JOURNAL OF MULTIVARIATE ANALYSIS 42, 141–161 (1992)

Bias-Robust Estimators of Multivariate Scatter Based on Projections*

RICARDO A. MARONNA

Universidad de la Plata and C.I.C.P.B.A., La Plata, Argentina

WERNER A. STAHEL

Swiss Federal Institute of Technology (ETH), Zurich, Switzerland

AND

VICTOR J. YOHAI

Universidad de Buenos Aires, CONICET, and CEMA, Buenos Aires, Argentina

Communicated by the Editors

Equivariant estimation of the multivariate scatter of a random vector X can be derived from a criterion of (lack of) spherical symmetry $g(X)$. The scatter matrix is $V=(A^{\mathrm{T}}A)^{-1}$, where A is the transformation matrix which makes AX as spherical as possible, that is, which minimizes $g(AX)$. The new class of projection estimators is based on making the spread of univariate projections as constant as possible by choosing $g(X)=\sup_{\|u\|=1}|s(u^{\mathrm{T}}X)-1|$, where s is any robust scale functional. The breakdown point of such an estimator is at least that of s, independently of the dimension p of X. In order to study the bias, we calculate condition numbers based on asymptotics and on simulations of finite samples for a spherically symmetric X, contaminated by a point mass, with the median absolute deviation as the scale measure. The simulations are done for an algorithm which is designed to approximate the projection estimator. The bias is much lower than the one of Rousseeuw's MVE-estimator, and compares favorably in most cases with two M-estimators.

Received August 15, 1991.

AMS 1980 subject classification: 62H12.

Key words and phrases: robust estimation, projection estimator, high breakdown point estimation, covariance matrix.

* Research partly supported by Mathematisches Forschungsinstitut while the first author was visiting ETH Zürich, and by ONR Contract N00014-91-J-1074 while the third author was visiting the Department of Statistics, University of Washington.

0047-259X/92 $5.00

1. INTRODUCTION

The classical approach to large parts of multivariate analysis is based on the multivariate normal distribution $N_p(\mu, V)$ with location vector μ and scatter matrix V. The maximum likelihood estimators of these parameters are the sample mean and covariance matrix, respectively. However, it is well known that slight departures from this model, in the form of a small proportion of outliers, may completely distort the estimators. Some interesting examples are shown by Devlin *et al.* [2]. It is thus desirable to study estimators which remain stable when the data are "contaminated" by a small proportion of outliers.

The breakdown point ε^* of an estimator (to be defined more precisely below) is a useful concept in robustness. Loosely speaking, ε^* is the largest proportion of arbitrary outliers that may be added to the sample, such that the estimator remains bounded. The classical estimators have $\varepsilon^* = 0$.

Robust equivariant M-estimators of multivariate location and scatter were defined and studied by Maronna [9] and generalized by Huber [4, 5]. However, it was found that these scatter estimators have breakdown point ε^* not larger than $1/p$, even if the location vector is known, whereas if the scatter matrix is known, the location estimators have $\varepsilon^* = \frac{1}{2}$. This points out a lack of robustness for large p, which affects the scatter matrix rather than the location vector. Within this class, Tyler [14] studied an estimator with breakdown point exactly equal to $1/p$.

Stahel [13] and Donoho [3] defined a class of equivariant multivariate estimators with $\varepsilon^* = \frac{1}{2}$ independent of p, but little is known about their properties. Rousseeuw [11] defined the "minimum volume ellipsoid" (MVE) estimator which has $\varepsilon^* = \frac{1}{2}$ for all p; later Davies [1] generalized this idea by defining multivariate S-estimators, with ε^* independent of p. Li and Chen [6] proposed an estimator based on robustifying principal components by replacing the variance by a robust scale estimator like the median deviation. This estimator has a breakdown point independent of p, although it is not affine equivariant. The relationship between S- and M-estimators was investigated by Lopuhaä [7].

The asymptotic bias (in a sense to be made precise below) of Rousseeuw's estimator under contamination was calculated by Yohai and Maronna [16], and it turned out that, although bounded, it may be so high as to make the estimator unreliable, even for small amounts of contamination. Thus our goal is to develop estimators which do not only have a high breakdown point (i.e., a bounded bias) but also a moderate bias under contamination.

If the random vector X has a nonsingular covariance matrix V and A is such that $V = (A^{\mathrm{T}}A)^{-1}$, then $Z = AX$ has covariance matrix I. If X is normal, then the distribution of Z is spherically symmetric (invariant with

respect to orthogonal transformations). A basic idea for defining robust equivariant estimators is therefore as follows. Choose a suitable criterion $g(Z)$ of "degree of spherical asymmetry" of a random variable Z. Define the functional $A(X)$ on the distribution of X as the matrix which minimizes the criterion for the transformed variable $A(X)X$,

$$A(X) = \arg\min_{A \in \mathscr{A}} g(AX), \tag{1}$$

where A ranges over the set $\mathscr{A}$ of nonsingular matrices. Then define the scatter matrix of X as

$$V(X) = [A(X)^T A(X)]^{-1}. \tag{2}$$

For a sample $(x_1, x_2, ..., x_n)$ the estimate is defined, as usual, as $V(X)$ with $X \sim F_n$, where F_n is the empirical distribution.

Remark 1.1. Since spherical symmetry is a quality which remains invariant under orthogonal transformations Q, natural versions of g fulfill $g(QX) = g(X)$. Then $A(X)$ is defined only up to an orthogonal matrix. To mend this, $\mathscr{A}$ can be restricted to, for example, the lower triangular matrices with positive diagonal. $V(X)$ is often unique without such a restriction.

Remark 1.2. The idea of finding A rather than V is also used for computation of M-estimators [5, 8].

Estimators of this form are affine equivariant:

LEMMA 1.1. *For any g, $V(CX) = CV(X)C^T$. If X is spherically symmetric and $V(X)$ is unique, then $V(X)$ is a multiple of the identity matrix I_p.*

All proofs are given in the Appendix.

The criteria which we propose in this paper are based on examining the scatter of the projections of the data in all directions, i.e., the linear combinations $u^T Z$ of $Z = AX$, where u is any constant vector with norm $\|u\| = 1$ and u^T denotes the transpose of u. Note that if the random vector Z has covariance matrix I, then $\text{var}(u^T Z) = 1$. Let s be any (robust) estimator of scale. The idea is that $s(u^T Z)$, where $Z = A(X)X$, should remain "as constant as possible" when u ranges over all u with $\|u\| = 1$.

This idea is made precise in Section 2. Section 3 deals with the breakdown point and with asymptotic bias. As exact calculation of the estimates defined in Section 2 is not feasible, we propose an algorithm which is meant to approximate them (Section 4). A small simulation compares the results of this algorithm with some other estimators (Section 5).

2. Definitions and General Results

Let s be a scale equivariant functional of univariate distributions. We write $s(R)$ for the scale of a univariate random variable R, and $s\{r_i\}$ for the scale of a sample. Thus $s(R)=s\{r_i\}$ if $R\sim F_n\{r_i\}$ and $F_n\{r_i\}$ is the empirical distribution defined by the sample $(r_1, r_2, ..., r_n)$.

