DISSERTATIONES MATHEMATICAE UNIVERSITATIS TARTUENSIS
36

LINEAR MIXED MODELS WITH EQUIVALENT PREDICTORS

MÄRT MÖLS

TARTU 2004

DISSERTATIONES MATHEMATICAE UNIVERSITATIS TARTUENSIS
36

DISSERTATIONES MATHEMATICAE UNIVERSITATIS TARTUENSIS
36

LINEAR MIXED MODELS WITH EQUIVALENT PREDICTORS

MÄRT MÖLS

Faculty of Mathematics and Computer Science, University of Tartu, Tartu, Estonia

Dissertation is accepted for the commencement of the degree of Doctor of Philosophy (PhD) in mathematical statistics on June 18, 2004 by the Council of the Faculty of Mathematics and Computer Science, University of Tartu.

Supervisor:
Cand. Sc., Docent Tõnu Möls
University of Tartu
Tartu, Estonia

Opponents:
PhD, professor Kenneth Nordström
University of Oulu
Oulu, Finland

PhD, docent Simo Puntanen
University of Tampere
Tampere, Finland

The public defence will take place on September 10, 2004.

Publication of this dissertation is granted by the Institute of Mathematical Statistics of the University of Tartu.

Tartu Ülikooli Kirjastus
www.tyk.ut.ee
Tellimus nr. 378.

Ilmarile ja Kalevile

CONTENTS

List of Original Publications:

1. M. Möls (2004) On Reparameterization of Random Effects in Linear Mixed Models. *Acta et Commentationes Universitatis Tartuensis de Mathematica*, 7 pages. (accepted)

2. M. Möls (2003) Constraints on Random Effects and Mixed Linear Model Predictions. *Acta Applicandae Mathematicae*, **79**, 17–23.

3. Möls M., Nõges T., Nõges P. (2001). Võrtsjärve ökoloogia modelleerimisest. *Eesti Statistikaseltsi Teabevihik*, **11**, 44–53.

4. Frisk, T., Bilaletdin, Ä., Kaipanen, H., Malve, O., Möls, M. (1999). Modelling phytoplankton dynamics of the eutrophic Lake Võrtsjärv, Estonia. *Hydrobiologia*, **414**, 59–69.

5. Nõges, T., Bilaletdin, Ä., Frisk, T., Huttula, T., Järvet, A., Kaipainen, H., Kivimaa, R., Malve, O., Möls, M. & P. Nõges, (1999). Results of ecohydrodynamical investigations on large shallow eutrophic Lake Võrtsjärv, Estonia in 1995–1996. *EMI Report Series*, **10**, 97–100.

6. Frisk, T., Bilaletdin, Ä., Kaipanen, H., Malve, O., Möls, M. (1998). Water quality modelling. Present state and future fate of Lake Võrtsjärv. *The Finnish Environment*, **209**, Tampere, 112–131.

7. Lapp K., Möls M. (1997). Kliinilised katsed ja kaasaegne meditsiinistatistika. *Eesti Statistikaseltsi Teabevihik*, **9**, 34–41.

8. Tiit E.-M., Möls M. (1997). Rakendusstatistika algkursus, Tartu. 144 pages.

Introduction

Linear mixed models are used in many different research areas like biology, sociology, medicine, ecology etc. Linear mixed models together with generalized linear mixed models are one of the main techniques used in longitudinal and spatial data analysis, multilevel modelling, small area estimation etc. In applying linear mixed models one often encounters difficulties in choosing correct covariance structure. Most theoretical works barely touch the question. Even those rare coverages have sparked some controversies, see, for example, the article by Voss (1999). In more practice oriented work authors usually limit themselves to suggesting some plausible covariance structure. Proving the correctness of the suggested (or used) covariance structure is often quite limited at best.

Some work has been done to investigate the effects of misspecification of the covariance structure in linear mixed models (see for example Puntanen & Styan 1989, Harville 1997). However, several aspects, like the effect of covariance structure misspecification on mixed model predictions, have so far remain largely uncovered. Further theoretical results providing techniques to interpret and justify the assumptions made about the covariance structure are therefore needed. The work presented here, in these Thesis, intends to make step further on the road to provide these techniques.

In Chapter 3 the results are presented showing that a popular estimation technique — Restricted Maximum Likelihood (REML) — can be viewed as estimation with respect to a misspecified covariance structure.

In Chapter 4 it is shown that for solving typical prediction-related problems unique determination of covariance matrix is not necessary. As an alternative to fixing the sample covariance matrix (or covariance structure) a new approach, reparameterisation constraints for random effects, is suggested.

Replacing one covariance structure with another may or may not lead to different prediction results. A relatively easy to check condition for different covariance structures to yield equivalent prediction results for a wide class of prediction problems is presented in Chapter 5.

To derive the results presented in Chapters 3–5 several results from matrix algebra and linear mixed models theory are needed. These supplementary results are presented together with introduction of notation in Chapters 1–2. The main results presented in Chapters 4 and 5 are previously published by author (Möls 2003, Möls 2004), but the coverage here adds several details. The investigated ideas are used by author in Möls, Nõges & Nõges (2001) and in Frisk, Bilaletdin, Kaipainen, Malve & Möls (1999). Results presented in Chapter 3 are new and not published before.

1. Matrix Algebra

In this chapter we give basic definitions and results from matrix algebra, which are needed in the following chapters. Most proofs are omitted, they can be found in graduate textbooks of Matrix Algebra. An interested reader is referred to Harville (1997) and Rao (1998). A few proofs are included because of their exceptional beauty or rarity.

1.1 Basic Terminology

The *transposed* matrix of the $m \times n$ matrix A is denoted by A^T.

A matrix is said to be *square* if it has as many rows as columns. An $1 \times n$ matrix is called *row vector* and $n \times 1$ matrix is called *column vector*. If the dimension of a vector is clear from context, the attribute "row" or "column" can be omitted.

A square matrix A is *idempotent* if $AA = A$.

A real square matrix is said to be *orthogonal* if $A^T A = I$.

The *determinant* of a square matrix A is denoted by $|A|$.

Sum of all diagonal elements of a square matrix A is called *trace* and denoted by $\mathrm{tr}(A)$.

A symmetric $m \times m$ matrix A is called *positive definite* if

$$x^T A x > 0$$

for any vector $x \neq 0$.

A symmetric $m \times m$ matrix A is called *non-negative definite* if

$$x^T A x \geq 0$$

for any vector x.

The *square root* of a non-negative definite matrix A, denoted by $A^{1/2}$, is a symmetric matrix satisfying $A = A^{1/2} A^{1/2}$. The square root of A^{-1} is denoted by $A^{-1/2}$.

1.2 Linear Spaces

Definition 1.1. *The column space of an $m \times n$ matrix A is the set of all m-dimensional column vectors that can be expressed as linear combinations of the columns of A. The symbol $\mathcal{C}(A)$ will be used to denote the column space of matrix A.*

Proposition 1.1. *For any $m \times n$ matrix A and $m \times p$ matrix B, $\mathcal{C}(A) \subset \mathcal{C}(B)$ if and only if there exists an $n \times p$ matrix M such that $B = AM$.*

Definition 1.2. *The rank of matrix A is defined as the dimension of the column space of matrix A and it is denoted by rank(A).*

Proposition 1.2. *For an arbitrary $m \times n$ matrix A following statements hold:*

1. *rank(A) = rank(A^T).*
2. *rank(A) = rank$(A^T A)$.*
3. *Let B be an $m \times m$ nonsingular matrix and C be an $n \times n$ nonsingular matrix. Then*

$$\text{rank}(BA) = \text{rank}(A) \text{ and } \text{rank}(AC) = \text{rank}(A). \tag{1.1}$$

Definition 1.3. *Let $\mathcal{U}$ be a subspace of the Euclidean space $\mathbb{R}^m$ of all m-component real column vectors. The orthogonal complement of $\mathcal{U}$, denoted by $\mathcal{U}^\perp$, is the collection of all vectors in $\mathbb{R}^m$ that are orthogonal to every vector in $\mathcal{U}$; that is, $\mathcal{U}^\perp = \{x : x \in \mathbb{R}^m \text{ and } x^T y = 0 \text{ for all } y \in \mathcal{U}\}$.*

Proposition 1.3. *If $\mathcal{U}$ is a subspace of $\mathbb{R}^m$, then its orthogonal complement $\mathcal{U}^\perp$ is also a subspace of $\mathbb{R}^m$.*

Definition 1.4. *Let $\mathcal{U}$ and $\mathcal{W}$ be subspaces of the linear space $\mathbb{R}^m$. If $x^T y = 0$ for every $x \in \mathcal{U}$ and $y \in \mathcal{W}$, then $\mathcal{U}$ and $\mathcal{W}$ are said to be orthogonal.*

Definition 1.5. *Let $\mathcal{U}$ and $\mathcal{W}$ be subspaces of the linear space $\mathbb{R}^m$, and let A is a symmetric positive definite matrix. If $x^T A y = 0$ for every $x \in \mathcal{U}$ and $y \in \mathcal{W}$, then $\mathcal{U}$ and $\mathcal{W}$ are said to be orthogonal with respect to A.*

Definition 1.6. *Let $\mathcal{U}$ and $\mathcal{W}$ be subspaces of the linear space $\mathbb{R}^m$. If $\mathcal{U} \cap \mathcal{W} = \{0\}$ then $\mathcal{U}$ and $\mathcal{W}$ are said to be essentially disjoint.*

Proposition 1.4. *Let U represent an $m \times n$ matrix and V an $m \times p$ matrix. The column spaces $\mathcal{C}(U)$ and $\mathcal{C}(V)$ are essentially disjoint if and only if*

$$rank\begin{pmatrix} U \\ V \end{pmatrix} = rank(U) + rank(V). \tag{1.2}$$

Proof. See Harville (1997), Theorem 17.2.4. □

Proposition 1.5. *Let $\mathcal{U}$ and $\mathcal{W}$ represent essentially disjoint subspaces of $\mathbb{R}^m$. Then there exists a (not necessarily uniquely determined) matrix A such, that $\mathcal{U}$ and $\mathcal{W}$ are orthogonal with respect to A.*

Proof. See Harville (1997), Theorem 17.7.1. □

Proposition 1.6. *Let U represent an $m \times n$ matrix and V an $m \times p$ matrix. If matrices U and V satisfy the condition (1.2), then there exists a matrix A such that*

$$UAV^T = 0.$$

Proof. Follows directly from Proposition 1.4 and Proposition 1.5. □

1.3 Generalized Inverse

Definition 1.7. *A generalized inverse of $n \times m$ matrix A is any $m \times n$ matrix A^- satisfying*

$$A = AA^-A. \tag{1.3}$$

Unfortunately, a generalized inverse matrix may not be uniquely determined.

Proposition 1.7. *Let A be an $m \times n$ matrix. Then the following statements hold.*

1. *There exist a generalized inverse A^-.*
2. *$(A^-)^T$ is a generalized inverse of A^T.*
3. *Let B represent an $m \times k$ matrix and C an $n \times q$ matrix. If $\mathcal{C}(B^T) \subset \mathcal{C}(A^T)$ and $\mathcal{C}(C) \subset \mathcal{C}(A)$, then B^TA^-C does not depend on the choice of A^-.*

4. *Let V represent an $n \times n$ positive definite matrix. Then $A(A^TVA)^-A^T$ is invariant to the choice of the generalized inverse of $(A^TVA)^-$.*

5. *$rank(A^-) \geq rank(A^-A) = rank(AA^-) = rank(A)$*

6. *$A(A^TA)^-$ is a generalized inverse of A^T*

Proposition 1.8. *Let A be any $m \times n$ matrix, G any $n \times m$ matrix. Then GB is a solution of the linear system $AX = B$ for every $m \times p$ matrix B for which the linear system is consistent if and only if $G = A^-$.*

Definition 1.8. *The Moore-Penrose inverse of the $n \times m$ matrix A is an $m \times n$ matrix A^+ satisfying*

$$A = AA^+A \tag{1.4}$$
$$A^+ = A^+AA^+ \tag{1.5}$$
$$AA^+ = (AA^+)^T \tag{1.6}$$
$$A^+A = (A^+A)^T \tag{1.7}$$

Proposition 1.9. *Corresponding to each $n \times m$ matrix A there exists one and only one $m \times n$ matrix A^+ satisfying conditions (1.4)–(1.7).*

Proposition 1.10. *Let A be a symmetric $m \times m$ matrix. Then A^+ is also symmetric.*

Proposition 1.11. *Let A be an $n \times p$ matrix and V an $n \times n$ symmetric positive definite matrix. Then $A(A^TVA)^-A^T$ is uniquely defined, symmetric and non-negative definite.*

Proof. Matrix A^TVA is symmetric and non-negative definite. Hence $(A^TVA)^+$ is also symmetric (Proposition 1.10). The equality

$$A(A^TVA)^-A^T = A(A^TVA)^+A^T$$

follows directly from Proposition 1.7, statement 4. It remains to prove that $A(A^TVA)^+A^T$ is non-negative definite. But

$$A(A^TVA)^+A^T = A(A^TVA)^+(A^TVA)(A^TVA)^+A^T.$$

Hence, for any vector v,

$$vA(A^TVA)^+A^Tv^T = v_*Vv_*^T, \tag{1.8}$$

where $v_* = vA(A^TVA)^+A^T$. But $v_*Vv_*^T \geq 0$ for any vector v_*, because V is positive definite. □

1.4 Projectors

Definition 1.9. *Let X be an $n \times m$ matrix. An $n \times n$ matrix P is a projector matrix onto column space of X if for arbitrary n-vector v*

$$Pv \in \mathcal{C}(X) \tag{1.9}$$

and for any vector $u \in \mathcal{C}(X)$

$$Pu = u. \tag{1.10}$$

Definition 1.10. *An $n \times n$ matrix P is called projector if there exists a matrix X so that P is a projector matrix onto $\mathcal{C}(X)$.*

These conditions can be presented also in a slightly different form. From Proposition 1.1 it follows that condition (1.9) holds if and only if there exists an $m \times n$ matrix M such that

$$P = XM. \tag{1.11}$$

The condition (1.10) is equivalent to the condition

$$PX = X. \tag{1.12}$$

Proposition 1.12. *A matrix P is a projector if and only if it is idempotent.*

Proof. *P is projector $\Rightarrow$ P is idempotent*

Follows immediately from (1.9) and (1.10).

