^{x_1} (\eta(x))^{-1}dx.$$

Put

$$N_\lambda = \sup_{[x_0,x_1]} \left|\tfrac{1}{2}\xi^2(x,\lambda) - \frac{\sigma'}{\sigma}(x)\xi(x,\lambda)\right|,$$

$$M_\lambda = \sup_{[x_0,x_1]} \xi(x,\lambda), \quad \alpha = \sup_{[x_0,x_1]} \left|\frac{\sigma'}{\sigma}(x)\right|.$$

Then

$$0 < \frac{\lambda}{2}\int_{x_0}^{x_1} (\eta(x))^{-1}dx = \int_{x_0}^{x_1} [\tfrac{1}{2}\xi^2(x,\lambda) - \frac{\sigma'}{\sigma}(x)\xi(x,\lambda)]dx + (\xi(x_0,\lambda)-\xi(x_1,\lambda))$$

$$\leq (x_1-x_0)N_\lambda + M_\lambda \leq (x_1-x_0)[\tfrac{1}{2}M_\lambda^2+\alpha M_\lambda] + M_\lambda,$$

and so $M_\lambda \to \infty$ as $\lambda \to \infty$. Since x_0 and x_1 are arbitrary, the result follows by B2).

Now let $\lambda > 0$ be fixed, and assume that B2$'$) holds. If x_0 is any critical point of ξ, then

$$\xi(x_0,\lambda) = \frac{\sigma'}{\sigma}(x_0) + \left(\left(\frac{\sigma'}{\sigma}(x_0)\right)^2 + \frac{\lambda}{\eta(x_0)}\right)^{\frac{1}{2}},$$

so that the values of ξ at such points are decreasing. If $\lim_{x\to 0} \xi(x,\lambda) = \infty$, or $\xi(0,\lambda) < \infty$ but $\xi'(0,\lambda) < 0$, then since $\xi(r,\lambda) = 0$, any interval on which ξ is not strictly decreasing must have as endpoints critical points violating the previous statement. If $\xi(0,\lambda) < \infty$ and $\xi'(0,\lambda) \geq 0$, then there exists x_0, $0 < x_0 < r$, with $\xi(0,\lambda) \leq \xi(x_0,\lambda)$ and $\xi'(x_0,\lambda) = 0$; i.e.

$$\xi(x_0,\lambda) = \frac{\sigma'}{\sigma}(x_0) + \left(\left(\frac{\sigma'}{\sigma}(x_0)\right)^2 + \frac{\lambda}{\eta(x_0)}\right)^{\frac{1}{2}}$$

$$\geq \xi(0,\lambda) = \frac{\sigma'}{\sigma}(0) + \left(\left(\frac{\sigma'}{\sigma}(0)\right)^2 + 2\xi'(0,\lambda) + \frac{\lambda}{\eta(0)}\right)^{\frac{1}{2}}$$

$$\geq \frac{\sigma'}{\sigma}(0) + \left(\left(\frac{\sigma'}{\sigma}(0)\right)^2 + \frac{\lambda}{\eta(0)}\right)^{\frac{1}{2}},$$

contradicting B2$'$). □

Define, for $\lambda \geq 0$;

$$A_\lambda = \{x \in (0,r) \mid J(\zeta) < -\lambda\},$$

$$B_\lambda = \{x \in (0,r) \mid \xi(x,\lambda) < \zeta(x)\},$$

and note that $\{A_\lambda\}$ is non-increasing, and that $\{B_\lambda\}$ is strictly decreasing from $(0,r)$ to $\emptyset$ as λ increases from 0 to ∞.

Under certain conditions on ζ, condition iv) of Theorem 4.2 is satisfied by any pair (A_λ, B_λ).

<u>LEMMA 4.6</u> If either of the following conditions holds:

a) ζ is non-decreasing on $(0,r)$,

b) $J(\zeta)$ is non-increasing on A_0;

then $A_\lambda \subseteq B_\lambda$ for all $\lambda \geq 0$ and B_λ is a single interval (a_λ, r).

<u>Proof:</u> Suppose that the first part of the conclusion is false. Put

$$\lambda_0 = \inf\{\lambda \mid A_\lambda \setminus B_\lambda \neq \emptyset\}.$$

We first show that $\lambda_0 > 0$. If $\lambda_0 = 0$, then there exists a sequence $x_\lambda \in A_\lambda \setminus B_\lambda$ as $\lambda \downarrow 0$. Note that B_λ contains an interval of the form (a_λ, r) for all λ, since $\zeta(r) > 0 = \xi(r,\lambda)$, and that $a_\lambda \downarrow 0$ as $\lambda \downarrow 0$. Thus $x_\lambda \downarrow 0$. Let (c_λ, d_λ) be the largest interval in A_λ containing x_λ. Then $c_\lambda < x_\lambda \leq a_\lambda$. As $\lambda \downarrow 0$, c_λ is non-increasing, d_λ is non-decreasing, and $a_\lambda \downarrow 0$. Thus $a_\lambda \in A_\lambda$ for some $\lambda > 0$.

Since a_λ is a boundary point of B_λ, $\xi(a_\lambda,\lambda) - \zeta(a_\lambda) = 0$. This, together with $a_\lambda \in A_\lambda$, represented as at (7), implies that $(\xi-\zeta)'(a_\lambda) > 0$, so that $\xi-\zeta$ is increasing through zero at a_λ. But this contradicts $a_\lambda^+ \in B_\lambda$.

Thus $\lambda_0 > 0$. Let $\lambda \in (0,\lambda_0)$ be arbitrary. Then $A_\lambda \subseteq B_\lambda$. Since $\xi(r,\lambda) - \zeta(r) < 0$ for all $\lambda > 0$, and since both A_λ and B_λ are open, we can represent them as

$$B_\lambda = \left[\biguplus_{j=1}^{N(\lambda)} (a_\lambda^j, b_\lambda^j)\right] \uplus (a_\lambda, r),$$

where $N(\lambda) \lesssim \infty$ and the $a_\lambda^j(b_\lambda^j)$ are strictly increasing (decreasing) functions of λ; and

$$A_\lambda = \left[\biguplus_{j=1}^{M(\lambda)} \biguplus_{l=1}^{k_j} (c_\lambda^{j,l}, d_\lambda^{j,l})\right] \uplus \left[\biguplus_{l=1}^{k} (c_\lambda^l, d_\lambda^l)\right],$$

where $M(\lambda) \lesssim N(\lambda)$, $\bigcup_{l=1}^{k_j} (c_\lambda^{j,l}, d_\lambda^{j,l}) \subseteq (a_\lambda^j, b_\lambda^j)$, and $\bigcup_{l=1}^{k} (c_\lambda^l, d_\lambda^l) \subseteq (a_\lambda, r)$.

We claim that for $\lambda < \lambda_0$, $N(\lambda) = 0$. Suppose that $N(\lambda) > 0$. Let (a_λ, b_λ) be any set in the first union. Since b_λ is a boundary point of B_λ, $\xi(b_\lambda,\lambda) - \zeta(b_\lambda) = 0$. Since $b_\lambda \notin A_\lambda$, $(\xi-\zeta)'(b_\lambda) \lesssim 0$ (see (7)). If this last inequality is strict, then $\xi-\zeta$ is decreasing through zero at b_λ, contradicting $b_\lambda^- \in B_\lambda$. Thus

$$(\xi-\zeta)(b_\lambda) = (\xi-\zeta)'(b_\lambda) = 0 \text{ for } \underline{\text{all}}\ \lambda < \lambda_0.$$

Fix $\lambda_1 \in (0,\lambda_0)$, and put $G(x,\lambda) = \xi(x,\lambda) - \zeta(x)$ for $\lambda \in (0,\lambda_0)$. Then $G(b_{\lambda_1},\lambda_1) = 0$, the second partial derivative $G_2(x,\lambda)$ is continuous and positive, and $G_1(x,\lambda)$ exists and is continuous, in a neighborhood S of $(\lambda_1,b_{\lambda_1})$. The implicit function theorem then applies: within S, λ is a continuously differentiable function of b, with

$$\lambda'(b) = -\left.\frac{G_1(x,\lambda)}{G_2(x,\lambda)}\right|_{(x,\lambda)=(b_\lambda,\lambda)} \equiv 0.$$

But this implies that λ is independent of b throughout S, contradicting the fact that b is a monotonic, hence invertible, function of λ.

