The Annals of Statistics
1993, Vol. 21, No. 3, 1385–1400

ASYMPTOTICS FOR THE MINIMUM COVARIANCE DETERMINANT ESTIMATOR

BY R. W. BUTLER,[1] P. L. DAVIES AND M. JHUN

Colorado State University, Universität Gesamthochschule Essen and Korea University

Consistency is shown for the minimum covariance determinant (MCD) estimators of multivariate location and scale and asymptotic normality is shown for the former. The proofs are made possible by showing a separating ellipsoid property for the MCD subset of observations. An analogous property is shown for the MCD subset computed from the population distribution.

1. Introduction. Let $(X_i)_1^\infty$ be independently and identically distributed random variables in $\mathbb{R}^k$ and S_n be an arbitrary subset of $\{1, \dots, n\}$ of size $s_n = [n\gamma]$, $0 < \gamma < 1$. We denote the empirical mean and matrix covariance based on this subset by $\hat{\mathfrak{M}}(S_n)$ and $\hat{\mathfrak{S}}(S_n)$, respectively,

$$\hat{\mathfrak{M}}(S_n) = \frac{1}{s_n} \sum_{i \in S_n} X_i, \qquad \hat{\mathfrak{S}}(S_n) = \frac{1}{s_n} \sum_{i \in S_n} \left(X_i - \hat{\mathfrak{M}}(S_n)\right)\left(X_i - \hat{\mathfrak{M}}(S_n)\right)^T.$$

Consider the subset $\hat{S}_n$ of $\{1, \dots, n\}$ for which the determinant of $\hat{\mathfrak{S}}(S_n)$, $|\hat{\mathfrak{S}}(S_n)|$, attains its minimum value over all subsets S_n of $\{1, \dots, n\}$ of size s_n. In the univariate case, $k = 1$, the mean $\hat{\mathfrak{M}}(\hat{S}_n)$ and the variance $\hat{\mathfrak{S}}(\hat{S}_n)$ correspond to the least trimmed squares (LTS) estimator. The asymptotics of this case have been given by Butler (1982) and Rousseeuw (1983); see also Rousseeuw and Leroy (1987). The general case corresponds to the minimum covariance determinant (MCD) estimate of Rousseeuw (1983). We will show that there exists a separating ellipsoid containing all data points X_j with $j \in \hat{S}_n$ and excluding all those with $j \notin \hat{S}_n$, that the center of the ellipsoid is $\hat{\mathfrak{M}}(\hat{S}_n) + O_p(1/n)$ and that the shape is given by a matrix of the form $\hat{\mathfrak{S}}(\hat{S}_n) + O_p(1/n)$.

In Section 2 we define the corresponding problem for the common distribution $\mathbb{P}$ of the X_j's. We show that if $\mathbb{P}$ is a unimodal elliptical distribution with a density, then the solution is an ellipsoid whose center is the point of symmetry of $\mathbb{P}$ and whose shape is that of $\mathbb{P}$. It will turn out that the empirical counterparts converge at the rate of $n^{-1/2}$ to the population values. The MCD estimator is affinely equivariant. It is a further example of estimators with a high breakdown point and $n^{-1/2}$ rate of convergence. Other examples are the S-estimator of Rousseeuw and Yohai (1984) and Davies (1987) and k-step

Received December 1987; revised July 1992.
[1]Research supported by NSF Grant DMS-89-96150.
AMS 1991 *subject classifications*. Primary 62H05; secondary 62G35, 62G15.
Key words and phrases. Minimum covariance determinant estimator, tolerance region.

M-estimators based on the minimum volume ellipsoid [Rousseeuw (1983) treated by Davies (1992a)]. The MVE estimator itself has an $n^{-1/3}$ rate of convergence [Davies (1992b) and Nolan (1991)].

The separating ellipsoids for the empirical distribution provide a natural methodology for constructing robust multivariate tolerance regions for the next X_{n+1} observation. The ellipsoid can be blown up to give and approximate 95% guarantee of 90% coverage for X_{n+1} thus extending the univariate methodology of Butler (1982) and Cho and Miller (1987). Example 2 in Section 4 has further details.

2. The MCD problem for the theoretical distribution. We assume that the common distribution $\mathbb{P}$ of the random variables $(X_j)_1^\infty$ has a density of the form

$$\frac{1}{|\Sigma|} f\big((x-\mu)^T \Sigma^{-1}(x-\mu)\big),$$

where $\mu \in \mathbb{R}^k$ and Σ is a symmetric positive definite $k \times k$ matrix. As the estimators we consider are affinely equivariant, we may assume without loss of generality that $\mu = 0$ and $\Sigma = I_k$. The function $f\colon \mathbb{R}_+ \to \mathbb{R}_+$ is assumed to be nonincreasing so that $\mathbb{P}$ is a unimodal distribution.

For any bounded Borel set B in $\mathbb{R}^k$ with $\mathbb{P}(B) = \gamma$, we define

$$\mathfrak{M}(B) = \frac{1}{\gamma} \int_B x \, d\mathbb{P}(x)$$

and

$$\mathfrak{S}(B) = \frac{1}{\gamma} \int_B (x - \mathfrak{M}(B))(x - \mathfrak{M}(B))^T \, d\mathbb{P}(x).$$

The restriction to bounded sets is simply to ensure the existence of the relevant moments. It could be weakened to those sets with well-defined second moments but the gain in generality is only slight. We denote the family of all such sets by $\mathfrak{C}(\gamma)$.

For $B \in \mathfrak{C}(\gamma)$ we define an ellipsoid $\mathscr{E}(B)$ by

$$\mathscr{E}(B) = \big\{x\colon (x - \mathfrak{M}(B))^T \mathfrak{S}(B)^{-1}(x - \mathfrak{M}(B)) \le r^2(B)\big\},$$

where $r^2(B)$ is chosen so that $\mathbb{P}(\mathscr{E}(B)) = \gamma$.

We write

$$D = \inf_{B \in \mathfrak{C}(\gamma)} |\mathfrak{S}(B)|. \tag{2.1}$$

Let $K(x, r)$ denote the ball with center x and radius r. On choosing r so that $K(0, r) \in \mathfrak{C}(\gamma)$, we see that

$$D \le |\mathfrak{S}(K(0, r))| < \infty. \tag{2.2}$$

The ball $K(0, r(\gamma))$ with $\mathbb{P}(K(0, r(\gamma))) = \gamma$ will turn out to be the solution of the MCD problem for the population distribution $\mathbb{P}$. The value of $r(\gamma)$ is determined by

$$\frac{2\pi^{k/2}}{\Gamma(k/2)}\int_0^{r(\gamma)} r^{k-1} f(r^2)\,dr = \gamma. \tag{2.3}$$

We require the following result.

