International Journal of Pure and Applied Mathematics

Volume 45 No. 2 2008, 205-216

ON ESTIMATION IN MIXED MODELS

Radosław Kala

Department of Mathematical and Statistical Methods
Agricultural University of Poznań
28, Ul. Wojska Polskiego, Poznań, PL-60-637, POLAND
e-mail: kalar@au.poznan.pl

Abstract: In the paper the main results of estimation theory in mixed models are collected and shortly explained. The estimation methods of unknown parameters corresponding to the expectation of the observed vector random variable as well as its dispersion matrix are presented in a unified form. The possibility of transforming the variance-covariance components model to the simpler one is also indicated.

AMS Subject Classification: 62F10, 62J05, 62F11
Key Words: unbiasedness, translation invariance, BLUE, MINQE

1. Linear Models

By a linear model we will consider a vector random variable $\mathbf{y}$ with the expectation $E(\mathbf{y}) = \mu$ belonging to a given subspace $\mathcal{E}$ of R^n, the n-dimensional Euclidean space, and with the dispersion matrix $D(\mathbf{y}) = \mathbf{V}$ belonging to a given set $\mathcal{V}$ of S^n, the space of symmetric matrices of order n. Usually,

$$\mathcal{E} = \{\mu : \mu = \beta_1\mathbf{x}_1 + \beta_2\mathbf{x}_2 + \cdots + \beta_k\mathbf{x}_k = \mathbf{X}\beta\} = \mathcal{R}(\mathbf{X}),$$

where $\mathbf{X} = (\mathbf{x}_1 : \cdots : \mathbf{x}_k)$ is a known matrix, $\mathcal{R}(\mathbf{X})$ is its range, while $\beta = (\beta_1, \ldots, \beta_k)'$ is a vector of unknown parameters. Similarly,

$$\mathcal{V} = \{\mathbf{V} : \mathbf{V} = \alpha_1\mathbf{S}_1 + \alpha_2\mathbf{S}_2 + \cdots + \alpha_l\mathbf{S}_l = \mathbf{V}(\alpha) \geqslant \mathbf{0}\}, \qquad (1)$$

where $\mathbf{S}_1, \ldots, \mathbf{S}_l$ are known symmetric matrices, $\alpha = (\alpha_1, \ldots, \alpha_l)'$ is a vector of unknown parameters, while the inequality $\mathbf{V}(\alpha) \geqslant \mathbf{0}$ indicates that $\mathbf{V}(\alpha)$ is a non-negative definite matrix.

Received: February 12, 2008

The expectation of $\mathbf{y}$, as well the dispersion matrix of $\mathbf{y}$, are both expressed as linear combinations of known elements belonging to appropriate vector spaces, i.e. to R^n and to S^n, respectively. The coefficients of these combinations are parameters which are to be estimated with the use of $\mathbf{y}$ and the knowledge of the structures of $\mathcal{E}$ and of $\mathcal{V}$.

If the matrices spanning $\mathcal{V}$ are all non-negative definite, we write $\mathbf{V}_i$ instead of $\mathbf{S}_i$. If in addition α_i are considered as non-negative, then we write σ_i^2 instead α_i and call them variance components. The matrix

$$\mathbf{V}(\sigma) = \sigma_1^2\mathbf{V}_1 + \sigma_2^2\mathbf{V}_2 + \cdots + \sigma_l^2\mathbf{V}_l,$$

where $\sigma = (\sigma_1^2, \ldots, \sigma_l^2)'$, is obviously non-negative definite. The triplet

$$\{\mathbf{y}, \mathbf{X}\beta, \mathbf{V}(\sigma)\}$$

is then termed as the variance components model. To distinguish, the triplets with the dispersion structure as in (1) are called variance-covariance components models. Both types are called mixed models if, in addition, $\mathcal{E}$ is of dimension two at least. Such models were considered by many authors (see e.g. Rao and Kleffe [10], Sengupta and Jammalamadaka [11], and other references contained there). If $\mathcal{E}$ is spanned only by the vector of ones, $\mathcal{E} = \mathcal{R}(\mathbf{1})$, the model $\{\mathbf{y}, \mu\mathbf{1}, \mathbf{V}(\sigma)\}$ is termed as random, and finally, if $\mathcal{V}$ is of dimension one, i.e. $\mathbf{V}(\sigma) = \sigma^2\mathbf{V}$, the model $\{\mathbf{y}, \mathbf{X}\beta, \sigma^2\mathbf{V}\}$ is termed as fixed.

The expectation subspace $\mathcal{E}$ as well as the set $\mathcal{V}$ of dispersion matrices can be specified in many ways by choosing the spanning set of vectors in R^n and in S^n, respectively. These sets are usually determined by particular experimental goals, by the experimental conditions, or directly follows from the conducted experiment. When these sets are specified, the unknown parameters, receive specific practical meanings, which enable to express the final conclusions in a clear and concise form.

2. Sets of Parameters

The unknown coefficients $\beta_1, \ldots, \beta_k$ in the linear combination

$$\mu = \beta_1\mathbf{x}_1 + \beta_2\mathbf{x}_2 + \cdots + \beta_k\mathbf{x}_k = \mathbf{X}\beta,$$

are not restricted in any way. Thus the parameter set corresponding to the expectation μ coincide here with the whole R^k space.

