LEAST SQUARES THEORY USING AN ESTIMATED DISPERSION MATRIX AND ITS APPLICATION TO MEASUREMENT OF SIGNALS

C. RADHAKRISHNA RAO
INDIAN STATISTICAL INSTITUTE

1. Introduction

In this paper are considered some problems in the estimation and inference on unknown parameters in a linear model under various assumptions on the error term. We write the linear model in the form

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\tau} + \mathbf{e} \tag{1}$$

where $\mathbf{Y}$ is a $p \times 1$ vector of observable random variables, $\mathbf{X}$ is $p \times m$ matrix of known coefficients, $\boldsymbol{\tau}$ is a $p \times 1$ vector of unknown (nonstochastic) parameters, and $\mathbf{e}$ is a $p \times 1$ vector of errors. If $\boldsymbol{\Sigma}$, the dispersion matrix of $\mathbf{e}$, is known, then there is no problem, as the method of least squares (for the correlated case) can be applied to estimate and draw inferences on linear parametric functions of $\boldsymbol{\tau}$. We shall consider the case where $\boldsymbol{\Sigma}$ is unknown but an estimate $\hat{\boldsymbol{\Sigma}}$ of $\boldsymbol{\Sigma}$ is available, which may be computed from previous data or from the present data without making any assumption on $\boldsymbol{\tau}$, and discuss how this information can be used. In other words, we will discuss the theory of *least squares using an estimated dispersion matrix*. It is shown that the estimator of $\boldsymbol{\tau}$, obtained by merely substituting $\hat{\boldsymbol{\Sigma}}$ for $\boldsymbol{\Sigma}$ in the least squares estimator of $\boldsymbol{\tau}$ when $\boldsymbol{\Sigma}$ is known, is not necessarily the best. Certain improvements can be made depending on the *known or inferred structure* of $\boldsymbol{\Sigma}$.

Let us denote by E, D, and C the operators for expectation, dispersion, and covariance respectively. We consider the following specific structures for $\boldsymbol{\Sigma}$.

Case 1. The matrix $D(\mathbf{Y}) = \boldsymbol{\Sigma}$ is an unknown arbitrary positive definite matrix.

Case 2. The matrix $\boldsymbol{\Sigma} = \mathbf{X}\boldsymbol{\Gamma}\mathbf{X}' + \mathbf{Z}\boldsymbol{\Theta}\mathbf{Z}' + \sigma^2\mathbf{I}$, where $\boldsymbol{\Gamma}$, $\boldsymbol{\Theta}$, and σ^2 are unknown, and $\mathbf{Z}$ is a matrix such that $\mathbf{X}'\mathbf{Z} = \mathbf{0}$. Such a situation arises when we consider the mixed model

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\tau} + \mathbf{X}\boldsymbol{\gamma} + \mathbf{Z}\boldsymbol{\xi} + \mathbf{e} \tag{2}$$

where $\boldsymbol{\gamma}$, $\boldsymbol{\xi}$, and $\mathbf{e}$ are all uncorrelated random vectors such that $E(\boldsymbol{\gamma}) = \mathbf{0}$, $D(\boldsymbol{\gamma}) = \boldsymbol{\Gamma}$, $E(\boldsymbol{\xi}) = \mathbf{0}$, $D(\boldsymbol{\xi}) = \boldsymbol{\Theta}$, and $D(\mathrm{e}) = \sigma^2\mathbf{I}$.

Case 3. The matrix $\boldsymbol{\Sigma} = \mathbf{C}\boldsymbol{\Gamma}\mathbf{C}' + \sigma^2\mathbf{I}$, where $\boldsymbol{\Gamma}$ and σ^2 are unknown and $\mathbf{C}$ is a known matrix. Such a situation arises when we consider the mixed model

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\tau} + \mathbf{C}\boldsymbol{\gamma} + \mathbf{e} \tag{3}$$

where $\boldsymbol{\gamma}$ is a random vector such that $E(\boldsymbol{\gamma}) = \mathbf{0}$, $D(\boldsymbol{\gamma}) = \boldsymbol{\Gamma}$, $D(\mathbf{e}) = \sigma^2\mathbf{I}$ and $C(\boldsymbol{\gamma}, \mathbf{e}) = \mathbf{0}$. A model of the type (3) has been considered by Duncan [3], Henderson *et al.* [5], and others under a different set of assumptions on the variables $\boldsymbol{\gamma}$ and $\mathbf{e}$.

Case 4. The matrix $\boldsymbol{\Sigma} = \mathbf{C}\boldsymbol{\Gamma}\mathbf{C}' + \sigma^2\mathbf{I}$, where $\mathbf{C}$, $\boldsymbol{\Gamma}$, and σ^2 are unknown, but the rank of $\mathbf{C}$ is known or can be inferred from an estimate of $\boldsymbol{\Sigma}$. Or, in other words, the error vector $\mathbf{e}$ in (1) has a factor analytic structure with a common specific variance for the components.

Case 5. Let $\mathbf{Y}' = (y_1, \cdots, y_p)$ and $\mathbf{e}' = (e_1, \cdots, e_p)$. The component y_t has the representation

$$y_t = P_k(t) + e_t \tag{4}$$

where the nonrandom part is a polynomial of the k-th degree in time, $P_k(t) = \beta_0 + \beta_1 t + \cdots + \beta_k t^k$, and the error terms e_t have an autoregressive scheme

$$e_t = \rho_1 e_{t-1} + \cdots + \rho_m e_{t-m} + \eta_t \tag{5}$$

where η_t are uncorrelated errors with a common variance σ^2. The parameters β_i representing the coefficients of the polynomial $P_k(t)$, the autoregressive parameters ρ_j, and σ^2 are all unknown. The problem is to estimate the parameters β_i from a single series of observations on y_t and in the absence of an independent estimate of $\boldsymbol{\Sigma}$.

In practice, we have the additional problem of checking the accuracy of an assumed model before estimating the unknown parameters. Appropriate tests for this purpose have been suggested in each case. Such tests are possible if an independent estimate of $\boldsymbol{\Sigma}$ is available.

An independent estimate of $\boldsymbol{\Sigma}$ may be available from past data or from multiple observations on vector $\mathbf{Y}$ of model (1). In the latter case, the observations are replaced by the average vector for which model (1) is true and the sample variance covariance (dispersion) matrix provides an estimate of $\boldsymbol{\Sigma}$. Note that if $\overline{\mathbf{Y}}$ and $(n-1)^{-1}\mathbf{S}$ represent the sample average and dispersion matrix, then the model (1) applied to $\overline{\mathbf{Y}}$ is written $\overline{\mathbf{Y}} = \mathbf{X}\boldsymbol{\tau} + \mathbf{e}$, and an estimate of $D(\overline{\mathbf{Y}})$ is $[n(n-1)]^{-1}\mathbf{S}$ where n is the sample size. Thus, the problem is reduced to the standard form with a linear model for a single vector random variable for which an estimate of the dispersion matrix is available.

In the general case we shall represent the dispersion matrix of $\mathbf{Y}$ by $\boldsymbol{\Sigma}$ and its estimator by $f^{-1}\mathbf{S}$. For purposes of tests of significance and computing confidence intervals for unknown parameters we shall assume the following distributions for $\mathbf{Y}$ and $\mathbf{S}$:

$$\mathbf{Y} \sim N_p(\boldsymbol{\tau}, \boldsymbol{\Sigma}), \tag{6}$$

$$\mathbf{S} \sim W_p(f, \boldsymbol{\Sigma}), \tag{7}$$

where $N_p(\boldsymbol{\tau}, \boldsymbol{\Sigma})$ denotes a p-variate normal distribution with mean and dispersion matrix as indicated in the brackets, and $W_p(f, \boldsymbol{\Sigma})$ denotes Wishart's distribution on degrees of freedom and expected matrix as indicated in the brackets. The symbol $\sim$ is used for "distributed as."

We let p denote the dimension (or the number of components) of $\mathbf{Y}$, and $\mathbf{k}$ that of $\boldsymbol{\tau}$. Without loss of generality, we shall assume that rank $\mathbf{X}$ is also k.

The methods discussed in this paper have wide applicability, although the specific problem of signal measurement is considered in the last section. Other areas in which these methods may be applied are in the estimation of polynomial trends of growth curves and time series, and prediction in time series.

