Computational Statistics & Data Analysis 12 (1991) 221–230
North-Holland

Covariance adjustment in biased estimation

Jerzy K. Baksalary

Tadeusz Kotarbiński Pedagogical University, 65-069 Zielona Góra and Academy of Agriculture in Poznań, 60-637 Poznań, Poland

Götz Trenkler

University of Dortmund, 4600 Dortmund 50, Germany

Received March 1989
Revised February 1990

Abstract: The covariance adjustment technique is usually applied to optimal unbiased estimation of a vector parameter $\boldsymbol{\theta} \in \mathscr{A}$ via linearly combining an unbiased estimator of $\boldsymbol{\theta}$, $\boldsymbol{T}_1$ say, and an unbiased estimator of a zero vector, $\boldsymbol{T}_2$ say, with the use of a known joint dispersion matrix of $\boldsymbol{T}_1$ and $\boldsymbol{T}_2$. In this paper, it is shown that covariance adjusted estimators may be regarded as competitors to the best linear unbiased estimator of $\boldsymbol{\theta}$ based on $\boldsymbol{T}_1$ alone, $\boldsymbol{T}_1(\mathscr{A})$, also when the expectation of $\boldsymbol{T}_2$ is a nonzero vector, in which case the covariance adjusted estimators are biased. Their admissibility among all linear combinations of $\boldsymbol{T}_1$ and $\boldsymbol{T}_2$ with respect to the mean-square-error criterion is examined and then all values of the parameters for which an admissible covariance adjusted estimator behaves at least as well as $\boldsymbol{T}_1(\mathscr{A})$ with respect to the mean-square-error-matrix criterion are characterized. The comparison between two admissible covariance adjusted estimators with respect to the latter criterion is also considered.

1. Introduction and preliminaries

Suppose we wish to estimate a $k \times 1$ vector of parameters $\boldsymbol{\theta}$ having available two statistics: $\boldsymbol{T}_1$, which is an unbiased estimator of $\boldsymbol{\theta}$, and $\boldsymbol{T}_2$, which takes values in the set of $m \times 1$ real matrices, $\mathbb{R}^{m \times 1}$, with expectation $E(\boldsymbol{T}_2) = \boldsymbol{\tau}$. The problem consists in combining the information contained in the two statistics when their joint dispersion matrix,

$$D\begin{pmatrix} \boldsymbol{T}_1 \\ \boldsymbol{T}_2 \end{pmatrix} = \sigma^2 \begin{pmatrix} \boldsymbol{V}_{11} & \boldsymbol{V}_{12} \\ \boldsymbol{V}_{12}' & \boldsymbol{V}_{22} \end{pmatrix} = \sigma^2 \boldsymbol{V}, \tag{1.1}$$

is positive definite and known entirely or except for a positive scalar σ^2.

An estimator of $\boldsymbol{\theta}$, in which the knowledge of $\boldsymbol{V}$ is reflected in the form

$$\boldsymbol{T}_0 = \boldsymbol{T}_1 - \boldsymbol{V}_{12}\boldsymbol{V}_{12}^{-1}\boldsymbol{T}_2, \tag{1.2}$$

is called the covariance adjusted estimator; cf. Rao (1967, Section 3) and Lewis and Odell (1971, Section 8.3). Its dispersion matrix is

$$D(T_0) = \sigma^2 S, \qquad \text{where} \tag{1.3}$$

$$S = V_{11} - V_{12}V_{22}^{-1}V_{12}' \tag{1.4}$$

is the Schur complement of V_{22} in V partitioned as in (1.1). From (1.1) and (1.3) it follows that $D(T_0)$ compares favourably with $D(T_1)$ in the sense that

$$D(T_1) - D(T_0) = \sigma^2 V_{12}V_{22}^{-1}V_{12}'$$

is a nonnegative definite matrix. In other words, $D(T_0)$ is below $D(T_1)$ with respect to the Löwner partial ordering, which will henceforth be denoted by $D(T_0) \leq D(T_1)$. Actually, it is known [cf. Baksalary and Kala (1982, p. 281)] that if $\tau = 0$ and θ is free to vary over the entire $\mathbb{R}^{k\times 1}$, then T_0 is the best (i.e. minimum dispersion) unbiased estimator of θ among the set

$$\mathscr{T} = \{L_1T_1 + L_2T_2 : L_1 \in \mathbb{R}^{k\times k},\ L_1 \in \mathbb{R}^{k\times m}\} \tag{1.5}$$

of all linear combinations of T_1 and T_2. See also Rao (1966, Section 6) for a discussion of the applicability of the covariance adjustment technique to estimating the matrix of parameters in the model of Potthoff and Roy (1964).

Baksalary and Kala (1979) discussed an extension of the covariance adjustment to the case where θ is known to lie in a given subspace, $\mathscr{A}$ say, of $\mathbb{R}^{k\times 1}$. (Such additional information may for instance be available from restrictions of the form $R\theta = 0$, in which case $\mathscr{A}$ coincides with the null space of R.) If $\tau = 0$ and the variability of θ is constrained to $\mathscr{A}$, then the best unbiased estimator of θ among the set (1.5) is

$$T_0(\mathscr{A}) = A(A'S^{-1}A)^{-}A'S^{-1}T_0, \tag{1.6}$$

where A is any matrix whose range $\mathscr{R}(A)$ coincides with $\mathscr{A}$, the minus superscript denotes any generalized inverse in the sense of Rao (1962), and T_0 is as defined in (1.2). See also Baksalary and Kala (1983a, Section 3) for the corresponding result when V specified in (1.1) may be singular.

