Ann. Inst. Statist. Math.
Vol. 43, No. 4, 689–705 (1991)

ROBUST M-ESTIMATION OF A DISPERSION MATRIX WITH A STRUCTURE

MADHUSUDAN BHANDARY

Department of Statistics, North Dakota State University, Fargo, ND 58105, U.S.A.

(Received June 2, 1988; revised April 13, 1990)

Abstract. An iterative algorithm for the robust M-estimation of the dispersion matrix of the form $\Gamma + \sigma^2 I_p$ has been given. This algorithm converges after some steps and reduces the effect of outliers on the covariance matrix. The consistency and asymptotic normality of the estimator are established.

Key words and phrases: Robust estimation, M-estimation, elliptically symmetric distribution, outliers, signal processing.

1. Introduction

The covariance and correlation matrices are used for a variety of purposes. They give a simple description of the overall shape of a point-cloud in p-space. They are used in principal component analysis, factor analysis, discriminant analysis, canonical correlation analysis, tests of independence, etc. Unfortunately, sample covariance matrices are excessively sensitive to outliers. Chen *et al.* (1974) gave an example of how a principal component analysis can be sensitive to a few outlying observations. The remedy for this difficulty has been found by robust estimation of dispersion matrices, i.e. an estimate of the dispersion matrix from the sample data which will reduce the effect of outlying observations. In this paper, we consider the problem of the robust estimation of the dispersion matrix $\Gamma + \sigma^2 I_p$, where Γ is n.n.d. of rank q ($< p$), $\sigma^2 > 0$ and both are unknowns and I_p is the identity matrix of order p. The above form of the dispersion matrix arises in the area of signal processing.

2. Model and assumptions

Model. In general, the model in signal processing is as follows:

$$\boldsymbol{X}(t) = A\boldsymbol{S}(t) + \boldsymbol{n}(t) \tag{2.1}$$

where

$\boldsymbol{X}(t) = (X_1(t), \dots, X_p(t))'$: $p \times 1$ observation vector at time t,
$\boldsymbol{S}(t) = (S_1(t), \dots, S_q(t))'$: $q \times 1$ vector of unknown random signals at time t,
$\boldsymbol{n}(t) = (n_1(t), \dots, n_p(t))'$: $p \times 1$ random noise vector at time t

and

$A = [A(\Phi_1), \dots, A(\Phi_q)]$,
$A(\Phi_i)$: $p \times 1$ vector of functions of the elements of unknown Φ_i associated with i-th signal

and

$$q < p.$$

In model (2.1) $\boldsymbol{X}(t)$ is assumed to be distributed with mean vector zero and scale matrix $A\Psi A' + \sigma^2 I_p = \Gamma + \sigma^2 I_p$, where $\Gamma = A\Psi A'$ and Ψ = covariance matrix of $\boldsymbol{S}(t)$. We are interested in the robust estimation of $\Gamma + \sigma^2 I_p$.

ASSUMPTIONS.
(a) $w(s)$ is a non-negative, non-increasing and continuous function for $s \geq 0$.
(b) $\phi(s) = sw(s)$ is bounded and let $K = \text{Sup}_{s \geq 0} \phi(s)$.
(c) ϕ is non-decreasing and is strictly increasing in the interval $\phi < K$.
(d) There exists a s_0 such that $\phi(s_0) > p$ for $s \leq s_0$ and hence $K > p$.
(e) There exists a $a > 0$ such that for every hyperplane H, $P(H) \leq 1 - p/K - a$.

3. Literature review

There are several procedures in the literature about the robust estimation of dispersion matrices. Details can be found in Huber (1981).

A few procedures are mentioned as follows:

(i) Mosteller and Tukey (1977, Chapter 10) suggested a procedure based on the robust regression calculation that result in a robust covariance matrix estimate S^*.

(ii) Gnanadesikan and Kettenring (1972) and Devlin *et al.* (1975) suggested a multivariate trimming (MVT) procedure as follows:

(a) Calculate $\bar{X}$, S from the given data $\boldsymbol{X}_i$ $(i = 1, \dots, n)$.
(b) Calculate $d_i^2 = (\boldsymbol{X}_i - \bar{\boldsymbol{X}})' S^{-1} (\boldsymbol{X}_i - \bar{\boldsymbol{X}})$, the squared distances of the observations from $\bar{\boldsymbol{X}}$ in the metric of S, $i = 1, \dots, n$.
(c) If d_i^2 is very large (w.r.t. the distribution of d_i^2) then throw the i-th observation and calculate $\bar{\boldsymbol{X}}$ and S on the basis of the remaining observations, and so on, for $i = 2, \dots, n$.
(d) Continue this way until $|z^{(u)} - z^{(u-1)}| < 10^{-3}$ or after the 25th iteration, where $z^{(u)}$ = Fisher z-transform of the correlation between two variables at the u-th iteration.

(iii) Following Marona (1976), the estimates of $\boldsymbol{t}$ and Σ are given by

$$\hat{\boldsymbol{t}} = \frac{\sum_{i=1}^{n} w_1(d_i) \boldsymbol{x}_i}{\sum_{i=1}^{n} w_1(d_i)}, \tag{3.1}$$
$$\hat{\Sigma} = \frac{1}{n} \sum_{i=1}^{n} w_2(d_i^2)(\boldsymbol{x}_i - \boldsymbol{m}^*)(\boldsymbol{x}_i - \boldsymbol{m}^*)'$$

where

$$w_1(d_i) = \frac{p+f}{f+d_i^2} w_2(d_i^2), \qquad i = 1, \ldots, n.$$

f is the number of degrees of freedom of the p-variate t-distribution, and $\boldsymbol{t}$ and Σ are the mean vector and the scale matrix of the distribution, respectively.

(iv) An alternative procedure considered by Huber (1977*a*, 1977*b*) is the same as that suggested by, Marona (1976) except that

$$w_1(d_i) = \begin{cases} 1 & \text{if } d_i \le k \\ k/d_i & \text{if } d_i > k \end{cases} \quad \text{and} \quad w_2(d_i^2) = [w_1(d_i)]/\beta \tag{3.2}$$

where k^2 is the 90% point of a χ_p^2 distribution and β is chosen so as to make S^* an asymptotically unbiased estimator of the covariance matrix in a multivariate normal situation.

Devlin *et al.* (1975) compared the above four methods, (i), (ii), (iii) and (iv), through numerical study and obtained the robust estimate of the principal components.

Let us now discuss the robust estimation of the dispersion matrix $\Gamma + \sigma^2 I_p$. We will follow Marona's way (1976) and that of Jöreskog (1967).

In Section 4 we will discuss the derivation of the estimate. Asymptotic normality and strong consistency of the estimate are discussed in Sections 5 and 6. Section 7 describes, through numerical study, the convergence of the iteration process for the estimate.

