NONNEGATIVE DEFINITENESS OF THE ESTIMATED DISPERSION MATRIX IN A MULTIVARIATE LINEAR MODEL

BY

FRIEDRICH PUKELESHEIM and GEORGE P. H. STYAN

TECHNICAL REPORT NO. 125
MAY 1978

PREPARED UNDER THE AUSPICES
OF
NATIONAL SCIENCE FOUNDATION GRANT
MPS 75-09450

Ingram Olkin, Project Director

DEPARTMENT OF STATISTICS
STANFORD UNIVERSITY
STANFORD, CALIFORNIA
NONNEGATIVE DEFINITENESS OF THE ESTIMATED DISPERSION MATRIX IN A MULTIVARIATE LINEAR MODEL

BY

FRIEDRICH PUKELSHEIM and GEORGE P.H. STYAN

TECHNICAL REPORT NO. 125

MAY 1978

PREPARED UNDER THE AUSPICES

OF

NATIONAL SCIENCE FOUNDATION GRANT

MPS 75-09450

Ingram Olkin, Project Director

This report is also issued as Technical Report No. 32. Partially supported by the Office of Naval Research Contract No. N00014-75-C-0442.

DEPARTMENT OF STATISTICS

STANFORD UNIVERSITY

STANFORD, CALIFORNIA
Nonnegative Definiteness of the Estimated Dispersion Matrix in a Multivariate Linear Model

by

Friedrich Pukelsheim
Universität Freiburg im Breisgau
and Stanford University

and

George P. H. Styan
McGill University

Summary.

Estimation is considered in a model where both the mean vector and the dispersion matrix have linear decompositions. It is shown that after an invariance reduction with respect to mean translation, MINQUE provides a nonnegative definite estimate of the dispersion matrix, when the decomposing matrices span a quadratic subspace of symmetric matrices. With normality, MINQUE is seen to equal the restricted maximum likelihood estimate and to be of uniformly minimum variance.

KEY WORDS: MINQUE, Noniterative solution of likelihood equations, Quadratic subspaces, REML, Special Jordan algebra, UMVU, Uniqueness of maximum likelihood estimate.
1. Introduction. Consider independent and identically distributed random \(\mathbb{R}^n\)-vectors \(y_{\pi\alpha} \in \mathbb{R}^n\), \(\alpha = 1, \ldots, N\), with common mean vector \(\Sigma^p_{\pi=1} b_{\pi} x_{\pi}\) and common dispersion matrix \(\Sigma^k_{k=1} t^t v_{k}\), where interest concentrates on estimating the vector \(t := (t_{1}, \ldots, t_{k})'\) of dispersion coefficients. Various procedures have been discussed in the literature. Among those are: (i) minimum norm unbiased quadratic invariant estimation (MINQUE, C.R. Rao [8, p. 302]), and, under normality, (ii) uniform minimum variance unbiased invariant estimation (UMVU, Seely [9]), and (iii) restricted (by invariance) maximum likelihood estimation (REML, Corbeil & Searle [2]). In this paper invariance is to be understood with respect to the group of all mean translations

\[
\{ y + \Sigma_{b_{1}} x_{1}, \ldots, b_{p} \} \in \mathbb{R}^p, \]

being \(\Sigma_{b_{1}}\) where \(\Sigma\) projects orthogonally onto the orthogonal complement of the space spanned by \(x_{1}, \ldots, x_{p}\); hence reduction by invariance yields the residual vectors \(\Sigma_{b_{1}} y_{\pi\alpha}\) with mean \(0\) and dispersion matrix \(\Sigma_{b_{1}} \Sigma_{b_{1}}\).

Our main result may be roughly summarized as follows: If estimates according to each of the three procedures above exist, then they coincide, and the common estimate \(t\) yields a nonnegative definite estimate \(\Sigma_{t} \Sigma_{b_{1}} \Sigma_{b_{1}}\) of the dispersion matrix in the invariance reduced model. This holds true for any finite sample size \(N \geq v := \text{rank } \Sigma\), in contrast to asymptotic results on consistency as \(N \to \infty\), cf., Anderson [1].

In Section 2, the invariance reduced model is discussed in a normal setting, and Section 3 is concerned with the linear model situation.
The vital assumption is the condition of Seely [9] that
\[ \mathbf{M}_1 \cdots \mathbf{M}_k \] span a \( k \)-dimensional quadratic subspace \( \mathcal{B} \) of symmetric
\( n \times n \) matrices. The subspace \( \mathcal{B} \) is quadratic if and only if \( \mathbf{A}^2 \in \mathcal{B} \)
whenever \( \mathbf{A} \in \mathcal{B} \), i.e., \( \mathcal{B} \) is closed under the multiplication
\[ \mathbf{A} \circ \mathbf{B} := \frac{1}{2} (\mathbf{A} \mathbf{B} + \mathbf{B} \mathbf{A}). \] Jensen [4] points out that the latter property makes
\( \mathcal{B} \) into a \( k \)-dimensional special Jordan algebra, and we shall adopt this
more informative terminology. For a discussion with no initial invariance
reduction, see Gnot, Klonecki & Zmyślony [3].