The present paper deals with a family of estimators of the form (1.6), viz.

$$T_0(\mathscr{F}) = F(F'S^{-1}F)^{-}F'S^{-1}T_0, \tag{1.7}$$

where F is any matrix such that $\mathscr{F} = \mathbb{R}(F) \subseteq \mathscr{A}$. It is shown that also in the cases where $E(T_2) \neq 0$, the covariance adjusted estimators (1.7) may be regarded as competitors to the best linear unbiased estimator of θ based on T_1 alone, i.e.,

$$T_1(\mathscr{A}) = A(A'V_{11}^{-1}A)^{-}A'V_{11}^{-1}T_1. \tag{1.8}$$

Since this actually corresponds to the classical situation in which an estimator is improved with respect to the dispersion matrix at the cost of introducing a bias, the conclusion above is not unexpected, but the contribution of the paper is to make it precise.

This goal is achieved by examining the admissibility of covariance adjusted estimators (1,7) among the class (1.5) with respect to the mean-square-error criterion and by determining those values of $\boldsymbol{\theta}$, $\boldsymbol{\tau}$, and σ^2 for which, from the point of view of the mean-square-error-matrix criterion, these estimators behave at least as well as the best linear unbiased estimator based on $\boldsymbol{T}_1$ alone. The latter criterion is adopted also to compare two estimators from the family (1.7). It is defined as $M(\hat{\boldsymbol{\theta}}) = E[(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})']$, and hence

$$M(\hat{\boldsymbol{\theta}}) = D(\hat{\boldsymbol{\theta}}) + b(\hat{\boldsymbol{\theta}})b(\hat{\boldsymbol{\theta}})', \tag{1.9}$$

where $b(\hat{\boldsymbol{\theta}}) = E(\hat{\boldsymbol{\theta}}) - \boldsymbol{\theta}$ denotes the bias vector. Then, for given values of the parameters involved, $\boldsymbol{K}_1\boldsymbol{T}_1 + \boldsymbol{K}_2\boldsymbol{T}_2 \in \mathscr{T}$ is said to dominate $\boldsymbol{L}_1\boldsymbol{T}_1 + \boldsymbol{L}_2\boldsymbol{T}_2 \in \mathscr{T}$ if the mean-square-error-matrices corresponding to these estimators are ordered as $M(\boldsymbol{K}_1\boldsymbol{T}_1 + \boldsymbol{K}_2\boldsymbol{T}_2) \leq M(\boldsymbol{L}_1\boldsymbol{T}_1 + \boldsymbol{L}_2\boldsymbol{T}_2)$ in the Löwner sense. The criterion adopted for evaluating the admissibility is $MSE(\hat{\boldsymbol{\theta}}) = E[(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})'(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})]$, so that

$$MSE(\hat{\boldsymbol{\theta}}) = \text{trace } M(\hat{\boldsymbol{\theta}}) = \text{trace } D(\hat{\boldsymbol{\theta}}) + b(\hat{\boldsymbol{\theta}})'b(\hat{\boldsymbol{\theta}}). \tag{1.10}$$

Then an estimator $\boldsymbol{K}_1\boldsymbol{T}_1 + \boldsymbol{K}_2\boldsymbol{T}_2$ is said to be admissible for $\boldsymbol{\theta}$ among $\mathscr{F}$ if there does not exist $\boldsymbol{L}_1\boldsymbol{T}_1 + \boldsymbol{L}_2\boldsymbol{T}_2 \in \mathscr{F}$ such that the inequality $MSE(\boldsymbol{L}_1\boldsymbol{T}_1 + \boldsymbol{L}_2\boldsymbol{T}_2) \leq MSE(\boldsymbol{K}_1\boldsymbol{T}_1 + \boldsymbol{K}_2\boldsymbol{T}_2)$ holds for all values of $\boldsymbol{\theta}$, $\boldsymbol{\tau}$, σ^2, being strict for some $\boldsymbol{\theta}$, $\boldsymbol{\tau}$, σ^2.

The two theorems of this paper refer to the general situation when $\boldsymbol{\theta}$ is assumed to lie in a given subspace $\mathscr{A} = \mathscr{R}(\boldsymbol{A})$ of $\mathbb{R}^{k\times 1}$, i.e., $\boldsymbol{\theta} = \boldsymbol{A}\boldsymbol{\alpha}$. The vector of nuisance parameters $\boldsymbol{\tau}$ is considered free to vary over $\mathbb{R}^{m\times 1}$, so that the corresponding model is

$$\left\{ \begin{pmatrix} \boldsymbol{T}_1 \\ \boldsymbol{T}_2 \end{pmatrix}, \begin{pmatrix} \boldsymbol{A} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{I}_m \end{pmatrix} \begin{pmatrix} \boldsymbol{\alpha} \\ \boldsymbol{\tau} \end{pmatrix}, \sigma^2 \begin{pmatrix} \boldsymbol{V}_{11}, & \boldsymbol{V}_{12} \\ \boldsymbol{V}_{12}' & \boldsymbol{V}_{22} \end{pmatrix} \right\}. \tag{1.11}$$

Notice that the statistics $\boldsymbol{T}_0(\mathscr{A})$ and $\boldsymbol{T}_0(\mathscr{F})$ defined in (1.6) and (1.7) have simple interpretations under the model

$$\left\{ \begin{pmatrix} \boldsymbol{T}_1 \\ \boldsymbol{T}_2 \end{pmatrix}, \begin{pmatrix} \boldsymbol{A} \\ \boldsymbol{0} \end{pmatrix} \boldsymbol{\alpha}, \sigma^2 \begin{pmatrix} \boldsymbol{V}_{11}, & \boldsymbol{V}_{12} \\ \boldsymbol{V}_{12}' & \boldsymbol{V}_{22} \end{pmatrix} \right\}, \tag{1.12}$$

or, equivalently, under the model (1.11) with the restrictions $\boldsymbol{\tau} = \boldsymbol{0}$. Each statistic $\boldsymbol{T}_0(\mathscr{F})$ is the S^{-1} orthogonal projection of $\boldsymbol{T}_0(\mathscr{A})$ on $\mathscr{F} \subseteq \mathscr{A}$, and $\boldsymbol{T}_0(\mathscr{A})$ is the best linear unbiased estimator of $\boldsymbol{\theta} = \boldsymbol{A}\boldsymbol{\alpha}$, which may be represented alternatively in the form