A criterion g for spherical asymmetry, suitable for the minimization task (1), is

$$g_1(X)=\sup_{\|u\|=1}|s(u^TX)-1|.$$

Alternatively, one can minimize

$$g_2(X)=\sup_{\|u\|=1}s(u^TX)$$

under a restriction of the form $g_3(AX)\geqslant 1$, where

$$g_3(X)=\inf_{\|u\|=1}s(u^TX)$$

—or vice versa: maximize $g_3(AX)$ under $g_2(AX)\leqslant 1$. When

$$g_4(X)=g_2(X)/g_3(X)$$

is used as a criterion without restrictions, then A (and hence V) is determined only up to a scalar factor. This can be mended by requiring $g_2(AX)=1$ or $g_3(AX)=1$.

These definitions lead to the same estimators, up to a scalar factor:

Lemma 2.1. *There are some constant c_2, c_3 (which depend on X) such that the following statements are equivalent:*

(P1) *A solves* (1) *with $g=g_1$*

(P2) *c_2A solves* (1) *with $g=g_2$ under $g_3(AX)\geqslant 1$*

(P3) *c_3A solves* (1) *with $g=-g_3$ under $g_2(AX)\leqslant 1$*

(P4) *c_2A solves* (1) *with $g=g_4$ under $g_3(AX)=1$*

(P5) *c_3A solves* (1) *with $g=g_4$ under $g_2(AX)=1$*

(P6) *A solves* (1) *with $g=g_4$ under $\frac{1}{2}[g_2(AX)+g_3(AX)]=1$.*

The estimators can also be formulated directly in terms of V, avoiding A:

LEMMA 2.2. *If A is nonsingular and solves* (P1), *then* $V=(A^{\mathrm{T}}A)^{-1}$ *minimizes*

$$\max_{u\neq 0} |s(u^{\mathrm{T}}X)/(u^{\mathrm{T}}Vu)^{1/2}-1|.$$

over the set $\mathscr{V}$ of all symmetric, positive definite matrices.

Here, one looks for a matrix V such that the quadratic form $u^{\mathrm{T}}Vu$ is "as close as possible" to $s(u^{\mathrm{T}}X)$.

DEFINITION 2.1. The *projection estimator* V based on the scale functional s maps X to the set

$$V(X)=\{[A(X)^{\mathrm{T}}A(X)]^{-1}\,|\,A(X) \text{ solves (1) with } g=g_4 \text{ under } g_3(AX)=1\}.$$

THEOREM 2.3. *If* $g_2(X)=\sup_{\|u\|=1}\{s(u^{\mathrm{T}}X)\}$ *is finite and* $g_3(X)=\inf_{\|u\|=1}\{s(u^{\mathrm{T}}X)\}>0$, *then a solution to all problems* (P1) *through* (P6) *exists.*

Remark 2.1. If $g_3(X)=0$, then solutions to (P1) and (P3) still exist, but they are virtually meaningless.

The assumption about $g_2(X)$ is met if X stands for a sample and $s\{x_i\}$ is continuous or monotone in all x_i and finite.

LEMMA 2.4. *In general, these estimators are not unique.*

This can be seen by considering the distribution in $\mathbb{R}^3$ which is uniform on the two circles $\mathscr{C}_1=\{x\,|\,x_2=0,\ x_1^2+x_3^2=1\}$ and $\mathscr{C}_2=\{x\,|\,x_1=0,\ x_2^2+x_3^2=1\}$ and $s(X)=\text{range}\,(X)/2$. Clearly, $g_2(X)=1$, and $g_3(X)=s((c,c,0)^{\mathrm{T}}X)=c$ with $c=2^{-1/2}$. Two solutions of (P3) are I and diag$(1,1,c)$, as is shown in the Appendix.

Although this example sounds very artificial, it illustrates the basic difficulty: In a typical situation, the minimum of $g_4(AX)$ may essentially fix the directions u_0 and u_1 for which $s(u^{\mathrm{T}}AX)$ assumes its minimum and maximum. The elements of A which determine the components of AX orthogonal to these two directions can then vary as long as no $s(u^{\mathrm{T}}AX)$ falls outside these limits.

For the special situation of a contaminated elliptical distribution

$$X\sim F_\varepsilon=(1-\varepsilon)F+\varepsilon H, \tag{3}$$

which we shall study in more detail below, we find that the estimator is unique if s has the following property.

DEFINITION 2.2. The functional s is *monotone* (non-decreasing) if it has the following property: If $|X|$ is stochastically smaller than $|\tilde{X}|$ in the sense that $F(x) \geqslant \tilde{F}(x)$ for all x, $|X| \sim F$ and $|\tilde{X}| \sim \tilde{F}$, then $s(X) \leqslant s(\tilde{X})$.

If s is monotone in this sense, then $s\{x_i\}$ is monotone in each x_i. Unfortunately, if s is translation invariant (and scale equivariant), then it cannot be monotone. The following theorem therefore applies only if location is fixed or estimated separately.

THEOREM 2.5. *If s is monotone and $X \sim F_\varepsilon$ with F spherical and H concentrated on a straight line through* 0, *say the $x^{(1)}$-axis, then $V(X)$ is unique and diagonal with $v_{jj} = v_{22}$ for $j \geqslant 2$.*

These results still leave the question of uniqueness largely open. Our research leads us to the following conjecture.

Conjecture. The projection estimator, if applied to samples, is unique with probability 1 for $p = 2$. For $p > 2$ it is, with probability 1, non-unique, but equal to a connected, compact set.

Some basic properties of $V(X)$—unique or non-unique—are given by Lemma 1.1 and the following statements.

LEMMA 2.6. *If the scale functional s is translation invariant, then so is the scatter V.*

LEMMA 2.7. *If X is spherically symmetric, then $V(X)$ is unique and equals a multiple of the identity. Moreover, V is Fischer-consistent for elliptical distributions, in the sense that, if $X = CZ + t$ with spherical Z, then $V(X)$ is a multiple of CC^{T}. In particular, V is a multiple of the covariance matrix of X if the latter exists.*

3. RESISTANCE AND ASYMPTOTIC BIAS

The gross error breakdown point of V at F is defined as

$$\varepsilon^*(V, F) = \sup \{\varepsilon \mid \text{for some compact } K \subset \mathscr{V},\ V(X) \subset K \text{ if } X \sim (1-\varepsilon)F + \varepsilon H \text{ with arbitrary } H\} \tag{4}$$

(where $\mathscr{V}$ is the set of all symmetric, positive definite matrices as before). It measures the amount of contamination needed to carry V over all bounds or making it "almost singular." The condition number of V is the ratio $\gamma(V) = \lambda_1(V)/\lambda_p(V)$, where $\lambda_j(V)$ $(j = 1, \ldots, p)$ are the eigenvalues of V in descending order. The matrix V remains in a compact set of non-

singular matrices if and only if $\lambda_1(V)$ and $\lambda_p(V)$ are bounded away from ∞ and 0. This implies that the condition number is bounded, too.

The breakdown point of projection estimators can be obtained from a suitably defined breakdown point of the underlying scale functional.

DEFINITION 3.1. The *uniform breakdown point* $\delta^*(s, F)$ of a scale functional s at a distribution F on $\mathbb{R}^p$ is

$$\delta^*(s, F) = \sup \{\varepsilon \geq 0 \mid \text{there exist } 0 < s_1 \leq s_2 < \infty \text{ such that } (X \sim (1-\varepsilon)F + \varepsilon H, H \text{ arbitrary}, \|u\| = 1 \text{ implies } s_1 \leq s(u^T X) \leq s_2)\}.$$

For example, if s is the median deviation and F is such that $F(u^T X \neq 0) \geq \alpha$ for all $u \neq 0$, then $\delta^*(s, F) \geq \alpha - 0.5$.

