MINIMAX BIAS-ROBUST ESTIMATION OF THE DISPERSION MATRIX OF A MULTIVARIATE DISTRIBUTION

By Jorge G. Adrover

Universidad Nacional de Córdoba

Maronna defines affine equivariant M-estimators for multivariate location and scatter. They are particularly suited for estimating the pseudocovariance or scatter matrix of an elliptical population. By defining the bias of a dispersion matrix properly, we consider the maximum bias of an M-estimator over an ε-neighborhood of the underlying elliptical distribution (location known). We find that Tyler's estimator minimizes the maximum bias.

1. Introduction. The classical approach of multivariate analysis is based on the multivariate normal distribution $\mathcal{N}_m(\mu, \mathbf{V})$ with location vector μ and dispersion matrix $\mathbf{V}$. The maximum likelihood estimators for these parameters are the sampling mean and covariance matrix, respectively. However, it is well known that slight departures from this model or the presence of a small fraction of outliers have a large influence on the estimators. The maximum asymptotic bias over an ε-neighborhood of contamination from the target model gives a measure of the global robustness of an estimator. The smallest level of contamination for which the maximum asymptotic bias becomes infinite is called *the gross-error breakdown point* of the estimator. The concept of breakdown point only guarantees that the bias remains bounded over an ε-contamination neighborhood of distributions. Formalizing the concept outlined above, maximum likelihood estimators have a zero breakdown point. Several robust methods to estimate location and scatter have been proposed. Maronna (1976) defines affine equivariant M-estimators of multivariate location and scatter, showing existence, uniqueness, consistency and asymptotic normality. Huber (1977, 1981) generalizes Maronna's M-estimators. Unfortunately, M-estimators may be deficient for high-dimensional problems, since such estimators have a gross-error breakdown point smaller than $1/(m + 1)$, where m is the dimension of the multivariate vector. Tyler (1987) considers a limiting case of M-estimators for multivariate scatter whose asymptotic distribution is distribution free with respect to the class of continuous elliptically distributed populations. In addition, it is the most robust estimator in the sense of minimizing the maximum asymptotic variance.

Maronna and Yohai (1990) define the asymptotic bias for dispersion matrices. They establish that the choice of Tyler's estimator within the class of

M-estimators was suggested by the results given in Martin, Yohai and Zamar (1989). More precisely, a weighted $L1$ estimator obtained from an implicit equation using a sign function minimizes the maximum asymptotic bias for a given ε among the M-estimators for regression. They conjecture that a similar result should hold for the robust covariance problem. In this paper we prove that Tyler's M-functional is bias-minimax among M-functionals. In other words, we show that Tyler's M-functional minimizes the maximum bias within the class of M-functionals of the dispersion matrix of a multivariate distribution. The main points of the paper are the following.

1. Let $\mathbf{V}(G)$ be a matrix-valued M-functional of scatter of a distribution G on $\Re^m$. Let F_0 be spherically symmetric around zero. Define the maximum bias as

$$\underline{b}(\mathbf{V}, F_0, \varepsilon)$$
$$= \sup\left\{\frac{\lambda_1(\mathbf{V}(G))}{\lambda_m(\mathbf{V}(G))}, G = (1 - \varepsilon)F_0 + \varepsilon H, H \text{ any distribution on } \Re^m\right\}$$

for $0 < \varepsilon < 1$, where λ_1, λ_m denote maximal and minimal eigenvalues, respectively. Then, under some conditions on $\mathbf{V}$ and F_0, the maximum bias occurs under point mass contaminations. That is,

$$\underline{b}(\mathbf{V}, F_0, \varepsilon) = \sup\left\{\frac{\lambda_1(\mathbf{V}(G))}{\lambda_m(\mathbf{V}(G))}, G = (1 - \varepsilon)F_0 + \varepsilon\delta_{\mathbf{x}}, \mathbf{x} \in \Re^m\right\},$$

where $\delta_{\mathbf{x}}$ is Dirac measure at some point $\mathbf{x} \in \Re^m$.
2. Let $\mathbf{V}$ and $\mathbf{V}_T$ denote Maronna's and Tyler's M-functionals, respectively. Then, under certain conditions on $\mathbf{V}$ and F_0,

$$\underline{b}(\mathbf{V}_T, F_0, \varepsilon) \leq \underline{b}(\mathbf{V}, F_0, \varepsilon).$$

The heuristics behind formal statement (1) are the following: restricting to point mass contamination seems to be the most suitable way to force the M-estimators to yield the largest degree of asymmetry while the solution under a spherically symmetric distribution results in a multiple of the identity matrix. The bias as defined above detects such an asymmetry through the condition number of the dispersion matrix, measuring its departure from the identity. The actual importance of discrete measures in the bias is formally established through Lemma 3.3, which was proved by a referee.

The wide range of applications for robust affine equivariant estimators of multivariate scatter matrix is not confined to classical areas of multivariate analysis such as discriminant analysis, principals components, canonical correlations, multivariate linear model and outlier detection. Robust dispersion matrix estimators are also useful to compute some robust regression estimators such as GM or to compute projection-based estimators which require the covariance matrix of random regressors in order to be optimal bias-robust [see Martin, Yohai and Zamar (1989) and Maronna and Yohai (1993)].

In Section 2 we introduce the notation and define the basic concepts. In Section 3 we obtain the maximum bias of M-estimators over the ε-contamination

neighborhood. Section 4 deals with M-functionals under point mass contamination. In Section 5 we show that Tyler's functional estimator is bias-minimax. All the proofs are deferred to the Appendix.

2. Definitions and notation. The Euclidean norm of a vector $\mathbf{z} \in \Re^m$ will be denoted by $\|\mathbf{z}\|$, $\mathbf{I}$ will represent the identity matrix in $\Re^{m \times m}$ and $\mathbf{C}'$ the transpose of $\mathbf{C}$. If $\mathbf{a} \in \Re^m$, $\mathbf{A} = \mathrm{diag}(\mathbf{a}')$ denotes the diagonal matrix such that $\mathbf{A}_{ii} = \mathbf{a}_i$. The relation "$\mathbf{A} - \mathbf{B}$ is positive definite" will be written as "$\mathbf{A} > \mathbf{B}$."

The indicator function of the set A will be written as $1\{A\}$. Given a probability measure $\mathrm{P} \in \Re^m$, we define the support of P by $\mathrm{supp}(\mathrm{P}) = \{\mathbf{z} \in \Re^m$ such that for all neighborhoods $\mathrm{V}_0(\mathbf{z})$ of $\mathbf{z}$, $\mathrm{P}(\mathrm{V}_0(\mathbf{z})) > 0\}$. Then δ_K is the point mass contamination at $(K^{1/2}, 0, \ldots, 0)$ with $K > 0$.

Let $\mathbf{z}$ be an m-dimensional vector with ellipsoidal distribution, that is, $\mathbf{z} = \mathbf{A}\mathbf{x} + \mathbf{t}_0$ where $\mathbf{A}$ is a nonsingular $m \times m$ matrix, $\mathbf{t}_0$ is an m-dimensional vector and $\mathbf{x}$ is spherically symmetric distributed with density $f_0 \colon \Re^+ \to \Re^+$. If $\mathbf{V}_0 = \mathbf{A}\mathbf{A}'$, the density of $\mathbf{z}$ is

$$f(\mathbf{z}, \mathbf{t}_0, \mathbf{V}_0, f_0) = (\det(\mathbf{V}_0))^{-1/2} f_0((\mathbf{z} - \mathbf{t}_0)' \mathbf{V}_0^{-1} (\mathbf{z} - \mathbf{t}_0)).$$

Let $\mathscr{E}$ be the set of positive definite symmetric $m \times m$ matrices. Given $\mathbf{V} \in \mathscr{E}$, we denote by $\lambda_1(\mathbf{V}) \geq \cdots \geq \lambda_m(\mathbf{V})$ the eigenvalues of $\mathbf{V}$ and by $\gamma(\mathbf{V})$ its *condition number*,

$$\gamma(\mathbf{V}) = \frac{\lambda_1(\mathbf{V})}{\lambda_m(\mathbf{V})}.$$

If F_0 denotes the distribution under the elliptical model, we define the contaminated ε-neighborhood as

$$\mathscr{F}_\varepsilon = \big\{G = (1-\varepsilon)F_0 + \varepsilon H, \ \text{where H denotes an arbitrary distribution in } \Re^m\big\},$$

for $0 < \varepsilon < 1$. Let $\mathbf{z}$ be distributed as $G = (1 - \varepsilon)F_0 + \varepsilon H$. For the sake of simplicity we will often write $(1 - \varepsilon)E_{F_0} h + \varepsilon E_H h$ instead of $E_G h(\mathbf{z}) = (1 - \varepsilon)E_{F_0} h(\mathbf{z}) + \varepsilon E_H h(\mathbf{z})$, provided the expectation exists. Let $\mathbf{V}$ be a multivariate dispersion functional with the following properties: its domain contains $\mathscr{F}_\varepsilon$ for every elliptical F_0, and $\mathbf{V}(G)$ is a symmetric definite positive matrix. $\mathbf{V}$ will be assumed to be an affine equivariant dispersion functional; that is, let $\mathbf{z}$ have a distribution function G, take any nonsingular $m \times m$ matrix $\mathbf{C}$ and $\mathbf{t}$ any vector in $\Re^m$. If H is the distribution function of $\mathbf{C}\mathbf{z}+\mathbf{t}$, then $\mathbf{V}(H) = \mathbf{C}\mathbf{V}(G)\mathbf{C}'$. The *gross-error asymptotical breakdown point* of the functional $\mathbf{V}$ measures the amount of contamination needed to carry $\mathbf{V}$ over all bounds or making it "almost singular." More precisely, it is defined as

$$\varepsilon^*(\mathbf{V}, F_0) = \sup \left\{ \begin{array}{l} \varepsilon > 0 \colon \text{ there exists a compact set } K(\varepsilon) \subset \Re^+ \text{ such that} \\ \lambda_j(\mathbf{V}(G)) \in K(\varepsilon) \text{ for all } 1 \leq j \leq m \text{ and } G \in \mathscr{F}_\varepsilon \end{array} \right\}.$$

Asymptotic bias can be measured in the following affine invariant way. Let us assume that $\mathbf{V}(F_0) = \mathbf{B}\mathbf{B}'$. Define *the asymptotic bias* of the dispersion matrix

$\mathbf{V}$ in $G \in \mathscr{F}_\varepsilon$ as $\gamma(\mathbf{B}^{-1}\mathbf{V}(G)(\mathbf{B}')^{-1})$ and the *maximum asymptotic bias* of the dispersion matrix $\mathbf{V}$ in the neighborhood $\mathscr{F}_\varepsilon$ as

$$\underline{b}(\mathbf{V}, F_0, \varepsilon) = \sup_{G \in \mathscr{F}_\varepsilon} \gamma(\mathbf{B}^{-1}\mathbf{V}(G)(\mathbf{B}')^{-1}).$$

Under such a definition, the bias is invariant under affine transformation of the data and does not depend on the decomposition $\mathbf{V}(F_0) = \mathbf{BB}'$ [see Muirhead (1982), page 589]. Only the shape of $\mathbf{V}$ matters. This measure of bias reflects the degree of spherical asymmetry caused by contamination for a spherically symmetric target model.