$$\boldsymbol{T}_0(\mathscr{A}) = \boldsymbol{A}(\tilde{\boldsymbol{A}}'\boldsymbol{V}^{-1}\tilde{\boldsymbol{A}})^{-}\tilde{\boldsymbol{A}}'\boldsymbol{V}^{-1}\boldsymbol{T},$$

where $\tilde{\boldsymbol{A}} = (\boldsymbol{A}' : \boldsymbol{0})$, $\boldsymbol{T}' = (\boldsymbol{T}_1' : \boldsymbol{T}_2')$ and

$$\boldsymbol{V}^{-1} = \begin{pmatrix} \boldsymbol{X}^{-1} & -\boldsymbol{S}^{-1}\boldsymbol{V}_{12}\boldsymbol{V}_{22}^{-1} \\ -\boldsymbol{V}_{22}^{-1}\boldsymbol{V}_{12}'\boldsymbol{S}^{-1} & \boldsymbol{V}_{22}^{-1} + \boldsymbol{V}_{22}^{-1}\boldsymbol{V}_{12}'\boldsymbol{S}^{-1}\boldsymbol{V}_{22}^{-1} \end{pmatrix}.$$

Consequently, the problem of investigating properties of these estimators under the model (1.11) may be viewed as a version of the general problem of validity of estimators developed under the Gauss–Markov model with a misspecified model matrix; cf. Mitra and Rao (1969), Rao and Mitra (1971, Chapter 8), and Mathew and Bhimasankaram (1983).

2. Admissibility

In view of the interpretation given at the end of section 1, admissibility of the estimators $T_1(\mathscr{A})$ and $T_0(\mathscr{F})$ defined in (1.8) and (1.7), respectively, may be examined utilizing the general results on admissible linear estimation in the Gauss–Markov model. The cornerstone of this theory was laid in the article of Cohen (1966), and the fundamental work is included in the 1975 Wald Memorial Lectures of Rao (1976).

In the context of covariance adjusted estimators, Baksalary and Kala (1982, Theorem 1) showed that if $\boldsymbol{\tau} = \boldsymbol{0}$, then $\boldsymbol{K}_1\boldsymbol{T}_1 + \boldsymbol{K}_2\boldsymbol{T}_2$ is admissible for $\boldsymbol{\theta}$ among the class $\mathscr{F}$ defined in (1.5) if ad only if

$$\mathscr{R}(\boldsymbol{S}\boldsymbol{K}_1') \subseteq \mathscr{R}(\boldsymbol{A}),\ \boldsymbol{K}_1\boldsymbol{A}(\boldsymbol{A}'\boldsymbol{S}^{-1}\boldsymbol{A})^{-}\boldsymbol{A}' = \boldsymbol{A}(\boldsymbol{A}'\boldsymbol{S}^{-1}\boldsymbol{A})^{-}\boldsymbol{A}'\boldsymbol{K}_1',$$

$$\boldsymbol{K}_1\boldsymbol{A}(\boldsymbol{A}'\boldsymbol{S}^{-1}\boldsymbol{A})^{-}\boldsymbol{A}'(\boldsymbol{I}_k - \boldsymbol{K}_1') \geq \boldsymbol{0} \quad \text{and} \quad \boldsymbol{K}_2 = -\boldsymbol{K}_1\boldsymbol{V}_{12}\boldsymbol{V}_{22}^{-1},$$

where $\boldsymbol{S}$ is the Schur complement of $\boldsymbol{V}_{22}$ in $\boldsymbol{V}$ as defined in (1.4). In particular, it is hence clear that, except in the trivial case of $\boldsymbol{V}_{12} = \boldsymbol{0}$, the condition $\boldsymbol{\tau} = \boldsymbol{0}$ implies inadmissibility of every linear estimator of $\boldsymbol{\theta}$ based on $\boldsymbol{T}_1$ alone. However, this paper is concerned with the situation where $\boldsymbol{\tau}$ may take any value in $\mathbb{R}^{m\times 1}$. Then the admissibility of estimators is examined under the model (1.11) using Lemma 1 below, which is obtained as a corollary to Theorem 6.6 of Rao (1976).

Lemma 1. *If $E(\boldsymbol{T}_1)$ $\boldsymbol{\theta}$ and $E(\boldsymbol{T}_2) = \boldsymbol{\tau}$ take values in $\mathscr{A} \subseteq \mathbb{R}^{k\times 1}$ and $\mathbb{R}^{m\times 1}$, respectively, then an estimator $\boldsymbol{K}_1\boldsymbol{T}_1 + \boldsymbol{K}_2\boldsymbol{T}_2$ is admissible for $\boldsymbol{\theta}$ among $\mathscr{F} = \{\boldsymbol{L}_1\boldsymbol{T}_1 + \boldsymbol{L}_2\boldsymbol{T}_2 : \boldsymbol{L}_1 \in \mathbb{R}^{k\times k},\ \boldsymbol{L}_2 \in \mathbb{R}^{k\times m}\}$ if and only if*

$$\mathscr{R}(\boldsymbol{K}_1\boldsymbol{V}_{11} + \boldsymbol{K}_2\boldsymbol{V}_{12}') \subseteq \mathscr{A}, \tag{2.1}$$

$$\boldsymbol{K}_1\boldsymbol{V}_{11} + \boldsymbol{K}_2\boldsymbol{V}_{12}' = \boldsymbol{V}_{11}\boldsymbol{K}_1' + \boldsymbol{V}_{12}\boldsymbol{K}_2', \quad \text{and} \tag{2.2}$$

$$(\boldsymbol{K}_1\boldsymbol{V}_{11} + \boldsymbol{K}_2\boldsymbol{V}_{12}')(\boldsymbol{I}_k - \boldsymbol{K}_1') - (\boldsymbol{K}_1\boldsymbol{V}_{12} + \boldsymbol{K}_2\boldsymbol{V}_{22})\boldsymbol{K}_2' \geq \boldsymbol{0}. \tag{2.3}$$