THEOREM 3.1. $\varepsilon^*(V, F) \geq \delta^*(s, F)$.

Remark 3.1. Intuitively, the reversed inequality would be more plausible. In fact, equality can be shown up to a rather strange potential behavior of the estimator: We could not exclude the case that the value of the criterion g_2 might tend to infinity while the solution A remains bounded.

Asymptotic bias can be measured in the following affine equivariant way. Assume that the solution for $X \sim F$ is unique, $V(X) = V_0$, say. Then define for any G with the set $V(Y)$, $Y \sim G$, of solutions to (P4),

$$b(F, G) = \sup_{V \in V(Y)} \gamma(B^{-1} V [B^T]^{-1}),$$

where B is such that $BB^T = V_0$. The log of this function b is a measure of the discrepancy between $V(Y)$ and V_0. Since the condition number remains the same if a matrix is multiplied by a factor, only the shape of V is relevant for this measure of bias, and the bias is the same for the solutions to problems (P1) through (P6).

We shall study the bias of the estimator V_α based on the α-quantile of absolute values as the scale functional: Let F_1 be the distribution of a real random variable R, while $\tilde{F}_1$ denotes the distribution of its absolute value $|R|$. Use $s(R) = \tilde{F}_1^{-1}(\alpha)$. When F_1 is contaminated by a mass ε, then the one-dimensional scale estimate is bounded by the quantiles corresponding to $\alpha_0 = (\alpha - \varepsilon)/(1-\varepsilon)$ and to $\alpha_1 = \alpha/(1-\varepsilon)$. We denote these quantiles by $\sqrt{b_0} = \tilde{F}_1^{-1}(\alpha_0)$ and $\sqrt{b_1} = \tilde{F}_1^{-1}(\alpha_1)$. The following results for the bias are valid.

THEOREM 3.2. *Let F be elliptical. Then, with the foregoing definitions,*

$$b(F, (1-\varepsilon)F+\varepsilon H) \leqslant (b_1/b_0)^2$$

for all H.

THEOREM 3.3. *Let F be elliptical and H the point mass at $x_0 \in \mathbb{R}^p$ with $x_0^T V(F)^{-1} x_0 = c$. Then*

$$b(F, (1-\varepsilon)F+\varepsilon H) = \begin{cases} 1 & \text{if } c < b_0 \\ c/b_0 & \text{if } c \in [b_0, b_1] \\ 1 + c(b_1 - b_0)/b_0 b_1 & \text{if } c \in [b_1, 2b_1] \\ b_1(c-b_0)/[b_0(c-b_1)] & \text{if } c \geqslant 2b_1. \end{cases}$$

Hence, the maximal bias for pointwise contamination corresponds to $c = 2b_1$ and equals $2b_1/b_0 - 1$.

The maximal biases for pointwise contamination of the standard normal distribution are given in Table I for three estimators:

PE. The projection estimator taking for s the median ($\alpha = 0.5$);

MVE. Rousseeuw's [11] minimum volume ellipsoid estimator; and

TY. The extreme M-estimator described by Tyler [14].

The results show that the maximum bias of our estimator is much lower than that of Rousseeuw's, and for $p > 2$ compares favorably with Tyler's estimator.

TABLE I

Maximum Asymptotic Condition Numbers

p		0.05	0.10	0.20	0.25
All	PE	1.5	2.3	5.5	9.1
2	TY	1.2	1.6	2.8	4.0
	MVE	6.3	14.2	58.3	129.0
3	TY	1.3	1.8	4.3	8.7
	MVE	4.6	9.5	32.0	60.5
4	TY	1.4	2.1	9.0	∞
	MVE	4.1	8.1	25.8	46.9
5	TY	1.5	2.5	∞	
	MVE	3.9	7.5	23.1	
10	TY	2.2	∞	∞	
	MVE	3.6	6.7	19.3	
20	TY	∞	∞	∞	
	MVE	3.7	6.8	19.0	

4. COMPUTATION

The computation of our estimators for samples $\{x_i : i = 1, ..., n\}$ seems extremely awkward, because of the double optimization required. In general, there are multiple local extrema in both optimizations, thus making ordinary optimization methods useless.

The following algorithm is based on subsampling ideas similar to those proposed for multivariate estimators by Stahel [13] and used for regression estimators by Maronna and Yohai [10] and Rousseeuw and Leroy [12], to yield an approximate solution to problem (P4). The idea is to restrict the sets of u's and of A's over which extrema are calculated to finite sets. To make the idea more precise, we first define a more general version of the estimator, which allows for such restrictions.

Let $\mathscr{A}^* \subset \mathscr{A}$ be a set of potential solutions to (P4), possibly depending on the sample. For each $A \in \mathscr{A}^*$ let $\mathscr{U}^*(A) \subset \{u : \|u\| = 1\}$ be a set of projection directions, depending on A and possibly on the sample. Define $g_2^*(A)$ and $g_3^*(A)$ as the maximum and minimum of $s(u^T AX)$ over $u \in \mathscr{U}^*(A)$, and $g_4^*(A) = g_2^*(A)/g_3^*(A)$. Finally, find $A^* = \arg\min_{A \in \mathscr{A}^*} g_4^*(A)$ and $\tilde{A} = A^*/g_3^*(A^*)$ and define the estimator as $V^* = (\tilde{A}^T \tilde{A})^{-1} = (A^{*T} A^*)^{-1} g_3^*(A^*)^2$. If $\mathscr{A}^* = \mathscr{A}$ and $\mathscr{U}^*(A) = \{u : \|u\| = 1\}$, then V^* is the estimator defined by (P4).

We suggest generating finite sets $\mathscr{A}^*$ and $\mathscr{U}^*(A)$ as follows: For a number L of random subsamples of size $p+1$ from $\{x_1, ..., x_n\}$, determine the covariance matrix C_l, $l = 1, ..., L$. Let $\mathscr{A}^* = \{A_1, ..., A_L\}$, where A_l is a square root of C_l^{-1}, $(A_l^T A_l)^{-1} = C_l$. The set $\mathscr{A}^*$ is also used to generate $\mathscr{U}^*(A)$. Given A, compute the eigenvectors corresponding to the smallest and the largest eigenvalues of $A(A_k^T A_k)^{-1} A^T = AC_k A^T$ for $k = 1, ..., L$ with $A_l \neq A$. Then, $\mathscr{U}^*(A)$ will be the set of these $2(L-1)$ unit vectors.

The rationale for this choice of $\mathscr{U}^*(A)$ is the following. Let A be the exact solution to (P4). If n and L are large, some subsamples will have an empirical covariance matrix C_l approximately proportional to $V = (A^T A)^{-1}$, even if the sample is contaminated. In order to evaluate g_4 for A_l corresponding to such a C_l, we need to find directions u which approximately minimize and maximize $s\{u^T z_i\}$, where $z_i = A_l x_i$. In the case of contamination by outliers, many subsamples $k \neq l$ will be contaminated by "extreme" observations. The first principal component of such a subsample, transformed to Z coordinates, is a candidate for a large value of $s\{u^T z_i\}$. If the contamination is concentrated near a linear subspace through the center of the sample, then the last principal component of the transformed contaminated subsample is a candidate for a small value of $s\{u^T z_i\}$. These principal components are given, in Z coordinates, by the eigenvectors of $A_l C_k A_l^T$. Similarly, if l is a contaminated subsample, the first and last principal components of $A_l C_k A_l^T$ are candidates for extreme

values of $s\{u^T z_i\}$ if C_k corresponds to a "good" sample (C_k is proportional to V). Of course, this is not a "proof" that the procedure will yield a sensible result—this has to be checked experimentally.