  We will study the effect of contamination for M-functionals. Maronna (1976) defines M-functionals of multivariate location and scatter as the solutions of the following equations:

$$E_{\mathrm{P}} w\big(((\mathbf{z} - \mathbf{t})\mathbf{V}^{-1}(\mathbf{z} - \mathbf{t}))^{1/2}\big)(\mathbf{z} - \mathbf{t}) = \mathbf{0},$$

$$E_{\mathrm{P}} u\big((\mathbf{z} - \mathbf{t})'\mathbf{V}^{-1}(\mathbf{z} - \mathbf{t})\big)(\mathbf{z} - \mathbf{t})(\mathbf{z} - \mathbf{t})' = \mathbf{V},$$

where $(\mathbf{t}, \mathbf{V}) \in \Re^m \times \mathscr{E}$ and the functions $w(s)$ and $u(s)$ are defined for $s \geq 0$. Existence and uniqueness of solutions can be proved under regularity conditions on w, u and P. Tyler (1987) considers the estimation of parameters of the form $H(\mathbf{V})$, where H is such that $H(\mathbf{V}) = H(c\mathbf{V})$ for any $\mathbf{V} > \mathbf{0}$ and $c > 0$. He treats a limiting form of a Huber-type M-estimator of scatter defined by $u(s) = \alpha$ if $s \leq r^2$ and $u(s) = \alpha r^2/s$ otherwise, for a chosen r and a scaling factor α. Then r and α can be taken so as to get Fisher-consistency under normality. More precisely, the estimator is chosen by solving

$$(2.1) \qquad \frac{m}{\mathrm{P}(\mathbf{z} \neq \mathbf{t})} E_{\mathrm{P}} \frac{(\mathbf{z} - \mathbf{t})(\mathbf{z} - \mathbf{t})'}{(\mathbf{z} - \mathbf{t})'\mathbf{V}^{-1}(\mathbf{z} - \mathbf{t})} 1\{\mathbf{z} \neq \mathbf{t}\} = \mathbf{V},$$

or equivalently,

$$(2.2) \qquad \frac{m}{\mathrm{P}(\mathbf{z} \neq \mathbf{t})} E_{\mathrm{P}} \frac{\mathbf{V}^{-1/2}(\mathbf{z} - \mathbf{t})(\mathbf{z} - \mathbf{t})'\mathbf{V}^{-1/2}}{(\mathbf{z} - \mathbf{t})'\mathbf{V}^{-1}(\mathbf{z} - \mathbf{t})} 1\{\mathbf{z} \neq \mathbf{t}\} = \mathbf{I},$$

where $\mathbf{t}$ is either known or assumed to be a robust affine equivariant location estimator satisfying several conditions given in Tyler (1987) that assure consistency and asymptotic normality. He proves existence of solutions and uniqueness up to a positive factor in the case of the empirical distribution function being based on a random sample. Consequently, additional conditions such as $\mathrm{trace}(\mathbf{V}) = m$ or $\det(\mathbf{V}) = 1$ are required to assure uniqueness. In Adrover (1993) it is shown that these results also hold within the contaminated ε-neighborhood, getting existence for $\varepsilon < 1/(m + 1)$ and uniqueness unless there is a positive factor. As mentioned, Tyler's estimator is optimum in the following sense: let us take the parameter $H(\mathbf{V})$, where H is a continuously differentiable real function. Let $\mathscr{D}$ be the set of elliptical continuous distribution functions F. If $\mathbf{V}_{T, n}(F)$ denotes the Tyler's estimator obtained from sampling a distribution $F \in \mathscr{D}$ then

$$\max_{F \in \mathscr{D}} \mathrm{Var}\, H(\mathbf{V}_{T, n}(F)) \leq \max_{F \in \mathscr{D}} \mathrm{Var}\, H(\mathbf{V}_n(F)),$$

where $\mathrm{Var}(\cdot)$ denotes the asymptotic variance and $H(\mathbf{V}_n(F))$ is any consistent, asymptotically normal estimator of $H(\mathbf{V})$.

Throughout, $\mathbf{t}$ will be assumed to be known and without loss of generality it will be taken equal to $\mathbf{0}$. Since we consider only the case of fixed location, the M-functional of scatter to consider will be the solution to

$$(2.3) \qquad E_{\mathrm{P}}\psi(\mathbf{z}'\mathbf{V}^{-1}\mathbf{z})\frac{\mathbf{z}\mathbf{z}'}{\mathbf{z}'\mathbf{V}^{-1}\mathbf{z}}1\{\mathbf{z}\neq\mathbf{0}\}=\mathbf{V}$$

for a function $\psi\geq 0$, or equivalently,

$$(2.4) \qquad S(\mathrm{P},\mathbf{V})=E_{\mathrm{P}}\psi(\mathbf{z}'\mathbf{V}^{-1}\mathbf{z})\frac{\mathbf{V}^{-1/2}\mathbf{z}\mathbf{z}'\mathbf{V}^{-1/2}}{\mathbf{z}'\mathbf{V}^{-1}\mathbf{z}}1\{\mathbf{z}\neq\mathbf{0}\}=\mathbf{I}.$$

On the other hand, (2.2) becomes

$$(2.5) \qquad S_T(\mathrm{P},\mathbf{V})=\frac{m}{\mathrm{P}(\mathbf{z}\neq\mathbf{0})}E_{\mathrm{P}}\frac{\mathbf{V}^{-1/2}\mathbf{z}\mathbf{z}'\mathbf{V}^{-1/2}}{\mathbf{z}'\mathbf{V}^{-1}\mathbf{z}}1\{\mathbf{z}\neq\mathbf{0}\}=\mathbf{I}.$$

We will impose the following conditions on P, u and ψ.

(A1) $u(s)$ is a nonnegative, nonincreasing and continuous function for $s>0$.
(A2) $\psi(s)=su(s)$ is nonnegative, continuous and bounded for $s\geq 0$; let $\psi(\infty)=\sup_{s\geq 0}\psi(s)$.
(A3) $\psi(s)$ is nondecreasing and strictly increasing on the interval $[0,x_0(\psi)]$, where $x_0(\psi)=\sup\{x:\psi(x)<\psi(\infty)\}$.
(A4) There exists s_0 such that $\psi(s_0^2)>m$.
(A5) $\lim_{s\to 0^+}\psi(s)=\psi(0)<m$.
(A6) For all H hyperplane in $\Re^m$, $\mathrm{P}(H)<1-m/\psi(\infty)$. For all H hyperplane in $\Re^m$, $\mathrm{P}(H)\leq 1/m$.

3. Maximum bias of M-functionals of multivariate scatter. In order to analyze the bias of M-functionals of multivariate scatter, it is necessary to determine its breakdown point. This is done by the following lemma [see Adrover (1993)].

LEMMA 3.1. (a) *Suppose* (A1)–(A6) *hold. Then the breakdown point of an M-functional of multivariate scatter* $\mathbf{V}$, *defined by the solution to* (2.3), *is*

$$\varepsilon^*(\mathbf{V},F_0)=\min\left\{\frac{1}{\psi(\infty)},\frac{m-\psi(0)}{\psi(\infty)-\psi(0)},1-\frac{m}{\psi(\infty)}\right\}.$$

(b) *Let* $\mathbf{V}$ *be as in* (a) *and* $\mathbf{V}_T$ *the solution of* (2.1). *Then* $\varepsilon^*(\mathbf{V},F_0)\leq 1/(m+1)\leq\varepsilon^*(\mathbf{V}_T,F_0)\leq 1/m$.

The quantity $1-m/\psi(\infty)$ is related to implosion due to inlier contamination making the smallest eigenvalue 0. Here $1/\psi(\infty)$ and $(m-\psi(0))/(\psi(\infty)-\psi(0))$ are related to explosion because certain type outlier contamination makes the largest eigenvalue go to infinity. In order to define bias we need to compare the M-functional under the target model and the M-functional under a contaminated distribution. Then we will require, without loss of generality, the following condition.

(A7) F_0 is spherically symmetric if F_0 denotes the distribution under the target model.

REMARK. The condition $\varepsilon < 1 - m/\psi(\infty)$ is also necessary to guarantee the existence of M-functionals over the whole ε-neighborhood. It follows after taking $\mathrm{P}_0 = (1-\varepsilon)F_0 + \varepsilon\delta_0$. If there exists $\mathbf{V} \in \mathscr{E}$ such that $S(\mathrm{P}_0, \mathbf{V}) = \mathbf{I}$, then $\operatorname{trace}(S(\mathrm{P}_0, \mathbf{V})) = (1-\varepsilon)E_{F_0}\psi(\mathbf{z}'\mathbf{V}^{-1}\mathbf{z}) = m$ and consequently $\varepsilon < 1-m/\psi(\infty)$. As was mentioned, Tyler's functional is well defined for any distribution in the contaminated neighborhood, provided $\varepsilon < 1/(m+1)$.