Sometimes it may be possible to make a preliminary transformation of model (1) by multiplying both sides by $\boldsymbol{\Sigma}_0^{-1/2}$, where $\boldsymbol{\Sigma}_0$ is a guessed, or an a priori dispersion matrix of $\mathbf{e}$. The new model is $\mathbf{Y}^* = \mathbf{X}^*\boldsymbol{\tau} + \mathbf{e}^*$ where $\mathbf{Y}^* = \boldsymbol{\Sigma}_0^{-1/2}\mathbf{Y}$, $\mathbf{e}^* = \boldsymbol{\Sigma}_0^{-1/2}\mathbf{e}$ and $\mathbf{X}^* = \boldsymbol{\Sigma}_0^{-1/2}\mathbf{X}$. The estimated dispersion matrix of $\mathbf{Y}^*$ is $f^{-1}\mathbf{S}^*$ where $\mathbf{S}^* = \boldsymbol{\Sigma}_0^{-1/2}\mathbf{S}\boldsymbol{\Sigma}_0^{-1/2}$. We can then apply the methods of this paper assuming similar models for $D(\mathbf{Y}^*)$.

2. Some algebraic lemmas

Now we will prove some algebraic lemmas which are used in later sections of the paper.

LEMMA 2a. *Let* $\mathbf{A}$ *be a positive definite matrix partitioned as*

$$\begin{pmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{A}_{21} & \mathbf{A}_{22} \end{pmatrix} \tag{8}$$

with its inverse as

$$\begin{pmatrix} \mathbf{A}^{11} & \mathbf{A}^{12} \\ \mathbf{A}^{21} & \mathbf{A}^{22} \end{pmatrix}. \tag{9}$$

Then $\mathbf{A}^{11} - (\mathbf{A}_{11})^{-1}$ *is nonnegative definite.*

Multiplying (8) by (9) we have

$$\mathbf{A}_{11}\mathbf{A}^{11} + \mathbf{A}_{12}\mathbf{A}^{21} = \mathbf{I}, \qquad \mathbf{A}_{11}\mathbf{A}^{12} + \mathbf{A}_{12}\mathbf{A}^{22} = \mathbf{0}. \tag{10}$$

Multiplying both the equations in (10) by $\mathbf{A}_{11}^{-1}$,

$$\mathbf{A}^{11} + \mathbf{A}_{11}^{-1}\mathbf{A}_{12}\mathbf{A}^{21} = \mathbf{A}_{11}^{-1}, \qquad \mathbf{A}^{12} = -\mathbf{A}_{11}^{-1}\mathbf{A}_{12}\mathbf{A}^{22}. \tag{11}$$

Rearranging the terms in the first and substituting for $\mathbf{A}^{21} = (\mathbf{A}^{12})'$ from the second equation of (11), we have

$$\mathbf{A}^{11} - \mathbf{A}_{11}^{-1} = -\mathbf{A}_{11}^{-1}\mathbf{A}_{12}\mathbf{A}^{21} = \mathbf{A}_{11}^{-1}\mathbf{A}_{12}\mathbf{A}^{22}\mathbf{A}_{21}\mathbf{A}_{11}^{-1}. \tag{12}$$

The last matrix in (12) is nonnegative definite, which proves the required result.

As a corollary we have the following result. Let the partitioned matrix

$$\begin{pmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} & \mathbf{A}_{13} & \cdots \\ \mathbf{A}_{21} & \mathbf{A}_{22} & \mathbf{A}_{23} & \cdots \\ \mathbf{A}_{31} & \mathbf{A}_{32} & \mathbf{A}_{33} & \cdots \\ \cdot & \cdot & \cdot & \cdots \end{pmatrix} \tag{13}$$

be positive definite. Denote by $\mathbf{A}_i^{11}$ the partition in the leading position in the reciprocal of the submatrix of (13) obtained by considering the first i row and column partitions. Then the matrix $\mathbf{A}_i^{11} - \mathbf{A}_j^{11}$ is nonnegative definite for any i and j such that $i \geq j$.

LEMMA 2b. *Let* $\mathbf{X}$ *be a* $p \times k$ *matrix of rank* k, *and let* $\mathbf{Z}$ *be a* $p \times (p - k)$ *matrix of rank* $(p - k)$ *such that* $\mathbf{X}'\mathbf{Z} = \mathbf{0}$. *Then*

$$(\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\Sigma}^{-1} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' - (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\Sigma}\mathbf{Z}(\mathbf{Z}'\boldsymbol{\Sigma}\mathbf{Z})^{-1}\mathbf{Z}' \tag{14}$$

where $\boldsymbol{\Sigma}$ *is any* $p \times p$ *positive definite matrix.*

Multiplying both sides of (14) by $\mathbf{X}$, it is easily seen that the equality holds. If multiplication by $\mathbf{Z}$ also results in equality, then (14) is true. Multiplying by $\mathbf{Z}$ from the right and by $(\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{X})$ from the left we have

$$\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{Z} = -\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\Sigma}\mathbf{Z}[(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\boldsymbol{\Sigma}\mathbf{Z}]^{-1}, \tag{15}$$

$$\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{Z}[(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\boldsymbol{\Sigma}\mathbf{Z}] = -\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\Sigma}\mathbf{Z}, \tag{16}$$

$$\mathbf{X}'\boldsymbol{\Sigma}^{-1}[\mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}' + \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}']\boldsymbol{\Sigma}\mathbf{Z} = 0, \tag{17}$$

which is true, since the expression within the square brackets of (17) is $\mathbf{I}$ and $\mathbf{X}'\mathbf{Z} = 0$

LEMMA 2c. *With* $\mathbf{X}$ *and* $\mathbf{Z}$ *as in lemma 2b, the matrix*

$$(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\Sigma}\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} - (\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{X})^{-1} \tag{18}$$

is nonnegative definite.

Consider the matrix

$$\begin{pmatrix} \mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{X} & \mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{Z} \\ \mathbf{Z}'\boldsymbol{\Sigma}^{-1}\mathbf{X} & \mathbf{Z}'\boldsymbol{\Sigma}^{-1}\mathbf{Z} \end{pmatrix} = \begin{pmatrix} \mathbf{X}' \\ \mathbf{Z}' \end{pmatrix} \boldsymbol{\Sigma}^{-1}(\mathbf{X} \vdots \mathbf{Z}). \tag{19}$$

The reciprocal of the right-hand side is

$$(\mathbf{X} \vdots \mathbf{Z})^{-1}\boldsymbol{\Sigma}\begin{pmatrix} \mathbf{X}' \\ \mathbf{Z}' \end{pmatrix}^{-1}. \tag{20}$$

But

$$(\mathbf{X} \vdots \mathbf{Z})^{-1} = \begin{pmatrix} (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' \\ (\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}' \end{pmatrix}. \tag{21}$$

Substituting the result (21) in (20), we have the leading partition in the reciprocal of (19) as

$$(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\Sigma}\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}. \tag{22}$$

Hence, (18) follows by applying the result of lemma 1a.

Finally we need some results on restricted eigenvectors of a symmetric matrix as developed by the author elsewhere (Rao [12]).

LEMMA 2d. *Let* $\boldsymbol{\Sigma}$ *and* $\mathbf{X}$ *be as in lemma 2b and consider the nonzero eigenvalues and right eigenvectors of the matrix* $(\mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}')\boldsymbol{\Sigma}$. *Let* $\lambda_1 \geq \lambda_2, \cdots \geq \lambda_r \geq 0$ *be the eigenvalues and* $\mathbf{L}_1, \cdots, \mathbf{L}_r$ *be the corresponding eigenvectors which can be chosen to be mutually orthogonal. Then we have the following results:*

$$\begin{aligned} \mathbf{L}_i'\mathbf{X} &= 0, \qquad i = 1, \cdots, r, \\ \mathbf{L}_i'\boldsymbol{\Sigma}\mathbf{L}_j &= 0, \qquad i \neq j. \end{aligned} \tag{23}$$

The results are easy to prove. The $\mathbf{L}_i$ are said to be restricted eigenvectors of $\boldsymbol{\Sigma}$ with the restriction $\mathbf{L}_i'\mathbf{X} = 0$.

LEMMA 2e. *Let $\boldsymbol{\Sigma} = \mathbf{C}\boldsymbol{\Gamma}\mathbf{C}' + \sigma^2\mathbf{I}$ and let $\mathbf{L}$ be a restricted eigenvector of $\boldsymbol{\Sigma}$ corresponding to the restricted eigenvalue σ^2 so that $\mathbf{L}'\mathbf{X} = \mathbf{0}$. Then*

$$\mathbf{L}'\boldsymbol{\Sigma}\mathbf{X} = 0. \tag{24}$$

By definition,

$$(\mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}')(\mathbf{C}\boldsymbol{\Gamma}\mathbf{C}' + \sigma^2\mathbf{I})\mathbf{L} = \sigma^2\mathbf{L}, \tag{25}$$

which on simplification gives

$$(\mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}')\mathbf{C}\boldsymbol{\Gamma}\mathbf{C}'\mathbf{L} = 0. \tag{26}$$

Multiplying by $\mathbf{L}'$ and putting $\mathbf{L}'\mathbf{X} = 0$, we have

$$\mathbf{L}'\mathbf{C}\boldsymbol{\Gamma}\mathbf{C}'\mathbf{L} = 0 \Rightarrow \mathbf{C}\boldsymbol{\Gamma}\mathbf{C}'\mathbf{L} = 0 \Rightarrow \mathbf{X}'\boldsymbol{\Sigma}\mathbf{L} = 0. \tag{27}$$

3. Covariance adjustment

In this paper, we frequently refer to covariance adjustment in an estimator using another statistic (with zero expectation) as a concomitant variable. The procedure is described as follows.

