PROCEEDINGS OF A CONFERENCE
IN HONOR OF C. R. HENDERSON
JULY 16-17, 1979

VARIANCE COMPONENTS AND ANIMAL BREEDING

EDITED BY
L. Dale Van Vleck
Shayle R. Searle

CORNELL UNIVERSITY, *ITHACA, N.Y.*
1979

CLASSES OF LINEAR MODELS

Friedrich Pukelsheim

Institut für Mathematische Stochastik der Universität
D-7800 Freiburg im Breisgau, Federal Republic of Germany

Summary

Point estimation is used to classify linear models with the help of the matrices that define mean and dispersion structure. Emphasis is laid on those aspects that are common to both mean and dispersion estimation. The problems of non-negative definite estimates of the dispersion matrix, and of non-negative estimates of variance components receive special attention. Some open questions conclude the paper.

List of Sections

Key words and phrases: Point estimation of mean and dispersion parameters; LS, BLUE, UMVU; dispersion mean correspondence; MINQUE, MIVQUE, REML; non-negative definite estimates of the dispersion matrix; non-negative estimates of variance components; linear model, residual model, Aitken model, derived model, special model.

AMS 1970 subject classification: 62J05, 62J10.

1. *The Moment Representation of Linear Models*

For many purposes of estimation the essentials of a *linear model* are reflected best by its moment representation

$$Y \sim \left(\sum_{i=1}^{k} \beta_i x_i ; \quad \sum_{j=1}^{\ell} \tau_j V_j \right). \qquad \text{(LM)}$$

Here it is implicitly understood that

- n real observations $Y_1,\dots,Y_n$ form the random R^n-vector Y,
- the k real R^n-vectors $x_1,\dots,x_k$, and the ℓ real symmetric $n \times n$ matrices $V_1,\dots,V_\ell$ are given and fixed,
- and Y has mean vector $\Sigma\beta_i x_i$ and dispersion matrix $\Sigma\tau_j V_j$.

Greek characters indicate unknown quantities, i.e.,

- $\beta = (\beta_1,\dots,\beta_k)'$ is the mean parameter,
- $\tau = (\tau_1,\dots,\tau_\ell)'$ is the dispersion parameter, and
- $\theta = (\beta,\tau)$ is the full parameter of the linear model (LM).

For β the natural parameter set then is the unrestricted space R^k, and for τ it is the set $\bar{G}$ of those R^ℓ-vectors t for which $\Sigma t_j V_j$ is non-negative definite. It is assumed that at least one combination $\Sigma t_j V_j$ is positive definite. As usual, X is the $n \times k$ matrix with columns $x_1,\dots,x_k$.

Certainly a *normal linear model* $Y \sim N_n(X\beta;\Sigma\tau_j Y_j)$ satisfies (LM), and there are more interesting examples. Seely & Zyskind (1971, p. 693) point out that the "mixed model" of the analysis of variance falls under (LM), including "fixed effect models" and "random effect models". It also covers the model where Y equals $X + Zu + e$, with a zero-mean random part $Zu + e$. Harville (1977, p. 321) stresses the applicability of this "general linear model" to the analysis of variance, multivariate statistical analysis, time series, and factor analysis, and those remarks then also pertain to (LM). In the multivariate case the grand observation vector Y has a particular structure in that it is built up from p-variate subvectors $Y_1,\dots,Y_N$; for examples see Kleffe (1977, p. 214), Searle (1978, p. 183).

In a linear model (LM) interest often concentrates on a function $c'\beta$ or the mean parameter β, where the coefficient R^k-vector c is specified in advance. With the distributional assumptions underlying (LM) this interest is unambiguous only if the function $c'\beta$ is identifiable. This means that when the expected value of the observations Y with respect to some fixed underlying distribution may be represented as both $\Sigma a_i x_i$ and $\Sigma b_i x_i$ with two vectors $a = (a_1,\dots,a_k)'$ and $b = (b_1,\dots,b_k)'$ in R^k, then the values $c'a$ and $c'b$ also coincide. Otherwise Y might be governed by some distribution to which there corresponds no unique value of the function $c'\beta$, and this function would be meaningless in the assumed model.