The following lemma will show that the solution of (2.3) under the target model is a multiple of the identity matrix (i.e., Fisher consistent) provided (A7) holds.

LEMMA 3.2. *Let us suppose that* (A2)–(A5) *hold and* F_0 *satisfies* (A7). *Let* $\mathbf{a} = (a_1, \ldots, a_m)'$, $a_i > 0$ *for all* $1 \leq i \leq m$. *Then*:

(a) *If* $\mathbf{V} = \operatorname{diag}(\mathbf{a}')$ *then* $S(F_0, \mathbf{V})$ *is a diagonal matrix.*
(b) *If* $\mathbf{V} = c_1\mathbf{I}$, $c_1 > 0$, *then* $S(F_0, \mathbf{V}) = c_2\mathbf{I}$, *for some* $c_2 > 0$.
(c) *If* $\mathbf{V} = \operatorname{diag}(\mathbf{a}')$ *and* $a_1 \geq a_2 \geq \cdots \geq a_m$, *then* $S(F_0, \mathbf{V})$ *is a diagonal matrix with nondecreasing diagonal elements.*
(d) *The solution of* (2.3) *is a multiple of the identity matrix.*

The following lemma shows why we can restrict ourselves to discrete measures to yield any bias.

LEMMA 3.3. *Under* (A1)–(A6) *and given* $\mathbf{V} \in \mathscr{E}$:

(a) $\{S(H, \mathbf{V}): H \text{ some distribution on } \Re^m\} = \mathscr{A} = \{\text{symmetric } \mathbf{A} \in \Re^{m \times m}: \lambda_m(\mathbf{A}) \geq 0, \operatorname{trace}(\mathbf{A}) < \psi(\infty)\}$ *if* $x_0(\psi) = \infty$.
(b) $\{S(H, \mathbf{V}): H \text{ some distribution on } \Re^m\} = \mathscr{B} = \{\text{symmetric } \mathbf{A} \in \Re^{m \times m}: \lambda_m(\mathbf{A}) \geq 0, \operatorname{trace}(\mathbf{A}) \leq \psi(\infty)\}$ *if* $x_0(\psi) < \infty$,
(c) $\{S_T(H, \mathbf{V}): H \text{ some distribution on } \Re^m, H(\{\mathbf{0}\}) < 1\} = \mathscr{A}_T = \{\text{symmetric } \mathbf{A} \in \Re^{m \times m}: \lambda_m(\mathbf{A}) \geq 0, \operatorname{trace}(\mathbf{A}) = m\}$.
(d) *Let* $S(H, \mathbf{V})$ *with* $\mathbf{V} \in \mathscr{E}$ *and* H *some distribution on* $\Re^m$. *Then there exists a distribution* H_m *supported by* m *or fewer points such that* $S(H_m, \mathbf{V}) = S(H, \mathbf{V})$.

Let us note that $\mathscr{A}$, $\mathscr{B}$ and $\mathscr{A}_T$ do not depend on $\mathbf{V}$.

A sufficient condition to obtain the results is given by the following condition, connecting ψ and the distribution F_0.

(A8) Either $x_0(\psi) = \infty$ or $0 \in \operatorname{supp}(F_0)$.

Let us restrict ourselves to the case $x_0(\psi) = \infty$, since the case $x_0(\psi) < \infty$ is completely similar. We will also assume that $E_{F_0}\psi(\mathbf{z}'\mathbf{z}) = m$. With the representations for $S(H, \mathbf{V})$ and $S_T(H, \mathbf{V})$ given by Lemma 3.3, one can easily

conclude that $\underline{b}(\mathbf{V}, F_0, \varepsilon)$ exceeds b if, and only if, there exists a $\mathbf{V} \in \mathcal{E}$ such that $\lambda_1/\lambda_m(\mathbf{V}) > b$ and

(3.1) $$\mathbf{A}(\mathbf{V}) = \varepsilon^{-1}(\mathbf{I} - (1 - \varepsilon)S(F_0, \mathbf{V})) \in \mathscr{A}.$$

But $\mathbf{A}(\mathbf{V}) \in \mathscr{A}$ is easily seen to be equivalent to

$$\text{trace}(S(F_0, \mathbf{V})) > \frac{m - \varepsilon\psi(\infty)}{1 - \varepsilon} \quad \text{and} \quad \lambda_1(S(F_0, \mathbf{V})) \leq \frac{1}{1 - \varepsilon}.$$

On the other hand,

$$S_T((1 - \varepsilon)F_0 + \varepsilon H, \mathbf{V})$$
$$= \begin{cases} \dfrac{1 - \varepsilon}{1 - \varepsilon H(\{\mathbf{0}\})} S_T(F_0, \mathbf{V}) + \dfrac{\varepsilon(1 - H(\{\mathbf{0}\}))}{1 - \varepsilon H(\{\mathbf{0}\})} S_T(H, \mathbf{V}), & \text{if } H(\{\mathbf{0}\}) < 1, \\ S_T(F_0, \mathbf{V}), & \text{if } H(\{\mathbf{0}\}) = 1, \end{cases}$$

and $\mathbf{A}_T(\mathbf{V}) \in \mathscr{A}_T$, if and only if,

$$\lambda_1(S_T(F_0, \mathbf{V})) \leq \frac{1}{1 - \varepsilon}.$$

Thus we end up with the representations

$$\underline{b}(\mathbf{V}, F_0, \varepsilon) = \sup\left\{ \frac{\lambda_1}{\lambda_m}(\mathbf{V}) \colon \mathbf{V} \in \mathcal{E}, \text{trace}(S(F_0, \mathbf{V})) \right.$$

(3.2) $$\left. > \frac{m - \varepsilon\psi(\infty)}{1 - \varepsilon} \text{ and } \lambda_1(S(F_0, \mathbf{V})) \leq \frac{1}{1 - \varepsilon} \right\},$$

$$\underline{b}(\mathbf{V}_T, F_0, \varepsilon) = \sup\left\{ \frac{\lambda_1}{\lambda_m}(\mathbf{V}) \colon \mathbf{V} \in \mathcal{E}, \lambda_1(S(F_0, \mathbf{V})) \leq \frac{1}{1 - \varepsilon} \right\}.$$

We could get more precision in the previous representation. Let $R = \|\mathbf{z}\|$ and $\mathbf{y} = \|\mathbf{z}\|^{-1}\mathbf{z}$. Let us define

$$H_{t, b} = \frac{R^2}{t}(1 - y_1^2(1 - b^{-1})) \quad \text{and} \quad D_b = \frac{1 - y_1^2}{1 - y_1^2(1 - b^{-1})}.$$

Then we can state the following result.

THEOREM 3.1. *Let* $\varepsilon < \min\{1/\psi(\infty), (m - \psi(0))/(\psi(\infty) - \psi(0)), 1 - m/\psi(\infty)\}$ *and* $E_{F_0}\psi(\mathbf{z}'\mathbf{z}) = m$. *Assume* (A1)–(A8). *Then*

$$\underline{b}(\mathbf{V}, F_0, \varepsilon) = \max\left\{ b \geq 1 \colon \text{for some } t > 0, E_{F_0}\psi(H_{t, b}) \geq \frac{m - \varepsilon\psi(\infty)}{1 - \varepsilon} \text{ and} \right.$$

$$\left. E_{F_0}\psi(H_{t, b})D_b \leq \frac{m - 1}{1 - \varepsilon} \right\},$$

$$\underline{b}(\mathbf{V}_T, F_0, \varepsilon) = \max\left\{ b \geq 1 \colon mE_{F_0}D_b \leq \frac{m - 1}{1 - \varepsilon} \right\}.$$

REMARK. Note that under F_0, the random variables R and $\mathbf{y}$ are independent with $\mathbf{y}$ being uniformly distributed on the unit sphere in $\Re^m$. Thus $E_{F_0} D_b$ and $\underline{b}(\mathbf{V}_T, F_0, \varepsilon)$ do not depend on F_0. Then we can denote $\underline{b}(\mathbf{V}_T, F_0, \varepsilon)$ by $b_{\mathbf{T}}(\varepsilon)$.

Let us observe that maximal bias is caused (approximately) by point mass contaminations. Let

$$t(b) = \max\left\{ t \in [0, \infty]\colon E_{F_0}\psi(H_{t, b}) \geq \frac{m - \varepsilon\psi(\infty)}{1 - \varepsilon} \right\}.$$

Note first that

$$E_{F_0}\psi(H_{t(b), b})D_b = \frac{m - 1}{1 - \varepsilon} < m \quad \text{for } b = \underline{b}(\mathbf{V}, F_0, \varepsilon),$$

because $E_{F_0}\psi(H_{t(b), b})D_b$ is continuous in $b \in [1, \infty]$. For arbitrary $b' < \underline{b}(\mathbf{V}, F_0, \varepsilon)$, one can find a pair (t, b) with $b' < b < \underline{b}(\mathbf{V}, F_0, \varepsilon)$ and $0 < t < t(\underline{b}(\mathbf{V}, F_0, \varepsilon))$ such that

$$\frac{m - \varepsilon\psi(\infty)}{1 - \varepsilon} > E_{F_0}\psi(H_{t, b}).$$

But this implies that the matrix $\mathbf{A}(\mathbf{V})$ defined in (3.1) corresponding to $\mathbf{V} = \mathrm{diag}(tb, t, \ldots, t)$ satisfies that $\lambda_2(\mathbf{A}(\mathbf{V})) = \cdots \lambda_m(\mathbf{A}(\mathbf{V})) = 0$. The proof of Lemma 3.3 will show that if we take the degenerate distribution $H(A(\mathbf{V}), \mathbf{V}) = \delta_{rtb}$, where r is chosen so that $\psi(r^2) = \varepsilon^{-1}(m - (1 - \varepsilon)E_{F_0}\psi(\mathbf{z}'\mathbf{V}^{-1}\mathbf{z}))$, then $\mathbf{V}$ solves $S((1 - \varepsilon)F_0 + \varepsilon H, \mathbf{V}) = \mathbf{I}$. Then we will get some insight into the behavior of M-functionals under point mass contamination through the next section.