P is idempotent $\Rightarrow$ P is projector onto some column space

If P is idempotent then it is a projector onto the column space of $\mathcal{C}(P)$: $P = PI$ and therefore condition (1.11) is satisfied, and because $PP = P$, the equality (1.12) also holds. □

Definition 1.11. *A symmetric projector matrix P is called orthogonal projector.*

Proposition 1.13. *The $n \times n$ matrix P_X is an orthogonal projector onto the subspace $\mathcal{C}(X)$ if and only if*

$$P_X = X(X^T X)^- X^T. \tag{1.13}$$

Proof. *P_X is orthogonal projector onto $\mathcal{C}(X) \Rightarrow P_X = X(X^TX)^-X^T$.*

If P_X is an orthogonal projector then it is idempotent, $P_XP_X = P_X$, and symmetric, $P_X^T = P_X$. Hence $P_X = P_X(P_X)^-P_X = P_X(P_XP_X)^-P_X = P_X(P_X^TP_X)^-P_X^T$. Because P_X is projector to the subspace of $\mathcal{C}(X)$ it has to have the form $P_X = XM$ for some matrix M (Proposition 1.1). Hence

$$P_X = XM(M^TX^TXM)^-M^TX^T. \tag{1.14}$$

This does not depend on the choice of generalized inverse (Proposition 1.7 statement 4). One choice for the generalized inverse is $(M^TX^TXM)^- = X(X^TX)^-X^T$:

$$(M^TX^TXM)X(X^TX)^-X^T(M^TX^TXM) = M^TX^TXM,$$

because $XMX = P_XX = X$. Using this generalized inverse in (1.14) leads to

$$P_X = X(X^TX)^-X^T,$$

which is uniquely determined because of Proposition 1.7. Therefore any orthogonal projector into the subspace of $\mathcal{C}(X)$ has to have the form (1.13).

$P_X = X(X^TX)^-X^TP_X \Rightarrow P_X$ is orthogonal projector onto $\mathcal{C}(X)$. As matrix $P_X = X(X^TX)^-X^T$ is idempotent and symmetric (see Proposition 1.11), it is an orthogonal projector. For any vector v with appropriate length $P_Xv \in \mathcal{C}(X)$ because it can presented as a linear combination of columns of X: $P_Xv = Xv_*$ for $v_* = (X^TX)^-X^Tv$. For any vector $u \in \mathcal{C}(X)$ (which can be expressed as $u = Xu_*$) we have $P_Xu = u$ because

$$P_Xu = P_XXu_* = X(X^TX)^-X^TXu_*.$$

From Proposition 1.7 it follows that $X(X^TX)^-X^TX = X$ and hence $Xu_* = u$, which proves the lemma. □

Proposition 1.14. *If P is a projector then also $I - P$ is a projector.*

Proposition 1.15. *Let X be an arbitrary $n \times p$ matrix, denote rank(X) by k, and let V be an $n \times n$ symmetric positive definite matrix. Then for the matrices*

$$P_{X,V} = X(X^TV^{-1}X)^-X^TV^{-1} \tag{1.15}$$

and

$$P_{X^\perp,V} = I - P_{X,V} \tag{1.16}$$

the following equalities hold:

$$P_{X,V} \cdot P_{X,V} = P_{X,V} \text{ and } P_{X^\perp,V} \cdot P_{X^\perp,V} = P_{X^\perp,V} \tag{1.17}$$

$$P_{X,V}X = X \text{ and } P_{X^\perp,V}X = 0 \tag{1.18}$$

$$rank(P_{X,V}) = k; \tag{1.19}$$

$$rank(P_{X^\perp,V}) = n - k; \tag{1.20}$$

$$V^{-1}P_{X,V} = P_{X,V}^T V^{-1} \tag{1.21}$$

$$V^{-1}P_{X^\perp,V} = P_{X^\perp,V}^T V^{-1} \tag{1.22}$$

$$P_{X,V}V = VP_{X,V}^T \tag{1.23}$$

$$P_{X^\perp,V}V = VP_{X^\perp,V}^T \tag{1.24}$$

$$X^T V^{-1} P_{X,V} = X^T V^{-1} \text{ and } X^T V^{-1} P_{X^\perp,V} = 0 \tag{1.25}$$

Theorem 1.1. *Let X be an $n \times p$ matrix, $rank(X) = k$, A be an $(n-k) \times n$ matrix, $rank(A) = n - k$, and let V be an $n \times n$ positive definite matrix. If $AX = 0$ then*

$$A^T(AVA^T)^{-1}A = V^{-1} - V^{-1}X(X^TV^{-1}X)^-X^TV^{-1}. \tag{1.26}$$

Proof given below is taken from Searle (1992), and has attributed by him to Pukelsheim.

Proof. Matrices $A^T(AA^T)^{-1}A$ and $X(X^TX)^-X^T$ are both symmetric and idempotent. Define a matrix

$$T = I - A^T(AA^T)^{-1}A - X(X^TX)^-X^T.$$

The matrix T is also symmetric and idempotent (because $AX = 0$). Therefore

$$\text{tr}(TT^T) = tr(T^2) = \text{tr}(I) - \text{tr}(A^T(AA^T)^{-1}A) - \text{tr}(X(X^TX)^-X^T).$$

Trace of an idempotent matrix is equal to its rank, and

$$\begin{aligned} \text{rank}(X(X^TX)^-X^T) &= \text{rank}(X) = k, \\ \text{rank}(A^T(AA^T)^{-1}A) &= \text{rank}(A) = n - k. \end{aligned}$$

Therefore

$$tr(TT^T) = n - (n - k) - k = 0.$$

Matrix T is real, therefore from $tr(TT^T) = 0$ follows $T = 0$. Hence $I - X(X^TX)^-X^T = A^T(AA^T)^{-1}A$. One may replace simultaneously A with $AV^{1/2}$ and X with $V^{-1/2}X$, because $AV^{1/2}V^{-1/2}X = AX = 0$. Making these replacements gives

$$I - V^{-1/2}X(X^TV^{-1}X)^-X^TV^{-1/2} = V^{1/2}A^T(AVA^T)^{-1}AV^{1/2},$$

what gives us, after multiplying both sides from left and right with $V^{-1/2}$, the desired equality

$$V^{-1} - V^{-1}X(X^TV^{-1}X)^-X^TV^{-1} = A^T(AVA^T)^{-1}A.$$

□

Definition 1.12. *One can remove k linearly dependent rows from matrix $P_{X^\perp} = I - X(X^TX)^-X^T$ to derive an $(n-k)\times n$ matrix with rank equal to $n-k$. This (not necessarily uniquely defined) matrix is denoted throughout the thesis by symbol K.*

Notice, that because $P_{X^\perp}X = 0$ also $KX = 0$.

There exist a useful relationship between matrices $P^\perp_{X,V}$ and K, which is stated in the following proposition.

Proposition 1.16. *Consider matrices $P^T_{X^\perp,V}$ — as defined in (1.16) — and K, defined by Definition 1.12. Then the following equality holds:*

$$K^T(KVK^T)^{-1}K = P^T_{X^\perp,V}V^{-1}P_{X^\perp,V} \tag{1.27}$$

Proof. To prove (1.27) use the equality

$$K^T(KVK^T)^{-1}K = V^{-1} - V^{-1}X(X^TV^{-1}X)^-X^TV^{-1}, \tag{1.28}$$

which follows from Theorem 1.1. Because of (1.28) and the properties (1.17) and (1.22), on can write

$$\begin{aligned} K^T(KVK^T)^{-1}K &= V^{-1} - V^{-1}X(X^TV^{-1}X)^-X^TV^{-1} \\ &= V^{-1}P_{X^\perp,V} \\ &= V^{-1}P_{X^\perp,V}P_{X^\perp,V} \\ &= P^T_{X^\perp,V}V^{-1}P_{X^\perp,V}, \end{aligned}$$

which proves the equality (1.27) and, hence, also the property 1.16. □

1.5 Eigenvalues and Eigenvectors

Definition 1.13. *Let A be a $n \times n$ matrix. The eigenvalues of A are defined as roots of the characteristic equation*

$$|A - \lambda I_n| = 0. \tag{1.29}$$

Equation (1.29) has n roots, in general complex.

Definition 1.14. *If λ is eigenvalue of A, then there exists vector v ($v \neq 0$) such that*

$$Av = \lambda v. \tag{1.30}$$

The vector v in (1.30) is called eigenvector of A associated with the eigenvalue λ.

Proposition 1.17. *Let A be a real $n \times n$ matrix. Then the following statements hold.*

1. *A real symmetric matrix has only real eigenvalues.*
2. *If A is an $n \times n$ matrix and G a nonsingular $n \times n$ matrix, then A and $G^{-1}AG$ have the same eigenvalues.*
3. *Matrices A and A^T have the same eigenvalues.*
4. *Matrices AB and BA have the same nonzero eigenvalues.*
5. *If $\lambda_1, \ldots, \lambda_n$ are eigenvalues of a nonsingular $n \times n$ matrix A then $\lambda_1^{-1}, \ldots, \lambda_n^{-1}$ are eigenvalues of A^{-1}.*
6. *An idempotent matrix has only eigenvalues 0 or 1.*
7. *If $\lambda_1, \ldots, \lambda_n$ are eigenvalues of a $n \times n$ matrix A, then $|A| = \lambda_1 \cdot \ldots \cdot \lambda_n$ and $tr(A) = \lambda_1 + \ldots + \lambda_n$.*

Definition 1.15. *An $n \times n$ matrix A is said to be diagonalizable if there exists an $n \times n$ nonsingular matrix Q such that $Q^{-1}AQ = D$ for some diagonal matrix D, in which case Q is said to diagonalize A (or A is said to be diagonalized by Q).*

An $n \times n$ matrix A is said to be orthogonally diagonalizable if it is diagonalizable by an orthogonal matrix; that is, if there exists an $n \times n$ orthogonal matrix Q such that Q^TAQ is diagonal.

Proposition 1.18. *An $n \times n$ matrix A is diagonalizable by an $n \times n$ nonsingular matrix Q if and only if the columns of Q are linearly independent eigenvectors of A.*

Proposition 1.19. *Every symmetric matrix is orthogonally diagonalizable.*

Proposition 1.20. *Let A represent an $n \times n$ symmetric matrix, and let $d_1, \ldots, d_n$ represent the (not necessarily distinct) eigenvalues of A (in arbitrary order). Then there exists an $n \times n$ orthogonal matrix Q such that*

$$Q^T AQ = diag(d_1, \ldots, d_n).$$

Proposition 1.21. *Let A be an $n \times n$ symmetric matrix. And let Q represent an $n \times n$ orthogonal matrix and $D = diag(d_1, \ldots, d_n)$ an $n \times n$ diagonal matrix such that $Q^T AQ = D$. Then matrix A can be expressed as*

$$A = QDQ^T. \tag{1.31}$$

The decomposition (1.31), also known under the name of spectral decomposition, is unique aside from the ordering of the diagonal elements (eigenvalues) and corresponding columns in Q (eigenvectors).

Comment: From Proposition 1.20 and Proposition 1.21 it follows: If a symmetric $n \times n$ matrix A has a decomposition $A = QDQ^T$, where Q is an orthogonal matrix and D is a diagonal matrix, then the diagonal elements of D have to be eigenvalues of matrix A.