In the same sense a function $q'\tau$ of the dispersion parameter τ is identifiable if $\Sigma s_j V_j = \Sigma t_j V_j$ implies $q's = q't$ for all vectors $s = (s_1,\dots,s_\ell)'$ and $t = (t_1,\dots,t_\ell)'$ in $\bar{G}$. The following lemma gives a characterization of identifiable functions $c'\beta$ and $q'\tau$.

LEMMA 1. *For a function $c'\beta$ the following three statements are equivalent:*

(1) *$c'\beta$ is identifiable in the linear model* (LM).

(2) *There exists some R^n-vector a such that for all $i = 1,\dots,k$ one has $c_i = a'x_i$.*

(3) *c lies in the range of the $k \times k$ Gramian matrix $X'X = ((x_i'x_j))$.*

For a function $q'\tau$ the following three statements are equivalent:

(4) *$q'\tau$ is identifiable in the linear model* (LM).

(5) *There exists some real $n \times n$ matrix A such that for all $j = 1,\dots,\ell$ one has $q_j = \operatorname{trace} A'V_j$.*

(6) *q lies in the range of the $\ell \times \ell$ Gramian matrix $S = ((\operatorname{trace} V_iV_j))$.*

PROOF. For $c'\beta$, identifiability means that the nullspace of X be contained in the nullspace of c'. Equivalently, c must be a member of the range of X', or of the range of $X'X$. For $q'\tau$ identifiability reduces to the same argument. Namely, assume $\Sigma a_j V_j$ is 0 for some vector a in R^ℓ. Since the interior of $\bar{G}$ is non-empty there exist s, $t \in \bar{G}$ such that $a = s - t$. Hence $\Sigma s_j V_j = \Sigma t_j V_j$, and $q'a = q's - q't = 0$. Now follow the argument for $c'\beta$. _/

2. *Classes of Estimators*

In the general theory any measurable function $a(Y)$ from the sample space R^n into the range of the function $c'\beta$ is taken to be an estimate for $c'\beta$. However, linear model theory faces the following alternative before a manageable theory is possible: Either attention is restricted to a precisely defined class of distributions, this is done in the normal linear model; or the class of all estimates is drastically restricted to some subclass which admits an investigation solely on the grounds of the moment assumptions (LM). It is this second course of action which is now adopted: Only estimates are considered which are *linear* functions of the observations, i.e., which can be written as $a'Y$ for some R^n-vector a.

This *restriction by linearity* has its analogue when a function $q'\tau$ of the dispersion parameter τ is to be estimated: Only estimates are considered which are *quadratic* functions of the observations, i.e., which can be written as $Y'AY$ for some real symmetric $n \times n$ matrix A. For the image $q'(\overline{G})$ of the natural parameter set $\overline{G}$ under a non-zero function $q'\tau$ one faces two possibilities: Either it is the closed half-ray $R_+ = [0, +\infty[$ of all non-negative numbers, then the function $q'\tau$ will be said to be *non-negative* on $\overline{G}$; or it is the full real line R. For a function $q'\tau$ which is non-negative on $\overline{G}$ any quadratic estimate $Y'AY$ must have a non-negative definite matrix A.

One often wishes to treat mean parameter β and dispersion parameter τ separately in two stages. Such a separation is the effect of a *restriction by translation-invariance*, i.e., a restriction to those estimates for $q'\tau$ that remain unchanged whenever y is translated into $y + Xb$. If M is defined to be the symmetric idempotent matrix $I_n - XX^+$, then the statistic MY is maximal invariant with respect to this translation group $\{y \to y + Xb | b \in R^k\}$ (Seely 1971, p. 718). Since the zero-mean statistic $Z = MY$ coincides with the residual vector from a simple least squares fit for $X\beta$, the model it generates will be called the *residual model:*

$$Z = MY \sim (0; \Sigma\tau_j MV_j M) \quad , \qquad M = I_n - X(X'X)^- X' \quad . \qquad \text{(RM)}$$

Here the sole remaining parameter is τ, and its new natural parameter set in this model (RM) is the set $\overline{G_M}$ of those R^ℓ-vectors t for which $\Sigma t_j MV_j M$ is non-negative definite. A priori, this parameter set $\overline{G_M}$ — which is larger than $\overline{G}$ — does not have any interpretation in the original model (LM), but it turns out to be closely related to the existence of translation-invariant quadratic estimates for $q'\tau$ which are also unbiased.

