IMPROVED ESTIMATION OF A PATTERNED COVARIANCE MATRIX

BY

DIPAK K. DEY and ALAN E. GELFAND

TECHNICAL REPORT NO. 431
JULY 18, 1990

PREPARED UNDER CONTRACT
NO0014-89-J-1627 (NR-042-267)
FOR THE OFFICE OF NAVAL RESEARCH

Reproduction in Whole or in Part is Permitted for any purpose of the United States Government

Approved for public release; distribution unlimited.

DEPARTMENT OF STATISTICS
STANFORD UNIVERSITY
STANFORD, CALIFORNIA
IMPROVED ESTIMATION OF A PATTERNED COVARIANCE MATRIX

BY

DIPAK K. DEY and ALAN E. GELFAND

TECHNICAL REPORT NO. 431
JULY 18, 1990

Prepared Under Contract
N00014-89-J-1627 (NR-042-267)
For the Office of Naval Research

Herbert Solomon, Project Director

Reproduction in Whole or in Part is Permitted for any purpose of the United States Government

Approved for public release; distribution unlimited.

DEPARTMENT OF STATISTICS
STANFORD UNIVERSITY
STANFORD, CALIFORNIA
ABSTRACT

Suppose a random vector \( X \) has a multinormal distribution with covariance matrix \( \Sigma \) of the form \( \Sigma = \sum_{i=1}^{k} \Theta_i M_i \), where the \( M_i \)'s form a known complete orthogonal set and \( \Theta_i \)'s are the distinct unknown eigenvalues of \( \Sigma \). The problem of estimation of \( \Sigma \) is considered under several plausible loss functions. The approach is to establish a duality relationship: estimation of the patterned covariance matrix \( \Sigma \) is dual to simultaneous estimation of scale parameters of independent chi-square distributions. This duality allows simple estimators which, for example, improved upon the MLE of \( \Sigma \). It also allows improved estimation of \( \text{tr} \Sigma \). Examples are given in the case when \( \Sigma \) has equicorrelated structure.

1. INTRODUCTION AND SUMMARY

Recently there has been considerable interest in the estimation of the covariance matrix of a multivariate normal distribution. This problem is addressed extensively in Stein (1975, 1977), Olkin and Selliah (1977), Haff (1977, 1979, 1982), and Dey and Srinivasan (1985, 1986) under plausible loss functions. However, there is no work of our kind available in the literature when the covariance matrix has an assumed structure.

Suppose a random vector \( X \) has a multinormal distribution with mean zero and covariance matrix \( \Sigma \), which has the form

\[
\Sigma = \sum_{i=1}^{k} \Theta_i M_i \tag{1.1}
\]

where the \( \Theta_i \)'s are the distinct but unknown eigenvalues of \( \Sigma \) and the \( M_i \)'s are a known complete orthogonal set of projection matrices. Such a structure for \( \Sigma \) arises in many practical
situations. A familiar example is the equicorrelated case, that is, \( \hat{\Sigma} = \sigma^2 [(1-\rho)I + \rho J] \) where \( I \) is the identity matrix and \( J \) is a \( p \times p \) matrix of 1's. This is often referred to as an intraclass correlation structure. More generally, patterned covariance matrices of the form (1.1) arise naturally in variance component models. See Albert (1976) for details.

From the classical viewpoint one would estimate \( \hat{\Sigma} \) by obtaining the maximum likelihood estimates of the \( \theta_i \)'s using the normality of \( X \). In this paper, however, we take a decision theoretic approach for the estimation of \( \hat{\Sigma} \) using the following loss structures:

\[
L_q(\hat{\Sigma}, \Sigma) = \text{tr}(\hat{\Sigma} - \Sigma)^2 \tag{1.2}
\]

and

\[
L_e(\hat{\Sigma}, \Sigma) = \text{tr} \hat{\Sigma}^{-1} - \log |\hat{\Sigma}^{-1}| - p \tag{1.3}
\]

The loss (1.2) is the usual extension of squared error loss (SEL) and the loss (1.3) is based on entropy measure of distance. Under these losses the MLE is inadmissible and substantial improvement is available (see Table 1). We may show that for estimators of the form \( \sum_{i=1}^{k} \hat{\theta}_i M_i \), these losses become, respectively,

\[
L(\hat{\theta}, \theta) = \sum_{i=1}^{k} p_i (\hat{\theta}_i - \theta_i)^2 \tag{1.4}
\]

\[
L(\hat{\theta}, \theta) = \sum_{i=1}^{k} p_i [\hat{\theta}_i / \theta_i - \log (\hat{\theta}_i / \theta_i) - 1] \tag{1.5}
\]

where \( p_i = \text{rank}(M_i) \), \( \theta = (\theta_1, \ldots, \theta_k) \) and \( \sum_{i=1}^{k} p_i = p \).