Definition 1.16. *Let $A_1, \ldots, A_k$ represent k matrices of dimensions $n \times n$. If there exists an $n \times n$ nonsingular matrix Q such that $Q^{-1}A_1Q = D_1, \ldots Q^{-1}A_kQ = D_k$ for some diagonal matrices $D_1, \ldots D_k$, then matrices $A_1, \ldots, A_k$ are simultaneously diagonalizable.*

Proposition 1.22. *If $n \times n$ matrices $A_1, \ldots, A_k$ are simultaneously diagonalizable then they commute in pairs, i.e.*

$$A_s A_i = A_i A_s \, (s > i = 1, \ldots, k).$$

If $n \times n$ symmetric matrices $A_1, \ldots, A_k$ commute in pairs, then they can be simultaneously diagonalized by an orthogonal matrix; that is there exist an orthogonal matrix P and diagonal matrices $D_1, \ldots, D_k$ such that for $i = 1, \ldots, k$

$$P^T A_i P = D_i.$$

Proof. For proof see for example Harville (1997), Theorem 21.13.1. □

2. Linear Mixed Model

2.1 Notation

Consider the model

$$Y = X\beta + Z\gamma + \varepsilon, \tag{2.1}$$

where Y is a vector of n observable random variables, β is a vector of p unknown parameters having fixed values (*fixed effects*), γ is a random vector of length r (*random effects*) and ε is a random n-vector of errors. Matrix X is an $n \times p$ and matrix Z is an $n \times r$ matrix. Both X and Z is assumed to be known and are sometimes referred to as *design* or *model matrices*. Models in the form of (2.1) are called *Linear Mixed Models*.

We assume that the expectations of γ and ε are zero, $E(\gamma) = 0$, $E(\varepsilon) = 0$ and, hence, $EY = X\beta$. In addition we assume

$$\operatorname{Var}\begin{bmatrix}\gamma \\ \varepsilon\end{bmatrix} = \begin{bmatrix}G & 0 \\ 0 & R\end{bmatrix}.$$

The (co-)variance matrix of Y can be expressed as

$$V = R + ZGZ^T. \tag{2.2}$$

Throughout the thesis it is assumed that V and R are nonsingular matrices.

For some results additional distributional assumptions are needed. These frequently used additional assumptions require γ and ε to have multivariate normal distribution with covariance matrices G and R respectively:

$$\gamma \sim \mathcal{N}(0, G); \tag{2.3}$$
$$\varepsilon \sim \mathcal{N}(0, R). \tag{2.4}$$

The distribution of Y is determined by (2.1)–(2.4):

$$Y \sim \mathcal{N}(X\beta, V). \tag{2.5}$$

2.1.1 Examples

Variety of models can be considered as special cases of linear mixed models. By choosing $r = 0$ (no random effects) and by taking $V = \sigma^2 I$ we get the traditional linear model from (2.1). The special cases of traditional linear model like linear regression and analysis of variance (ANOVA) models are therefore also special cases of linear mixed model. Small area estimation methodology used in survey sampling makes heavy use of mixed models, being particularly interested in predicting random effects γ. Multilevel models, popular in human and biological sciences, are also basically mixed models with diagonal matrix G, as are variance components models. Linear mixed models are also an essential technique in longitudinal and spatial data analysis. In spatial or longitudinal data analysis usually considerable effort is directed to modelling and interpretation of matrix R.

2.2 Estimation and prediction

Definition 2.1. *An estimator $A(Y)$ of parameter vector Θ is called Best Linear Unbiased Estimator (BLUE) if*

- *$A(Y)$ is linear estimator:*

$$A(Y) = BY \text{ for some matrix } B; \tag{2.6}$$

- *$A(Y)$ is unbiased:*

$$EA(Y) = \Theta \tag{2.7}$$

- *$A(Y)$ has minimum variance among all linear unbiased estimators:*

$$Var(BY) \leq Var(B_*Y) \tag{2.8}$$

  for any fixed matrix B_* for which $EB_*Y = \Theta$.

Even if the covariance matrix V is known, there exists a BLUE estimator for fixed effects in linear mixed model (2.1) only if matrix X is of full column rank. However, for vector $X\beta$ there exists a BLUE estimator as stated by the extended Gauss-Markov Theorem.

Theorem 2.1. *Consider linear mixed model (2.1) and assume the covariance matrix V to be known. Then the Best Linear Unbiased Estimator (BLUE) of $X\beta$ is given by*

$$X\hat{\beta} = X(X^T V^{-1} X)^- X^T V^{-1} Y. \tag{2.9}$$

There exist extensions of this result to the case where V is singular (see Searle, 1994). In this thesis we assume that V is nonsingular, which allows us to ignore the relatively complex form of the general BLUE estimator.

Definition 2.2. *A predictor $A(Y)$ of a random variable θ is called Best Linear Unbiased Predictor (BLUP) if*

- *$A(Y)$ is linear predictor:*

$$A(Y) = BY \text{ for some matrix } B; \tag{2.10}$$

- *$A(Y)$ is unbiased:*

$$EA(Y) = E\theta \tag{2.11}$$

- *$A(Y)$ has minimum mean square error among all linear unbiased predictors:*

$$E(BY - \theta)^2 \leq E(B_*Y - \theta)^2 \tag{2.12}$$

  for any fixed matrix B_* for which $EB_*Y = E\theta$.

Theorem 2.2. *Consider linear mixed model (2.1) and assume the covariance matrices V and G to be known. Then the Best Linear Unbiased Predictor (BLUP) of γ is given by*

$$\hat{\gamma} = GZ^TV^{-1}(Y - X\hat{\beta}). \tag{2.13}$$

Proof. See for example Henderson (1963) or Searle (1997a). □

Note: Formulas for BLUE (2.9) and BLUP (2.13) are derived without using the distributional assumptions (2.3) and (2.4). However, if one is willing to assume (2.3) and (2.4), then it is possible to use maximum likelihood method to estimate the unknown quantities. The results obtained by maximum likelihood are equal to (2.9) and (2.13) as stated by the following two theorems.

Theorem 2.3. *Consider the linear mixed model (2.1) and assume that the distributional assumption (2.5) holds and that the covariance matrix V is known. Then the maximum likelihood estimator $X\hat{\beta}$ of $X\beta$ is given by (2.9).*

Proof. The well-known proof is presented here in detail because of the author's desire to refer later to some intermediate results.

If the distributional assumptions hold, it is possible to write down likelihood function L for Y:

$$L = |2\pi V|^{-1/2} \exp\left(-\frac{1}{2}(Y - X\beta)^T V^{-1}(Y - X\beta)\right), \tag{2.14}$$

and the log-likelihood function:

$$l = -\frac{1}{2}\ln(|2\pi V|) - \frac{1}{2}(Y - X\beta)^T V^{-1}(Y - X\beta). \tag{2.15}$$

To derive maximum likelihood estimates for β one has to take derivative from l with respect to β,

$$\frac{\partial l}{\partial \beta} = XV^{-1}(Y - X\beta)$$

and equate it to zero. From $\partial l/\partial\beta = 0$ we get the equation

$$XV^{-1}Y = XV^{-1}X\beta. \tag{2.16}$$

This equation has unique solution for β if and only if matrix X is of full column rank. More generally the solution can be written as

$$\hat{\beta} = (XV^{-1}X)^- XV^{-1}Y,$$

where the generalized inverse $(XV^{-1}X)^-$ is not uniquely defined. However, the estimate of $X\beta$,

$$X\hat{\beta} = X(XV^{-1}X)^- XV^{-1}Y$$

is always unique because of the properties of generalized inverse (see Lemma 1.7). □

Theorem 2.4. *Consider the linear mixed model (2.1). We assume that the distributional assumptions (2.3) and (2.4) hold and that the covariance matrices G and R (and hence also V) are known. Then the value of γ which maximizes the joint likelihood function $f_{Y,\gamma}$ of Y and γ is given by (2.13) and the value of β maximizing $f_{Y,\gamma}$ is given by (2.9).*

Proof. Assume for now G to be nonsingular. Then the joint density of Y and γ can be written down as

$$\begin{aligned} f_{Y,\gamma} &= f_{Y|\gamma} \cdot f_\gamma \\ &= |2\pi R|^{-1/2} \exp\left(-\frac{1}{2}(Y - X\beta - Z\gamma)^T R^{-1}(Y - X\beta - Z\gamma)\right) \\ &\quad \times |2\pi G|^{-1/2} \exp\left(-\frac{1}{2}\gamma^T G^{-1}\gamma\right). \end{aligned} \tag{2.17}$$

It is easier to maximize the log-density $\ln(f_{Y,\gamma})$ than the joint density itself. The logarithm of the joint density function is

$$\begin{aligned}\ln(f_{Y,\gamma}) &= -\frac{1}{2}\ln(|2\pi R|) - \frac{1}{2}(Y - X\beta - Z\gamma)^T R^{-1}(Y - X\beta - Z\gamma) \\ &\quad -\frac{1}{2}\ln(|2\pi G|) - \frac{1}{2}\gamma^T G^{-1}\gamma.\end{aligned}$$

Partial derivative of $\ln(f_{Y,\gamma})$ with respect to γ is

$$\frac{\partial f_{Y,\gamma}}{\partial \gamma} = Z^T R^{-1}(Y - X\beta - Z\gamma) + G^{-1}\gamma$$

and partial derivative of $\ln(f_{Y,\gamma})$ with respect to β is

$$\frac{\partial f_{Y,\gamma}}{\partial \beta} = X^T R^{-1}(Y - X\beta - Z\gamma).$$

To find the values of γ and β maximizing (2.17) we equate the partial derivatives to zero and solve the resulting system of equations:

$$\begin{aligned} Z^T R^{-1}Y - Z^T R^{-1}X\beta - (Z^T R^{-1}Z + G^{-1})\gamma &= 0, \qquad (2.18) \\ X^T R^{-1}Y - X^T R^{-1}X\beta - X^T R^{-1}Z\gamma &= 0 \qquad (2.19)\end{aligned}$$

which can be rewritten in the matrix form as

$$\begin{bmatrix} Z^T R^{-1}X & Z^T R^{-1}Z \\ X^T R^{-1}X & X^T R^{-1}Z + G^{-1} \end{bmatrix} \begin{bmatrix} \beta \\ \gamma \end{bmatrix} = \begin{bmatrix} Z^T R^{-1}Y \\ X^T R^{-1}Y \end{bmatrix}. \qquad (2.20)$$

Equations (2.20) are called *Mixed Models Equation* (MME). To solve the MME one can solve (2.19) for γ to derive

$$\gamma = (Z^T R^{-1}Z + G^{-1})^{-1} Z^T R^{-1}(Y - X\beta). \qquad (2.21)$$

The derived value of γ can now be plugged into (2.18):

$$X^T R^{-1}Y - X^T R^{-1}X\beta - X^T R^{-1}Z(Z^T R^{-1}Z + G^{-1})^{-1} Z^T R^{-1}(Y - X\beta) = 0.$$

After some simple algebra we get

$$\begin{aligned} &X^T(R^{-1} - R^{-1}Z(Z^T R^{-1}Z + G^{-1})^{-1} Z^T R^{-1})Y = \\ &\quad X^T(R^{-1} - R^{-1}Z(Z^T R^{-1}Z + G^{-1})^{-1} Z^T R^{-1})X\beta. \end{aligned}$$

Now one may use the equality (which can be easily shown by multiplying with $V = ZGZ^T + R$):

$$R^{-1} - R^{-1}Z(Z^T R^{-1}Z + G^{-1})^{-1} Z^T R^{-1} = V^{-1}$$

to derive

$$X^T V^{-1} Y = X^T V^{-1} X\beta. \tag{2.22}$$

From (2.22) follows

$$\hat{\beta} = (X^T V^{-1} X)^- X^T V^{-1} Y. \tag{2.23}$$

It is worth to notice that $\hat{\beta}$ is not uniquely determined — different generalized inverses of $X^T V^{-1} X$ can lead to different values for $\hat{\beta}$, all of which maximize the likelihood function. Now we can plug (2.23) into (2.21) and use some simple algebra:

$$\begin{aligned}
\hat{\gamma} &= (Z^T R^{-1} Z + G^{-1})^{-1} Z^T R^{-1} (Y - X\hat{\beta}) \\
&= (Z^T R^{-1} Z + G^{-1})^{-1} Z^T R^{-1} V V^{-1} (Y - X\hat{\beta}) \\
&\overset{(2.2)}{=} (Z^T R^{-1} Z + G^{-1})^{-1} Z^T R^{-1} (ZGZ^T + R) V^{-1} (Y - X\hat{\beta}) \\
&= (Z^T R^{-1} Z + G^{-1})^{-1} (Z^T R^{-1} ZGZ^T + Z^T) V^{-1} (Y - X\hat{\beta}) \\
&= (Z^T R^{-1} Z + G^{-1})^{-1} (Z^T R^{-1} Z + G^{-1}) GZ^T V^{-1} (Y - X\hat{\beta}) \\
&= GZ^T V^{-1} (Y - X\hat{\beta}).
\end{aligned} \tag{2.24}$$

This result, together with (2.23), completes the proof for nonsingular G.