When the columns of $\mathbf{X}$ are linearly dependent, the same expectation vector $\mu = \mathbf{X}\beta$ can be expressed with the use of many different vectors β. This means, that the parametrization of the expectation is overdone. In result, the sensible inference can be conducted only with respect to μ, or to any linear function of β which is expressible by μ, i.e. any function $\mathbf{p}'\beta$ such that $\mathbf{p}'= \mathbf{s}'\mathbf{X}$ for some vector $\mathbf{s} \in R^n$. Such functions fulfil the condition of identifiability in the sense of Rao and Kleffe [10]. This concept coincide with the estimability of $\mathbf{p}'\beta$, which means that there exists a linear statistic $\mathbf{s}'\mathbf{y}$ such that $E(\mathbf{s}'\mathbf{y}) = \mathbf{p}'\beta$.

The parametrization of the dispersion matrix looks alike,

$$\mathbf{V} = \alpha_1\mathbf{S}_1 + \alpha_2\mathbf{S}_2 + \cdots + \alpha_l\mathbf{S}_l = \mathbf{V}(\alpha),$$

but the unknown coefficients $\alpha_1, \ldots, \alpha_l$ are not free to vary over the whole R^l space. They are restricted by the condition that $\mathbf{V}(\alpha)$ is a non-negative definite matrix. Let $\mathcal{A}$ denote a set of acceptable parameters, i.e.

$$\mathcal{A} = \{\alpha \in R^l, \mathbf{V}(\alpha) \geqslant \mathbf{0}\}.$$

It is easy to note, that $\mathcal{A}$ is a convex cone. Indeed, if $\alpha \in \mathcal{A}$, then $s\alpha \in \mathcal{A}$ for any non-negative scalar s, and if $\alpha_1, \alpha_2 \in \mathcal{A}$, then $s\alpha_1 + (1-s)\alpha_2 \in \mathcal{A}$ for any $s \in [0,1]$. These features correspond to the properties of the set $\mathcal{V}$ which actually is also a convex cone in S^n.

The relations between $\mathcal{A}$ and $\mathcal{V}$ are more close. Let $\mathbf{S}_1$, $\mathbf{S}_2$, $\mathbf{S}_3$ be three matrices in S^2,

$$\mathbf{S}_1 = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}, \ \mathbf{S}_2 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \ \mathbf{S}_3 = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}.$$

If $\mathcal{V} = \{\alpha_1\mathbf{S}_1 + \alpha_2\mathbf{S}_2 + \alpha_3\mathbf{S}_3 \geqslant \mathbf{0}\}$, then

$$\mathcal{A} = \{(\alpha_1,\alpha_2,\alpha_3)', \alpha_1 \geqslant 0, \ \alpha_2 \geqslant 0, \ \alpha_1\alpha_3 \geqslant \alpha_2^2\}.$$

In this case

$$\dim S^2 = r(\{\mathbf{S}_1, \mathbf{S}_2, \mathbf{S}_3\}) = r(\mathcal{V}) = r(\mathcal{A}) = 3,$$

where $r(.)$ denotes the rank, i.e. the number of linearly independent elements of the set argument.

Now, let $\mathcal{V}_r$ will be spanned only by two matrices $\mathbf{R}_1 = \mathbf{S}_1 + \mathbf{S}_3$ and $\mathbf{R}_2 = \mathbf{S}_1 + \mathbf{S}_2$. Then

$$\mathcal{V}_r = \{r_1\mathbf{R}_1 + r_2\mathbf{R}_2 \geqslant \mathbf{0}\},$$

$$\mathcal{A}_r = \{(r_1,r_2)', \ r_1 + r_2 \geqslant 0, \ r_1 \geqslant 0, \ r_1(r_1 + r_2) \geqslant r_2^2\}$$

and

$$\dim S^2 > r(\{\mathbf{R}_1, \mathbf{R}_2\}) = r(\mathcal{V}_r) = r(\mathcal{A}_r) = 2.$$

The set $\mathcal{V}_r$ can also be spanned by its members, $\mathbf{R}_1$ and $\mathbf{R}_1 + \mathbf{R}_2$, which are non-negative definite. Then $\mathcal{V}_r$ is equal to

$$\mathcal{V}_s = \{s_1\mathbf{R}_1 + s_2(\mathbf{R}_1 + \mathbf{R}_2) \geqslant \mathbf{0}\},$$

where $s_1 = r_1 - r_2$ and $s_2 = r_2$. The set $\mathcal{A}_r$ is also equal to

$$\mathcal{A}_s = \{(s_1,s_2)',\ s_1 + 2s_2 \geqslant 0,\ s_1 + s_2 \geqslant 0,\ (s_1 + s_2)(s_1 + 2s_2) \geqslant s_2^2\}.$$

It is so, since the transformation $(s_1,s_2) \to (r_1 - r_2, r_2)$ is one-to-one and preserves the corresponding inequality constraints. Note also that the inequalities of the set $\mathcal{A}_s$ are fulfilled by any non-negative scalars s_1 and s_2. This is not true for the scalars r_1 and r_2 and the inequalities of the set $\mathcal{V}_r$. In result

$$\mathcal{A}_0 = \{(s_1,s_2)', s_1 \geqslant 0, s_2 \geqslant 0\} \subset \mathcal{A}_s.$$

However, the reverse inclusion does not hold.