This *restriction by unbiasedness*, both for functions $c'\beta$ of the mean parameter β and for functions $q'\tau$ of the dispersion parameter τ, is a commonly accepted constraint, particularly in the analysis of variance. Note that unbiasedness and non-negative definiteness of a quadratic estimate $Y'AY$ for $q'\tau$ automatically entail translation-invariance (Balestra 1973, p. 25). In general, however, the requirement of translation - invariance means a genuine restriction. Given two functions $c'\beta$ and $q'\tau$, the classes of estimates to be investigated now are formed by

- all unbiased linear estimates for $c'\beta$, and
- all unbiased translation-invariant quadratic estimates for $q'\tau$.

A natural first question is when estimates with these properties exist.

THEOREM 1. *For a function* $c'\beta$ *the following two statements are equivalent:*

(1) *There exists an unbiased linear estimate for* $c'\beta$.

(2) $c'\beta$ *is identifiable in the original model* (LM).

For a function $q'\tau$ *which is* not *non-negative on* $\overline{G}$ *the following two statements are equivalent:*

(3) *There exists an unbiased translation-invariant quadratic estimate for* $q'\tau$.

(4) $q'\tau$ *is identifiable in the residual model* (RM).

For a function $q'\tau$ *which is non-negative on* $\overline{G}$ *the following two statements are equivalent:*

(5) *There exists an unbiased translation-invariant non-negative definite quadratic estimate for* $q'\tau$.

(6) $q'\tau$ *is non-negative even on* $\overline{G_M}$, *i.e.,* $q't \geq 0$ *for all* $t \in \overline{G_M}$.

PROOF. The first two equivalences are standard, the last two are proved in Pukelsheim (1979b). _/

3. *The Dispersion Mean Correspondence in the Residual Model*

Once a certain class of estimates is seen to be non-empty a natural second question is to ask for optimal estimates in this class. A less problematic situation for the mean parameter β arises in the *Aitken model*

$$Y \sim (X\beta;\ \sigma^2 V) \ . \qquad \text{(AM)}$$

This is the special case $\ell = 1$ of the general linear model (LM), but suffices to illustrate the relevant arguments needed below. The central result is the following well known Gauss-Markov type theorem, for more details and related versions see the discussion by Zyskind (1975, pp. 653-661).

THEOREM 2. *Assume an Aitken model* $Y \sim (X\beta;\ \sigma^2 V)$.

(a) *The weighted least squares estimate* $c'(X'V^+X)^+X'V^+Y$ *is unbiased for* $c'\beta$ *and of minimal variance among all unbiased linear estimates for* $c'\beta$, *for all* R^k*-vectors* c, *if and only if* $\text{rank } X = k$ *and* $\text{range } X \subset \text{range } V$.

(b) *The simple least squares estimate* $c'X^+Y$ *is unbiased for* $c'\beta$ *and of minimal variance among all unbiased linear estimates for* $c'\beta$, *for all* R^k*-vectors* c, *if and only if* $\text{rank } X = k$ *and* $\text{range } VX \subset \text{range } X$. _/

The simplest example is provided by the classical linear model $Y \sim (X\beta;\ \sigma^2 I_n)$. However, Theorem 2 also covers estimation of functions $q'\tau$ of the dispersion parameter τ. To this end recall the following. For some R^n-vector a and some R^p-vector b their Kronecker product $a \otimes b$ is the R^{np}-vector $(a_1b_1, a_1b_2, \ldots, a_1b_p, a_2b_1, \ldots, a_nb_p)'$. The same entries a_ib_j occur in the matrix ab', and the transition between both types of arrangement is easily handled by means of the transformation $\text{vec } ab' = a \otimes b$. More general, $\text{vec } A$ is taken to be the column vector obtained from writing the rows of the matrix A one behind the other and transposing the resulting row vector. The identity $\text{trace } A'B = (\text{vec } A)'(\text{vec } B)$ then is immediate. This formalism applies to translation-invariant quadratic estimates $Z'AZ$ in the residual model (RM). Abbreviating $\text{vec } A$ by a, the estimate $Z'AZ$ may be rewritten as $a'(Z \otimes Z)$. Moreover, the random variable $Z \otimes Z$ which occurs here has expectation $\Sigma\tau_j \text{ vec } MV_jM$.