LEMMA 1. *For all $\eta > 0$ and $\theta, y \in \mathbb{R}^k$, we have*

$$\mathbb{P}(\{x: \|x\| \le \eta\}) \ge \mathbb{P}(\{x: \|x - y\| \le \eta\}) \tag{i}$$

and

$$\mathbb{P}(\{x: |\theta^T x| \le \eta\}) \ge \mathbb{P}(\{x: |\theta^T(x - y)| \le \eta\}). \tag{ii}$$

PROOF. We have

$$\int (\{\|x - y\| \le \eta\} - \{\|x\| \le \eta\})\big(f(\|x - y\|^2) - f(\|x\|^2)\big)\,dx \ge 0$$

as the expressions in the products are either both positive or both negative. This may be seen as follows. If $\|x - y\| \le \|x\|$, then $\{\|x - y\| \le \eta\} - \{\|x\| \le \eta\} \ge 0$ and $f(\|x - y\|^2) - f(\|x\|^2) \ge 0$ showing

$$(\{\|x - y\| \le \eta\} - \{\|x\| \le \eta\})\big(f(\|x - y\|^2) - f(\|x\|^2)\big) \ge 0.$$

Similarly, if $\|x\| \le \|x - y\|$, then both terms in the product are nonpositive.

On multiplying out we obtain

$$2\mathbb{P}(\{\|x\| \le \eta\}) \ge \mathbb{P}(\|x - y\| \le \eta) + \mathbb{P}(\|x + y\| \le \eta) = 2\mathbb{P}(\|x - y\| \le \eta),$$

proving (i).

To prove (ii), we note that if $\theta \ne 0$ the distribution of $\theta^T X$ is unimodal and the result follows from (i) for the special case $k = 1$. □

The smallest and largest eigenvalues of a matrix Σ will be denoted by $\lambda_{\min}(\Sigma)$ and $\lambda_{\max}(\Sigma)$, respectively.

LEMMA 2. *Let $(B_n)_1^\infty$ be a sequence of sets in $\mathfrak{C}(\gamma)$ with*

$$\lim_{n\to\infty} |\mathfrak{S}(B_n)| = D.$$

Then

$$\liminf \lambda_{\min}(\mathfrak{S}(B_n)) > 0, \tag{i}$$

$$\limsup \lambda_{\max}(\mathfrak{S}(B_n)) < \infty \tag{ii}$$

and

$$\limsup \|\mathfrak{M}(B_n)\| < \infty. \tag{iii}$$

PROOF. Let θ_n, $\|\theta_n\| = 1$, denote the eigenvector associated with $\lambda_{\min}(\mathfrak{S}(B_n))$. We have

$$\begin{aligned}\lambda_{\min}(\mathfrak{S}(B_n)) &= \frac{1}{\gamma}\int_{B_n}\left(\theta_n^T(x - \mathfrak{M}(B_n))\right)^2 d\mathbb{P}(x)\\ &\geq \frac{\eta^2}{\gamma}\mathbb{P}\left(B_n \cap \left\{\left|\theta_n^T(x - \mathfrak{M}(B_n))\right| \geq \eta\right\}\right)\\ &\geq \frac{\eta^2}{\gamma}\left(\mathbb{P}(B_n) - \mathbb{P}\left(\left\{\left|\theta_n^T(x - \mathfrak{M}(B_n))\right| \leq \eta\right\}\right)\right)\\ &\geq \frac{\eta^2}{\gamma}\left(\mathbb{P}(B_n) - \mathbb{P}\left(\left\{\left|\theta_n^T x\right| \leq \eta\right\}\right)\right)\end{aligned}$$

by (ii) of Lemma 1. We may choose η so small that $\mathbb{P}(\{|\theta_n^T x| \leq \eta\}) \leq \gamma/2$. This gives

$$\lambda_{\min}(\mathfrak{S}(B_n)) \geq \tfrac{1}{2}\eta^2,$$

proving (i).

(ii) follows from (i) and (2.2).

To prove (iii), we set $\theta_n = \mathfrak{M}(B_n)/\|\mathfrak{M}(B_n)\|$. Then

$$\begin{aligned}\lambda_{\max}(\mathfrak{S}(B_n)) &\geq \frac{1}{\gamma}\int_{B_n}\left(\theta_n^T(x - \mathfrak{M}(B_n))\right)^2 d\mathbb{P}\\ &= \frac{1}{\gamma}\int_{B_n}\left(\|\mathfrak{M}(B_n)\| - \theta_n^T x\right)^2 d\mathbb{P}.\end{aligned}$$

Let r be such that $\mathbb{P}(K(0, r)) = 1 - \gamma/2$ implying $\mathbb{P}(B_n \cap K(0, r)) \geq \gamma/2$. We have

$$\begin{aligned}\lambda_{\max}(\mathfrak{S}(B_n)) &\geq \frac{1}{\gamma}\int_{B_n \cap K(0,r)}\left(\|\mathfrak{M}(B_n)\| - \theta_n^T x\right)^2 d\mathbb{P}\\ &\geq \left(\|\mathfrak{M}(B_n)\|^2 - 2\|\mathfrak{M}(B_n)\| r\right)\mathbb{P}(B_n \cap K(0, r))/\gamma\\ &\geq \tfrac{1}{2}\left(\|\mathfrak{M}(B_n)\|^2 - 2\|\mathfrak{M}(B_n)\| r\right).\end{aligned}$$

This together with (ii) proves (iii). □

LEMMA 3. *Let $E \in \mathfrak{E}(\gamma)$ be an ellipsoid with $\mathscr{E}(E) = E$. Then $E = K(0, r(\gamma))$, where $r(\gamma)$ is given by* (2.3).

PROOF. Let $E = \{x\colon (x - m)^T\Gamma^{-1}(x - m) \leq 1\}$. Then

$$m = \frac{1}{\gamma}\int_E y f\left(\|y\|^2\right) dy,$$

which gives

$$\int_{E'} y\, f(\|y + m\|^2)\, dy = 0,$$

with $E' = \{x\colon x^T\Gamma^{-1}x \le 1\}$. In particular,

$$\int_{E'} m^T y\, f(\|y + m\|^2)\, dy = 0$$

and hence

$$\begin{aligned}\int_{E'} m^T y\{m^T y \ge 0\} f(\|y + m\|^2)\, dy &= \int_{E'} - m^T y\{m^T y \le 0\} f(\|y + m\|^2)\, dy \\ &= \int_{E'} m^T y\{m^T y \ge 0\} f(\|y - m\|^2)\, dy.\end{aligned}$$

We obtain

$$\int_{E'} m^T y\{m^T y \ge 0\}\big(f(\|y + m\|^2) - f(\|y - m\|^2)\big)\, dy = 0.$$

If $m^T y \ge 0$, then $\|y + m\|^2 \ge \|y - m\|^2$ so that the integrand is always non-positive and strictly negative for some y unless $m = 0$. This proves $m = 0$.