Now consider a case where some elements of γ are almost surely linearly dependent, so that covariance matrix G becomes singular. If $\operatorname{rank}(G)$, denoted here by g, is smaller than the number of random effects r, then there exists a (normally distributed) random vector γ_* of length g such, that $\gamma = L\gamma_*$ (and $G = LG_*L^T$, where $\operatorname{Var}(\gamma_*) = G_*$). We can rewrite the mixed model (2.1) in the following way:

$$Y = X\beta + Z_*\gamma_* + \varepsilon, \tag{2.25}$$

where $Z_* = ZL$. There is no problem in writing down BLU predictor for γ_* in (2.25), because the covariance matrix G_* is nonsingular:

$$\hat{\gamma}_* = G_* Z_*^T V^{-1} (Y - X\hat{\beta}). \tag{2.26}$$

But if (2.26) is the best linear unbiased predictor for γ_* then $L\hat{\gamma}_*$ is also the best linear unbiased predictor for $L\gamma_*$. Hence, BLU predictor for $\gamma = L\gamma_*$ is

$$\begin{aligned}
\hat{\gamma} &= LG_* Z_*^T V^{-1} (Y - X\hat{\beta}) \\
&= LG_* L^T Z^T V^{-1} (Y - X\hat{\beta}) \\
&= GZ^T V^{-1} (Y - X\hat{\beta}).
\end{aligned}$$

Therefore, the formula for BLUP (2.13) holds also for a singular G. □

The derivation of BLUE and BLUP assume the variance matrices G and R (and hence V) are known. In practical situation this is rarely the case. One frequently used solution to the problem is to use estimated variance matrices $\tilde{R}, \tilde{G}$ and $\tilde{V}$ instead of the unknown true variance matrices in equations (2.9) and (2.13). The estimator of $X\beta$ derived in this way,

$$X\hat{\beta}_{EBLUE} = X(X^T\tilde{V}^{-1}X)^-X^T\tilde{V}^{-1}Y,$$

is called *Estimated Best Linear Unbiased Estimator* or *EBLUE* and the predictor for γ in the form

$$\hat{\gamma}_{EBLUP} = \tilde{G}Z^T\tilde{V}^{-1}(Y - X\hat{\beta}_{EBLUE})$$

is called *Estimated Best Linear Unbiased Predictor* or *EBLUP*. Because there are more than one possibility to estimate the unknown variance matrices, the $EBLUP$ and $EBLUE$ are relatively wide concepts. The two most frequently used methods to obtain the estimates of variance matrices are maximum likelihood and restricted maximum likelihood (REML). Both of these methods require the additional distributional assumption (2.5). There exist less restrictive methods for estimating G and V to use in EBLUP and EBLUE. For example one can use ANOVA or Minimum Norm Quadratic Estimation (MINQE) to obtain plausible estimates for variance parameters without using the assumption of normality, and use these estimates to derive EBLUP and EBLUE.

Comment on terminology. The term "best linear unbiased predictor" was made popular by Henderson, who started to use it since 1973 to evade criticism of BLUP (Robinson, 1991). Robinson argues that phrases like "estimator of random effects" or "estimate of the realized value" would be more correct. Even though the author of this thesis fully supports the arguments of Robinson, in this thesis the more widespread terminology is used and the estimation of random effects is called prediction of random effects.

2.3 Inference

The questions related to inference remain outside of the main focus of these thesis. Still one basic result and a concept are useful as tools for understanding and interpreting some of the main results presented. The first proposition concerns the sampling variability of the predictors/estimators.

Proposition 2.1. *Let U be a particular generalized inverse $(XV^{-1}X)^-$. The MSE of $\hat{\beta} = UX^TV^{-1}Y$ and BLU predictor $\hat{\gamma}$ of unknown parameters can be calculated using the following result*

$$Var\begin{bmatrix}\hat{\beta}-\beta\\ \hat{\gamma}-\gamma\end{bmatrix} = \begin{bmatrix}D_{11} & D_{12}\\ {D_{12}}^T & D_{22}\end{bmatrix},$$

where

$$\begin{aligned}
D_{11} &= U;\\
D_{12} &= -UX^TR^{-1}ZG^{1/2}(I+G^{1/2}Z^TR^{-1}ZG^{1/2})^{-1}G^{1/2};\\
D_{22} &= G^{1/2}(I+G^{1/2}Z^TR^{-1}ZG^{1/2})^{-1}G^{1/2} -\\
&\quad -D_{12}^TX^TR^{-1}ZG^{1/2}(I+G^{1/2}Z^TR^{-1}ZG^{1/2})^{-1}G^{1/2}.
\end{aligned}$$

Proof. The proof of a slightly more general result (incorporating also some cases where V is singular) can be found in Harville (1976). This particular result has been obtained from the more general case by choosing the decomposition $G = STU$, used by Harville, to be following: $T = I$, $S = G^{1/2}$, $U = G^{1/2}$; and by assuming the existence of V^{-1} and R^{-1}. The result can be further simplified, as is done in Lemma 5.5. The form presented here tries to follow the form given by Harville (1976). □

Notice that because $\hat{\beta}$ is not uniquely determined, also the variance matrix given in Proposition 2.1 is not uniquely determined. However, if one calculates MSE for a prediction of

$$l_1\beta + l_2\gamma, \tag{2.27}$$

where $l_1\beta$ is an estimable (and therefore uniquely determined) linear combination of parameters, then formula given in Proposition 2.1 leads to a uniquely determined variance for (2.27).

There are no exact formula available for calculating MSE for EBLUE or EBLUP estimates. As a first approximation one can use the formulas for MSE of BLUP/BLUE estimates given in Proposition 2.1 together with estimated variance/covariance matrices. This approach, which is known to underestimate the variability of prediction results, is often used by software to estimate mixed models (for example, PROC MIXED in SAS). More exact approximations are available, see for example Prasad & Rao (1990) and Lahiri & Rao (1995). To illustratively compare the accuracy of EBLUP/EBLUE and BLUP/BLUE estimators, two simple sample

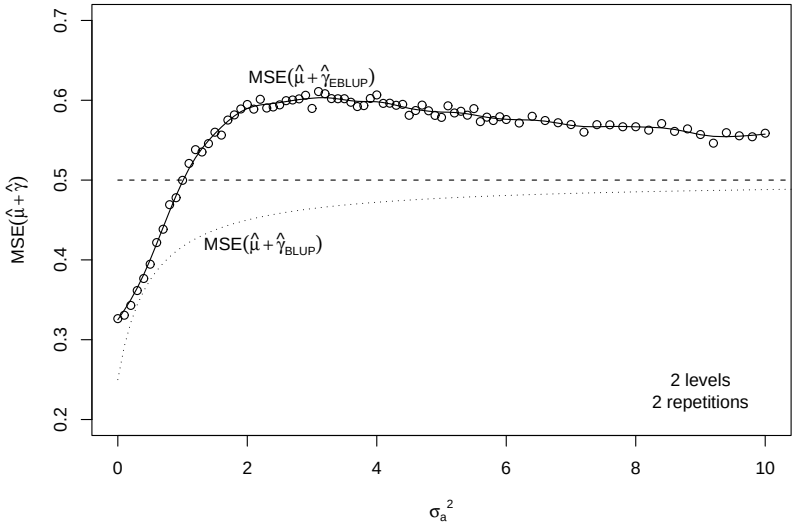

Figure 2.1: Precision of BLU and EBLU predictors. Random factor with 2 levels

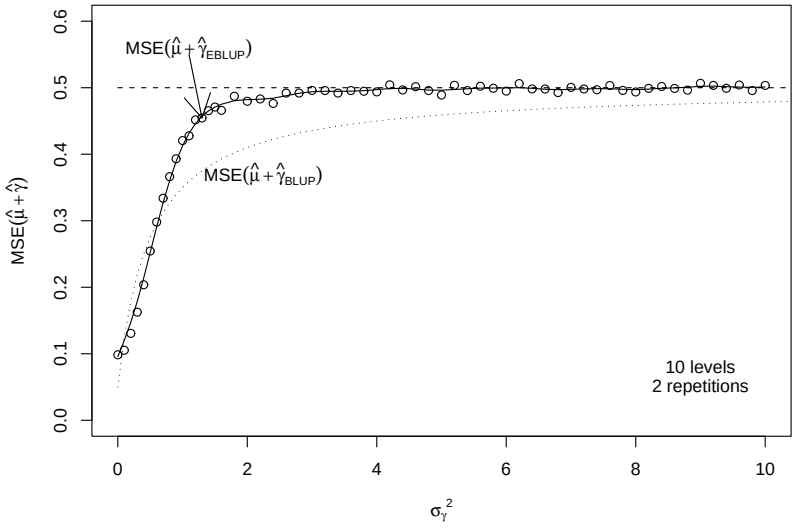

Figure 2.2: Precision of BLU and EBLU predictors. Random factor with 10 levels

designs are considered. Figure 2.1 describes the accuracy of EBLU predictor (obtained using ML estimates of variance components) for a model with one random factor with two observed levels. On each level there are two observations available. Figure 2.2 describes also a model with one random factor, but with ten levels sampled. If there are sampled only a few levels from a random factor, then the prediction accuracy achieved by including the factor to the model as a fixed factor can be considerably better than the accuracy of EBLUP predictions. The prediction accuracy of a random factor achieved by considering it as a fixed factor is illustrated by the straight dashed line in both figures.

Several other aspects of inference, like hypothesis testing, are not covered here. Interested reader is referred to Khuri, Mathew & Sinha (1998) for more detailed presentation of the topic.

3. REML Estimation

The REML (REsidual Maximum Likelihood or REstricted Maximum Likelihood) method was put on a broad basis by Patterson & Thompson (1971). Often REML is considered as a method providing the estimates of unknown variance parameters only. For example, Searle *et al* (1992) state: "REML estimation includes no procedure for estimating fixed effects". However, in practice, REML estimates of variance parameters are often used to derive estimates for fixed parameters. In this chapter an overview of REML methodology is presented. Also a new approach is proposed to look at the REML estimation of variance and fixed parameters in a unified way.

Majority of leading statistical software packages use REML as default method of estimating covariance parameters and fixed effects in mixed models (SAS version 8.2 — procedure MIXED, R version 1.9 — function lme, S-Plus 2000 — function lme, SPSS version 12.0 — procedure MIXED).

3.1 REML estimation of variance parameters

One way to present the idea of REML estimation is the following. Estimate the fixed effects using ordinary least squares estimator (OLSE): $X\hat{\beta}_{OLSE} = P_X Y = X(X^T X)^- XY$. This estimator does not depend on unknown variance parameters. One may then calculate the OLSE residuals $\hat{\varepsilon}_{OLSE} = Y - P_X Y = P_{X^\perp} Y$. These residuals are normally distributed, $\hat{\varepsilon}_{OLSE} \sim \mathcal{N}(0, P_{X^\perp} V P_{X^\perp}^T)$. The distribution of OLSE residuals does not depend any more on unknown fixed effects. Now estimate the unknown variance parameters by maximizing the likelihood of OLSE residuals. This is why the method is called REsidual Maximum Likelihood or REML. The OLSE residuals are linearly dependent, $\mathrm{rank}(P_{X^\perp}) = n - \mathrm{rank}(X) = n - k$, and therefore their distribution is singular multinormal distribution. It is often easier to work with nonsingular variance matrix, therefore traditionally instead of n dependent residuals one often works with $n-k$ independent residuals, or more generally, with $n - k$ independent linear combinations of residuals. Let L be a $(n - k) \times n$ matrix selecting $n - k$ independent

residuals (or independent linear combinations of residuals, or error contrasts). Then matrix K, $K = LP_{X^\perp}$, is an $(n-k) \times n$ matrix with full row rank, $\text{rank}(LP_{X^\perp}) = n - k$. Hence one may maximize the likelihood (or log-likelihood) of KY instead of the likelihood of $P_{X^\perp}Y$. This leads to the most widespread definition of REML — see for example Searle *et al* (1992), Verbeke & Molenberghs (2000), Patterson & Thompson (1971). Suppose that each element of V is a differentiable function of covariance parameters $\nu_1, \ldots, \nu_t$. Choose unknown covariance parameters to maximize the log-likelihood

$$l_{KY} = -\frac{1}{2} log|2\pi KVK^T| - \frac{1}{2} Y^T K^T (KVK^T)^{-1} KY \tag{3.1}$$

of KY. To maximize l_{KY}, we differentiate (3.1) with respect to $\nu_1, \ldots, \nu_t$:

$$\begin{aligned} \frac{\partial l_{KY}}{\partial \nu_i} &= -\frac{1}{2}\text{tr}((KVK^T)^{-1} K \frac{\partial V}{\partial \nu_i} K^T) \\ &\quad -\frac{1}{2} Y^T K^T (KVK^T)^{-1} K \frac{\partial V}{\partial \nu_i} K^T (KVK^T)^{-1} KY. \end{aligned} \tag{3.2}$$

To obtain the REML estimates one has to equate (3.2) to zero and solve the resulting equations. Sometimes additional restrictions are added to covariance parameters (for example in variance components models the estimates of covariance parameters are often required to be positive, $\nu_i > 0$). One should notice, that (3.2) depends on the matrix K only via the matrix $K^T(KVK^T)^{-1}K$, because

$$\text{tr}((KVK^T)^{-1} K \frac{\partial V}{\partial \nu_i} K^T) = tr(K^T (KVK^T)^{-1} K \frac{\partial V}{\partial \nu_i}).$$

But, according to Theorem 1.1,

$$K^T (KVK^T)^{-1} K = V^{-1} - V^{-1} X (X^T V^{-1} X)^- X^T V^{-1},$$

which does not depend on the choice of K. Therefore the REML estimates do not depend on the particular choice of matrix K.