4. Derivation of the estimate

Let $\boldsymbol{X}(t_1), \boldsymbol{X}(t_2), \ldots, \boldsymbol{X}(t_n)$ be n observed p-component signals at n different time points which are independently and identically distributed as an elliptically symmetric distribution (Kelker (1970)) with mean vector zero and scale matrix $\Gamma + \sigma^2 I_p$, where Γ, σ^2 and I_p are explained in Section 3.

Since Γ is of rank q ($< p$), we can assume $\Gamma = BB'$, where B is a $p \times q$ matrix of rank q and $B'B = \mathrm{Diag}(\theta_1, \ldots, \theta_q)$, where $\theta_1 \ge \theta_2 \ge \cdots \ge \theta_q$ are the non-zero eigenvalues of Γ. Hence, we can write the log-likelihood as follows:

$$\log L = -\frac{n}{2} \log |BB' + \sigma^2 I_p| + \sum_{i=1}^{n} \log h(\mathrm{tr}(BB' + \sigma^2 I_p)^{-1} \boldsymbol{x}_i \boldsymbol{x}_i') \tag{4.1}$$

where $\boldsymbol{x}_i = x(t_i)$, $i = 1, \ldots, n$ and $h(\cdot)$ is a convex function from $[0, \infty)$ to $[0, \infty)$. Following Lawley and Maxwell ((1963), Chapter 2), we have

$$\begin{aligned} \frac{\partial \log L}{\partial B} &= \Bigg[-\frac{n}{2}(BB' + \sigma^2 I_p)^{-1} \\ &\quad + \sum_{i=1}^{n} \frac{h'(d_i^2)}{h(d_i^2)} (-(BB' + \sigma^2 I)^{-1} \boldsymbol{x}_i \boldsymbol{x}_i' (BB' + \sigma^2 I)^{-1}) \Bigg] 2B = 0 \quad \text{or} \\ &\Sigma^{-1}(\Sigma - T)\Sigma^{-1} B = 0 \end{aligned} \tag{4.2}$$

and

$$\frac{\partial \log L}{\partial \sigma^2} = \operatorname{tr}(\Sigma^{-1}(\Sigma - T)\Sigma^{-1}) = 0 \tag{4.3}$$

where

$$\begin{aligned} &d_i^2 = \boldsymbol{x}_i'(BB' + \sigma^2 I)^{-1}\boldsymbol{x}_i, \\ &h'(d_i^2) = \frac{\partial h(d_i^2)}{\partial d_i^2}, \qquad w(d_i^2) = -\frac{h'(d_i^2)}{h(d_i^2)}, \qquad i = 1, \dots, n, \\ &\Sigma = BB' + \sigma^2 I, \qquad T = \frac{2}{n}\sum_{i=1}^{n} w(d_i^2)\boldsymbol{x}_i\boldsymbol{x}_i'. \end{aligned} \tag{4.4}$$

Using Rao (1983, p. 33) we can write

$$\begin{aligned} \Sigma^{-1} &= (BB' + \sigma^2 I)^{-1} \\ &= \frac{1}{\sigma^2}\left(I_p - B(I + D)^{-1}\frac{B'}{\sigma^2}\right) \end{aligned} \tag{4.5}$$

where

$$D = \frac{B'B}{\sigma^2} = \operatorname{Diag}(\theta_1/\sigma^2, \dots, \theta_q/\sigma^2).$$

Using (4.5) in (4.2) we get

$$\begin{aligned} &(\Sigma - T)\frac{1}{\sigma^2}\left[I_p - B(I + D)^{-1}\frac{B'}{\sigma^2}\right]B = 0, \qquad (\Sigma - T)\frac{B}{\sigma^2}(I + D)^{-1} = 0, \\ &(\Sigma - T)\tilde{B} = 0 \end{aligned} \tag{4.6}$$

or

$$T\tilde{B} = \tilde{B}\operatorname{Diag}(\theta_1 + \sigma^2, \dots, \theta_q + \sigma^2, \sigma^2, \dots, \sigma^2). \tag{4.7}$$

From (4.7) it is clear that the columns of B are the eigen vectors of T and that the estimates of $\theta_1 + \sigma^2, \dots, \theta_q + \sigma^2$ and σ^2 are based on the eigenvalues of T. We choose the eigen vectors of T such that these are orthonormal. From (4.3), (4.5) and (4.6), we have

$$\operatorname{tr}\left[\Sigma^{-1}\left(I - \frac{1}{\sigma^2}\left\{T - \Sigma\tilde{B}(I + D)^{-1}\frac{\tilde{B}'}{\sigma^2}\right\}\right)\right] = 0. \tag{4.8}$$

It can be shown after calculation that (4.8) can be deduced to

$$\operatorname{tr}\left[\frac{1}{\sigma^2}\left(I + \frac{\tilde{B}\tilde{B}'}{\sigma^2} - \frac{T}{\sigma^2}\right)\right] = 0 \quad \text{or} \quad \operatorname{tr}\left[\frac{1}{\sigma^2}\left(I + \frac{\tilde{B}\tilde{B}'}{\sigma^2} - \frac{T}{\sigma^2}\right)\right] = 0 \tag{4.9}$$

where

$$\tilde{B} = \hat{B}(\Theta - \sigma^2 I_q)^{1/2} \qquad \text{and}$$

$\Theta = \mathrm{Diag}(l_1, \ldots, l_q)$,
(4.10) l_i = i-th ordered eigenvalue of T when these are in decreasing order and $\hat{B}'\hat{B} = I_q$.

From (4.6) to (4.9) we are using the fact that

$$\left.\frac{\delta \log L}{\delta B}\right|_{B=\hat{B}} = 0.$$

From (4.9) we get

$$\frac{1}{\sigma^2}\left[p + \sum_{i=1}^{q}\frac{(l_i - \sigma^2)}{\sigma^2} - \sum_{i=1}^{p}\frac{l_i}{\sigma^2}\right] = 0 \quad \text{or} \quad \hat{\sigma}^2 = \frac{\sum_{i=q+1}^{p} l_i}{p-q}. \tag{4.11}$$

Thus,

$$\hat{\Sigma} = \hat{B}(\hat{\Lambda} - \hat{\sigma}^2 I)\hat{B}' + \hat{\sigma}^2 I_p \tag{4.12}$$

where,

$\hat{B} = (\boldsymbol{w}_1 : \cdots : \boldsymbol{w}_q)$,
$\hat{\Lambda} = \mathrm{Diag}(l_1, \ldots, l_q)$,
l_i = i-th ordered eigenvalue of T,
$\boldsymbol{w}_i$ = i-th orthonormal eigen vector of T corresponding to l_i, $i = 1, \ldots, p$.