Let $\mathbf{T}_1$ and $\mathbf{T}_2$ be two vector statistics of orders k and r such that $E(\mathbf{T}_1) = \boldsymbol{\tau}$ and $E(\mathbf{T}_2) = \mathbf{0}$, where $\boldsymbol{\tau}$ is a vector of k unknown parameters. The vector $\mathbf{T}_1$ is an estimator of $\boldsymbol{\tau}$, but if $C(\mathbf{T}_1, \mathbf{T}_2) \neq \mathbf{0}$, then a better estimator of $\boldsymbol{\tau}$ can be found when the dispersion matrix

$$\boldsymbol{\Lambda} = \begin{pmatrix} \boldsymbol{\Lambda}_{11} & \boldsymbol{\Lambda}_{12} \\ \boldsymbol{\Lambda}_{21} & \boldsymbol{\Lambda}_{22} \end{pmatrix} \tag{28}$$

of $(\mathbf{T}_1, \mathbf{T}_2)$ is known. Thus, if we consider the estimator

$$\boldsymbol{\tau}^* = \mathbf{T}_1 - \boldsymbol{\Lambda}_{12}\boldsymbol{\Lambda}_{22}^{-1}\mathbf{T}_2, \tag{29}$$

then

$$D(\boldsymbol{\tau}^*) = \boldsymbol{\Lambda}_{11} - \boldsymbol{\Lambda}_{12}\boldsymbol{\Lambda}_{22}^{-1}\boldsymbol{\Lambda}_{21} = D(\mathbf{T}_1) - \boldsymbol{\Lambda}_{12}\boldsymbol{\Lambda}_{22}^{-1}\boldsymbol{\Lambda}_{21}; \tag{30}$$

that is, $D(\mathbf{T}_1) - D(\boldsymbol{\tau}^*)$ is always nonnegative definite. Hence $\boldsymbol{\tau}^*$ is more efficient than $\mathbf{T}_1$.

If only an estimate of $\boldsymbol{\Lambda}$,

$$\mathbf{U} = \begin{pmatrix} \mathbf{U}_{11} & \mathbf{U}_{12} \\ \mathbf{U}_{21} & \mathbf{U}_{22} \end{pmatrix}, \tag{31}$$

is available, we may substitute $\mathbf{U}_{ij}$ for $\boldsymbol{\Lambda}_{ij}$ in the formula (29) and obtain the (covariance) adjusted estimator as

$$\hat{\boldsymbol{\tau}} = \mathbf{T}_1 - \mathbf{U}_{12}\mathbf{U}_{22}^{-1}\mathbf{T}_2. \tag{32}$$

It is seen that when $\mathbf{U}_{ij}$ are distributed independently of $\mathbf{T}_i$,

$$E(\hat{\tau}) = E(\mathbf{T}_1 - \mathbf{U}_{12}\mathbf{U}_{22}^{-1}\mathbf{T}_2) = \tau \tag{33}$$

so that the adjusted estimator $\hat{\tau}$ is also unbiased for τ. Now

$$\begin{aligned} D(\hat{\tau}) &= D(\mathbf{T}_1) + E[\mathbf{U}_{12}\mathbf{U}_{22}^{-1}D(\mathbf{T}_2)\mathbf{U}_{22}^{-1}\mathbf{U}_{21}] - E[\mathbf{U}_{12}\mathbf{U}_{22}^{-1}C(\mathbf{T}_2, \mathbf{T}_1)] \\ &\qquad - E[C(\mathbf{T}_1, \mathbf{T}_2)\mathbf{U}_{22}^{-1}\mathbf{U}_{21}] \\ &= \Lambda_{11} + E[\mathbf{U}_{12}\mathbf{U}_{22}^{-1}\Lambda_{22}\mathbf{U}_{22}^{-1}\mathbf{U}_{21}] - E[\mathbf{U}_{12}\mathbf{U}_{22}^{-1}\Lambda_{21}] - E[\Lambda_{12}\mathbf{U}_{22}^{-1}\mathbf{U}_{21}] \end{aligned} \tag{34}$$

where the expectations are taken over the variations of $\mathbf{U}_{ij}$. We are no longer in a position to claim that $D(\mathbf{T}_1) - D(\hat{\tau})$ is always nonnegative definite as in the case of $D(\mathbf{T}_1) - D(\tau^*)$. As a matter of fact, it is seen from (34) that if $\Lambda_{21} = \mathbf{0}$, or very nearly $\mathbf{0}$, then $D(\mathbf{T}_1) - D(\hat{\tau})$ is negative definite; that is, $\mathbf{T}_1$ is more efficient than $\hat{\tau}$.

Thus, covariance adjustment can result in a decrease in efficiency when an estimated dispersion matrix is used in the place of the unknown matrix. However, if Λ_{12} is not close to zero, we should expect $D(\mathbf{T}_1) - D(\hat{\tau})$ to be nonnegative definite.

There is, however, an important problem. It is possible that the use of $\mathbf{T}_2$ as a whole for covariance adjustment is not optimum, and a suitable choice of functions of $\mathbf{T}_2$ for this purpose may provide maximum efficiency. In the absence of an exact knowledge about Λ the optimum solution cannot be found. However, an estimate of Λ may provide some guidance in the choice of suitable functions of $\mathbf{T}_2$. We consider such problems in the rest of the sections.

We were able to draw some conclusions on the basis of the formula (34) for $D(\hat{\tau})$ without making an explicit evaluation of the expectations involved. We shall now complete the discussion by making the following assumptions on the distributions of $(\mathbf{T}_1, \mathbf{T}_2)$ and $\mathbf{U}$:

$$\begin{pmatrix}\mathbf{T}_1\\ \mathbf{T}_2\end{pmatrix} \sim N_{k+r}\left[\begin{pmatrix}\tau\\ 0\end{pmatrix}, \Lambda\right], \tag{35}$$

$$f\mathbf{U} \sim W_{k+r}(f, \Lambda). \tag{36}$$

In the rest of the present section we lay down procedures for drawing inferences on τ on the basis of the estimator $\hat{\tau}$ obtained by adjusting a *given* estimator T_1 with respect to a *given* concomitant variable $\mathbf{T}_2$ under the assumptions (35) and (36). An important result in this direction is contained in lemma 3a.

LEMMA 3a. *The conditional distributions of* $\hat{\tau}$ *and* $\mathbf{G} = f(\mathbf{U}_{11} - \mathbf{U}_{12}\mathbf{U}_{22}^{-1}\mathbf{U}_{21})$, *given* $\mathbf{T}_2$ *and* $\mathbf{U}_{22}$, *are independent, and the conditional distributions are*

$$\hat{\tau} \sim N_k(\tau, (1 + T_\tau^2)\Gamma), \tag{37}$$

$$\mathbf{G} \sim W_k(f - r, \Gamma), \tag{38}$$

where $fT_\tau^2 = \mathbf{T}_2'\mathbf{U}_{22}^{-1}\mathbf{T}_2$ *and* $\Gamma = \Lambda_{11} - \Lambda_{12}\Lambda_{22}^{-1}\Lambda_{21}$.

As a consequence of lemma 3a, we have the results of lemma 3b.

LEMMA 3b. *Let*

$$U_k = \frac{(\hat{\tau} - \tau)'\mathbf{G}^{-1}(\hat{\tau} - \tau)}{1 + T_\tau^2}, \tag{39}$$

$$V_k = (\hat{\tau} - \tau)'\mathbf{G}^{-1}(\hat{\tau} - \tau). \tag{40}$$

Then

$$\frac{f - r - k + 1}{k} U_k \sim F(k, f - r - k + 1); \tag{41}$$

that is, a variance ratio distribution on k *and* $f - r - k + 1$ *degrees of freedom, and* V_k *has the distribution*

$$\text{const. } V_k^{k/2}(1 + V_k)^{-(f-k-r+3)/2}{}_2F_1\left(\frac{r}{2}, \frac{f - r + 1}{2}; \frac{f + k + 1}{2}, \frac{V_k}{1 + V_k}\right) dV_k. \tag{42}$$

The distributions of lemmas 3a and 3b follow from the basic results derived in earlier papers by the author (Rao [7], [8], [9], [11]).

