The Annals of Statistics
1981, Vol. 9, No. 2, 278–292

NONNEGATIVE MINIMUM BIASED INVARIANT ESTIMATION IN VARIANCE COMPONENT MODELS[1]

BY JOACHIM HARTUNG

Universität Dortmund

In a general variance component model, nonnegative quadratic estimators of the components of variance are considered which are invariant with respect to mean value translations and have minimum bias, analogously to estimation theory of mean value parameters. Here the minimum is taken over an appropriate cone of positive semidefinite matrices, after having made a reduction by invariance.

Among these estimators, which always exist, the one of minimum norm is characterized. This characterization is achieved by systems of necessary and sufficient conditions, and by a nonlinear cone-restricted pseudoinverse. A representation of this pseudoinverse is given, that allows computation without consideration of the boundary.

In models where the decomposing covariance matrices span a commutative quadratic subspace, a representation of the considered estimator is derived that requires merely to solve an ordinary convex quadratic optimization problem. As an example, we present the two-way nested classification random model.

In the case that unbiased nonnegative quadratic estimation is possible, this estimator automatically becomes the "nonnegative MINQUE".

Besides this, a general representation of the MINQUE is given, that involves just one matrix pseudoinversion in the reduced model.

1. Introduction. The fundamental defect in estimating variance components by unbiased quadratic estimators is the fact that the estimators can take on negative values while estimating nonnegative variances. Much attention has been given to this problem and several approaches are taken; see Thompson (1962), Federer (1968), McHugh and Milke (1968), J. N. K. Rao and K. Subrahmaniam (1971), C. R. Rao (1972), Drygas (1972), LaMotte (1973), J. N. K. Rao (1973), Harville (1977), Pukelsheim (1977, 1979, 1981), Hartley, et al. (1978), and the discussion and references given by Searle (1971), pages 406–408. The proposed nonnegative estimators either lack some desirable optimality properties or they exist only for special models, resp. are applicable only under particular assumptions.

The MINQUE (minimum norm quadratic unbiased estimator), introduced by C. R. Rao (1970; 1972; 1973), pages 303–305, is usually defined on the whole space of appropriate symmetric matrices. In order to get nonnegative estimates C. R. Rao (1972), Section 7, suggested restricting the class of possible estimators to the corresponding cone of positive semidefinite matrices, mentioning the resulting problem of finding a "nonnegative MINQUE", if it exists, as likely a difficult one. This problem is further considered, e.g., by LaMotte (1973b) and Pukelsheim (1977, 1979, 1981). However, such estimators exist only in very special cases. For example, in the analysis of variance (ANOVA) models, besides the overall variance σ_e^2, none of the other variance components permit the existence of a

Received October 1978; revised November 1979.

[1] This research has been supported by the Sonderforschungsbereich 72 at the University of Bonn.

AMS 1970 *subject classification.* 62J10.

Key words and phrases. Negative estimates of variances, MINQUE, invariance, minimum bias, invariant quadratic nonnegative minimum bias minimum norm estimator, restricted pseudoinverse, ANOVA models, balanced two-way nested classification model.

positive semidefinite matrix that is an unbiased estimator, as pointed out by LaMotte (1973b).

In this paper we consider minimum bias estimators (as introduced by Chipman (1964) for estimating mean value parameters), which are invariant under the group of mean value translations. Here the minimum is taken over the appropriate cone of positive semidefinite matrices after having made a reduction by invariance. These estimators always exist, and of course they guarantee nonnegative estimates. Moreover, they are unbiased if nonnegative unbiased quadratic estimation is possible. To get a unique estimator, we choose the one with minimal norm.

We characterize the minimum norm minimum bias invariant positive semidefinite estimator by introducing a cone-restricted pseudoinverse, and we obtain a limit formula that partially extends a result for null-space-restricted pseudoinverses given by the author (1979). Further, useful necessary and sufficient conditions for a matrix to be the desired estimator are derived.

In models where the decomposing covariance matrices span a commutative quadratic subspace the computation of this estimator is reduced to an ordinary convex quadratic optimization problem. As an example, the balanced two-way nested classification model with random effects is considered where the estimators for the three variance components are stated explicitly.

For the MINQUE a representation is given that involves just one matrix pseudoinversion, in the reduced (by invariance) model.

Some results of this paper were presented by the author at the 11th European meeting of statisticians in Oslo, Norway in August, 1978.

2. Formulation of the problem, definitions, and a representation of the MINQUE. Let us denote by Sym the Hilbert space of all real symmetric $n \times n$-matrices, where the inner product is given by $(A \mid B)_{\text{Sym}} = \text{trace } AB =: \text{tr } AB$, defining the norm $\|A\|^2_{\text{Sym}} := \text{tr } AA = \text{tr } A^2$, for $A, B \in \text{Sym}$. Further let PSD denote the closed convex cone of positive semidefinite matrices in Sym,

$$\text{PSD} := \{A \mid A \in \text{Sym}, x'Ax \geq 0 \text{ for all } x \in \mathbb{R}^n\}.$$

We consider the linear variance component model

$$z \sim (X\beta; \textstyle\sum_{i=1}^m \alpha_i U_i), \tag{2.1}$$

that consists of an n-dimensional random variable z with mean value

$$Ez = X\beta$$

and variance-covariance matrix

$$\text{Cov}(z) = \textstyle\sum_{i=1}^m \alpha_i U_i,$$

where the $n \times k$ design matrix X and the m symmetric positive semidefinite $n \times n$ matrices U_i are known, $U_i \in \text{PSD}$, $i = 1, \ldots, m$, while the parameter β varies in $\mathbb{R}^k$ and the parameter $\alpha = (\alpha_1, \ldots, \alpha_m)'$ varies in $\mathbb{R}^m_+$, the nonnegative orthant of $\mathbb{R}^m$.

The problem considered here is to find quadratic estimates for the variance components $\alpha_1, \ldots, \alpha_m$, that are nonnegative and invariant with respect to the group G of mean value translations,

$$G = \{z \mapsto z + X\beta \mid \beta \in \mathbb{R}^k\}.$$

A maximal invariant liner statistic y with respect to G is given then by (Seely (1971))

$$y = \text{Proj}_{R(X)^\perp} z = (I - XX^+)z,$$

where I is the $n \times n$ identity matrix, X^+ is the Moore-Penrose generalized inverse, and $R(X)$ denotes the range of X. We get the reduced (by invariance) linear variance component model

(2.2) $$y \sim (0; \textstyle\sum_{i=1}^m \alpha_i V_i), \qquad \alpha \in \mathbb{R}_+^m, \quad V_i \in \text{PSD},$$

where $V_i = (I - XX^+)U_i(I - XX^+)$, $i = 1, \ldots, m$. Often a model of the kind (2.2) is given also by the experimental arrangement, for instance by grouping and measuring of differences.

Now let $A \in \text{Sym}$, then a quadratic invariant estimate for a linear form $p'\alpha$, $p \in \mathbb{R}^m$, is given by $y'Ay$, with the bias

$$Ey'Ay - p'\alpha = \textstyle\sum_{i=1}^m \alpha_i(\operatorname{tr} A\ V_i - p_i).$$

Here $y'\tilde{A}y$ is an unbiased estimate of $p'\alpha$ if

(2.3) $$\operatorname{tr} \tilde{A}\ V_i = p_i, \qquad \text{for all} \quad i = 1, \ldots, m.$$

It is additionally of "minimum norm" if $\tilde{A}$ solves the problem

(2.4) $$\text{minimize}\{\operatorname{tr} B^2 \mid B \in \text{Sym}, \operatorname{tr} B\ V_i = p_i, i = 1, \ldots, m,\}.$$

Then $\tilde{A}$ is called the MINQUE, minimum norm quadratic unbiased estimator.

However, the MINQUE doesn't always exist. By analogy with estimation theory of mean value parameters (cf., for instance, Chipman (1964)), the condition of unbiasedness (2.3) may be weakened to that of finding a best approximate solution of (2.3), i.e., minimizing the discrepancy $\sum_i (\operatorname{tr} A\ V_i - p_i)^2$ over Sym, cf., Pukelsheim (1976).

DEFINITION 2.1. For estimating the linear form $p'\alpha$ the matrix $\bar{A} \in \text{Sym}$ is the MINQE that gives minimum bias with respect to Sym, if $\bar{A}$ solves the following problem.