We have shown $E = \{x\colon x^T\Gamma^{-1}x \le 1\}$. By an orthogonal transformation if necessary, we may assume that Γ is a diagonal matrix Λ with diagonal elements $\Lambda_1, \dots, \Lambda_k$. We have

$$\Lambda = \lambda \int_{\{x^T\Lambda^{-1}x \le 1\}} x x^T f(\|x\|^2)\, dx,$$

for some $\lambda > 0$. On writing $y = \Lambda^{-1/2}x$, it is sufficient to show that all solutions of

$$I_k = \lambda' \int_{\{\|x\| \le 1\}} x x^T f\left(\sum_1^k \Lambda_i x_i^2 \right) dx,$$

for some $\lambda' > 0$ satisfy $\Lambda_1 = \cdots = \Lambda_k$.

We have

$$\int_{\{\|x\| \le 1\}} x_1^2 f\left(\sum_1^k \Lambda_i x_i^2 \right) dx = \int_{\{\|x\| \le 1\}} x_2^2 f\left(\sum_1^k \Lambda_i x_i^2 \right) dx \tag{2.4}$$

and hence

$$\begin{aligned}\int_{\{\|x\| \le 1\}} (x_1^2 - x_2^2)\Bigg(f\left(\Lambda_1 x_1^2 + \Lambda_2 x_2^2 + \sum_3^k \Lambda_j x_j^2 \right) \\ - f\left(\Lambda_2 x_1^2 + \Lambda_1 x_2^2 + \sum_3^k \Lambda_j x_j^2 \right)\Bigg) dx = 0\end{aligned} \tag{2.5}$$

as may be seen by interchanging the roles of x_1 and x_2. Suppose $\Lambda_1 > \Lambda_2$. Then if $x_1^2 > x_2^2$ it follows that $\Lambda_1 x_1^2 + \Lambda_2 x_2^2 > \Lambda_2 x_1^2 + \Lambda_1 x_2^2$. Similarly, if

$x_1^2 < x_2^2$, then $\Lambda_1 x_1^2 + \Lambda_2 x_2^2 < \Lambda_2 x_1^2 + \Lambda_1 x_2^2$. Thus if $\Lambda_1 > \Lambda_2$ the integral in (2.5) is always nonpositive and strictly negative at some x_1, x_2. This contradicts (2.4) showing $\Lambda_1 = \Lambda_2$ and in general $\Lambda_1 = \cdots = \Lambda_k = \Lambda$. □

We are now in a position to prove the main theorem of this section.

THEOREM 1. *The ball* $K(0, r(\gamma))$ *solves the MCD problem, that is,*

$$\mathbb{P}(K(0,r(\gamma))) = \gamma \quad \textit{and} \quad |\mathfrak{S}(K(0,r(\gamma)))| = \inf_{B \in \mathfrak{C}(\gamma)} |\mathfrak{S}(B)|.$$

Furthermore, $K(0, r(\gamma))$ *is essentially unique in that for any* $B \in \mathfrak{C}(\gamma)$ *with* $\mathbb{P}(K(0, r(\gamma)) \triangle B) > 0$ *we have*

$$|\mathfrak{S}(B)| > |\mathfrak{S}(K(0,r(\gamma)))|.$$

PROOF. Let $(B_n)_1^\infty$ be a sequence of sets in $\mathfrak{C}(\gamma)$ with

$$\lim_{n\to\infty} |\mathfrak{S}(B_n)| = D = \inf_{B \in \mathfrak{C}(\gamma)} |\mathfrak{S}(B)|.$$

It follows from Lemma 2 that the sequence of ellipsoids $(\mathscr{E}(B_n))_1^\infty$ lies in a compact subset of $\mathbb{R}^{k+(1/2)k(k+1)}$. We choose a subsequence which we continue to denote by $(B_n)_1^\infty$ such that $\lim_{n\to\infty} \mathscr{E}(B_n) = E$. As $\mathscr{E}(B_n) \in \mathfrak{C}(\gamma)$ for all n, it follows that $E \in \mathfrak{C}(\gamma)$.

Suppose that $\liminf_{n\to\infty} \mathbb{P}(B_n \triangle E) = \eta > 0$ and let n_0 be such that $\mathbb{P}(B_n \triangle E) \geq \frac{1}{2}\eta$ for all $n \geq n_0$. As $\mathbb{P}(B_n) = \mathbb{P}(E) = \gamma$, we have $\mathbb{P}(B_n \setminus E) = \mathbb{P}(E \setminus B_n) \geq \frac{1}{4}\eta$ for all $n \geq n_0$.

Let ε, $0 < \varepsilon < \frac{1}{8}\eta$, be given. We choose points $x_n \in B_n \setminus E$ and $x_n' \in E \setminus B_n$ and balls $K_n(x_n, r_n)$ and $K_n'(x_n', r_n')$ such that $\mathbb{P}(K_n \cap (B_n \setminus E)) = \mathbb{P}(K_n' \cap (E \setminus B_n)) = \varepsilon$. As $\mathbb{P}(B_n \setminus E) \geq \frac{1}{4}\eta$, we may choose the sequence $(x_n)_{n_0}^\infty$ in such a manner that it is bounded away from the ellipsoid E. It is clear that both r_n and r_n' are of order $o(1)$ as ε tends to 0. We write $N_n = K_n \cap (B_n \setminus E)$, $N_n' = K_n' \cap (E \setminus B_n)$ and $B_n' = (B_n \setminus N_n) + N_n'$. By construction $\mathbb{P}(B_n') = \gamma$ so that $B_n' \in \mathfrak{C}(\gamma)$.