It is seen that for the best linear unbiased estimator of $\boldsymbol{\theta} = \boldsymbol{A}\boldsymbol{\alpha}$ under the submodel $\{\boldsymbol{T}_1, \boldsymbol{A}\boldsymbol{\alpha}, \sigma^2\boldsymbol{V}_{11}\}$, i.e. [cf. (1.8)], when $\boldsymbol{K}_1 = \boldsymbol{A}(\boldsymbol{A}'\boldsymbol{V}_{11}^{-1}\boldsymbol{A})^{-}\boldsymbol{A}'\boldsymbol{V}_{11}^{-1}$ and $\boldsymbol{K}_2 = \boldsymbol{0}$, the conditions (2.1), (2.2), and (2.3) are obviously fulfilled, thus leading to the conclusion that

$$\boldsymbol{T}_1(\mathscr{A}) \text{ is admissible for } \boldsymbol{\theta} \in \mathscr{A} \text{ among } \mathscr{F} \text{ under (1.11).} \tag{2.4}$$

It is also seen that for the covariance adjusted estimators of the type (1.7), i.e., when $\boldsymbol{K}_1 = \boldsymbol{F}(\boldsymbol{F}'\boldsymbol{S}^{-1}\boldsymbol{F})^{-}\boldsymbol{F}'\boldsymbol{S}^{-1}$ and $\boldsymbol{K}_2 = -\boldsymbol{K}_1\boldsymbol{V}_{12}\boldsymbol{V}_{22}^{-1}$, the conditions (2.2) and (2.3) are satisfied for any matrix $\boldsymbol{F}$. Consequently, from the condition (2.1) it follows that

$$\boldsymbol{T}_0(\mathscr{F}) \text{ is admissible for } \boldsymbol{\theta} \in \mathscr{A} \text{ among } \mathscr{F} \text{ under (1.11)}$$

$$\Leftrightarrow \mathscr{F} = \mathscr{R}(\boldsymbol{F}) \subseteq \mathscr{A}. \tag{2.5}$$

The results (2.4) and (2.5) provide a preliminary evaluation of the estimators $T_1(\mathscr{A})$ and $T_0(\mathscr{F})$. Following it, comparisons between $T_1(\mathscr{A})$ and $T_0(\mathscr{F})$ in section 3 and between $T_0(\mathscr{A})$ and $T_0(\mathscr{F})$ in section 4 are discussed under the assumption that $\mathscr{F} \subseteq \mathscr{A}$. It may be mentioned that if the admissibility of $T_1(\mathscr{A})$ and $T_0(\mathscr{F})$ was examined under the misspecified (reduced) model (1.12), then $T_1(\mathscr{A})$ would become inadmissible unless $V_{12} = 0$, whereas $T_0(\mathscr{F})$ would remain admissible in exactly the same cases in which it is admissible under the model (1.11) i.e., when $\mathscr{F} \subseteq \mathscr{A}$.

3. Comparison of $T_1(\mathscr{A})$ with $T_0(\mathscr{F})$

The unbiased estimator of $\theta \in \mathscr{A}$ defined in (1.8) has the dispersion matrix

$$D[T_1(\mathscr{A})] = \sigma^2 A(A'V_{11}^{-1}A)^- A'. \tag{3.1}$$

On the other hand, the biased estimator defined in (1.7) has the dispersion matrix

$$D[T_0(\mathscr{F})] = \sigma^2 F(F'S^{-1}F)^- F' \tag{3.2}$$

and the bias vector

$$b[T_0(\mathscr{F})] = -S^{1/2}\eta(\mathscr{F}), \qquad \text{with} \tag{3.3}$$

$$\eta(\mathscr{F}) = (I_k - P_{S^{-1/2}F})S^{-1/2}\theta + P_{S^{-1/2}F}H\tau, \qquad \text{where} \tag{3.4}$$

$$H = S^{-1/2}V_{12}V_{22}^{-1} \tag{3.5}$$

and $P_{S-1/2F}$ denotes the orthogonal projector on $\mathscr{R}(S^{-1/2}F)$, i.e.,

$$P_{S^{-1/2}F} = S^{-1/2}F(F'S^{-1}F)^- F'S^{-1/2}.$$

In view of (1.9), the difference between the mean-square-error-matrices corresponding to the two estimators compared is

$$M[T_1(\mathscr{A})] - M[T_0(\mathscr{F})] = S^{1/2}[\sigma^2 D - \eta(\mathscr{F})\eta(\mathscr{F})']S^{1/2}, \qquad \text{where} \tag{3.6}$$

$$D = S^{-1/2}A(A'V_{11}^{-1}A)^- A'S^{-1/2} - P_{S^{-1/2}F}. \tag{3.7}$$

The nonnegative definiteness of the difference (3.6) will be evaluated with the use of the following lemma; cf. Baksalary and Kala (1983b, Theorem 1) and Trenkler (1985, Corollary 1).

Lemma 2. *Let C be an $n \times n$ symmetric matrix, let c be an $n \times 1$ vector, and let γ be a positive real number. Then $\gamma C - cc' \geqslant 0$ if and only if $C \geqslant 0$, $c \in \mathscr{R}(C)$, and $c'C^- c \leq \gamma$, were the choice of C^- is arbitrary.*

Theorem 1. *If $E(T_1) = \theta$ and $E(T_2) = \tau$ take values in $\mathscr{A} \subseteq \mathbb{R}^{k \times 1}$ and $\mathbb{R}^{m \times 1}$, respectively, then the covariance estimator $T_0(\mathscr{F})$, with $\mathscr{F} = \mathscr{R}(F) \subseteq \mathscr{R}(A) = \mathscr{A}$,* *dominates the best linear unbiased estimator* $T_1(\mathscr{A})$ *(based on* T_1 *alone) with respect to the mean-square-error-matrix criterion if and only if*

$$\eta(\mathscr{F})'S^{1/2}\left[A\left(A'V_{11}^{-1}A\right)^{-}A' - F\left(F'S^{-1}F\right)^{-}F'\right]^{-}S^{1/2}\eta(\mathscr{F}) \le \sigma^2, \tag{3.8}$$

where $S^{1/2}\eta(\mathscr{F}) = [I_k - F(F'S^{-1}F)^{-}F'S^{-1}]\theta + \mathrm{F}(\mathrm{F}'\mathrm{S}^{-1}\mathrm{F})^{-}\mathrm{F}'\mathrm{S}^{-1}\mathrm{V}_{12}\mathrm{V}_{22}^{-1}\tau$.