This procedure would require $L(L-1)$ eigen analysis calculations, but this number can be reduced as follows. Define $g_{l,k} = \max_{j \leqslant k} s(u_j^T A_l X) / \min_{j \leqslant k} s(u_j^T A_l X)$. This is a non-decreasing function of k. If $g_{l,k}$ becomes larger than the current lowest value of $g_4^*(A_j)$, $j < l$, we may discard A_l as a candidate for the minimum. It is easy to prove, as in Maronna and Yohai [16], that the average number of computations becomes $\sim L \log L$, which is remarkable saving.

Summing up, the algorithm is the following:

1. For $l = 1, ..., L$ do:
   (a) Take a random subsample $\{x_{i_j} : j = 1, ..., p+1\}$ of size $p+1$ without replacement from the original sample.
   (b) Compute its covariance matrix C_l.
   (c) Let A_l be such that $A_l^{-1}(A_l^{-1})^T = C_l$.
   end do.
2. Let $g^* := \infty$. For $l = 1, ..., L$ do:
   (a) Let $z_i := A_l x_i$, $i = 1, ..., n$, and $t^{(0)} := \infty$, $t^{(1)} := 0$.
   (b) For $k = 1, ..., L$, $k \neq l$ do until $g \geqslant g^*$:
      i. Compute the eigenvectors $u^{(0)}$ and $u^{(1)}$ of $A_l C_k A_l^T$ corresponding to the smallest and the largest eigenvalue, respectively.
      ii. For $h = 0, 1$, let $s^{(h)} := s\{u^{(h)T} z_i : i = 1, ..., n\}$, and $t^{(0)} := \min(t^{(0)}, s^{(0)}, s^{(1)})$, $t^{(1)} := \max(t^{(1)}, s^{(0)}, s^{(1)})$,
      iii. Let $g := t^{(1)}/t^{(0)}$.
      end do.
   (c) If $k = L$ and $g < g^*$, then let $g^* := g$, $r := t^{(0)}$, and $m := l$.
   end do.
3. Estimate V by $C_m r^2$.

Remark 4.1. It is easy to verify that this choice of $\mathscr{A}^*$ and $\mathscr{U}^*(A)$ makes the result of the algorithm affine equivariant as well as independent of the particular choice of the matrix square root A_l of C_l.

Remark 4.2. The choice of L depends on p and n. A criterion might be to choose it such that, if the maximal proportion of outliers in the sample is assumed to be known, then the probability of picking a prescribed amount of subsamples free of outliers be larger than some prescribed value.

However, numerical experiments showed that, even for samples without outliers, a substantial number of subsamples is needed to get a sensible result.

We have thus used a more empirical approach to choose L: For each p, take L such that increasing it further produces remarkable changes in V only very rarely. Numerical experiments showed that $L = 200$ suffices for $p = 2$, and $L = 500$ for $p = 5$. Of course, a more rational approach deserves being developed.

The algorithm was implemented in Fortran on an IBM 3032 computer, using the IMSL routines for linear algebra. The computer time required for $p = 5$, $n = 40$, and $L = 500$ was about one minute.

In order to compare the behavior of this subsampling approximation with the "exact'" estimator, a better approximation was also computed for $p = 2$. Here it suffices to take u of the form $u^{\mathrm{T}} = (\cos \varphi, \sin \varphi)$ for $\varphi \in \Gamma = (-\pi/2, \pi/2)$. A symmetric positive definite 2×2 matrix V with unit determinant may be represented by its largest eigenvalue λ_1 and the direction of the corresponding eigenvector v_1, i.e.,

$$V = \lambda_1 v_1 v_1^{\mathrm{T}} + \lambda_2 v_2 v_2^{\mathrm{T}},$$

where $\lambda_2 = 1/\lambda_1$, $v_1 = (\cos \psi, \sin \psi)^{\mathrm{T}}$, and $v_2 = (\cos \psi, -\sin \psi)^{\mathrm{T}}$, with $\psi \in \Gamma$. The extrema in u were computed by means of a grid search over N_φ evenly spaced φ's in Γ; the minimum in V was also computed over N_ψ evenly spaced ψ's in Γ and N_λ values of λ_1 in $(1, \Lambda)$. To save computing effort, the same trick for discarding V's as in Step 2 of the former algorithm was used; a random permutation of the φ's the ψ's, and the λ_1's was performed previous to calculations. The values $N_\varphi = 100$, $N_\psi = N_\lambda = 30$, and $\Lambda = 4$ were found to suffice. In general, the estimators obtained by the subsampling algorithm were not very different from the more exact ones obtained in this way.

5. Simulation Results

The following estimators were compared in a simulation study. Location was not assumed to be known, but was estimated along with the scatter matrix.

PE. The approximate P-estimator as obtained by the subsampling algorithm, taking as scale s the median deviation.

MVE. Rousseeuw's [11] minimum volume ellipsoid, also computed by a subsampling scheme. In analogy with PE, the algorithm used was the following:

1. Same as Part 1 of the algorithm for the P-estimators.
2. For $l = 1, ..., L$ do:

   Let $y_i := A_l x_i$ $(i = 1, ..., n)$. Let μ be the vector of coordinatewise medians of the y_i, and $v_l := \text{median}\{\|y_i - \mu\|\}$, $s_l := v_l \det(C_l)^{1/p}$.
3. The estimator is $v_{l_0} C_{l_0}$, where $l_0 := \arg\min_l s_l$.

CAU. The maximum likelihood estimator for the p-variate Cauchy distribution defined by

$$\begin{aligned} \text{ave}\{v(d_i^2)(x_i - \mu)(x_i - \mu)^T\} &= V, \\ \text{ave}\{w(d_i^2)(x_i - \mu)\} &= 0, \end{aligned} \tag{5}$$

where $d_i^2 = (x_i - \mu)^T V^{-1}(x_i - \mu)$, and the weight functions v and w are defined by

$$v(u) = w(u) = (p + 1)/(u + 1)$$

[9]. The estimators are computed by means of an iterative reweighting algorithm (see Tyler [15] for a convergence proof).

TY. Tyler's [14] M-estimator. To make it shift-invariant, a location vector μ is needed. Although Tyler in his paper does not recommend any particular location estimator as a "natural" companion to his proposed matrix, we thought that a sensible choice would be to define μ and V as solutions to a system of the form (5) with $v(u) = w(u^2) = p/u$. For this particular choice, V is determined only up to a scalar factor, which can be fixed by requiring $\text{median}(d_i^2) = 1$, for example. Thus if $z_i = \mu + (p/d_i)(x_i - \mu)$, then μ and V (times a constant) are the mean and covariance matrix of the z_i's, respectively.

COV. The ordinary empirical covariance matrix.