The $(1 - \alpha)$ probability concentration ellipsoids for the unknown parameter τ based on U_k and V_k are

$$(\hat{\tau} - \tau)'\mathbf{G}^{-1}(\hat{\tau} - \tau) \le \frac{k}{f - k - r + 1} F_\alpha(1 + T_\tau^2), \tag{43}$$

$$(\hat{\tau} - \tau)'\mathbf{G}^{-1}(\hat{\tau} - \tau) \le V_{k\alpha}, \tag{44}$$

where F_α and $V_{k\alpha}$ are the upper α-probability points of the F and V_k distributions given in (41) and (42), respectively. It has been shown in the earlier paper (Rao [8]) that the inferences based on the F and V_k distributions are not very different, although there is slight advantage in using the V_k distribution. However, the percentage points of the V_k distribution are not yet available.

Observing that

$$(\hat{\tau} - \tau)'\mathbf{G}^{-1}(\hat{\tau} - \tau) = \max_{\mathbf{P}} \frac{(\mathbf{P}'\hat{\tau} - \mathbf{P}'\tau)^2}{\mathbf{P}'\mathbf{G}\mathbf{P}}, \tag{45}$$

we find that the simultaneous confidence intervals for linear functions $\mathbf{P}'\tau$ of τ are provided by

$$\mathbf{P}'\hat{\tau} \pm [\mathbf{P}'\mathbf{G}\mathbf{P}kF_\alpha(1 + T_\tau^2)/(f - k - r + 1)]^{1/2} \tag{46}$$

$$\mathbf{P}'\hat{\tau} \pm [\mathbf{P}'\mathbf{G}\mathbf{P}V_{k\alpha}]^{1/2} \tag{47}$$

using (43) or (44).

If the confidence interval for a particular linear function $\mathbf{P}'\tau$ is needed, we replace k by unity in the expressions (46) and (47). The number F_α is then the upper α-probability value of F on 1 and $f - r$ degrees of freedom, and $V_{1\alpha}$ is the upper α-probability value of the V_1 distribution.

The knowledge of the actual distributions of $(\mathbf{T}_1, \mathbf{T}_2)$ and $\mathbf{U}$ enable us to find the exact expression for $D(\hat{\tau})$ which is left in a symbolic form in (34). Using the result (37),

$$D(\hat{\tau}|\mathbf{T}_2, \mathbf{U}_{22}) = (1 + T_\tau^2)\mathbf{\Gamma}. \tag{48}$$

Now, observing that

$$\frac{f - r - 1}{r} T_\tau^2 \sim F(r, f - r + 1), \tag{49}$$

we find

$$D(\hat{\boldsymbol{\tau}}) = E[(1 + T_r^2)]\boldsymbol{\Gamma} = \frac{f-1}{f-r-1}\boldsymbol{\Gamma}. \tag{50}$$

When $\boldsymbol{\Lambda}$ is known, the best estimator $\boldsymbol{\tau}^*$ has the dispersion matrix $\boldsymbol{\Gamma}$ so that the loss of efficiency in using an estimate of $\boldsymbol{\Lambda}$ is $r/(f - r - 1)$, which is zero when $r = 0$, and which tends to zero as f tends to infinity.

Further, if $\boldsymbol{\Lambda}_{12} = \mathbf{0}$, $\boldsymbol{\Gamma} = \boldsymbol{\Lambda}_{11}$ and therefore $\mathbf{T}_1$ is more efficient than $\hat{\boldsymbol{\tau}}$ for any fixed values of r and f. The situation will be the same if $\boldsymbol{\Lambda}_{12}$ is close to $\mathbf{0}$.

In the cases considered in this paper we shall first reduce the problem to that of making covariance adjustment in an estimator $\mathbf{T}_1$ using a concomitant variable $\mathbf{T}_2$ and an estimated dispersion matrix of $\mathbf{T}_1$, $\mathbf{T}_2$. Then the inference follows on the lines discussed in this section.

4. Case 1: An arbitrary matrix $\boldsymbol{\Sigma}$

4.1. *Test for specification of the model.* Let us recall that the linear model is $\mathbf{Y} = \mathbf{X}\boldsymbol{\tau} + \mathbf{e}$ where $D(\mathbf{e}) = \boldsymbol{\Sigma}$, an arbitrary positive definite matrix. An independent estimate $f^{-1}\mathbf{S}$ of $\boldsymbol{\Sigma}$ is available. Assume that $\mathbf{Y}$ and $\mathbf{S}$ have the distributions (7) and (8) respectively. The theory and appropriate statistical methods in such a case have been worked out in an earlier paper (Rao [11]). However, we shall make some important comments and also discuss an alternative way of expressing the precision of the estimators.

LEMMA 4a. *Let k be the rank of* $\mathbf{X}$, $r = p - k$, *and*

$$T_r^2 = \min_{\boldsymbol{\tau}} (\mathbf{Y} - \mathbf{X}\boldsymbol{\tau})'\mathbf{S}^{-1}(\mathbf{Y} - \mathbf{X}\boldsymbol{\tau}). \tag{51}$$

Then under the assumptions (7) *and* (8) *on the distributions of* $\mathbf{Y}$ *and* $\mathbf{S}$,

$$\frac{f-r+1}{r} T_r^2 \sim F(r, f-r+1). \tag{52}$$

The result of lemma 4a is proved in [11]. The test criterion T_r^2 examines the adequacy of the model, $\mathbf{Y} = \mathbf{X}\boldsymbol{\tau} + \mathbf{e}$, with respect to the nonrandom part.

4.2. *Estimation of parameters.* If $\boldsymbol{\Sigma}$ were known, the least squares estimator of the unknown vector $\boldsymbol{\tau}$ is obtained by minimizing

$$(\mathbf{Y} - \mathbf{X}\boldsymbol{\tau})'\boldsymbol{\Sigma}^{-1}(\mathbf{Y} - \mathbf{X}\boldsymbol{\tau}) \tag{53}$$

with respect to $\boldsymbol{\tau}$. A natural method of estimation when only an estimate of $\boldsymbol{\Sigma}$ is available is to apply the method of least squares, substituting $f^{-1}\mathbf{S}$ for $\boldsymbol{\Sigma}$. Thus, we are led to minimize the expression

$$(\mathbf{Y} - \mathbf{X}\boldsymbol{\tau})'\mathbf{S}^{-1}(\mathbf{Y} - \mathbf{X}\boldsymbol{\tau}), \tag{54}$$

and to obtain the normal equations

$$(\mathbf{X}'\mathbf{S}^{-1}\mathbf{X})\boldsymbol{\tau} = \mathbf{X}'\mathbf{S}^{-1}\mathbf{Y}. \tag{55}$$

Then we have the estimator

$$\hat{\boldsymbol{\tau}} = (\mathbf{X}'\mathbf{S}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{S}^{-1}\mathbf{Y} \tag{56}$$

under the assumption that rank $\mathbf{X}$ is k, the dimension of $\boldsymbol{\tau}$ (without loss of generality).

To study the properties of $\hat{\boldsymbol{\tau}}$ and to draw inferences on $\boldsymbol{\tau}$, we make the following transformation of the model:

$$\mathbf{T}_1 = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}, \qquad E(\mathbf{T}_1) = \boldsymbol{\tau}, \tag{57}$$

$$\mathbf{T}_2 = \mathbf{Z}'\mathbf{Y}, \qquad E(\mathbf{T}_2) = \mathbf{0}, \tag{58}$$

where $\mathbf{Z}$ is $p \times r$ matrix of rank r such that $\mathbf{Z}'\mathbf{X} = \mathbf{0}$. The estimated dispersion matrix of $(\mathbf{T}_1, \mathbf{T}_2)$ is $\mathbf{U}$ where

$$f\mathbf{U} = \begin{pmatrix} (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{S}\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} & (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{S}\mathbf{Z} \\ \mathbf{Z}'\mathbf{S}\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} & \mathbf{Z}'\mathbf{S}\mathbf{Z} \end{pmatrix}. \tag{59}$$

Then the theory and methods developed in section 3 apply. The adjusted estimator, according to the formula (32), is

$$\begin{aligned} \mathbf{T}_1 - \mathbf{U}_{12}\mathbf{U}_{22}^{-1}\mathbf{T}_2 &= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y} - (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{S}\mathbf{Z}(\mathbf{Z}'\mathbf{S}\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{Y} \\ &= (\mathbf{X}'\mathbf{S}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{S}^{-1}\mathbf{Y}, \end{aligned} \tag{60}$$

using the identity of lemma 2b. Thus, the estimators (56) and (60) are the same. The formula (56) is useful in that it provides the estimator directly in terms of given quantities $\mathbf{X}$, $\mathbf{S}$, and $\mathbf{Y}$ (that is, not involving $\mathbf{Z}$).