So in order to get, the robust estimate of Σ we can apply the following algorithm:

(i) The observations are $\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_n$. Calculate $S = (1/n)\sum_{i=1}^{n}\boldsymbol{x}_i\boldsymbol{x}_i' = \hat{\Sigma}^{(0)}$.

(ii) Calculate $d_{i(1)}^2 = \boldsymbol{x}_i'S^{-1}\boldsymbol{x}_i$ and $T^{(1)} = (1/n)\sum_{i=1}^{n} w(d_{i(1)}^2)\boldsymbol{x}_i\boldsymbol{x}_i'$ where $w(d_{i(1)}^2)$ is a decreasing function and we choose, for practical purposes, $w(d_{i(1)}^2) = (m+p)/(m+d_{i(1)}^2)$ which is the maximum likelihood estimator from a p-variate t-distribution with m.d.f.

(iii) Compute the eigenvalues of $T^{(1)}$ and order them as follows: $l_1 \geq l_2 \geq \cdots \geq l_q \geq l_{q+1} \geq \cdots \geq l_p$ and the corresponding eigen vectors are $\boldsymbol{w}_1, \ldots, \boldsymbol{w}_q, \boldsymbol{w}_{q+1}, \ldots, \boldsymbol{w}_p$. Finally, compute $\hat{\Sigma}^{(1)} = \sum_{i=1}^{q}(l_i - \hat{\sigma}^2)\boldsymbol{w}_i\boldsymbol{w}_i' + \hat{\sigma}^2 I_p$, where $\hat{\sigma}^2 = (\sum_{i=q+1}^{p} l_i)/(p-q)$.

(iv) Calculate $d_{i(2)}^2 = \boldsymbol{x}_i'\hat{\Sigma}(1)^{-1}\boldsymbol{x}_i$ and $T^{(2)} = (1/n)\sum_{i=1}^{n} w(d_{i(2)}^2)\boldsymbol{x}_i\boldsymbol{x}_i'$.

(v) Repeat steps (iii) and (iv) until the iteration converges, i.e., $\|\hat{\Sigma}^{(r)} - \hat{\Sigma}^{(r-1)}\| < \epsilon$, where $\hat{\Sigma}^{(r)}$ = estimate of Σ at the r-th iteration and ϵ = some pre-assigned small numbers.

Remark. Here, we assumed that the value of q is known. If q is unknown, Zhao *et al.* (1986*a*, 1986*b*) gave some model selection method to estimate q.

5. Asymptotic normality of the estimate

The asymptotic normality of the estimate $(\hat{\sigma}_n^2, \hat{B}_n)$ will be proved by using the result of Huber (1967).

LEMMA 5.1. *Let*

(i) Ψ *be the function from* $R^p \times \Theta^0$ *into* Θ^0 *defined by* $\Psi(x, \theta) = (\Psi_1(x, \theta), \Psi_2(x, \theta))$ *where*

$$\Psi_1(x, \theta) = \frac{\partial \log L}{\partial \sigma^2} = \mathrm{tr}[\Sigma^{-1}(\Sigma - T)\Sigma^{-1}], \tag{5.1}$$

$$\Psi_2(x, \theta) = \frac{\partial \log L}{\partial B} = \Sigma^{-1}(\Sigma - T)\Sigma^{-1}B, \tag{5.2}$$

$$\theta = \left(\sigma^2, \underset{p\times q}{B}\right),$$

$$\Theta^0 = R_+ \times \Theta,$$

$$\Theta = \textit{Set of } p \times q \textit{ reql matrices},$$

$$R_+ = \textit{Set of positive real numbers}.$$

(ii) $\lambda(\theta) = (\lambda_1(\theta), \lambda_2(\theta)) = E_P\Psi(x, \theta)$ *where P is the underlying distribution and E denotes the expectation operator.*

(iii) *The vector space* Θ^0 *is normed with* $\|\theta\| = \max\{\sigma^2, \|B\|\}$.

(iv) $U_j(x, \theta, \delta) = \mathrm{Sup}_{\|\theta_1 - \theta\| < \delta} \|\Psi_j(x, \theta_1) - \Psi_j(x, \theta)\|$, $j = 1, 2$.

(v) *There exist positive numbers b, c and* δ_0 *such that* $EU_j(x, \theta, \delta) \leq b\delta$ *and* $EU_j^2(x, \theta, \delta) \leq c\delta$ *for* $\|\theta - \theta_0\| + \delta \leq \delta_0$, $j = 1, 2$, *where* θ_0 *is the true parameter.*

(vi) *The derivative* $(D\lambda)_{\theta_0}$ *is non-singular, then* $\sqrt{n}(\hat{\theta}_n - \theta_0) \xrightarrow{L}$ *Normal distribution with zero mean and covariance matrix* $(D\lambda)_{\theta_0}^{-1} C (D\lambda)_{\theta_0}^{-1'}$, *where C is the covariance matrix of* $\Psi(x, \theta_0)$ *and* $\hat{\theta}_n = (\hat{\sigma}_n^2, \hat{B}_n)$.

PROOF. Since $E_P((\partial \log L)/\partial \theta_0) = 0$, we have $\lambda(\theta_0) = 0$ and this satisfies the assumption $(N-2)$ of Huber (1967). When P_n is the emperical distribution, the estimator $\hat{\theta}_n = (\hat{\sigma}_n^2, \hat{B}_n)$ is defined by $E_{P_n}\Psi(x, \theta) = 0$. Lemma 5.1 follows from Huber's (1967) theorem and its corollary.

LEMMA 5.2.

$$E_P\left[\frac{1}{n}\sum_{i=1}^{n} \phi(|\boldsymbol{x}_i|^2)\right] = p, \qquad \textit{where} \quad \phi(s) = sw(s), \quad s \geq 0.$$

PROOF. We have from the likelihood equation,

$$\mathrm{tr}[\Sigma^{-1}(\Sigma - T)\Sigma^{-1}] = 0. \tag{5.3}$$

If $\Sigma = I_p$ then (5.1) can be deduced to

$$E \operatorname{tr}\left(I_p - \frac{1}{n}\sum_{i=1}^{n} w(|\boldsymbol{x}_i|^2)\boldsymbol{x}_i\boldsymbol{x}_i'\right) = 0 \quad \text{or}$$

$$E\left[\frac{1}{n}\sum_{i=1}^{n}\phi(|\boldsymbol{x}_i|^2)\right] = p.$$

LEMMA 5.3. *Let*

$$\begin{aligned} C &= E\frac{\partial\bar{\Psi}_2}{\partial\bar{B}'}\bigg|_{\Sigma=I} \\ &= [-(B_0 \otimes I_p)'\{(B_0 \otimes I)P^* + (I_p \otimes B_0)\} + I_{pq}][I_{pq} - (I_q \otimes ET\,|_{\Sigma=I})] \\ &\quad + (B_0 \otimes I_p)'(ET\,|_{\Sigma=I} \otimes I_p)'\{(B_0 \otimes I_p)P^* + (I \otimes B_0)\} \\ &\quad - (B_0 \otimes I_p)'E\frac{\partial\bar{T}}{\partial\bar{B}'}\bigg|_{\Sigma=I} \end{aligned}$$

where P^* is the permutation matrix given in Rao (1982) and B_0 is such that $\theta_0 = (\sigma_0^2, B_0)$ is a unique solution of the likelihood equations with $\sigma_0^2 \neq 1$. Then C is a non-negative definite matrix.