Thus $N(\lambda) = 0$ for $\lambda < \lambda_0$; i.e. for $\lambda < \lambda_0$,

$$B_\lambda = (a_\lambda,r) \supseteq A_\lambda = \bigcup_{l=1}^{k} (c_\lambda^l,d_\lambda^l).$$

As λ increases beyond λ_0, the inclusion then fails in one of two ways:

i) a_λ "moves inside" c_λ^1 - a possibility ruled out in the proof that $\lambda_0 > 0$;

ii) $N(\lambda)$ jumps to 1 due to the removal of a point from B_λ, and failure occurs immediately thereafter. In this case, $B_{\lambda_0} = (a_\lambda^1,b_{\lambda_0}^1) \cup (a_{\lambda_0},r)$, with $b_{\lambda_0}^1 = a_{\lambda_0}$; and a_{λ_0} is both a zero and a local maximum of $\xi-\zeta$; i.e.

$$(\xi-\zeta)(a_{\lambda_0}) = (\xi-\zeta)'(a_{\lambda_0}) = 0.$$

But this contradicts condition a), using B2). In any event, a_{λ_0} is then a boundary point of A_{λ_0}. If b) holds, then for all $\lambda \geqslant 0$, A_λ is of the form (c_λ, r). Thus at $\lambda = \lambda_0$,

$$A_\lambda = (c_\lambda, r) \subseteq (a_\lambda^1, b_\lambda^1) \cup (a_\lambda, r) = B_\lambda,$$

where $a_{\lambda_0} = b_{\lambda_0}^1 = c_{\lambda_0}$ (since a_{λ_0} is a boundary point of A_{λ_0}). But as λ increases from λ_0, c_λ is non-decreasing, b_λ^1 is decreasing, and a_λ is increasing; so that, again, failure can only occur if a_λ moves inside A_λ - a contradiction.

Thus $\lambda_0 = \infty$, $N(\lambda) = 0$ for all λ, and $A_\lambda \subseteq B_\lambda$ for all λ. □

The statistical problem of minimaxing $V(\psi, f)$ over $\Psi_r \times F$ is meaningless unless the least informative member f_0 of F has $I_r(f_0) > 0$. Suppose that

$$I_r(f_0) = \int_{-r}^{r} \left(\frac{f_0'}{f_0}\right)^2 f_0 \sigma dx = 0.$$

Then on $[-r,r]$, $f_0' \equiv 0$, so that $f_0(x) \geqslant (1-\varepsilon)h(0)$, and

$$\int_{-r}^{r} \nu g_0 dx \geqslant \frac{1-\varepsilon}{\varepsilon} \int_{-r}^{r} \nu(x)[h(0)-h(x)]dx.$$

If ε is sufficiently small, then this last term exceeds unity, contradicting $g_0 \in G$. This bound on ε, say ε^*, is clearly also necessary, and so $I_r(f_0) > 0$ iff $\varepsilon < \varepsilon^*$. Although $\psi_0 = -\dfrac{f_0'}{f_0}$ will not, in general, belong to Ψ_r, it is easy to construct a sequence $\{\psi_n\} \subseteq \Psi_r$, with $\psi_n \to \psi_0$

pointwise on $(-r,r)$, such that

$$V(\psi_n,f_0) \to V(\psi_0,f_0) = \frac{1}{I(f_0)} < \infty.$$

We thus have

LEMMA 4.7 In order that $\inf_{\Psi_r} \sup_{F} V(\psi,f)$ be finite, it is necessary and sufficient that ε satisfy $\varepsilon < \varepsilon^*$, where

$$\frac{1-\varepsilon^*}{\varepsilon^*} \int_{-r}^{r} \nu(x)[h(0)-h(x)]dx = 1. \qquad \square$$

Now assume that either condition of Lemma 4.6 holds, so that B_λ is of the form (a_λ, r) for all $\lambda > 0$. Let $k(x,\lambda)$ satisfy $\xi = -\frac{k'}{k}$. Define

$$\psi(x,\lambda) = -\psi(-x,\lambda) = \begin{cases} \zeta(x), & 0 < x \leq a_\lambda, \\ \xi(x,\lambda), & a_\lambda \leq x \leq r, \\ 0, & x \geq r; \end{cases}$$

$$f(x,\lambda) = f(-x,\lambda) = \begin{cases} (1-\varepsilon)h(x), & 0 \leq x \leq a_\lambda, \\ (1-\varepsilon)\frac{h(a_\lambda)}{k(a_\lambda,\lambda)}\, k(x,\lambda), & a_\lambda \leq x \leq r, \\ 0, & x > r. \end{cases}$$

It follows as in Remark 1) in Case I that $\sup_{B_\lambda} \frac{h}{k}(x) = \frac{h}{k}(a_\lambda)$, so that condition ii) of Theorem 4.2 is satisfied. We then need only verify that condition iii) is met by some λ.

<u>LEMMA 4.8</u> Assume that $\varepsilon < \varepsilon^*$, and put

$$\alpha(\lambda) = \frac{1-\varepsilon}{\varepsilon}\int_{a_\lambda}^{r} \nu(x)\left[\frac{h(a_\lambda)}{k(a_\lambda,\lambda)}k(x,\lambda)-h(x)\right]dx - \frac{1}{2}\,.$$

There exists a unique $\bar{\lambda} > 0$ such that $\alpha(\bar{\lambda}) = 0$.

<u>Proof</u>:

i) By Lemma 4.5, $a_\lambda \to r$ as $\lambda \to \infty$, and $k(x,\lambda)$ is decreasing in x, so that

$$\alpha(\lambda) \leq \frac{1-\varepsilon}{\varepsilon}\int_{a_\lambda}^{r} \nu(x)[h(a_\lambda)-h(x)]dx - \frac{1}{2}$$

$$\to -\frac{1}{2} \text{ as } \lambda \to \infty.$$

ii) A rather lengthy calculation, using

$$\frac{k(x,\lambda)}{k(a_\lambda,\lambda)} = e^{-\int_{a_\lambda}^{x} \xi(y,\lambda)dy}$$

yields

$$\alpha'(\lambda) = -\frac{1}{\varepsilon}\int_{a_\lambda}^{r} \bar{F}_\nu(y,\lambda)\frac{d}{d\lambda}\,\xi(y,\lambda)dy < 0,$$

where $\bar{F}_\nu(y,\lambda) = \int_{y}^{r} \nu(x)f(x,\lambda)dx$.

iii) Since $a_\lambda \to 0$ as $\lambda \to 0$,

$$\lim_{\lambda\to 0} \alpha(\lambda) = \frac{1-\varepsilon}{\varepsilon}\left[h(0) \lim_{\lambda\to 0} \int_{a_\lambda}^{r} \nu(x) \frac{k(x,\lambda)}{k(a_\lambda,\lambda)}dx - \int_{a_\lambda}^{r} \nu(x)h(x)\right] - \frac{1}{2},$$

which is positive iff

$$\lim_{\lambda\to 0} \int_{a_\lambda}^{r} \nu(x)\frac{k(x,\lambda)}{k(a_\lambda,\lambda)}dx > \frac{\frac{1}{2}\cdot\frac{\varepsilon}{1-\varepsilon} + \int_0^r \nu(x)h(x)dx}{h(0)}$$

iff

$$\lim_{\lambda\to 0} \int_{a_\lambda}^{r} \nu(x)\left[1 - \frac{k(x,\lambda)}{k(a_\lambda,\lambda)}\right]dx < \frac{\int_0^r \nu(x)[h(0)-h(x)]dx - \frac{1}{2}\cdot\frac{\varepsilon}{1-\varepsilon}}{h(0)} .$$

The right hand side above is positive since $\varepsilon < \varepsilon^*$, and the left hand side is zero:

$$\begin{aligned}
0 &\le \int_{a_\lambda}^{r} \nu(x)\left[1 - \frac{k(x,\lambda)}{k(a_\lambda,\lambda)}\right]dx \\
&= \int_{a_\lambda}^{r} \nu(x)\left[1 - e^{-\int_{a_\lambda}^{x} \xi(y,\lambda)dy}\right]dx \\
&\le \int_{a_\lambda}^{r} \nu(x) \int_{a_\lambda}^{x} \xi(y,\lambda)dy\ dx \\
&\le \int_0^r \nu(x)dx \cdot \int_{a_\lambda}^{r} \xi(y,\lambda)dy \\
&\le r \cdot \int_0^r \nu(x)dx \cdot \int_0^1 \xi((1-t)a_\lambda+tr,\lambda)dt
\end{aligned}$$

$\to 0$ as $\lambda \to 0$.

The last statement is justified by the bounded convergence theorem, since $\xi((1-t)a_\lambda+tr,\lambda)$ converges to zero for all $t \in (0,1]$, and is bounded by the bounded function $\zeta((1-t)a_\lambda+tr)$ for all $t \in [0,1]$. Thus α has a unique zero $\bar\lambda$. □

The set of sufficient conditions of Theorem 4.2 having been met, we now have

<u>THEOREM 4.9</u> Assume that $\varepsilon < \varepsilon^*$, and that B1), B2) and either condition of Lemma 4.6 hold. Let $(\bar\lambda, a_{\bar\lambda}) \in (0,\infty) \times (0,r)$ be the unique pair satisfying $\alpha(\bar\lambda) = 0$, $\xi(a_{\bar\lambda},\bar\lambda) = \zeta(a_{\bar\lambda})$. Put $\psi_0(x) = \psi(x,\bar\lambda)$, $f_0(x) = f(x,\bar\lambda)$. Then

$$V(\psi_0,f) \leq \frac{1}{I_r(f_0)} = V(\psi_0,f_0) \leq V(\psi,f_0)$$

for all $\psi \in \Psi_r$ and $f \in F$. Furthermore,

$$0 < I_r(f_0) = 2(1-\varepsilon)\left\{\int_0^{a_{\bar\lambda}} \left(\frac{h'}{h}\right)^2 h\sigma dx - \bar\lambda[H_\nu(r)-H_\nu(a_{\bar\lambda})]-2(h'\sigma)(a_{\bar\lambda})\right\} - \bar\lambda\varepsilon. \quad □$$

Remark:

5) Even when solutions to the minimax problem exist in Ψ_r, for all $r < \infty$, it is not generally the case that the solution in Ψ coincides with the limiting version, as $r \to \infty$, of the solutions in Ψ_r. The proof, in Lemma 4.8, that the condition

$$\int_{a_\lambda}^{r} \nu g_0 dx = \frac{1}{2}$$

is always attainable by some $\lambda > 0$ makes crucial use of a number of properties unique to Ψ_r. In particular, the property $\lim_{\lambda \to 0} \xi(x,\lambda) = 0$ may fail in Ψ. As a consequence, the above condition cannot always be met in Ψ. This leads to the use of $\lambda = 0$ and sub-stochastic contamination - the excepted possibility in Theorem 4.2 iii).