In addition, we note the following result essentially given in Albert (1976).
Theorem 1.1. Suppose \( X \sim N(0, \Sigma) \). Define \( Q_i = X'M_iX \), where
\[
I = \sum_{i=1}^{k} M_i, \quad M_i \text{ being orthogonal projection matrices free from}
\]
\( \theta_i \)'s, having rank\( (M_i) = p_i, i=1, \ldots, k (\leq p) \). Then a necessary and
sufficient condition for \( Q_i \sim \theta_i X_i^2, i=1, \ldots, k, \) and
\[
\begin{align*}
\sum_{i=1}^{k} Q_i = \Sigma \theta_i M_i.
\end{align*}
\]
(2) \( Q_i \)'s mutually independent is \( \frac{\Sigma}{\sum_i} \theta_i M_i \).

The equivalence of (1.2) and (1.4) and of (1.3) and (1.5) for
\[
k \text{estimators of the form} \ \Sigma \hat{\theta}_i M_i \text{ along with Theorem 1.1}
\]
establishes the following duality. Estimation of the patterned
covariance matrix \( \frac{\Sigma}{\Sigma} \theta_i M_i \) in (1.1) under loss (1.2) ((1.3)) is dual to
simultaneous estimation of the scale parameters of independent
chi-square random variables under loss (1.4) ((1.5)).

It is to be noted that in the decomposition of \( \frac{\Sigma}{\Sigma} \theta_i M_i \),
\( \theta_i \)'s are the distinct eigenvalues of \( \frac{\Sigma}{\Sigma} \theta_i M_i \) with multiplicity \( p_i \).
Thus, for example, in the equicorrelated model
\[
\begin{align*}
\frac{\Sigma}{\Sigma} &= \sigma^2 \mathbf{I} + \rho \mathbf{J}, \quad \theta_1 = \sigma^2 (1 - \rho), \quad \theta_2 = \sigma^2 (1 + (p-1)\rho),
\end{align*}
\]
\[
M_1 = \mathbf{I} - \rho \mathbf{J}, \quad M_2 = \rho \mathbf{J}, \quad p_1 = p - 1, \quad \text{and} \quad p_2 = 1.
\]

In Section 2, we study the estimation of \( \frac{\Sigma}{\Sigma} \) under loss (1.2)
and also tr\( \frac{\Sigma}{\Sigma} \) under SEL. We illustrate for \( \frac{\Sigma}{\Sigma} \) with equicorrelated
structure. In this case, improved estimation of \( \sigma^2 \) is discussed.
Additionally, improved estimation of \( \rho \sigma^2 \) is also considered.
Section 3 is devoted to estimation of $\hat{\theta}$ under loss (1.3). Finally, in Section 4, some encouraging numerical results are given for the equicorrelated model.

2. IMPROVED ESTIMATION UNDER $L_q$ LOSS

2.1. Estimation of $\hat{\theta}$

Here we assume the conditions of Theorem 1.1 and generalize slightly the above discussion to estimation of $\hat{\theta} = \Sigma_{i=1}^{k} \theta_i M_i$ under the loss

$$L_q(\hat{\theta}, \theta) = \text{tr}(\hat{\theta} - \theta)^2,$$

which, using estimators of the form $\Sigma_{i=1}^{k} \hat{\theta}_i M_i$, is dual to the estimation of $\hat{\theta} = (\hat{\theta}_1, \ldots, \hat{\theta}_k)$ under the loss

$$L(\hat{\theta}, \theta) = \Sigma_{i=1}^{k} p_i (\hat{\theta}_i - \theta_i)^2.$$  \hfill (2.2)

For the estimation of $\hat{\theta}$, $s = 1$; for the estimation of the precision matrix $\hat{\theta}^{-1}$, $s = -1$. Let $Q = (Q_1, \ldots, Q_k)$ where $Q_i \sim \chi^2_{2} - \chi^2_{p_i}$ and are independent. Suppose $\delta^0(Q)$ is an estimator of $\theta$ given componentwise as $\delta^0_i(Q) = a_i Q_i^s$, $i = 1, \ldots, k$. For example, $a_i = p_i - s$ gives the MLE of $\theta_i$.