Proof. The matrix D in (3.7) may be represented as $D = D_* + D_\#$, where

$$D_* = S^{-1/2}A\left(A'V_{11}^{-1}A\right)^{-}A'S^{-1/2} - S^{-1/2}A\left(A'S^{-1}A\right)^{-}A'S^{-1/2} \quad \text{and} \tag{3.9}$$

$$D_\# = P_{S^{-1/2}A} - P_{S^{-1/2}F}. \tag{3.10}$$

Since $S \le V_{11}$, Theorem 3 in Pukelsheim and Styan (1983) asserts that $(A'S^{-1}A)^{+} \le (A'V_{11}^{-1}A)^{+}$, where the plus superscript denotes the Moore-Penrose inverse. Combining this observation with the invariance of the formula (3.9) with respect to the choice of the generalized inverses involved [cf. Rao and Mitra (1971, Lemma 2.2.4)] shows that $D_* \ge 0$. Further, since $\mathscr{R}(F) \subseteq \mathscr{R}(A)$ entails $\mathscr{R}(S^{-1/2}F) \subseteq \mathscr{R}(S^{-1/2}A)$, it follows that

$$D_\# \text{ is the orthogonal projector on } \mathscr{R}(S^{-1/2}A) \cap \mathscr{R}^{\perp}(S^{-1/2}F) \tag{3.11}$$

[cf., e.g., Rao and Mitra (1971, Theorem 5.1.3)], and thus $D_\# \ge 0$. Consequently $D \le 0$ and

$$\mathscr{R}(D) = \mathscr{R}(D_*) + \mathscr{R}(D_\#) = \mathscr{R}\left[(I_k - D_\#)D_*\right] + \mathscr{R}(D_\#). \tag{3.12}$$

It is known that

$$S^{-1} = V_{11}^{-1} + V_{11}^{-1}V_{12}\,U^{-1}V_{12}'V_{11}^{-1}, \tag{3.13}$$

where $U = V_{22} - V_{12}'V_{11}^{-1}V_{12}$ is the Schur complement of V_{11} in V partitioned as in (1.1). From (3.9 –) and (3.13) we have

$$\begin{aligned}
\mathscr{R}(D_*) &= \mathscr{R}\left[S^{-1/2}A\left(A'V_{11}^{-1}A\right)^{-}A'V_{11}^{-1}A - S^{-1/2}A\left(A'S^{-1}A\right)^{-}A'V_{11}^{-1}A\right] \\
&= \mathscr{R}\left[S^{-1/2}A - S^{-1/2}A\left(A'S^{-1}A\right)^{-}A'S^{-1}A\right. \\
&\qquad \left. + S^{-1/2}A\left(A'S^{-1}A\right)^{-}A'V_{11}^{-1}V_{12}U^{-1}V_{12}'V_{11}^{-1}A\right] \\
&= \mathscr{R}\left[S^{-1/2}A\left(A'S^{-1}A\right)^{-}A'S^{-1}V_{12}V_{22}^{-1}\right] \\
&= \mathscr{R}\left(P_{S^{-1/2}A}H\right),
\end{aligned}$$

where H is as defined in (3.5). Hence $\mathscr{R}[(I_k - D_\#)D_*] = \mathscr{R}(P_{S^{-1/2}F}H)$, which shows that the second component of $\eta(\mathscr{F})$ in (3.4) lies in $\mathscr{R}[(I_k - D_\#)D_*]$. Moreover, since $\theta \in \mathscr{A}$, the first component of $\eta(\mathscr{F})$ may be reexpressed as

$$\left(P_{S^{-1/2}A} - P_{S^{-1/2}F}\right)S^{-1/2}\theta,$$

and thus it is seen that it lies in $\mathscr{R}(\boldsymbol{D}_{\#})$. Consequently, the decomposition (3.12) shows that $\eta(\mathscr{F}) \in \mathscr{R}(\boldsymbol{D})$ and then, in view of (3.6), Lemma 2 implies that $M[\boldsymbol{T}_0(\mathscr{F})] \leq M[\boldsymbol{T}_1(\mathscr{A})]$ if and only if $\eta(\mathscr{F})'\boldsymbol{D}^-\eta(\mathscr{F}) \leq \sigma^2$. This leads straightforwardly to (3.8), since the condition $\eta(\mathscr{F}) \in \mathscr{R}(\boldsymbol{D})$ assures the invariance of $\eta(\mathscr{F})'\boldsymbol{D}^-\eta(\mathscr{F})$ with respect to the choice of $\boldsymbol{D}^-$ and, for any matrix $\boldsymbol{G}$ and nonsingular matrices $\boldsymbol{H}_1$ and $\boldsymbol{H}_2$, a choice of $(\boldsymbol{H}_1\boldsymbol{G}\boldsymbol{H}_2)^-$ is $\boldsymbol{H}_2^{-1}\boldsymbol{G}^-\boldsymbol{H}_1^{-1}$. *Q.E.D.*