In the next section, we give the solutions in Ψ_r to the minimax problem for m-dimensional estimation of location and scale. The solutions in Ψ are also given, based upon the considerations in Case I above. For both location and scale, when the ideal density is the normal density, the condition above fails for sufficiently large ε or m, and the least favourable f_0 is sub-stochastic.

5. THE MINIMAX SOLUTIONS

We return now to the optimization problem for $\mathcal{U}_{\varepsilon,r}$, outlined at the beginning of Section 4, using the parameterization introduced at (3.15) and in Example 2, Section 4. We must first check that the assumptions of Lemma 3.4 are satisfied by some pair (ψ_0,ψ_1), for all $u \in \mathcal{U}_{\varepsilon,r}$. That $\beta \neq 0$ is clear, as long as $\psi_1 \geq 0$, $\psi_1 \not\equiv 0$. In the presence of the conditions $\psi_i \not\equiv 0$, the assumptions that α and $\beta+2\gamma$ be bounded away from zero are equivalent to $\inf_{\psi_i} \sup_{F} V(\psi_i,f) < \infty$, and so are satisfied if ε is less than the bound ε^* of Lemma 4.7. In the notation of Theorem 1.6, $\varepsilon^* = \varepsilon_4^* \leq \varepsilon_1^* \leq \varepsilon_2^*$, and so those bounds which are necessary and sufficient for identifiability are sufficient to ensure that the conditions of Lemma 3.4 are met.

CASE I: *LOCATION*

Here, $\eta_0(x) = \frac{1}{m}$ and $\frac{\sigma_0'}{\sigma_0}(x) = \frac{m-1}{x}$, so that for $m > 1$ the solution to $J(\xi) \equiv -\lambda$, passing through $(r,0)$, is strictly decreasing by B2') of Lemma 4.5. This holds for $m = 1$ as well - see Example 3, Section 4. For notational convenience we work with $J(\xi) \equiv -\frac{4\lambda^2}{m}$, i.e.

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

Putting $\xi = -\frac{2v'}{v} = -\frac{k'}{k}$, where $k = v^2$, gives

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

which in turn transforms into Bessel's equation (see Watson [11]). The general solution is

$$v(x) = x^{(2-m)/2}(\omega_1 I_{(m-2)/2}(\lambda x) + \omega_2 K_{(m-2)/2}(\lambda x)),\ \omega_1, \omega_2 \in R;$$

where I and K are the modified Bessel functions. Solving for $k = k_0$ and $\xi = \xi_0$, and choosing the solution satisfying $\xi_0(r,\lambda) = 0$ gives

$$k_0(x,\lambda) = x^{2-m}(\omega K_{(m-2)/2}(\lambda x) + I_{(m-2)/2}(\lambda x))^2, \tag{1}$$

$$\xi_0(x,\lambda) = \frac{2\lambda(\omega K_{m/2}(\lambda x) - I_{m/2}(\lambda x))}{(\omega K_{(m-2)/2}(\lambda x) + I_{(m-2)/2}(\lambda x))}\ ; \tag{2}$$

where $\omega = \dfrac{I_{m/2}(\lambda r)}{K_{m/2}(\lambda r)}$.

The construction of (ψ_0, f_0) then requires one to solve the equations $\xi(a,\lambda) = \zeta(a)$, $\alpha(\lambda) = 0$, from Theorem 4.9. Putting

$$\delta(a) = \frac{\frac{\varepsilon}{1-\varepsilon} + H_m(r^2) - H_m(a^2)}{a h_m(a^2)}, \tag{3}$$

these equations become

$$\xi(a,\lambda) = \zeta(a), \tag{4}$$

$$\frac{2}{a^{m-1}k(a,\lambda)} \int_a^r x^{m-1}k(x,\lambda)dx = \delta(a). \tag{5}$$

By Theorem 4.9, we then have

<u>THEOREM 5.1</u> Assume that $\varepsilon < \varepsilon_1^*$, that B1) holds, and that either

i) $\zeta(x)$ is non-decreasing on $(0,r)$; or

ii) $J(\zeta)$ is non-increasing where it is negative.

With $k_0(x,\lambda)$ and $\xi_0(x,\lambda)$ defined by (1) and (2), and (λ,a) determined from (4) and (5), put

$$\psi_0(x) = -\psi_0(-x) = \begin{cases} \zeta(x), & 0 < x \leq a, \\ \xi_0(x,\lambda), & a \leq x \leq r, \\ 0, & x \geq r; \end{cases}$$

$$f_0(x) = f_0(-x) = \begin{cases} (1-\varepsilon)h(x), & 0 \leq x \leq a, \\ (1-\varepsilon)\dfrac{h(a)}{k_0(a,\lambda)}k_0(x,\lambda), & a \leq x \leq r, \\ 0, & x > r; \end{cases}$$

$$a_0^*(\underline{y}'\underline{y}) = \frac{\psi_0(|\underline{y}|)}{|\underline{y}|}, \quad u_0^*(\underline{y};\underline{0},I) = f_0(|\underline{y}|).$$

Then $V(a_0^*,u) \leq \dfrac{1}{I_r(f_0)} = V(a_0^*,u_0^*) \leq V(a_0,u_0^*)$,

for all $a_0 \in \{a(\underline{y}'\underline{y}) = \dfrac{\psi(|\underline{y}|)}{|\underline{y}|};\ \psi\in\Psi_r\}$ and all $u \in \mathcal{U}_{\varepsilon,r}$. □

Remarks:

1) Consider the following families of m-variate densities:

$$P_1 = \{w|w(\underline{y}) = h(|\underline{y}|) = (\frac{c}{\sqrt{\pi}})^m \frac{\Gamma(m/2)}{\Gamma(m/(2k))}\exp(-|c\underline{y}|^{2k});\ c,k>0,\ m\geq 1\},$$

$$P_2 = \{w|w(\underline{y}) = h(|\underline{y}|) = \frac{\Gamma((m+k)/2)}{(\pi c)^{m/2}\Gamma(k/2)}\left(1 + \frac{|\underline{y}|^2}{c}\right)^{-(m+k)/2};\ c,k>0,\ m\geq 1\}.$$

The assumptions on ζ, in Theorem 5.1, hold for

a) P_1; if $k \geq \frac{1}{2}$,

b) P_2; if $k \leq m-4$, or if $k > m-4$ and $r^2 < c(1 + \frac{4m}{k-m+4})$.

Incidentally, this answers a conjecture of Collins [2], who considered the problem of multivariate estimation of location, with scale known and with $w(\underset{-}{y})$ the normal density. It was conjectured that (ψ_0, f_0) would be of the above form, with, in the three-dimensional case, the constants determined from six equations and inequalities. See also Remark 3) below.

2) The integral at (5) can be evaluated explicitly. With $p = \frac{m-2}{2}$, (5) becomes

$$\frac{2}{\lambda^2 a^{m-1} k_0(a,\lambda)} \int_{\lambda a}^{\lambda r} t(\omega K_p(t)+I_p(t))^2 dt = \delta(a). \tag{6}$$

By the method of Lommel integrals, one can derive the identity

$$2\int t(\omega K_p(t)+I_p(t))^2 dt = (t^2+p^2)(I_p(t)+\omega K_p(t))^2 - [p(I_p(t)+\omega K_p(t)) - t(\omega K_{p+1}(t)-I_{p+1}(t))]^2$$

$$= \left(\frac{t}{\lambda}\right)^{2p+2} k_0\left(\frac{t}{\lambda},\lambda\right)\left[\lambda^2 + \frac{p\lambda}{t}\,\xi_0\left(\frac{t}{\lambda},\lambda\right) - \frac{1}{4}\,\xi_0^2\left(\frac{t}{\lambda},\lambda\right)\right]. \tag{7}$$

The identity $I_{p+1}(z)K_p(z) + I_p(z)K_{p+1}(z) = \frac{1}{z}$ gives

$$k_0(r,\lambda) = [\lambda^2 r^m K_{m/2}^2(\lambda r)]^{-1}. \tag{8}$$

Using (4), (8) and $\xi_0(r,\lambda) = 0$ to evaluate (7), we find that (4) and (5) are together equivalent to

$$\xi_0(a,\lambda) = \zeta(a), \tag{9}$$

$$\frac{r^m k_0(r,\lambda)}{a^m k_0(a,\lambda)} = \frac{\zeta(a)}{4\lambda^2}\left(\frac{2(m-2)}{a} - \zeta(a)\right) + \frac{\delta(a)}{a} + 1. \tag{10}$$