Now define

1. $r_{i,s} = E(Q_i^s | \theta_i = 1) / E(Q_i^s | \theta_i = 1) = 2^{-1/2} \frac{\Gamma((p_i+2\alpha)/2)}{\Gamma((p_i+2\beta)/2)}.$

2. For $\alpha = (\alpha_1, \ldots, \alpha_k)$, $\nu = \prod_{i=1}^{k} E(Q_i^s | \theta_i = 1) = \prod_{i=1}^{k} 2^{-1/2} \frac{\Gamma((p_i+2\alpha_i)/2)}{\Gamma((p_i+2\beta)/2)}.$

Note that $a_i = r_{i,s} s$ gives the best invariant estimator of $\theta_i$ under squared error loss.
Thus \( \hat{\theta}_M^S = \Sigma p_i^{-S} Q_{i}^S \) is the maximum likelihood estimator of \( \hat{\theta} \) and \( \hat{\theta}_0 = \Sigma r_{i, s, 2s} Q_{i}^S \) is the estimator obtained by combining the best invariant estimators of \( \theta_i^S \). The following lemma shows that \( \hat{\theta}_0^S \) dominates \( \hat{\theta}_M^S \) under the risk criterion.

**Lemma 2.1.** \( R(\hat{\theta}_0^S, \hat{\theta}^S) \leq R(\hat{\theta}_M^S, \hat{\theta}^S) \forall \hat{\theta} \).

**Proof.** Immediate from the duality (2.1), (2.2) and the fact that \( r_{i, s, 2s} Q_{i}^S \) dominates \( p_i^{-S} Q_{i}^S \) in estimating \( \theta_i^S \).

In view of Lemma 2.1, it is sufficient to find estimators which improve \( \hat{\theta}_0^S \). We have the following theorem whose proof is a special case of a result in Dey and Gelfand (1987).

**Theorem 2.1.** Consider the estimator \( \delta(Q) = (\delta_1(Q), \ldots, \delta_k(Q)) \) given componentwise as

\[
\delta_i(Q) = \delta_i^0(Q) = b(\Pi Q_{j})^{s/k}, \quad i = 1, \ldots, k (\geq 2),
\]

(2.4)

where \( \delta_i^0(Q) = r_{i, s, 2s} Q_{i}^S \). Then provided all expectations exist, \( \delta(Q) \) dominates \( \delta^0(Q) \) under loss (2.2) if

\[0 < b < 2^{\nu_s/k} \frac{d(1)}{\nu_s} \frac{d(2)}{\nu_s/k}\]

where \( d(1) = \min(p_i | d_i |) \) and \( d(2) = \max(p_i) \) with

\[d_i = r_{i, s, 2s} r_i, (k+1)s/k, s/k - 1, \quad i = 1, \ldots, k.\]

In view of Theorem 2.1, it follows that under the loss (2.1), an improved estimator of \( \hat{\theta}_i^S \) is given as

\[
\hat{\theta}_i^S = \Sigma \delta_i(Q) M_i = \hat{\theta}_0^S + b(\Pi Q_{j})^{s/k} \tilde{p}.
\]

(2.5)

**Remark 2.1.** Theorem 2.1 requires only that \( Q_{i} \) follow a distribution with \( \theta_i \) as scale parameter. In this setting Dey and Gelfand
(1987) offer more general results than Theorem 2.1 including estimators which provide maximum improvement along a ray determined by a specified vector \( \theta_0 \). These estimators shrink (expand) differently on each coordinate. Using the fact that the \( Q_i \) are distributed as multiples of chi-square random variables enables Klonecki and Zontek (1987) to provide necessary and sufficient conditions for the existence of an estimator of the form (2.4) to dominate \( \delta^0 \) given componentwise as

\[
\delta^0_i = a_i Q_i^s \quad \text{for any specified } a_i, \ i = 1, \ldots, k.
\]