(2.5) $$\text{minimize } \operatorname{tr} A^2 \text{ subject to: } A \in \text{Sym, and } \textstyle\sum_{i=1}^m (\operatorname{tr} A\ V_i - p_i)^2 = \min!_{A \in \text{Sym}}.$$

Let us introduce the linear operator

(2.6) $$g\colon \text{Sym} \to \mathbb{R}^m, \qquad A \mapsto gA = \begin{pmatrix} \operatorname{tr} A\ V_1 \\ \vdots \\ \operatorname{tr} A\ V_m \end{pmatrix}.$$

Then (2.5) is equivalent to

(2.7) $$\text{minimize}\{\operatorname{tr} A^2 \mid A \in \text{Sym}, \|gA - p\|_{\mathbb{R}^m} = \min!_{A \in \text{Sym}}\},$$

respectively, find a best approximate solution $\bar{A}$, of minimum norm of the linear equation $gA = p$, $A \in \text{Sym}$. By definition of a pseudoinverse operator, e.g., Holmes (1972), page 220,

(2.8) $$\bar{A} = g^+ p,$$

g^+ the pseudoinverse of g, which for matrices is identical to the Moore-Penrose generalized inverse, cf., Mitra (1975). The estimator $\bar{A}$ always exists and $\bar{A}$ is equal to the MINQUE if $p \in R(g)$.

We now show how to compute g^+. The adjoint g^* of g is given by

(2.9) $$g^*\colon \mathbb{R}^m \to \text{Sym}, a \mapsto g^* a = \textstyle\sum_{i=1}^m a_i V_i, \qquad a = (a_1, \ldots, a_m)'.$$

Then

(2.10) $$gg^* = (\operatorname{tr} V_i V_j)_{i=1,\ldots,m;\, j=1,\ldots,m},$$

and

(2.11) $$g^*g\colon \text{Sym} \to \text{Sym}, A \mapsto \textstyle\sum_{i=1}^m (\operatorname{tr} A\ V_i) V_i.$$

Using now the following properties of pseudoinverses in Hilbert spaces (e.g., Holmes (1972), page 222),

$$g^+ = g^*(gg^*)^+, \tag{2.12}$$

$$g^+ = (g^*g)^+ g^*, \tag{2.13}$$

we get a computational representation of g^+, resp. of $\bar{A}$ or of the MINQUE $\tilde{A}$ if $p \in R(g)$.

THEOREM 2.1. *The estimator $\bar{A}$ satisfies the "normal equation"*

$$\textstyle\sum_{i=1}^m (\operatorname{tr} \bar{A}\ V_i) V_i = \sum_{i=1}^m [p]_i V_i, \tag{2.14}$$

and permits the computational representation

$$\bar{A} = g^+p = \textstyle\sum_{i=1}^m [(gg^*)^+p]_i V_i, \tag{2.15}$$

where $[w]_i$ denotes the ith component of vector w.

Now, an estimate $y'Ay$ with $A \in$ Sym can be negative, while estimating nonnegative variance components. Therefore, C. R. Rao (1972), Section 7, suggested finding a MINQUE over the cone PSD, i.e., find a solution $\tilde{A}^0$ of the problem

$$\text{minimize}\{\operatorname{tr} A^2 \,|\, A \in \text{PSD}, gA = p\}, \tag{2.16}$$

a problem further considered for instance by LaMotte (1973b) and Pukelsheim (1977, 1979, 1981). However, the kind of models where such an estimator $\tilde{A}^0$ exists for all variance components is very limited. For instance, as pointed out by LaMotte (1973b), in ANOVA models the only component that might be estimable in this way is the overall variance σ_e^2; cf. also the 2-way nested layout considered in Section 5.

Thus we are led to exchange the unbiasedness condition in (2.16), $gA = p$, by the claim to minimize the discrepancy $\|gA - p\|_{\mathbb{R}^m}$ over PSD.

DEFINITION 2.2. $\hat{A} \in$ PSD is the nonnegative MINQ minimum bias estimator of the linear form $p'\alpha$ if $\hat{A}$ solves the following problem,

$$\text{minimize } \operatorname{tr} A^2 \text{ subject to: } A \in \text{PSD, and } \|gA - p\| = \min_{B \in \text{PSD}} \|gB - p\|. \tag{2.17}$$

THEOREM 2.2. *$\hat{A}$ always exists and is uniquely determined.*

PROOF. The cone PSD is closed and convex, g is a continuous linear mapping with closed range, $R(g_{|\text{PSD}})$ is closed and convex, and so there exists a best $R(g_{|\text{PSD}})$-approximation to p, say p_a, that is unique. Now PSD $\cap$ $\{A \,|\, gA = p_a\}$ is nonempty, closed and convex, so has a unique element of minimum norm and this is just $\hat{A}$.

Analogously to a nullspace-restricted pseudoinverse (cf., Minamide and Nakamura (1970), Holmes (1972), Section 35, Hartung (1979)) let us introduce now a cone-restricted pseudoinverse, which, of course, in general is a nonlinear operator.

DEFINITION 2.3. The operator $g^+_{|\text{PSD}}: \mathbb{R}^m \to \text{PSD}$ is the PSD-restricted pseudoinverse of g if for every $q \in \mathbb{R}^m$ the best approximate solution $A(q)$ of minimum norm of the linear equation

$$gA = q \qquad \text{subject to} \quad A \in \text{PSD}$$

is given by $A(q) = g^+_{|\text{PSD}}\, q$.

So the solution $\hat{A}$ of (2.17) is given by $\hat{A} = g_{|\text{PSD}^+}\, p$, and by Theorem 2.2 $g_{|\text{PSD}^+}$ exists and is a single valued function.

If nonnegative unbiased estimability is given then $\hat{A}$ automatically becomes the "nonnegative MINQUE" $\tilde{A}^0$.

REMARK 2.1. In our approach to finding an estimation function $d_A(y) = y'Ay$, $A \in$ Sym, of $p'\alpha$ we have simultaneously three goals in mind: (i) $d_A(y)$ should be nonnegative for all realizations of y, i.e., $A \in$ PSD, resp. $A = \text{Proj}_{\text{PSD}} A$; (ii) $d_A(y)$ should be unbiased, i.e., $gA = p$; (iii) A should be of minimum norm. Thus the general decision problem here consists of minimizing "simultaneously" the three functions,

1. $r_1(A) := \operatorname{tr}(A - \text{Proj}_{\text{PSD}} A)^2$;
2. $r_2(A) := \|gA - p\|^2$;
3. $r_3(A) := \operatorname{tr} A^2$.

Choosing a lexicographical preference in the indicated order, then $\hat{A} = g_{|\text{PSD}^+}p$ gives the lexicographical minimum and, of course, $d_{\hat{A}}$ is admissible, resp. $\hat{A}$, in the sense that there is no function d_{A_1} such that $r(A_1) \leq r(\hat{A})$, $r = (r_1, r_2, r_3)'$, $A_1 \in$ Sym.

With arbitrary weights $\tau_i > 0$, $i = 1, 2, 3$, we get by minimizing $\sum_{i=1}^{3} \tau_i r_i(A)$ over Sym a unique solution $A(\tau)$. $d_{A(\tau)}$ is also admissible, and it is known that each admissible function d_{A_a} can be approximated by a sequence $\{d_{A(\tau^{(\nu)})}\}_{\nu \in \mathbb{N}}$ of such functions, i.e., $\operatorname{tr}(A(\tau^{(\nu)}) - A_a)^2 \to 0$ for $\nu \to \infty$, where $\tau_i^{(\nu)} > 0$, $i = 1, 2, 3$, $\nu \in \mathbb{N}$. Hence in particular to $\hat{A}$ there belongs a sequence $\{\tau^{(\nu)}(\hat{A})\}$, so that $\hat{A}$ is approximated by admissible $A(\tau^{(\nu)}(\hat{A}))$, and in section 4 such a sequence $\{\tau^{(\nu)}(\hat{A})\}$ is specified, where for convenience a norming condition of the weights is omitted. Note that finding an $A(\tau)$ is an unrestricted optimization problem, which is in general much easier to solve than restricted problems. Thus the convergence properties mentioned above also provide a method for determining $\hat{A}$. But first necessary and sufficient optimality conditions are derived for the three-stage minimization problem corresponding to $\hat{A}$.