Finally, let $\mathcal{V}_* = \{\alpha_1 \mathbf{S}_1 + \alpha_2 \mathbf{S}_2 \geqslant \mathbf{0}\}$. The corresponding set of parameters has now the form $\mathcal{A}_* = \{(\alpha_1,0)', \alpha_1 \geqslant 0\}$. Thus $\mathcal{V}_* = \{\alpha_1 \mathbf{S}_1\}$ and

$$\dim S^2 > r(\{\mathbf{S}_1, \mathbf{S}_2\}) > r(\mathcal{V}_*) = r(\mathcal{A}_*) = 1.$$

The above examples show that the number of linearly independent matrices $\mathbf{S}_1, \ldots, \mathbf{S}_l$ used in description of the set of dispersion matrices in not in general equal to $r(\mathcal{V})$. It is so, if the matrices $\mathbf{S}_1, \ldots, \mathbf{S}_l$ are all non-negative definite. The direct relation between the ranks of $\mathcal{V}$ and of $\mathcal{A}$ is exhibited in the following

Theorem 1. *Let the matrices spanning the set $\mathcal{V}$ be linearly independent. If $\alpha_1, \ldots, \alpha_k \in \mathcal{A}$ are linearly independent, then*

$$\mathbf{V}(\alpha_1), \mathbf{V}(\alpha_2), \ldots, \mathbf{V}(\alpha_k) \in \mathcal{V}$$

are also linearly independent and vice versa.

From the assumption, $\mathbf{V}_1 = \mathbf{V}(\alpha_1), \ldots, \mathbf{V}_k = \mathbf{V}(\alpha_k) \in \mathcal{V}$ and $k \leqslant l$. Now, by contradiction,

$$\sum\nolimits_i^k s_i \mathbf{V}_i = \sum\nolimits_i^k \sum\nolimits_j^l s_i \alpha_{ij} \mathbf{S}_j = \sum\nolimits_j^l r_j \mathbf{S}_j = \mathbf{0}$$

implies that all $r_j = \sum_i^k s_i \alpha_{ij}$ are equal to zero, which contradicts the assumption that $\alpha_1, \ldots, \alpha_k$ are linearly independent. The converse implication follows by similar arguments.

The result of Theorem 1 implies that $r(\mathcal{V}) = r(\mathcal{A})$. Moreover, there always exists a set of non-negative definite matrices $\mathbf{V}_1, \ldots, \mathbf{V}_q$ such that

$$\begin{aligned} \mathcal{V} &= \{\mathbf{V} = \alpha_1 \mathbf{S}_1 + \alpha_2 \mathbf{S}_2 + \cdots + \alpha_l \mathbf{S}_l \geqslant \mathbf{0}\} \\ &= \{\mathbf{V} = \gamma_1 \mathbf{V}_1 + \gamma_2 \mathbf{V}_2 + \cdots + \gamma_q \mathbf{V}_q \geqslant \mathbf{0}\}, \end{aligned}$$

where $q = r(\mathcal{V})$.

3. Estimation in Fixed Models

The model $\{\mathbf{y}, \mathbf{X}\beta, \sigma^2\mathbf{V}\}$ is termed as fixed. Usually it is written in the form

$$\mathbf{y} = \mathbf{X}\beta + \varepsilon, \tag{2}$$

where ε is an unobservable random variable such that $E(\varepsilon) = \mathbf{0}$ and $D(\varepsilon) = \sigma^2\mathbf{V}$. The components of the vector β are called the fixed effects while that of ε are the random effects. The scalar σ^2 is called the error variance. If $r(\mathbf{V}) < n$, then (2) is referred as the singular model.

The singularity implies some restrictions on $\mathbf{y}$. Namely, if $\mathbf{V}$ is singular, then

$$\mathbf{y} - \mathbf{X}\beta \in \mathcal{R}(\mathbf{V}) \text{ with probability one.}$$

This inclusion, noticed by Rao [5] (see also Feuerverger and Fraser [1]), is known as a consistency condition of the model $\{\mathbf{y}, \mathbf{X}\beta, \sigma^2\mathbf{V}\}$. Obviously, the non-singular models are always consistent.

In what follows we will assume that the models under consideration fulfil this consistency condition, i.e. that $\mathbf{y} \in \mathcal{T}$, where $\mathcal{T}$ is a subspace spanned by the columns of $\mathbf{V}$ and of $\mathbf{X}$, i.e. $\mathcal{T} = \mathcal{R}(\mathbf{V} : \mathbf{X})$. The subspace $\mathcal{T}$ is called the support of the model.

It is known (see Rao [5]), that the uniformly minimum variance unbiased estimator, also called the Best Linear Unbiased Estimator (BLUE), of any estimable function of fixed parameters $\mathbf{p}'\beta$ can be expressed by any solution of the so called normal equations.