3) For odd m, these equations simplify somewhat, since the modified Bessel functions of order half an odd integer have terminating series expansions. For $m = 1$, see Example 3, Section 4. If $m = 3$, then

$$\xi_0(x,\lambda) = \frac{2}{x}\left[\frac{(\lambda^2 xr-1)\tanh\,(\lambda r-\lambda x)+(\lambda r-\lambda x)}{\lambda r-\tanh\,(\lambda r-\lambda x)}\right],$$

$$k_0(x,\lambda) = \frac{2e^{2\lambda r}}{\pi}\left[\frac{\lambda r\cosh\,(\lambda r-\lambda x)-\sinh\,(\lambda r-\lambda x)}{\lambda^2 x(r+1)}\right]^2.$$

In particular, if $w(y) = (2\pi)^{-3/2}e^{-\frac{1}{2}\underline{y}'\underline{y}}$, so that $\zeta(x) = x$, then (9) and (10) become

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

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

where

$$\delta(a) = \frac{\frac{\varepsilon}{1-\varepsilon} + X_3^2(r^2) - X_3^2(a^2)}{a\chi_3^2(a^2)} .$$

4) For $w(\underline{y})$ the normal density:

$$(V(a_0^*,u_0^*))^{-1} = I_r(f_0)$$

$$= (1-\varepsilon)\left[X_{m+2}^2(a^2) + 4\chi_{m+2}^2(a^2) - \frac{4\lambda^2}{m}(X_m^2(r^2) - X_m^2(a^2))\right] - \frac{4\lambda^2\varepsilon}{m} .$$

5) In the case $r = \infty$, we find

$$\xi_0(x,\lambda) = \frac{2\lambda K_{m/2}(\lambda x)}{K_{(m-2)/2}(\lambda x)} ,$$

$$k_0(x,\lambda) = x^{2-m}K^2_{(m-2)/2}(\lambda x).$$

For $m = 1$ and $m = 3$,

$$\xi_0(x,\lambda) = 2(\lambda + \frac{m-1}{2x}),\ k_0(x,\lambda) = \frac{\pi e^{-2\lambda x}}{2\lambda x^{m-1}} .$$

For all m, ξ_0 and k_0 are asymptotic to the right hand members above, as $x \to \infty$. The constants satisfy either

$$\lambda > 0,\ \xi_0(a,\lambda) = \zeta(a),$$

$$\frac{\zeta(a)}{2\lambda} \cdot \frac{K_{(m-4)/2}(\lambda a)}{K_{(m-2)/2}(\lambda a)} = 1 + \frac{\frac{1}{1-\varepsilon} - H_m(a^2)}{a^2 h_m(a^2)} ;$$

or, for ε and m sufficiently large,

$$\lambda = 0, \quad \zeta(a) = \frac{2(m-2)}{a} \quad (= \xi_0(a,0)).$$

In particular, for $w(\underline{y}) = (2\pi)^{-3/2} e^{-\frac{1}{2}(\underline{y}'\underline{y})} = h(|\underline{y}|)$, the optimal pair is that given by Huber [5]:

$$\psi_0(x) = -\psi_0(-x) = \begin{cases} x, & 0 < x \leq a, \\ (a - \frac{2}{a}) + \frac{2}{x}, & x \geq a; \end{cases}$$

$$f_0(x) = f_0(-x) = \begin{cases} (1-\varepsilon)h(x), & 0 \leq x \leq a \\ (1-\varepsilon)(\frac{a}{x})^2 h(a) \exp(-(a - \frac{2}{a})(x-a)), & x \geq a; \end{cases}$$

where

$$\frac{1}{1-\varepsilon} = \chi_3^2(a^2) + \frac{2a^2\chi_3^2(a^2)}{a^2-2}$$

and $a^2 \geq 2$.

CASE II: SCALE

Here, $\eta_1(x) = \frac{2x^2}{m(m+2)}$ and $\frac{\sigma_1'}{\sigma_1}(x) = \frac{m+1}{x}$, so that again the solution to $J(\xi) \equiv -\lambda$ is strictly decreasing in x. We work with $J(\xi) \equiv -\frac{2m\lambda}{m+2}$ instead, i.e.

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

Put $v_1(x) = x\xi - m$, to get $v_1^2 - 2xv_1' \equiv m^2(1+\lambda)$. A particular solution is $v_1(x) = m\sqrt{1+\lambda}$. The general solution is obtained by setting $v_1(x) = m\sqrt{1+\lambda} + \frac{1}{v(x)}$, and solving the resulting equation $v' + \frac{m\sqrt{1+\lambda}}{x} v = -\frac{1}{2x}$, with the integrating factor $x^{m\sqrt{1+\lambda}}$. Unravelling these transformations and choosing the solution through $(r,0)$ gives

$$\xi_1(x,\lambda) = \frac{(R+1)m}{x}\left[\frac{1-\left(\frac{x}{r}\right)^{mR}}{1+\frac{R+1}{R-1}\left(\frac{x}{r}\right)^{mR}}\right], \text{ where } R = \sqrt{1+\lambda}; \tag{11}$$

whence

$$k_1(x,\lambda) = \frac{\left[1+\frac{R+1}{R-1}\left(\frac{x}{r}\right)^{mR}\right]^2}{\left(\frac{x}{r}\right)^{m(R+1)}} . \tag{12}$$

<u>THEOREM 5.2</u> With assumptions as in Theorem 5.1, with $\xi_1(x,\lambda)$ and $k_1(x,\lambda)$ defined by (11) and (12), and with (λ,a) determined from (4) and (5), put

$$\psi_1(x) = -\psi_1(-x) = \begin{cases} \zeta(x), & 0 < x \leq a, \\ \xi_1(x,\lambda), & a \leq x \leq r, \\ 0, & x \geq r; \end{cases}$$

$$f_1(x) = f_1(-x) = \begin{cases} (1-\varepsilon)h(x), & 0 \leq x \leq a, \\ (1-\varepsilon)\frac{h(a)}{k_1(a,\lambda)}k_1(x,\lambda), & a \leq x \leq r, \\ 0, & x > r; \end{cases}$$

$$a_1^*(\underline{y}'\underline{y}) = |\underline{y}|\psi_1(|\underline{y}|), \quad u_1^*(\underline{y};\underline{0},I) = f_1(|\underline{y}|).$$

Then $V(a_1^*,u) \leq \frac{1}{I_r(f_0)} = V(a_1^*,u_1^*) \leq V(a_1,u_1^*)$, for all $a_1 \in \{a(\underline{y}'\underline{y}) = |\underline{y}|\psi(|\underline{y}|);\ \psi \in \Psi_r\}$ and all $u \in U_{\varepsilon,r}$. □

Remarks:

6) In the families P_1 and P_2, the assumptions on ζ, in Theorem 5.2, hold for

a) P_1; if $k \geq \frac{1}{2}$;

b) P_2; all members.

7) For $w(\underline{y})$ the normal density:

$$(V(a_1^*,u_1^*))^{-1} = I_r(f_0)$$

$$= 2(1-\varepsilon)\left[X^2_{m+4}(a^2) + 4\chi^2_{m+4}(a^2) - \frac{m\lambda}{m+2}(X^2_m(r^2) - X^2_m(a^2))\right] - \frac{2m\lambda\varepsilon}{m+2}\ .$$

8) If $r = \infty$, then $\xi_1(x,\lambda) = \frac{(R+1)m}{x}$, $k_1(x,\lambda) = x^{-(R+1)m}$. If $w(\underline{y})$ is the normal density, then the solution is

$$\psi_1(x) = -\psi_1(-x) = \begin{cases} x, & x \leq b, \\ \frac{b^2}{x}, & x \geq b; \end{cases}$$

$$f_1(x) = f_1(-x) = \begin{cases} (1-\varepsilon)h(x), & x \leq b; \\ (1-\varepsilon)h(b)\left(\frac{b}{x}\right)^{b^2}, & x \geq b; \end{cases}$$

where $h(x) = (2\pi)^{-m/2}e^{-\frac{1}{2}x^2}$. Here, $\sqrt{1+\lambda} = \frac{b^2-m}{m}$, so that $b^2 \geq 2m$. If $b^2 > 2m$, it satisfies

$$\frac{1}{1-\varepsilon} = \frac{2b^2\chi_m^2(b^2)}{b^2-m} + X_m^2(b^2).$$

For large ε or m, this is then not satisfiable, so that $b^2 = 2m$ and the contamination is sub-stochastic. In this case Huber [5] adopts another approach. For $-1 < \lambda < 0$, A_λ also includes an interval about the origin. Huber's method is to contaminate this interval as well, to get

$$\psi_1(x) = -\psi_1(-x) = \begin{cases} \frac{a^2}{x}, & 0 < x \leq a, \\ x, & a \leq x \leq b, \\ \frac{b^2}{x}, & b \leq x; \end{cases}$$

$$f_1(x) = f_1(-x) = \begin{cases} (1-\varepsilon)h(a)(\frac{a}{x})^{a^2}, & 0 < x \leq a, \\ (1-\varepsilon)h(x), & a \leq x \leq b, \\ (1-\varepsilon)h(b)(\frac{b}{x})^{b^2}, & b \leq x; \end{cases}$$

where $a^2 < m < b^2 = 2m-a^2$, implying $J(\psi_1) \equiv -m^2\lambda > 0$ on both extreme intervals, are determined from

$$\frac{1}{1-\varepsilon} = \left[\frac{2a^2}{m-a^2}\chi_m^2(a^2) - X_m^2(a^2)\right] + \left[\frac{2b^2}{b^2-m}\chi_m^2(b^2) + X_m^2(b^2)\right].$$