There exists another interpretation of REML. Frequently the main interest concerns the covariance parameters, not the fixed effects (for example in estimating heritability in genetics). One possibility to eliminate the nuisance parameters (fixed effects) is by conditioning on an appropriate sufficient statistic. Cox & Reid (1987) have shown using approximate methods that conditioning on a sufficient statistic for fixed effects leads to REML estimates of variance parameters (parameters of interest). Smyth & Verbyla (1996) have shown, that the proposed equality is exact.

3.2 REML estimation of fixed effects

Some authors consider the REML estimation method as a method providing the estimates of unknown variance parameters only. For example, Searle *et al* (1992) state: "REML estimation includes no procedure for estimating fixed effects". Later the same authors argue, that it seems to be reasonable to use maximum likelihood estimate of $X\beta$ together with covariance matrix estimate $\tilde{V}$ obtained by REML.

Others provide a method to estimate the $X\beta$ in the context of REML, results of which coincide with the previous suggestion. Patterson and Thompson in their article separate the likelihood of Y into two parts, $l_Y = l_1 + l_2$. By maximizing the first part l_1, which is equivalent to maximization of (3.1), they obtain the REML estimate $\tilde{V}$ of V. Then the second part of the likelihood l_2 (which depends on V) is maximized with respect to β to find the REML estimates of fixed effects. It is noted that given the covariance matrix V, maximizing l_2 and l_Y with respect to β will yield equivalent answers.

In practice the REML estimation of fixed effects tends to work reasonably well as illustrated in the following example.

Example 3.1. Consider following random effects model:

$$Y = \mu + \gamma_i + \varepsilon_{ij}, i = 1 \dots 3, j = 1 \dots n_i,$$

where $n_1 = 1, n_2 = 2, n_3 = 10$, γ_i are independent, $\gamma_i \sim \mathcal{N}(0, \sigma_\gamma^2)$, and ε_{ij} are independent, $\varepsilon_{ij} \sim \mathcal{N}(0, \sigma_\varepsilon^2)$. The parameters for this unbalanced model can be estimated using both ML or REML. In a simulation study the relative precision of REML estimator of μ was compared to the precision of ML estimator. For each value of $\sigma_\gamma/\sigma_\varepsilon$ there were performed 20 000 simulations. Simulation results are presented in Figure 3.1.

It may be noticed that neither REML nor ML are uniformly better with respect to precision measured by Mean Square Error.

3.3 REML estimation as ML estimation with respect to a misspecified covariance structure

For a covariance structure V one can derive the ML estimates of covariance and fixed effects parameters. It is also possible to use the REML estimation methodology described in sections 3.1 and 3.2. In this section a modified

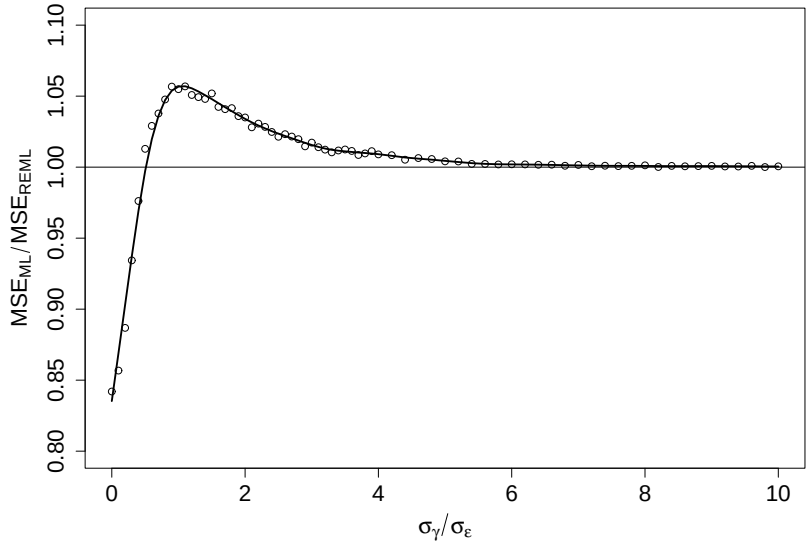

Figure 3.1: Comparison of precision of REML and ML estimators: estimation of intercept μ.

covariance structure $V_* = V_*(V)$ is constructed and some of it's properties are described. It is shown that if one uses V_* instead of V during the estimation of unknown parameters (fixed effects and covariance parameters), then the ML estimates for V_* are equal to the REML estimates obtained using V. In other words, REML estimation methodology can be viewed as ML estimation with respect to a misspecified covariance structure.

3.3.1 Definition of the covariance structure $V_*(V)$

In this subsection we construct an alternative covariance structure $V_* = V_*(V)$ for Y. It will be shown in the next subsection, that if one uses this misspecified covariance structure V_* to obtain ML estimates for covariance parameters and fixed effects, then the results are equal to REML estimates obtained using the true covariance structure V.

Define

$$V_* = 1/c \cdot X(X^T U X)^- X^T + P_{X^\perp,V} V P_{X^\perp,V}^T, \tag{3.3}$$

where U is any symmetric positive definite $n \times n$ matrix not depending on β and the scalar $c = c(V)$ is defined by formula

$$c = \left(\frac{|2\pi KVK^T|}{|D_1^*| \cdot |2\pi V|} \right)^{-1/k}, \tag{3.4}$$

where $k = \text{rank}(X)$ and D_1^* denotes the diagonal matrix of nonzero eigenvalues of matrix A,

$$A = V^{-1/2} X (X^T U X)^- X^T V^{-1/2}. \tag{3.5}$$

Remark: the matrix A in (3.5) is non-negative definite (see Proposition 1.11) and, hence, it's eigenvalues are non-negative. Therefore matrices KVK^T, D_1^* and V are all symmetric positive definite, and their determinants are hence all greater than 0. Therefore c is correctly defined and $c > 0$.

The Definition (3.3) describes a relatively wide class of covariance matrices. To determine V_* uniquely, one has to choose some particular value for matrices U and K.

It may not be obvious, whether the alternative covariance structure presented in (3.3) is positive definite or not. The following lemma proves that proposed covariance structure is positive definite indeed.

Lemma 3.1. *If V_* has the form given in (3.3), then V_*^{-1} exists and*

$$V_*^{-1} = c \cdot P_{X,V}^T U P_{X,V} + P_{X^\perp,V}^T V^{-1} P_{X^\perp,V}. \tag{3.6}$$

Proof. Define $B = c \cdot P_{X,V}^T U P_{X,V} + P_{X^\perp,V}^T V^{-1} P_{X^\perp,V}$. Then

$$\begin{aligned} V_* \cdot B &= [1/c \cdot X(X^T U X)^- X^T + P_{X^\perp,V} V P_{X^\perp,V}^T] \\ &\quad \cdot [c \cdot P_{X,V}^T U P_{X,V} + P_{X^\perp,V}^T V^{-1} P_{X^\perp,V}] \\ &= X(X^T U X)^- X^T P_{X,V}^T U P_{X,V} \\ &\quad + P_{X^\perp,V} V P_{X^\perp,V}^T P_{X^\perp,V}^T V^{-1} P_{X^\perp,V} \end{aligned} \tag{3.7}$$

because $P_{X^\perp,V}^T P_{X,V}^T = 0$ and $X^T P_{X^\perp,V}^T = 0$ (property 1.18).

Using $P_{X^\perp,V}^T P_{X^\perp,V}^T = P_{X^\perp,V}^T$ (property (1.17)) and $X^T P_{X,V}^T = X^T$ (property (1.18)) one may simplify the expression (3.7) further to obtain

$$V_* \cdot B = X(X^T U X)^- X^T U P_{X,V} + P_{X^\perp,V} V P_{X^\perp,V}^T V^{-1} P_{X^\perp,V}. \tag{3.8}$$

Notice now that, via (1.15) and (1.18),

$$\begin{aligned} X(X^TUX)^-X^TUP_{X,V} &= P_{X,U^{-1}}P_{X,V} \\ &= P_{X,U^{-1}}X(X^TV^{-1}X)^-X^TV^{-1} \\ &= X(X^TV^{-1}X)^-X^TV^{-1} \\ &= P_{X,V}, \end{aligned}$$

which together with property (1.24) yields

$$\begin{aligned} (3.8) \quad &= P_{X,V} + P_{X^\perp,V}VV^{-1}P_{X^\perp,V}P_{X^\perp,V} \\ &= P_{X,V} + P_{X^\perp,V} \\ &= I. \end{aligned}$$

Therefore $V_* \cdot B = I$ and hence $B = V_*^{-1}$. □

The misspecified model for Y can be defined in the following way:

$$Y \sim \mathcal{N}(X\beta, V_*),$$

and the log-likelihood for the misspecified model is

$$l_* = -\frac{1}{2}\ln(|2\pi V_*|) - \frac{1}{2}(Y - X\beta)^T V_*^{-1}(Y - X\beta). \tag{3.9}$$

If the covariance matrix V depends on unknown covariance parameters $\nu_1, \ldots, \nu_t$, then the covariance matrix V_* is also a function of these unknown parameters.

One can consider some particular choice for matrix U. For example, one may choose $U = V^{-1}$. Then matrix A, defined in (3.5), is idempotent and non-negative definite. Therefore, all it's non-zero eigenvalues are equal to one and, hence, $|D_1^*| = 1$. This leads to a simplified formula for c in (3.4):

$$c = \left(\frac{|2\pi KVK^T|}{|2\pi V|}\right)^{1/k}$$

and the alternative covariance structure V_* has the form

$$V_* = \left(\frac{|2\pi KVK^T|}{|2\pi V|}\right)^{1/k} P_{X,V}V + P_{X^\perp,V}V \tag{3.10}$$

and

$$V_*^{-1} = \left(\frac{|2\pi KVK^T|}{|2\pi V|}\right)^{-1/k} P_{X,V}^T V^{-1} + P_{X^\perp,V}^T V^{-1}. \tag{3.11}$$

Example 3.2. Consider n independent normally distributed random variables with common mean,

$$Y \sim \mathcal{N}(\mathbf{1}\mu, I\sigma^2),$$

where $\mathbf{1}$ denotes a column of ones. We construct a matrix $V_*(V)$ for this particular example, using $U = V^{-1}$ and taking K equal to first $n-1$ rows from the matrix $\kappa^{1/(2n-2)} P_{X^\perp,V}$, where κ is any strictly positive constant. Then

$$c = \frac{\kappa}{2\pi\sigma^2 n} \tag{3.12}$$

and $V_*(V)$ has the form of

$$V_*(V) = \left(I - \frac{1}{n}J\right)\sigma^2 + \frac{\kappa}{2\pi n^2}J, \tag{3.13}$$

where J denotes an $n \times n$ matrix of ones. By choosing $\kappa = 2\pi n^2$ we can further simplify Equation 3.13 to derive

$$V_*(V) = \left(I - \frac{1}{n}J\right)\sigma^2 + J.$$

3.3.2 Derivation of ML estimates using covariance structure $V_*(V)$

First we prove a technical lemma.

Lemma 3.2. *For V_* defined in (3.3) the following equality holds:*

$$|2\pi V_*| = |2\pi K V K^T|.$$

Proof. On can express V_* in the following way:

$$\begin{aligned} V_* &= V^{1/2} V^{-1/2} \left[1/c \cdot X(X^T U X)^- X^T + P_{X^\perp,V} V P_{X^\perp,V}^T\right] V^{-1/2} V^{1/2} \\ &= V^{1/2}[1/c \cdot V^{-1/2} X(X^T U X)^- X^T V^{-1/2} \\ &\qquad + V^{-1/2} P_{X^\perp,V} V P_{X^\perp,V}^T V^{-1/2}] V^{1/2} \\ &= V^{1/2}[1/c \cdot A + B] V^{1/2}, \end{aligned}$$

where

$$A = V^{-1/2} X(X^T U X)^- X^T V^{-1/2}$$

and

$$\begin{aligned} B &= V^{-1/2} P_{X^\perp,V} V P_{X^\perp,V}^T V^{-1/2} \\ &= I - V^{-1/2} X((V^{-1/2}X)^T V^{-1/2} X)^- (V^{-1/2}X)^T. \end{aligned}$$

Matrices A and B are symmetric and they commute because of $AB = 0$ and $BA = 0$. Hence they are simultaneously diagonalizable by an orthogonal matrix O (Proposition 1.22):

$$OAO^T = D_1 \text{ and } OBO^T = D_2,$$

where D_1 and D_2 are diagonal matrices of the eigenvalues of A and B, correspondingly. This result can be rewritten in the following form:

$$A = O^T D_1 O \text{ and } B = O^T D_2 O.$$

Therefore, it follows from $AB = 0$ that $D_1 D_2 = 0$.

Because A and B are simultaneously diagonalizable by orthogonal matrix O, the following equality holds:

$$\begin{aligned} |V_*| &= |V^{1/2} [1/c \cdot A + B] V^{1/2}| \\ &= |V||O^T [1/c \cdot D_1 + D_2] O| \\ &= |V||1/c \cdot D_1 + D_2|, \end{aligned}$$

because the determinant of an orthogonal matrix is ± 1.