**Example:** Suppose \( \hat{\eta} \) has equicorrelated structure. In this case \( \theta_1 = \sigma^2(1 - \rho), \sigma_2 = \sigma^2(1 + (p - 1)\rho) \), are the distinct eigenvalues of \( \hat{\eta} \). The best invariant estimate of \( \theta_1 \) is \( \delta^0_i(Q) = (p_i + 2)^{-1} Q_i \) with \( p_1 = p - 1 \) and \( p_2 = 1 \), a class of improved estimators of \( \theta = (\theta_1, \theta_2) \) is given componentwise as

\[
\delta_i(Q) = (p_i + 2)^{-1} Q_i + b \left( \prod_{j=1}^{2} Q_j \right)^{\frac{1}{2}}, \ i = 1, 2, \tag{2.6}
\]

where \( 0 < b < 2 \sqrt{\frac{d_1}{d_2}} / \sqrt{\frac{d_2}{d_1}} \) and the upper bound on \( b \) simplifies to

\[
2 \Gamma(p/2)/3(p-1)\Gamma((p+1)/2)\sqrt{\pi}.
\]

The corresponding improved estimator of \( \hat{\eta} \) has the simple form

\[
\hat{\eta} = \hat{\eta}_0 + b(Q_1 Q_2)^{1/2} I_p. \tag{2.7}
\]

**Remark 2.2.** Improved estimation of \( \theta^{-1} = (\theta_1^{-1}, \ldots, \theta_k^{-1}) \), hence of \( \theta^{-1} \), follows directly from Theorem 2.1. We only need the existence of appropriate reciprocal moments of the \( Q_i \). Unfortunately, in the equicorrelated case \( E Q_i^{-\alpha} \) does not exist for \( \alpha > 1 \) and, hence, our approach does not provide a dominating estimator.
2.2. Estimation of trace of \( \frac{1}{\theta} \)

Consider now estimation of the trace of \( \frac{1}{\theta} \) under the SEL given as

\[
L(a, \text{tr}\frac{1}{\theta}) = (a - \text{tr}\frac{1}{\theta})^2.
\] (2.8)

Since \( \text{tr}\frac{1}{\theta} = \sum_{i=1}^{k} p_i \theta_i \) our duality converts estimation of the trace to estimation of a linear combination of chi-square scale parameters.

The following theorem gives a class of admissible estimators of

\[
\sum_{i=1}^{k} \frac{1}{\theta_i} \text{if } \theta_i > 0.
\]

**Theorem 2.2.** If \( \alpha_i > 0 \), \( i = 1, \ldots, k \), known then under the loss

\[
(2.9) \quad \delta_\alpha(Q) = \frac{\sum_{i=1}^{k} \frac{1}{\alpha_i}}{p+2} \sum_{i=1}^{k} \frac{Q_i}{\alpha_i}
\]

is admissible for \( \sum_{i=1}^{k} \frac{1}{\theta_i} \).

**Proof.** Consider the subset of the parameter space

\[
C = \{ (\theta_1, \ldots, \theta_k) : \theta_i = \frac{\theta}{\alpha_i}, \alpha_i > 0, i = 1, \ldots, k \}.
\]

Then \( Q_i/\alpha_i \sim \theta X_i^2, i = 1, \ldots, k \). Thus, on \( C \), \( \sum_{i=1}^{k} Q_i/\alpha_i \) is sufficient for \( \theta \) and \( \sum_{i=1}^{k} Q_i/\alpha_i \sim \theta X^2 \). Thus, by a theorem of Karlin (1958),

\[
\sum_{i=1}^{k} \frac{Q_i}{\alpha_i} \quad \text{is admissible for } \theta \text{ under SEL.}
\]

Then on \( C \), \( \delta_\alpha(Q) \)

\[
= \left\{ \left( \sum_{i=1}^{k} \frac{1}{\theta_i} \right) \sum_{i=1}^{k} \frac{Q_i}{\alpha_i} \right\}/(p+2) \text{ is admissible for } \sum_{i=1}^{k} \frac{1}{\theta_i} \theta = \sum_{i=1}^{k} \frac{1}{\theta_i} \theta.
\]