Theorem 2. *The BLUE of* $\mathbf{p}'\beta$ *estimable in the model* $\{\mathbf{y}, \mathbf{X}\beta, \sigma^2\mathbf{V}\}$ *has the form* $\mathbf{p}'\hat{\beta}$, *where* $\hat{\beta}$ *is any solution of the normal equations*

$$\mathbf{X}'\mathbf{T}^+\mathbf{X}\beta = \mathbf{X}'\mathbf{T}^+\mathbf{y}, \tag{3}$$

with $\mathbf{T} = \mathbf{V} + \delta\mathbf{X}\mathbf{X}'$, $\delta > 0$ *if* $\mathcal{R}(\mathbf{X}) \nsubseteq \mathcal{R}(\mathbf{V})$ *and* $\delta = 0$ *otherwise, whereas* $\mathbf{T}^+$ *is the Moore-Penrose inverse of* $\mathbf{T}$.

The BLUE of μ $(=\mathbf{X}\beta)$ is $\mathbf{P}^{\mathbf{T}}\mathbf{y}$, where

$$\mathbf{P}^{\mathbf{T}} = \mathbf{X}(\mathbf{X}'\mathbf{T}^+\mathbf{X})^+\mathbf{X}'\mathbf{T}^+.$$

This operator is idempotent and $\mathcal{R}(\mathbf{P}^{\mathbf{T}}) = \mathcal{R}(\mathbf{X})$. These properties indicate that $\mathbf{P}^{\mathbf{T}}$ is a projector on $\mathcal{E}$. Since $\mathbf{T}$ is a symmetric and non-negative definite matrix, $\mathbf{P}^{\mathbf{T}}$ can be considered as the orthogonal projector under a semi-inner product defined with the use of $\mathbf{T}^+$. The BLUEs of μ can also be characterized as in the following result due to Kala and Pordzik [3].

Theorem 3. *A statistic* $\mathbf{R}\mathbf{y}$ *is the BLUE of* μ *in the model* $\{\mathbf{y}, \mathbf{X}\beta, \sigma^2\mathbf{V}\}$

if and only if $\mathbf{R} = \mathbf{R}^2$, $\mathcal{R}(\mathbf{R}) = \mathcal{R}(\mathbf{X})$, *and* $\mathbf{RV} = \mathbf{VR}'$.

If $\mathbf{V} = \mathbf{I}$, then the model $\{\mathbf{y}, \mathbf{X}\beta, \sigma^2\mathbf{I}\}$ is termed as simple and the BLUE of μ is delivered by $\mathbf{Py}$, where $\mathbf{P} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{+}\mathbf{X}' = \mathbf{X}\mathbf{X}^{+}$. It is called the Simple Least Squares Estimator (SLSE), since it minimizes the Euclidean norm $||\mathbf{y} - \mathbf{X}\beta||$ with respect to all vectors β. The question when the SLSE is the BLUE of μ also in the model $\{\mathbf{y}, \mathbf{X}\beta, \sigma^2\mathbf{V}\}$ was considered by many authors. The first who raised it, in the frames of non-singular models, was probably Zyskind [13] (see also Zyskind [14], Watson [12] and Kruskal [4]). The answer to this question is very simple and elegant.

Theorem 4. *The SLSE coincides with the BLUE of* μ *in the model* $\{\mathbf{y}, \mathbf{X}\beta, \sigma^2\mathbf{V}\}$ *if and only if* $\mathbf{V}\mathcal{E} \subset \mathcal{E}$, *i.e. when* $\mathcal{E}$ *is an invariant subspace of the dispersion matrix* $\mathbf{V}$.

As to estimation of the variance error, first note that σ^2, being a measure of diversity of observations, is independent on the translations of the observed random variable along the expectation subspace $\mathcal{E}$. The same feature is required of its estimate $g(\mathbf{y})$. The translation invariance of $g(\mathbf{y})$ means that

$$g(\mathbf{y}) = g(\mathbf{y} + \mathbf{d}) \quad \text{for any } \mathbf{d} \in \mathcal{E}. \tag{4}$$

But $\mathbf{y} = \mathbf{Qy} + \mathbf{d}$, where $\mathbf{Q} = \mathbf{I} - \mathbf{P}$ and $\mathbf{d} = \mathbf{Py} \in \mathcal{E}$. Therefore (4) implies that $g(\mathbf{y})$ depends on $\mathbf{y}$ only through $\mathbf{Qy}$, and if it is so, the statistic $g(\mathbf{y}) = g(\mathbf{Qy})$ is invariant with respect to translations along $\mathcal{E}$.

The function $g(\mathbf{y})$ is usually a quadratic form, which is the simplest statistic with the expectation dependent on the dispersion structure of the observed vector random variable. If $g(\mathbf{y}) = \mathbf{y}'\mathbf{G}\mathbf{y}$, then the equality $g(\mathbf{y}) = g(\mathbf{Qy})$ for all $\mathbf{y}$, implies that

$$\mathbf{G} = \mathbf{QGQ}. \tag{5}$$

In view of (5), we have also

$$E(g(\mathbf{y})) = \mathrm{tr}(E(\mathbf{y}\mathbf{y}')\mathbf{G}) = \sigma^2 \mathrm{tr}(\mathbf{VQGQ}).$$

Thus each Quadratic Invariant Unbiased Estimator (QIUE) of the error variance σ^2 has the form

$$\mathrm{QIUE}(\sigma^2) = \frac{1}{\mathrm{tr}(\mathbf{VQGQ})}\mathbf{y}'\mathbf{QGQy}.$$