The sampling situations considered were point mass contaminations of the standard normal distribution. For given p, n, ε, and k, $(1 - \varepsilon)n$ observations x_i were generated according to $N_p(0, I)$, and the remaining εn equal to $k(1, 0, ..., 0)^T$. The positions k of the contaminating points were chosen in order to include the worst behavior of the estimators. Since, as Table I shows, PE and MVE have a "redescending" behavior (i.e., the bias is not a monotonic function of the position of the contamination), some trials were needed.

For each estimator, the condition number of the matrix was computed as a measure of its departure from the identity. While in the asymptotic case this served as a measure of bias, now it was affected by both bias and variability, as would happen in other situations with mean squared error. It was observed that the empirical distribution of the condition numbers

was very asymmetric and heavy tailed—especially for PE and MVE. Averages of logarithms of the condition numbers are therefore reported. (Medians were also computed. In general, they did not differ much from these means.)

The number of replications was 200 in all cases. For each sampling situation, all estimators were computed from the same set of samples. For each p and ε, the same samples were used for the different k's.

The results in Table II show that:

- MVE is the worst estimator in many cases.
- TY is generally better than the "high breakdown point estimators" PE and MVE for $p=2$ and for $p=4$ and $\varepsilon=0.10$, but it behaves worse for $p=4$ and $\varepsilon=0.20$, i.e., near its breakdown point.
- CAU is very efficient, but it is far from robust. For $\varepsilon=0.20$, it behaves even worse than COV! This coincides with the results of Devlin *et al.* [2], whose simulations pointed out a lack of robustness of CAU for $p=6$ under asymmetric contamination.

TABLE II

Monte Carlo Averages of Log Condition Numbers

p	ε	k	PE	MVE	TY	CAU	COV
2	0		0.71	1.33	0.62	0.45	0.43
	0.1	1	0.81	1.45	0.80	0.51	0.45
		2	0.97	1.97	0.80	0.64	0.54
		4	0.88	1.24	0.81	0.98	1.15
		8	0.80	1.22	0.80	1.10	2.12
		15	0.80	1.22	0.80	1.17	3.12
	0.2	1	0.95	1.45	1.34	0.56	0.46
		2	1.71	2.42	1.34	0.94	0.71
		3	1.59	2.89	1.34	1.30	1.13
		4	1.57	2.60	1.33	1.60	1.51
		6	1.39	1.20	1.34	2.02	2.11
		15	1.32	1.14	1.34	3.43	3.92
4	0		1.75	2.29	1.17	0.93	0.92
	0.1	2	2.10	2.60	1.70	1.38	1.38
		4	2.30	3.19	1.74	1.67	1.61
		8	2.13	2.19	1.74	2.65	2.68
		15	2.06	2.16	1.74	3.73	3.97
	0.2	4	3.15	3.59	4.30	2.11	1.97
		6	3.23	4.36	4.99	2.76	2.62
		8	3.28	4.05	5.50	3.27	3.13
		10	2.99	3.48	5.84	3.62	3.50
		20	2.35	2.15	7.12	5.00	4.89

- PE outperforms the other estimators for $p = 4$ and $\varepsilon = 0.20$. It is for situations with highly contaminated data in moderate dimensions that it ought to be comparatively most advantageous.

APPENDIX: PROOFS

Notation. We denote $(A^{-1})^T$ by A^{-T} and by $\lambda_j(A)$, the absolute eigenvalues of A in descending order.

Proof of Lemma 1.1. Let $Y = CX$. Because the set $\mathscr{A}$ of all A equals the set of all AC, $g(A(X)C^{-1}Y) = g(A(X)X) = \min_A g(AX) = \min_A g(ACX) = \min_A g(AY)$, and, hence, $A(Y) = A(X)C^{-1}$. The first result follows from (2). If X is spherically symmetric and Q orthogonal, X and QX have the same distribution, and therefore $V(X) = QV(X)Q^T$ for every orthogonal Q. Using the Q's which correspond to a change of sign in a single coordinate and an interchange of two coordinates, we obtain the second result.

Proof of Lemma 2.1. If A solves (P2), it has $g_3(AX) = 1$, since $g_3(AX) > 1$ would imply $g_2([A/g_3(AX)]X) < g_2(AX)$. Therefore, (P4) is equivalent to (P2). It is clear that if A solves (P4), then $A/g_2(AX)$ solves (P5) and $A/c(A)$ with $c(A) = \frac{1}{2}[g_2(A) + g_3(A)]$ solves (P6). If A solves (P1), then $c(A) = 1$. This is seen as follows: $c(A) > 1$ implies $g_1(AX) = g_2(AX) - 1$, $g_1([A/c(A)]X) = g_2(AX)/c(A) - 1 > g_2(AX) - 1 = g_1(AX)$, and A does not minimize g_1. A similar argument holds for $c(A) < 1$. Since $c(A) = 1$ implies $g_4(AX) = (1 + g_1(AX))/(1 - g_1(AX))$, (P6) and (P1) are equivalent.

Proof of Lemma 2.2. Noting that $\mathscr{V} = \{(A^TA)^{-1} \mid A \in \mathscr{A}\}$ the proof follows from

$$\max_{\|u\|=1} |s(u^TAX) - 1| = \max_{u \neq 0} \left| \frac{s(u^TAX)}{(u^Tu)^{1/2}} - 1 \right| = \max_{v \neq 0} \left| \frac{s(v^TX)}{v^TA^{-1}A^{-T}v} - 1 \right|. \quad (6)$$

Proof of Theorem 2.3. Let $c = g_2(X)$. Then $g_2(X/c) = 1$ and $g_3(X/c) = \tilde{c} > 0$. In view of Lemma 1.1 it suffices to show that problem (P3) has a solution for $Y = X/c$ instead of X. We first show that the maximization of g_3 can be restricted to a compact set of A's,

$$\mathscr{A}_0 = \{A \mid \tilde{c} \leqslant \lambda_p(A) \leqslant \cdots \leqslant \lambda_1(A) \leqslant \tilde{c}^{-1} \text{ and } g_2(AX) = 1\},$$

and then use the standard continuity argument.

First note that

$$\frac{s(v^TY)}{\|A^{-T}v\|} \geqslant \min_{\|\tilde{v}\|=1} s(\tilde{v}^TY) \cdot \frac{\|v\|}{\|A^{-T}v\|}$$

for all $v \neq 0$. Therefore the restriction in (P3) leads to

$$1 \geqslant g_2(AY) = \max_v \left(\frac{s(v^T Y)}{\|A^{-T} v\|} \right) \geqslant g_3(Y) \max_v \frac{\|v\|}{\|A^{-T} v\|}$$
$$= \tilde{c} \left(\min_{\|v\|=1} \|A^{-T} v\| \right)^{-1} = \tilde{c} \lambda_1(A).$$

Thus, $\lambda_1(A) \leqslant 1/\tilde{c}$. Similarly,

$$\frac{s(v^T Y)}{\|A^{-T} v\|} \leqslant \max_{\|\tilde{v}\|=1} s(\tilde{v}^T Y) \cdot \frac{\|v\|}{\|A^{-T} v\|} \qquad \text{for all} \quad v \neq 0,$$

$$g_3(AY) \leqslant g_2(Y) \min_v \frac{\|v\|}{\|A^{-T} v\|} = \lambda_p(A).$$

Thus, $\lambda_p(A) < \tilde{c}$ would lead to $g_3(AY) < g_3(Y)$, such that A could not solve (P3). Therefore, the eigenvalues of all solutions of (P3) are bounded away from 0 and ∞, and the maximization can be restricted to $\mathscr{A}_0$.