We shall now apply the formulae of section 3 to obtain explicit expressions in terms of $\mathbf{X}$, $\mathbf{S}$, and $\mathbf{Y}$ for drawing inferences on $\boldsymbol{\tau}$. The quantities that appear in the formulae (43), (44), (46), and (47) providing confidence intervals are

$$T_r^2 = f^{-1}\mathbf{T}_2'\mathbf{U}_{22}^{-1}\mathbf{T}_2 = \mathbf{Y}'\mathbf{S}^{-1}\mathbf{Y} - \hat{\boldsymbol{\tau}}'\mathbf{X}'\mathbf{S}^{-1}\mathbf{Y}, \tag{61}$$

$$\mathbf{G} = f(\mathbf{U}_{11} - \mathbf{U}_{12}\mathbf{U}_{22}^{-1}\mathbf{U}_{21}) = (\mathbf{X}'\mathbf{S}^{-1}\mathbf{X})^{-1}. \tag{62}$$

The identity in (62) is derived from the identity (14) of lemma 2b. We now have $\hat{\boldsymbol{\tau}}$, T_r^2, and $\mathbf{G}$ all expressed in terms of $\mathbf{X}$, $\mathbf{S}$, and $\mathbf{Y}$, and the methods of section 3 can be applied using the computed values of $\hat{\boldsymbol{\tau}}$, T_r^2, and $\mathbf{G}$.

5. A lemma on least squares estimators

In section 4 we have exhibited the estimator (56) as derived from $\mathbf{T}_1 = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}$ after making covariance adjustment with respect to $\mathbf{T}_2 = \mathbf{Z}'\mathbf{Y}$. As a matter of fact, the choice of $\mathbf{T}_1$ can be arbitrary subject to the condition that $E(\mathbf{T}_1) = \boldsymbol{\tau}$, and the choice of $\mathbf{Z}$ defining $\mathbf{T}_2$ can be arbitrary subject to the condition $\mathbf{Z}'\mathbf{X} = \mathbf{0}$; the adjusted estimator using the appropriate dispersion matrix in each case is the same. We have seen that there are situations where covariance adjustment may not lead to better estimators, but such questions cannot be examined unless there are preassigned choices of $\mathbf{T}_1$ and $\mathbf{T}_2$. We can then raise the specific question as to what components or functions of $\mathbf{T}_2$ would be useful for covariance adjustment. We shall consider some special structures for $\boldsymbol{\Sigma}$ which enable us to make a choice of $\mathbf{T}_1$ and $\mathbf{T}_2$, and then look for relevant concomitant variables.

Let us observe that the choice $\mathbf{T}_1 = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}$ is the best linear estimator of $\boldsymbol{\tau}$ when $\boldsymbol{\Sigma}$, the dispersion matrix of $\mathbf{Y}$, has the special form $\sigma^2\mathbf{I}$ (that is, when the components of the error vector $\mathbf{e}$ in the model (1) are uncorrelated). We shall now determine the class of $\boldsymbol{\Sigma}$ matrices for which $(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}$ is the best linear estimator of $\boldsymbol{\tau}$. The object is to characterize the class of $\boldsymbol{\Sigma}$ matrices for which covariance adjustment in $\mathbf{T}_1$ results in a loss of efficiency. For given $\boldsymbol{\Sigma}$, the best linear estimator of $\boldsymbol{\tau}$ is $(\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{Y}$. Then the question raised is equivalent to the problem of determining $\boldsymbol{\Sigma}$ so that it satisfies the equation

$$(\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\Sigma}^{-1} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'. \tag{63}$$

Lemma 5a provides the set of solutions to (63).

LEMMA 5a. *Let* $\mathbf{Z}$ *be a* $p \times r$ *matrix of rank* $r = (p - \text{rank}\ \mathbf{X})$ *such that* $\mathbf{Z}'\mathbf{X} = 0$, *and* R *be the set of* $\boldsymbol{\Sigma}$ *matrices of the form*

$$\boldsymbol{\Sigma} = \mathbf{X}\boldsymbol{\Gamma}\mathbf{X}' + \mathbf{Z}\boldsymbol{\Theta}\mathbf{Z}' + \sigma^2\mathbf{I} \tag{64}$$

where $\boldsymbol{\Gamma}$, $\boldsymbol{\Theta}$, *and* σ^2 *are arbitrary. Then the necessary and sufficient condition that the least squares estimator of* $\boldsymbol{\tau}$, *in the model* $\mathbf{Y} = \mathbf{X}\boldsymbol{\tau} + \mathbf{e}$ *with* $D(\mathbf{e}) = \boldsymbol{\Sigma}$, *is the same as that for the special choice* $D(\mathbf{e}) = \sigma^2\mathbf{I}$ *is that* $\boldsymbol{\Sigma} \in R$.

The result of lemma 5a shows that for any $\boldsymbol{\Sigma}$ of the form (64) the least squares estimator of $\boldsymbol{\tau}$ is $(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}$ (which is well known for the special case of uncorrelated errors) and vice versa.

We note that the complete class of linear functions of $\mathbf{Y}$ with zero expectation is provided by $\mathbf{Z}'\mathbf{Y}$. Hence, if $(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}$ is the least squares estimate of $\boldsymbol{\tau}$, then

$$\text{cov}\,[(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}, \mathbf{Z}'\mathbf{Y}] = 0; \tag{65}$$

that is,

$$(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\Sigma}\mathbf{Z} = 0 \Leftrightarrow \mathbf{X}'\boldsymbol{\Sigma}\mathbf{Z} = 0. \tag{66}$$

Then it is easy to verify that

$$\mathbf{X}'\boldsymbol{\Sigma}\mathbf{Z} = 0 \Leftrightarrow \boldsymbol{\Sigma} = \mathbf{X}\boldsymbol{\Gamma}\mathbf{X}' + \mathbf{Z}\boldsymbol{\Theta}\mathbf{Z}' + \sigma^2\mathbf{I}. \tag{67}$$

In the proof $\mathbf{X}$ is taken to be of full rank. But this is unnecessary as all the steps are valid with a general inverse of $\mathbf{X}'\mathbf{X}$ (Rao [14]). Further, $\boldsymbol{\Sigma}$ may be singular.

As a corollary, we find that for any $\boldsymbol{\Sigma}$ of the form

$$\boldsymbol{\Sigma} = \mathbf{X}\boldsymbol{\Gamma}\mathbf{X}' + \boldsymbol{\Sigma}_0\mathbf{Z}\boldsymbol{\Theta}\mathbf{Z}'\boldsymbol{\Sigma}_0 + \boldsymbol{\Sigma}_0, \tag{68}$$

the least squares estimator is the same as that for Σ_0.

Let us compare the estimators $\mathbf{T}_1 = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}$ and $\hat{\boldsymbol{\tau}} = (\mathbf{X}'\mathbf{S}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{S}^{-1}\mathbf{Y}$ of $\boldsymbol{\tau}$ under the assumption that $\boldsymbol{\Sigma} \in R$. It is seen that

$$D[(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}] = (\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{X})^{-1}, \tag{69}$$

$$D[(\mathbf{X}'\mathbf{S}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{S}^{-1}\mathbf{Y}] = \frac{f-1}{f-r-1}(\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{X})^{-1}, \tag{70}$$

so that the effect of using an estimate of $\boldsymbol{\Sigma}$, when in fact $\boldsymbol{\Sigma} \in R$, or to a slightly extended class, is to decrease efficiency. For $\boldsymbol{\Sigma} \notin R$ equation (70) remains the same, while (69) changes to

$$D[(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}] = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\Sigma}\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}. \tag{71}$$

From the result (18) we find that the estimator using an estimate of $\boldsymbol{\Sigma}$ has a smaller dispersion matrix only if

$$(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\Sigma}\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} \tag{72}$$

is somewhat larger than

$$(\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{X})^{-1} \tag{73}$$

to compensate for the multiplying factor in (70).

6. Case 2: $D(Y) = \mathbf{X}\boldsymbol{\Gamma}\mathbf{X}' + \mathbf{Z}\,\boldsymbol{\Theta}\,\mathbf{Z}' + \sigma^2\mathbf{I}$

As mentioned in the introduction, the dispersion matrix of case 2 arises from the mixed model

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\tau} + \mathbf{X}\boldsymbol{\gamma} + \mathbf{Z}\boldsymbol{\xi} + \mathbf{e} \tag{74}$$

where $\boldsymbol{\gamma}$, $\boldsymbol{\xi}$, $\mathbf{e}$ are all uncorrelated random vectors with zero expectations and dispersion matrices $\boldsymbol{\Gamma}$, $\boldsymbol{\Theta}$, and $\sigma^2\mathbf{I}$, respectively. These $\boldsymbol{\Gamma}$, $\boldsymbol{\Theta}$, and σ^2 are all unknown, but an independent estimate $f^{-1}\mathbf{S}$ of $\boldsymbol{\Sigma}$ is available. By choosing $\boldsymbol{\Gamma}$ or $\boldsymbol{\Theta}$ or both to be zero we obtain special cases.