The standard choice of $\mathbf{G}$ is $\mathbf{T}^{+}\mathbf{Q}^{\mathbf{T}}$, where $\mathbf{Q}^{\mathbf{T}} = \mathbf{I} - \mathbf{P}^{\mathbf{T}}$. Then

$$\mathrm{QIUE}(\sigma^2) = \frac{1}{r(\mathbf{T}) - r(\mathbf{X})}\mathbf{y}'\mathbf{T}^{+}\mathbf{Q}^{\mathbf{T}}\mathbf{y}, \tag{6}$$

which formula can be traced back to Rao [9]. Indeed, $\mathbf{Q}^{\mathbf{T}}$ is idempotent, $\mathbf{T}^{+}\mathbf{Q}^{\mathbf{T}}$

is symmetric, and $\mathbf{Q^T} = \mathbf{Q^T Q}$. In result

$$\mathbf{T}^+\mathbf{Q^T} = (\mathbf{Q^T})'\mathbf{T}^+\mathbf{Q^T} = \mathbf{Q}\mathbf{T}^+\mathbf{Q^T}\mathbf{Q}$$

and tr $\{\mathbf{T}^+\mathbf{Q^T}\mathbf{V}\} = r(\mathbf{T}) - r(\mathbf{X})$, since $\mathbf{Q^T V} = \mathbf{Q^T T}$.

Finally note that $\mathbf{Q^T y}$ as well as $\mathbf{Qy}$ are vectors of residuals in the model $\{\mathbf{y}, \mathbf{X}\beta, \sigma^2\mathbf{V}\}$ and the simple model $\{\mathbf{y}, \mathbf{X}\beta, \sigma^2\mathbf{I}\}$, respectively.

4. Estimation in Mixed Models

From the previous section it follows that the simple least squares procedure provides the BLUE of μ in all fixed models $\{\mathbf{y}, \mathbf{X}\beta, \sigma^2\mathbf{V}\}$ for which $\mathcal{E}$ is an invariant subspace of the dispersion matrix $\mathbf{V}$. This condition is linear, i.e. if $\mathcal{E}$ is an invariant subspace of $\mathbf{V}_1$ and of $\mathbf{V}_2$, then it is also invariant subspace of $\mathbf{V} = \alpha_1\mathbf{V}_1 + \alpha_2\mathbf{V}_2$ for any α_1,α_2. Thus, if $\mathbf{V}\mathcal{E} \subset \mathcal{E}$, then $\mathbf{Py}$ is the BLUE of μ in the model $\{\mathbf{y}, \mathbf{X}\beta, \sigma^2\mathbf{V}\}$ as well as in the variance components model

$$\{\mathbf{y}, \mathbf{X}\beta, \sigma_1^2\mathbf{V} + \sigma^2\mathbf{I}\}.$$

In general, however, $\mathbf{P^T y}$ differs from $\mathbf{Py}$. None is the estimator of minimum dispersion uniformly in both variance components σ_1^2 and σ^2. Such an estimator exists if and only if $\mathbf{Py}$ equals $\mathbf{P^T y}$, i.e. when $\mathbf{PV} = \mathbf{VP}$. Of course, in such case the BLUE of μ is delivered by the simple least squares procedure.

This remark applies also to any mixed variance components model,

$$\{\mathbf{y}, \mathbf{X}\beta, \mathbf{V}(\sigma) = \sigma_1^2\mathbf{V}_1 + \sigma_2^2\mathbf{V}_2 + \cdots + \sigma_l^2\mathbf{V}_l\}. \tag{7}$$

In such case a linear statistic $\mathbf{Ry}$ is the unbiased estimator with minimum dispersion matrix uniformly with respect to all variance components σ_1^2, ..., σ_l^2, if it is the BLUE of μ in each submodel $\{\mathbf{y}, \mathbf{X}\beta, \sigma_i^2\mathbf{V}_i\}$, as well as in the model

$$\{\mathbf{y}, \mathbf{X}\beta, \sigma^2\mathbf{V}_w\}, \tag{8}$$

where $\mathbf{V}_w$ is a maximal element in $\mathcal{V}$, i.e. such non-negative definite matrix that $\mathcal{R}(\mathbf{V}) \subseteq \mathcal{R}(\mathbf{V}_w)$ for all $\mathbf{V} \in \mathcal{V}$. One choice of $\mathbf{V}_w$ is

$$\mathbf{V}_w = w_1\mathbf{V}_1 + w_2\mathbf{V}_2 + \cdots + w_l\mathbf{V}_l = \mathbf{V}(\mathbf{w}), \tag{9}$$

where $\mathbf{w} = (w_1, \ldots, w_l)'$ with $w_i > 0$. The scalars w_i, $i = 1, 2, \ldots, l$, can be considered as weights, since we can assume that they sum to one.