Second, we show that g_3 is continuous in A if A is non-singular. We have

$$g_3(AX) = \min_u \frac{s(u^T AX)}{\|u\|} = \min_v \frac{s(v^T X)}{\|A^{-T} v\|} \tag{7}$$

$$\geqslant \min_v \frac{s(v^T X)}{\|v\|} \Big/ \max_v \frac{\|A^{-T} v\|}{\|v\|} = g_3(X) \cdot \lambda_p(A). \tag{8}$$

Interchanging X and AX, this implies $g_3(AX) \leqslant g_3(X) \lambda_1(A)$. Choosing $A = BC^{-1}$, these two bounds for $\tilde{X} = CX$ prove $\lim_{B \to C} g_3(BX) = \lim_{A \to I} g_3(A\tilde{X}) = g_3(\tilde{X}) = g_3(CX)$ which proves the continuity of g_3, and similarly of g_2. The theorem follows since an argument A which maximizes a continuous function on a compact set $\mathscr{A}_0$ always exists.

Proof of Lemma 2.4. For the example mentioned after the lemma, we need to show that the two matrices I and $C = \mathrm{diag}(1, 1, c)$ do indeed solve (P3). A sketch of proof is as follows:

First, $s(u^T AX) \leqslant 1$ with $\|u\| = 1$ leads to $\|Ax\| \leqslant 1$ for all $x \in \mathscr{C}_1 \cup \mathscr{C}_2$. If $\|Ax\| < 1$ for such an x, there is an $\tilde{A} \geqslant A$ with $\|\tilde{A}x\| = 1$ for all such x. Since $g_3(\tilde{A}X) \geqslant g_3(AX)$, A can only solve (P3) if $\tilde{A}$ does, and therefore, there must be a solution with $\|Ax\| = 1$ for all $x = \mathscr{C}_1 \cup \mathscr{C}_2$. It is clear that $g_3(AX)$ is only a function of the angle φ between the planes through $A\mathscr{C}_1$ and $A\mathscr{C}_2$, and it can be shown that $\varphi = \pi/2$ maximizes $g_3(AX)$. Therefore, I is a solution. Finally, it is easy to check that $g_2(CX) \leqslant 1$ (since $C \leqslant I$) and that $g_3(CX) = g_3(X)$.

Proof of Theorem 2.5. We prove the statement for problem (P2). Note that

$$g_2(AX) = \sup_{\|v\|=1} \frac{s(v^T X)}{\|Bv\|},$$

where $B = A^{-T}$, and similarly for g_3.

We treat the case of $p = 2$ first. Write $v_\varphi = (\cos\varphi, \sin\varphi)^T$, and

$$B = Q_\psi \begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix} Q_\psi^T, \qquad Q_\psi = \begin{bmatrix} \cos\psi & -\sin\psi \\ \sin\psi & \cos\psi \end{bmatrix}$$

with $\lambda_1 \geq \lambda_2$, $\psi \in \Gamma = (-\pi/2, \pi/2]$. Then

$$\begin{aligned} \|Bv_\varphi\|^2 &= \lambda_1^2 \cos^2(\varphi - \psi) + \lambda_2^2 \sin^2(\varphi - \psi) \\ &= h^2(\varphi - \psi, \lambda_1, \lambda_2) \end{aligned}$$

and $s(v_\varphi^T X) = \tilde{s}(\varphi)$, say. In order to solve (P2), we need to find $\psi, \lambda_1, \lambda_2$ which minimize

$$\tilde{g}(\psi, \lambda_1, \lambda_2) = \sup_{\varphi \in \Gamma} \{\tilde{s}(\varphi)/h(\varphi - \psi, \lambda_1, \lambda_2)\}$$

subject to

$$\tilde{s}(\varphi)/h(\varphi - \psi, \lambda_1, \lambda_2) \geq 1.$$

The task is illustrated in Fig. 1. We want to find a pair of proportional functions of the form $h(\cdot - \psi, \lambda_1, \lambda_2)$ and $gh(\cdot - \psi, \lambda_1, \lambda_2)$ which bound the function $\tilde{s}$ from below and above, such that g is minimal. Note that $\tilde{s}(\varphi)$

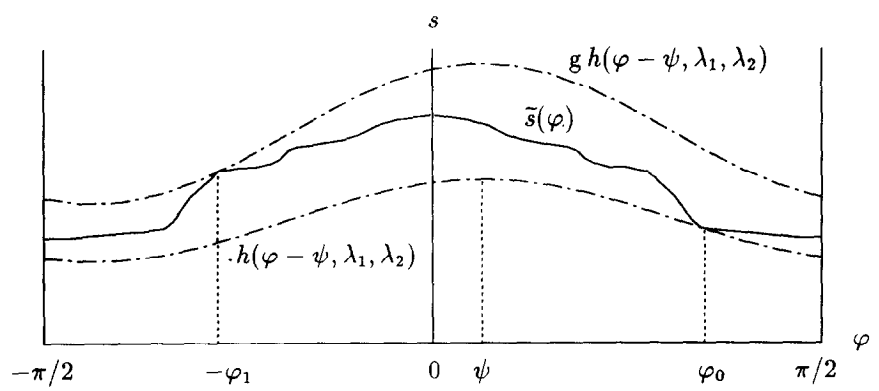

FIG. 1. Illustration of the optimization task for the proof of Theorem 2.5.

is symmetric and monotone non-increasing in $0 \leqslant \varphi \leqslant \pi/2$, since the projections of F and H on v_φ are constant and decreasing, respectively.

Let $s^+(\varphi) = \lim_{\alpha \uparrow \varphi} \tilde{s}(\alpha)$ and $s^-(\varphi) = \lim_{\alpha \downarrow \varphi} \tilde{s}(\alpha)$ be left- and right-continuous versions of $\tilde{s}$ for $0 \leqslant \varphi \leqslant \pi/2$ (which exist for monotone functions), and $s^+(-\varphi) = s^+(\varphi)$, $s^-(-\varphi) = s^-(\varphi)$, $s^+(0) = \tilde{s}(0)$, $s^-(\pi/2) = \tilde{s}(\pi/2)$.

Clearly, for any pair of functions which solve the minimization problem there is a φ_0 for which $h(\varphi_0 - \psi, \lambda_1, \lambda_2) = s^-(\varphi_0)$ and a φ_1 with $gh(\varphi_1 - \psi, \lambda_1, \lambda_2) = s^+(\varphi_1)$. Otherwise, g could be decreased by rescaling the lower or upper bounding function.