It was seen in earlier sections that the problem considered is essentially one of making covariance adjustments in the estimator $(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}$ using the concomitant variables $\mathbf{Z}'\mathbf{Y}$. The decrease in efficiency arises when the association between the estimator $(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}$ and the concomitant variables $\mathbf{Z}'\mathbf{Y}$ is weak. Our aim is then to make a selection of suitable concomitant variables or their functions for covariance adjustment on the basis of a given structure for $\boldsymbol{\Sigma}$. It is clear that any selection made on the basis of observed association (using $f^{-1}\mathbf{S}$) does not improve the situation, for the effect of such a selection has to be considered in the estimation of precision of the adjusted estimator.

In the case of model (74), we find that

$$C[(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}, \mathbf{Z}'\mathbf{Y}] = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\Sigma}\mathbf{Z} = \mathbf{0}, \tag{75}$$

and therefore, there is definite loss in efficiency by covariance adjustment. Indeed $\mathbf{T}_1 = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}$ is the least squares estimate of $\boldsymbol{\tau}$ since $\boldsymbol{\Sigma}$ has the structure of lemma 5a. Now

$$\mathbf{T}_1 \sim N_k(\boldsymbol{\tau}, \mathbf{H}) \tag{76}$$

where $\mathbf{H} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\Sigma}\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}$. For drawing inferences on $\boldsymbol{\tau}$, we need an estimate of $\mathbf{H}$. This is supplied by

$$\hat{\mathbf{H}} = f^{-1}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{S}\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} \tag{77}$$

with the distribution

$$f\hat{\mathbf{H}} \sim W_k(f, \mathbf{H}). \tag{78}$$

Hence the inference on $\boldsymbol{\tau}$ follows on standard lines using the distributions (76) and (78), as indicated in section 3.

7. Case 3: $D(\mathbf{Y}) = \mathbf{C}\boldsymbol{\Gamma}\mathbf{C}' + \sigma^2\mathbf{I}$ (C known)

Such a situation arises when we consider a mixed model

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\tau} + \mathbf{C}\boldsymbol{\gamma} + \mathbf{e} \tag{79}$$

where $\boldsymbol{\gamma}$ and $\mathbf{e}$ are uncorrelated, $E(\boldsymbol{\gamma}) = \mathbf{0}$, $D(\boldsymbol{\gamma}) = \boldsymbol{\Gamma}$ and $E(\mathbf{e}) = \mathbf{0}$, $D(\mathbf{e}) = \sigma^2\mathbf{I}$. The matrices **X** and **C** are known. Such a model was studied by a number of authors (see Duncan [2]) under a different set of assumptions on $\boldsymbol{\Gamma}$. In the present problem, lemma 7a provides a first reduction of the problem.

LEMMA 7a. *Let* **B** *be a matrix such that* $\mathbf{X}'(\mathbf{C} - \mathbf{X}\mathbf{B}) = \mathbf{0}$; *that is,*

$$\mathbf{B} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{C},$$

and let the rank of $\mathbf{C} - \mathbf{X}\mathbf{B}$ *be* m. *Consider the linear functions*

$$\begin{aligned} \mathbf{T}_1 &= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}, \\ \mathbf{T}_2 &= (\mathbf{C} - \mathbf{X}\mathbf{B})'\mathbf{Y}, \\ \mathbf{T}_3 &= \mathbf{G}'\mathbf{Y}, \quad \text{where} \quad \mathbf{G}'\mathbf{X} = \mathbf{0}, \quad \mathbf{G}'(\mathbf{C} - \mathbf{X}\mathbf{B}) = \mathbf{0}, \end{aligned} \tag{80}$$

which provide a linear transformation of **Y**. *Then* $E(\mathbf{T}_3) = \mathbf{0}$, $C(\mathbf{T}_1, \mathbf{T}_3) = \mathbf{0}$, *and* $C(\mathbf{T}_2, \mathbf{T}_3) = \mathbf{0}$. *Further,* $\mathbf{T}_1$ *and* $\mathbf{T}_2$ *are correlated, unless* $\mathbf{C}'\mathbf{X} = \mathbf{0}$ *or* $\mathbf{C} - \mathbf{X}\mathbf{B} = \mathbf{0}$.

The results are easy to verify. Lemma 7a shows that $\mathbf{T}_3$ does not throw any information on $\boldsymbol{\tau}$ and should be discarded. Since $E(\mathbf{T}_1) = \boldsymbol{\tau}$, $E(\mathbf{T}_2) = \mathbf{0}$, and $\mathbf{T}_1$ and $\mathbf{T}_2$ are possibly correlated, covariance adjustment in $\mathbf{T}_1$ using $\mathbf{T}_2$ as concomitant variable might provide good estimators. The estimated dispersion matrix of $\mathbf{T}_1$, $\mathbf{T}_2$ is

$$\mathbf{U} = f^{-1}\begin{pmatrix} (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{S}\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} & (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{S}(\mathbf{C} - \mathbf{X}\mathbf{B}) \\ (\mathbf{C} - \mathbf{X}\mathbf{B})'\mathbf{S}\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} & (\mathbf{C} - \mathbf{X}\mathbf{B})'\mathbf{S}(\mathbf{C} - \mathbf{X}\mathbf{B}) \end{pmatrix} \tag{81}$$

with the distribution

$$f\mathbf{U} \sim W_{k+m}(f, \boldsymbol{\Lambda}) \tag{82}$$

where $\boldsymbol{\Lambda}$ is the true dispersion matrix of $(\mathbf{T}_1, \mathbf{T}_2)$. The inference on $\boldsymbol{\tau}$, then proceeds on standard lines making covariance adjustment in $(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}$ using the concomitant variable $(\mathbf{C} - \mathbf{X}\mathbf{B})'\mathbf{Y}$ and the estimated dispersion matrix (81).

Note 1. The difference between case 1 and case 3 is that $(\mathbf{C} - \mathbf{X}\mathbf{B})'\mathbf{Y}$ constitutes a subset of all linear functions of $\mathbf{Z}'\mathbf{Y}$. In case 1, no selection out of $\mathbf{Z}'\mathbf{Y}$ was possible, as nothing was known about the structure of $\boldsymbol{\Sigma}$. The structure for $\boldsymbol{\Sigma}$, as in case 3, enabled us to choose suitable functions of $\mathbf{Z}'\mathbf{Y}$ for covariance adjustment.

Note 2. The validity of the structure for $\boldsymbol{\Sigma}$ as in case 3 can be examined on the basis of S by testing the equality of the last $(p - b)$ eigenvectors of **S**, where b is the rank of **C**. Appropriate test criteria for this purpose have been given by Bartlett [1] and Rao [10].

8. Case 4: $D(\mathbf{Y}) = \mathbf{C}\boldsymbol{\Gamma}\mathbf{C}' + \sigma^2\mathbf{I}$ (C unknown)

Let us first consider the case of $\boldsymbol{\Sigma}$ known. Then we can determine the restricted eigenvalues and eigenvectors of $\boldsymbol{\Sigma}$ subject to the condition that the eigenvectors are orthogonal to the columns of $\mathbf{X}$, as considered in lemmas 2d and 2e of section 2. The restricted eigenvalues and eigenvectors are the nonzero eigenvalues and their corresponding eigenvectors of the matrix $(\mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}')\boldsymbol{\Sigma}$.

For the special choice $\boldsymbol{\Sigma} = \mathbf{C}\boldsymbol{\Gamma}\mathbf{C}' + \sigma^2\mathbf{I}$, we have the following results. The smallest nonzero eigenvalue of $(\mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}')\boldsymbol{\Sigma}$ is σ^2. Let the multiplicity of this root be m and represent the corresponding eigenvectors by $\mathbf{G}_1, \cdots, \mathbf{G}_m$. Let $\mathbf{G}$ be the matrix with $\mathbf{G}_i$ as its columns.

There are $(p - k)$ nonzero eigenvalues on the total. Let the eigenvectors corresponding to the eigenvalues different from σ^2 be $\mathbf{B}_1, \cdots, \mathbf{B}_{p-k-m}$. Let $\mathbf{B}$ be the matrix with $\mathbf{B}_i$ as its columns.

With $\mathbf{B}$ and $\mathbf{G}$ as defined above, we have from lemma 2e

$$\mathbf{X}'\mathbf{G} = \mathbf{X}'\boldsymbol{\Sigma}\mathbf{G} = \mathbf{0}, \qquad \mathbf{B}'\mathbf{G} = \mathbf{B}'\boldsymbol{\Sigma}\mathbf{G} = \mathbf{0}. \tag{83}$$

The conditions (83) imply that the linear functions

$$\mathbf{T}_3 = \mathbf{G}'\mathbf{Y} \tag{84}$$

have zero expectation and are uncorrelated with the linear functions

$$\mathbf{T}_1 = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}, \qquad \mathbf{T}_2 = \mathbf{B}'\mathbf{Y}. \tag{85}$$

Further, $E(\mathbf{T}_2) = \mathbf{0}$, but $\mathbf{T}_2$ is possibly correlated with $\mathbf{T}_1$. Hence, the best estimate of $\boldsymbol{\tau}$ can be obtained by making covariance adjustment in $\mathbf{T}_1$ using $\mathbf{T}_2$ only as concomitant variables (that is, discarding $\mathbf{T}_3$).