The rank of matrix A equals to the rank of X:

$$\begin{aligned} \text{rank}(A) &= \text{rank}(V^{-1/2} X (X^T U X)^- X^T V^{-1/2}) \\ &\overset{(1.1)}{=} \text{rank}(X (X^T U X)^- X^T) \\ &\overset{(1.1)}{=} \text{rank}(X (X^T U X)^- X^T U) \\ &\overset{(1.19)}{=} \text{rank}(X) = k \end{aligned}$$

and the rank of matrix B is $n - k$:

$$\begin{aligned} \text{rank}(I - A) &= rank\left(I - (V^{-1/2}X)\left[X^T V^{-1} X\right]^- (V^{-1/2}X)^T\right) \\ &\overset{(1.20)}{=} n - \text{rank}(V^{-1/2}X) \\ &\overset{(1.1)}{=} n - \text{rank}(X). \end{aligned}$$

Therefore, matrix A has k and matrix B has $n-k$ nonzero (positive) eigenvalues. From $D_1 D_2 = 0$ one can conclude, that these nonzero elements occupy different positions on the diagonal and hence $|D_1 + D_2|$ can be calculated as a product of nonzero eigenvalues of matrices A and B. Because matrix B is idempotent, it's eigenvalues are either 0 or 1, and the product of nonzero eigenvalues of matrices A and B is therefore equal to the product of nonzero eigenvalues of matrix A.

After denoting with D_1^* a diagonal matrix of the nonzero eigenvalues of matrix A, we have

$$\begin{aligned} |2\pi V_*| &= |c^{-1} D_1^*| \cdot |2\pi V| \\ &= c^{-k} |D_1^*| |2\pi V|. \end{aligned} \tag{3.14}$$

After substituting the value of c,

$$c = \left(\frac{|2\pi K V K^T|}{|D_1^*| \cdot |2\pi V|} \right)^{-1/k},$$

(3.14) can be written as

$$\begin{aligned} |2\pi V_*| &= \frac{|2\pi K V K^T|}{|D_1^*| \cdot |2\pi V|} |D_1^*| |2\pi V| \\ &= |2\pi K V K^T|, \end{aligned}$$

which completes the proof. □

Theorem 3.1. *The ML estimates of covariance and fixed effects parameters obtained using misspecified model $Y \sim \mathcal{N}(X\beta, V_*)$, where V_* is given in (3.3), are equal to the REML estimates for the true model $Y \sim \mathcal{N}(X\beta, V)$.*

Proof. If l_* is the log-likelihood for the misspecified model (3.9), then

$$\begin{aligned} 2l_* &= -\ln(|2\pi V_*|) - (Y - X\beta)^T V_*^{-1} (Y - X\beta) \\ &\overset{(3.6)}{=} -\ln(|2\pi V_*|) - c(Y - X\beta)^T P_{X,V}^T U P_{X,V} (Y - X\beta) \\ &\quad -(Y - X\beta)^T P_{X^\perp,V}^T V^{-1} P_{X^\perp,V} (Y - X\beta) \\ &\overset{(1.18)}{=} -\ln(|2\pi V_*|) \end{aligned}$$

$$
\begin{aligned}
& - \left(c(P_{X,V}Y - X\beta)^T U (P_{X,V}Y - X\beta) + (P_{X^\perp,V}Y)^T V^{-1} (P_{X^\perp,V}Y)\right) \\
= \quad & - \ln(|2\pi K V K^T|) - \\
& - \left(c(P_{X,V}Y - X\beta)^T U (P_{X,V}Y - X\beta) + K^T (KVK^T)^{-1} K\right).
\end{aligned}
$$

Last equality follows from Lemma 3.2 and (1.27). Clearly, the likelihood can be maximized with respect to β by choosing $X\hat{\beta} = P_{X,V}Y$ what is equivalent to choosing a generally non-uniquely determined $\hat{\beta}$ of the form

$$\hat{\beta} = (X^T V^{-1} X)^- X^T V^{-1} Y.$$

After the maximization with respect to β one has to maximize the likelihood with respect to the covariance parameters, but the function to maximize is equal to the likelihood of KY used by REML to estimate the covariance parameters (see Equation 3.1). Hence, also the estimates of covariance parameters will coincide with those obtained by REML. □

4. Reparameterisation of Fixed and Random Effects

Identifiability problems for fixed effects have been under great attention. The overview of the developments in the area, given in Section 4.1, is mainly based on Scheffe (1959) and Searle (1971). Interested reader is referred to these sources for in-depth coverage of the topic. Problems related to reparameterisation of random effects have largely remained neglected. Some results on reparameterisation of random effects are given in Möls (2004). These results, together with added details, are presented in Section 4.2.

4.1 Estimable functions and Reparameterisation Constraints for Fixed Effects

The BLU estimator for fixed effects,

$$\hat{\beta} = (X^T V^{-1} X)^- X^T V^{-1} Y,$$

is not uniquely determined if matrix X is not of full column rank. Namely, if X is not of full column rank, then there exist infinity many different generalized inverses of matrix $X^T V^{-1} X$ leading to different estimates of β. Two strategies are commonly used to cope with the problem of identifiability. One approach is to restrict the interest to estimable parameter functions only.

Definition 4.1. *The estimable parameter function is a linear combination of fixed effects which can be expressed as $kX\beta$, where k is row-vector of length n.*

Even though $\hat{\beta}$ may not be uniquely determined, the functions of the form $X\hat{\beta}$ are unique because the value of $X(X^T V^{-1} X)^- X^T$ does not depend on the choice of generalized inverse (see Proposition 1.7).

Another commonly used solution of the parameter identifiability problem is to use additional restrictions to determine uniquely the parameter values.

These restrictions are sometimes called *usual constraints* or *reparameterisation constraints*. For example, for one-way ANOVA model

$$Y_{ij} = \mu + \alpha_i + \varepsilon_{ij},\ i = 1 \dots m,\ j = 1 \dots n_i$$

the two most frequently used reparameterisation constraints are $\sum \alpha_i = 0$ and $\alpha_m = 0$.

Not all possible constraints are valid reparameterisation constraints. A constraint in the form of $L\hat{\beta} = 0$, where L is some row-vector, is reparameterisation constraint if and only if it is non-estimable function of parameters. To determine uniquely fixed effects parameter estimates one needs $p - k$ additional linearly independent constraints, where p is the number of fixed effects' parameters and $k = \text{rank}(X)$, e.g. the matrix H has to have $n - k$ linearly independent rows, and each row has to be a non-estimable function of parameters. These requirements are rephrased in the following proposition.

Proposition 4.1. *Suppose that X is $n \times p$ matrix and H is $t \times p$, rank$(X) = k$. Then the system*

$$\begin{aligned} X^T V^{-1} X\beta &= X^T V^{-1} Y \\ H\beta &= 0 \end{aligned}$$

has a unique solution for any value of Y if and only if the following conditions hold

$$\begin{aligned} \textit{rank}(X) + \textit{rank}(H) &= \textit{rank}\begin{pmatrix} X \\ H \end{pmatrix} \\ \textit{rank}(X) + \textit{rank}(H) &= p. \end{aligned}$$

In the following part of the chapter we show that similar conditions are required to determine the EBLU predictor uniquely.

4.2 Reparameterisation Constraints for Random Effects

To derive unique estimates for fixed effects one had to use additional "usual constraints" as described in the previous section. Random effects are often (mistakenly) considered to be free of such indeterminacy problems. The BLU predictor of γ,

$$\hat{\gamma}(R, G) = GZ^T V^{-1}(Y - X\hat{\beta}), \tag{4.1}$$

where $V = ZGZ^T + R$, is a straightforward uniquely determined function of Y. The BLU predictors certainly follow some rules. For each random effects factor the sum of the BLUP's of it's effects is zero as stated by Searle (1997b); McLean, Sanders & Stroup (1991). These restrictions are the consequence of the BLU prediction process as clearly described by Searle (1997b): "These restrictions are a consequence of the very form of BLUP; they are not a consequence of any definitional restrictions such as $\sum \alpha_i = 0$ often seen as part of the model equation $Y_{ij} = \mu + \alpha_i + \varepsilon_{ij}$ for the one-way classification."

However, situation can change radically, if one does not know exactly the dispersion matrices G and V. If one has to estimate the variance matrices, then additional restrictions are required to determine uniquely the EBLU predictor

$$\hat{\gamma}(\hat{R}, \hat{G}) = \hat{G}Z^T\hat{V}^{-1}(Y - X\hat{\beta}(\hat{V})),$$

as will be seen in this chapter. It will be shown that these required additional restrictions may be stated exactly the same way as in classical restrictions for fixed effects.

For fixed effects estimable parameter functions are of special importance. For random effects it will be also shown, in this section, that predictors for linear combinations depending on random effects only as functions of $Z\gamma$ can be uniquely determined for wide class of problems where G (and γ) cannot be determined uniquely.

4.2.1 Existence of a model with constrained random factors

First imagine two separate mixed models which differ from each other only by having different covariance matrices for γ. For Model 1 (reference model) $\mathrm{Var}(\gamma) = G$ and for Model 2 (alternative model) $\mathrm{Var}(\gamma) = G^*$. The first theorem proves the existence of such alternative model for which BLU predictors satisfy "reparameterisation constraints" in the form $H\hat{\gamma} = 0$ and for which the BLU predictors for linear combinations of $Z\gamma$ ("predictable linear combinations") would be the same as for the reference model.

Theorem 4.1. *Let H be a matrix for which the equality*

$$\mathit{rank}\begin{pmatrix} Z \\ H \end{pmatrix} = \mathit{rank}(Z) + \mathit{rank}(H) \tag{4.2}$$

holds. Then there exists a matrix G^* such, that

$$H\hat{\gamma}(R, G^*) \quad = \quad 0,$$

$$Z\hat{\gamma}(R, G^*) \quad = \quad Z\hat{\gamma}(R, G),$$

where $\hat{\gamma}(R, G)$ is the BLU predictor as defined in Equation 4.1.

Proof. From (4.2) and Proposition 1.6 there follows existence of a symmetric positive definite matrix A such that

$$HAZ^T = 0.$$

Using matrix A, define a projector matrix $P_{Z^T,A}$ projecting onto column space of Z^T:

$$P^T_{Z^T,A} = AZ^T(ZAZ^T)^- Z.$$

Obviously

$$HP^T_{Z^T,A} = HAZ^T(ZAZ^T)^- Z = 0,$$

and from (1.18) it follows that

$$ZP^T_{Z^T,A} = (P_{Z^T,A}Z^T)^T = (Z^T)^T = Z.$$

As the final step, define matrix G^* as the following covariance matrix:

$$G^* = \mathrm{Var}(P^T_{Z^T,A}\gamma) = P^T_{Z^T,A} G P_{Z^T,A}. \tag{4.3}$$

Then

$$ZG^*Z^T = ZP^T_{Z^T,A} G (ZP^T_{Z^T,A})^T = ZGZ^T$$

and, hence,

$$V(G^*) = ZG^*Z^T + R = ZGZ^T + R = V.$$

If $V(G^*) = V$ then also $\hat{\beta}(V(G^*)) = \hat{\beta}(V)$, and

$$\begin{aligned} Z\hat{\gamma}(R, G^*) &= ZG^*Z^T V^{-1}(Y - X\hat{\beta}(V(G^*))) \\ &= ZGZ^T V^{-1}(Y - X\hat{\beta}(V)) \\ &= Z\hat{\gamma}(R, G). \end{aligned}$$

Vector $\hat{\gamma}(R, G^*)$ of predicted random effects satisfies condition $H\hat{\gamma}(R, G^*) = 0$ because of $HP^T_{Z^T,A} = 0$:

$$\begin{aligned} H\hat{\gamma}(R, G^*) &= HG^*V(G^*)^-(Y - X\hat{\beta}(V(G^*)) \\ &\overset{(4.3)}{=} HP^T_{Z^T,A} G P_{Z^T,A} V(G^*)^-(Y - X\hat{\beta}(V(G^*)) \\ &= 0. \end{aligned}$$

□

4.2.2 Identifiability of random effects

Consider the likelihood function for normally distributed Y,

$$L(G|Y) = |2\pi(R+ZGZ^T)|^{-1/2} \exp(-\frac{1}{2}(Y-X\beta)^T(R+ZGZ^T)^{-1}(Y-X\beta)).$$

Replacing G with some other covariance matrix may change the likelihood function, but does not have to. Let us consider a class of covariance matrices $\mathcal{G}(G)$ which all lead to the same likelihood function (for any possible value of Y):

$$\mathcal{G}(G) = \{G_i : L(G|Y) = L(G_i|Y) \mid \text{ for all Y}\}.$$

The set $\mathcal{G}(G)$ consists of covariance matrices which are in the equally good agreement with observed data and, hence, can not be preferred one over another on the basis of observed data alone. Note, that the matrix G^* defined in (4.3) also belongs to this set, $G^* \in \mathcal{G}(G)$.

Two covariance matrices $G_1, G_2 \in \mathcal{G}(G)$ may lead to different predictors for γ. However, if $G_1, G_2 \in \mathcal{G}(G)$ then $Z\hat{\gamma}(R, G_1) = Z\hat{\gamma}(R, G_2)$, as is proved in the following lemma.