A straightforward calculation gives

$$\begin{aligned}\mathfrak{S}(B_n') = \mathfrak{S}(B_n) &- (\mu_n' - \mu_n)(\mu_n' - \mu_n)^T \\ &- \frac{1}{\gamma}\int_{N_n} (x - \mathfrak{M}(B_n))(x - \mathfrak{M}(B_n))^T \, d\mathbb{P} \\ &+ \frac{1}{\gamma}\int_{N_n'} (x - \mathfrak{M}(B_n))(x - \mathfrak{M}(B_n))^T \, d\mathbb{P},\end{aligned}$$

where

$$\mu_n' = \frac{1}{\gamma}\int_{N_n'} x \, d\mathbb{P}(x) \quad \text{and} \quad \mu_n = \frac{1}{\gamma}\int_{N_n} x \, d\mathbb{P}(x).$$

This implies

$$\mathfrak{S}(B'_n) \le \mathfrak{S}(B_n) - \frac{\varepsilon}{\gamma}(x_n - \mathfrak{M}(B_n))(x_n - \mathfrak{M}(B_n))^T(1 + o(1))$$
$$+ \frac{\varepsilon}{\gamma}(x'_n - \mathfrak{M}(B_n))(x'_n - \mathfrak{M}(B_n))^T(1 + o(1)),$$

where $A \le B$ means that $B - A$ is nonnegative definite. We may deduce

$$|\mathfrak{S}(B'_n)| \le \left|\mathfrak{S}(B_n) - \frac{\varepsilon}{\gamma}(x_n - \mathfrak{M}(B_n))(x_n - \mathfrak{M}(B_n))^T \right.$$
$$\left. + \frac{\varepsilon}{\gamma}(x'_n - \mathfrak{M}(B_n))(x'_n - \mathfrak{M}(B_n))^T\right|(1 + o(1))$$
$$= |\mathfrak{S}(B_n)|\left(1 - \frac{\varepsilon}{\gamma}(x_n - \mathfrak{M}(B_n))^T \mathfrak{S}(B_n)^{-1}(x_n - \mathfrak{M}(B_n))\right.$$
$$\left. + \frac{\varepsilon}{\gamma}(x'_n - \mathfrak{M}(B_n))^T \mathfrak{S}(B_n)^{-1}(x'_n - \mathfrak{M}(B_n))\right)$$
$$\times(1 + o(1)).$$

By construction $x'_n \in E$ and x_n is bounded away from E so that

$$(x'_n - \mathfrak{M}(B_n))^T \mathfrak{S}(B_n)^{-1}(x'_n - \mathfrak{M}(B_n))$$
$$< (x_n - \mathfrak{M}(B_n))^T \mathfrak{S}(B_n)^{-1}(x_n - \mathfrak{M}(B_n)) - \delta,$$

for all n sufficiently large with $\delta > 0$. It follows that

$$|\mathfrak{S}(B'_n)| < |\mathfrak{S}(B_n)|(1 - \eta),$$

for all n sufficiently large with $\eta > 0$. This yields

$$\liminf |\mathfrak{S}(B'_n)| \le D(1 - \eta) < D = \inf_{B \in \mathfrak{C}(\gamma)} |\mathfrak{S}(B)|,$$

contradicting $B'_n \in \mathfrak{C}(\gamma)$. We have therefore shown that

$$\lim_{n \to \infty} \mathbb{P}(B_n \triangle E) = 0. \tag{2.6}$$

Because of Lemma 3 it is sufficient to show that $\mathscr{E}(E) = E$. Let

$$E = \{x: (x - m)^T \Gamma^{-1}(x + m) \le 1\}.$$

Then

$$\mathfrak{S}(B_n) = \frac{1}{\gamma}\int_{B_n}(x - \mathfrak{M}(B_n))(x - \mathfrak{M}(B_n))^T \, d\mathbb{P}$$
$$\ge \frac{1}{\gamma}\int_{E \cap B_n}(x - \mathfrak{M}(B_n))(x - \mathfrak{M}(B_n))^T \, d\mathbb{P},$$

which implies

$$D = \lim |\mathfrak{S}(B_n)| \geq \left| \frac{1}{\gamma} \int_E (x - m)(x - m)^T \, d\mathbb{P} \right|$$

$$\geq \left| \frac{1}{\gamma} \int_E (x - \mathfrak{M}(E))(x - \mathfrak{M}(E))^T \, d\mathbb{P} \right|$$

$$= |\mathfrak{S}(E)|.$$

From this it is seen that

$$|\mathfrak{S}(E)| = \inf_{B \in \mathfrak{C}(\gamma)} |\mathfrak{S}(B)|.$$

As $E \in \mathfrak{C}(\gamma)$ we may choose $B_n = E$ for all n. Then $\mathscr{E}(B_n) = \mathscr{E}(E) = E'$ and from (2.6) we have $\mathbb{P}(E \triangle E') = 0$ giving $E = E'$ as was to be shown. □

3. Consistency and weak convergence. The empirical distribution associated with $(X_j)_1^n$ will be denoted by $\hat{\mathbb{P}}_n$. A finite subset S_n of $\{1, \dots, n\}$ will be called a local minimum if $|\hat{\mathfrak{S}}(S_n)|$ cannot be reduced by interchanging a point of S_n with a point in $\{1, \dots, n\} \setminus S_n$.

THEOREM 2. *If S_n^* is a local minimum, then there exists an ellipsoid E_n^* containing all points X_j with $j \in S_n^*$ and excluding all points X_j with $j \notin S_n^*$. Furthermore, the ellipsoid E_n^* may be chosen to be of the form*

$$E_n^* = \left\{ x \colon (x - m_n^*)^T \Gamma_n^{*-1} (x - m_n^*) \leq r_n^{*2} \right\},$$

where $m_n^* = \hat{\mathfrak{M}}(S_n^* \setminus \{j^*\})$, $\Gamma_n^* = \hat{\mathfrak{S}}(S_n^* \setminus \{j^*\})$ for some $j^* \in S_n^*$ and $r_n^{*2} = (X_{j^*} - m_n^*)^T \Gamma_n^{*-1} (X_{j^*} - m_n^*)$. The integer j^* may be taken to be any $j \in S_n^*$ which maximizes

$$\left(X_j - \hat{\mathfrak{M}}(S_n^* \setminus \{j\})\right)^T \hat{\mathfrak{S}}(S_n^* \setminus \{j\})^{-1} \left(X_j - \hat{\mathfrak{M}}(S_n^* \setminus \{j\})\right).$$

PROOF. We show first that E_n^* contains all the points in $\{X_j \colon j \in S_n^*\}$. We write

$$d(i, j) = \left(X_i - \hat{\mathfrak{M}}(S_n^* \setminus \{j\})\right)^T \hat{\mathfrak{S}}(S_n^* \setminus \{j\})^{-1} \left(X_i - \hat{\mathfrak{M}}(S_n^* \setminus \{j\})\right).$$