If $\boldsymbol{\Sigma}$ is unknown, we proceed as follows. First determine the eigenvalues and eigenvectors of the matrix $(\mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}')\mathbf{S}$ and choose all the (say q) eigenvectors corresponding to dominant roots. Let $\mathbf{B}_1, \cdots, \mathbf{B}_q$ be the eigenvectors chosen and $\mathbf{B}$ the matrix with $\mathbf{B}_i$ as columns. Now consider the linear functions:

$$\mathbf{T}_1 = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}, \qquad \mathbf{T}_2 = \mathbf{B}'\mathbf{Y}. \tag{86}$$

Note that since $\mathbf{B}$ *is an estimated matrix, no exact theory exists for making covariance adjustment using* $\mathbf{T}_2$. For covariance adjustment we consider the estimated dispersion matrix of $\mathbf{T}_1$, $\mathbf{T}_2$, treating $\mathbf{B}$ as fixed, which is

$$f^{-1}\begin{pmatrix} (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{S}\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} & (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{S}\mathbf{B} \\ \mathbf{B}'\mathbf{S}\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} & \mathbf{B}'\mathbf{S}\mathbf{B} \end{pmatrix}. \tag{87}$$

The basic theory of covariance adjustment (in $\mathbf{T}_1$ using $\mathbf{T}_2$ as a concomitant variable) and expressions for the precision of the estimates as discussed in section 3 are applicable, in an approximate way. Note that $\mathbf{B}'\mathbf{Y}$ is selected not on the basis of the observed association with $(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}$, but in a manner which does not overestimate precision due to covariance adjustment. The value of q is taken as the number of eigenroots of $(\mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X})\mathbf{S}$ judged to be dominant by an appropriate test, if necessary.

9. Case 5: Autoregressive errors

The model assumed is

$$y_t = P_k(t) + e_t, \qquad t = 1, \cdots, p \tag{88}$$

where $P_k(t)$ is a k-th degree polynomial in time

$$P_k(t) = \beta_0 + \beta_1 t + \cdots + \beta_k t^k, \tag{89}$$

and e_t have the autoregressive scheme

$$e_t = \rho_1 e_{t-1} + \cdots + \rho_m e_{t-m} + \eta_t. \tag{90}$$

The proposed method of estimation of the coefficients of the polynomial trend and the autoregressive parameters $\rho_1, \cdots, \rho_m$ is an extension of the least squares method considered by Mann and Wald [6]. The estimating equations are obtained by minimizing $\Sigma\eta_t^2$ where

(91)

$$\begin{aligned}\eta_t &= e_t - \rho_1 e_{t-1} - \cdots - \rho_m e_{t-m}\\ &= y_t - P_k(t) - \rho_1[y_{t-1} - P_k(t-1)] - \cdots - \rho_m[y_{t-m} - P_k(t-m)]\\ &= y_t - \rho_1 y_{t-1} - \cdots - \beta_0[1 - \rho_1 - \cdots] - \beta_1[t - \rho_1(t-1) - \cdots] - \cdots.\end{aligned}$$

When ρ_i and β_j are all unknown, η_t is nonlinear in the parameters, and therefore, the estimating equations obtained by minimizing $\Sigma\eta_t^2$ become nonlinear. Fortunately the problem can be reduced to yield a definitive solution. First we transform the parameters in such a way that η_t is linear in the new parameters. Thus we can write

$$\eta_t = y_t - \rho_1 y_{t-1} - \cdots - \rho_m y_{t-m} - \gamma_0 - \gamma_1 t - \cdots - \gamma_k t^k, \tag{92}$$

which is linear in ρ_i and γ_j, where γ_j are defined as follows. Let

$$\begin{aligned}\delta_i &= \rho_1 + 2^i\rho_2 + \cdots + m^i\rho_m \qquad \text{for } i \geq 1\\ \delta_0 &= 1 - \rho_1 - \cdots - \rho_n\end{aligned} \tag{93}$$

Then

$$\begin{aligned}\gamma_k &= \beta_k\delta_0\\ \gamma_{k-1} &= \binom{k}{1}\beta_k\delta_1 + \beta_{k-1}\delta_0\\ \gamma_{k-2} &= -\binom{k}{2}\beta_k\delta_2 + \binom{k-1}{1}\beta_{k-1}\delta_1 + \beta_{k-2}\delta_0\\ &\cdots\cdots\\ \gamma_{k-i} &= (-1)^{i+1}\binom{k}{i}\beta_k\delta_i + (-1)^i\binom{k-1}{i-1}\beta_{k-1}\delta_{i-1} + \cdots\\ &\cdots\cdots\\ \gamma_0 &= (-1)^{k+1}\beta_k\delta_k + (-1)^k\beta_{k-1}\delta_{k-1} + \cdots.\end{aligned} \tag{94}$$

First we estimate ρ_i and γ_j by minimizing

$$\sum_{m+1}^{p} \eta_t^2 = \sum_{m+1}^{p} (y_t - \rho_1 y_{t-1} - \cdots - \gamma_0 - \gamma_1 t - \cdots)^2 \tag{95}$$

with respect to ρ_i, $i = 1, \cdots, m$, and γ_j, $j = 1, \cdots, k$. Let $\hat{\rho}_i$, $\hat{\gamma}_j$ be the solutions of the normal equations. Then we obtain $\hat{\delta}_i$ from the equation (93), and then successively $\hat{\beta}_k, \hat{\beta}_{k-1}, \cdots, \hat{\beta}_0$ from the equations (94).

The normal equations obtained by minimizing (95) provide consistent estimators of ρ_i, γ_j, and the least sum of squares a consistent estimator of σ^2. The inverse of the matrix of normal equations multiplied by the estimator of σ^2 provides the asymptotic dispersion matrix of the estimators $\hat{\rho}_i$, $\hat{\gamma}_j$. We observe that $\hat{\beta}_r$ are simple functions of $\hat{\rho}_i$, $\hat{\gamma}_j$ and, therefore, the asymptotic dispersion matrix of $\hat{\beta}_r$ can be computed by the well known formula (see [13], p. 322).

We shall give the explicit formula for the estimated dispersion matrix of $\hat{\beta}_k, \cdots, \hat{\beta}_0$ in terms of the estimated dispersion matrix of $\hat{\gamma}_k, \cdots, \hat{\gamma}_0; \hat{\rho}_1, \cdots, \hat{\rho}_m$,

$$\hat{\sigma}^2 \begin{pmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{A}_{21} & \mathbf{A}_{22} \end{pmatrix}. \tag{96}$$

Note that to obtain the matrix in (96), it is convenient to write down the normal equations by deriving (95) with respect to $\gamma_k, \cdots, \gamma_0, \rho_1, \cdots, \rho_m$ in the order indicated. Let **B** be the matrix of coefficients of $\beta_k, \cdots, \beta_0$ on the right-hand side of (94), and **C** be the matrix of coefficients of $\delta_0, \cdots, \delta_k$ on the right-hand side of (94). Finally, let **F** be the matrix

$$\begin{pmatrix} -1 & -1 & \cdots & -1 \\ 1 & 2 & \cdots & m \\ \cdot & \cdot & \cdots & \cdot \\ 1 & 2^k & \cdots & m^k \end{pmatrix}. \tag{97}$$

Then we have the relation connecting the differentials $d\hat{\boldsymbol{\gamma}}$, $d\hat{\boldsymbol{\rho}}$ and $d\hat{\boldsymbol{\beta}}$ as

$$d\hat{\boldsymbol{\gamma}} - \mathbf{CF}\, d\hat{\boldsymbol{\rho}} = \mathbf{B}\, d\hat{\boldsymbol{\beta}} \tag{98}$$

where the vectors $\hat{\boldsymbol{\gamma}}$, $\hat{\boldsymbol{\rho}}$, and $\hat{\boldsymbol{\beta}}$ are

$$\hat{\boldsymbol{\gamma}} = \begin{pmatrix} \hat{\gamma}_k \\ \cdot \\ \cdot \\ \cdot \\ \hat{\gamma}_0 \end{pmatrix}, \quad \hat{\boldsymbol{\rho}} = \begin{pmatrix} \hat{\rho}_1 \\ \cdot \\ \cdot \\ \cdot \\ \hat{\rho}_m \end{pmatrix}, \quad \hat{\boldsymbol{\beta}} = \begin{pmatrix} \hat{\beta}_k \\ \cdot \\ \cdot \\ \cdot \\ \hat{\beta}_0 \end{pmatrix}. \tag{99}$$