4. M-functionals under point mass contamination. It will be shown that the solution of (2.3) when $\mathrm{P} = (1 - \varepsilon)F_0 + \varepsilon\delta_K$, $0 < K < \infty$, is a positive definite diagonal matrix $\mathbf{V} = \mathrm{diag}(tb, t, \ldots, t)$ where $b > 1$. This could be intuitively expected: the estimator of scatter should reflect the lack of associations between the variables (the contamination just affects one particular direction). One eigenvector follows the contaminated direction and its eigenvalue should be clearly affected, differing from the indistinguishable remaining eigenvalues.

Instead of using some general theorems about existence of solutions [see Huber (1981) and Maronna (1976)], we will give a proof tailored to the case of point mass contamination. The proof will not require assumption (A6). The equations defining the M-functionals under point mass contamination are taken according to the solution we are seeking and then we show that the M-functionals are actually as they were supposed. Taking $\mathrm{P}_K = (1 - \varepsilon)F_0 + \varepsilon\delta_K$, we are looking for $\mathbf{V} \in \mathscr{E}$ such that

(4.1) $$S(\mathrm{P}_K, \mathbf{V}) = (1 - \varepsilon)S(F_0, \mathbf{V}) + \varepsilon S(\delta_K, \mathbf{V}) = \mathbf{I}.$$

Then, by Lemma 3.2, we may restrict the search for solutions of (4.1) to diagonal matrices $\mathbf{V} = \mathrm{diag}(tb, t, \dots, t)$ where $t > 0$, $b > 0$. Then we search for the solutions to

$$(4.2) \quad (1 - \varepsilon)E_{F_0}\psi(H_{t,b})\frac{y_1^2}{y_1^2 + (1 - y_1^2)b} + \varepsilon\psi\left(\frac{K}{tb}\right)1\{0 < K < \infty\} = 1,$$

$$(1 - \varepsilon)E_{F_0}\psi(H_{t,b})\frac{y_j^2}{1 - y_1^2(1 - b^{-1})} = 1,$$

where $2 \leq j \leq m$.

If we call $g_1^{(1)}(t, b) = \mathrm{trace}(S(\mathrm{P}_K, \mathbf{V}))$ and $g_2(t, b) = m - \lambda_1(S(\mathrm{P}_K, \mathbf{V}))$, we get

$$(4.3) \quad g_1^{(1)}(t, b) = (1 - \varepsilon)E_{F_0}\psi(H_{t,b}) + \varepsilon\psi\left(\frac{K}{tb}\right)1\{0 < K < \infty\} = m,$$

$$g_2(t, b) = (1 - \varepsilon)E_{F_0}\psi\left(H_{t,b}\right)D_b = m - 1.$$

According to Theorem 3.1, $(\underline{b}(\mathbf{V}, F_0, \varepsilon), t(\underline{b}(\mathbf{V}, F_0, \varepsilon)))$ solve a similar system,

$$(4.4) \quad g_1^{(2)}(t, b) = (1 - \varepsilon)E_{F_0}\psi(H_{t,b}) + \varepsilon\psi(\infty) = m,$$

$$g_2(t, b) = (1 - \varepsilon)E_{F_0}\psi(H_{t,b})D_b = m - 1.$$

From Lemma 3.2, we may deduce that (4.1) and (4.3) are actually equivalent.

The solutions of (4.3) and (4.4) (whose existence and uniqueness will follow from Theorem 4.1) will be denoted by $(t(K, \varepsilon), b(K, \varepsilon))$ and $(t(\varepsilon), b(\varepsilon))$, respectively. Although the existence of solutions of (4.4) is given by Theorem 3.1, we treat it simultaneously with the case of point mass contaminations. We see that we can actually omit condition (A1), in order to assure the existence of solutions of (4.4), and we specify the inequality that ε should satisfy, which is less stringent than in Theorem 3.1.

Theorem 4.1 proves the existence of solutions of (4.1) without imposing condition (A6). Uniqueness follows as in Maronna (1976). But we first need some lemmas.

LEMMA 4.1. *Let ψ satisfy* (A2)–(A5). *If* (A7) *and* (A8) *hold, then*:

(a) $g_1^{(i)}(t, b)$, $i = 1, 2$, *is strictly decreasing in t for a fixed b and strictly decreasing in b for a fixed t; $g_2(t, b)$ is strictly decreasing in t for a fixed b.*

(b) $g_1^{(i)}(t, b)$, $i = 1, 2$, *is strictly decreasing in b for a fixed t and $g_2(t, b)$ is strictly increasing in b for a fixed t provided* (A1) *also holds.*

LEMMA 4.2. *Let $h(b) = E_{F_0}D_b$. Then*:

(a) $h(b)$ *is a continuous, bounded and strictly increasing function.*

(b) *Let $\varepsilon < 1/m$. Then $b^* = \inf\{b: (1 - \varepsilon)\psi(\infty)h(b) \geq m - 1\}$. Then $b^* \geq 1$ if and only if $K_2 \geq m/(1 - \varepsilon)$.*

(c) *Let* $b^{**} = \sup\{b: (1-\varepsilon)\psi(0)h(b) \le m - 1\}$. *Then* $b^{**} = \infty$ *if and only if* $\psi(0) \le (m-1)/(1-\varepsilon)$.

All the verifications for the previous lemma are straightforward.

Let $t_1^{(i)}(b)$ and $t_2(b)$ be any solutions to $g_1^{(i)}(t, b) = m$ and $g_2(t, b) = m - 1$, $i = 1, 2$, respectively. The following lemma shows that $t_1^{(i)}(b)$ and $t_2(b)$ are unique for each b, defining continuous functions.

LEMMA 4.3.   *Let us suppose that* (A2)–(A5), (A7) *and* (A8) *hold. Let* $\varepsilon < 1/m$. *Then*:

(a) $t_1^{(1)}(b)$ *exists for every* b *and* $t_2(b)$ *exists for* $b \in (b^*, b^{**})$. *If* $\varepsilon < \min \{1/\psi(\infty), (m - \psi(0))/(\psi(\infty) - \psi(0))\}$, $t_1^{(2)}(b)$ *exists for every* b. *Moreover,* $t_1^{(1)}(b)$, $t_1^{(2)}(b)$ *and* $t_2(b)$ *are unique if they exist, defining continuous functions.*
(b) $\lim_{b \to b^{**}} t_2(b) = \infty$.
(c) $\lim_{b \to b^*} t_2(b) = 0$.

THEOREM 4.1.   *Let us suppose that* (A2)–(A5), (A7) *and* (A8) *hold. Then*:

(a) *If* $\varepsilon < 1/m$, *then* (4.3) *admits a solution.*
(b) *If* $\varepsilon < \min\{1/\psi(\infty), (m - \psi(0))/(\psi(\infty) - \psi(0))\}$, *then there exists a solution of* (4.4).

Theorem 4.1 and Lemma 3.2(b) say that (4.1) with a diagonal matrix **V** admits solution. The uniqueness of the solution of (4.1) follows using the same arguments as in Maronna (1976). Therefore the solutions of (4.3) and (4.4) are unique.

REMARK.   When $\varepsilon < 1 - m/\psi(\infty)$, existence of solutions for a distribution G belonging to an ε-neighborhood follows from results in Huber (1981) and Maronna(1976).

We next describe the behavior of $b(K, \varepsilon)$ and $t(K, \varepsilon)$. We show that both estimators increase monotonically with the contamination δ_K. The following lemma determines the values of (K, ε) for which $(t(K, \varepsilon), b(K, \varepsilon))$ remains bounded away from 0 and ∞ in $\Re^+ \times \Re^+$.

LEMMA 4.4.   *Let us assume* (A2)–(A5), (A7) *and* (A8) *hold and* $\varepsilon_0 < 1/m$. *Then*:

(a) $\liminf_{K \to K_0, \varepsilon \to \varepsilon_0} b(K, \varepsilon) > 0$ *for all* $K_0 \in [0, \infty]$.
(b) $\limsup_{K \to K_0, \varepsilon \to \varepsilon_0} t(K, \varepsilon) < \infty$ *if and only if* $K_0 < \infty$ *or* $\varepsilon_0 < \min \{1/\psi(\infty), (m - \psi(0))/(\psi(\infty) - \psi(0))\}$.
(c) $\limsup_{K \to K_0, \varepsilon \to \varepsilon_0} b(K, \varepsilon) < \infty$ *if and only if* $K_0 < \infty$ *or* $\varepsilon_0 < \min \{1/\psi(\infty), (m - \psi(0))/(\psi(\infty) - \psi(0))\}$.
(d) $\liminf_{K \to K_0, \varepsilon \to \varepsilon_0} t(K, \varepsilon) > 0$ *if and only if* $\varepsilon_0 < 1 - (m - \psi(0))/(\psi(\infty) - \psi(0))$ *or* $K_0 > 0$.

LEMMA 4.5. *Under* (A2)–(A5), (A7) *and* (A8), $(t(K, \varepsilon), b(K, \varepsilon))$ *is continuous in* $\Re^+ \times (0, 1/m)$.

The following lemma shows that $b(K, \varepsilon)$ is nondecreasing in K, which implies that there exist contaminations which produce biases as close to the maximum as possible.

LEMMA 4.6. *Under* (A2)–(A5), (A7) *and* (A8), $b(K, \varepsilon)$ *is nondecreasing in* $K > 0$.

The estimator $t(K, \varepsilon)$ shows the same qualitative behavior as $b(K, \varepsilon)$ under point mass contamination.