3. A characterization of $\hat{\mathbf{A}}$. First we consider the class of nonnegative minimum bias estimators. For the following convex program, describing the nonnegative minimum bias estimators for a linear form $q'\alpha$,

$$\text{minimize}\{\|gA - q\|^2 \,|\, A \in \text{PSD}\}, \tag{3.1}$$

with $q \in \mathbb{R}^m$ given, we define the correspondent Lagrange function

$$L(A, B) = \|gA - q\|^2 - \operatorname{tr} AB, \qquad A, B \in \text{Sym}. \tag{3.2}$$

By Theorem 2.2 there always exists a solution of (3.1).

THEOREM 3.1. *There exists a "Lagrange multiplier"* $B^0 \in$ PSD *such that for any solution* A^0 *of* (3.1)

$$\|gA^0 - q\|^2 = L(A^0, B^0) = \min_{A \in \text{Sym}} L(A, B^0) \tag{3.3}$$

and

$$\operatorname{tr} A^0 B^0 = 0. \tag{3.4}$$

Further, (A^0, B^0) *is a saddle point of* $L(A, B)$ *with respect to* Sym $\times$ PSD, i.e.,

$$L(A^0, B) \leq L(A^0, B^0) \leq L(A, B^0) \tag{3.5}$$

for all $A \in$ Sym, $B \in$ PSD. *Conversely, if* $(A^1, B^1) \in$ Sym $\times$ PSD *is a saddle point of* L *in* (3.5), *then* A^1 *solves the problem* (3.1).

PROOF. Since Sym with the inner product tr AB for $A, B \in$ Sym is a Hilbert space, the "positive" dual cone of the " positive" cone PSD is given by

$$\text{PSD}^* = \{B \in \text{Sym} \,|\, \operatorname{tr} AB \geq 0 \text{ for all } A \in \text{PSD}\},$$

cf., Luenberger (1969), page 215, Berman (1973), page 5. Now PSD is self-dual, and its interior consists of PD, the set of positive definite matrices in Sym, cf., Berman (1973), page 55, i.e.,

$$\text{PSD}^* = \text{PSD}, \text{int PSD} = \text{PD}, \tag{3.6}$$

and PD is nonempty.

The cone PSD introduces a partial order on Sym, by saying $A \ge B$ if $A - B \in$ PSD and $A > B$ if $A - B \in$ PD. Then $A \in$ PSD, $B \in$ PD correspond to $A \ge 0$, $B > 0$, or equivalently $-A \le 0$, $-B < 0$, i.e., $-A \in -$PSD, the "negative" cone in Sym, $-B \in -$PD, respectively. Now with these reformulations and (3.6) we may apply the theorems of Luenberger (1969), pages 217, 219, 221, to extract our statements.

LEMMA 3.1. *The matrix $A^0 \in$ PSD is a solution of* (3.1) *if and only if*
(i) $g^*gA^0 - g^*q \in$ PSD *and*
(ii) $\text{tr}\, A^0(g^*gA^0 - g^*q) = 0$.

PROOF. Let A^0 solve (3.1), then by Theorem 3.1 there exists a $B^0 \in$ PSD such that minimum $L(A, B^0)$, $A \in$ Sym, is achieved at A^0. This implies that the gradient of $L(A, B^0)$ with respect to A vanishes at A^0, i.e., A^0 and B^0 are connected by

$$g^*gA^0 - g^*q = \tfrac{1}{2}B^0, \tag{3.7}$$

and by (3.4) $\text{tr}\, A^0B^0 = 0$, so that (i) and (ii) follow.

Now let $A^0 \in$ PSD satisfy (i), (ii) and define B^0 by the relation (3.7), then $L(A, B^0)$ is stationary at A^0 and by a convexity argument we have

$$L(A^0, B^0) = \min_{A\in\text{Sym}} L(A, B^0).$$

With (ii) this means,

$$\begin{aligned} \|gA^0 - q\|^2 &= \min_{A\in\text{Sym}} \{\|gA - q\|^2 - \text{tr}\, AB^0\} \\ &\le \min_{A\in\text{PSD}} \{\|gA - q\|^2 - \text{tr}\, AB^0\} \\ &\le \min_{A\in\text{PSD}} \|gA - q\|^2, \end{aligned}$$

because $\text{tr}\, AB^0 \ge 0$ for $A, B^0 \in$ PSD by the self-duality (3.6) of PSD.

Since A^0 is assumed to be in PSD it minimizes $\|gA - q\|$ over PSD, and the lemma is proved.

Using the self-duality of PSD, we may state the following corollary.

COROLLARY 3.1. *$A^0 \in$ PSD solves* (3.1) *if and only if A^0 solves the variational inequalities*

$$\text{tr}\, B(g^*gA^0 - g^*q) \ge 0, \qquad \textit{for all} \quad B \in \text{PSD},$$

and

$$\text{tr}\, A^0(g^*gA^0 - g^*q) = 0.$$

Now we give necessary and sufficient conditions for a matrix A to be an estimator $\hat{A}$, i.e., to be the minimum norm solution of the program (3.1) for a $q \in \mathbb{R}^m$, denoted by $g^+_{|\text{PSD}}\, q$, cf., Definitions (2.2) and (2.3).

THEOREM 3.2. *Let $q \in \mathbb{R}^m$, then $\hat{A} = g^+_{|\text{PSD}}\, q$ if the following conditions hold*:

$$\hat{A} \in \text{PSD}, \tag{3.8}$$

$$g^*g\hat{A} - g^*q \in \text{PSD}, \tag{3.9}$$

$$\text{(3.10)} \qquad \operatorname{tr} \hat{A}(g^*g\hat{A} - g^*q) = 0,$$

and for some $b^0 \in \mathbb{R}^m$,

$$\text{(3.11)} \qquad \hat{A} + g^*b^0 \in \text{PSD},$$

$$\text{(3.12)} \qquad \operatorname{tr} \hat{A}(\hat{A} + g^*b^0) = 0.$$

PROOF. Let $\hat{A}$ satisfy (3.8)–(3.12), then by Lemma 3.1 conditions (3.8)–(3.10) imply that $\hat{A}$ is a solution of (3.1), i.e., $\hat{A}$ minimizes $\|gA - q\|$ over PSD. Denote by $S(q)$ the set of all solutions of (3.1) and let $q_a := g\hat{A}$, then q_a is the unique best $R(g_{|\text{PSD}})$-approximation to q and

$$\text{(3.13)} \qquad S(q) = \{A \mid A \in \text{PSD}, gA = q_a\}, \qquad \hat{A} \in S(q).$$

We have to show that $\hat{A}$ is the element of minimum norm in $S(q)$. For A, $B \in$ Sym, $b \in \mathbb{R}^m$ we define the Lagrange function

$$\text{(3.14)} \qquad M(A, b, B) := \tfrac{1}{2} \operatorname{tr} A^2 + (gA - q_a)'b - \operatorname{tr} AB.$$

Let a fixed b^0 satisfy (3.11), (3.12) and take

$$B^0 := \hat{A} + g^*b^0,$$

then the function $M_0(A) := M(A, b^0, B^0)$ is stationary at $\hat{A}$. Note that grad $\{(gA)'b^0\} =$ grad $\{\operatorname{tr} A(g^*b^0)\} = g^*b^0$, and grad $M_0(A) = A + g^*b^0 - B^0$ vanishes in $\hat{A}$. Now $M_0(A)$ is convex and so is minimal at $\hat{A}$. Further, by (3.12) $\operatorname{tr} \hat{A}B^0 = 0$, and $\operatorname{tr} AB^0 \geq 0$ for all $A \in$ PSD by (3.11) together with the self-duality of PSD (3.6). Thus we get for all $A_1 \in S(q)$

$$\tfrac{1}{2} \operatorname{tr} \hat{A}^2 = \tfrac{1}{2} \operatorname{tr} \hat{A}^2 + (g\hat{A} - q_a)'b^0 - \operatorname{tr} \hat{A}B^0 = M_0(\hat{A}) = \min_{A \in \text{Sym}} M_0(A)$$
$$\leq \min_{A \in S(q)} M_0(A) \leq \tfrac{1}{2} \operatorname{tr} A_1^2 - \operatorname{tr} A_1B^0 \leq \tfrac{1}{2} \operatorname{tr} A_1^2,$$

and since $\hat{A} \in S(q)$ we have $\hat{A} = g^+_{|\text{PSD}}\, q$.