Assuming the dispersion matrix of $Z \otimes Z$ to be F_M one arrives at the *derived model*

$$Z \otimes Z \sim (\Sigma\tau_j \text{ vec } MV_jM;\ F_M) \ . \qquad \text{(DM)}$$

The dispersion parameter τ of the original model (LM) now appears as the mean parameter in the derived model (DM), and estimates for $q'\tau$ which are quadratic functions of the residual statistic Z appear as linear functions of the derived random variable $Z \otimes Z$. Hence Theorem 2 applies and yields a full battery of estimates for $q'\tau$, depending on the choice of the fourth moment matrix F_M.

An example may illustrate this approach: In an Aitken model $Y \sim (X\beta;\ \sigma^2 V)$ the quadratic form $Y'(V^+ - V^+X(X'V^+X)^+X'V^+)Y/(n - \text{rank } X)$ is the unbiased translation-invariant non-negative definite quadratic estimate for σ^2 which under normality is of minimal variance in its class, provided the range of V contains the range of X. To see this note that under normality the dispersion matrix F_M may be written as $2\sigma^4 MVM \otimes MVM$. In Theorem 2(a) the first condition $\text{rank vec } MVM = 1$ translates into $MVM \neq 0$ and this is satisfied when $n > \text{rank } X$. The second condition $\text{vec } MVM \in \text{range } MVM \otimes MVM$ holds true because of $\text{vec } MVM = (MVM \otimes MVM)(\text{vec } (MVM)^+)$. Thus the weighted least squares estimate for σ^2 in the derived model has the desired optimality properties, and a short calculation shows it to be $Z'(MVM)^+Z/(\text{trace } (MVM)^+MVM)$. Using the assumption $\text{range } X \subset \text{range } V$ this is the quadratic form given above.

More detailed formulae are given in Pukelsheim (1976). The fourth moment matrix F_M takes on a special form when certain assumptions are made concerning the coefficients of skewness and kurtosis, see the discussion of *Hsu's model* in Pukelsheim (1977). In fact, the idea of derived models also extends to estimating these third and fourth moment coefficients, as demonstrated in Pukelsheim (1979a). Seely (1970) first elaborated that the problems of mean and dispersion estimation share the same structure. The common denominator is found to be regression analysis in finite dimensional linear spaces with inner product; points may then be realized as column vectors for mean estimation, or as symmetric matrices for dispersion estimation.

Kronecker products appear in both multivariate analysis (Eaton 1970) and in the dispersion mean correspondence above, but this happens for two quite different reasons. By its intrinsic nature the Kronecker product provides a correspondence between bilinear — and hence quadratic — forms and linear forms (Lang 1966, pp. 221-227). Only by mere coincidence it can also be used to simplify direct sums as they appear in multivariate analysis, for instance writing a block diagonal matrix Diag(V:...:V) as $I_N \otimes V$.

The dispersion mean correspondence may also serve to classify methods of dispersion estimation analoguously to those methods known for mean estimation: C.R. Rao's (1973, p. 303) MINQUE theory corresponds to simple least squares estimation in the derived model (DM). Weighted MINQUE, or MINQUE given V in the terminology of Kleffe (1977, p. 223) is weighted least squares estimation with the fourth moment matrix taken to be $V \otimes V$, hence under normality this coincides with minimum variance estimation (MIVQUE) when the true dispersion matrix is V.

4. *A Corresponding Look at Henderson's Methods of Estimating Variance Components*

Motivated by applied problems statistical practitioners have preceded or initiated most of the development which has been achieved on a more formal level. For example, Kelm (1978) and Welsch (1978) noted that a procedure for estimating variance components described in a 1907 textbook on geodesic measurements (Helmert 1907, pp. 358-363) coincides, in fact, with the MINQUE method introduced and formalized by C.R. Rao (1970) more than half a century later. Similarly methods I, II, III of Henderson (1953) have proved useful for variance component estimation long before their formal properties were explored in detail. R.D. Anderson (1978) meticulously recalls the history of variance component estimation.

Employing the dispersion mean correspondence one may translate the rationale underlying Henderson's methods into the terminology of mean estimation, thus gaining insight into their intrinsic properties.