It is shown in the Appendix that

$$\max_{i \in S_n^*} d(i, j^*) = d(j^*, j^*) = r_n^{*2}. \tag{3.1}$$

Let $l \notin S_n^*$ and suppose we interchange X_l and an X_j with $j \in S_n^*$. Let $\tilde{S}_n = S_n^* \setminus \{j\} + \{l\}$. Then [see, e.g., Butler (1983), equation 3.13]

$$\begin{aligned} s_n \hat{\mathfrak{S}}(S_n^*) &= (s_n - 1) \hat{\mathfrak{S}}(S_n^* \setminus \{j\}) \\ &\quad + (1 - s_n^{-1}) \left(X_j - \hat{\mathfrak{M}}(S_n^* \setminus \{j\})\right) \left(X_j - \hat{\mathfrak{M}}(S_n^* \setminus \{j\})\right)^T, \end{aligned} \tag{3.2}$$

where s_n is the size of S_n^*. This gives

$$(3.3)\qquad \left|\hat{\mathfrak{S}}(S_n^*)\right| = \left(1 - s_n^{-1}\right)^k \left|\hat{\mathfrak{S}}(S_n^* \setminus \{j\})\right|\left(1 + d(j,j)/s_n\right).$$

Similarly,

$$\left|\hat{\mathfrak{S}}(\tilde{S}_n)\right| = \left(1 - s_n^{-1}\right)^k \left|\hat{\mathfrak{S}}(S_n^* \setminus \{l\})\right|\left(1 + d(l,j)/s_n\right).$$

As S_n^* is a local minimum

$$d(l,j) \geq d(j,j)$$

and on putting $j = j^*$, we see that X_l must lie on the surface or outside of E_n^*. As E_n^* is defined solely in terms of X_j, $j \in S_n^*$, and $\mathbb{P}$ has a density function, it is clear that no X_j with $j \notin S_n^*$ lies on the surface of E_n^* with probability 1. □

In the case $k = 1$ the ellipsoids become intervals and it is possible to show that all locally minimal intervals remain bounded as n increases. We indicate a proof which can be made precise. If the interval becomes arbitrarily large, then one endpoint becomes large whilst the other must remain bounded. Theorem 2 shows that if we exclude one of the endpoints, then the remaining $s_n - 1$ observations are contained in an interval whose midpoint is the mean of the observations in it. This is not possible as the observations will be concentrated in that part of the interval nearest the origin. The boundedness of the intervals together with the argument used for the global minimum to be given in Theorem 3 shows that all local minima will yield consistent estimates for the location and spread of $\mathbb{P}$ in the one-dimensional case. This is an improvement of the corresponding result in Butler (1982) where it was assumed that the observations have a finite mean.

The argument collapses for $k \geq 2$ but can be rescued by assuming that $\mathbb{P}$ has finite second moments. As this is at odds with the spirit of robustness, we do not pursue the problem of locally minimal ellipsoids further.

The situation is different for globally minimal ellipsoids. As we now show they remain bounded as n tends to ∞ and provide consistent estimators of the location and shape of $\mathbb{P}$.

Let $(\hat{E}_n)_1^\infty$ denote the sequence of globally minimal ellipsoids. We require the following result.

LEMMA 4. *With probability* 1 *there exists a compact set* C *of* $\mathbb{R}^k$ *such that* $\hat{E}_n \subset C$ *for all* n *sufficiently large.*

PROOF. According to Theorem 2, we can write

$$\hat{E}_n = \left\{x\colon (x - \hat{m}_n)^T \hat{\Gamma}_n^{-1}(x - \hat{m}_n) \leq \hat{r}_n^2\right\},$$

where $\hat{\Gamma}_n = \hat{\mathfrak{S}}(E_n')$, $\hat{m}_n = \hat{\mathfrak{M}}(E_n')$ and $E_n' = \hat{E}_n \setminus \{X_j\}$ for some $X_j \in \hat{E}_n$.

Consider a ball $K(0,r)$ with $\mathbb{P}(K(0,r)) > \gamma$. It follows that $\hat{\mathbb{P}}_n(K(0,r)) > \gamma$ for all n sufficiently large. For any set S_n of size $s_n = [n\gamma]$ with $X_j \in K(0,r)$

for $j \in S_n$, we have

$$\begin{aligned}\hat{\mathfrak{S}}(S_n) &= \frac{1}{s_n} \sum_{j \in S_n} \left(X_j - \hat{\mathfrak{M}}(S_n)\right)\left(X_j - \hat{\mathfrak{M}}(S_n)\right)^T \\ &\le \frac{n}{s_n} \int_{K(0,r)} (x - \hat{\mathfrak{M}}(S_n))(x - \hat{\mathfrak{M}}(S_n))^T \, d\hat{\mathbb{P}}_n \\ &\le cr^2 I_k,\end{aligned}$$

for all n sufficiently large. Thus $|\hat{\mathfrak{S}}(S_n)| \le c < \infty$ for all n and we may deduce

$$\limsup_{n \to \infty} \left|\hat{\mathfrak{S}}(\hat{E}_n)\right| < \infty. \tag{3.4}$$

For the sequence $(\lambda_{\min}(\hat{\mathfrak{S}}(\hat{E}_n)))_1^\infty$, we have

$$\lambda_{\min}\left(\hat{\mathfrak{S}}(\hat{E}_n)\right) = \frac{n}{s_n} \int_{\hat{E}_n} \left(\left(x - \hat{\mathfrak{M}}(\hat{E}_n)\right)^T \theta_n\right)^2 d\hat{\mathbb{P}}_n,$$

where θ_n, $\|\theta_n\| = 1$, is an eigenvector associated with $\lambda_{\min}(\hat{\mathfrak{S}}(\hat{E}_n))$. This gives

$$\begin{aligned}\lambda_{\min}\left(\hat{\mathfrak{S}}(\hat{E}_n)\right) &= \frac{n}{s_n} \int_{\hat{E}_n \cap K(0,r)} \left(\theta_n^T\left(x - \hat{\mathfrak{M}}(\hat{E}_n)\right)\right)^2 d\hat{\mathbb{P}}_n \\ &\ge \frac{n}{s_n} \int_{\hat{E}_n \cap K(0,r)} \left(\theta_n^T\left(x - \hat{\mathfrak{M}}\left(\hat{E}_n \cap K(0,r)\right)\right)\right)^2 d\hat{\mathbb{P}}_n.\end{aligned}$$

Using the fact that ellipsoids form a class with polynomial discrimination or a Vapnic–Cervonenkis class, standard results in the theory of empirical processes [Pollard (1984), Chapter 2] give

$$\lambda_{\min}\left(\hat{\mathfrak{S}}(\hat{E}_n)\right) \ge \frac{n}{s_n} \int_{\hat{E}_n \cap K(0,r)} \left(\theta_n^T\left(x - \hat{\mathfrak{M}}\left(\hat{E}_n \cap K(0,r)\right)\right)\right)^2 d\mathbb{P} + o_p(1).$$