THEOREM 3.3. *Let* $\hat{A} = g_{|\text{PSD}^+}\, q$, $q \in \mathbb{R}^m$, *then there hold* (3.8), (3.9), (3.10) *and*

$$\text{(3.15)} \qquad \inf_{B \in \text{PSD}} \{\operatorname{tr} \hat{A}B + \tfrac{1}{2} \operatorname{tr}[\operatorname{Proj}_{(R(g^*)^\perp)} (B - \hat{A})]^2\} = 0;$$

if the infimum in (3.15) *is achieved for some* $B^0 \in$ PSD, *then* $\hat{A}$ *also satisfies* (3.11) *and* (3.12), *with* $g^*b^0 := B^0 - \hat{A}$.

PROOF. Now let $\hat{A} = g^+_{|\text{PSD}}\, q$, and $S(q)$ may be defined as in (3.13). Of course, $\hat{A} \in S(q)$ and so Lemma 3.1 gives conditions (3.8)–(3.10).

For the Lagrange function (3.14) define

$$u(A) := \sup\{M(A, b, B) \mid b \in \mathbb{R}^m, B \in \text{PSD}\},$$

and

$$v(b, B) := \inf\{M(A, b, B) \mid A \in \text{Sym}\}.$$

Then $u(A_1)$ is positive, and finite if and only if $gA_1 = q_a$ and $\operatorname{tr} A_1B \geq 0$ for all $B \in$ PSD, i.e., $A_1 \in S(q)$, in which case we get

$$\text{(3.16)} \qquad u(A_1) = \tfrac{1}{2} \operatorname{tr} A_1^2, \qquad A_1 \in S(q),$$

and there exists

$$\text{(3.17)} \qquad \min_{A \in \text{Sym}} u(A) = \min_{A \in S(q)} u(A) = u(\hat{A}).$$

$v(b, B)$ is finite for all $b \in \mathbb{R}^m$, $B \in$ PSD. Differentiating $M(A, b, B)$ with respect to A

gives $\text{grad}_A\, M(A, b, B) = A + g^*b - B$, such that for $A(b, B) := B - g^*b$ we have

$$\begin{aligned} v(b, B) &= M(A(b, B), b, B) \\ &= \tfrac{1}{2}\operatorname{tr}(B - g^*b)^2 + (gB - gg^*b - q_a)'b - \operatorname{tr} B^2 + \operatorname{tr} B(g^*b) \\ &= -\tfrac{1}{2}\operatorname{tr}(g^*b - (B - \hat{A}))^2 + \tfrac{1}{2}\operatorname{tr}\hat{A}^2 - \operatorname{tr}\hat{A}B. \end{aligned} \tag{3.18}$$

With (3.17) we obtain the trivial inequality,

$$\sup\{v(b, B) \mid (b, B) \in \mathbb{R}^m \times \text{PSD}\} \le u(\hat{A}). \tag{3.19}$$

Now for some fixed $b_1 \in \mathbb{R}^m$, $B_1 \in \text{PSD}$ consider the function

$$\begin{aligned} M_1(A) &:= M(A, b_1, B_1) \\ &= \tfrac{1}{2}\operatorname{tr} A^2 + \operatorname{tr} A(g^*b_1) - q_a'b_1 - \operatorname{tr} AB_1 \\ &= \tfrac{1}{2}\operatorname{tr}(A - (B_1 - g^*b_1))^2 - \tfrac{1}{2}\operatorname{tr}(B_1 - g^*b_1)^2 - q_a'b_1. \end{aligned}$$

Then the level sets $\{A \mid A \in \text{Sym}, M_1(A) \le c\}$, $c \in \mathbb{R}$, are compact. Hence the "inf-sup"-theorem of Moreau (1964) yields the existence of a saddle value of $M(A, b, B)$ with respect to $A \in \text{Sym}$ and $(b, B) \in \mathbb{R}^m \times \text{PSD}$, i.e., in (3.19) there holds equality, which, with the help of (3.16) and (3.18), gives

$$\inf\{\operatorname{tr}\hat{A}B + \tfrac{1}{2}\operatorname{tr}[g^*b - (B - \hat{A})]^2 \mid (b, B) \in \mathbb{R}^m \times \text{PSD}\} = 0. \tag{3.20}$$

For $g^*b^1 := \text{Proj}_{R(g^*)}(B - \hat{A}) = g^*(g^*)^+(B - \hat{A})$ the infimum in (3.20) with respect to $b \in \mathbb{R}^m$, and B fixed, is achieved, such that (3.15) follows. Now if the infimum in (3.15) is achieved for some $B^0 \in \text{PSD}$, then each term of the function vanishes, i.e., $\operatorname{tr}\hat{A}B^0 = 0$ and $B^0 - \hat{A} \in R(g^*)$, so (3.11) and (3.12) hold, with, for instance, $b^0 = (g^*)^+(B^0 - \hat{A})$, and the theorem is proved.

REMARK 3.1. If $\hat{A} \in R(g^*)$, then $B^0 = 0$ is minimizing point of (3.15). Further, defining the pair of dual programs

$$\{P\}: \text{minimize}_{A \in S(q)}\, u(A), \qquad \{D\}: \text{maximize}_{(b,B) \in \mathbb{R}^m \times \text{PSD}}\, v(b, B) \tag{3.21}$$

we have implicitly proved the following corollary.

COROLLARY 3.2. *For the programs $\{P\}$ and $\{D\}$ there holds the strong duality, in the form*: min $\{P\}$ = sup $\{D\}$, *and if (b^0, B^0) solves $\{D\}$, then the solution $\hat{A}$ of $\{P\}$ satisfies*: $\hat{A} = B^0 - g^*b^0$.

Here, by (3.16), $u(A) = \tfrac{1}{2}\operatorname{tr} A^2$ and from (3.18) we derive $v(b, B) = \operatorname{tr} B(g^*b) - \tfrac{1}{2}\operatorname{tr} B^2 - \tfrac{1}{2}\operatorname{tr}(g^*b)^2 - q_a'b$. Noting that the estimator $\hat{A}$ for $q'\alpha$ is the "nonnegative MINQUE" $\tilde{A}^0$ for $q_a'\alpha$, a further characterization via dual programs is given in Pukelsheim (1977).

4. A representation of $g_{|\text{PSD}}^+$.

Here we give a limit formula for $g_{|\text{PSD}^+}$ that partially extends a result of Hartung (1979) obtained for nullspace-restricted pseudoinverses and permits the pointwise computation.

For this let the spectral decomposition of a matrix $A \in \text{Sym}$ be given by

$$A = \textstyle\sum_{\lambda \in \sigma(A)} \lambda E(\lambda), \tag{4.1}$$

where $\sigma(A)$ is the spectrum of A and $E(\lambda)$ is the projection onto the eigenspace associated with λ. Define the "positive part" A_+ and "negative part" A_- of A by

$$A_+ = \textstyle\sum_{\lambda \in \sigma(A)} \lambda_+ E(\lambda), \qquad A_- = \sum_{\lambda \in \sigma(A)} \lambda_- E(\lambda), \tag{4.2}$$

where $\lambda_+ = \max\{0, \lambda\}$ and $\lambda_- = \min\{0, \lambda\}$. Then

$$A = A_+ + A_-, A_+, -A_- \in \mathrm{PSD}, \ \mathrm{tr}\, A_+A_- = 0, \tag{4.3}$$

$$A_+ = \mathrm{Proj}_{\mathrm{PSD}}\, A, \text{ i.e., } \mathrm{tr}(A - A_+)^2 = \min_{B\in\mathrm{PSD}}\mathrm{tr}(A - B)^2. \tag{4.4}$$

Denote this projection operator by P,

$$P:\mathrm{Sym} \to \mathrm{PSD}, \qquad A \mapsto PA = A_+. \tag{4.5}$$

For a positive real nullsequence $\{r_n\}_{n\in\mathbb{N}}$ we define the functionals, cf., Remark 2.1,

$$f_n(A) := \mathrm{tr}(A - A_+)^2 + r_n^2\|gA - q\|^2_{\mathbb{R}^m} + r_n^3\mathrm{tr}\, A^2, \tag{4.6}$$

where $A \in \mathrm{Sym}$, and $q \in \mathbb{R}^m$ fixed.