PROOF. We have for $\boldsymbol{l} \neq \boldsymbol{0}$,

$$\begin{aligned} \boldsymbol{l}'(I_p - ET\,|_{\Sigma=I})\boldsymbol{l} &= \boldsymbol{l}'\left(I_p - E\frac{1}{n}\sum_{i=1}^{n} w(|\boldsymbol{x}_i|^2)\boldsymbol{x}_i\boldsymbol{x}_i'\right)\boldsymbol{l} \\ &= \boldsymbol{l}'\boldsymbol{l} - \frac{1}{n}\sum_{i=1}^{n} E|\boldsymbol{x}_i|^2 w(|\boldsymbol{x}_i|^2)E(\boldsymbol{l}'\boldsymbol{z}_i)^2 \qquad (5.4) \\ &= \boldsymbol{l}'\boldsymbol{l} - E\left(\frac{1}{n}\sum_{i=1}^{n}\phi(|\boldsymbol{x}_i|^2)\right)\frac{|\boldsymbol{l}|^2}{p} \qquad (5.5) \\ &= 0 \quad \text{(by Lemma 5.2).} \end{aligned}$$

Note that (5.4) and (5.5) are true because $\boldsymbol{z}_i = \boldsymbol{x}_i/|\boldsymbol{x}_i| \sim$ uniform distribution over a unit sphere, independently of $|\boldsymbol{x}_i|$ and hence, $E(\boldsymbol{l}'\boldsymbol{z}_i)^2 = |\boldsymbol{l}|^2/p$. Hence the matrix $I_q \otimes (I_p - ET\,|_{\Sigma=I})$ can be ignored w.r.t.the definiteness of the matrix and it is easy to show that the matrix $-(B_0 \otimes I_p)'E(\partial\bar{T}/\partial\bar{B})$ is n.n.d. and the matrix $[B_0'ET\,|_{\Sigma=I} \otimes I_p][(B_0 \otimes I)P^* + I \otimes B_0]$ is n.n.d.; hence, Lemma 5.3.

LEMMA 5.4.

$$\frac{\overline{\partial U(x)V(x)}}{\partial\bar{x}'} = (V \otimes I_p)'\frac{\partial\bar{U}'}{\partial\bar{x}'} + (I_r \otimes U)\frac{\partial\bar{V}'}{\partial\bar{x}'}$$

where

$U(x)$ is a $p \times q$ matrix function of x and

$V(x)$ is a $q \times r$ matrix function of x.

PROOF. The proof of Lemma 5.4 is given in Rao (1982).

LEMMA 5.5.

$$\frac{\partial\bar{\Psi}_2}{\partial\bar{B}'} = [-(B\otimes I_p)'(\Sigma^{-1}\otimes\Sigma^{-1})\{(B\otimes I_p)P^* + (I_p\otimes B)\} + (I_q\otimes\Sigma^{-1})]$$
$$\cdot [I_{pq} - (I_q\otimes\Sigma^{-1}T)]$$
$$+ (\Sigma^{-1}B\otimes I_p)'(T\otimes I_p)'(\Sigma^{-1}\otimes\Sigma^{-1})\{(B\otimes I_p)P^* + (I_p\otimes B)\}$$
$$- (\Sigma^{-1}B\otimes I_p)'(I_p\otimes\Sigma^{-1})\frac{\partial\bar{T}}{\partial\bar{B}'}$$

where P^* is the permutation matrix given in Rao (1982).

PROOF.

$$\frac{\partial\bar{\Psi}_2}{\partial\bar{B}'} = \frac{\partial\overline{\Sigma^{-1}B}}{\partial\bar{B}'} - \frac{\partial\overline{\Sigma^{-1}T\Sigma^{-1}B}}{\partial\bar{B}'}. \tag{5.6}$$

Now,

$$\frac{\partial\overline{\Sigma^{-1}B}}{\partial\bar{B}'} = (B\otimes I_p)'\frac{\partial\overline{\Sigma^{-1}}}{\partial\bar{B}'} + (I_q\otimes\Sigma^{-1})\frac{\partial\bar{B}}{\partial\bar{B}'} \quad \text{(by Lemma 5.3)} \tag{5.7}$$
$$= -(B\otimes I_p)'(\Sigma^{-1}\otimes\Sigma^{-1})\frac{\partial\bar{\Sigma}}{\partial\bar{B}'} + (I_q\otimes\Sigma^{-1}).$$

Similarly,

$$\frac{\partial\overline{\Sigma^{-1}T\Sigma^{-1}B}}{\partial\bar{B}'} = (\Sigma^{-1}B\otimes I_p)'\frac{\partial\overline{\Sigma^{-1}T}}{\partial\bar{B}'} + (I_q\otimes\Sigma^{-1}T)\frac{\partial\overline{\Sigma^{-1}B}}{\partial\bar{B}'} \tag{5.8}$$
$$= (\Sigma^{-1}B\otimes I_p)'\left[(T\otimes I_p)'\frac{\partial\bar{\Sigma}^{-1}}{\partial\bar{B}'} + (I_p\otimes\Sigma^{-1})\frac{\partial\bar{T}}{\partial\bar{B}'}\right]$$
$$+ (I_q\otimes\Sigma^{-1}T)\frac{\partial\overline{\Sigma^{-1}B}}{\partial\bar{B}'}.$$

Lemma 5.5 follows from (5.6), (5.7) and (5.8).

THEOREM 5.1. *Let*
(i) *the function $s\phi(s)$ be bounded*
(ii) *P be a radial distribution*
(iii) *the likelihood equations have a unique solution $\theta_0 = (\sigma_0^2, B_0)$, where $\sigma_0^2 \neq 1$*
(iv) $E_P\{|\boldsymbol{x}|^2\phi'(|\boldsymbol{x}|^2)\} > 0$.
Then $\sqrt{n}(\hat{\theta}_n - \theta_0)$ has a limit normal distribution.