As in the proof of Theorem 4.2, however, the use of negative values of λ implies that condition (4.3) can fail for sub-stochastic members of G_ν - put $g = \alpha g_1$, $\alpha < 1$, in the above to get, using $\int_0^\infty \nu g_1 dx = \frac{1}{2}$,

$$\int J(\psi_1)\nu(g-g_1)dx = \frac{\alpha-1}{2}(-m^2\lambda) = \frac{\alpha-1}{2}(ab)^2 < 0.$$

APPENDIX

This appendix contains a more detailed exposition of the steps leading to Lemma 4.1, and an expanded proof of a variant of Theorem 4.2. Both the assumptions and conclusions are somewhat weaker than in the original version. The assumption that all distributions under consideration have densities, and that these densities are absolutely continuous and symmetric, may be weakened to apply only to the known distribution H_ν. On the other hand, the necessity of condition vi) of Theorem 4.2 is lost. It was previously overlooked that the definition of information used there does not apply to densities which are only piecewise absolutely continuous. With respect to the more generally applicable definition used here, the distribution corresponding to the function f_1 constructed in the previous proof has infinite information. This renders the proof of necessity of vi) invalid. Consequently, some of our examples pertaining to the class Ψ must be amended. Our results for Ψ_r require no changes, however. See also the remarks following the proof of Theorem 3 below.

The problem is to minimax variance functionals of the form

$$V(\psi,f) = \frac{\int_{-\infty}^{\infty} \psi^2(x)\eta(x)f(x)\nu(x)dx}{\left[\int_{-\infty}^{\infty} \psi(x)\eta(x)f'(x)\nu(x)dx\right]^2} \quad ;$$

where

i) ψ is chosen from a set Ψ of functions which are typically piecewise smooth, skew-symmetric, and continuous everywhere except possibly at $x = 0$.

ii) ν, η and $\sigma = \eta\nu: R \to [0,\infty)$ are positive on $R \setminus \{0\}$.

iii) $f\nu$ is an absolutely continuous probability density.

We define

$$\Psi_r = \{\psi \in \Psi \mid \psi(x)=0 \text{ for } |x| \geq r\}.$$

We do not specify the set Ψ precisely, and restrict the definition of $V(\psi,f)$ to d.f.s which possess absolutely continuous densities, for reasons which will later become clear, when $V(\psi,f)$ is related to the information functional.

In the definitions which follow we do not assume the existence of densities.

For fixed ν and η, define

$$\mathsf{G} = \{G \mid \int_{-\infty}^{x} \nu(y)dG(y) = G_\nu(x) \text{ is a, possibly sub-stochastic,}$$

$$\text{d.f. on } R, \text{ with } \int_{-\infty}^{\infty} \eta(x)dG_\nu(x) = \int_{-\infty}^{\infty} \sigma(x)dG(x) < \infty\},$$

$$\mathsf{G}_\nu = \{G_\nu \mid G \in \mathsf{G}\}.$$

For fixed $\varepsilon \in [0,1)$, and a fixed member H of G, put

$$\mathsf{F} = \{F \mid F = (1-\varepsilon)H + \varepsilon G \text{ for some } G \in \mathsf{G}\},$$

$$\mathsf{F}_\nu = \{F_\nu \mid F \in \mathsf{F}\}.$$

Example: Let the r. vec. $\underline{x} \in R^m$ have a spherically symmetric density $g(|\underline{x}|)$. Put $\nu(x) = \frac{C_m m}{2}|x|^{m-1}$. Then $X = |\underline{x}|$ has density

$$g^m(x) = 2\nu(x)g(x) \text{ and d.f.}$$

$$G_\nu^*(x) = \int_0^x 2\nu(y)g(y)dy, \quad 0 \leqslant x \leqslant \infty.$$

Put

$$G(x) = \int_{-\infty}^x g(|y|)dy,$$

$$G_\nu(x) = \frac{1}{2} + \left(\frac{\text{sign } x}{2}\right)G_\nu^*(|x|)$$

$$= \int_{-\infty}^x \nu(y)dG(y), \quad -\infty \leqslant x \leqslant \infty.$$

Then G_ν is a symmetric d.f. on R. Conversely, if $\int_{-\infty}^x \nu(y)dG(y)$ is a d.f., then $G(|\underline{x}|)$ is the d.f. of a r. vec. on R^m, with

$$P(\underline{x} \in A) = \int_{\{|\underline{a}|\,|\,\underline{a}\in A\}} dG(|\underline{x}|),$$

for all Borel sets A, and G_ν^* is the d.f. of $|\underline{x}|$.

Defining $\eta_i(x) = \left(\frac{2}{m+2}\right)^i \frac{x^{2i}}{m}$ and $\psi_i(x) = x^{1-2i}a_i(x^2)$ $(i=0,1)$, we then have $V(\psi_i,f) = V(a_i,u)$, as at (3.15), in the cases where $f = F'$ is absolutely continuous. If $m = 1$ and $i = 0$, then $V(\psi_0,f)$ is the well known variance functional for univariate estimation of location.

Define the information of F_ν (with respect to η) by

$$I(F_\nu) = \sup_\psi \frac{\left(\int_{-\infty}^{\infty} (\psi\sigma)' dF\right)^2}{\int_{-\infty}^{\infty} \psi^2 \sigma dF} ,$$

where the sup is over the set C_k^1 of all continuously differentiable functions with compact support satisfying $\int \psi^2 \sigma dF > 0$.

Simple modifications to the proofs of Theorem 4.2, and Propositions 4.3, 4.5 of Huber [7] give

THEOREM 1:

A) The following are equivalent:

i) $I(F_\nu) < \infty$.

ii) F has an absolutely continuous derivative f satisfying $\int \left(\frac{f'}{f}\right)^2 f\sigma dx < \infty$.

In either case, $I(F_\nu) = \int \left(\frac{f'}{f}\right)^2 f\sigma dx$.

B) There is an $F_\nu^0 \in \mathsf{F}_\nu$ minimizing $I(F_\nu)$.

Define F_0 by $F_\nu^0(x) = \int_{-\infty}^{x} \nu(y) dF_0(y)$.

C) If $0 < I(F_\nu^0) < \infty$, and the set where $f_0 = F_0'$ is strictly positive is convex, then F_ν^0 is unique.

Now assume that $I(H_\nu) < \infty$, so that H has an absolutely continuous derivative h satisfying $I(H_\nu) = \int \left(\frac{h'}{h}\right)^2 h\sigma dx < \infty$. Assume further that h is strictly positive on R. Put

$$\mathsf{F}_\nu' = \{F_\nu \in \mathsf{F}_\nu \mid I(F_\nu) < \infty\}.$$

Then $\mathcal{F}_\nu'$ is dense in $\mathcal{F}_\nu$.

The conditions imposed on H ensure that there is a unique $F_\nu^0 \in \mathcal{F}_\nu'$ minimizing $I(F_\nu)$ over $\mathcal{F}_\nu$. Since $I(F_\nu)$ is a convex functional of F_ν (see Lemma 4.4, Huber [7]), F_ν^0 minimizes $I(F_\nu)$ iff $\frac{d}{dt} I(F_\nu^t)\Big|_{t=0} \geq 0$ for all $F_\nu^t = (1-t)F_\nu^0 + tF_\nu^1$ with $F_\nu^1 \in \mathcal{F}_\nu'$. Performing the differentiation and setting $\psi_0 = -\frac{f_0'}{f_0}$ gives the necessary and sufficient condition

$$0 \leq \int_{-\infty}^{\infty} [2(f_0'-f')\psi_0 + (f_0-f)\psi_0^2]\, c\,dx \tag{1}$$

for all $f = F'$ with $F_\nu \in \mathcal{F}_\nu'$.

We can now relate the information and variance functionals via Theorem 2 of Huber [3].

<u>THEOREM 2</u>: Under the assumptions on H, there is a unique $F_\nu^0 \in \mathcal{F}_\nu'$ minimizing $I(F_\nu)$.

i) If $\psi_0 = -\frac{f_0'}{f_0}$ is contained in Ψ, then (ψ_0, f_0) is a saddlepoint of $V(\psi, f)$ in $\Psi \times \mathcal{F}_\nu'$:

$$V(\psi_0, f) \leq \frac{1}{I(F_\nu^0)} = V(\psi_0, f_0) \leq V(\psi, f_0), \tag{2}$$

for all $\psi \in \Psi$ and $F_\nu \in \mathcal{F}_\nu'$.

ii) Conversely, if (ψ_0, f_0) is a saddlepoint, and Ψ contains a non-zero multiple of $-\frac{f_0'}{f_0}$, then $I(F_\nu^0) \leq I(F_\nu)$ for all $F_\nu \in \mathsf{F}_\nu$, F_ν^0 is uniquely determined, and ψ_0 is $[F_\nu^0]$-equivalent to a multiple of $-\frac{f_0'}{f_0}$.

iii) Necessary and sufficient for F_ν^0 to minimize $I(F_\nu)$ is that (1) above be satisfied.