2. The Normal Model. We will use the isomorphism \( \text{vec} \) that maps a
matrix into a vector by ordering its entries lexicographically, see
Fukelsheim [7].

**Theorem 1.** Consider independent and identically normally distrib-
uted random \( \mathbb{R}^v \)-vectors \( \mathbf{z}_i \in \mathbb{R}^v \) with common mean \( \mathbf{\mu} \) and common dispersion
matrix \( \Sigma_k \mathbf{W}_k \), where \( N \geq v \). Assume that the \( k \) decomposing matrices
\( \mathbf{W}_k \) span a \( k \)-dimensional special Jordan algebra \( \mathcal{B} \). Define \( G \subseteq \mathbb{R}^k \) to
be the region of those values \( t \) of the dispersion parameter such that
\( \Sigma_k \mathbf{W}_k \) is positive definite, and assume \( G \neq \emptyset \). Then:

(a) The maximum likelihood estimator for \( t \in G \) is almost surely
equal to the uniform minimum variance unbiased estimator
\[ \hat{t} := (D' D)^{-1} D' \cdot \text{vec}_\Sigma, \] where \( D := [ \text{vec}_1 \cdots \text{vec}_k ] \), and
\[ \Sigma := \Sigma \sum_{i=1}^{k} \mathbf{z}_i / N. \]

(b) The estimated dispersion matrix \( \hat{\mathbf{W}} := \Sigma_k \mathbf{W}_k \) is nonnegative
definite; in fact, if the sample dispersion matrix \( \Sigma \) is positive definite,
so is \( \hat{\mathbf{W}} \).
Proof. (a) Since $G$ is open and connected it is a region, and its boundary $\partial G$ consists of those $\xi \in \mathbb{R}^k$ such that $\sum_{k=1}^K W_k$ is nonnegative definite and singular. The sample dispersion matrix $\Sigma_\Xi$ is almost surely positive definite. If $\xi$ tends to $\partial G$, or $\|\xi\|_\Xi$ tends to $\infty$, the likelihood function $L$ tends to zero [1, p. 5]. Since $L$ is positive in $G$ there exists a maximum in $G$, and no maximum lies on the boundary $\partial G$. Hence the maximum likelihood estimate is a solution of the likelihood equations

\begin{equation}
(1) \quad D' \Sigma_\Xi^{-1} D \hat{\xi} = D' \Sigma_\Xi^{-1} \text{vec} \Sigma_\Xi,
\end{equation}

where the matrix of fourth moments

\begin{equation}
(2) \quad \Sigma_\Xi = \Sigma_\Xi(\hat{\xi}) := (\sum_{k=1}^K W_k) \otimes (\sum_{k=1}^K W_k).
\end{equation}

If $\Sigma_\Xi$ in (1) is put equal to $\Sigma_\Xi(\xi_0)$ for some given $\xi_0 \in G$, then (1) is a set of weighted normal equations, cf. [7, p. 628], and hence yields a minimum variance unbiased estimator for the vector parameter $\xi$. Since the matrices $W_k$ span a special Jordan algebra, there exists an almost surely unique uniform minimum variance unbiased invariant estimator which does not depend on the choice of $\xi_0 \in G$. Thus

\begin{equation}
(3) \quad \hat{\xi} = (D'D)^{-1} D' \text{vec} \Sigma_\Xi,
\end{equation}

since $G \neq \emptyset$ implies the existence of a nonsingular matrix $\Xi \in \mathcal{B}$, and so $\Xi^{-1} \in \mathcal{B}$ and $\Xi^{-1} = \Xi \circ \Xi^{-1} \in \mathcal{B}$; the matrix $\Sigma_\Xi$ in (1) may, therefore, be set equal to $I_{\Xi^2} = \Xi \otimes \Xi$. 

4
(b) As a linear operator on the space of symmetric matrices, $\hat{w}_S$ is surjective and hence open, and so if for some positive definite matrix $S_{\geq 0}$ the value $\hat{w}_S(S_{\geq 0}) \not\subset G$, the same is true for an open neighbourhood of $S_{\geq 0}$, i.e., for a set of positive Lebesgue measure. This contradicts part (a) that $\hat{w}_S$ maps into $G$ almost surely. For a singular sample dispersion matrix $S$, consider the limit $S + \varepsilon I_{\mathbb{I}_V}$ as $\varepsilon$ tends to zero. Q.E.D.