PROOF. Here, we will apply Lemma 5.1 in order to prove Theorem 5.1. So, we have to verify the conditions in Lemma 5.1. If we can show that $\|D\Psi(x, \theta_0)\| <$

d, where d (> 0) is some constant, then by the Mean Value Theorem, the condition (v) in Lemma 5.1 will be verified. From (5.1) we have,

$$(5.9)\quad \frac{\partial \Psi_1}{\partial \sigma^2} = \frac{\partial}{\partial \sigma^2}[\operatorname{tr}\Sigma^{-1} - \operatorname{tr}\Sigma^{-1}T\Sigma^{-1}]$$
$$= \operatorname{tr}\left[\Sigma^{-1}(2T\Sigma^{-1} - I_p)\Sigma^{-1} + \Sigma^{-1}\Sigma^{-1}\sum_{i=1}^{n} w'(d_i^2)\boldsymbol{x}_i\boldsymbol{x}_i'\Sigma^{-1}\boldsymbol{x}_i\boldsymbol{x}_i'\Sigma^{-1}\right].$$

Similarly,

$$(5.10)\qquad \frac{\partial \Psi_1}{\partial B} = 2\Sigma^{-1}(2T\Sigma^{-1} - I_p)\Sigma^{-1}B + \Sigma^{-1}\Sigma^{-1}\sum_{i=1}^{n} d_i^2 w'(d_i^2)\boldsymbol{x}_i\boldsymbol{x}_i'\Sigma^{-1}B.$$

Hence,

$$(5.11)\qquad \frac{\partial \Psi_1}{\partial \bar{B}'} = \operatorname{Vec}\left(\frac{\partial \Psi_1}{\partial B}\right)$$
$$= 2(\Sigma^{-1}(2T\Sigma^{-1} - I_p) \otimes B')\operatorname{Vec}(\Sigma^{-1})$$
$$+ 2\left(\Sigma^{-1}\Sigma^{-1}\sum_{i=1}^{n} d_i^2 w'(d_i^2)\boldsymbol{x}_i\boldsymbol{x}_i' \otimes B'\right)\operatorname{Vec}(\Sigma^{-1}).$$

From (5.2) we get,

$$(5.12)\quad \frac{\partial \Psi_2}{\partial \sigma^2} = \frac{\partial \Sigma^{-1}B}{\partial \sigma^2} - \frac{\partial}{\partial \sigma^2}(\Sigma^{-1}T\Sigma^{-1}B)$$
$$= -\Sigma^{-1}\Sigma^{-1}B + \Sigma^{-1}\Sigma^{-1}T\Sigma^{-1}B + \Sigma^{-1}\Sigma^{-1}\sum_{i=1}^{n} d_i^2 w'(d_i^2)\boldsymbol{x}_i\boldsymbol{x}_i'\Sigma^{-1}B$$
$$+ \Sigma^{-1}T\Sigma^{-1}\Sigma^{-1}B.$$

So,

$$(5.13)\qquad \frac{\partial \bar{\Psi}_2}{\partial \sigma^2} = \operatorname{Vec}\left(\frac{\partial \Psi_2}{\partial \sigma^2}\right)$$
$$= -(\Sigma^{-1} \otimes B')\operatorname{Vec}(\Sigma^{-1}) + (\Sigma^{-1}\Sigma^{-1}T \otimes B')\operatorname{Vec}(\Sigma^{-1})$$
$$+ \left(\Sigma^{-1}\Sigma^{-1}\sum_{i=1}^{n} d_i^2 w'(d_i^2)\boldsymbol{x}_i\boldsymbol{x}_i' \otimes B'\right)\operatorname{Vec}(\Sigma^{-1})$$
$$+ (\Sigma^{-1}T\Sigma^{-1} \otimes B')\operatorname{Vec}(\Sigma^{-1})$$

The expression for $\partial\bar{\Psi}_2/\partial\bar{B}'$ is given in Lemma 5.5.

Now,

$$(5.14)\qquad \left\|\begin{bmatrix} \frac{\partial \Psi_1}{\partial \sigma^2} & \frac{\partial \Psi_1}{\partial \bar{B}'} \\ \frac{\partial \bar{\Psi}_2}{\partial \sigma^2} & \frac{\partial \bar{\Psi}_2}{\partial \bar{B}'} \end{bmatrix}\right\| = \left|\frac{\partial \Psi_1}{\partial \sigma^2}\right| + \left\|\frac{\partial \Psi_1}{\partial \bar{B}'}\right\| + \left\|\frac{\partial \bar{\Psi}_2}{\partial \sigma^2}\right\| + \left\|\frac{\partial \bar{\Psi}_2}{\partial \bar{B}'}\right\|$$
$$< B_1$$

where B_1 is a constant depending on $\|\Sigma\|$ and $\|B\|$. We get (5.14) by using

(i) (5.9), Lemma 5.5, (5.11) and (5.13),

(ii) the fact that $d_i^{-1}|\boldsymbol{x}_i| \le \|\Sigma\|^{1/2}$,

(iii) $\|A \otimes B\| = \|A\| \cdot \|B\|$,

(iv) the function $\phi(s)$ is bounded for $s \ge 0$. Now, we will verify the other conditions in Lemma 5.1, i.e. $ED\Psi(x, \theta)$ is non-singular.

Without loss of generality, we assume that P is spherically symmetric and we take $\Sigma = I_p$ so that $\theta_0 = (\sigma_0^2, B_0)$, where $\sigma_0^2 \neq 1$.

Note that

$$w'(s) = s^{-1}(\phi'(s) - w(s)). \tag{5.15}$$

Since P is spherically symmetric, we will use the property that $\boldsymbol{z}_i = \boldsymbol{x}_i/|\boldsymbol{x}_i|$ and $|\boldsymbol{x}_i|$ are independently distributed and $\boldsymbol{z}_i = \boldsymbol{x}_i/|\boldsymbol{x}_i| \sim$ uniform distribution over the unit sphere. Hence,

$$\begin{aligned} ET\,|_{\Sigma=I} &= E\frac{1}{n}\sum_{i=1}^{n} w(|\boldsymbol{x}_i|^2)\boldsymbol{x}_i\boldsymbol{x}_i' \\ &= \frac{1}{n}\sum_{i=1}^{n} E|\boldsymbol{x}_i|^2 w(|\boldsymbol{x}_i|^2)\cdot E\boldsymbol{z}_i\boldsymbol{z}_i' \\ &= \frac{1}{n}\sum_{i=1}^{n} E\phi(|\boldsymbol{x}_i|^2)\cdot E\boldsymbol{z}_i\boldsymbol{z}_i'. \end{aligned} \tag{5.16}$$