Note that since $V(\psi,f) = V(c\psi,f)$ for any $c \neq 0$, we may take the multiplying constant in Theorem 2 to be unity. Note also that since $I(F_\nu)$ is convex, the symmetrized distribution $\tilde{F}_\nu(x) = \frac{1}{2}[F_\nu(x) + 1 - F_\nu(-x)]$ has smaller information than F_ν. If H_ν is symmetric, then $\tilde{F}_\nu = (1-\varepsilon)H_\nu + \varepsilon\tilde{G}_\nu \in \mathsf{F}_\nu$, so that the least informative F_ν^0 will then arise from that subset of G_ν whose members are symmetric. If as well η and ν are symmetric, then (1) becomes

$$0 \leq \int_{-\infty}^{\infty} [2(g_0'-g')\psi_0+(g_0-g)\psi_0^2]\sigma dx \tag{3}$$

for all $G \in \tilde{\mathsf{G}} = \{G \in \mathsf{G} | G'(x)=G'(-x),\ (1-\varepsilon)H_\nu+\varepsilon G_\nu \in \mathsf{F}_\nu'\}$.

Suppose that, as will be shown, ψ_0 is differentiable and bounded. The differentiability allows for an integration by parts in (3). Our assumption that $\int_{-\infty}^{\infty} \sigma(x)dG(x) < \infty$ for all $G \in \mathsf{G}$, implying that $\sigma(x)g(x) \to 0$ as $x \to \pm\infty$, together with the boundedness of ψ_0, yields the vanishing of the end-effects at $\pm\infty$. Then (3) becomes

$$0 \leq \int_{-\infty}^{\infty} [2(\psi_0\sigma)'-(\psi_0^2\sigma)](g-g_0)dx + 2 \lim_{x\downarrow 0} (\psi_0\sigma)(g-g_0)(x) - 2 \lim_{x\uparrow 0} (\psi_0\sigma)(g-g_0)(x)$$

$$= 2\left\{\int_0^{\infty} [2\psi_0'-\psi_0^2+2\psi_0\frac{\sigma'}{\sigma}]\eta\nu(g-g_0)dx + 2 \lim_{x\downarrow 0} (\psi_0\eta)(\nu g-\nu g_0)(x)\right\}. \quad (4)$$

Defining

$$J(\psi) = (2\psi'-\psi^2+2\psi \frac{\sigma'}{\sigma})\eta$$

we then have that, if ψ_0 is differentiable and bounded, and ν and η are symmetric, then a pair of sufficient conditions for F_ν^0 to minimize $I(F_\nu)$ is

S1) $\int_0^{\infty} J(\psi_0)\nu(g-g_0)dx \geq 0$ for all $G \in \tilde{G}$;

S2) $(\nu g_0)(0) = 0 \leq (\psi_0\eta)(0^+)$, or $(\psi_0\eta)(0^+) = 0 \leq (\nu g_0)(0)$.

Define $\zeta(x) = -\frac{h'}{h}(x)$, and assume:

A0) The functions ν,η are symmetric, and are positive and smooth on $R \setminus \{0\}$. The function $h(x)$ is symmetric and positive on R, and $I(H_\nu) = \int_{-\infty}^{\infty} \zeta^2 h\sigma dx < \infty$.

A1) The function $\zeta(x)$ is smooth.

<u>THEOREM 3</u>: There is a unique, symmetric $F_\nu^0 \in \mathcal{F}_\nu'$ minimizing $I(F_\nu)$ over $\mathcal{F}_\nu$.

A) The function $\psi_0 = -\frac{f_0'}{f_0}$ is bounded, continuous, and piecewise smooth.

B) In order that F_ν^0 minimizes $I(F_\nu)$, it is necessary that there exist $\lambda \geq 0$, and a set $B_\lambda = \bigcup_{j=1}^{N(\lambda)} B_{\lambda,j} \subseteq (0,\infty)$, where $N(\lambda) \leq \infty$ and the $B_{\lambda,j}$ are non-overlapping open intervals, such that:

N1) $$\psi_0(x) = -\psi_0(-x) = \begin{cases} \zeta(x), & x \in B_\lambda^c \\ \xi(x;\omega_j,\lambda), & x \in B_{\lambda,j}, \end{cases}$$

where $J(\xi(\cdot;\omega_j,\lambda)) \equiv -\lambda$ on $B_{\lambda,j}$ for any fixed $\omega_j \in R$;

N2) $$f_0(x) = f_0(-x) = \begin{cases} (1-\varepsilon)h(x), & x \in B_\lambda^c, \\ (1-\varepsilon)\left(\sup_{B_{\lambda,j}} \dfrac{h(x)}{k(x;\omega_j,\lambda)}\right) k(x;\omega_j,\lambda), & x \in B_{\lambda,j}, \end{cases}$$

where each k satisfies $\xi = -\frac{k'}{k}$ and $\sup_{B_{\lambda,j}} \frac{h}{k}(x)$ is attained at each non-zero, finite endpoint of $B_{\lambda,j}$;

N3) $\int_{B_\lambda} \nu g_0 dx \leq \frac{1}{2}$, with equality if $\lambda > 0$, where $g_0 = \dfrac{f_0-(1-\varepsilon)h}{\varepsilon}$;

N4) $A_\lambda = \{x\in(0,\infty) \mid J(\zeta)<-\lambda\} \subseteq B_\lambda$.

C) If the pair (ψ_0, f_0) satisfies N1) - N4) and S2), and if as well ψ_0 is bounded, and $\psi_0 \in \Psi$, then (2) holds.

D) If C) holds, and $\dfrac{(\psi_0\sigma)'}{\nu}$ is either continuous or non-negative upper semi-continuous, and vanishes at infinity, then (ψ_0, f_0) is not only a saddlepoint with respect to $\mathcal{F}_\nu'$, but with respect to $\mathcal{F}_\nu$ as well.

Proof: The existence and uniqueness of F_ν^0 follow from Theorem 2). Part D) may be proven as in Theorem 5) of Huber [3], by writing $V(\psi,f)$ as

$$\frac{\int \frac{\psi^2\sigma}{\nu}\, dF_\nu}{\left(\int \frac{(\psi\sigma)'}{\nu}\, dF_\nu\right)^2}\,.$$

The assumptions in C) ensure that if S1) is satisfied, then $I(F_\nu^0)$ minimizes $I(F_\nu)$, so that (ψ_0,f_0) is a saddlepoint by Theorem 2. Using first N1), then N1) and N4), then N3), the integral in S1) is

$$\begin{aligned}
\int_0^\infty J(\psi_0)\nu(g-g_0)dx &= \int_{B_\lambda^c} J(\zeta)\nu g dx + \int_{B_\lambda} J(\xi)\nu(g-g_0)dx \\
&\geq -\lambda\left[\int_{B_\lambda^c}\nu g dx + \int_{B_\lambda}\nu(g-g_0)dx\right] \\
&= -\lambda\left[\int_0^\infty \nu g dx - \int_{B_\lambda}\nu g_0\right] \\
&\geq 0.
\end{aligned}$$

It remains to prove A) and B).

With $f_0 = (1-\varepsilon)h + \varepsilon g_0$, where $F_\nu^0(x) = \int_{-\infty}^x f_0(y)\nu(y)dy$, define

$$B = \{x\in(0,\infty)\,|\,g_0(x)>0\},$$

and put

$$\xi(x) = \begin{cases} -\dfrac{((1-\varepsilon)h+\varepsilon g_0)'}{((1-\varepsilon)h+\varepsilon g_0)}(x), & x\in B, \\ 0, & x\in B^c. \end{cases}$$

Then

$$f_0(x) = f_0(-x) = \begin{cases} (1-\varepsilon)h(x), & x \in B^c, \\ \\ (1-\varepsilon)h(x)+\varepsilon g_0(x), & x \in B. \end{cases}$$

Define

$$\frac{\lambda}{2} = \int_B [2\psi_0 g_0' + \psi_0^2 g_0]\sigma dx.$$

We first show that $\lambda < \infty$, by showing that otherwise the derivative $\frac{d}{dt} I(F_\nu^t)_{|t=0}$ fails to exist in at least one direction $F_\nu^t = (1-t)F_\nu^0 + tF_\nu^1$, with $F_\nu^1 \in \mathsf{F}_\nu'$. We have, as at (1),

$$\begin{aligned} 0 \leq \frac{d}{dt} I(F_\nu^t)_{|t=0} &= \int_{-\infty}^{\infty} [2(f_0'-f_1')\psi_0 + (f_0-f_1)\psi_0^2]\sigma dx \\ &= \varepsilon \int_{-\infty}^{\infty} [2(g_0'-g_1')\psi_0 + (g_0-g_1)\psi_0^2]\sigma dx \\ &= 2\varepsilon\left\{\frac{\lambda}{2} - \int_0^{\infty} [2g_1'\psi_0 + g_1\psi_0^2]\sigma dx\right\}, \end{aligned} \tag{5}$$

if $G_1 \in \tilde{G}$. The finiteness of λ will then follow if the last integral is finite for at least one $G_1 \in \tilde{G}$. Let $[a,b] \subset (0,\infty)$ be any compact on which $\int_a^b \psi^2(x)dx < \infty$. Let g_1 be any continuously differentiable function vanishing off of $[a,b] \cup [-b,-a]$, with $G_1 \in \tilde{G}$. Then

$$\left|\int_0^\infty [2g_1'\psi_0+g_1\psi_0^2]\sigma dx\right| = \left|\int_a^b [2g_1'\psi_0+g_1\psi_0^2]\sigma dx\right|$$

$$\leq 2 \sup_{[a,b]} |g_1'(x)\sigma(x)| \left|\int_a^b |\psi_0(x)|dx + \sup_{[a,b]} |g_1(x)\sigma(x)| \left|\int_a^b \psi_0^2(x)dx < \infty.\right.\right.$$

Thus $\lambda < \infty$. Also, if ψ_0 is discontinuous, then by taking a sequence $\{G_n\} \subset \tilde{G}$, with $g_n\nu$ tending to point mass at a discontinuity, one can obtain a contradiction to (5). Similarly, ψ_0 must be bounded.