Suppose \( \delta_\alpha(Q) \) is inadmissible for \( \sum_{i=1}^{k} \frac{1}{\theta_i} \), Then there exists \( \delta_\alpha^*(Q) \)

which dominates \( \delta_\alpha(Q) \). But \( \delta_\alpha(Q) \) being admissible on \( C \) implies

\[
\delta_\alpha^*(Q) = \delta_\alpha(Q) \text{ a.s.}
\]
Note that since \( p_i > 0 \) in the expression for \( \text{tr}^+ \) Theorem 2.2 applies.

Remark 2.3. Consider the equicorrelated structure. As a special case, \( \alpha_i = 1 \) gives \( \frac{\mathbf{X}'\mathbf{X}}{p+2} \) admissible for trace \( \frac{1}{2} \) and hence \( \mathbf{X}'\mathbf{X}/(p+2) \) admissible for \( \sigma^2 \). Similarly \( \alpha_i = p_i/(p_i+2) \) implies \( \frac{\sum_{i=1}^{k} p_i^2/(p_i+2)}{p+2} \sum_{i=1}^{k} p_i Q_i/(p_i+2) \) is admissible for \( \sigma^2 \), i.e., an appropriate linear combination of the componentwise best invariant estimator is admissible.

Now we will demonstrate a general method for improving on a linear estimator of a linear combination. The improved estimators are nonlinear, and may shrink or expand the given linear estimator. Work of Das Gupta (1986), Dey and Gelfand (1987), and Klonecki and Zontek (1987) is relevant here. A general result is:

**Theorem 2.3.** Provided expectations exist, an estimator of \( \text{tr}^+ \) of the form \( \delta_0 = \frac{\sum_{i=1}^{k} \xi_i Q_i}{\xi_i = 1} \) yields \( \frac{\sum_{i=1}^{k} Q_i}{\frac{\sum_{i=1}^{k} Q_i}{\sum_{i=1}^{k} Q_i}} \) the MLE, which is also UMVUE) is dominated by

\[
\delta_{r,c} = \delta_0 + c \sum_{j=1}^{k} \frac{r_j}{Q_j}
\]

(2.10)

where \( r_j > 0, \sum_{j=1}^{k} r_j = 1 \), if and only if either

(i) \( d(1) > 0, r_i = 0 \) if \( d_i = 0 \) and \( c > 0 \) sufficiently small,

(ii) \( d(k) < 0, r_i = 0 \) if \( d_i = 0 \) and \( c < 0 \) sufficiently large,

where \( d_i = (1-\xi_i)p_i/2 - \xi_i r_i, i = 1, \ldots, k \) and \( d(1) = \min d_i, d(k) = \max d_i \).

**Proof.** The risk difference between (2.10) and \( \delta_0 \) is

\[
\Delta(\theta) = \sum_{j=1}^{k} \frac{2r_j}{2r} \prod_{\theta_j} \frac{r_j}{\prod_{\theta_j}} - \sum_{j=1}^{k} \frac{r_j}{\prod_{\theta_j}} \sum_{j=1}^{k} \theta_j d_i
\]
Suppose (i) holds (the proof for (ii) is similar) and that all
j such that \( r_j = d_j = 0 \) have been deleted in \( \Delta(\theta) \). Then
\[
\Delta(\theta) = \sum_{i=1}^{k} 2r_i \prod_{i=1}^{k} \theta_i \left[ c - \sum_{j=1}^{r} \frac{\theta_j d_j}{2r_j} \prod_{j=1}^{r} \theta_j \right].
\]
But \( \prod_{j=1}^{r} \frac{d_j}{r_j} \geq \prod_{j=1}^{r} \frac{d_j}{r_j} \). Hence, \( \Delta(\theta) \leq 0, \forall \theta \), if c is positive
and sufficiently small.