The inequality in (3.8) admits an operational interpretation. Under the assumption that $(\boldsymbol{T}_1' : \boldsymbol{T}_2')'$ is normally distributed, the expression on the left-hand side of (3.8) divided by $2\sigma^2$ is the noncentrality parameter of the distribution of the usual F-statistic for testing $H_0 : \eta(\mathscr{F}) = \boldsymbol{0}$ in the model (1.11), where $\eta(\mathscr{F})$ is as given in (3.4). This follows from the fact that the best linear estimator of $\boldsymbol{S}^{1/2}\eta(\mathscr{F})$ in the model (1.11) is

$$\left[\boldsymbol{A}\left(\boldsymbol{A}'\boldsymbol{V}_{11}^{-1}\boldsymbol{A}\right)^-\boldsymbol{A}'\boldsymbol{V}_{11}^{-1} - \boldsymbol{F}(\boldsymbol{F}'\boldsymbol{S}^{-1}\boldsymbol{F})^-\boldsymbol{F}'\boldsymbol{S}^{-1}\right]\boldsymbol{T}_1$$
$$+ \boldsymbol{F}(\boldsymbol{F}'\boldsymbol{S}^{-1}\boldsymbol{F})^-\boldsymbol{F}'\boldsymbol{S}^{-1}\boldsymbol{V}_{12}\boldsymbol{V}_{22}^{-1}\boldsymbol{T}_2 \tag{3.14}$$

and the dispersion matrix of this estimator is

$$\sigma^2\left[\boldsymbol{A}\left(\boldsymbol{A}'\boldsymbol{V}_{11}^{-1}\boldsymbol{A}\right)^-\boldsymbol{A}' - \boldsymbol{F}(\boldsymbol{F}'\boldsymbol{S}^{-1}\boldsymbol{F})^-\boldsymbol{F}'\right].$$

The formula (3.14) has been derived applying Gauss–Markov theorem to the model (1.11) and the results of Marsaglia and Styan (1974) for determining explicit forms of generalized inverses of the partitioned matrices involved. Its validity follows also from the fact that $\boldsymbol{K}_1\boldsymbol{T}_1 + \boldsymbol{K}_2\boldsymbol{T}_2$ is the best linear unbiased estimator of an estimable vector of parametric functions in the model (1.11) if and only if it is unbiased and satisfies the condition $\mathscr{R}(\boldsymbol{V}_{11}\boldsymbol{K}_1' + \boldsymbol{V}_{12}\boldsymbol{K}_2') \subseteq \mathscr{R}(\boldsymbol{A})$. The operational character of the condition (3.8) may be utilized to construct a preliminary test estimator of $\boldsymbol{\theta}$ along the lines proposed by Toro–Vizcarrondo and Wallace (1968) in the context of a linear regression model; see also Judge and Bock (1978, Chapter 5).

4. Comparison of $\boldsymbol{T}_0(\mathscr{A})$ with $\boldsymbol{T}_0(\mathscr{F})$

Since $\boldsymbol{T}_0(\mathscr{A})$ is the best linear unbiased estimator of $\boldsymbol{\theta} = \boldsymbol{A\alpha}$ under the model (1.12), and since $\boldsymbol{T}_0(\mathscr{F})$ is admissible for $\boldsymbol{\theta}$ under this model whenever $\mathscr{F} \subseteq \mathscr{A}$ [cf. the comment to (2.5)], the existence of certain values of $\boldsymbol{\theta}$ and σ^2 for which $\boldsymbol{T}_0(\mathscr{F})$ dominates $\boldsymbol{T}_0(\mathscr{A})$ with respect to the mean-square-error-matrix criterion follows from Theorem 2 of Baksalary, Liski, and Trenkler (1989), and a precise characterization of these values may be obtained with the use of Lemma 2. But if the estimators $\boldsymbol{T}_0(\mathscr{A})$ and $\boldsymbol{T}_0(\mathscr{F})$ are compared under the model (1.11), then, except in the trivial case of $\boldsymbol{V}_{12} = \boldsymbol{0}$, they both are biased, and thus the two results mentioned above are no longer applicable to evaluating the nonnegative definiteness of the difference $M[T_0(\mathscr{A})] - M[T_0(\mathscr{F})]$, which is now of the form

$$M[T_0(\mathscr{A})] - M[T_0(\mathscr{F})] = D[T_0(\mathscr{A})] - D[T_0(\mathscr{F})] + b[T_0(\mathscr{A})]b[T_0(\mathscr{A})]' - b[T_0(\mathscr{F})]b[T_0(\mathscr{F})]'. \tag{4.1}$$

However, the problem can be solved with the use of Lemma 3 below, which follows by combining the results in Table 1 of Caron and Gould (1986).

Lemma 3. *Let C be an $n \times n$ symmetric nonnegative definite matrix, let c_1 and c_2 be $n \times 1$ vectors, and let γ be a positive real number. Then $\gamma C + c_1c_1' - c_2c_2' \geq 0$ if and only if*

(*a*) *either $c_1 \in \mathscr{R}(C)$, $c_2 \in \mathscr{R}(C)$, and*

$$c_2'C^-c_2 - c_1'C^-c_1 + \sqrt{(c_1'C^-c_1 + c_2'C^-c_2)^2 - 4(c_1'C^-c_2)^2} \leq 2\gamma, \tag{4.2}$$

(*b*) *or $c_1 \notin \mathscr{R}(C)$, $c_2 \in \mathscr{R}(C : c_1)$, and*

$$(c_2 - \delta c_1)'C^-(c_2 - \delta c_1) \leq \gamma(1 - \delta^2), \tag{4.3}$$

where $\delta = c_1'(I_n - P_C)c_2/c_1'(I_n - P_C)c_1$ and the choice of C^- is arbitrary.