Now (5) becomes

$$0 \leq -\int_B [2\xi g'+\xi^2 g]\sigma dx - \int_{B^c} [2\zeta g'+\zeta^2 g]\sigma dx + \frac{\lambda}{2}, \tag{6}$$

for all $G \in \tilde{G}$. First consider these G's for which the support of g is finite, and is contained in B^c. Then the first integral in (6) vanishes, and the assumed differentiability of ζ, together with the finiteness of the support of g, allows for an integration by parts. This yields

$$\int_{B^c} J(\zeta)\nu g dx \geq -\frac{\lambda}{2}$$

for all such g, whence $J(\zeta) \geq -\lambda$ on B^c and $A_\lambda \subseteq B$.

On the other hand, if $G \in \tilde{G}$ satisfies $g \equiv 0$ on B^c, $\int_B \nu g dx \leq \frac{1}{2}$, then (6) becomes

$$\int_B [2\xi g'+\xi^2 g]\sigma dx \leq \frac{\lambda}{2} = \int_B [2\xi g_0'+\xi^2 g_0]\sigma dx. \tag{7}$$

Putting $g = \frac{1}{2}g_0$ in (7) shows that $\lambda \geq 0$. Putting

$$g = \frac{g_0}{2\int_B g_0 \nu dx}$$

shows that $\int_B g_0 \nu dx = \frac{1}{2}$ if $\lambda > 0$.

Now let $[a,b]$ be any compact in B, and let $g\nu$ have all of its mass $(\frac{1}{2})$ on $[a,b] \cup [-b,-a]$. Define a function ϕ, continuous on B, by

$$\phi(t) = 2(\xi\eta)(t) - \int_0^t [-2(\xi\eta)(x)\frac{\nu'}{\nu}(x)+(\xi^2\eta)(x)]dx + \lambda t.$$

An integration by parts gives

$$\int_a^b \phi(t)(g\nu)'(t)dt$$

$$= \int_a^b 2(\xi\eta)(t)(g\nu)'(t)dt - \int_a^b (g\nu)'(t) \int_0^t [-2(\xi\eta)(x)\frac{\nu'}{\nu}(x)+(\xi^2\eta)(x)]dxdt + \lambda\int_a^b t(g\nu)'(t)dt$$

$$= \int_a^b \{2(\xi\eta)(t)(g\nu)'(t)+(g\nu)(t)[-2(\xi\eta)(t)\frac{\nu'}{\nu}(t)+(\xi^2\eta)(t)]\}dt - \lambda\int_a^b (g\nu)(t)dt$$

$$= \int_a^b [2\xi g'+\xi^2 g]\sigma dt - \frac{\lambda}{2}$$

$$\leq 0,$$

by (7). Since a and b are arbitrary,

$$\int_B \phi(g\nu)'dt \leq 0 \qquad (8)$$

for all $g\nu$ with mass $\frac{1}{2}$ on B. The equality at (7) becomes

$$\int_B \phi(g_0\nu)'dt = 0, \tag{9}$$

upon approximating g_0 by functions with compact support. By multiplying the left hand side of (9) by a constant, if necessary, we may assume that $\int_B g_0\nu dt = \frac{1}{2}$.

Suppose that, for some $[a,b] \subset B$ and some $g_1\nu$ with mass $\frac{1}{2}$ on $[a,b]$, the inequality at (8) is strict:

$$\int_a^b \phi(g_1\nu)'dt < 0. \tag{10}$$

Choose α such that

$$0 < \alpha < \min\left\{1, \inf_{[a,b]} \frac{g_0(t)}{g_1(t)}\right\},$$

and put $g_2 = \dfrac{g_0-\alpha g_1}{1-\alpha}$. Then $\int_B g_2\nu dt = \frac{1}{2}$, $g_2\nu \geq 0$ on B, and by (8) - (10),

$$0 \geq \int_B \phi(g_2\nu)'dt = \frac{1}{1-\alpha}\int_B \phi(g_0\nu)'dt - \frac{\alpha}{1-\alpha}\int_B \phi(g_1\nu)'dt$$

$$= \frac{-\alpha}{1-\alpha}\int_a^b \phi(g_1\nu)'dt > 0,$$

a contradiction. Thus

$$\int_a^b \phi(g\nu)'dt = 0 \tag{11}$$

for all $[a,b] \subset B$ and all g with $G \in \tilde{G}$ and $\int_a^b g\nu dt = \frac{1}{2}$.

We now show that (11) implies that ϕ is constant on B. Define a set of symmetric functions $\{g(x;a_0,b_0) \mid a<a_0<b_0<b\}$ by

$$g(x;a_0,b_0)\nu(x) = \begin{cases} \dfrac{c_0(x-a)}{a_0-a}, & a \leq x \leq a_0, \\ c_0, & a_0 \leq x \leq b_0, \\ \dfrac{c_0(b-x)}{b-b_0}, & b_0 \leq x \leq b, \\ 0, & 0 \leq x \leq a,\ b \leq x, \\ g(-x;a_0,b_0), & x < 0; \end{cases}$$

where c_0 is a normalizing constant.

Recall that $\nu(x)$ is smooth, hence absolutely continuous on every compact. The absolute continuity of the $g(x;a_0,b_0)$ follows. Put

$$f(x;a_0,b_0) = (1-\varepsilon)h(x) + \varepsilon g(x;a_0,b_0),$$

and recall that h is strictly positive on $[a,b]$. A straightforward calculation then shows that $\int_a^b \left(\frac{f'}{f}\right)^2 f\sigma dx < \infty$, whence $I(F_\nu) < \infty$ and $G(x;a_0,b_0) \in \tilde{G}$. Substituting $g(x;a_0,b_0)$ into (11) gives

$$\frac{1}{a_0-a}\int_a^{a_0} \phi dt - \frac{1}{b-b_0}\int_{b_0}^{b} \phi dt = 0$$

for all $a < a_0 < b_0 < b$. Letting $(a_0,b_0) \to (a,b)$ gives $\phi(a) = \phi(b)$, so that ϕ is constant on B. It follows that $\xi\eta$ is differentiable. Since η is smooth and positive on B, ξ is differentiable. Furthermore,

$$0 \equiv \phi'(t) = 2(\xi\eta)'(t) + 2(\xi\eta)(t)\frac{\nu'}{\nu}(t) - (\xi^2\eta)(t) + \lambda$$

$$= J(\xi) + \lambda,$$

i.e. ξ is continuously differentiable on B and satisfies $J(\xi) \equiv -\lambda$ there. This proves part A) of the theorem.

Now define $B_\lambda = B$, and represent it as

$$B_\lambda = \bigcup_{j=1}^{N(\lambda)} B_{\lambda,j},$$

where $N(\lambda) \leq \infty$ and the $B_{\lambda,j}$ are non-overlapping open intervals. On $B_{\lambda,j}$, $\psi_0 = \xi$ satisfies $J(\xi) \equiv -\lambda$, by the above. Since this (Ricatti) equation has a solution passing through any pre-specified point, ψ_0 must have the form given in N1). Parts N3) and N4) have already been established.

That f_0 has the form given in N2) on B_λ^c follows from the definition of B_λ. On $B_{\lambda,j}$, $\psi_0 = \xi(x;\omega_j,\lambda)$, so that $f_0(x) = \alpha_j k(x;\omega_j,\lambda)$, where k is any function satisfying $-\frac{k'}{k} = \xi$, and $\alpha_j \in R$. Since $f_0 \geq (1-\varepsilon)h$ on $\bar{B}_{\lambda,j}$, $\alpha_j \geq (1-\varepsilon)\sup_{B_{\lambda,j}} \frac{h}{k}(x)$. The continuity of f_0 then forces equality in this last relation, and forces the sup to be attained at each non-zero, finite endpoint of $B_{\lambda,j}$.

This completes the proof.

Remarks.

1. Theorem 3 is valid for the class Ψ_r, with $(0,\infty)$ and Ψ replaced by $(0,r)$ and Ψ_r throughout.

2. I am so far unable to find a correct proof of the necessity of the condition "$A_\lambda \cap B_{\lambda,j} \neq \phi$ for all j." I do, however, believe that the condition is necessary. This belief rests primarily on philosophical grounds. The optimality principle at work in the construction of ψ_0 is that of replacing ζ by ξ wherever $J(\zeta)$ threatens to become too small. The placement of contaminating mass g_0 on an interval on which $J(\zeta)$ is already large contradicts this principle. At the very least, it makes no contribution to the optimality of ψ_0 and hence cannot, in my opinion, be a feature of the solution.