COROLLARY 4.1. *Assume* (A1)–(A5), (A7) *and* (A8). *Then* $t(K, \varepsilon)$ *is nondecreasing in* $K > 0$.

The next corollary shows that the maximum eigenvalue corresponds to the direction of the point mass contamination.

COROLLARY 4.2. *Assume* (A1)–(A5), (A7) *and* (A8). *Then* $b(K, \varepsilon) > 1$ *for all* $K > 0$.

REMARK. Lemma 4.6 and Corollary 4.1 imply that $\lim_{K \to \infty} b(K, \varepsilon) = \sup_K b(K, \varepsilon) = b^*$ and $\lim_{K \to \infty} t(K, \varepsilon) = \sup_K t(K, \varepsilon) = t^*$. By the dominated convergence theorem, (t^*, b^*) verifies (4.4). By uniqueness, $(t^*, b^*) = (t(\varepsilon), b(\varepsilon))$.

5. Optimality of Tyler's functional. In this section we prove that Tyler's functional as defined by (2.1) is bias-minimax. Let $\varepsilon_1 < \varepsilon_2 < \varepsilon^*(\mathbf{V}, F_0)$. Therefore, $b(\varepsilon_1) \le b(\varepsilon_2)$ since $\mathscr{F}_{\varepsilon_1} \subseteq \mathscr{F}_{\varepsilon_2}$. In Lemma 3.1 we have seen that $\varepsilon^*(\mathbf{V}, F_0)$ depends on the quantity $1 - m/\psi(\infty)$, which is related to implosion, while $b(\varepsilon)$ is just related to explosion. Then, the following lemma shows that the function $b(\varepsilon)$ is monotone for ε beyond the breakdown point.

LEMMA 5.1. *Assume* (A2)–(A5), (A7) *and* (A8) *hold. Then the maximum bias function* $b(\varepsilon)$: $[0, \min\{1/\psi(\infty), (m - \psi(0))/(\psi(\infty) - \psi(0))\}) \to [1, \infty)$ *is nondecreasing*.

Figure 1 displays the bias function $\sqrt{b_{\mathbf{T}}(\varepsilon)}$ for Tyler's functional and the bias function $\sqrt{b(\varepsilon)}$ for the M-functional obtained by using $u(t) = (m + 1)/(t + 1)$. This M-estimator coincides with the maximum likelihood estimator for the m-variate Cauchy distribution. In our case $m = 2$, and we assume that F_0 is Gaussian at the target model. The vertical line drawn at $1/\psi(\infty) = 0.269$ marks the breakdown point of the M-functional.

We observe that the maximum bias $b(\varepsilon)$ of the M-functional is always greater than that of Tyler's depicted by the function $b_{\mathbf{T}}$. The behavior displayed at the picture is not an isolated case as will be proved in Theorem 5.1.

The example takes a decreasing density function in $\|\mathbf{z}\|$. We will obtain the optimality result under such a condition.

(A9) $f_0(u)$ is decreasing in u for $u \geq 0$.

Theorem 3.1 seems to indicate the convenience of choosing a ψ penalizing severely a contamination at the infinity, since it provides the worst case for maximum bias. Besides, we have the evidence given in Figure 1. Hence, Tyler's functional with $\psi = m$ seems to be the most natural candidate. Let $\mathscr{C}$ be the set of bounded, continuous, nonnegative and nondecreasing functions defined on $\Re^+$. The idea of the proof is to define a curve $\tilde{\psi}_2^{(c)} \in \mathscr{C}$ so that ψ and m are connected by such a curve. If the M-functional solving (4.4) depends continuously and monotonically on the parameter c, we will be able to prove that the maximum bias of Tyler's functional is a lower bound for the maximum bias of any M-functional. More precisely, let us consider $\tilde{\psi}_2^{(c)}(x) = m\psi(cx)(E_{F_0}\psi(c\mathbf{z}'\mathbf{z}))^{-1}$ for $c \geq 1$, where ψ satisfies sufficient conditions for existence and uniqueness of the estimators. The solutions to (4.4) will be denoted as $(t_c(\psi), b_c(\psi))$. However, the solutions to (4.4) for $\psi^{(c)}(x) = m\psi(x)(E_{F_0}\psi(c\mathbf{z}'\mathbf{z}))^{-1}$, are $(c^{-1}t_c(\psi), b_c(\psi))$, and consequently $\tilde{\psi}_2^{(c)}$ as well as $\psi^{(c)}$ yield the same maximum bias. Then we use $\psi^{(c)}$ instead of $\tilde{\psi}_2^{(c)}$ and we can establish the following result.

THEOREM 5.1.   *Tyler's functional is bias-minimax in the class of M-functionals of multivariate scatter which satisfy conditions* (A1)–(A9), *that is,* $\underline{b}(\mathbf{V}_T, F_0, \varepsilon) \leq \underline{b}(\mathbf{V}, F_0, \varepsilon)$.
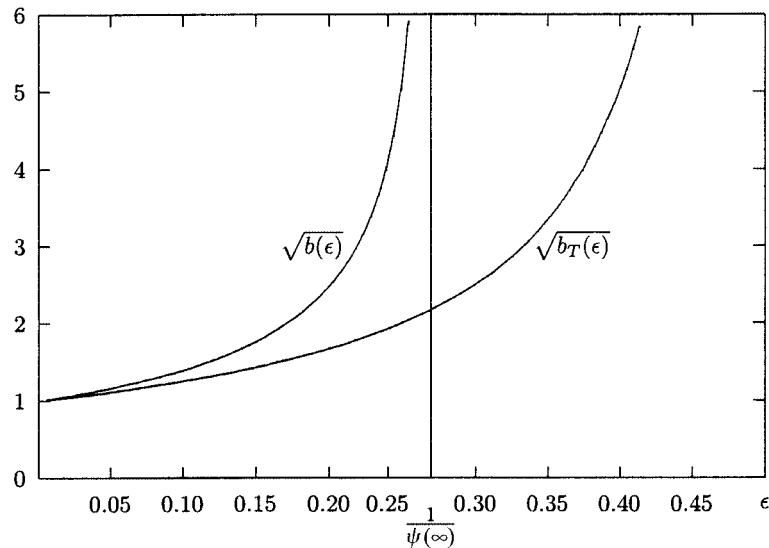

FIG. 1.   *Maximum bias functions of Tyler's functional and an M-functional.*

APPENDIX

PROOF OF LEMMA 3.2. (a) Let us take $\mathbf{B}^{(i,\,j)}$ the orthogonal matrix obtained by transposing the ith. and the jth row in the identity matrix. The result follows by considering $\mathbf{B}^{(i,\,j)}\mathbf{z}$ and the spherical symmetry of $\mathbf{z}$.

(b) Let $\mathbf{B}^{(j)}$ be a diagonal matrix such that every diagonal element is 1 unless the jth element which is -1 and consider $\mathbf{B}^{(j)}\mathbf{z}$. Hence, the result follows by the spherical symmetry of $\mathbf{z}$.

(c) The jth diagonal element of $S(F_0, \mathbf{V})$ is

$$E_{F_0}\psi\left(\sum_{i=1}^m \frac{z_i^2}{a_i}\right)\frac{z_j^2/a_j}{\sum_{i=1}^m z_i^2/a_i}.$$

Since $a_j^{-1} \leq a_{j+1}^{-1}$, then the result follows.

(d) By (b), $S(F_0, s\mathbf{I})$ is a multiple of the identity for all $s > 0$. Let $c = E_{F_0}\psi(s^{-1}\mathbf{z}'\mathbf{z})/m$. By choosing s_0 such that $c = 1$, $S(F_0, s_0\mathbf{I}) = \mathbf{I}$. □

PROOF OF LEMMA 3.3. (a) The inclusion $\subseteq$ is obvious. To see the other one let us take $\mathbf{A} = \sum_{i=1}^m \lambda_i(\mathbf{A})\mathbf{t}_i\mathbf{t}_i'$ where $\mathbf{t}_1,\ldots\mathbf{t}_m$ are a basis of orthonormal vectors in $\Re^m$. Therefore taking the discrete measure

$$H = H(\mathbf{A}, \mathbf{V}) = \begin{cases} \displaystyle\sum_{i=1}^m \frac{\lambda_i(\mathbf{A})}{\mathrm{trace}(\mathbf{A})}\delta_{r\mathbf{V}^{1/2}\mathbf{t}_i}, & \text{if } \mathrm{trace}(\mathbf{A}) > 0, \\[2mm] \delta_{\mathbf{0}}, & \text{if } \mathrm{trace}(\mathbf{A}) = 0, \end{cases}$$

where $r > 0$ such that $\psi(r^2) = \mathrm{trace}(\mathbf{A}) > 0$, we get that $S(H, \mathbf{V}) = \mathbf{A}$.

(b) Follows completely similarly.