The model (8) is fixed. The BLUE of μ in this model is delivered by $\mathbf{P}^{\mathbf{T}_w}\mathbf{y}$, where $\mathbf{T}_w = \mathbf{V}_w + \delta\mathbf{XX}'$, with $\delta > 0$ if $\mathcal{R}(\mathbf{X}) \nsubseteq \mathcal{R}(\mathbf{V}_w)$ and $\delta = 0$ otherwise. Since $\mathbf{P}^{\mathbf{T}_w}$ is a projector and $\mathcal{R}(\mathbf{P}^{\mathbf{T}_w}) = \mathcal{R}(\mathbf{X})$, then directly from Theorem 3 we have the following

Theorem 5. *The BLUE of μ in the model $\{\mathbf{y}, \mathbf{X}\beta, \mathbf{V}(\sigma)\}$ exists if and only if*

$$\mathbf{P}^{\mathbf{T}_w}\mathbf{V}_i = \mathbf{V}_i(\mathbf{P}^{\mathbf{T}_w})' \quad \text{for} \quad i = 1, 2, \ldots, l. \tag{10}$$

If it is the case, $\mathbf{P}^{\mathbf{T}_w}\mathbf{y}$ is the BLUE of μ.

The condition (10) is burdensome, which may result in non-existence of the BLUE. If the BLUE does not exist, the residual vectors in the sub-models $\{\mathbf{y}, \mathbf{X}\beta, \sigma_i^2\mathbf{V}_i\}$ are different, which may lead to some inconsistency among the individual estimates of $\sigma_1^2, \ldots, \sigma_l^2$ following from the sub-models treated separately.

Let $g(\mathbf{y})$ be a translation invariant quadratic form, $g(\mathbf{y}) = \mathbf{y}'\mathbf{G}\mathbf{y}$, and consider its unbiasedness for $\mathbf{p}'\sigma$ in frames of the model (7). Because $\mathbf{G} = \mathbf{QGQ}$, we have

$$E\{g(\mathbf{y})\} = \sum\nolimits_i^l \sigma_i^2 \mathrm{tr}(\mathbf{V}_i\mathbf{QGQ}),$$

But,

$$\mathrm{tr}(\mathbf{A}'\mathbf{BCD}') = (\mathrm{vec}\mathbf{A})'(\mathbf{D} \otimes \mathbf{B})\mathrm{vec}\mathbf{C}, \tag{11}$$

where $\otimes$ denotes the Kronecker product, while vec operation forms a vector from the matrix by writing its columns one below the other (for details see e.g. Harville [2]). This implies that

$$E\{g(\mathbf{y})\} = \sigma'\mathbf{H}'(\mathbf{Q} \otimes \mathbf{Q})\mathbf{g}, \tag{12}$$

where $\mathbf{g} = \mathrm{vec}\ (\mathbf{G})$, while $\mathbf{H}$ is an $n^2 \times l$ matrix of the form

$$\mathbf{H} = (\mathrm{vec}\mathbf{V}_1, \mathrm{vec}\mathbf{V}_2, \ldots, \mathrm{vec}\mathbf{V}_l). \tag{13}$$

Now, comparing (12) with $\mathbf{p}'\sigma$ for all $\sigma \in R^l_{\geq}$, leads to the condition

$$\mathbf{H}'(\mathbf{Q} \otimes \mathbf{Q})\mathbf{g} = \mathbf{p}. \tag{14}$$

If it holds, then the statistic $g(\mathbf{y}) = \mathbf{y}'\mathbf{G}\mathbf{y}$ is translation invariant and unbiased for $\mathbf{p}'\sigma$. In such case we say that $\mathbf{p}'\sigma$ is translation invariant estimable or, shortly, I-estimable.

The condition (14) implies that only those functions $\mathbf{p}'\sigma$ are I-estimable which have the vectors of coefficients being linear combinations of columns of $\mathbf{H}'(\mathbf{Q} \otimes \mathbf{Q})$. So, there is only $r(\mathbf{H}'(\mathbf{Q} \otimes \mathbf{Q})) \leqslant l$ linearly independent and I-estimable functions of variance components. The complete set of them is spanned by the $l \times l$ matrix $\mathbf{H}'(\mathbf{Q} \otimes \mathbf{Q})\mathbf{H} = (h_{ij})$, where

$$h_{ij} = (\mathrm{vec}(\mathbf{Q}\mathbf{V}_i\mathbf{Q}))'\mathrm{vec}(\mathbf{Q}\mathbf{V}_j\mathbf{Q}) = \mathrm{tr}(\mathbf{Q}\mathbf{V}_i\mathbf{Q}\mathbf{V}_j).$$

Theorem 6. *The QIUE of $\mathbf{p}'\sigma$ I-estimable in the model $\{\mathbf{y}, \mathbf{X}\beta, \mathbf{V}(\sigma)\}$ has the form $\mathbf{p}'\hat{\sigma}$, where $\hat{\sigma}$ is any solution of the variance components normal* *equation*

$$\mathbf{H}'(\mathbf{Q}\otimes\mathbf{Q})\mathbf{H}\sigma = \mathbf{H}'(\mathbf{Q}\otimes\mathbf{Q})vec(\mathbf{y}\mathbf{y}'), \quad (15)$$

with $\mathbf{H}$ *as given in (13) and* $\mathbf{Q} = \mathbf{I} - \mathbf{X}\mathbf{X}^{+}$.

Note, that the left hand side of (15) is the expectation of the right hand side. It follows by applying (11) to $g(\mathbf{Q}\mathbf{y}) = \text{tr}\,(\mathbf{G}'\mathbf{Q}\mathbf{y}\mathbf{y}'\mathbf{Q})$ and then comparing its expectation with (12) for all vectors $\mathbf{g}$. Thus

$$E\{(\mathbf{Q}\otimes\mathbf{Q})\text{vec}(\mathbf{y}\mathbf{y}')\} = \ (\mathbf{Q}\otimes\mathbf{Q})\mathbf{H}\sigma.$$

The same property is also true for the normal equation (3).