As an abbreviation introduce the $n^2 \times \ell$ matrix D_M with column vectors vec MV_jM. Then in the derived model $Z \otimes Z \sim (D_M\tau; F_M)$ Searle's (1968, p. 749) summary of Henderson's methods reads as follows: *All of Henderson's three methods involve (i) calculating some ℓ linear forms $L'(Z \otimes Z)$, with an $n^2 \times \ell$ matrix L, say, (ii) obtaining their expectations $L'D_M\tau$, and (iii) solving linear equations in the unknown variance components, derived from equating these linear forms to their expected values, i.e., solving*

$$L'D_M\hat{\tau} = L'(Z \otimes Z) \quad . \qquad [*]$$

Seely (1970, p. 1744) points out that there is a question of whether this equation is consistent and whether its solution(s) yield unbiased estimates for τ. In any case an approximate solution is $\hat{\tau} = (L'D_M)^+L'(Z \otimes Z)$, and this is an unbiased estimate for τ if and only if $(L'D_M)^+L'D_M\tau = \tau$, or equivalently, (1) the $\ell \times \ell$ matrix $L'D_M$ is non-singular. In this case, equation [*] is consistent and has the unique solution $\hat{\tau} = (L'D_M)^{-1}L'(Z \otimes Z)$.

As mentioned by H.O. Hartley in the discussion of Searle's (1968, p. 780) paper, there is a considerable freedom which choice of the matrix L is found to be appropriate. Guided by the general theory it certainly seems advantageous to convert [*] into a normal equation. Hence for some non-negative definite matrix V let

$$L = (MVM \otimes MVM)^+D_M$$

be a particular choice. Then $L'D_M$ becomes the matrix $S_{(MVM)^+}$ with (i,j)-th entry trace $(MVM)^+V_i(MVM)^+V_j$, and $L'(Z \otimes Z)$ consists of the ℓ quadratic forms $Y'(MVM)^+V_j(MVM)^+Y$. The question of whether the estimate $\hat{\tau}$ thus obtained does have any optimality properties is answered by Theorem 2. Still demanding (1), i.e., rank $S_{(MVM)^+} = \ell$, the first condition in part (a) is satisfied. The second condition reads range $D_M \subset$ range $MVM \otimes MVM$, and follows if (2a) range $M \subset$ range V, as in Kleffe (1977, p. 223), or if (2b) all matrices V_j are non-negative definite and V equals Σt_jV_j with all coefficients t_j being positive, or if (2c) range $MV_jM \subset$ range MVM for all $j = 1,\ldots,\ell$. Under conditions (1), and (2a) or (2b) or (2c), $\hat{\tau}$ is under normality of minimum variance when the true dispersion matrix is V (Kleffe 1977, p. 223).

5. *Quadratic Subspaces of Symmetric Matrices*

The estimation space range $X = \{ \Sigma b_i x_i \mid b_1,\dots,b_k \in \mathbb{R} \}$ of the original model (LM) has its analogue in the derived model (DM), namely $\{ \text{vec } \Sigma t_j MV_j M \mid t_1,\dots,t_\ell \in \mathbb{R} \}$. For what follows it is preferable to write this latter set as vec $\mathcal{B}_M$, where

$$\mathcal{B}_M = \{ \Sigma\, t_j MV_j M \mid t_1,\dots,t_\ell \in \mathbb{R} \}$$

comprises all linear combinations of the matrices $MV_1M,\dots,MV_\ell M$. Under normality the fourth moment matrix F_M may be written as $2(\Sigma\tau_j MV_j M) \otimes (\Sigma\tau_j MV_j M)$. Whereas Theorem 2 initially applied to two separate matrices X and V in the Aitken model $Y \sim (X\beta;\ \sigma^2 V)$, it now quite differently sets the matrix $\Sigma\tau_j MV_j M$ into relation with itself. The notions and ideas for a comprehensive discussion of this situation are due to Seely (1971): $\mathcal{B}_M$ is said to be a *quadratic subspace of symmetric matrices* if $B \in \mathcal{B}_M$ implies $B^2 \in \mathcal{B}_M$. In the analysis of variance models one invariably finds that one combination $\Sigma t_j V_j$ yields the identity matrix I_n, thus justifying the assumption $M \in \mathcal{B}_M$. Theorem 2(b) now has an easy corollary on uniformly minimum variance unbiased (UMVU) estimation of the dispersion parameter τ.