We choose r sufficiently large so that $\mathbb{P}(K(0,r)) > 1 - \gamma/2$. Now

$$\lambda_{\min}\left(\hat{\mathfrak{S}}(\hat{E}_n)\right) \ge \frac{n}{s_n} \inf \int_{E \cap K(0,r)} (\theta^T(x-y))^2 \, d\mathbb{P} + o(1),$$

where the infimum is taken over all $E \in \mathfrak{E}(\gamma)$, $y \in \mathbb{R}^k$ and all θ, $\|\theta\| = 1$. As $\mathbb{P}$ has a density we have

$$\inf \int_{E \cap K(0,r)} (\theta^T(x-y))^2 \, d\mathbb{P} > 0$$

and consequently

$$\liminf_{n \to \infty} \lambda_{\min}\left(\hat{\mathfrak{S}}(\hat{E}_n)\right) > 0. \tag{3.5}$$

From (3.4) and (3.5) we have immediately

$$\limsup_{n \to \infty} \lambda_{\max}\left(\hat{\mathfrak{S}}(\hat{E}_n)\right) < \infty. \tag{3.6}$$

Let $\hat{S}_n$ be the set of indices of those X_j, $1 \le j \le n$, with $X_j \in \hat{E}_n$. By Theorem 2 there exists a $j_n \in \hat{S}_n$ such that

$$\hat{E}_n = \left\{x: (x - \hat{m}_n)^T \hat{\Gamma}_n^{-1}(x - \hat{m}_n) \le \hat{r}_n^2\right\},$$

with $\hat{m}_n = \mathfrak{M}(\hat{S}_n \setminus \{j_n\})$, $\hat{\Gamma}_n = \hat{\mathfrak{S}}(\hat{S}_n \setminus \{j_n\})$ and

$$\hat{r}_n^2 = \left(X_{j_n} - \hat{m}_n\right)^T \hat{\Gamma}_n^{-1}\left(X_{j_n} - \hat{m}_n\right).$$

The proof leading to (3.4) also shows that

$$\liminf_{n\to\infty} \lambda_{\min}\left(\hat{\Gamma}_n\right) > 0. \tag{3.7}$$

From (3.3) we have

$$\left|\hat{\mathfrak{S}}\left(\hat{E}_n\right)\right| \ge \left(1 - \frac{1}{s_n}\right)^k |\hat{\Gamma}_n|$$

and hence

$$\limsup_{n\to\infty} \lambda_{\max}\left(\hat{\Gamma}_n\right) < \infty. \tag{3.8}$$

From (3.7) and (3.8) it is clear that the trace of $\hat{\Gamma}_n$ is bounded giving

$$\begin{aligned}
\infty &> \limsup_{n\to\infty} \operatorname{trace}\left(\hat{\Gamma}_n\right) \\
&= \frac{n}{s_n}\int_{\hat{E}_n\setminus\{X_{j_n}\}} \|x - \hat{m}_n\|^2 \, d\hat{\mathbb{P}}_n \\
&\ge \frac{n}{s_n}\int_{(\hat{E}_n\setminus\{X_{j_n}\})\cap K(0,r)} \|x - \hat{m}_n\|^2 \, d\mathbb{P}_n \\
&\ge \frac{n}{s_n}\left(\|\hat{m}_n\|^2 - 2\|\hat{m}_n\| r\right)\hat{\mathbb{P}}_n\left(\left(\hat{E}_n \setminus \{X_{j_n}\}\right) \cap K(0,r)\right).
\end{aligned}$$

By choosing r so large that $\liminf_{n\to\infty} \hat{\mathbb{P}}_n((\hat{E}_n \setminus \{X_{j_n}\}) \cap K(0,r)) > 0$, we obtain

$$\limsup_{n\to\infty} \|\hat{m}_n\| < \infty. \tag{3.9}$$

Finally, from (3.7) through (3.9) and $\hat{\mathbb{P}}_n(\hat{E}_n) = s_n/n \to \gamma$, we see that

$$\limsup_{n\to\infty} \hat{r}_n^2 < \infty, \tag{3.10}$$

which completes the proof of the lemma. □

THEOREM 3. *With probability* 1,

$$\lim_{n\to\infty} \left(\mathfrak{M}\left(\hat{E}_n\right), \hat{\mathfrak{S}}\left(\hat{E}_n\right)\right) = \left(0, \rho(\gamma) I_k\right)$$

and

$$\lim_{n\to\infty}\left(\hat{m}_n, \hat{r}_n^2, \hat{\Gamma}_n\right) = \left(0, r^2(\gamma), \rho(\gamma) I_k\right),$$

where

$$\hat{E}_n = \left\{x: (x - \hat{m}_n)^T \hat{\Gamma}_n^{-1}(x - \hat{m}_n) \le \hat{r}_n^2\right\},$$

$r(\gamma)$ *is given by* (2.3) *and*

$$\rho(\gamma) = \frac{2\pi^{k/2}}{k\Gamma(k/2)} \int_0^{r(\gamma)} r^{k+1} f(r^2)\, dr.$$

PROOF. It follows from Lemmas 3 and 4 that it is sufficient to prove the first statement of the theorem.

By Lemma 4 the sequence $(\hat{\mathfrak{M}}(\hat{E}_n), \hat{\mathfrak{S}}(\hat{E}_n))_1^\infty$ is contained in a compact subset of $\mathbb{R}^{k+(1/2)k(k+1)}$. Therefore, there exists a convergent subsequence which we continue to denote by $(\hat{\mathfrak{M}}(\hat{E}_n), \hat{\mathfrak{S}}(\hat{E}_n))_1^\infty$ such that

$$\lim_{n\to\infty}\left(\hat{\mathfrak{M}}(\hat{E}_n), \hat{\mathfrak{S}}(\hat{E}_n)\right) = (\mu, \Sigma),$$

for some μ and Σ. By Theorem 2

$$\hat{\mathfrak{M}}(\hat{E}_n) = \frac{n}{s_n} \int_{\{(x - \hat{\mathfrak{M}}(\hat{E}_n))^T \hat{\mathfrak{S}}(\hat{E}_n)^{-1}(x - \hat{\mathfrak{M}}(\hat{E}_n)) \le \hat{r}_n^2\}} x\, d\hat{\mathbb{P}}_n + o\left(\frac{1}{n}\right)$$

and

$$\hat{\mathfrak{S}}(\hat{E}_n) = \frac{n}{s_n} \int_{\{(x - \hat{\mathfrak{M}}(\hat{E}_n))^T \hat{\mathfrak{S}}(\hat{E}_n)^{-1}(x - \hat{\mathfrak{M}}(\hat{E}_n)) \le \hat{r}_n^2\}} \left(x - \hat{\mathfrak{M}}(\hat{E}_n)\right) \times \left(x - \hat{\mathfrak{M}}(\hat{E}_n)\right)^T d\hat{\mathbb{P}}_n + o\left(\frac{1}{n}\right).$$