**Remark 2.4.** In particular, when \( l_i = 1 \) any set of nonnegative
\( r_i \) such that \( \sum_{i} r_i = 1 \) and at least two \( r_i \) differ from zero will work.
Here \( c < 0 \) so that the dominating estimator is a shrinker. Since
\( \text{tr} \hat{\Sigma} > 0 \), \( \hat{\delta}^+ = \max(\hat{\delta}, 0) \) will dominate \( \hat{\delta} \) (using a lemma of Klotz,

**Example.** Again consider the equicorrelated structure. Clearly
\( \text{tr} \hat{\Sigma} = p \sigma^2 \). Thus, using (2.10), we can explicitly dominate \( X'X \) in esti-
mating \( p \sigma^2 \), hence \( X'X/p \) in estimating \( \sigma^2 \) by nonlinear estimators. For
instance, the estimator
\[
\hat{\delta}^* = \frac{X'X}{p} + b(Q_1Q_2)^{\frac{1}{2}}
\]  
(2.11)
dominates the MLE under loss function \((\hat{\delta} - \sigma^2)^2\) if
\[-4 \Gamma(p/2)/p \Gamma(p/2 + 1) \sqrt{p} < b < 0.\]

**Remark 2.5.** If we attempt to apply Theorem 2.3 to linear
estimators of the form (2.9), we will discover that all \( d_i \)'s are
equal to zero. In particular, in the above example, we cannot
dominate \( X'X/(p+2) \) in estimating \( \sigma^2 \).

Continuing with our example, suppose \( p \geq 0 \) and consider estimation
of \( p \sigma^2 \) which may be viewed as variance component. (See Gelfand and Dey
(1988) for more detailed discussion of improved estimation of variance components.) Since $\rho\sigma^2 = (\theta_2 - \theta_1)/p$, we consider the estimator
\[ \delta^0 = p^{-1}(a_2Q_2 - a_1Q_1) \]
where \(a_i = (p_i + 2\epsilon_i)^{-1}, i = 1,2\), with \(0 \leq \epsilon_i \leq 1\). For example, taking \(\epsilon_i = 0, i = 1,2\), \(\delta^0\) becomes the MLE of \(\rho\sigma^2\) and taking \(\epsilon_i = 1, \delta^0\) is formed from the best invariant estimator of \(\theta_i\), under SEL, \(i = 1,2\). A class of improved estimators of \(\rho\sigma^2\), is given as
\[ \delta = \delta^0 + b(a_1Q_1)^r(a_2Q_2)^{1-r}, \] (2.12)
using Theorem 2.3, provided \(r > 0\) can be chosen such that either (i) \(r < \min(\epsilon_1, 1 - \epsilon_2)\) whence \(b\) must be positive and sufficiently small or (ii) \(r > \max(\epsilon_1, 1 - \epsilon_2)\) whence \(b\) must be positive and small. In fact, we would use \(\delta^+\).

Remark 2.6. Note that Theorem 2.2 does not provide admissible estimators of \(\rho\sigma^2\) since \(l_i < 0\).

3. IMPROVED ESTIMATION UNDER L\(_0\) LOSS

In this section we consider the estimation of patterned \(\frac{1}{\lambda}\) under loss (1.3). Using the aforementioned duality for estimators of the form \(\sum_{i=1}^{k} \theta_i M_i\), we convert this problem to simultaneous estimation of the eigenvalues \(\theta = (\theta_1, \ldots, \theta_k)\) of \(\frac{1}{\lambda}\) under the loss (1.5). In fact, our results can be extended to the estimation of \(\theta^S = (\theta_1^S, \ldots, \theta_k^S)\) and hence the estimation of \(\frac{1}{\lambda}^S\). However, it is not clear how to apply loss (1.3) in estimating \(tr\frac{1}{\lambda}\) which is not a scale parameter.

Our approach is that of Berger (1980) and Dey, Ghosh and Srinivasan (1987). Assume \(Q_i\) are independent Gamma\((\alpha_i, \eta_i)\) \((\alpha_i > 0, \eta_i > 0)\) random variables, having density
\[ f(Q_i | \eta_i) = \eta_i^{\alpha_i - 1} Q_i^{-\alpha_i} e^{-Q_i \eta_i / \Gamma(\alpha_i)}. \] (3.1)

In our case, we have \( \alpha_i = \frac{p_i}{2}, \eta_i = (2\theta_i)^{-1} \), whence the loss (1.5) corresponds to simultaneous estimation of the \( \eta_i^{-1} \); that is,

\[ L(\delta, \eta^{-1}) = \sum_{i=1}^{k} p_i [\delta_n \eta_i - \log(\delta_n \eta_i) - 1]. \]