Now, from (5.9) we have,

$$E_P\frac{\partial\Psi_1}{\partial\sigma^2}\bigg|_{\Sigma=I} = E_P \operatorname{tr}\left[(2T\,|_{\Sigma=I} - I_p) + \sum_{i=1}^{n} w'(|\boldsymbol{x}_i|^2)\boldsymbol{x}_i|\boldsymbol{x}_i|^2\boldsymbol{x}_i'\right]. \tag{5.17}$$

Using (5.15) and (5.16) in (5.17) we get,

$$\begin{aligned} E\frac{\partial\Psi_1}{\partial\sigma^2}\bigg|_{\Sigma=1} &= E\frac{1}{n}\sum_{i=1}^{n}\phi(|\boldsymbol{x}_i|^2) + E\frac{1}{n}\sum_{i=1}^{n}|\boldsymbol{x}_i|^2\phi'(|\boldsymbol{x}_i|^2) - p \\ &> 0 \quad \text{(using Lemma 5.2 and Condition (iv) of Theorem 5.1).} \end{aligned} \tag{5.18}$$

Now, from (5.13) we have,

$$\begin{aligned} E\frac{\partial\bar{\Psi}_2}{\partial\sigma^2}\bigg|_{\Sigma=I} &= \Bigg[\Bigg\{(2ET\,|_{\Sigma=I} - I_p) \\ &\qquad + E\frac{1}{n}\sum_{i=1}^{n}|\boldsymbol{x}_i|^2 w'(|\boldsymbol{x}_i|^2)\boldsymbol{x}_i\boldsymbol{x}_i'\Bigg\}\otimes B_0'\Bigg]\operatorname{Vec}(I_p) \\ &= \Bigg[\Bigg\{\left(2\frac{1}{n}\sum_{i=1}^{n}E\phi(|\boldsymbol{x}_i|^2)\cdot E\boldsymbol{z}_i\boldsymbol{z}_i' - I_p\right) \\ &\qquad + \frac{1}{n}\sum_{i=1}^{n}E|\boldsymbol{x}_i|^2\phi'(|\boldsymbol{x}_i|^2)\cdot E\boldsymbol{z}_i\boldsymbol{z}_i' \\ &\qquad - \frac{1}{n}\sum_{i=1}^{n}E\phi(|\boldsymbol{x}_i|^2)\cdot E\boldsymbol{z}_i\boldsymbol{z}_i'\Bigg\}\otimes B_0'\Bigg]\operatorname{Vec}(I_p). \end{aligned} \tag{5.19}$$

Using (5.11) we get

$$E\frac{\partial\Psi_1}{\partial\bar{B}'}\bigg|_{\Sigma=I} = 2E\frac{\partial\bar{\Psi}_2}{\partial\sigma^2}\bigg|_{\Sigma=I}. \tag{5.20}$$

It can be shown using Lemma 5.5 that

$$\begin{aligned} C = {} & E\frac{\partial\bar{\Psi}_2}{\partial\bar{B}'}\bigg|_{\Sigma=I} \\ = {} & -(B_0\otimes I_p)'\bigg[\{(B_0\otimes I_p)P^* + (I_p\otimes B_0)\} \\ & \qquad\cdot\{I_q\otimes(I_p - ET\,|_{\Sigma=I})\} + E\frac{\partial\bar{T}}{\partial\bar{B}'}\bigg|_{\Sigma=I}\bigg] \\ & + I_q\otimes(I_p - ET\,|_{\Sigma=I}) \\ & + (B_0'ET\,|_{\Sigma=I}\otimes I_p)[(B_0\otimes I_p)P^* + (I_p\otimes B_0)]. \end{aligned} \tag{5.21}$$

Now,

$$\begin{aligned} D(x,\theta) &= \begin{bmatrix} b & E\dfrac{\partial\Psi_1}{\partial\bar{B}'}\bigg|_{\Sigma=I} \\ E\dfrac{\partial\bar{\Psi}_2}{\partial\sigma^2}\bigg|_{\Sigma=I} & C \end{bmatrix} \\ &= \begin{bmatrix} b & 2\boldsymbol{a}' \\ \boldsymbol{a} & C \end{bmatrix}, \end{aligned} \tag{5.22}$$

where $b = E(\partial\Psi_1/\partial\sigma^2)\,|_{\Sigma=I}$, $\boldsymbol{a} = E(\partial\bar{\Psi}_2/\partial\sigma^2)\,|_{\Sigma=I}$ and C is given by (5.21).

Now, we will evaluate the function $D\Psi(x,\theta)$ at some point $(\sigma_0^2, \mathrm{Vec}(B_0))$ where $\sigma_0^2\in R_+$ and $B_0\in\Theta$.

Now,

$$\begin{aligned} & (\sigma_0^2, \{\mathrm{Vec}(B_0)\}')\begin{bmatrix} b & 2\boldsymbol{a}' \\ \boldsymbol{a} & C\end{bmatrix}\begin{pmatrix}\sigma_0^2 \\ \mathrm{Vec}(B_0)\end{pmatrix} \\ & \quad = b\sigma_0^4 + 3\sigma_0^2[\mathrm{Vec}(B_0)]'\boldsymbol{a} + [\mathrm{Vec}(B_0)]'C[\mathrm{Vec}(B_0)]. \end{aligned} \tag{5.23}$$

Thus, in order to show that (5.22) is non-singular, it is enough to show by using (5.18) and Lemma 5.3 that $[\mathrm{Vec}(B_0)]'\boldsymbol{a}\geq 0$. From (5.19) we have,

$$\begin{aligned} \boldsymbol{a} &= E\frac{\partial\bar{\Psi}_2}{\partial\sigma^2}\bigg|_{\Sigma=I} \\ &= \Bigg[\Bigg\{\frac{1}{n}\sum_{i=1}^n E\phi(|\boldsymbol{x}_i|^2)\cdot E\boldsymbol{z}_i\boldsymbol{z}_i' - I_p \\ &\qquad + \frac{1}{n}\sum_{i=1}^n E|\boldsymbol{x}_i|^2\cdot\phi'(|\boldsymbol{x}_i|^2)\cdot E\boldsymbol{z}_i\boldsymbol{z}_i'\Bigg\}\otimes B_0'\Bigg]\mathrm{Vec}(I_p) \\ &= [M\otimes B_0']\,\mathrm{Vec}(I_p), \end{aligned} \tag{5.24}$$

where

$$M = \frac{1}{n}\sum_{i=1}^{n}\{E\phi(|\boldsymbol{x}_i|^2) + E|\boldsymbol{x}_i|^2\phi'(|\boldsymbol{x}_i|^2)\}E\boldsymbol{z}_i\boldsymbol{z}_i' - I_p. \tag{5.25}$$

It is easy to show that M is positive definite. Hence,

$$\begin{aligned}[\mathrm{Vec}(B_0)]'\boldsymbol{a} &= [\mathrm{Vec}(B_0)]'[M \otimes B_0']\,\mathrm{Vec}(I_p) \\ &= [\mathrm{Vec}(I_p)]'[I_p \otimes B_0][M \otimes B_0']\,\mathrm{Vec}(I_p) \\ &\geq 0 \qquad (\text{since } [I_p \otimes B_0][M \otimes B_0'] \text{ is n.n.d.}).\end{aligned}$$

Hence Theorem 5.1 follows by using Lemma 5.1.