Because $\mathbf{Q}\otimes\mathbf{Q}$ is an idempotent matrix, the equation (15) can be extended to the weighted form

$$\mathbf{H}'(\mathbf{Q}\otimes\mathbf{Q})\mathbf{\Lambda}(\mathbf{Q}\otimes\mathbf{Q})\mathbf{H}\sigma = \mathbf{H}'(\mathbf{Q}\otimes\mathbf{Q})\mathbf{\Lambda}(\mathbf{Q}\otimes\mathbf{Q})\text{vec}(\mathbf{y}\mathbf{y}'), \quad (16)$$

where $\mathbf{\Lambda}$ is a non-negative definite matrix of weights. Of course, this equation preserves the properties of the equation (15) if $r(\mathbf{H}'(\mathbf{Q}\otimes\mathbf{Q})\mathbf{\Lambda}) = r(\mathbf{H}'(\mathbf{Q}\otimes\mathbf{Q}))$.

5. MINQE Approach

The unbiasedness is an important property, but usually there are many unbiased estimators for the same parametric function. In the case of fixed parameters, the unique result of estimation is ensured by the second criterion of uniformly, in parameters, minimal variance. For functions of variance components the similar approach requires the additional assumptions about third and fourth moments of the observed random variable. A statistic being a quadratic function of observations that is unbiased for a given linear function of variance components and has minimum variance among all its unbiased estimators is termed as Best Quadratic Unbiased Estimator (BQUE). Such estimation procedure was developed by Rao [6]. Rao [7, 8] has also proposed some alternative principle which does not depend on the additional assumptions. It leads to the MInimum-Norm Quadratic Estimators (MINQE), which, in a given class, are the closest, in a sense of the distance, to the so called natural estimator. We will consider only the class of translation invariant unbiased estimators.

The resulting estimation procedure as formulated in Rao and Kleffe [10, Theorem 5.2.1] and expressed in symbols of this paper takes the following form.

Theorem 7. *The MINQE of* $\mathbf{p}'\sigma$ *I-estimable in the model* $\{\mathbf{y}, \mathbf{X}\beta, \mathbf{V}(\sigma)\}$ *has the form* $g(\mathbf{y}) = \mathbf{y}'\mathbf{G}_*\mathbf{y}$, *where*

$$\mathbf{G}_* = \sum\nolimits_i^l \lambda_i \mathbf{T}_w^{+}\mathbf{Q}^{\mathbf{T}_w}\mathbf{V}_i\mathbf{T}_w^{+}\mathbf{Q}^{\mathbf{T}_w}, \quad (17)$$

with $\mathbf{T}_w = \mathbf{V}_w + \mathbf{X}\mathbf{X}'$, $\mathbf{Q}^{\mathbf{T}_w} = \mathbf{I} - \mathbf{P}^{\mathbf{T}_w}$, *and* $\mathbf{V}_w$ *given in (9), while* $\lambda = (\lambda_1, \ldots, \lambda_l)'$ *is any solution of the equation*

$$(tr(\mathbf{T}_w^+\mathbf{Q}^{\mathbf{T}_w}\mathbf{V}_i\mathbf{T}_w^+\mathbf{Q}^{\mathbf{T}_w}\mathbf{V}_j))\lambda = \mathbf{p}. \tag{18}$$

This result, however, can be rewritten as in the following

Theorem 8. *The MINQE of* $\mathbf{p}'\sigma$ *I-estimable in the model* $\{\mathbf{y}, \mathbf{X}\beta, \mathbf{V}(\sigma)\}$ *has the form* $\mathbf{p}'\hat{\sigma}$, *where* $\hat{\sigma}$ *is any solution of the variance components normal equations*

$$\mathbf{H}'\mathbf{W}\mathbf{H}\sigma = \mathbf{H}'\mathbf{W}\,vec(\mathbf{y}\mathbf{y}'), \tag{19}$$

with $\mathbf{W} = \mathbf{T}_w^+\mathbf{Q}^{\mathbf{T}_w} \otimes \mathbf{T}_w^+\mathbf{Q}^{\mathbf{T}_w}$ *and* $\mathbf{H}$ *given in (13).*

Indeed, using (11) in (18) and (17), we have

$$g(\mathbf{y}) = \sum\nolimits_i^l \lambda_i \mathrm{tr}(\mathbf{V}_i\mathbf{T}_w^+\mathbf{Q}^{\mathbf{T}_w}\mathbf{y}\mathbf{y}'\mathbf{T}_\mathbf{w}^+\mathbf{Q}^{\mathbf{T}_\mathbf{w}}) = \lambda'\mathbf{H}'\mathbf{W}\mathrm{vec}(\mathbf{y}\mathbf{y}')$$

and

$$\mathbf{H}'\mathbf{W}\mathbf{H}\lambda = \mathbf{p}.$$

Solving the last equation with respect to λ, the MINQE of $\mathbf{p}'\sigma$, takes the form $g(\mathbf{y}) = \mathbf{p}'\hat{\sigma}$, where

$$\hat{\sigma} = (\mathbf{H}'\mathbf{W}\mathbf{H})^+\mathbf{H}'\mathbf{W}\mathrm{vec}(\mathbf{y}\mathbf{y}')$$

can be considered as a solution of (19).