By empirical process theory the integrals on the right converge to

$$\frac{1}{\gamma}\int_E x\, d\mathbb{P} \quad \text{and} \quad \frac{1}{\gamma}\int_E (x - \mu)(x - \mu)^T\, d\mathbb{P},$$

respectively, where

$$E = \left\{(x - \mu)^T \Sigma^{-1}(x - \mu) \le r^2\right\}$$

and r is such that $\mathbb{P}(E) = \gamma$.

We obtain

$$\mu = \frac{1}{\gamma}\int_E x\, d\mathbb{P},$$

$$\Sigma = \frac{1}{\gamma}\int_E (x - \mu)(x - \mu)^T\, d\mathbb{P},$$

so that $\mathscr{E}(E) = E$. From Lemma 3 we deduce $(\mu, \Sigma) = (0, \rho(\gamma) I_k)$ and, as this holds for every convergent subsequence, the claim of the theorem follows. □

4. Central limit theorem. We revert to the general case where $\mathbb{P}$ has a density of the form

$$\frac{1}{|\Sigma|} f\big((x-\mu)^T \Sigma^{-1}(x-\mu)\big),$$

for μ in $\mathbb{R}^k$ and Σ some positive definite symmetric $k \times k$ matrix.

We shall restrict attention to the asymptotic distribution of $\sqrt{n}\,(\hat{m}_n - \mu)$ and $\sqrt{n}\,(\hat{\mathfrak{M}}(\hat{E}_n) - \mu)$. The distributions of $\sqrt{n}\,(\hat{m}_n - \mu)$ and $\sqrt{n}\,(\hat{\mathfrak{S}}(\hat{E}_n) - \rho(\gamma)\Sigma)$ can be shown to be asymptotically independent. The latter distribution is rather complex and will be reported elsewhere.

THEOREM 4. *Suppose that f has a continuous first derivative $f^{(1)}$ which is strictly negative and let $r(\gamma)$ be as in* (2.3).

Then $\sqrt{n}\,(\hat{m}_n - \mu)$ and $\sqrt{n}\,(\hat{\mathfrak{M}}(\hat{E}_n) - \mu)$ both converge weakly to a Gaussian random variable with mean 0 *and covariance matrix $\kappa(\gamma)\Sigma$, where*

$$\kappa(\gamma) = \frac{k\Gamma(k/2)\int_0^{r(\gamma)} r^{k+1} f(r^2)\,dr}{8\pi^{k/2}\big(\int_0^{r(\gamma)} r^{k+1} f^{(1)}(r^2)\,dr\big)^2}.$$

PROOF. Suppose without loss of generality that $\mu = 0$ and $\Sigma = I_k$. Let C be a compact set such that the sequence of ellipsoids $(E_n)_1^\infty$ is eventually contained in C with probability 1. The class of functions $\mathfrak{G} = \{x\{x \in E \cap C\}$: E ellipsoid$\}$ has graphs with polynomial discrimination [Pollard (1984), page 27]. For $g \in \mathfrak{G}$ we write $\mathfrak{G} = \{x\{x \in E \cap C\}$: E ellipsoid$\}$,

$$\hat{\mathbb{P}}_n(g) = \int g\,d\hat{\mathbb{P}}_n = \frac{1}{n}\sum_1^n g(X_j)$$

and

$$\mathbb{P}(g) = \int g\,d\mathbb{P}(x).$$

Standard empirical process theory [Pollard (1984), page 27, Sections 7.4 and 7.5] gives $\sqrt{n}\,(\hat{\mathbb{P}}_n(g) - \mathbb{P}(g)) \Rightarrow Z(g)$, $g \in \mathfrak{G}$, where Z is a continuous Gaussian process with mean 0 and covariance structure given by

$$\mathbb{E}(Z(g)Z(g')) = \mathbb{P}(gg') - \mathbb{P}(g)\mathbb{P}(g').$$

As $\lim_{n\to\infty} E_n = E$ the continuity of Z gives

$$\sqrt{n}\,\big(\hat{\mathbb{P}}_n(g_n) - \mathbb{P}(g_n)\big) \Rightarrow Z(g),$$

where $g_n(x) = x\{x \in E_n \cap C\}$ and $g(x) = x\{x \in E \cap C\} = x\{x \in E\}$. From Theorem 2 we have $\hat{\mathbb{P}}_n(g_n) = \hat{m}_n + O_p(1/n)$ and consequently

$$\sqrt{n}\,(\gamma\hat{m}_n - \mathbb{P}(g_n)) \Rightarrow Z(g).$$

Furthermore,

$$\mathbb{P}(g_n) = \int_{E_n} x \, d\mathbb{P}(x)$$

$$= \int_{\{x:\,(x-\hat{m}_n)^T \hat{\Gamma}_n^{-1}(x-\hat{m}_n) \le r_n^2\}} x f(\|x\|^2)\, dx$$

and as $E_n \to E$ almost surely a Taylor expansion gives

$$\mathbb{P}(g_n) = \gamma(\hat{m}_n + \zeta(\gamma)\hat{m}_n)(1 + o_p(1)),$$

where

$$\zeta(\gamma) = \frac{4\pi^{k/2}}{\gamma k \Gamma(k/2)} \int_0^{r(\gamma)} r^{k+1} f^{(1)}(r^2)\, dr.$$

We therefore obtain

$$\sqrt{n}\,\gamma\zeta(\gamma)\hat{m}_n \Rightarrow Z(g),$$

where $Z(g)$ is a normal random variable with mean 0 and covariance matrix

$$\mathbb{E}\big(Z(g)Z(g)^T\big) = \int_E xx^T \, d\mathbb{P}(x)$$

$$= \frac{2\pi^{k/2}}{k\Gamma(k/2)} \int_0^{r(\gamma)} r^{k+1} f(r^2)\, dr\, I_k.$$

This proves the first statement of the theorem. The second follows from Theorem 2 as $\hat{\mathfrak{M}}(E_n) = \hat{m}_n + O_p(1/n)$ and $\hat{\mathfrak{S}}(E_n) = \hat{\mathfrak{S}}(E_n \cap C) + o_p(1)$. □

EXAMPLE 1. Suppose $\mathbb{P}$ is $N_2(0, I_2)$. Then with $\gamma = 0.9$, $k_\gamma = 1.493$ and $\hat{m}_n$ and $\hat{\mathfrak{M}}(\hat{E}_n)$ are 67.0% asymptotically efficient relative to the sample mean.