6. Consistency of the estimate

LEMMA 6.1.

$$\mathrm{tr}(\Sigma^{-1}T) = p.$$

PROOF. From likelihood equations we have,

$$\mathrm{tr}(\Sigma^{-1}(\Sigma - T)\Sigma^{-1}) = 0 \quad \text{and} \tag{6.1}$$

$$\Sigma^{-1}(\Sigma - T)\Sigma^{-1}B = 0. \tag{6.2}$$

From (6.2) we can write,

$$\begin{aligned}&(I_p - \Sigma^{-1}T)\Sigma^{-1}BB' = 0, \qquad I_p - \Sigma^{-1}T - \sigma^2\Sigma^{-1} + \sigma^2\Sigma^{-1}T\Sigma^{-1} = 0, \\ &p - \mathrm{tr}(\Sigma^{-1}T) - \sigma^2\,\mathrm{tr}(\Sigma^{-1}(\Sigma - T)\Sigma^{-1}) = 0 \quad \text{or} \quad \mathrm{tr}(\Sigma^{-1}T) = p.\end{aligned}$$

LEMMA 6.2.

$$\boldsymbol{x}'\Sigma^{-1}\boldsymbol{x} \geq \frac{(\boldsymbol{z}'\boldsymbol{x})^2}{l} \quad \textit{where}$$

l = *smallest eigenvalue of* Σ,

$\boldsymbol{z}$ = *corresponding unit eigen vector.*

PROOF.

$$\begin{aligned}\boldsymbol{x}'\Sigma^{-1}\boldsymbol{x} = \mathrm{tr}\,\Sigma^{-1}\boldsymbol{x}\boldsymbol{x}' &\geq \frac{(\boldsymbol{x}'\boldsymbol{x})}{l} \qquad (\text{by the Von-Neumann (1937) inequality}) \\ &= \frac{(\boldsymbol{x}'\boldsymbol{x})(\boldsymbol{z}'\boldsymbol{z})}{l} \\ &\geq \frac{(\boldsymbol{z}'\boldsymbol{x})^2}{l} \qquad (\text{by the C.S. inequality}).\end{aligned}$$

THEOREM 6.1. Let $\boldsymbol{x}_1, \dots, \boldsymbol{x}_n$ be i.i.d. with a common distribution P. Let $\Sigma_n = B_nB_n' + \sigma_n^2 I_p$, where σ_n^2 and B_n are the solutions of the likelihood equations based on $\boldsymbol{x}_1, \dots, \boldsymbol{x}_n$ with P_n being the emperical distribution. Then $\lim_{n\to\infty}\Sigma_n = \Sigma$ a.s.

PROOF. Wolfowitz (1954) established that for $S \subseteq R^m$, the $\lim_{n\to\infty} P_n(S) = P(S)$ uniformly in S with probability one. It is easy to verify that P_n satisfies the assumption (e) given in Section 2. We will prove the consistency of Σ_n by using Huber's (1967) result. It is easy to verify the conditions (B-1), (B-2′) and (B-3) of Section 3 in Huber. According to Huber's Theorem 2, it is enough to show the existence of a compact set $K \subseteq \Theta$ such that with probability one, the sequence Σ_n ultimately stays in K. It is very difficult to check condition (B-4) of Huber, which would entail the desired result. It suffices to prove the existence of a finite constant B such that with probability one,

$$\limsup_{n\to\infty} \|\Sigma_n^{-1}\| < B \quad \text{or} \tag{6.3}$$

$$\text{equivalently } l_n^{-1} \le B' < \infty, \tag{6.4}$$

where l_n = smallest eigenvalue of Σ_n.

$$\text{Suppose that (6.4) is not true, i.e. } \frac{1}{l_n} > B'; \tag{6.5}$$

then we have from Lemma 6.1

$$\begin{aligned}
p &= \operatorname{tr}(\Sigma_n^{-1}T) \\
&= \frac{1}{n}\sum_{i=1}^{n} \phi(\boldsymbol{x}_i'\Sigma_n^{-1}\boldsymbol{x}_i) \\
&= E_{P_n}\phi(\boldsymbol{x}'\Sigma^{-1}\boldsymbol{x}) \\
&= \int_{\{\boldsymbol{x}:|\boldsymbol{z}'\boldsymbol{x}|\le c\}} \phi(\boldsymbol{x}'\Sigma^{-1}\boldsymbol{x})dP_n + \int_{\{\boldsymbol{x}:|\boldsymbol{z}'\boldsymbol{x}|>c\}} \phi(\boldsymbol{x}'\Sigma^{-1}\boldsymbol{x})dP_n \\
&\qquad\qquad \text{(where } c > 0 \text{ is a constant)} \\
&> P_n\{\boldsymbol{x}: |\boldsymbol{z}'\boldsymbol{x}| > c\}\phi(c^2B') \\
&\qquad \text{(by Lemmas 6.2 and (6.5), and Assumption (c) of Section 2)} \\
&> \left(\frac{p}{K} + \frac{a}{2}\right)(K-b) \quad \text{(by Assumption (e) of Section 2)} \\
&\qquad\qquad \text{for some } b > 0.
\end{aligned} \tag{6.6}$$

There exist $b > 0$ such that the expression in (6.6) is greater than p. Hence, we get a contradiction.

7. Convergence of iteration in Section 4 through simulation

In Section 4 we have introduced an algorithm for the estimation of a dispersion matrix, which involves an iterative procedure. In this section we give a numerical example to show that the iteration converges after some steps.

We give some simulation results as follows:

$S^{(0)}$ = Initial Covariance matrix,

$\hat{S}$ = Final Covariance matrix after iteration.

Table 7.1.

Number of simulations: 1001

$n = 10$, $p = 2$, $q = 1$, $m = 4$, $\epsilon = 0.10$

$$S^{(0)} = \begin{bmatrix} 0.366 & 0.032 \\ & 0.175 \end{bmatrix}, \qquad \hat{S} = \begin{bmatrix} 0.099 & 0.023 \\ & 0.015 \end{bmatrix}$$

Iteration No.	Epsilon
1	0.3116
2	0.0571

Maximum of Iteration No. in the remaining 1000 simulations = 4.

Table 7.2.