From (98),

$$d\hat{\boldsymbol{\beta}} = \mathbf{B}^{-1}\, d\hat{\boldsymbol{\gamma}} - \mathbf{B}^{-1}\mathbf{CF}\, d\hat{\boldsymbol{\rho}} = \mathbf{H}\, d\hat{\boldsymbol{\gamma}} - \mathbf{G}\, d\hat{\boldsymbol{\rho}} \tag{100}$$

writing $\mathbf{H} = \mathbf{B}^{-1}$ and $\mathbf{G} = \mathbf{B}^{-1}\mathbf{CF}$. Then the asymptotic dispersion matrix of $\hat{\boldsymbol{\beta}}$ is

$$D(\hat{\boldsymbol{\beta}}) = \hat{\sigma}^2(\mathbf{HA}_{11}\mathbf{H}' + \mathbf{GA}_{22}\mathbf{G}' - \mathbf{HA}_{12}\mathbf{G}' - \mathbf{GA}_{21}\mathbf{H}'). \tag{101}$$

Let us note that, in practice, the order of the autoregressive scheme for the errors may not be preassigned and may have to be inferred from data. If the chosen order is higher than the true one, then an application of lemma 2a shows that the estimators of γ_i and ρ_j lose in efficiency. For determining the appropriate order, the residual sum of squares providing the estimate of σ^2 has to be examined. If there is no significant reduction in the residual sum of squares by increasing the number of autoregressive parameters, then a lower order is indicated.

We observe that the linearization of the parameters in η_t is not possible if the nonstochastic part in y_t is not a polynomial in t, but is simply linear in the unknown parameters β_j with given coefficients. An appropriate method in such a case is given by Durbin [3].

10. Estimation of phase and amplitude

In the estimation of signal parameters, models of the type (1) are used. Thus, if β and θ represent amplitude and phase of a signal, the t-th observation is written

$$y_t = x_{t1}\beta \sin \theta + x_{t2}\beta \cos \theta + e_t, \qquad t = 1, \cdots, p \tag{102}$$

where x_{ti} are known coefficients. Writing $\tau_1 = \beta \sin \theta$, $\tau_2 = \beta \cos \theta$, and $\boldsymbol{\tau}' = (\tau_1, \tau_2)$, $\mathbf{X} = (x_{ti})$, $\mathbf{Y}' = (y_1, \cdots, y_p)$, and $\mathbf{e}' = (e_1, \cdots, e_p)$, we have the linear model

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\tau} + \mathbf{e}. \tag{103}$$

Let us suppose that under suitable assumption on $\boldsymbol{\Sigma}$, estimates $\hat{\tau}_1$, $\hat{\tau}_2$ of τ_1, τ_2 have been obtained as discussed in the present paper. Further, let

$$\begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} \tag{104}$$

be the estimated dispersion matrix of the estimates $\hat{\tau}_1$, $\hat{\tau}_2$.

We are interested in estimating β and θ which are nonlinear functions of τ_1 and τ_2. We may estimate β and θ by

$$\begin{aligned} \hat{\beta} &= (\hat{\tau}_1^2 + \hat{\tau}_2^2)^{1/2}, \\ \hat{\theta} &= \tan^{-1} (\hat{\tau}_1/\hat{\tau}_2), \end{aligned} \tag{105}$$

and compute their large sample standard errors (or asymptotic variance). Using the formula (see [13], p. 322) for asymptotic variance

$$\begin{aligned} V(\hat{\beta}) &= (a_{11}\hat{\tau}_1^2 + 2a_{12}\hat{\tau}_1\hat{\tau}_2 + a_{22}\hat{\tau}_2^2)/(\hat{\tau}_1^2 + \hat{\tau}_2^2), \\ V(\hat{\theta}) &= (a_{11}\hat{\tau}_2^2 - 2a_{12}\hat{\tau}_1\hat{\tau}_2 + a_{22}\hat{\tau}_1^2)/(\hat{\tau}_1^2 + \hat{\tau}_2^2)^2. \end{aligned} \tag{106}$$

Exact confidence intervals of a given probability level $(1 - \alpha)$ can be obtained for $\lambda = \tan \theta$ (and hence for θ), provided that under the assumptions made on the errors there exists an exact test for the linear hypothesis $\tau_1 - \lambda\tau_2 = 0$. In all cases, except that of autoregressive errors, such an exact test is possible using a t distribution on appropriate degrees of freedom. Then the confidence limits for λ are obtained by solving the equation in λ

$$\frac{(\hat{\tau}_1 - \lambda\hat{\tau}_2)^2}{a_{11} - 2\lambda a_{12} + \lambda^2 a_{22}} = t_{\alpha/2} \tag{107}$$

where $t_{\alpha/2}$ is the upper $(\alpha/2)$ probability value of t on degrees of freedom applicable for a t test of a linear hypothesis on $\boldsymbol{\tau}$. When an exact test is not available, we may use the upper $(\alpha/2)$ probability value of the standard normal distribution in the place of $t_{\alpha/2}$ in (107). Thus we obtain asymptotic confidence limits for λ from which those of θ can be computed.

There is, however, no exact method for determining confidence intervals for β, even when exact tests of linear hypotheses on $\boldsymbol{\tau}$ are possible. One may have to use the estimator for β as in (105), and the asymptotic variance $V(\hat{\beta})$ as in (106), to obtain an asymptotic confidence interval.

Simultaneous confidence intervals for β and θ may be deduced from a confidence ellipsoid (exact or asymptotic) of τ_1 and τ_2. Let

$$a^{11}(\hat{\tau}_1 - \tau_1)^2 + 2a^{12}(\hat{\tau}_1 - \tau_1)(\hat{\tau}_2 - \tau_2) + a^{22}(\hat{\tau}_2 - \tau_2)^2 \leq c \tag{108}$$

be the confidence ellipsoid of τ_1, τ_2 with a given probability $(1 - \alpha)$. The confidence interval for β with probability $\geq 1 - \alpha$ is

$$\left(\sqrt{\min\,(\tau_1^2 + \tau_2^2)},\ \sqrt{\max\,(\tau_1^2 + \tau_2^2)}\right) \tag{109}$$

where the minimum and maximum are obtained subject to the restriction (108) on τ_1 and τ_2 for given $\hat{\tau}_1$ and $\hat{\tau}_2$. A satisfactory solution to such a problem of finding the extrema of a quadratic form on an ellipsoid in many dimensions has been recently given by Forsyth and Golub [4]. The confidence interval for θ is given by the angles which the pair of tangents from the origin to the ellipsoid (108) make with the τ_2 axis.

REFERENCES

[1] M. S. Bartlett, "Tests of significance in factor analysis," *British J. Psych.* (Statist. Sec.), Vol. 3 (1950), pp. 77–85.

[2] D. B. Duncan, "On some theory of mixed linear models and its application in the estimation of missile trajectory," paper No. 332 from the Department of Biostatistics, The Johns Hopkins University.

[3] J. Durbin, "Estimation of parameters in time-series regression models," *J. Roy. Statist. Soc.*, Vol. 22 (1960), pp. 139–153.

[4] G. E. Forsyth and G. Golub, "On the stationary values of a second-degree polynomial on the unit sphere," *J. SIAM*, to appear.

[5] C. R. Henderson, Oscar Kempthorne, S. R. Searle, and C. M. Krosigk, "The estimation of environmental and genetic trends from records subject to culling," *Biometrics*, Vol. 15 (1959), pp. 192–218.

[6] H. B. Mann and A. Wald, "On the statistical treatment of linear stochastic difference equations," *Econometrics*, Vol. 11 (1943), pp. 173–220.

[7] C. R. Rao, "Tests with discriminant functions in multivariate analysis," *Sankhyā*, Vol. 7 (1946), pp. 407–414.

[8] ———, "On some problems arising out of discrimination with multiple characters," *Sankhyā*, Vol. 9 (1949), pp. 343–364.

[9] ———, "A note on the distribution of $D_{p+q}^2 - D_p^2$ and some computational aspects of D^2 statistic and discriminant function," *Sankhyā*, Vol. 10 (1950), pp. 257–268.

[10] ———, "Estimation and tests of significance in factor analysis," *Psychometrika*, Vol. 20 (1955), pp. 93–111.

[11] ———, "Some problems involving linear hypotheses in multivariate analysis," *Biometrika*, Vol. 46 (1959), pp. 49–58.

[12] ———, "The use and interpretation of principal component analysis in applied research," *Sankhyā*, Vol. 26 (1964), pp. 329–358.

[13] ———, *Linear Statistical Inference and Its Applications*, New York, Wiley, 1965.

[14] ———, "Generalized inverses for matrices and its applications in mathematical statistics," *Research Papers in Statistics* (Festschrift volume for J. Neyman), pp. 263–280, New York, Wiley, 1966.