Since the MLE of \( \theta_i \) is \( \delta_i^0(Q) = Q_i/p_i \), the MLE and the unbiased estimator of \( \theta_i = \eta_i^{-1}/2 \) is \( \delta_i^0(Q) = Q_i/2\alpha_i \), \( i = 1, \ldots, k \). From Dey, Ghosh and Srinivasan (1987), it follows that \( \delta_i^0(Q) = (\delta_i^0(Q), \ldots, \delta_k^0(Q)) \) is admissible for \( k = 2 \) if \( \min(\alpha_1, \alpha_2) > 4 \). To seek a dominating estimator, we require \( k > 2 \).

Now assuming the conditions in Lemma 1 of Berger (1980), it follows that if

\[ \delta(Q) = \delta_i^0(Q) + \phi(Q) \] (3.2)

is a rival estimator, the risk difference is

\[ \Delta(\delta) = R(\delta, \theta) - R(\delta_i^0, \theta) = E_{\theta} \Delta_0(Q), \]

where \( \Delta_0(Q) \) is the unbiased estimate of the risk difference given as

\[ \Delta_0(Q) = \sum_{i=1}^{k} p_i [\phi_i^1(Q) + (\alpha_i - 1) \phi_i(Q)/Q_i - \log(1 + \alpha_i \psi_i(Q)/Q_i)]. \]

with \( \phi_i^1(Q) = \partial \phi_i(Q)/\partial Q_i \). Defining \( \phi_i(Q) = Q_i \psi_i(Q) \), one gets

\[ \Delta_0(Q) = \sum_{i=1}^{k} p_i [Q_i \phi_i^1(Q) + \alpha_i \psi_i(Q) - \log(1 + \alpha_i \psi_i(Q))]. \] (3.3)

In order to obtain an improved estimator \( \delta(Q) \), it is sufficient to find a solution \( \Delta_0(Q) \leq 0 \) with strict inequality for some set of \( Q \).
The following theorem gives a class of dominating shrinkage estimators.

Theorem 3.1. Suppose \( S = \sum \log^2(Q_i/2) \). Consider an estimator

\[
\delta(Q) = (\delta_1(Q), \ldots, \delta_k(Q))
\]
given componentwise as

\[
\delta_i(Q) = p_i^{-1}Q_i - Q_i \tau(S) \log(Q_i/2)/2(b+S), \quad i = 1, \ldots, k,
\]

with \( b > (5.76)(k-2)/p^* \), where \( p^* = \max p_i \) and \( \tau(S) \) is a function satisfying

(i) \( 0 < \tau(S) < 4.8(k-2)/p^* \)

(ii) \( \tau(S) \) in \( S \) and

(iii) \( E[\tau'(S)] < \infty \). \hspace{1cm} (3.5)

Then \( \delta(Q) \) dominates \( \delta^0(Q) \) for \( k \geq 3 \), in terms of risk.

Proof. The argument is similar to that of Theorem 3.1 of Dey, Ghosh and Srinivasan (1987).

Remark 3.1. Using Theorem 3.2 and 3.3 of Dey, Ghosh and Srinivasan (1987), adaptive estimators and trimmed shrinkage estimators of \( \theta \) can be obtained as well.

In concluding this section, we observe another illustration of our duality relationship. Improved estimation of patterned \( \hat{\theta} \) under the scale invariant loss

\[
L^2(\hat{\theta}, \Theta) = \text{tr}(\hat{\Sigma}^{-1} - I)^2
\]

using estimators of the form \( \sum_{i=1}^k \Theta_i M_i \), converts to simultaneous estimation of \( \Theta = (\Theta_1, \ldots, \Theta_k) \) under loss

\[
L(\hat{\Theta}, \Theta) = \sum_{i=1}^k p_i (\hat{\Theta}_i/\Theta_i - 1)^2.
\]

(3.7)
Again, Berger's approach yields a differential inequality (the only difference will be the presence of the weights, \( p_i \)) whose solution leads to dominating estimators similar to those in (3.4). Details are omitted.