(c) Let $\mathbf{A} \in \mathscr{A}_T$. Then $S_T(H_T, \mathbf{V}) = \mathbf{A}$, where

$$H_T = H_T(\mathbf{A}, \mathbf{V}) = \sum_{i=1}^m \frac{\lambda_i(\mathbf{A})}{m}\delta_{\mathbf{V}^{1/2}\mathbf{t}_i}.$$

(d) It follows from (a), (b) and (c). □

PROOF OF LEMMA 4.3. (a) (A2) and (A3) imply that for a fixed b, the functions $g_1^{(2)}(t, b)$ and $g_2(t, b)$ are continuous in $t \in (0, \infty)$. By (A5) we get that $\lim_{t\to 0} g_1^{(i)}(t, b) = \psi(\infty) > m$, $i = 1, 2$ and $\lim_{t\to\infty} g_1^{(1)}(t, b) = \psi(0) < m$. Since $\varepsilon < (m - \psi(0))/(\psi(\infty) - \psi(0))$, it holds that

$$\lim_{t\to\infty} g_1^{(2)}(t, b) = (1 - \varepsilon)\psi(0) + \varepsilon\psi(\infty) < m.$$

By Bolzano's theorem, $t_1^{(i)}(b)$ exists for all b. On the other hand, $\lim_{t\to 0} g_2(t, b) = (1 - \varepsilon)\psi(\infty)h(b) > m - 1$ if $b > b^*$ and $\lim_{t\to\infty} g_2(t, b) = (1 - \varepsilon)\psi(0)h(b) < m - 1$ if $b < b^{**}$. Thus, by the continuity of $g_2(t, b)$ there exists $t_2(b)$ for $b^* < b < b^{**}$. Let us observe that $0 < \liminf_{b_n\to b} t_1^{(i)}(b_n) \leq \limsup_{b_n\to b} t_1^{(i)}(b_n) < \infty$ if $0 < b < \infty$ and $0 < \liminf_{b_n\to b} t_2(b_n) \leq \limsup_{b_n\to b} t_2(b_n) < \infty$ if $b^* <$

$b < b^{**}$. Then we conclude that $t_1^{(i)}(b)$ and $t_2(b)$ are continuous by using that $g_1^{(i)}(t, b)$ and $g_2(t, b)$ are strictly decreasing and continuous in t.

(b) Suppose that $t_1 = \liminf_{b \to b^{**}} t_2(b) < \infty$. If $\psi(0) \leq (m-1)/(1-\varepsilon)$, then $\limsup_{b \to b^{**}} g_2(t_2(b), b) = (1-\varepsilon)\psi(\infty) = m - 1$. This implies $\varepsilon = 1 - (m-1)/\psi(\infty) > 1/m$ contradicting the assumptions on ε. If $\psi(0) > (m-1)/(1-\varepsilon)$ then $m - 1 = (1-\varepsilon)\psi(0)h(b^{**}) < g_2(t_1, b^{**})$ which is again a contradiction and $\lim_{b \to b^{**}} t_2(b) = \infty$. On the other hand,

$$(A.1) \qquad\qquad m - 1 = g_2(t_2(b), b) < (1-\varepsilon)\psi(\infty)h(b).$$

Therefore, $b^* > 0$. If $\liminf_{b \to b^*} t_2(b) > 0$, using (A.1) we contradict that b^* is an infimum and $\liminf_{b \to b^*} t_2(b) > 0$. $\square$

PROOF OF THEOREM 4.1.   Lemma 4.3 implies that $t_1^{(1)}(b)$ and $t_2(b)$ intersect for some $b > b^*$. Hence, (4.3) can be solved and (a) holds; (b) follows, using a similar argument. $\square$

PROOF OF LEMMA 4.4.   (a) Suppose that (a) does not hold. Then taking limits in (4.3) leads to the obvious contradiction $m - 1 = 0$.

(b) We simply prove by contradiction using the first equation of (4.2) and (4.3). Assertions (c) and (d) follow, using a similar argument. $\square$

PROOF OF LEMMA 4.5.   Let $\lim_{n \to \infty} K_n = K_0 \in (0, \infty)$ and $\lim_{n \to \infty} \varepsilon_n = \varepsilon_0 \in (0, 1/m)$. By Lemma 4.4, both $\{t(K_n, \varepsilon_n)\}_{n=1}^{\infty}$ and $\{b(K_n, \varepsilon_n)\}_{n=1}^{\infty}$ are bounded. By taking a subsequence if necessary, we may assume that $\lim_{n \to \infty} t(K_n, \varepsilon_n) = t_0$ and $\lim_{n \to \infty} b(K_n, \varepsilon_n) = b_0$. The dominated convergence theorem implies that (t_0, b_0) solves (4.3) if $K = K_0$ and $\varepsilon = \varepsilon_0$. Therefore, by uniqueness of the solution, $t(K_0, \varepsilon_0) = t_0$ and $b(K_0, \varepsilon_0) = b_0$. This implies that $(t(K, \varepsilon), b(K, \varepsilon))$ is continuous. $\square$

PROOF OF LEMMA 4.6.   Let $K_0 < K_1 < \infty$. Thus

$$g_1^{(1)}(t, b, K_0) \leq g_1^{(1)}(t, b, K_1),$$

and consequently $g_1^{(1)}(t_1^{(1)}(b, K_0), b, K_1) \geq g_1^{(1)}(t_1^{(1)}(b, K_0), b, K_0) = m$. Since $g_1^{(1)}(t, b, K)$ is decreasing in t,

$$(A.2) \qquad\qquad t_1^{(1)}(b, K_1) \geq t_1^{(1)}(b, K_0).$$

Let (t_0, b_0) and (t_1, b_1) be the solutions of (4.3) when the contaminations are δ_{K_0} and δ_{K_1} respectively. Then $t_0 = t_1^{(1)}(b_0, K_0)$ and $t_1 = t_1^{(1)}(b_1, K_1)$. By the

existence and uniqueness of solutions of (4.3) we obtain

(A.3)

$$t_1^{(1)}(b, K_0) < t_2(b, K_0) \quad \text{if } b > b_0,$$

$$t_1^{(1)}(b, K_0) > t_2(b, K_0) \quad \text{if } b < b_0,$$

$$t_1^{(1)}(b, K_1) < t_2(b, K_1) = t_2(b, K_0) \quad \text{if } b > b_1,$$

$$t_1^{(1)}(b, K_1) > t_2(b, K_1) = t_2(b, K_0) \quad \text{if } b < b_1.$$

When $b_1 < b_0$, (A.2) and (A.3) imply

$$t_1 = t_1^{(1)}(b_1, K_1) \geq t_1^{(1)}(b_1, K_0) > t_2(b_1, K_0)$$

$$= t_2(b_1, K_1) = t_1^{(1)}(b_1, K_1) = t_1,$$

which is a contradiction. Therefore, $b_0 \leq b_1$ and $b(K_0, \varepsilon) \leq b(K_1, \varepsilon)$ for $K_0 < K_1$. □

PROOF OF COROLLARY 4.2.   Let $\tilde{K} = \inf\{K \colon b(K, \varepsilon) \geq 1\}$. If $\tilde{K} \in (0, \infty)$, then the continuity of $b(K, \varepsilon)$ entails that $b(\tilde{K}, \varepsilon) = 1$ and (4.2) becomes

$$(1 - \varepsilon)E_{F_0}\psi\left(\frac{R^2}{t}\right)y_1^2 + \varepsilon\psi\left(\frac{\tilde{K}}{t}\right) = 1,$$

$$(1 - \varepsilon)E_{F_0}\psi\left(\frac{R^2}{t}\right)y_j^2 = 1, \qquad 2 \leq j \leq m.$$

Consequently, we get that $\psi\left(\tilde{K}/t\right) = 0$. This implies that either $\tilde{K} = 0$ and $b(K, \varepsilon) > 1$ for all $K > 0$ or $\tilde{K} = \infty$ and $b(K, \varepsilon) < 1$ for all $K > 0$. Since $b(\varepsilon) > 1$ and $\lim_{K \to \infty} b(K, \varepsilon) = b(\varepsilon)$, the corollary follows. □

PROOF OF LEMMA 5.1.   If $\varepsilon_0 < \varepsilon_1$ then $g_1^{(2)}(t, b, \varepsilon_0) \leq g_1^{(2)}(t, b, \varepsilon_1)$, and consequently $m = g_1^{(2)}(t_1^{(2)}(b, \varepsilon_0), b, \varepsilon_0) \leq g_1^{(2)}(t_1^{(2)}(b, \varepsilon_0), b, \varepsilon_1)$. Since $g_1^{(2)}(t, b, \varepsilon)$ is decreasing in t,

$$t_1^{(2)}(b, \varepsilon_1) \geq t_1^{(2)}(b, \varepsilon_0).$$

On the other hand, $g_2(t, b, \varepsilon_0) \geq g_2(t, b, \varepsilon_1)$ and $m - 1 = g_2(t_2(b, \varepsilon_0), b, \varepsilon_0) \geq g_2(t_2(b, \varepsilon_0), b, \varepsilon_1)$. Let $b^*(\varepsilon)$ and $b^{**}(\varepsilon)$ be defined as in Lemma 4.2. Then $b^*(\varepsilon)$ and $b^{**}(\varepsilon)$ are nondecreasing in ε. As in Lemma 4.3(a) we know that $t_2(b, \varepsilon)$ exists if $b \in (b^*(\varepsilon), b^{**}(\varepsilon))$. Since $g_2(t, b, \varepsilon)$ is decreasing in t, $t_2(b, \varepsilon_0) \geq t_2(b, \varepsilon_1)$ if $b \in (b^*(\varepsilon_1), b^{**}(\varepsilon_0))$. Let (t_0, b_0) and (t_1, b_1) be the solutions of (4.4) for ε_0 and ε_1, respectively. Then $t_0 = t_1^{(2)}(b_0, \varepsilon_0)$ and $t_1 = t_1^{(2)}(b_1, \varepsilon_1)$. By the existence and uniqueness of solutions of (4.4) we obtain that

$$t_1^{(2)}(b, \varepsilon_0) < t_2(b, \varepsilon_0) \quad \text{if } b > b_0,$$

$$t_1^{(2)}(b, \varepsilon_0) > t_2(b, \varepsilon_0) \quad \text{if } b < b_0,$$

$$t_1^{(2)}(b, \varepsilon_1) < t_2(b, \varepsilon_1) \leq t_2(b, \varepsilon_0) \quad \text{if } b > b_1.$$

In case that $b_1 < b_0$, we have

$$t_1 = t_1^{(2)}(b_1, \varepsilon_1) \geq t_1^{(2)}(b_1, \varepsilon_0) > t_2(b_1, \varepsilon_0) \geq t_2(b_1, \varepsilon_1) = t_1^{(2)}(b_1, \varepsilon_1) = t_1,$$

which is a contradiction. Therefore, $b_0 \leq b_1$ and $b(\varepsilon_0) \leq b(\varepsilon_1)$ for $\varepsilon_0 < \varepsilon_1$. □