With a result of Holmes (1972), page 63, regarding the derivative of a projection we get that $A_n \in \mathrm{Sym}$ minimizes $f_n(A)$ over Sym if and only if

$$A_n - (A_n)_+ + r_n^2 g^*gA_n + r_n^3A_n = r_n^2g^*q. \tag{4.7}$$

Now (4.7) uniquely defines a solution operator Q_n, which with (4.5) may be written as follows,

$$Q_n = (I - P + r_n^2g^*g + r_n^3I)^{-1}r_n^2g^*, \tag{4.8}$$

that is,

$$A_n = Q_nq \quad \text{if and only if } A_n \text{ solves (4.7)}, \qquad n \in \mathbb{N}. \tag{4.9}$$

A_n exists and is uniquely determined, because the level sets of f_n are compact,

$$\{A \in \mathrm{Sym}\,|\,f_n(A) \le c\} \subset \{A \in \mathrm{Sym}\,|\,r_n^3\mathrm{tr}\, A^2 \le c\}, \qquad c \in \mathbb{R},$$

and f_n is strictly convex. In general, Q_n is a nonlinear operator.

THEOREM 4.1. *With the operator Q_n defined in* (4.8) *there holds the following representation of the* PSD-*restricted pseudoinverse of g,*

$$g_{|\mathrm{PSD}^+} = \lim_{n\to\infty}Q_n, \tag{4.10}$$

in the sense of pointwise convergence, and for a $q \in \mathbb{R}^m$,

$$\mathrm{tr}(Q_nq - (Q_nq)_+)^2 = o(r_n^3), \tag{4.11}$$

$$\|gQ_nq - q\|^2_{\mathbb{R}^m} = \|gg_{|\mathrm{PSD}^+}q - q\|^2_{\mathbb{R}^m} + o(r_n) \tag{4.12}$$

for $r_n \to +0$, $n \to \infty$.

PROOF. Let $g \in \mathbb{R}^m$ be arbitrary but fixed, $A_0 = g_{|\mathrm{PSD}^+}q (= \hat{A})$ and $A_n = Q_nq$, $(A_n)_- = A_n - (A_n)_+$, $n \in \mathbb{N}$. Noting that $A_- = 0$ if $A \in \mathrm{PSD}$ we have

$$0 \le f_n(A_n) \le f_n(A) = r_n^2\|gA - g\|^2 + r_n^3\mathrm{tr}\, A^2 \qquad \text{for} \quad A \in \mathrm{PSD}, \tag{4.13}$$

and so $\lim_{n\to\infty} f_n(A_n) = 0$. Together with

$$0 \le f_n(A_n) - \mathrm{tr}(A_n)^2_- \le f_n(A_n)$$

this yields

$$\lim_{n\to\infty}\mathrm{tr}(A_n)^2_- = 0. \tag{4.14}$$

By Theorem 3.1 there exists a matrix $B_0 \in \mathrm{PSD}$ such that for $A_0 = g_{|\mathrm{PSD}^+}q$ and for all

$A \in \mathrm{Sym}$, cf. (3.3),

$$\begin{aligned} \|gA_0 - q\|^2 &\le \|gA - q\|^2 - \operatorname{tr} AB_0 \\ &= \|gA - q\|^2 - \operatorname{tr}(A_+ + A_-)B_0 \\ &\le \|gA - q\|^2 - \operatorname{tr} B_0 A_- \\ &\le \|gA - q\|^2 + (\operatorname{tr} B_0^2)^{1/2}(\operatorname{tr} A_-^2)^{1/2}, \end{aligned} \tag{4.15}$$

where use is made of the Cauchy-Schwarz inequality. Now taking (4.13) for $A = A_0$ and (4.15) for $A = A_n$ we get with $k_n := (\operatorname{tr} B_0^2)^{1/2}(\operatorname{tr}(A_n)_-^2)^{1/2}$,

$$\begin{aligned} \operatorname{tr}(A_n)_-^2 + r_n^2\|gA_n - q\|^2 + r_n^3 \operatorname{tr} A_n^2 &= f_n(A_n) \\ &\le r_n^2\|gA_0 - q\|^2 + r_n^3 \operatorname{tr} A_0^2 \\ &\le r_n^2(\|gA_n - q\|^2 + k_n) + r_n^3 \operatorname{tr} A_0^2, \end{aligned} \tag{4.16}$$

and by subtracting $r_n^2\|gA_n - q\|^2$,

$$\operatorname{tr}(A_n)_-^2 + r_n^3 \operatorname{tr} A_n^2 \le r_n^2 k_n + r_n^3 \operatorname{tr} A_0^2. \tag{4.17}$$

By (4.14), $k_n \to 0$ for $n \to \infty$ and so dividing (4.17) by r_n^2 gives $\lim_{n\to\infty} r_n^{-2}\operatorname{tr}(A_n)_-^2 = 0$, respectively

$$\lim_{n\to\infty} r_n^{-1} k_n = 0. \tag{4.18}$$

Further from (4.17),

$$\operatorname{tr} A_n^2 \le r_n^{-1} k_n + \operatorname{tr} A_0^2, \tag{4.19}$$

hence by (4.18) the sequence $\{A_n\}_{n\in\mathbb{N}}$ is uniformly bounded and so possesses an accumulation point $A^0 \in \mathrm{Sym}$ and a subsequence converging to it. By (4.14) it follows with the continuity of the projection

$$\operatorname{tr}(A^0)_-^2 = 0, \qquad \text{i.e., } A^0 \in \mathrm{PSD}, \tag{4.20}$$

and by (4.13) with $r_n^2\|gA_n - q\|^2 \le f_n(A_n)$,

$$\|gA_n - q\|^2 \le \|gA - q\|^2 + r_n \operatorname{tr} A^2 \qquad \text{for all} \quad A \in \mathrm{PSD}. \tag{4.21}$$

Thus together with (4.20),

$$A^0 \text{ minimizes } \quad \|gA - q\| \quad \text{over PSD}. \tag{4.22}$$

By (4.19) we have with (4.18)

$$\operatorname{tr}(A^0)^2 \le \operatorname{tr} A_0^2,$$

where A_0 is the unique minimum norm solution of $\min_{A\in\mathrm{PSD}}\|gA - q\|$, so with (4.22) there holds

$$A^0 = A_0$$

and this is valid for all accumulation points of $\{A_n\}_{n\in\mathbb{N}}$, each subsequence of which has an accumulation point by (4.19). Thus the whole sequence $\{A_n\}_{n\in\mathbb{N}}$ converges to A_0, and (4.10) is shown. Then (4.17) and (4.18) yield $\lim_{n\to\infty} r_n^{-3}\operatorname{tr}(A_n)_-^2 = 0$, which states (4.11). Now subtracting $r_n^2\|gA_n - q\|^2 + r_n^3 \operatorname{tr} A_0^2$ we derive from (4.16)

$$r_n^3(\operatorname{tr} A_n^2 - \operatorname{tr} A_0^2) \le r_n^2(\|gA_0 - q\|^2 - \|gA_n - q\|^2) \le r_n^2 k_n,$$

hence with (4.18) we get

$$\lim_{n\to\infty} r_n^{-1}(\|gA_0 - q\|^2 - \|gA_n - q\|^2) = 0,$$

which completes the proof.

5. Models where $V_1, \cdots, V_m$ span a commutative quadratic subspace. Here we assume now that $\sum := \text{span}[V_1, \cdots, V_m] = R(g^*)$ forms an m-dimensional commutative quadratic subspace in Sym, i.e., $A, B \in \sum$ implies $A^2 \in \sum$, $AB = BA$. By Lemma 6 of Seely (1971), page 714 a necessary and sufficient condition for a subspace $\sum$ to be an m-dimensional commutative quadratic subspace is the existence of m pairwise orthogonal projection matrices $P_1, \cdots, P_m$ that form a basis for $\sum$. Let these $P_1, \cdots, P_m$ be given, then there is a regular matrix $\Phi = (\phi_{ij})_{i=1,\cdots,m; j=1,\cdots,m}$ defined by

$$V_i = \textstyle\sum_{j=1}^m \phi_{ij} P_j \qquad \text{for} \quad i = 1, \cdots, m. \tag{5.1}$$

The problem of computing an estimation matrix $\hat{A}$ is here now reduced to an ordinary quadratic optimization problem in $\mathbb{R}^m$.