Theorem 2. *If $E(T_1) = \theta$ and $E(T_2) = \tau$ take values in $\mathscr{A} \subseteq \mathbb{R}^{k\times 1}$ and $\mathbb{R}^{m\times 1}$, respectively, then the covariance adjusted estimator $T_0(\mathscr{F})$, with some $\mathscr{F} = \mathscr{R}(F)$ being a proper subspace of $\mathscr{A} = \mathscr{R}(A)$, dominates the covariance adjusted estimator $T_0(\mathscr{A})$ with respect to the mean-square-error-matrix criterion if and only if*

(*a*) *either the equality*

$$F'S^{-1}V_{12}V_{22}^{-1}\tau = 0 \tag{4.4}$$

holds along with

$$\phi_{22} - \phi_{11} + \sqrt{(\phi_{11} + \phi_{22})^2 - 4\phi_{12}^2} \leq 2\sigma^2, \qquad \textit{where} \tag{4.5}$$

$$\phi_{11} = \tau'V_{22}^{-1}V_{12}'S^{-1}A(A'S^{-1}A)^-A'S^{-1}V_{12}V_{22}^{-1}\tau,$$

$$\phi_{12} = \tau'V_{22}^{-1}V_{12}'S^{-1}\theta,$$

$$\phi_{22} = \theta'\left[S^{-1} - S^{-1}F(F'S^{-1}F)^-F'S^{-1}\right]\theta,$$

(*b*) *or, if $F'S^{-1}V_{12}V_{22}^{-1}\tau \neq 0$, then*

$$\theta - V_{12}V_{22}^{-1}\tau \in \mathscr{R}\left[I_k - A(A'S^{-1}A)^-A'S^{-1}\right] \oplus \mathscr{R}(F), \tag{4.6}$$

i.e., $b[T_0(\mathscr{F})] = b[T_0(\mathscr{A})]$.

Proof. From (3.2) it is clear that

$$D[T_0(\mathscr{A})] - D[T_0(\mathscr{F})] = \sigma^2S^{1/2}D_{\#}S^{1/2}, \tag{4.7}$$

where $\boldsymbol{D}_{\#}$ is specified in (3.10). In view of (4.7) and (3.3), the difference of the mean-square-error-matrices in (4.1) may be expressed as

$$M[\boldsymbol{T}_0(\mathscr{A})] - M[\boldsymbol{T}_0(\mathscr{F})] = \boldsymbol{s}^{1/2}\big[\sigma^2\boldsymbol{D}_{\#} + \eta(\mathscr{A})\eta(\mathscr{A})' - \eta(\mathscr{F})\eta(\mathscr{F})'\big]\boldsymbol{S}^{1/2},$$

where the general representation of $\eta(\mathscr{F})$ is given in (3.4) and, due to $\boldsymbol{\theta} \in \mathscr{A}$, its particular case is

$$\eta(\mathscr{A}) = \boldsymbol{P}_{\boldsymbol{S}^{-1/2}\boldsymbol{A}}\boldsymbol{H}\boldsymbol{\tau}. \tag{4.8}$$

Now we apply Lemma 3 with $\boldsymbol{C} = \boldsymbol{D}_{\#}$, $\boldsymbol{c}_1 = \eta(\mathscr{A})$, $\boldsymbol{c}_2 = \eta(\mathscr{F})$, and $\gamma = \sigma^2$. Its part (a) requires that $\eta(\mathscr{A}) \in \mathscr{R}(\boldsymbol{D}_{\#})$. According to (3.11), this is equivalent to

$$(\boldsymbol{I}_k - \boldsymbol{P}_{\boldsymbol{S}^{-1/2}\boldsymbol{A}} + \boldsymbol{P}_{\boldsymbol{S}^{-1/2}\boldsymbol{F}})\boldsymbol{P}_{\boldsymbol{S}^{-1/2}\boldsymbol{A}}\boldsymbol{H}\boldsymbol{\tau} = \boldsymbol{0},$$

and hence to (4.4). If (4.4) holds, then (3.4) simplifies to

$$\eta(\mathscr{F}) = (\boldsymbol{P}_{\boldsymbol{S}^{-1/2}\boldsymbol{A}} - \boldsymbol{P}_{\boldsymbol{S}^{-1/2}\boldsymbol{F}})\boldsymbol{S}^{-1/2}\boldsymbol{\theta}. \tag{4.9}$$

In view of (3.10) and (4.9), the relation $\eta(\mathscr{F}) \in \mathscr{R}(\boldsymbol{D}_{\#})$ is trivially fulfilled. With the choice of $\boldsymbol{D}_{\#}^{-} = \boldsymbol{D}_{\#}$ and the use of (4.8) and (4.9), the condition (4.2) transforms to (4.5).

From part (b) of Lemma 3 we get $\boldsymbol{F}'\boldsymbol{S}^{-1}\boldsymbol{V}_{12}\boldsymbol{V}_{22}^{-1}\boldsymbol{\tau} \neq \boldsymbol{0}$ and, since for $\eta(\mathscr{A})$ and $\eta(\mathscr{F})$ given in (4.8) and (3.4) we have

$$\delta = \frac{\eta(\mathscr{A})'(\boldsymbol{I}_k - \boldsymbol{D}_{\#})\eta(\mathscr{F})}{\eta(\mathscr{F})'(\boldsymbol{I}_k - \boldsymbol{D}_{\#})\eta(\mathscr{F})} = 1,$$

the condition (4.3) reduces to the equality

$$(\boldsymbol{P}_{\boldsymbol{S}^{-1/2}\boldsymbol{A}} - \boldsymbol{P}_{\boldsymbol{S}^{-1/2}\boldsymbol{F}})(\boldsymbol{S}^{-1/2}\boldsymbol{\theta} - \boldsymbol{H}\boldsymbol{\tau}) = \boldsymbol{0}. \tag{4.10}$$

In view of (3.10), (3.11), and due to the fact that the orthocomplement of $\mathscr{R}(\boldsymbol{D}_{\#})$ coincides with $\mathscr{R}^{\perp}(\boldsymbol{S}^{-1/2}\boldsymbol{A}) \oplus \mathscr{R}(\boldsymbol{S}^{-1/2}\boldsymbol{F})$, the condition (4.10) may be replaced by

$$\boldsymbol{S}^{-1/2}(\boldsymbol{\theta} - \boldsymbol{V}_{12}\boldsymbol{V}_{22}^{-1}\boldsymbol{\tau}) \in \mathscr{R}\big[\boldsymbol{I}_k - \boldsymbol{S}^{-1/2}\boldsymbol{A}(\boldsymbol{A}'\boldsymbol{S}^{-1}\boldsymbol{A})^{-}\boldsymbol{A}'\boldsymbol{S}^{-1/2}\big] \oplus \mathscr{R}(\boldsymbol{S}^{-1/2}\boldsymbol{F}). \tag{4.11}$$