We now show that $\psi = 0$ if $\lambda_1 > \lambda_2$ for such pairs. Assume $0 < \psi \leqslant \pi/2$. We show that g can be decreased. For φ with $\psi < |\varphi| \leqslant \pi/2$,

$$h(\varphi, \lambda_1, \lambda_2) = h(|\varphi|, \lambda_1, \lambda_2) < h(|\varphi| - \psi, \lambda_1, \lambda_2) \leqslant s^-(\varphi),$$

and for $|\varphi| \leqslant \psi$,

$$h(|\varphi|, \lambda_1, \lambda_2) \leqslant \lambda_1 = h(\psi - \psi, \lambda_1, \lambda_2) \leqslant s^-(\psi) \leqslant s^-(\varphi).$$

If for all φ the corresponding inequality were strict, we could improve g as before. Equality can only hold for $\varphi = 0$, and only if $\tilde{s}(\varphi)$ is constant for $0 \leqslant \varphi \leqslant \psi$ and $h(0, \lambda_1, \lambda_2) = \lambda_1 = \tilde{s}(0)$. In this case, we still have $h(\varphi, \lambda_1, \lambda_2) < s^-(\varphi)$ for all $\varphi \neq 0$. Then, λ_2 can be increased: There is a $\bar{\lambda}_2$ with $\lambda_2 < \bar{\lambda}_2 < \lambda_1$ and $h(\varphi, \lambda_1, \bar{\lambda}_2) \leqslant s^-(\varphi)$ for all φ. Then $gh(\varphi, \lambda_1, \bar{\lambda}_2) > gh(\varphi, \lambda_1, \lambda_2) \geqslant \tilde{s}(\varphi)$ for $\varphi \neq 0$, and $gh(0, \lambda_1, \bar{\lambda}_2) > gh(\psi, \lambda_1, \bar{\lambda}_2) > gh(\psi, \lambda_1, \lambda_2) \geqslant \tilde{s}(\psi) = \tilde{s}(0)$. Therefore g can be decreased.

If $\lambda_1 = \lambda_2$, ψ is arbitrary. Thus, we can assume $\psi = 0$ in all cases.

To complete the proof for $p = 2$, we need to show that the pair (λ_1, λ_2) which solves the minimization problem is unique. Let

$$\Phi^- = \{0 \leqslant \varphi \leqslant \pi/2 \mid h(\varphi, \lambda_1, \lambda_2) = s^-(\varphi)\},$$

$$\Phi^+ = \{0 \leqslant \varphi \leqslant \pi/2 \mid gh(\varphi, \lambda_1, \lambda_2) = s^+(\varphi)\},$$

$$\varphi_0 = \min \Phi^-, \qquad \varphi_1 = \min \Phi^+.$$

We first show an auxiliary statement about φ_0 and φ_1: If $\varphi_0 \leqslant \varphi_1$, then $\varphi_1 < \pi/2$ and there is a $\varphi_2 > \varphi_1$ in Φ^-. In order to see this, assume the contrary, either $\varphi_1 = \pi/2$ or

$$h(\varphi, \lambda_1, \lambda_2) < s^-(\varphi) \qquad \text{for all} \quad \varphi > \varphi_1.$$

Let $\tilde{\varphi}_0 = \max(\Phi^- \cap \{\varphi \leqslant \varphi_1\})$. If $\tilde{\varphi}_0 = \varphi_1$, then $g = 1$, and uniqueness is obvious. Otherwise, the lower bound $h(\cdot, \lambda_1, \lambda_2)$ is not strict for $\varphi \geqslant \tilde{\varphi}_0$, and the upper bound $gh(\cdot, \lambda_1, \lambda_2)$ is not strict for $\varphi \leqslant \varphi_1$ with $\tilde{\varphi}_0 < \varphi_1$. Then it is easy to see in Fig. 1 and not difficult to prove that there is a pair

$(\tilde{\lambda}_1, \tilde{\lambda}_2)$ and a $\tilde{g} < g$ with the required properties. This proves the auxiliary statement.

Assume now that there were two different solutions $(\lambda_1, \lambda_2, g)$ and $(\tilde{\lambda}_1, \tilde{\lambda}_2, \tilde{g})$, $g = \tilde{g}$. Since the lower bounds both touch s^-, the lower curves $h(\cdot, \lambda_1, \lambda_2)$ and $h(\cdot, \tilde{\lambda}_1, \tilde{\lambda}_2)$ must have a point in common in $[0, \pi/2]$, and because of the nature of h, this is a unique point with $\varphi = \varphi_*$, say. Without loss of generality, assume $\lambda_1 > \tilde{\lambda}_1$. Then

$$h(\varphi, \lambda_1, \lambda_2) < h(\varphi, \tilde{\lambda}_1, \tilde{\lambda}_2) \leqslant s^-(\varphi) \qquad \text{for all} \quad \varphi > \varphi_*,$$

and therefore $\varphi_0 \leqslant \varphi_*$. Similarly, $\varphi_1 \geqslant \varphi_*$. Then, there is, by the auxiliary statement, a $\varphi_2 > \varphi_1 \geqslant \varphi_*$ with $s^-(\varphi_2) = h(\varphi_2, \lambda_1, \lambda_2)$, and this is $< h(\varphi_2, \tilde{\lambda}_1, \tilde{\lambda}_2) \leqslant s^-(\varphi_2)$, because of the last equation. The combined inequality is a contradiction. This proves uniqueness for $p = 2$.

Finally, let $p > 2$. The projection of F onto any two-dimensional subspace which includes the $x^{(1)}$-axis is the same. Let $(\lambda_1, \lambda_2, g_0)$ be the solution of the minimization problem for this subspace, and $B = \operatorname{diag}(\lambda_1, \lambda_2, ..., \lambda_2)$, $V = B^2$. For any $\tilde{V} = \tilde{B}^T \tilde{B} \neq V$ there is a v_* with $\|Bv_*\| \neq \|\tilde{B}v_*\|$. If $\tilde{V}$ satisfies the side condition in (P2), then $s(v^T X)/\|\tilde{B}v\| \geqslant 1$ for all v. Focussing on the v's in the subspace $\mathscr{S}$ spanned by $(1, 0, ..., 0)^T$ and v_*, we obtain, by the uniqueness in the case $p = 2$, that $\max_v \{s(v^T X)/\|Bv\|\} = g_0 < \max_{v \in \mathscr{S}} (s(v^T X)/\|\tilde{B}v\|)$, and therefore $\tilde{V}$ does not solve (P2).

Proof of Lemma 2.6. The statement follows immediately from the definitions.

Proof of Lemma 2.7. The first statement follows from the preceding theorem. An easy proof procedes by noting that $V = vI$ with $v = s(X^{(1)})$ is a solution of (P4), yielding the minimum 1, and showing that it is unique: Let $V \in \mathscr{V}$ have eigenvalues $\lambda_1^2 \geqslant \cdots \geqslant \lambda_p^2$ with respective eigenvectors $u_1, ..., u_p$. Then $A = \sum_{i=1}^p \lambda_i^{-1} u_i u_i^T$ is a square root of V^{-1}. The sphericity of X implies that $\lambda_1 u_1^T AX = u_1^T X$ has the same distribution as $u_p^T X = \lambda_p u_p^T AX$, and hence $s(u_1^T AX) = s(u_p^T AX)\lambda_p/\lambda_1$. If V is not a multiple of I, then $\lambda_1 > \lambda_p$, and hence $g_4(A) \geqslant \lambda_1/\lambda_p > 1$, which implies that A does not minimize g_4.

The second statement of the lemma follows from the first one and Lemma 1.1.