4. NUMERICAL RESULTS

To study the performance of the MLE \( \hat{\theta}_M = \sum_{i=1}^{2} p_i^{-1} Q_i \),
\( \hat{\theta}_0 = \sum_{i=1}^{2} (p_i + 2)^{-1} Q_i \) and \( \hat{\xi} = \hat{\theta}_0 + b(Q_{02})^{\frac{1}{2}} I_p \), we calculate risks for different values of \( p \) and \( \rho \) in the equicorrelated structure. We took \( \sigma^2 = 1 \) and \( b = \Gamma(p/2)/3(p-1)\Gamma((p+1)/2)/\pi \), which is the midpoint of the allowable range. We then computed the percentage improvements for selected values of \( p \) and \( \rho \) (\( -(p-1)^{-1} \leq \rho \leq 1 \)). The improvements of \( \hat{\xi} \) over the MLE are substantial. While the percentage improvements in risk of \( \hat{\xi} \) over \( \hat{\theta}_0 \) are small, the simplicity of \( \hat{\xi} \) encourages its use.
TABLE 1

PERCENTAGE IMPROVEMENTS OVER $\hat{\Sigma}_M$ AND $\hat{\Sigma}_0$

\[
\begin{align*}
p & \quad \text{PI1} = \frac{R(\hat{\Sigma}_M) - R(\hat{\Sigma})}{R(\hat{\Sigma}_M)} \times 100 \\
\rho & \quad \text{PI2} = \frac{R(\hat{\Sigma}_0) - R(\hat{\Sigma})}{R(\hat{\Sigma}_0)} \times 100
\end{align*}
\]

\begin{tabular}{cccc}
\hline
$p$ & \multicolumn{2}{c}{-0.75} & \multicolumn{2}{c}{3.83} \\
-0.50 & 67.94 & 5.31 \\
-0.25 & 68.78 & 6.75 \\
0 & 68.92 & 6.35 \\
0.25 & 68.78 & 5.31 \\
0.50 & 68.44 & 3.83 \\
0.75 & 67.94 & 3.83 \\
\hline
$p = 6$ \\
0 & 49.25 & 3.11 \\
0.25 & 63.77 & 2.45 \\
0.50 & 66.42 & 1.50 \\
0.75 & 66.85 & 0.85 \\
\hline
$p = 10$ \\
0 & 43.50 & 1.87 \\
0.25 & 64.66 & 1.23 \\
0.50 & 66.50 & 0.70 \\
0.75 & 66.75 & 0.39 \\
\hline
\end{tabular}

ACKNOWLEDGMENT

The authors wish to acknowledge Len Kelly for performing the computations.

AMS 1980 Subject Classifications. Primary 62F10; secondary 62C99.
BIBLIOGRAPHY


# Improved Estimation of A Patterned Covariance Matrix

**Authors:** Dipak K. Dey and Alan E. Gelfand

**Performing Organization:**
Department of Statistics
Stanford University
Stanford, CA 94305

**Controlling Office:**
Office of Naval Research
Statistics & Probability Program Code 1111

**Report Date:**
July 18, 1990

**Number of Pages:**
19

**DISTRIBUTION STATEMENT:**
APPROVED FOR PUBLIC RELEASE: DISTRIBUTION UNLIMITED

**Keywords:**
- patterned covariance matrix;
- loss function;
- simultaneous estimation;
- equicorrelated model.

**Abstract:**
PLEASE SEE FOLLOWING PAGE.
20. ABSTRACT

Suppose a random vector $X$ has a multinormal distribution with covariance matrix $\Sigma$ of the form $\Sigma = \sum_{i=1}^{k} \theta_i M_i$, where the $M_i$'s form a known complete orthogonal set and $\theta_i$'s are the distinct unknown eigenvalues of $\Sigma$. The problem of estimation of $\Sigma$ is considered under several plausible loss functions. The approach is to establish a duality relationship: estimation of the patterned covariance matrix $\Sigma$ is dual to simultaneous estimation of scale parameters of independent chi-square distributions. This duality allows simple estimators which, for example, improved upon the MLE of $\Sigma$. It also allows improved estimation of $\text{tr}\Sigma$. Examples are given in the case when $\Sigma$ has equicorrelated structure.