PROOF OF THEOREM 3.1.   Using the spectral decomposition of a matrix $\mathbf{V} \in \mathscr{E}$ and the spherical symmetry of F_0, one can easily show that it suffices to consider diagonal matrices $\mathbf{V} = \mathrm{diag}(\mathbf{a}')$ with vector $\mathbf{a}$ in $\Lambda = \{\mathbf{a} \in \Re^m : a_1 \geq a_2 \cdots \geq a_m > 0\}$. For $\mathbf{a} \in \Lambda$, define

$$\tilde{\mathbf{a}} = (a_1, a_m, \ldots, a_m)'.$$

From Lemma 3.2 we obtain that $S_T(F_0, \mathrm{diag}(\mathbf{a}'))$ and $S_T(F_0, \mathrm{diag}(\tilde{\mathbf{a}}'))$ are diagonal matrices with nondecreasing diagonal elements. Thus $a_1/a_m = \tilde{a}_1/\tilde{a}_m$ and

$$\frac{\lambda_1}{\lambda_m}(\mathrm{diag}(\tilde{\mathbf{a}}')) = \frac{\lambda_1}{\lambda_m}(\mathrm{diag}(\mathbf{a}')) = \frac{a_1}{a_m},$$

$$\begin{aligned}
\mathrm{trace}(S(F_0, \mathrm{diag}(\tilde{\mathbf{a}}'))) &= E_{F_0}\psi\big(\mathbf{z}'\,\mathrm{diag}(\tilde{\mathbf{a}}')^{-1}\mathbf{z}\big) \\
&\geq E_{F_0}\psi\big(\mathbf{z}'\,\mathrm{diag}(\mathbf{a}')^{-1}\mathbf{z}\big) \quad \text{[by (A3)]} \\
&= \mathrm{trace}(S(F_0, \mathrm{diag}(\mathbf{a}'))),
\end{aligned}$$

$$\begin{aligned}
\lambda_1(S(F_0, \mathrm{diag}(\tilde{\mathbf{a}}'))) &= E_{F_0}\frac{\psi\big(\mathbf{z}'\,\mathrm{diag}(\tilde{\mathbf{a}}')^{-1}\mathbf{z}\big)}{\mathbf{z}'\,\mathrm{diag}(\tilde{\mathbf{a}}')^{-1}\mathbf{z}}\frac{z_m^2}{a_m} \\
&\leq E_{F_0}u\big(\mathbf{z}'\,\mathrm{diag}(\tilde{\mathbf{a}}')^{-1}\mathbf{z}\big)\frac{z_m^2}{a_m} \quad \text{[by (A1)]} \\
&= \lambda_1(S(F_0, \mathrm{diag}(\mathbf{a}'))).
\end{aligned}$$

Similarly,

$$\lambda_1\big(S_T(F_0, \mathrm{diag}(\tilde{\mathbf{a}}'))\big) \leq \lambda_1\big(S_T(F_0, \mathrm{diag}(\mathbf{a}'))\big).$$

Hence it suffices to consider $\mathbf{V} = \mathrm{diag}(tb, t, \ldots, t)$ in (3.2), where $b \geq 1$ and $t > 0$. Then, we may write

$$\mathrm{trace}\big(S(F_0, \mathrm{diag}(tb, t, \ldots, t))\big) = E_{F_0}\psi\big(H_{t,b}\big),$$

$$\begin{aligned}
\lambda_1\big(S(F_0, \mathrm{diag}(tb, t, \ldots, t))\big) &= E_{F_0}\psi\big(H_{t,b}\big)\frac{y_m^2}{1 - y_1^2(1 - b^{-1})} \\
&= E_{F_0}\psi\big(H_{t,b}\big)\frac{(1 - y_1^2)/(m-1)}{1 - y_1^2(1 - b^{-1})},
\end{aligned}$$

$$\lambda_1\big(S_T(F_0, \mathrm{diag}(tb, t, \ldots, t))\big) = m E_{F_0}D_b/(m-1).$$

This entails the result. □

In order to prove Theorem 5.1, we need to show that the M-functionals are continuous, in a certain sense, with respect to the function $\psi \in \mathscr{C}$ which appears in their definition. The following lemma establishes formally the M-functionals' continuity and its proof can be found in Adrover (1993).

LEMMA A.1.   *Assume* (A1)–(A6) *hold. Let* $\mathbf{V} = \sum_{i=1}^{m} \lambda_i(\mathbf{V})\mathbf{t}_i\mathbf{t}_i' \in \mathscr{E}$, *where* $\lambda_1(\mathbf{V}) \geq \cdots \geq \lambda_m(\mathbf{V})$ *are the eigenvalues of* $\mathbf{V}$ *and* $\mathbf{t}_1, \ldots, \mathbf{t}_m$ *their corresponding eigenvectors. Then*:

(a) $\lambda_m(\mathbf{V}) \leq \mathbf{z}'\mathbf{z}/(\mathbf{z}'\mathbf{V}^{-1}\mathbf{z}) \leq \lambda_1(\mathbf{V})$ *for any* $\mathbf{z} \in \Re^m - \{\mathbf{0}\}$.
(b) *Suppose that* $\lim_{n\to\infty} \sup_{x\in\Re} |\psi^{(n)}(x) - \psi(x)| = 0$. *Let* $\mathbf{V}^{(n)}$ *and* $\mathbf{V}$ *denote the solutions of* (2.3) *associated with* $\psi^{(n)}$ *and* ψ *respectively. Suppose that there exists a compact set* K *in* $(\Re^+)^m$ *such that* $(\lambda_1(\mathbf{V}^{(n)}), \ldots, \lambda_m(\mathbf{V}^{(n)})) \in K$ *for all* $1 \leq i \leq m$ *and* n. *Then* $\lim_{n\to\infty} \lambda_i(\mathbf{V}^{(n)}) = \lambda_i(\mathbf{V})$.
(c) *If* $\lim_{n\to\infty} \sup_{x\in\Re^+} |\psi^{(n)}(x) - \psi(x)| = 0$, *then the solutions to* (4.4) *corresponding to* $\psi^{(n)}$ *remain bounded and the conclusion in part* (b) *still holds*.

We next define some function subsets in $\mathscr{C}$, which will be used in Theorem 5.1:

$$\mathscr{C}_c = \big\{\psi \in \mathscr{C}: x_0(\psi) < \infty\big\},$$

$$\mathscr{C}_d = \Big\{\psi \in \mathscr{C}: \text{ is continuously differentiable in } \Re^+,$$
$$\sup_{x\in\Re} \psi'(x) < \infty \text{ and } \psi' > 0 \text{ in } (0, x_0(\psi))\Big\},$$

$$\mathscr{C}_{c,d} = \mathscr{C}_c \cap \mathscr{C}_d.$$

Let us note that the set $\mathscr{C}_{c,d}$ is dense in $\mathscr{C}$ with respect to the uniform convergence topology. This fact will be used in the proof of Theorem 5.1.

PROOF OF THEOREM 5.1.   Let $\psi \in \mathscr{C}_{c,d}$. Let us consider the functions $\psi^{(c)}(x) = (E_{F_0}\psi(c\mathbf{z}'\mathbf{z}))^{-1}m\psi(x)$, $c \geq 1$. Then, putting $\psi^{(c)}$ in (4.4) we obtain

(A.4)
$$g_1(t, b, \varepsilon, c) = (1 - \varepsilon)E_{F_0}\psi\big(H_{t,b}\big) + \varepsilon\psi(\infty) - E_{F_0}\psi(c\mathbf{z}'\mathbf{z}) = 0,$$
$$g_2(t, b, \varepsilon, c) = (1 - \varepsilon)mE_{F_0}\psi\big(H_{t,b}\big)D_b - (m - 1)E_{F_0}\psi(c\mathbf{z}'\mathbf{z}) = 0.$$

Then, by differentiating (A.4) under the integral sign and using the monotonicity of the corresponding functions we get

$$g_{t,1}(t, b, \varepsilon) = -\frac{1 - \varepsilon}{t^2}E_{F_0}\psi'(H_{t,b})R^2(1 - y_1^2(1 - b^{-1})) < 0,$$

$$g_{t,2}(t, b, \varepsilon) = -\frac{(1 - \varepsilon)m}{t^2}E_{F_0}\psi'(H_{t,b})R^2(1 - y_1^2) < 0,$$

$$g_{b,1}(t, b, \varepsilon) = \frac{1 - \varepsilon}{t}E_{F_0}\psi'(H_{t,b})R^2(-y_1^2(b^{-2})) \leq 0,$$

$$g_{\varepsilon,1}(t, b, \varepsilon) = \psi(\infty) - E_{F_0}\psi(H_{t,b}),$$

$$g_{\varepsilon,2}(t, b, \varepsilon) = -mE_{F_0}\psi(H_{t,b})D_b,$$

where the first subscript refers to the variable and the second one is 1 or 2, referring to the first or the second equation in (A.4), respectively.