Proof of Theorem 3.1. We prove the result for problem (P2). Let $\varepsilon < \delta^*(s, F)$, $X \sim (1-\varepsilon)F + \varepsilon H$, and $V = V(X)$. Let A be the symmetric positive definite square root of V^{-1}, with eigenvalues $\lambda_1 \geqslant \cdots \geqslant \lambda_p$ and corresponding eigenvectors $u_1, ..., u_p$. The restriction on A implies that

$$1 \leqslant g_3(AX) \leqslant s(u_p^T AX) = \lambda_p s(u_p^T X) \leqslant \lambda_p g_2(X), \tag{9}$$

and, hence, $\lambda_p \geqslant 1/g_2(X) \geqslant 1/s_2 > 0$ for all H. On the other hand, note that the matrix $I/g_3(X)$ fulfills the restriction $g_3([I/g_3(X)]X) \geqslant 1$. Since A minimizes g_2 under this restriction, it follows that $g_2(AX) \leqslant g_2(X/g_3(X)) = g_2(X)/g_3(X)$. Since

$$g_2(AX) \geqslant s(u_1^T AX) = \lambda_1 s(u_1^T X) \geqslant \lambda_1 g_3(X),$$

it follows that

$$\lambda_1 \leqslant g_2(X)/[g_3(X)]^2 \leqslant s_2/s_1^2 \tag{10}$$

for all H. This shows that the eigenvalues of A and, therefore, those of V are bounded away from 0 and ∞.

Proof of Theorem 3.2. In view of Lemma 1.1 and Lemma 2.7 we assume that F is spherically symmetric. Combining (9) and (10) we obtain

$$\gamma(A) \leqslant (g_2(X)/g_3(X))^2.$$

We shall calculate a bound for the right-hand side. For any u with $\|u\| = 1$, $|u^T X|$ has distribution function $\tilde{F}_1$ under $X \sim F$. Hence under F_ε, the quantile $s(u^T X)$ satisfies $(1-\varepsilon)F_1(s) \leqslant \alpha \leqslant (1-\varepsilon)F_1(s) + \varepsilon$, and hence $\sqrt{b_0} \leqslant s(u^T X) \leqslant \sqrt{b_1}$. This implies $\sqrt{b_0} \leqslant g_3(X)$ and $g_2(X) \leqslant \sqrt{b_1}$, which yields the desired result since $\gamma(V) = \gamma(A)^2$.

Proof of Theorem 3.3. Since F is elliptical, we can assume F is spherically symmetric, $x_0 = \sqrt{c}\,(1, 0, ..., 0)^T$, and in view of Theorem 2.5, $V(X) = \operatorname{diag}(v_1, v_2, ..., v_2)$. Using the notation of the proof of Theorem 2.5, we are left with calculating the pair (λ_1, λ_2) which solves the minimization problem mentioned there for the special case, where s is the α-quantile and H is a point mass, since $\gamma(V) = (\lambda_1/\lambda_2)^2$. In this case, $\tilde{s}(\varphi) = \min(\max(\sqrt{c}\cos\varphi, \sqrt{b_0}), \sqrt{b_1})$.

If $c \leqslant b_0$, then $\tilde{s}(\varphi) = \sqrt{b_0}$ for all φ, and $V = I$, $\gamma(V) = 1$. Otherwise, $\tilde{s}$ has "corners" at

$$\eta_0 = \arccos\sqrt{b_0/c}$$

$$\eta_1 = \arccos(\min(1, \sqrt{b_1/c})).$$

Examining Fig. 1, modified to show this simple function $\tilde{s}$, and the argument given in the proof of Theorem 2.5, it is easy to see that for the solution $(\lambda_1, \lambda_2, g)$, the bounds $h(\varphi, \lambda_1, \lambda_2)$ and $gh(\varphi, \lambda_1, \lambda_2)$ must touch $\tilde{s}(\varphi)$ in three points

$$h(\eta_0, \lambda_1, \lambda_2) = \tilde{s}(\eta_0), \tag{11}$$

$$gh(\eta_1, \lambda_1, \lambda_2) = \tilde{s}(\eta_1) \tag{12}$$

and either

$$h(0, \lambda_1, \lambda_2) = \tilde{s}(0) = \sqrt{b_1} = \lambda_1 \tag{13}$$

or

$$gh(\pi/2, \lambda_1, \lambda_2) = \tilde{s}(\pi/2), \tag{14}$$

whichever gives the lower value of g.

If $b_0 < c \leqslant b_1$, then $\eta_1 = 0$ and $gh(0, \lambda_1, \lambda_2) = g\lambda_1 = \sqrt{c} = \tilde{s}(0)$. It can be checked by calculating both cases that (14) gives the lower value of g. Therefore $g\lambda_2 = \sqrt{b_0}$ and $\gamma(V) = c/b_0$. Finally, for $c > b_1$, we have, by (11) and (12),

$$\lambda_1^2 b_0/c + \lambda_2^2(1 - b_0/c) = b_0$$

$$g^2\lambda_1^2 b_1/c + g^2\lambda_2^2(1 - b_1/c) = b_1.$$

Then, (13) leads to $\lambda_1^2 = b_1$, $\lambda_2^2 = (b_0 - b_1 b_0/c)/(1 - b_0/c)$, and $g^2 = b_1/[b_1^2/c + b_0(1 - b_1/c)^2/(1 - b_0/c)] = b_1(1 - b_0/c)/(b_1^2/c - 2b_0 b_1/c + b_0)$, whereas (14) implies $g^2\lambda_2^2 = b_0$, $g^2\lambda_1^2 = [b_1 - b_0(1 - b_1/c)]/(b_1/c)$, $g^2 = 2 - b_0/b_1$. It is straightforward to check that the former g is larger than the latter for $c > 2b_1$ and vice versa, and that the ratios λ_1^2/λ_2^2 coincide with the formulas given in the theorem.

REFERENCES

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

[2] DEVLIN, S. J., GNANADESIKAN, R., AND KETTENRING, J. R. (1981). Robust estimation of dispersion matrices and principal components. *J. Amer. Statist. Assoc.* **76** 354–362.

[3] DONOHO, D. L. (1982). *Breakdown Properties of Multivariate Location Estimators.* Ph.D. qualifying paper, Department of Statistics, Harvard University, Cambridge, MA.

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

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

[6] LI, G., AND CHEN, Z. (1985). Projection-pursuit approach to robust dispersion matrices and principal components: Primary theory and Monte Carlo. *J. Amer. Statist. Assoc.* **80** 759–766.

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

[8] MARAZZI, A. (1988). Algorithms for the computation of weights in bounded influence regression. *Comput. Statist. Data Anal.* **6** 251–276.

[9] MARONNA, R. A. (1976). Robust M-estimators of location and scatter. *Ann. Statist.* **4** 51–67.

[10] MARONNA, R. A., AND YOHAI, V. J. (1990). A new class of bias-robust estimates of regression. Manuscript.

[11] ROUSSEEUW, P. J. (1985). Multivariate estimation with high breakdown point. In *Mathematical Statistics and its Applications* (W. Grossmann, G. Pflug, I. Vincze, and W. Wertz, Eds.), pp. 237–297. Reidel, Dordrecht.

[12] ROUSSEEUW, P. J., AND LEROY, A. M. (1987). *Robust Regression & Outlier Detection.* Wiley, New York.

[13] STAHEL, W. A. (1981). *Breakdown of Covariance Estimators.* Res. Rep. 31, Fachgruppe für Statistik ETH, Zurich.

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

[15] TYLER, D. E. (1988). Some results on the existence and computation of the M-estimates of multivariate location and scatter. *SIAM J. Sci. Statist. Comput.* **9** 354–362.

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