THEOREM 5.1. *Under the assumptions above the following is valid:*

(i) *The nonnegative minimum bias* MINQ *estimator $\hat{A}$ for $p'\alpha$ is given by*

$$\hat{A} = \textstyle\sum_{i=1}^m \hat{d}_i (\text{tr } P_i)^{-1} P_i \tag{5.2}$$

where $\hat{d} = (\hat{d}_1, \cdots, \hat{d}_m)' \in \mathbb{R}^m$ is the unique solution of the problem:

$$\text{minimize}_{d \in \mathbb{R}_+^m} (\Phi d - p)'(\Phi d - p), \tag{5.3}$$

respectively the unique solution of the following system:

$$\begin{cases} d \in \mathbb{R}_+^m \\ \Phi'(\Phi d - p) \in \mathbb{R}_+^m. \\ d'\Phi'(\Phi d - p) = 0 \end{cases} \tag{5.4}$$

(ii) *The corresponding estimation function has the representation*

$$\widehat{p'\alpha} := y'\hat{A}y = \textstyle\sum_{i=1}^m \hat{d}_i \|P_i z\|^2 / \text{tr } P_i, \tag{5.5}$$

where $y = (I - XX^+)z$ and z is defined in (2.1).

(iii) *If z is normally distributed, then $\widehat{p'\alpha}$ has uniformly minimum variance among all invariant estimates $y'A^0 y$ where A^0 is a nonnegative minimum bias estimator for $p'\alpha$.*

PROOF. Analogously to the definition of g let the linear operator h:Sym $\to \mathbb{R}^m$ be defined by $h: A \mapsto (\text{tr } AP_1, \cdots, \text{tr } AP_m)'$, then $h^*: \mathbb{R}^m \to$ Sym, $a \mapsto \sum_{i=1}^m a_i P_i$, and

$$hh^* = \begin{pmatrix} \text{tr } P_1 & & 0 \\ & \ddots & \\ 0 & & \text{tr } P_m \end{pmatrix}. \tag{5.6}$$

By (5.1), $\text{tr } AV_i = \sum_{j=1}^m \phi_{ij} \text{tr } AP_j$ for $i = 1, \cdots, m$, $A \in$ Sym, and so

$$g = \Phi h. \tag{5.7}$$

Denote $p_g := \text{Proj}_{R(g_{|\text{PSD}})} p$, then all matrices $A^0 \in$ PSD with $gA^0 = p_g$ give minimum bias with respect to PSD, $\|gA^0 - p\| = \min_{A \in \text{PSD}}!$, and of course, $gg^+p_g = p_g$. So with (5.7) we have $\Phi hA^0 = \Phi hg^+p_g$, respectively $hA^0 = hg^+p_g$ since Φ is regular, and $A^0 \in$ PSD implies $hA^0 \in \mathbb{R}_+^m$ and thus $hg^+p_g \in \mathbb{R}_+^m$. Now $R(g^+) = R(g^*)$ and $R(g^*) = \sum = R(h^*)$, hence there exists a $\hat{c} \in \mathbb{R}^m$ such that $g^+p_g = h^*\hat{c}$, and $hh^*\hat{c} \in \mathbb{R}_+^m$ yields with (5.6) that $\hat{c} \in \mathbb{R}_+^m$, implying $h^*\hat{c} \in$ PSD. Thus $g^+p_g \in$ PSD and since g^+p_g is the minimum norm solution of $gA = p_g$, $A \in$ Sym, we have

$$\hat{A} = g_{|\text{PSD}}^+ p = g^+p_g = h^*\hat{c}. \tag{5.8}$$

$\hat{c}$ is unique because $\hat{A}$ is unique and $P_1, \cdots, P_m$ are linearly independent. Then in minimizing $\|gA - p\|$ over PSD we can restrict ourselves to those A which are in $R(h^*_{|\mathbb{R}^m_+})$ and with (5.7) $\hat{c}$ minimizes (uniquely) $\|gh^*c - p\| = \|\Phi hh^*c - p\|$ over $\mathbb{R}^m_+$.

Note that $h^*c \in$ PSD implies tr $P_ih^*c = c_i$tr $P_i \geq 0$, $i = 1, \cdots, m$ and so $c \in \mathbb{R}^m_+$. Now the conditions given in Theorem 3.2 yield for $\hat{A} \in R(h^*)$:

$$\hat{A} = h^*\hat{c} \in \text{PSD} \Leftrightarrow \hat{c} \in \mathbb{R}^m_+,$$

$$g^*(g\hat{A} - p) = h^*\Phi'(\Phi hh^*\hat{c} - p) \in \text{PSD} \Leftrightarrow \Phi'(\Phi hh^*\hat{c} - p) \in \mathbb{R}^m_+,$$

$$\text{tr } \hat{A}g^*(g\hat{A} - p) = (\Phi hh^*\hat{c})'(\Phi hh^*\hat{c} - p) = 0;$$

(3.11), (3.12) are trivially fulfilled for $\hat{A} \in R(h^*) = R(g^*)$. The $\hat{c}$ satisfying these three conditions is unique because of (5.8). Putting $\hat{d} = hh^*\hat{c}$ then with (5.6) assertion (i) is shown. Denote $K = I - XX^+$, so $y = Kz$, $V_i = KU_iK$, $i = 1, \cdots, m$, cf., (2.2), and for all $B \in \sum$ we have $KB = BK = B$. Now $P_1, \cdots, P_m$ are in $\sum$, such that $y'P_iy = z'P_iz = (P_iz)'(P_iz)$ for $i = 1, \cdots, m$, which with (5.2) gives (ii). Note that since $\hat{A}K = \hat{A}$, minimizing tr A^2 subject to $gA = p_g$ leads to a K-locally minimum variance invariant unbiased estimate $y'\hat{A}y$ of $p_g'\alpha$, if z is normally distributed. Then because $\sum$ is a quadratic subspace, $y'\hat{A}y$ has uniformly minimum variance among all invariant unbiased estimates $y'Ay$, $A \in$ Sym, of $p_g'\alpha$, cf., for instance, Seely (1971), Section 3, Kleffe (1977), Section 7, and so in particular among all invariant unbiased estimates $y'A^0y$ of $p_g'\alpha$ with $A^0 \in$ PSD. Now these A^0 are the nonnegative minimum bias (invariant) estimators for $p'\alpha$, and the theorem is proved.

COROLLARY 5.1. (i) *With* $\hat{d}$ *defined in* (5.3), (5.4) *there holds*

$$\Phi\hat{d} = \text{Proj}_{R(g_{|\text{PSD}})}p(= p_g). \tag{5.9}$$

(ii) *The* MINQUE $\tilde{A}$ *for* $p'\alpha$ *is given by*, cf., *also Pukelsheim* (1979, 1981),

$$\tilde{A} = \textstyle\sum_{i=1}^m \tilde{d}_i(\text{tr } P_i)^{-1}P_i \tag{5.10}$$

where $\tilde{d} = (\tilde{d}_1, \cdots, \tilde{d}_m)' = \Phi^{-1}p$, *and*

$$\widetilde{p'\alpha} := y'\tilde{A}y = \textstyle\sum_{i=1}^m \tilde{d}_i\|P_iz\|^2/\text{tr } P_i. \tag{5.11}$$

Further we get useful formulas for computing expectation and covariance. Since $g\hat{A} = p_g$ and $E\widehat{p'\alpha} = \alpha'\alpha\,\hat{A}$ we have by (5.9).

$$E\widehat{p'\alpha} = \alpha'\Phi\hat{d}. \tag{5.12}$$

Let for $\rho = 1, 2$, $A_\rho = \sum_{i=1}^m d_{i\rho}(\text{tr } P_i)^{-1}P_i$, $\psi_\rho = y'A_\rho y$, then if y is normally distributed the covariance of ψ_1, ψ_2 is given by $\text{Cov}(\psi_1, \psi_2) = 2\,\text{tr}(A_1V_\alpha A_2V_\alpha)$ where $V_\alpha = \sum_{i=1}^m \alpha_iV_i$. But $V_\alpha = g^*\alpha = h^*\Phi'\alpha =: \sum_{i=1}^m \xi_iP_i$, $\xi = (\xi_1, \cdots, \xi_m)' = \Phi'\alpha$, and $A_\rho V_\alpha = \sum_{i=1}^m d_{i\rho}\xi_i(\text{tr } P_i)^{-1}P_i$, $A_1V_\alpha A_2V_\alpha = \sum_{i=1}^m d_{i1}d_{i2}\xi_i^2(\text{tr } P_i)^{-2}P_i$, so that

$$\text{Cov}(\psi_1, \psi_2) = 2\textstyle\sum_{i=1}^m d_{i1}d_{i2}\xi_i^2/\text{tr } P_i, \qquad \xi = \Phi'\alpha, \tag{5.13}$$

if y is normally distributed.