Number of simulations: 1001

$n = 10$, $p = 3$, $q = 1$, $m = 4$, $\epsilon = 0.10$

$$S^{(0)} = \begin{bmatrix} 0.170 & 0.006 & 0.355 \\ & 0.509 & 0.517 \\ & & 2.823 \end{bmatrix}, \qquad \hat{S} = \begin{bmatrix} 0.323 & 0.007 & -0.026 \\ & 0.063 & -0.001 \\ & & 0.065 \end{bmatrix}$$

Iteration No.	Epsilon
1	2.8956
2	0.1472
3	0.0931

Maximum of Iteration No. in the remaining 1000 simulations = 8.

Table 7.3.

Number of simulations: 1001

$n = 10$, $p = 3$, $q = 2$, $m = 4$, $\epsilon = 0.10$

$$S^{(0)} = \begin{bmatrix} 0.457 & 0.410 & 1.987 \\ & 1.278 & 3.418 \\ & & 13.369 \end{bmatrix}, \qquad \hat{S} = \begin{bmatrix} 2.040 & 0.038 & 0.010 \\ & 0.083 & 0.032 \\ & & 0.043 \end{bmatrix}$$

Iteration No.	Epsilon
1	14.3252
2	0.4838
3	0.3385
4	0.2943
5	0.2310
6	0.1626
7	0.1070
8	0.0675

Maximum of Iteration No. in the remaining 1000 simulations = 17.

Table 7.4.

Number of simulations: 1001

$n = 10$, $p = 4$, $q = 1$, $m = 4$, $\epsilon = 0.10$

$$S^{(0)} = \begin{bmatrix} 0.326 & 0.023 & 0.597 & 1.004 \\ & 0.394 & 0.705 & 0.939 \\ & & 4.324 & 5.939 \\ & & & 8.667 \end{bmatrix}, \quad \hat{S} = \begin{bmatrix} 0.400 & 0.143 & -0.808 & -0.231 \\ & 0.082 & -0.308 & -0.089 \\ & & 1.757 & 0.497 \\ & & & 0.170 \end{bmatrix}$$

Iteration No.	Epsilon
1	12.7429
2	0.3709
3	0.4288
4	0.3897
5	0.2868
6	0.1655
7	0.0711

Maximum of Iteration No. in the remaining 1000 simulations = 19.

Table 7.5.

Number of simulations: 1001

$n = 10$, $p = 4$, $q = 2$, $m = 4$, $\epsilon = 0.10$

$$S^{(0)} = \begin{bmatrix} 0.878 & 0.810 & 3.693 & 2.047 \\ & 1.449 & 5.058 & 1.999 \\ & & 21.801 & 10.062 \\ & & & 8.123 \end{bmatrix}, \quad \hat{S} = \begin{bmatrix} 0.991 & 0.163 & -1.317 & -0.397 \\ & 0.068 & -0.210 & -0.021 \\ & & 1.872 & 0.634 \\ & & & 0.479 \end{bmatrix}$$

Iteration No.	Epsilon
1	28.1115
2	0.8442
3	0.6229
4	0.3742
5	0.1961
6	0.0845

Maximum of Iteration No. in the remaining 1000 simulations = 22.

Table 7.6.

Number of simulations: 1001

$n = 10,\ p = 4,\ q = 3,\ m = 4,\ \epsilon = 0.10$

$$S^{(0)} = \begin{bmatrix} 0.979 & 0.956 & 3.959 & 2.319 \\ & 1.649 & 5.408 & 2.375 \\ & & 22.506 & 10.377 \\ & & & 8.893 \end{bmatrix}, \quad \hat{S} = \begin{bmatrix} 0.634 & 0.118 & -0.753 & -0.115 \\ & 0.106 & -0.227 & -0.027 \\ & & 1.036 & 0.342 \\ & & & 0.494 \end{bmatrix}$$

Iteration No.	Epsilon
1	29.2105
2	0.5352
3	0.2494
4	0.1375
5	0.0821

Maximum of Iteration No. in the remaining 1000 simulations $= 23$.

Acknowledgements

The author would like to thank the referees for very careful reading of the original version, which resulted in substantial improvements.

References

Chen, H.-J., Gnanadesikan, R. and Kettenring, J. R. (1974). Statistical methods for grouping corporations, *Sankhyā Ser. B*, **36**, 1–28.

Devlin, S. J., Gnanadesikan, R. and Kettenring, J. R. (1975). Robust estimation and outlier detection with correlation coefficients, *Biometrika*, **62**, 531–545.

Gnanadesikan, R. and Kettenring, J. R. (1972). Robust estimates, residuals and outlier detection with multiresponse data, *Biometrics*, **28**, 81–124.

Huber, P. J. (1967). The behavior of maximum likelihood estimates under non-standard conditions, *Proc. Fifth Berkeley Symp. on Math. Statist. Prob.*, Vol. 1, 222–233, Univ. of California Press, Berkeley.

Huber, P. J. (1977*a*). Robust covariances, *Statistical Decision Theory and Related Topics II* (eds. Shanti S. Gupta and David S. Moore), 165–191, Academic Press, New York.

Huber, P. J. (1977*b*). *Robust Statistical Procedures*, Society for Industrial Applied Mathematics, Philadelphia.

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

Jöreskog, K. G. (1967). Some contributions to maximum likelihood factor analysis, *Psychometrika*, **32**, 444–482.

Kelker, D. (1970). Distribution theory of spherical distribution and a location scale parameter generalization, *Sankhyā Ser. A*, **44**, 419–430.

Lawley, D. N. and Maxwell, A. E. (1963). *Factor Analysis as a Statistical Method*, Butterworths, London.

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

Mosteller, F. and Tukey, J. W. (1977). *Data Analysis and Regression*, Addison-Wesley, Reading, Massachusetts.

Rao, C. R. (1982). Matrix derivatives and its application to statistics, *Encyclopedia of Statistical Science*, Vol. 5 (eds. Samuel Kotz and Norman L. Johnson), 320–325, Wiley, New York.

Rao, C. R. (1983). *Linear Statistical Inference and Its Applications*, Wiley, New York.

Von-Neumann, J. (1937). Some matrix inequalities and metrization of metric space, *Tomsk Univ. Rev.*, **1**, 286–300.

Wolfowitz, J. (1954). Generalization of the theorem of Glivenko-Cantelli, *Ann. Math. Statist.*, **25**, 131–138.

Zhao, L. C., Krishnaiah, P. R. and Bai, Z. D. (1986*a*). On detection of the number of signals in presence of white noise, *J. Multivariate Anal.*, **20**, 1–25.

Zhao, L. C., Krishnaiah, P. R. and Bai, Z. D. (1986*b*). On detection of the number of signals when the noise covariance matrix is arbitrary, *J. Multivariate Anal.*, **20**, 26–49.