Because $\mathbf{Q}^{\mathbf{T}_w}\mathbf{Q} = \mathbf{Q}^{\mathbf{T}_w}$ and the matrix $\mathbf{T}_w^+\mathbf{Q}^{\mathbf{T}_w}$ is symmetric and non-negative definite, the matrix $\mathbf{W}$ satisfies the equality

$$\mathbf{W} = (\mathbf{Q} \otimes \mathbf{Q})\mathbf{W}(\mathbf{Q} \otimes \mathbf{Q}).$$

This means that the MINQE estimator actually follows by solving the equation (16) with $\mathbf{\Lambda} = \mathbf{W}$.

This procedure, after appropriate reparametrization as in Theorem 1, can also be applied to variance-covariance components models $\{\mathbf{y}, \mathbf{X}\beta, \mathbf{V}(\gamma)\}$. The only difference is that the negative estimates of parameters $\gamma_1, \ldots, \gamma_q$ are acceptable as long as $\mathbf{V}(\hat{\gamma}) \geqslant \mathbf{0}$.

The MINQE depends on the matrices $\mathbf{X}$ and $\mathbf{V}_1, \mathbf{V}_2, \ldots, \mathbf{V}_l$, and on a choice of the vector $\mathbf{w}$ of weights in $\mathbf{V}_w = \mathbf{V}(\mathbf{w})$. It is assumed that this vector reflect the values of true variance components. Thus the MINQE depends on the prior values of unknown parameters. Actually, it suffices that $\mathbf{w}$ is proportional to σ, since (19) does not change if $\mathbf{W}$ is replaced by $s\mathbf{W}$, where s is any positive scalar. In this sense, the MINQE is optimal only locally, on a prior chosen direction in the cone $\mathcal{V}$.

From this observations it follows that the estimator (6), which is also of the

MINQE type, is unique. It is so, because the cone of dispersion matrices for each fixed model is of rank one, i.e. it is connected only with one direction in $\mathcal{V}$.

When there is no independent prior information on variance components proportions, the equal weights are recommended. It means, that the directions in cone $\mathcal{V}$ pointed by $\mathbf{V}_1, \ldots, \mathbf{V}_l$ are all equally important.

Freedom of choice of the vector of weights gives also a possibility of computing MINQEs iteratively using the previous estimates of variance components as a prior values in the next step, which means changing successively the direction in $\mathcal{V}$. Although in such procedure each next step seems to be less depends on the initial weights, the main statistical properties are lost. However, if such estimates are non-negative and converge, then the whole estimation process, including also fixed parameters, is consistent in a sense that $\mathbf{P}^{\mathbf{T}_w}\mathbf{y} = (\mathbf{I} - \mathbf{Q}^{\mathbf{T}_w})\mathbf{y}$, obtained in the last step, can be considered as iterated BLUE of μ. Note also that the initial weights can be obtained estimating variance components from the sub-models $\{\mathbf{y}, \mathbf{X}\beta, \sigma_i^2\mathbf{V}_i\}$, $i = 1, 2, \ldots, l$, what can be done even if the BLUE of μ does not exist.

References

[1] A. Feuerverger, D.A.S. Fraser, Categorical information and the singular linear model, *Canad.J. Statist.*, **8** (1980), 41-45.

[2] A.D. Harville, *Matrix Algebra From a Statistician's Perspective* , Springer, New York (1997).

[3] R. Kala, P. Pordzik, Two local operators and the BLUE, *Linear Algebra Appl.*, **417** (2005), 134-139.

[4] W. Kruskal, When are the Gauss-Markov and least squares estimators the same? A coordinate free approach, *Ann. Math. Statist.*, **39** (1968), 70-75.

[5] C.R. Rao, Unified theory of linear estimation, *Sankhya Ser. A*, **33** (1971), 371-394.

[6] C.R. Rao, Estimation of variance and covariance components. MINQUE theory, *J. Multivariate Anal.*, **1** (1971), 256-275.

[7] C.R. Rao, Minimum variance quadratic unbiased estimation of variance components, *J. Multivariate Anal.*, **1** (1971), 445-456.

[8] C.R. Rao, Estimation of variance and covariance components in linear models, *J. Amer. Statist. Ass.*, **67** (1972), 112-115.

[9] C.R. Rao, Projectors, generalized inverses and the BLUE's, *J. Roy.Statist. Soc. Ser. B*, **36** (1974), 442-448.

[10] C.R. Rao, J. Kleffe, *Estimation of Variance Components and Applications*, North-Holland, Amsterdam (1988).

[11] D. Sengupta, S.R. Jammalamadaka, *Linear Models. An Integrated Approach*, World Scientific, New Jersey (2003).

[12] G.S. Watson, Linear least squares regression, *Ann. Math. Statist.*, **38** (1967), 1679-1699.

[13] G. Zyskind, On conditions for equality of best and simple linear least squares estmator (Abstract), *Ann. Math. Statist.*, **33** (1962), 1502-1503.

[14] G. Zyskind, On canonical forms, non-negative covariance matrices and best and simple linear estimators in linear models, *Ann. Math. Statist.*, **38** (1967), 1092-1109.