As an example let us now considerzhe balanced two-way nested classification model with random effects; cf., Pukelsheim (1979) for a characterization of the unbiasedly nonnegatively estimable linear forms.

EXAMPLE. The model is

$$z_{ij\nu} = \mu + a_i + b_{ij} + e_{ij\nu}, \qquad i = 1, \cdots, r > 1,$$
$$j = 1, \cdots, s > 1$$
$$\nu = 1, \cdots, t > 1$$
$$n = rst$$

where $a_1, \cdots, a_r, b_{11}, \cdots, b_{rs}, e_{111}, \cdots, e_{rst}$ are independent (1-dimensional) random variables with $Ea_i = Eb_{ij} = Ee_{ij\nu} = 0$ and $Ea_i^2 = \sigma_a^2$, $Eb_{ij}^2 = \sigma_b^2$, $Ee_{ij\nu}^2 = \sigma_e^2$, and $\mu \in \mathbb{R}$ is the mean value parameter. Denote $1_\kappa = (1, \cdots, 1)' \in \mathbb{R}^\kappa$, $J_\kappa = 1_\kappa 1_\kappa'$, $\bar{J}_\kappa = (1/\kappa)J_\kappa$, I_κ the $\kappa \times \kappa$ identity matrix, $K_\kappa = I_\kappa - \bar{J}_\kappa$, $\kappa \in \mathbb{N}$, and $\otimes$ the Kronecker product of two matrices. The model has then the equivalent representation

$$z = 1_n\mu + (I_r \otimes 1_s \otimes 1_t)a + (I_r \otimes I_s \otimes 1_t)b + (I_r \otimes I_s \otimes I_t)e,$$

respectively $z \sim (1_n\mu;\ \sigma_a^2 U_1 + \sigma_b^2 U_2 + \sigma_e^2 U_3)$, where $U_1 = I_r \otimes J_s \otimes J_t$, $U_2 = I_r \otimes I_s \otimes J_t$, $U_3 = I_r \otimes I_s \otimes I_t$. Since for a pseudoinverse there holds $X^+ = (X'X)^+X'$ we have $1_n^+ = (1/n)1_n'$ and so $\mathrm{Proj}_{R(1_n)^\perp} = I_n - (1/n)1_n1_n' = K_n$.

With the usual notation, e.g., $\bar{z}_{i\cdot\cdot} = (1/st)\sum_{j=1}^s \sum_{\nu=1}^t z_{ij\nu}$, there is the following identity, cf., Graybill (1976), page 634,

$$z_{ij\nu} - \bar{z}_{\cdot\cdot\cdot} = (\bar{z}_{i\cdot\cdot} - \bar{z}_{\cdot\cdot\cdot}) + (\bar{z}_{ij\cdot} - \bar{z}_{i\cdot\cdot}) + (\bar{z}_{ij\nu} - \bar{z}_{ij\cdot}),$$

respectively the orthogonal decomposition

$$K_n z = P_1 z + P_2 z + P_3 z,$$

where

$$P_1 = K_r \otimes \bar{J}_s \otimes \bar{J}_t, \quad P_2 = I_r \otimes K_s \otimes \bar{J}_t, \quad P_3 = I_r \otimes I_s \otimes K_t$$

are pairwise orthogonal projection matrices; note that I_κ, $\bar{J}_\kappa$, K_κ are projectors and $\bar{J}_\kappa K_\kappa = K_\kappa \bar{J}_\kappa = 0$. Further $\|P_1 z\|^2 = st \cdot \sum_{i=1}^r (\bar{z}_{i\cdot\cdot} - \bar{z}_{\cdot\cdot\cdot})^2$, $\|P_2 z\|^2 = t \cdot \sum_{i=1}^r \sum_{j=1}^s (\bar{z}_{ij\cdot} - \bar{z}_{i\cdot\cdot})^2$, $\|P_3 z\|^2 = \sum_{i=1}^r \sum_{j=1}^s \sum_{\nu=1}^t (z_{ij\nu} - \bar{z}_{ij\cdot})^2$, and $\mathrm{tr}\, P_1 = r - 1$, $\mathrm{tr}\, P_2 = r(s-1)$, $\mathrm{tr}\, P_3 = rs(t-1)$, $\mathrm{tr}\, K_n = n - 1$.

Using the above decomposition of K_n we easily get

$$V_1 = K_n U_1 K_n = K_r \otimes J_s \otimes J_t = stP_1,$$

$$V_2 = K_n U_2 K_n = K_r \otimes \bar{J}_s \otimes J_t + I_r \otimes K_s \otimes J_t = tP_1 + tP_2,$$

$$V_3 = K_n U_3 K_n = K_n = P_1 + P_2 + P_3,$$

so that

$$\Phi = \begin{pmatrix} st & 0 & 0 \\ t & t & 0 \\ 1 & 1 & 1 \end{pmatrix}, \quad \Phi'\Phi = \begin{pmatrix} s^2t^2 + t^2 + 1 & t^2 + 1 & 1 \\ t^2 + 1 & t + 1 & 1 \\ 1 & 1 & 1 \end{pmatrix}, \quad \Phi'p = \begin{pmatrix} p_1 st + p_2 t + p_3 \\ p_2 t + p_3 \\ p_3 \end{pmatrix}$$

and the system (5.4) is solved, for instance, for $p = p_a = (1, 0, 0)'$ by $\hat{d}(a) = (st/(s^2t^2 + t^2 + 1), 0, 0)'$, $p = p_b = (0, 1, 0)'$ by $\hat{d}(b) = (0, t/(t^2 + 1), 0)'$, $p = p_e = (0, 0, 1)'$ by $\hat{d}(e) = (0, 0, 1)'$. Then the estimates according to (5.5) are given by $\hat{\sigma}_a^2 = st\|P_1 z\|^2/(s^2t^2 + t^2 + 1)(r - 1)$, $\hat{\sigma}_b^2 = t\|P_2 z\|^2/(t^2 + 1)r(s - 1)$, $\hat{\sigma}_e^2 = \|P_3 z\|^2/rs(t - 1)$. With respect to (5.10) we get for the MINQUE's, $\tilde{d}(a) = \Phi^{-1}p_a = (1/st, -1/st, 0)'$, $\tilde{d}(b) = \Phi^{-1}p_b = (0, 1/t, -1/t)'$, $\tilde{d}(e) = \Phi^{-1}p_e = \hat{d}(e)$, and the correspondent estimates are given then by (5.11), which are the usual ANOVA estimates, cf., for instance, Corbeil and Searle (1976), Graybill (1976), page 635, and are denoted by $\tilde{\sigma}_a^2, \tilde{\sigma}_b^2, \tilde{\sigma}_e^2$. Now using (5.12) we get

$$E\hat{\sigma}_a^2 = \sigma_a^2/(1 + 1/s^2 + 1/s^2t^2) + \sigma_b^2/(s + 1/s + 1/st^2)$$
$$+ \sigma_e^2/(st + t/s + 1/st) \to \sigma_a^2 \quad \text{for} \quad s \to \infty,$$
$$E\hat{\sigma}_b^2 = \sigma_b^2/(1 + 1/t^2) + \sigma_e^2/(t + 1/t) \to \sigma_b^2 \quad \text{for} \quad t \to \infty.$$

If z is normally distributed we have by use of (5.13)

$$\mathrm{Var}(\hat{\sigma}_a^2) = 2\{(\sigma_a^2 + \sigma_b^2/s + \sigma_e^2/st)/(1 + 1/s^2 + 1/s^2t^2)\}^2/(r - 1)$$
$$\to \begin{cases} 0 & \text{for } r \to \infty, \\ 2(\sigma_a^2)^2/(r - 1) & \text{for } s \to \infty \end{cases}$$

$$\operatorname{Var}(\hat{\sigma}_b^2) = 2\{(\sigma_b^2 + \sigma_e^2/t)/(1 + 1/t^2)\}^2/r(s-1)$$
$$\to \begin{cases} 0 & \text{for } rs \to \infty, \\ 2(\sigma_b^2)^2/r(s-1) & \text{for } t \to \infty \end{cases}$$

and for the ANOVA estimates, cf., also Corbeil and Searle (1976),

$$\operatorname{Var}(\tilde{\sigma}_a^2) = 2\{[\sigma_a^2 + \sigma_b^2/s + \sigma_e^2/st]^2/(r-1) + [\sigma_b^2/s + \sigma_e^2/st]^2/r(s-1)\}$$
$$\to \begin{cases} 0 & \text{for } r \to \infty, \\ 2(\sigma_a^2)^2/(r-1) & \text{for } s \to \infty \end{cases}$$
$$\operatorname{Var}(\tilde{\sigma}_b^2) = 2\{[\sigma_b^2 + \sigma_e^2/t]^2/r(s-1) + (\sigma_e^2/t)^2/rs(t-1)\}$$
$$\to \begin{cases} 0 & \text{for } rs \to \infty, \\ 2(\sigma_b^2)^2/r(s-1) & \text{for } t \to \infty \end{cases}$$