COROLLARY 2.1. *For a derived model* (DM) *assume that the* $\ell \times \ell$ *matrix* $S_M = ((\text{ trace } MV_i MV_j))$ *is non-singular and that M is a member of* $\mathcal{B}_M$. *Then there exists an unbiased translation-invariant quadratic estimate for* $q'\tau$ *which under normality is of uniformly minimal variance, for all* $\mathbb{R}^\ell$*-vectors* q, *if and only if* $\mathcal{B}_M$ *forms a quadratic subspace of symmetric matrices. In this case the required estimate is* $Y'(\Sigma\lambda_j MV_j M)Y$, *with* $\lambda = S_M^{-1}q$.

PROOF. The estimate mentioned last has the minimum variance property in case the true dispersion matrix of Y is I_n, by Theorem 2 part (a). By part (b) this estimate is of uniformly minimum variance if and only if for all non-negative definite matrices A in $\mathcal{B}_M$ and for all B in $\mathcal{B}_M$ one has $ABA \in \mathcal{B}_M$. This is equivalent to $\mathcal{B}_M$ being a quadratic subspace (Seely 1971, Lemma 1(b), 2(c)). _/

A compact representation of the estimate of Corollary 2.1 is $q'\hat{\tau}$ where the vector statistic

$$\hat{\tau} = ((\text{ trace } MV_i MV_j))^{-1} \begin{pmatrix} Y'MV_1MY \\ \vdots \\ Y'MV_\ell MY \end{pmatrix}$$

will be called the *standard estimate for* τ. This $\hat{\tau}$ is the simple least squares estimate in the residual model (RM), note the resemblance with the simple least squares estimate $(X'X)^{-1}X'Y$ for β in the original model (LM).

When normality prevails a quadratic subspace condition is even more powerful than the above indicates. Seely (1971) proved that the class of normal distributions $N_n(0;\ \Sigma\tau_j MV_j M)$ then is an exponential family, and that the estimate $\hat{\tau}$ is a function of a complete sufficient statistic. Here the maximum likelihood estimate for τ also coincides with $\hat{\tau}$ if there are at least n - rank X replicates, see Pukelsheim & Styan (1979).

In summary, in a normal linear model with a quadratic subspace $\mathcal{B}_M$ the standard estimate $\hat{\tau}$ for τ is both the uniformly minimum variance unbiased (UMVU) translation-invariant estimate, thus coinciding with the estimate obtained from MINQUE, MIVQUE, or ANOVA methods, and the maximum likelihood translation-invariant (REML) estimate. It also respects the inherent constraint on τ and maps into the natural parameter set $\overline{G_M}$ as outlined in the next section.

The condition $M \in \mathcal{B}_M$ in Corollary 2.1 may be removed as in Drygas (1977). When $\mathcal{B}_M$ is not a quadratic subspace itself it still may be advantageous to study the smallest quadratic subspace containing $\mathcal{B}_M$, see Searle & H.V. Henderson (1979). The joint treatment of both mean and dispersion parameters becomes possible in a *special model* (SM), i.e., in a linear model (LM) in which the set $\mathcal{B} = \{ \Sigma t_j V_j \mid t_1,\dots,t_\ell \in \mathbb{R} \}$ forms a quadratic subspace of symmetric matrices, and range $VX \subset$ range X for all non-negative definite matrices V in $\mathcal{B}$; for details see the work of Seely (1971). As a rule of thumb any analysis of variance model with balanced numbers of observations satisfies these conditions, examples are given in Pukelsheim (1979b). Also many models from multivariate analysis are special models as just defined.

No method that exclusively relies on linear and multilinear algebra can adequately reflect the convex constraint that comes with the natural parameter $\overline{G_M}$. Nevertheless the estimated dispersion matrix does turn out to again be non-negative definite provided $\mathcal{B}_M$ forms a quadratic subspace of symmetric matrices and τ is estimated by its standard estimate $\hat{\tau}$. For with this estimate one has vec $\Sigma\hat{\tau}_j MV_j M = D_M D_M^+$ vec $MYY'M$, where the $n^2 \times \ell$ matrix D_M has columns vec MV_jM. This shows the estimated dispersion matrix $\Sigma\hat{\tau}_j MV_j M$ to be nothing but the projection of the sample dispersion matrix $MYY'M$ onto the subspace $\mathcal{B}_M$, and projections onto quadratic subspaces preserve non-negative definiteness (Pukelsheim 1979b, proof of Theorem 2).