Part (a) may also be obtained from a reparametrization by $\Theta = \Theta(t)$, where the bijection $\Theta$ from $G$ onto $G$ solves $\Sigma_{\kappa}^\Theta(t) \Sigma_{\kappa} = (\Sigma_t \Sigma_{\kappa})^{-1}$, as introduced by Seely [9, p. 715]. In this case one obtains an exponential family in the vector parameter $\Theta$ and standard theory applies, cf. Anderson [1]. A theorem proved by Mäkeläinen, Schmidt & Styan [6] may be used to obtain uniqueness of the solution to the likelihood equations (1).

3. The Multivariate Linear Model. We now return to the linear, but not necessarily normal, model discussed in Section 1.

**THEOREM 2.** Consider independent and identically distributed random $\mathbb{R}^n$-vectors $Y_{\alpha}, \alpha = 1, \ldots, N$, with common mean vector $\Sigma_0 \Xi_{\alpha}$ and common dispersion matrix $\Sigma_t \Xi_{\alpha}$, where $N \geq \nu = \text{rank } M$. Assume that the $k$ matrices $\Sigma_{\kappa} \Xi_{\alpha} M$ span a $k$-dimensional special Jordan algebra $\mathcal{B}$ that contains $M$. Let $D_{M} = \begin{bmatrix} \vec{\Sigma}_{\Xi_{\alpha} M} & \cdots & \vec{\Sigma}_{\Xi_{\alpha} M} \end{bmatrix}$. Then the MINQUE

\[ \hat{w}_S = (D'M)^{-1} D'M \cdot \vec{S} \]
for $\bar{t}$ yields a nonnegative definite estimate $\Sigma_{K=K=K=K=}^{M_{K=K=K=K} M}$ of the invariance reduced dispersion matrix, this estimate being of rank $v$ if $S := \Sigma_{M=K=K=K=K} A_{K=K=K=K} M$ is of rank $v$.

Proof. It is easily checked that $\hat{t}$ is the MINQUE in the enlarged model \[ \{ Y_{1,K=K=K=K=K=K=K}^M : \ldots : Y_{N=K=K=K=K=K=K=K}^M \} \sim (Q, \Sigma_{K=K=K=K=K=K=K=K}^{M_{K=K=K=K=K=K=K=K} M}). \] The rest will be proved by reference to Theorem 1. Choose an $n \times v$ full rank $v$ factor $Q$ of $M$, i.e., $M = QQ'$ and $Q'Q = I_v$; then $Q'Y$ is another maximal invariant statistic [5, p. 707]. For the sole reason the proof, add a normality assumption. Then Theorem 1 is applicable to $\Sigma := Q'YQ$, and yields the same $\hat{t}$ as in (4); and the results on $\Sigma_{K=K=K=K=K=K=K=K} Q$ imply the assertions on $\Sigma_{K=K=K=K=K=K=K=K}^{M_{K=K=K=K=K=K=K=K} M}$. Q.E.D.

If a normality assumption is added to Theorem 2, then using Theorem 1, we obtain the following:

**Corollary.** If the common distribution of $Y_1, \ldots, Y_N$ is normal, then $\hat{t}$ is the UMVU and REML estimate of $\bar{t}$, as well as the MINQUE.

Examples may be found in Corbeil and Searle [2]. In each one of their four cases a special Jordan algebra is present: equality of MINQUE (i.e., ANOVA estimators) and REML is implied by the Corollary and need not be checked explicitly, nor need the likelihood equations be solved iteratively.

4. Acknowledgements. This paper was presented at the Instytut Matematyczny PAN, Wrocław, by the first author, who would like to thank the Polish Academy of Sciences for their kind invitation. The authors would
also like to thank T.W. Anderson, K. Conradsen, and H. Witting for very helpful discussions. The second author's research was begun while he was a Canada Council fellow at the University of Helsinki.

INSTITUT FÜR MATHEMATISCHE STOCHASTIK, ALBERT-LUDWIGS-UNIVERSITÄT, D-7800 FREIBURG IM BREISGAU, FEDERAL REPUBLIC OF GERMANY, AND DEPARTMENT OF STATISTICS, STANFORD UNIVERSITY, STANFORD, CALIFORNIA 94305, USA.

DEPARTMENT OF MATHEMATICS, MCGILL UNIVERSITY, MONTREAL, QUEBEC, CANADA H3A 2K6.
REFERENCES