3. Theorem 4.4 is invalid without the proof of the necessity of the aforementioned condition. However, remarks 2) and 3) on p. 77 remain valid, as do remarks 1) and 4) in the cases in which B_λ is a single interval. Part i) of Example 3 remains valid - see also Huber [7]. Part ii) requires a little more thought.

The statements in Example 4 now become conjectures. It is of course still necessary to have contamination on the interval $A_\lambda = (c_\lambda, d_\lambda)$, and it is easy to rule out the possibility of further contamination to the right of A_λ. Contamination to the left of the interval (a_λ, b_λ) constructed in that example remains a possibility, however.

4. Part a) of Example 5 requires no changes. However, the "solution" for Ψ, given in part b), fails to meet condition S2). We give here the correct solution, based upon a suggestion by J. Collins.

a) <u>Small values of ε</u>:

Define a_0 ($\approx$2.218) by $\tanh(-\frac{a_0}{2}) = 1 - \frac{4}{a_0}$. For $a \in (0,4)$, define $\lambda = \lambda(a)$ by

i) $\sqrt{\lambda}\tanh(-\frac{\sqrt{\lambda}}{2}a) = 1 - \frac{4}{a}$.

It will be shown that for all $a \in (0,a_0)$:

ii) $\lambda \geq 1$ $(= \lambda(a_0))$,

iii) $J(\zeta)(a) = \frac{-a^2+8a-8}{a^2} \geq -\lambda$.

Put

$$\xi(x;\lambda) = \sqrt{\lambda}\tanh(-\frac{\sqrt{\lambda}}{2}x),$$

$$k(x;\lambda) = \cosh^2(-\frac{\sqrt{\lambda}}{2}x),$$

$$\psi_0(x) = -\psi_0(-x) = \begin{cases} \xi(x;\lambda), & 0 \leq x \leq a, \\ \zeta(x), & a \leq x; \end{cases}$$

$$f_0(x) = f_0(-x) = \begin{cases} (1-\varepsilon)\left[\frac{h(a)}{k(a;\lambda)}\right]k(x;\lambda), & 0 \leq x \leq a, \\ (1-\varepsilon)h(x), & a \leq x. \end{cases}$$

The continuity of ψ_0 follows from condition i) above. Condition ii) implies that A_λ is a single interval $(0,b)$ and condition iii) ensures that $A_\lambda \subseteq (0,a) = B_\lambda$. As in Remark 1) on p. 77, g_0 is non-negative. Also, $\psi_0(0) = 0$ and ψ_0 is bounded. The solution is then valid for those ε for which the condition $\int_{B_\lambda} g_0 dx = \frac{1}{2}$ can be met. With $\alpha(a)$ defined by

$$\alpha(a) = \frac{h(a)}{k(a;\lambda)} \int_0^a k(x;\lambda)dx - \int_0^a h(x)dx,$$

this condition becomes

iv) $\frac{\varepsilon}{2(1-\varepsilon)} = \alpha(a).$

We claim that α is increasing in a. From this it follows that the solution is valid for $\varepsilon \lesssim \varepsilon^*$, where

$$\frac{\varepsilon^*}{2(1-\varepsilon^*)} = \alpha(a_0) \quad (= (3 - \frac{4}{a_0})h(a_0) + \frac{1}{2} - H(a_0)).$$

We have

$$\alpha'(a) = \left(\frac{d}{da}\frac{h(a)}{k(a;\lambda)}\right) \int_0^a k(x;\lambda)dx;$$

$$\frac{d}{da}\frac{h(a)}{k(a;\lambda)} = \frac{h'(a)k(a;\lambda)-h(a)\frac{d}{da}k(a;\lambda)-h(a)\left(\frac{d}{d\lambda}k(a;\lambda)\right)\lambda'(a)}{k^2(a;\lambda)}$$

$$= \frac{h(a)(\xi(a;\lambda)-\zeta(a))}{k(a;\lambda)} - \frac{h(a)}{k^2(a;\lambda)}\left(\frac{d}{d\lambda}k(a;\lambda)\right)\lambda'(a);$$

$\xi(a;\lambda) - \zeta(a) = 0$ by i);

$$\frac{d}{d\lambda}\,k(a;\lambda) = \frac{a\sinh(a\sqrt{\lambda})}{4\sqrt{\lambda}} > 0;$$

and

$$\lambda'(a) = \frac{-2\sqrt{\lambda}[\lambda a^2+4+4\cosh(a\sqrt{\lambda})]}{a^2[a\sqrt{\lambda}+\sinh(a\sqrt{\lambda})]} < 0;$$

so that $\alpha'(a) > 0$. Also, condition ii) above follows from $\lambda'(a) < 0$. It remains only to verify condition iii). With $\lambda_1(a)$ defined by

$$\lambda_1(a) = \frac{a^2-8a+8}{a^2},$$

condition iii) is implied by

v) $\lambda(a) \geq \lambda_1(a)$ for $0 < a < 4$.

Since $\lambda(4) = 0 > \lambda_1(4)$, if v) fails then there must exist a such that $\lambda(a) = \lambda_1(a)$, i.e.

$$\lambda = \frac{a^2-8a+8}{a^2}, \quad \sqrt{\lambda}\tanh(-\frac{\sqrt{\lambda}}{2}a) - (1-\frac{4}{a}) = 0.$$

Equivalently, with ϕ defined by

$$\phi(x) = \left(\frac{x^2-8x+8}{x^2}\right)^{\frac{1}{2}}\tanh\left(-\frac{x}{2}\left(\frac{x^2-8x+8}{x^2}\right)^{\frac{1}{2}}\right) - 1 + \frac{4}{x},$$

we must have $\phi(a) = 0$. Note that $\lambda_1(a) = \lambda(a) > 0$ requires that a be less than a_1, where a_1 is the zero in $(0,4)$ of a^2-8a+8.

It can be shown that ϕ is strictly decreasing for $x \in (0,a_1)$. Then since $\phi(a_1) = -1 + \frac{4}{a_1} > 0$, ϕ has no zeroes. It follows that (ψ_0,f_0) is the optimal pair for $\varepsilon \lesssim \varepsilon^* \approx .054$, with (λ,a) determined from i) and iv).

b) Large values of ε:

For $b > 4$, define $a \in (2,4)$ by

vi) $\zeta(b) \tanh\left(-\frac{a\zeta(b)}{2}\right) = 1 - \frac{4}{a}$.

Here, $\zeta^2(b)$ plays the role of λ, and $A_{\zeta^2(b)}$ includes an additional, half-infinite, interval since $\zeta^2(b) < 1$. Put

$$\psi_1(x) = -\psi_1(-x) = \begin{cases} \xi(x;\zeta^2(b)), & 0 \leq x \leq a, \\ \zeta(x), & a \leq x \leq b, \\ \zeta(b), & b \leq x; \end{cases}$$

$$f_1(x) = f_1(-x) = \begin{cases} (1-\varepsilon)\left(\frac{h(a)}{k(a;\zeta^2(b))}\right)k(x;\zeta^2(b)), & 0 \leq x \leq a, \\ (1-\varepsilon)h(x), & a \leq x \leq b, \\ (1-\varepsilon)h(b)e^{-\zeta(b)(x-b)}, & b \leq x. \end{cases}$$

Determine b from

vii) $\int_0^\infty f_1(x)dx = \frac{1}{2}$.

With the exception of N4), the conditions in A) - C) of Theorem 3) are immediate. Condition N4) requires

viii) $J(\zeta)(a) \geq -\zeta^2(b)$, $J(\zeta)(b) \geq -\zeta^2(b)$.

The first part of viii) is equivalent to v) above, and the second is easily verified. Thus (ψ_1, f_1) is the optimal pair for $\varepsilon > \varepsilon^*$. The limiting case $\varepsilon = \varepsilon^*$ corresponds to $b = \infty$, $a = a_0$, $\lambda = 1$.

5. Our results for Ψ_r in Case II, Section 4 were obtained by verifying directly that conditions N1) - N4) may be satisfied. Under the conditions of Lemma 4.6, $(\nu g_0)(0) = 0 \leq (\psi_0\eta)(0^+)$, so that S2) is satisfied. Thus, all of the results for Ψ_r remain valid, as do those in Section 5.

BIBLIOGRAPHY

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

[2] Collins, J.R. (1982). Robust *M*-estimators of location vectors. To appear in *Multivariate Analysis*.

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

[4] Huber, P.J. (1971). Robust Statistics. *Unpublished manuscript.*

[5] Huber, P.J. (1976). Robust covariances. Statistical Decision Theory and Related Topics, II, S.S. Gupta and D.S. Moore, eds., *Academic Press, New York.*

[6] Huber, P.J. (1977). Robust Statistical Procedures. *SIAM.*

[7] Huber, P.J. (1981). Robust Statistics. *Wiley & Sons.*

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

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

[10] Rao, R.R. (1962). Relations between weak and uniform convergence of measures with applications. *Ann. Math. Statist.* 33 659-680.

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