Proper variance component estimation slightly changes the assumptions of a linear model (LM). Again using a moment representation a *variance component model* is given by

$$Y \sim (X\beta;\ \Sigma\sigma_j^2 V_j)\ ,\quad \text{all } V_j \text{ non-negative definite}\ , \qquad \text{(VCM)}$$

with parameter sets $\mathbb{R}^k$ for β, and the non-negative orthant $\mathbb{R}_+^\ell$ for $\tau = (\sigma_1^2,\ldots,\sigma_\ell^2)'$. Unbiased non-negative definite quadratic estimation of a single variance component σ_j^2 was first investigated by LaMotte (1973), a discussion for general functions $q'\tau$ is given in Pukelsheim (1979b). The results are most constructive again if $\mathcal{B}_M$ forms a quadratic subspace of symmetric matrices. Then the same *alternative on non-negative estimability* emerges regardless of whether $q'\tau$ is estimated in a linear model (LM) or in a variance component model (VCM): either the standard estimate $q'\hat{\tau}$ is unbiased and non-negative, or the concepts of unbiasedness and non-negative definiteness are incompatible.

As an example, take a 2-way nested classification random model $Y_{ij\ell} = \mu + \alpha_i + \beta_{ij} + e_{ij\ell}$. Let 1_n be the $\mathbb{R}^n$-vector consisting of one's only, and define $J_n = 1_n 1_n'$. This gives the representation $Y \sim (1_{abn}\mu;\ \sigma_\alpha^2 I_a \otimes J_b \otimes J_n + \sigma_\beta^2 I_a \otimes I_b \otimes J_n + \sigma_e^2 I_a \otimes I_b \otimes I_n)$ with $a, b, n > 1$. Then for a function $q_1\sigma_\alpha^2 + q_2\sigma_\beta^2 + q_3\sigma_e^2$ there exists an unbiased non-negative definite quadratic estimate if and only if $0 \le q_1/(bn) \le q_2/n \le q_3$ (Pukelsheim 1979b).

1. Is it possible to bring Henderson's methods into normal equation form [*] as outlined in Section 4? In other words: do there exist matrices V_h such that the ℓ quadratic forms that serve as a starting point for Henderson's method h collectively give the matrix $L = L(V_h)$ displayed in Section 4, for h = I, II, III? This would ensure local minimum variance properties of these estimates.

2. The restricted maximum likelihood estimate for τ always solves the likelihood equation (T.W. Anderson 1970, p. 5). Are there examples where the likelihood equation has more than one solution? This cannot happen when $\mathcal{B}_M$ forms a quadratic subspace of symmetric matrices (Pukelsheim & Styan 1979).

3. How should one decide the contest between unbiasedness and non-negative definiteness?

4. Which procedures result when with the dispersion mean correspondence biased (Ridge- and Stein-type) estimation of the mean is translated to dispersion estimation?

5. Are there more sensible loss functions for variance component estimation than squared error loss? To illustrate its deficiencies suppose that with the same squared error loss a true set of variance components $(\sigma_1^2,\ldots,\sigma_\ell^2)'$ is (i) overestimated to be $(\sigma_1^2+\varepsilon,\ldots,\sigma_\ell^2+\varepsilon)'$, or (ii) underestimated to be $(\sigma_1^2-\varepsilon,\ldots,\sigma_\ell^2-\varepsilon)'$. The statistician will then conclude that the data contain (i) more, or (ii) less variability than is actually true, and advice the experimenter to be (i) too cautious, or (ii) too trustful. Squared error weighs both reactions with the same loss $\ell\varepsilon^2$, although, in general, these decisions will have drastically different consequences: to be too cautious is inefficient, to be too trustful is dangerous.

References

Anderson, R.D. (1978). Studies on the estimation of variance components. Ph.D. Thesis, Cornell Univ.