EXAMPLE 2. Ellipsoid $\hat{E}_n$ provides a robust predictive region for observable X_{n+1}. In such usage it may be of interest to have some prescribed assurance (0.95 say) about a particular amount of predictive coverage (0.90 say). This is achieved by choosing γ so that

$$\mathbb{P}\left(\int_{\hat{E}_n} d\mathbb{P} < 0.9\right) \ge 0.95.$$

Such assurance can in general be achieved asymptotically by considering the limiting distribution of probability coverage. Since

$$0 = \sqrt{n}\,(\gamma - \gamma) = \sqrt{n}\left(\hat{\mathbb{P}}_n(\hat{E}_n) - \mathbb{P}(K(0, r(\gamma)))\right),$$

then adding and subtracting $\mathbb{P}(\hat{E}_n)$ we have

$$\begin{aligned} 0 &= \sqrt{n}\left(\hat{\mathbb{P}}_n(\hat{E}_n) - \mathbb{P}(\hat{E}_n)\right) + \sqrt{n}\left(\mathbb{P}(\hat{E}_n) - \gamma\right) \\ &= Z(K(0, r(\gamma))) + o_p(1) + \sqrt{n}\left(\mathbb{P}(\hat{E}_n) - \gamma\right). \end{aligned} \tag{4.1}$$

As $Z(K(0, r(\gamma))) \sim N(0, \gamma(1-\gamma))$ the approximation (4.1) with $n = 100$ allows one to choose $\gamma = 0.939$ to have approximately 95% assurance of 90% predictive coverage.

APPENDIX

PROOF OF (3.1) IN THE PROOF OF THEOREM 2. We show that $\max_{i \in S_n^*} d(i, j^*) = d(j^*, j^*)$ by computing $d(i, j^*)$ and $d(j^*, j^*)$ directly. It can be shown that

$$d(j^*, j^*) = r_s^{*2} = s_n(s_n - 1)^{-1} d_{j^*}/\left[1 - (s_n - 1)^{-1} d_{j^*}\right],$$

where

$$d_j = \left(X_j - \hat{\mathfrak{M}}(S_n^*)\right)^T \hat{\mathfrak{S}}(S_n^*)^{-1}\left(X_j - \hat{\mathfrak{M}}(S_n^*)\right).$$

Thus, not only does j^* maximize $d(j, j)$ over $j \in S_n^*$, it also maximizes d_j over $j \in S_n^*$. In computing $d(i, j^*)$ use the identity

$$X_j - \hat{\mathfrak{M}}(S_n^* \setminus \{j^*\}) = \left[X_i - \hat{\mathfrak{M}}(S_n^*)\right] + (s_n - 1)^{-1}\left[X_{j^*} - \hat{\mathfrak{M}}(S_n^*)\right]$$

and invert $\hat{\mathfrak{S}}(S_n^* \setminus \{j^*\})$ using (3.2) to derive that

$$\begin{aligned} d(i, j^*) = s_n^{-1}(s_n - 1)\Big\{ & d_i + 2(s_n - 1)^{-1} d_{ij^*} + (s_n - 1)^{-2} d_{j^*} \\ & + \left[(s_n - 1)^{-1} d_{ij^*}^2 + 2(s_n - 1)^{-2} d_{ij^*} d_{j^*} \right. \\ & \left. + (s_n - 1)^{-3} d_{j^*}^2\right]/\left[1 - (s_n - 1)^{-1} d_{j^*}\right]\Big\}, \end{aligned}$$

where

$$d_{ij} = \left(X_i - \hat{\mathfrak{M}}(S_n^*)\right)^T \hat{\mathfrak{S}}(S_n^*)^{-1}\left(X_j - \hat{\mathfrak{M}}(S_n^*)\right).$$

Now $\max_{i \in S_n^*} d_i = d_{j^*}$ and $\max_{i \in S_n^*} d_{ij^*} = d_{j^*j^*} = d_{j^*}$ so $\max_{i \in S_n^*} d(i, j^*)$ is the expression above for $d(i, j^*)$ with d_i replaced with d_{j^*} and d_{ij^*} with d_{j^*}. Upon simplification,

$$\max_{i \in S_n^*} d(i, j^*) = s_n(s_n - 1)^{-1} d_{j^*}/\left[1 - (s_n - 1)^{-1} d_{j^*}\right] = d(j^*, j^*)$$

as was to be shown. □

Acknowledgments. We thank a referee for suggestions. Discussions with David Tyler were helpful.

REFERENCES

BUTLER, R. W. (1982). Nonparametric interval and point prediction using data trimmed by a Grubbs type outlier rule. *Ann. Statist.* **10** 197–204.

BUTLER, R. W. (1983). Outlier discordancy tests in the normal linear model. *J. Roy. Statist. Soc. Ser. B* **45** 120–132.

CHO, S. and MILLER, R. B. (1987). Model-free one-step-ahead prediction intervals: asymptotic theory and small sample simulations. *Ann. Statist.* **15** 1064–1078.

DAVIES, L. (1992a). An efficient, Fréchet differentiable and high breakdown multivariate location and dispersion estimator. *J. Multivariate Anal.* **40** 311–327.

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

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

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

POLLARD, D. (1984). *Convergence of Stochastic Processes*. Springer, New York.

ROUSSEEUW, P. J. (1983). Multivariate estimation with high breakdown point. In *Mathematical Statistics and Applications B* (W. Grossmann, G. Pflug, I. Vincze and W. Wertz, eds.) 283–297. Reidel, Dordrecht.

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

ROUSSEEUW, P. J. and YOHAI, V. (1984). Robust regression by means of S-estimators. *Robust and Nonlinear Time Series. Lecture Notes in Statist.* **26** 267–272. Springer, New York.

R. W. BUTLER
DEPARTMENT OF STATISTICS
COLORADO STATE UNIVERSITY
FORT COLLINS, COLORADO 80523

P. L. DAVIES
FACHBEREICH MATHEMATIK
UNIVERSITÄT GESAMTHOCHSCHULE ESSEN
D-4300 ESSEN 1
GERMANY

M. JHUN
DEPARTMENT OF STATISTICS
KOREA UNIVERSITY
SEOUL, 136-701
KOREA