$\operatorname{Var}(\tilde{\tilde{\sigma}}_e^2) = 2(\sigma_e^2)^2/rs(t-1) \to 0$ for $rst \to \infty$; further $\hat{\sigma}_a^2$, $\hat{\sigma}_b^2$, $\tilde{\tilde{\sigma}}_e^2$ and $\tilde{\sigma}_a^2$, $\tilde{\tilde{\sigma}}_e^2$ are independent, whereas

$$\operatorname{Cov}(\tilde{\sigma}_a^2, \tilde{\sigma}_b^2) = -2(\sigma_b^2 t + \sigma_e^2)^2/rst^2(s-1),$$
$$\operatorname{Cov}(\tilde{\sigma}_b^2, \tilde{\tilde{\sigma}}_e^2) = -2(\sigma_e^2)^2/rst(t-1).$$

Now for the normally distributed z we get the following distributions from Scheffé (1959), Section 7.6, by putting there $\sigma_e^2 = 0$, $\sigma_T^2 = \sigma_e^2$, $\sigma_B^2 = \sigma_b^2$, $\sigma_C^2 = \sigma_a^2$, and $I = r$, $J = s$, $K = t$, $M = 1$,

$$\hat{\sigma}_a^2 \sim [(r-1)(st + t/s + 1/st)]^{-1}(\sigma_e^2 + t\sigma_b^2 + st\sigma_a^2)\chi_{r-1}^2,$$
$$\hat{\sigma}_b^2 \sim [r(s-1)(t + 1/t)]^{-1}(\sigma_e^2 + t\sigma_b^2)\chi_{r(s-1)}^2,$$
$$\tilde{\tilde{\sigma}}_e^2 \sim [rs(t-1)]^{-1}\sigma_e^2\chi_{rs(t-1)}^2.$$

These distributions can also be deduced from the general Theorem 7.2 of Rayner and Livington (1965), cf. also Searle (1971), page 69.

Similarly in other balanced ANOVA models the estimates can be derived explicitly using the well-known formulas for the "partitioning of the sum of squares"

Acknowledgment. The author would like to thank the referees for their comments and valuable suggestions, and F. Pukelsheim for a comment leading to an improvement in Section 3.

REFERENCES

[1] BERMAN, A. (1973). *Cones, Matrices and Mathematical Programming*. Springer, Berlin/Heidelberg/New York.
[2] BROWN, K. G. (1976). Asymptotic behavior of MINQUE-type estimators of variance components. *Ann. Statist.* **4** 746-754.
[3] CHIPMAN, J. S. (1964). On least squares with insufficient observations. *J. Amer. Statist. Assoc.* **59** 1078-1111.
[4] CORBEIL, R. R. and SEARLE, S. R. (1976). A comparison of variance component estimators. *Biometrics* **32** 779-791.
[5] DRYGAS, H. (1972). The estimation of residual variance in regression analysis. *Math. Operationsforsch. Statist.* **3** 373-388.
[6] DRYGAS, H. (1977). Best quadratic unbiased estimation in variance-covariance component models. *Math. Operationsforsch. Statist.* **8** 211-231.
[7] FEDERER, W. T. (1968). Non-negative estimators for components of variance. *J. Roy. Statist. Soc., Ser. C, Appl. Statist.* **17** 171-174.
[8] GRAYBILL, F. A. (1976). *Theory and Application of the Linear Model*. Duxbury Press, North Scituate, Massachusetts.
[9] HARTLEY, H. O., RAO, J. N. K. and LAMOTTE, L. R. (1978). A simple "synthesis" based method of variance component estimation. *Biometrics* **24** 233-242.

[10] HARTUNG, J. (1978). On nonnegative estimation of variance components. Presented at 11th European meeting of statisticians, Oslo.
[11] HARTUNG, J. (1979). A note on restricted pseudoinverses. *SIAM J. Math. Anal.* **10** 266–273.
[12] HARVILLE, D. N. (1977). Maximum likelihood approaches to variance component estimation and to related problems. *J. Amer. Statist. Assoc.* **72** 320–340.
[13] HOLMES, R. B. (1972). *A Course on Optimization and Best Approximation.* Springer, Berlin/Heidelberg/New York.
[14] KLEFFE, J. (1977). Optimal estimation of variance components —a survey. *Sankyā Ser. B* **39** 211–244.
[15] LAMOTTE, L. R. (1973a). Quadratic estimation of variance components. *Biometrics* **29** 311–330.
[16] LAMOTTE, L. R. (1973b). On nonnegative quadratic unbiased estimation of variance components. *J. Amer. Statist. Assoc.* **68** 728–730.
[17] LUENBERGER, D. G. (1969). *Optimization by Vector Space Methods.* Wiley, New York/London/Sydney/Toronto.
[18] MCHUGH, R. B. and MIELKE, P. W. JR. (1968). Negative variance estimates and statistical dependence in nested sampling. *J. Amer. Statist. Assoc.* **63** 1000–1003.
[19] MINAMIDE, N. and NAKAMURA, K. (1970). A restricted pseudoinverse and its application to constrained minima. *SIAM J. Appl. Math.* **19** 167–177.
[20] MITRA, S. K. (1975). Optimal inverse of a matrix. *Sankyā Ser. A* **37** 550–563.
[21] MOREAU, J. J. (1964). Théorèmes "inf-sup". *C. R. Acad. Sci. Paris* **258** 2720–2722.
[22] PUKELSHEIM, F. (1976). Estimating variance components in linear models. *J. Multivariate Anal.* **6** 626–629.
[23] PUKELSHEIM, F. (1977). Linear models and convex programs: unbiased nonnegative estimation in variance component models. Technical Report No. 104, Depart. Statist., Stanford Univ.
[24] PUKELSHEIM, F. (1979). Classes of linear models. In *Proceedings of Variance Components and Animal Breeding: A Conference in Honor of C. R. Henderson.* (S. R. Searle and L. D. Vanvleck, eds.). 69–83. Animal Science Dept., Cornell Univ.
[25] PUKELSHEIM, F. (1981). On the existence of unbiased nonnegative estimates of variance covariance components. *Ann. Statist.* **9** 293–299.
[26] RAO, C. R. (1970). Estimation of heteroscedastic variances in linear models. *J. Amer. Statist. Assoc.* **65** 161–172.
[27] RAO, C. R. (1972). Estimation of variance and covariance components in linear models. *J. Amer. Statist. Assoc.* **67** 112–115.
[28] RAO, C. R. (1973). *Linear Statistical Inference and its Applications, 2nd ed.* Wiley, New York/London/Sydney/Toronto.
[29] RAO, J. N. K. (1973). On the estimation of heteroscedastic variances. *Biometrics* **29** 11–24.
[30] RAO, J. N. K. and SUBRAHMANIAM, K. (1971). Combining independent estimators and estimation in linear regression with unequal variances. *Biometrics* **27** 971–990.
[31] RAYNER, A. A. and LIVINGSTONE, D. (1965). On the distribution of quadratic forms in singular normal variables. *S. African J. Agric. Sci.* **8** 357–370.
[32] SCHEFFÉ, H. (1959). *The Analysis of Variance.* Wiley, New York/London/Sydney.
[33] SEARLE, S. R. (1971). *Linear Models.* Wiley, New York/London/Sydney/Toronto.
[34] SEELY, J. (1971). Quadratic subspaces and completeness. *Ann. Math. Statist.* **42** 710–721.
[35] THOMPSON, W. A., JR. (1962). The problem of negative estimates of variance components. *Ann. Math. Statist.* **33** 273–289.

ABTEILUNG STATISTIK
UNIVERSITÄT DORTMUND
POSTFACH 500500
D-4600 DORTMUND 50
WEST GERMANY