Anderson, T.W. (1970). Estimation of covariance matrices which are linear combinations or whose inverses are linear combinations of given matrices. In *Essays in Probability and Statistics*. R.C. Bose, I.M. Chakravarti, P.C. Mahalanobis, C.R. Rao & K.J.C. Smith (eds.), The Univ. of North Carolina Press, Chapel Hill, 1-24.

Balestra, P. (1973). Best quadratic unbiased estimators of the variance-covariance matrix in normal regression. *Journal of Econometrics* 1, 17-28.

Drygas, H. (1977). Best quadratic unbiased estimation in variance-covariance component models. *Mathematische Operationsforschung und Statistik Series Statistics* 8, 211-231.

Eaton, M.L. (1970). Gauss-Markov estimation for multivariate linear models: a coordinate free approach. *The Annals of Mathematical Statistics* 41, 528-538.

Harville, D.A. (1977). Maximum likelihood approaches to variance component estimation and to related problems. *Journal of the American Statistical Association* 72, 320-340.

Helmert, F.R. (1907). *Die Ausgleichsrechnung nach der Methode der kleinsten Quadrate (Zweite Auflage)*. B.G. Teubner, Leipzig.

Henderson, C.R. (1953). Estimation of variance and covariance components. *Biometrics* 9, 226-252.

Kelm, R. (1978). Ist die Varianzschätzung nach Helmert MINQUE? *Allgemeine Vermessungs-Nachrichten* 85, 49-54.

Kleffe, J. (1977). Optimal estimation of variance components - a survey. *Sankhyā Series B* 39, 211-244.

LaMotte, L.R. (1973). On non-negative quadratic unbiased estimation of variance components. *Journal of the American Statistical Association* 68, 728-730.

Lang, S. (1966). *Linear Algebra*. Addison-Wesley Publishing Company, Reading, Massachusetts.

Pukelsheim, F. (1976). Estimating variance components in linear models. *Journal of Multivariate Analysis* 6, 626-629.

Pukelsheim, F. (1977). On Hsu's model in regression analysis. *Mathematische Operationsforschung und Statistik Series Statistics* 8, 323-331.

Pukelsheim, F. (1979a). Multilinear estimation of skewness and kurtosis in linear models. *Metrika*, to appear.

Pukelsheim, F. (1979b). On the existence of unbiased non-negative estimates of variance covariance components. Typescript, Univ. Freiburg im Breisgau.

Pukelsheim, F. & G.P.H. Styan (1979). Nonnegative definiteness of the estimated dispersion matrix in a multivariate linear model. *Bulletin de l'Académie Polonaise des Sciences. Série des Sciences Mathématiques, Astronomiques et Physiques*, to appear.

Rao, C.R. (1970). Estimation of heteroscedastic variance in linear models. *Journal of the American Statistical Association* 65, 161-172.

Rao, C.R. (1973). *Linear Statistical Inference and Its Applications. Second Edition*. John Wiley & Sons, New York.

Searle, S.R. (1968). Another look at Henderson's methods of estimating variance components. *Biometrics* 24, 749-787 (discussion 779-787).

Searle, S.R. (1978). A univariate formulation of the multivariate linear model. In *Contributions to Survey Sampling and Applied Statistics. Papers in Honor of H.O. Hartley*. H.A. David (ed.), Academic Press Inc., New York, 181-189.

Searle, S.R. & H.V. Henderson (1979). Dispersion matrices for variance components models. Paper No. BU-341 in the Biometrics Unit Mimeo Series, Cornell Univ.

Seely, J. (1970). Linear spaces and unbiased estimation - application to the mixed linear model. *The Annals of Mathematical Statistics* 41, 1735-1748.

Seely, J. (1971). Quadratic subspaces and completeness. *The Annals of Mathematical Statistics* 42, 710-721.

Seely, J. & G. Zyskind (1971). Linear spaces and minimum variance unbiased estimation. *The Annals of Mathematical Statistics* 42, 691-703.

Welsch, W. (1978). A posteriori Varianzschätzung nach Helmert. *Allgemeine Vermessungsnachrichten* 85, 55-63.

Zyskind, G. (1975). Error structures, projections and conditional inverses in linear model theory. In *A Survey of Statistical Design and Linear Models*. J.N. Srivastava (ed.), North-Holland Publishing Company, Amsterdam, 647-663.