The solution to (A.4) will be denoted by $(t(\varepsilon, c), b(\varepsilon, c))$ (for $c = 1$, $(t(\varepsilon, 1)$, $b(\varepsilon, 1) = (t(\varepsilon), b(\varepsilon)))$. For ε fixed, such functions are continuous in c by Lemma A.1(c). By Lemma 5.1, $b(\varepsilon, c)$ is nondecreasing in ε for fixed c. Let $t_2(b, \varepsilon, c)$ be such that $g_2(t_2(b, \varepsilon, c), b, \varepsilon, c) = m - 1$ and $t_1(b, \varepsilon, c)$ be such that $g_1(t_1(b, \varepsilon, c), b, \varepsilon, c) = m$. Using the implicit function theorem and Lemma 5.1, the function $t(\varepsilon, c) = t_2(b(\varepsilon, c), \varepsilon, c) = t_1(b(\varepsilon, c), \varepsilon, c)$ turns out to be differentiable with respect to ε except in a zero Lebesgue measure set for fixed c. Let $E_0(c) \subseteq [0, \min\{1/\psi(\infty), (m - \psi(0))/(\psi(\infty) - \psi(0))\})$ be such a set. Differentiating g_1 and g_2 with respect to ε, we obtain a system of equations whose determinant does not vanish if $\varepsilon \in E_0(c)$. Moreover, we can assure that the determinant does not vanish if $\varepsilon \in \mathbf{O}(c) = \bigcup_{\tilde{\varepsilon} \in E_0(c)} I(\tilde{\varepsilon})$ where $I(\tilde{\varepsilon})$ is an interval centered at $\tilde{\varepsilon}$ where the determinant does not vanish by the continuity of the functions involved. Here $\mathbf{O}(c)$ is an open dense set in $[0, \min\{1/\psi(\infty), (m - \psi(0))/(\psi(\infty) - \psi(0))\})$. Let $\{c_n\}_{n=1}^{\infty}$ be a countable dense set in $[1, \infty)$. Thus $\bigcap_{n=1}^{\infty} \mathbf{O}(c_n)$ is a dense set and $b(\varepsilon, c_j)$ is differentiable in ε for all $j \in \mathbf{N}$. Whence, we get

$$(g_{t,1}g_{b,2} - g_{t,2}g_{b,1})(t(\varepsilon, c_j), b(\varepsilon, c_j), \varepsilon, c_j) < 0 \quad \text{for all } c_j \text{ and } \varepsilon \in \bigcap_{n=1}^{\infty} \mathbf{O}(c_n).$$

Fixing ε and according to the implicit function theorem, there exist open sets $\mathbf{V}_1(c_j)$ and $\mathbf{V}_2(c_j)$ in $\Re$ such that $b(\varepsilon, c)$: $\mathbf{V}_1(c_j) \to \mathbf{V}_2(c_j)$ is continuously differentiable, and their derivative is given by

$$\frac{\partial}{\partial c} b(\varepsilon, c) = \frac{(g_{c,1}g_{t,2} - g_{c,2}g_{t,1})(t(\varepsilon, c), b(\varepsilon, c), \varepsilon, c)}{(g_{t,1}g_{b,2} - g_{t,2}g_{b,1})(t(\varepsilon, c), b(\varepsilon, c), \varepsilon, c)}.$$

Since $g_{c,2} = -(m - 1)(\partial/\partial c)E_{F_0}\psi(c\mathbf{z}'\mathbf{z}) = (m - 1)g_{c,1}$ and $g_{t,1} = g_{t,2}/m - (1 - \varepsilon/t^2)E_{F_0}\psi'(H_{t,b})R^2 y_1^2 b^{-1} = g_{t,2}/m - a(t, b) < g_{t,1}$, where $a(t, b) > 0$ for all $t > 0$, $b > 0$, then $g_{c,1}g_{t,2} - g_{c,2}g_{t,1} = g_{c,1}(g_{t,2}/m + (m - 1)a(t, b))$. Using that $g_{t,2}/m + (m - 1)a(t, b) \leq 0$ (see Lemma A.2), we get that $(\partial/\partial c)b(\varepsilon, c) \leq 0$ for $c \in \mathbf{V}_1(c_j)$. By noting that $\bigcup_{n=1}^{\infty} \mathbf{V}_1(c_n)$ is an open dense set in $[1, \infty)$, $(\partial/\partial c)b(\varepsilon, c) \leq 0$ for $c \in \bigcup_{n=1}^{\infty} \mathbf{V}_1(c_n)$ and $b(\varepsilon, c)$ is continuous for all $c \in [1, \infty)$, there results that $b(\varepsilon, c)$ is nonincreasing in $[1, \infty)$. Let us call $b_* = b_*(\varepsilon) = \lim_{c \to \infty} b(\varepsilon, c) = \inf_{c \geq 1} b(\varepsilon, c)$. Since $(\partial/\partial c)t(\varepsilon, c) \leq 0$ and $t(\varepsilon, c)$ is continuous for all $c \in [1, \infty)$, there results that $t(\varepsilon, c)$ is nonincreasing in $[1, \infty)$ and $\limsup_{c \to \infty} t(\varepsilon, c) < \infty$. In case that $0 < \liminf_{c \to \infty} t(\varepsilon, c)$, we get after taking the limit as $c \to \infty$ that $\psi(\infty) = (1 - \varepsilon)E_{F_0}\psi(H_{t_*, b_*}) + \varepsilon\psi(\infty) < \psi(\infty)$. Therefore, $\liminf_{c \to \infty} t(\varepsilon, c) = 0$ and we get that $(1 - \varepsilon)m\psi(\infty)h(b_*) = (m - 1)\psi(\infty)$. Consequently, $b_* = b_\mathbf{T}(\varepsilon)$. Since $b(\varepsilon, c)$ is nonincreasing in c, we conclude that $b(\varepsilon) \geq b_\mathbf{T}(\varepsilon)$.

The result follows for $\varepsilon \in \bigcap_{j=1}^{\infty} \mathbf{O}(c_j)$. Suppose now that $b(\varepsilon_0) < b_\mathbf{T}(\varepsilon_0)$ where $\varepsilon_0 \in (\bigcap_{j=1}^{\infty} \mathbf{O}(c_j))^c \cap [0, \min\{1/\psi(\infty), (m - \psi(0))/(\psi(\infty) - \psi(0))\})$. Using the continuity of $b(\varepsilon)$ and $b_\mathbf{T}(\varepsilon)$, there exists $\delta > 0$ such that $b(\varepsilon) < b_\mathbf{T}(\varepsilon)$ for all $\varepsilon \in (\varepsilon_0 - \delta, \varepsilon_0 + \delta)$ and this contradicts the fact that $\bigcap_{j=1}^{\infty} \mathbf{O}(c_j)$ is dense.

This proves the theorem for $\psi \in \mathscr{C}_{c,d}$. Since $\mathscr{C}_{c,d}$ is dense in $\mathscr{C}$, Lemma A.1(c) lets us conclude the result of the theorem for $\psi \in \mathscr{C}$. $\square$

LEMMA A.2. *Assume* $\psi \in \mathscr{C}$ *and that condition* (A9) *holds. Then*:

(a) $E_{F_0}\psi(1/s(z_1^2/v + v^{1/(m-1)}(R^2 - z_1^2)))$ *is nondecreasing for* $v \geq 1$ *when* $s > 0$ *is fixed.*

(b) *If* $\psi \in \mathscr{C}_{c,d}$, $g_{t,2}/m + (m-1)a(t,b) \leq 0$ *for* $b \geq 1$ *and* $t > 0$.

PROOF. (a) Maronna and Yohai [(1990), Lemma 1, page 3928] proved the lemma for the case $\psi(x) = 1 - I_{(0,t]}(x)$. More precisely, they show

$$\int I_{(0,t]}(\mathbf{z}'\Lambda^{-1}\mathbf{z})f_0(\mathbf{z}'\mathbf{z})\,d\mathbf{z} < \int I_{(0,t]}(\mathbf{z}'\Delta^{-1}\mathbf{z})f_0(\mathbf{z}'\mathbf{z})\,d\mathbf{z},$$

where $\Lambda = \text{diag}(v, a, \dots, a)$ and $\Delta = \text{diag}(v', a', \dots, a')$ with $v > v' > 1$. A straightforward argument generalizes the result first to step functions and then uses an approximation by step functions to any $\psi \in \mathscr{C}$.

(b) Let $s = tb^{1/m}$ and $v = b^{(m-1)/m}$. Then

$$g_{t,2}/m + (m-1)a(t,b)$$

$$= -\frac{1-\varepsilon}{t^2}E_{F_0}\psi'\left(\frac{R^2}{t}(1 - y_1^2(1 - b^{-1}))\right)R^2(1 - y_1^2 - (m-1)y_1^2 b^{-1})$$

$$= -v^{-(m-2)/(m-1)}\frac{1-\varepsilon}{s^2}E_{F_0}\psi'\left(\frac{1}{s}\left((R^2 - z_1^2)v^{1/(m-1)} + \frac{z_1^2}{v}\right)\right)$$

$$\quad \times \left((R^2 - z_1^2)v^{m/(m-1)} - (m-1)z_1^2\right)$$

$$= -v^{m/(m-1)}\frac{(1-\varepsilon)(m-1)}{s}\frac{\partial}{\partial v}E_{F_0}\psi\left(\frac{1}{s}\left((R^2 - z_1^2)v^{1/(m-1)} + \frac{z_1^2}{v}\right)\right) \leq 0$$

by item (a). $\square$

REFERENCES

ADROVER, J. G. (1993). Métodos robustos en modelo lineal y análisis multivariado. Ph.D. dissertation, Univ. Buenos Aires.

HUBER, P. J. (1977). Robust covariances In *Statistical Decision Theory and Related Topics* (S. S. Gupta and D. S. Moore, eds.) **2** 165–191. Academic Press, New York.

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

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

MARONNA, R. A. and YOHAI, V. J. (1990). The maximum bias of robust covariances. *Comm. Statist. Theory Methods* **19** 3925–3933.

MARONNA, R. A. and YOHAI, V. J. (1993). Bias-robust estimates of regression based on projections. *Ann. Statist.* **21** 965–990.

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

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

ROUSSEEUW, P. J. and YOHAI, V. J. (1984). Multivariate estimation with high breakdown point. In *Mathematical Statistics and Its Applications* (W. Grossmann, G. Pflug, I. Vincze and W. Wertz, eds.) 237–297. Reidel, Dordrecht.

TYLER, D. E. (1987). A distribution free M-estimator of multivariate scatter. *Ann. Statist.* **15** 234–251.

FACULTAD DE MATEMÁTICA
ASTRONOMÍA Y FÍSICA
UNIVERSIDAD NACIONAL DE CÓRDOBA
5000 CÓRDOBA
ARGENTINA
E–MAIL: adrover@mate.uncor.edu