Lemma 4.1. *If $G_1, G_2 \in \mathcal{G}(G)$ then $Z\hat{\gamma}(R, G_1) = Z\hat{\gamma}(R, G_2)$.*

Proof. Instead of the likelihood function, it is easier to work with log-likelihood function

$$\begin{aligned} l(G|Y) &= \ln(L(G|Y)) \\ &= \ln(|2\pi V|^{-1/2}) - \frac{1}{2}(Y - X\beta)^T V^{-1}(Y - X\beta). \end{aligned} \tag{4.4}$$

where $V(G) = R + ZGZ^T$. If $G_1, G_2 \in \mathcal{G}(G)$ then $l(G_1|Y) = l(G_2|Y)$ for any value of Y. Choosing $Y = X\beta$ one gets the equality

$$\ln(|2\pi V(G_1)|^{-1/2}) = \ln(|2\pi V(G_2)|^{-1/2}), \tag{4.5}$$

where $V(G_1) = R + ZG_1Z^T$ and $V(G_2) = R + ZG_2Z^T$. Choosing $Y = X\beta + v$ and using (4.5) it follows from the equality of likelihood functions that

$$v^T V(G_1)^{-1} v = v^T V(G_2)^{-1} v. \tag{4.6}$$

The equation

$$v^T(V(G_1)^{-1} - V(G_2)^{-1})v = 0 \tag{4.7}$$

holds for any value of v. In addition the matrix $V(G_1)^{-1} - V(G_2)^{-1}$ is symmetric. Consequently

$$V(G_1)^{-1} = V(G_2)^{-1}. \tag{4.8}$$

Due to (4.8) also $X\hat{\beta}(V(G_1)) = X\hat{\beta}(V(G_2))$. Because of

$$Z\hat{\gamma}(R, G_1) \;=\; ZG_1Z^TV(G_1)^{-1}(Y - X\hat{\beta}(V(G_1)))$$

and $ZG_1Z^T = V(G_1) - R$, it follows that

$$\begin{aligned} Z\hat{\gamma}(R, G_1) &= (Y - X\hat{\beta}(V(G_1))) - RV(G_1)^{-1}(Y - X\hat{\beta}(V(G_1))) \\ &= (Y - X\hat{\beta}(V(G_2))) - RV(G_2)^{-1}(Y - X\hat{\beta}(V(G_2))) \\ &= Z\hat{\gamma}(R, G_2), \end{aligned}$$

which proves the lemma. □

Theorem 4.2. *Let $n \times k$ matrix Z and $v \times k$ matrix H satisfy the condition*

$$rank\begin{pmatrix} Z \\ H \end{pmatrix} = rank(Z) + rank(H) = k.$$

Then the condition $H\hat{\gamma} = 0$ determines uniquely the predictor of γ within the set of covariance matrices $\mathcal{G}(G)$.

Proof. It follows from Theorem 4.1 that there exists a matrix $G^* \in \mathcal{G}(G)$ such that the condition $H\hat{\gamma}(R, G^*) = 0$ holds. Let two matrices $G_1, G_2 \in \mathcal{G}(G)$ satisfy the condition $H\hat{\gamma}(R, G_1) = H\hat{\gamma}(R, G_2) = 0$. Then one may write

$$\begin{bmatrix} Z \\ H \end{bmatrix} \hat{\gamma}(R, G_1) = \begin{bmatrix} Z \\ H \end{bmatrix} \hat{\gamma}(R, G_2). \tag{4.9}$$

Because the rank of $(n + v) \times k$ matrix $\begin{bmatrix} Z \\ H \end{bmatrix}$ is k, it is a full column rank matrix and therefore the inverse of the matrix $\begin{bmatrix} Z \\ H \end{bmatrix}^T \begin{bmatrix} Z \\ H \end{bmatrix}$ exists. It is sufficient to multiply the equation (4.9) from the left with the matrix

$$\left[\begin{bmatrix} Z \\ H \end{bmatrix}^T \begin{bmatrix} Z \\ H \end{bmatrix}\right]^{-1} \begin{bmatrix} Z \\ H \end{bmatrix}^T$$

to get the conclusion

$$\hat{\gamma}(R, G_1) = \hat{\gamma}(R, G_2).$$

which proves the uniqueness. □

4.2.3 Discussion

For fixed effects it is well known that if the matrix X is not of full column rank, then one needs to apply additional constraints to define the unknown parameters β uniquely. These constraints can be represented in the form $H\beta = 0$, where

$$\operatorname{rank}\begin{pmatrix} X \\ H \end{pmatrix} = \operatorname{rank}(X) + \operatorname{rank}(H).$$

The validity of these constraints on observed data cannot be assessed using the observed data. Different constraints will lead to different, but equivalent parameterisations of the linear model. Nevertheless linear combinations of the parameters in the form $X\beta$ will remain equal for all parameterisations.

It was shown in this chapter, that based on the observed data, one cannot identify exactly the covariance matrix $G = \operatorname{Var}(\gamma)$, if the matrix G does not have full column rank. Unidentifiability of G causes also the unidentifiability of γ. To determine γ uniquely one has to imply some assumptions that can not be tested. These assumptions may be presented as assumptions on the structure of G but they may also be represented in the form $H\gamma = 0$, where ranks satisfy (4.9).

The restrictions $H\gamma = 0$ are preferable since they are less restrictive – namely there exist several covariance matrices G yielding exactly the same predictors $\hat{\gamma}$. Choosing an exact form for the covariance matrix among those covariance matrices (structures) yielding identical prediction results may be unnecessary in practical applications.

As in the case of fixed effects, the correctness of implied restrictions cannot be assessed using only the observed data.

There are some aspects of Theorem 4.1 which should be mentioned. First, one can consider matrices G and R as some functions of unknown parameters $\nu_1, \ldots, \nu_k$. Then the matrix functions

$$V = ZG(\nu_1, \ldots, \nu_k)Z^T + R(\nu_1, \ldots, \nu_k)$$

and

$$V(G^*) = ZG^*(\nu_1, \ldots, \nu_k)Z^T + R(\nu_1, \ldots, \nu_k),$$

where

$$G^*(\nu_1, \ldots, \nu_k) = P_{Z^T,A}^T G(\nu_1, \ldots, \nu_k) P_{Z^T,A}, \tag{4.10}$$

are equal. Therefore, replacing G with G^* does not change any function which depends on G via the covariance matrix V only. For example, if Y

is normally distributed, then the likelihood function of Y depends on the covariance parameters via the matrix function of V only. Hence, the likelihood function does not change if one replaces G with G^*. In situations where the covariance parameters are unknown and one uses the likelihood function to derive the ML estimates of covariance parameters, the estimates remain unaffected if one uses G^* instead of G, because the likelihood function does not change.

Under REML, one maximizes the likelihood of a linearly transformed vector KY instead of the likelihood of Y. But the likelihood of transformed vector KY depends on G only via the matrix V (see for example Section 4.1). Hence, replacing G with G^* does not affect the estimates of covariance parameters (if Y is normally distributed and one uses REML or ML method). Therefore, for normally distributed data and matrices G and G^* related by (4.10) the EBLU (estimated BLU) estimators will be equal:

$$X\hat{\beta}(\hat{V}) = X\hat{\beta}(\hat{V}(\hat{G}^*)),$$

and EBLU (estimated BLU) predictors will also be equal

$$Z\hat{\gamma}(\hat{R}, \hat{G}) = Z\hat{\gamma}(\hat{R}, \hat{G}^*).$$

It is worth to note here that the derivation of G^* is not necessarily unique, because there may exist more than one matrix A satisfying the key condition $HAZ^T = 0$.

5. Covariance Matrix Classes preserving prediction results

5.1 Introduction

In Linear Mixed Models theory one is assumed to know the structure of random effects' covariance matrix. It is not always an easy task to determine the correct covariance matrix or correct covariance structure. Even in some relatively simple situations different authors can give contradictory suggestions. Consider, as an example, interactions between a random factor and a fixed factor. Some authors, like Searle (1971), and SAS software (SAS, 1999) use unconstrained parameter model, suggesting that these interaction terms should be treated as independent random variables. Others, like Sheffe (1959) and Neter, Wasserman & Kutner (1990) suggest that the interaction terms should be restricted to sum zero over the levels of fixed factor. These two approaches lead to quite different structures for G, if the mixed model considered contains these type of interactions. The problem has been recently addressed by Voss (1999); Wolfinger & Stroup (2000); Hinkelmann (2000).

When treating complex practical situations it is often even harder to choose a correct covariance structure. Therefore, it may be of interest to know, which different covariance structures can lead to the same answers. This question is considered in great detail for estimating fixed effects. For example Puntanen & Styan (1989) give an impressive list of conditions when Ordinary Least Squares Estimator — OLSE — is equal to the BLUE. Harville (1997) gives a condition when two different covariance (weight) matrices lead to the same estimators for fixed effects. However, question when two different covariance structures lead to the same prediction results for linear combinations involving both random and fixed effects has remained unanswered. In this chapter we intend to shed some light to this particular problem.

In the following we consider two mixed models differing from each other only by having different variance matrices:

$$V_1 = R_1 + ZG_1Z^T \text{ and } V_2 = R_2 + ZG_2Z^T.$$

The main purpose of this chapter is to understand the conditions under which these two different models lead to identical prediction (and estimation) results. The derivation and presentation of the results follows largely the idea used by author in Möls (2003).

5.2 Derivation

First we need three intermediate results, which are presented in the following lemmas.

Lemma 5.1. *The estimators*

$$\begin{aligned} \hat{\beta}(V_1) &= (X^T {V_1}^{-1} X)^{-1} X^T {V_1}^{-1} Y \\ \hat{\beta}(V_2) &= (X^T {V_2}^{-1} X)^{-1} X^T {V_2}^{-1} Y \end{aligned}$$

are equal if and only if

$$\mathcal{C}(V_1^{-1}X) = \mathcal{C}(V_2^{-1}X).$$

Proof. See Harville (1997), page 265–266. □

Lemma 5.2. $\mathcal{C}(V_1 - V_2) \subset \mathcal{C}(X) \Rightarrow \mathcal{C}(V_1^{-1}X) = \mathcal{C}(V_2^{-1}X)$.

Proof.

$$\begin{aligned} \mathcal{C}(V_1 - V_2) \subset \mathcal{C}(X) &\Leftrightarrow \exists M \; V_1 - V_2 = XM \\ &\Leftrightarrow \exists M \; I - V_1^{-1}V_2 = V_1^{-1}XM \\ &\Leftrightarrow \exists M \; V_2^{-1} - V_1^{-1} = V_1^{-1}XMV_2^{-1} \\ &\Rightarrow \exists M \; V_2^{-1}X - V_1^{-1}X = V_1^{-1}XMV_2^{-1}X \\ &\Leftrightarrow \exists M \; V_2^{-1}X = V_1^{-1}X(I + MV_2^{-1}X) \\ &\Leftrightarrow \mathcal{C}(V_2^{-1}X) \subset \mathcal{C}(V_1^{-1}X). \end{aligned}$$

Using similar reasoning one can also prove that $\mathcal{C}(V_1^{-1}X) \subset \mathcal{C}(V_2^{-1}X)$ and, hence,

$$\mathcal{C}(V_2^{-1}X) = \mathcal{C}(V_1^{-1}X).$$

□

Lemma 5.3. $\mathcal{C}(V_2 - V_1) \subset \mathcal{C}(X) \Leftrightarrow \mathcal{C}(V_1^{-1} - V_2^{-1}) \subset \mathcal{C}(V_1^{-1}X)$.

Proof.

$$\begin{aligned}
\mathcal{C}(V_2 - V_1) \subset \mathcal{C}(X) &\Leftrightarrow \exists M_1 :\ V_2 - V_1 = XM_1 \\
&\Leftrightarrow \exists M_1 :\ I - V_1V_2^{-1} = XM_1V_2^{-1} \\
&\Leftrightarrow \exists M_1 :\ V_1^{-1} - V_2^{-1} = V_1^{-1}XM_1V_2^{-1} \\
&\Leftrightarrow \mathcal{C}(V_1^{-1} - V_2^{-1}) \subset \mathcal{C}(V_1^{-1}X).
\end{aligned}$$

□

Now we are sufficiently equipped with technical results to prove a major theorem concerning predictors of random effects.

Theorem 5.1. *If*

$$\mathcal{C}(Z(G_1 - G_2)Z^T) \subset \mathcal{C}(X) \tag{5.1}$$

and

$$\mathcal{C}(R_1 - R_2) \subset \mathcal{C}(X), \tag{5.2}$$

then predictors

$$Z\hat{\gamma}(V_1) = ZG_1Z^TV_1^{-1}(Y - X\hat{\beta}(V_1))$$

and

$$Z\hat{\gamma}(V_2) = ZG_2Z^TV_2^{-1}(Y - X\hat{\beta}(V_2))$$

are equal.