Premultiplying in (4.11) by $\boldsymbol{S}^{1/2}$ leads to (4.6). Hence

$$\boldsymbol{A}(\boldsymbol{A}'\boldsymbol{S}^{-1}\boldsymbol{A})^{-}\boldsymbol{A}'\boldsymbol{S}^{-1}(\boldsymbol{\theta} - \boldsymbol{V}_{12}\boldsymbol{V}_{22}^{-1}\boldsymbol{\tau}) = \boldsymbol{F}(\boldsymbol{F}'\boldsymbol{S}^{-1}\boldsymbol{F})^{-}\boldsymbol{F}'\boldsymbol{S}^{-1}(\boldsymbol{\theta} - \boldsymbol{V}_{12}\boldsymbol{V}_{22}^{-1}\boldsymbol{\tau}),$$

which is equivalent to $b[\boldsymbol{T}_0(\mathscr{A})] = b[\boldsymbol{T}_0(\mathscr{F})]$. *Q.E.D.*

A part of the proof of Theorem 2 establishes that, for any $\mathscr{F} \subseteq \mathscr{A}$, the dispersion matrices of the estimators $\boldsymbol{T}_0(\mathscr{A})$ and $\boldsymbol{T}_0(\mathscr{F})$ are ordered as $D[\boldsymbol{T}_0(\mathscr{F})] \leq D[\boldsymbol{T}_0(\mathscr{A})]$ in the Löwner sense, which is necessary for considering the mean-square-error-matrix dominance of $\boldsymbol{T}_0(\mathscr{F})$ over $\boldsymbol{T}_0(\mathscr{A})$. Theorem 2 itself shows that the additional requirements concerning the bias vector of $\boldsymbol{T}_0(\mathscr{F})$ are rather strong: either $b[\boldsymbol{T}_0(\mathscr{F})]$ is independent of $\boldsymbol{\tau}$, which , combined with the fact that $b[\boldsymbol{T}_0(\mathscr{A})]$ depends merely on $\boldsymbol{\tau}$, assures the existence of the dominance region as specified in (4.5), or, if $b[\boldsymbol{T}_0(\mathscr{F})]$ is a function of $\boldsymbol{\tau}$ as well, then it must be identical with $b[\boldsymbol{T}_0(\mathscr{A})]$, thus being independent of $\boldsymbol{\theta}$.

References

Baksalary, J.K. and R. Kala (1979). Covariance adjustment when a vector of parameters is restricted to a given subspace. *SIAM J. Appl. Math.* **37**, 20–21.

Baksalary, J.K. and R. Kala (1982). Admissible estimation by covariance adjustment technique. *Sankhyā Ser.* **44**, 281–285.

Baksalary, J.K. and R. Kala (1983a). Estimation via linearly combining two given statistics. *Ann. Statist.* **11** 691–696.

Baksalary, J.K. and R. Kala (1983b). Partial orderings between matrices one of which is of rank one. *Bull. Polish Acad. Sci. Math.* **31**, 5–7.

Baksalary, J.K., E.P. Liski and G. Trenkler (1989). Mean square error matrix improvements and admissibility of linear estimators. *J. Statist. Plann. Inference* **23**, 313–325.

Caron, R.J. and N.I.M. Gould (1986). Finding a positive semidefinite interval for a parametric matrix. *Linear Algebra Appl.* **76**, 19–29.

Cohen, A. (1966). All admissible linear estimates of the mean vector. *Ann. Math. Statist.* **37**, 458–463.

Judge, G.G. and M.E. Bock (1978). *The Statistical Implications of Pre-test and Stein-rule Estimators in Econometrics* (North-Holland, Amsterdam).

Lewis, T.O. and P.L. Odell (1971). *Estimation in Linear Models*, Prentice-Hall, Englewood Cliffs, NJ.

Marsaglia, G. and G.P.H. Styan (1974). Rank conditions for generalized inverses of partitioned matrices. *Sankhyā Ser. A* **36**, 437–442.

Mathew, T. and P. Bhimasankaram (1983). Optimality of BLUE's in a general linear model with incorrect design matrix. *J. Statist. Plann. Inference* **8**, 315–329.

Mitra, S.K. and C.R. Rao (1969). Conditions for optimality and validity of simple least squares theory. *Ann. Math. Statist.* **40**, 1617–1624.

Potthoff, R.F. and S.N. Roy (1964). A generalized multivariate analysis of variance model useful especially for growth curve problems. *Biometrika* **51** 313–326.

Pukelsheim, F. and G.P.H. Styan (1983). Convexity and monotonicity properties of dispersion matrices of estimators in linear models. *Scand. J. Statist.* **10** 145–149.

Rao, C.R. (1962). A note on a generalized inverse of a matrix with applications to problems in mathematical statistics. *J. Roy. Statist. Soc. Ser. B.* **24** 152–158.

Rao, C.R. (1966). Covariance adjustment and related problems in multivariate analysis: In: P.R. Krishnaiah (Ed.), *Multivariate Analysis* (Academic Press, New York) 87–103.

Rao, C.R. (1967). Least squares theory using an estimated dispersion matrix and its application to measurement of signals. In: L.M. Le Cam and J. Neyman (Eds.), *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability*, Vol. 1, University of California Press, Berkeley, CA, 355–372.

Rao, C.R. (1976). Estimation of parameters in a linear model. (The 1975 Wald Memorial Lectures.) *Ann. Statist.*. **4**, 1023–1037. [Corrigendum: **7** (1979), 696.]

Rao, C.R. and S.K. Mitra (1971). *Generalized Inverse of Matrices and Its Applications* (Wiley, New York).

Toro-Vizcarrondo, C. and T.D. Wallace (1968). A test of the mean square error criterion for restrictions in linear regression. *J. Amer. Statist. Assoc.* **63** 558–572.

Trenkler, G. (1985). Mean square error matrix comparisons of estimators in linear regression. *Comm. Statist. Theory Methods* **14** 2495–2509.