Proof. The difference $V_1 - V_2$ equals

$$V_1 - V_2 = Z(G_1 - G_2)Z^T + (R_1 - R_2).$$

From (5.1) and (5.2) follows $\mathcal{C}(V_1 - V_2) \subset \mathcal{C}(X)$. Hence, the requirements of Lemma 5.2 are satisfied and therefore the equality $\mathcal{C}(V_1^{-1}X) = \mathcal{C}(V_2^{-1}X)$ holds. This allows us to apply Lemma 5.1 to derive

$$X\hat{\beta}(V_1) = X\hat{\beta}(V_2).$$

The common estimator for $X\beta$ will be denoted by $X\hat{\beta}$.

From $\mathcal{C}(V_1 - V_2) \subset \mathcal{C}(X)$ and Lemma 5.3 follows $\mathcal{C}(V_1^{-1} - V_2^{-1}) \subset \mathcal{C}(V_1^{-1}X)$ and, hence, there exists a matrix M for which

$$V_1^{-1} - V_2^{-1} = V_1^{-1}XM \tag{5.3}$$

holds. Because V_1^{-1} and V_2^{-1} are symmetric, their difference is also symmetric and therefore (5.3) is equivalent to

$$V_1^{-1} - V_2^{-1} = M^T X^T V_1^{-1}$$

or, equivalently,

$$V_2^{-1} = V_1^{-1} - M^T X^T V_1^{-1}. \tag{5.4}$$

Now consider the difference

$$\begin{aligned} & Z\hat{\gamma}(V_1) - Z\hat{\gamma}(V_2) = \\ & \quad = \quad (ZG_1Z^TV_1^{-1} - ZG_2Z^TV_2^{-1})(Y - X\hat{\beta}) \\ & \quad \overset{(5.4)}{=} \quad (ZG_1Z^TV_1^{-1} - ZG_2Z^TV_1^{-1} + ZG_2Z^TM^TX^TV_1^{-1})(Y - X\hat{\beta}) \\ & \quad = \quad (Z(G_1 - G_2)Z^TV_1^{-1} + ZG_2Z^TM^TX^TV_1^{-1})(Y - X\hat{\beta}). \end{aligned} \tag{5.5}$$

From (5.1) follows the existence of a matrix M_2 satisfying

$$Z(G_1 - G_2)Z^T = XM_2.$$

Because G_1 and G_2 are symmetric

$$Z(G_1 - G_2)Z^T = M_2^T X^T.$$

Therefore we can write (5.5) as

$$\begin{aligned} & Z\hat{\gamma}(V_1) - Z\hat{\gamma}(V_2) = \\ & \quad = \quad (M_2^TX^TV_1^{-1} + ZG_2Z^TM^TX^TV_1^{-1})(Y - X\hat{\beta}) \\ & \quad = \quad M_3X^TV_1^{-1}(Y - X\hat{\beta}) \end{aligned} \tag{5.6}$$

where $M_3 = M_2^T + ZG_2Z^TM^T$. But $Y - X\hat{\beta} = P_{X^\perp,V_1}Y$ and from Proposition 1.25 follows

$$X^TV_1^{-1}P_{X^\perp,V_1} = 0.$$

Hence, (5.6) is also zero and $Z\hat{\gamma}(V_1) = Z\hat{\gamma}(V_2)$. □

Note that if $R_1 = R_2$ then also (5.2) holds.

5.3 Examples

Example 5.1. Consider a model where column of 1's belongs to the subspace $\mathcal{C}(X)$, for example a model with intercept included. Denote with τ random variables corresponding to a random factor and with ω the remaining random effects parameters, which are supposed to be independent of τ, $\omega \perp \tau$. Then the choices of $\mathrm{Var}(\tau) = I\sigma_\tau^2$ and $\mathrm{Var}(\tau) = (I - \frac{1}{k}J)\sigma_\tau^2$, where k is the length of τ (number of factor levels), will lead to identical estimation and prediction results for predictable linear combinations of parameters.

This follows from the next argument. The random effects covariance matrices for the two models considered are

$$G_1 = \mathrm{Var}\begin{bmatrix}\tau\\ \omega\end{bmatrix} = \begin{bmatrix} I\sigma_\tau^2 & 0\\ 0 & D\end{bmatrix}$$

and

$$G_2 = \mathrm{Var}\begin{bmatrix}\tau\\ \omega\end{bmatrix} = \begin{bmatrix} (I - \frac{1}{k}J)\sigma_\tau^2 & 0\\ 0 & D\end{bmatrix}.$$

The design matrix of random effects can also be separated into two parts corresponding to τ and ω : $Z = (Z_\tau | Z_\omega)$. We will assume here the design matrix Z_τ to have exactly one 1 in every row (the position of the 1 indicates the factor level). Then

$$\begin{aligned} Z(G_1 - G_2)Z^T &= (Z_\tau|Z_\omega)\begin{bmatrix}\frac{1}{k}J_{k\times k}\sigma_\tau^2 & 0\\ 0 & 0\end{bmatrix}(Z_\tau|Z_\omega)^T\\ &= \frac{1}{k}J_{n\times n}\sigma_\tau^2. \end{aligned}$$

Because column of 1's belongs to the subspace $\mathcal{C}(X)$ and $R_1 = R_2$, the equality of estimated fixed effects parameters and predicted random effects follows from Theorem 5.1.

Remark. If $cov(\tau, \omega) \neq 0$ then $\mathrm{Var}(\tau) = I\sigma_\tau^2$ and $\mathrm{Var}(\tau) = (I - \frac{1}{k}J)\sigma_\tau^2$ will still lead to the same result as long as matrices G_1 and G_2 remain valid covariance matrices, e.g. they are non-negative definite.

Example 5.2. Let random factor – fixed factor interaction terms $\xi = (\xi_1, \ldots, \xi_{k'})^T$ correspond to the fixed factor level represented in the design matrix with column X_c. Denote by Z_ξ the columns from design matrix Z corresponding to ξ, so that $Z = [Z_\xi | \cdots]$. One can change the part of G corresponding to $\tau\beta$ from $I\sigma_\xi^2$ to $(I - \frac{1}{k'}J)\sigma_\xi^2$ without affecting the estimated and predicted parameter values, eg matrices

$$G_1 = \mathrm{Var}\begin{bmatrix}\xi\\ \cdots\end{bmatrix} = \begin{bmatrix} I\sigma_\xi^2 & 0\\ 0 & D^*\end{bmatrix}$$

and

$$G_2 = \mathrm{Var}\begin{bmatrix} \xi \\ \cdots \end{bmatrix} = \begin{bmatrix} (I - \frac{1}{k'} J_{k' \times k'})\sigma_\xi^2 & 0 \\ 0 & D^* \end{bmatrix}$$

lead to equal predictions (in the sense $Z\hat{\gamma}(G_1) = Z\hat{\gamma}(G_2)$) and estimations. One may notice that this change of covariance matrix restricts interaction terms to sum zero over the particular fixed factor level (with probability 1).

Similar to the discussion of the previous example we have now the following argument. Consider typical model parameterisation. Then matrix Z_ξ contain rows with one 1 or rows filled entirely by zeros. The rows containing ones correspond to the rows of X_c containing 1. Therefore

$$\begin{aligned} Z(G_1 - G_2)Z^T &= [Z_\xi | \cdots] \begin{bmatrix} \frac{1}{k'} J_{k' \times k'} \sigma_\xi^2 & 0 \\ 0 & 0 \end{bmatrix} [Z_{\tau\beta} | \cdots]^T \\ &= \frac{1}{k'} \sigma_\xi^2 Z_\xi J_{k' \times k'} Z_\xi^T = \frac{1}{k'} X_c \sigma_\xi^2. \end{aligned}$$

Thus the assumptions of Theorem 5.1 are satisfied and, consequently, the estimated and predicted parameters for both models are equal (in the sense $X\hat{\beta}(G_1) = X\hat{\beta}(G_2)$ and $Z\hat{\gamma}(G_1) = Z\hat{\gamma}(G_2)$).

Repeating the procedure for all levels of fixed effect we can conclude, that the predictions of linear combinations of model parameters in the form $l_1 X\hat{\beta} + l_2 Z\hat{\gamma}$ will remain the same regardless wether one constrains the fixed effect – random effect interaction terms to sum zero over fixed factor levels or not.

5.4 Comparison of covariance parameter estimates

Usually the covariance matrices R and G are not known, but have to be estimated from data. Here we restrict ourselves to the case, where covariance structure can be presented as a linear matrix function of unknown (covariance) parameters. A useful result will be given in the following lemma.

Lemma 5.4. *Let two mixed models differ only by their covariance structure. Assume these two alternative covariance structures V and V^* can be decomposed as*

$$\begin{aligned} V &= V_1\sigma_1^2 + V_2\sigma_2^2 + \ldots + V_t\sigma_t^2, \\ V^* &= V_1^*\sigma_1^2 + V_2^*\sigma_2^2 + \ldots + V_t^*\sigma_t^2. \end{aligned}$$

If

$$\mathcal{C}(V_1 - V_1^*) \subset \mathcal{C}(X), \mathcal{C}(V_2 - V_2^*) \subset \mathcal{C}(X), \ldots, \mathcal{C}(V_t - V_t^*) \subset \mathcal{C}(X) \quad (5.7)$$

and covariance parameters $\sigma_1^2, \sigma_2^2, \ldots, \sigma_t^2$ *can be uniquely estimated with REML for at least one model, then the REML estimates* $\hat{\sigma}_1^2, \hat{\sigma}_2^2, \ldots, \hat{\sigma}_t^2$ *of covariance parameters from both models are equal given the same observed data vector* Y*.*

Proof. The REML estimation method maximizes the likelihood of transformed data vector KY, where $KX = 0$ (see Chapter 3 for details). If the estimation in based on the covariance structure V, then REML assumes the distribution of KY to be

$$KY \sim \mathcal{N}(0, KVK^T).$$

From (5.7) follows the existence of matrices $M_1, M_2, \ldots, M_t$ such that

$$V_1 = V_1^* - XM_1, \ldots, V_t = V_t^* - XM_t.$$

Therefore, KVK^T can be written as

$$\begin{aligned} KVK^T &= K(V_1\sigma_1^2 + \ldots + V_t\sigma_t^2)K^T \\ &= K(V_1^*\sigma_1^2 - XM_1\sigma_1^2 + \ldots + V_t^*\sigma_t^2 - XM_t\sigma_t^2)K^T \\ &= K(V_1^*\sigma_1^2 + \ldots + V_t^*\sigma_t^2)K^T - K(XM_1\sigma_1^2 - XM_t\sigma_t^2)K^T \\ &= KV^*K^T - 0. \end{aligned}$$

Hence, the likelihood maximized by REML, is the same regardless which covariance structure, V or V^*, one uses. Therefore covariance parameter estimates obtained by REML are equal as long as the likelihood function has unique maximum. □

It may be worth to note that if the assumptions of Lemma 5.4 are satisfied, then also the assumptions of Theorem 5.1 are satisfied.

However, maximum likelihood estimates of covariance parameters obtained using V and V^* may differ from each other considerably, as can be seen in the following small simulation study.

5.4.1 Simulation results

To investigate the effect of changing structure of the assumed (working) covariance matrix, a simulation study was carried out. 10000 datasets were

Table 5.1: Simulation results ($\sigma^2_\gamma = 9$, $\sigma^2_\varepsilon = 1$).

Form of G	Method	E(MSE)	$\mathrm{E}\hat{\sigma}^2_\varepsilon$	$\mathrm{E}\hat{\sigma}^2_\gamma$
$I\sigma^2_\gamma$	ML	0.892 ± 0.002	0.999 ± 0.002	7.24 ± 0.05
$(I - \frac{1}{5}J)\sigma^2_\gamma$	ML	0.890 ± 0.002	0.975 ± 0.002	9.08 ± 0.07
$I\sigma^2_\gamma$	REML	0.890 ± 0.002	0.999 ± 0.002	9.07 ± 0.07
$(I - \frac{1}{5}J)\sigma^2_\gamma$	REML	0.890 ± 0.002	0.999 ± 0.002	9.07 ± 0.07

generated using a model with intercept and one random factor γ with 5 levels drawn from normal distribution, $G = \sigma^2_\gamma I_5, \sigma^2_\gamma = 9$. On each random factor level, 5 repetitions were generated, $R = \sigma^2_\varepsilon I_{25}, \sigma^2_\varepsilon = 1$.
Datasets were analyzed using both ML and REML estimation methods. For each method, G was considered in two forms: $I\sigma^2_\gamma$ and $(I - \frac{1}{5}J)\sigma^2_\gamma$. Simulations were performed with SAS version 8.

One may notice that the REML estimates are equal for the two forms for G (as they should in the light of Lemma 5.4), but the maximum likelihood estimates are different.

5.5 Sampling variability of estimates and predictions

Consider two mixed models, which differ only by their random effects' covariance matrices G_1 and G_2. We show, that if these two covariance matrices G_1 and G_2 satisfy the condition $\mathcal{C}(Z(G_1 - G_2)Z^T) \subset \mathcal{C}(X)$, then for a relatively wide class of linear combinations of model parameters the sampling variability will remain the same. First we prove some preliminary results.

Lemma 5.5. *The MSE of combined vector* $((X\hat{\beta} - X\beta)^T, (Z\hat{\gamma} - Z\gamma)^T)^T$ *is*

$$Var \begin{bmatrix} X\hat{\beta} - X\beta \\ Z\hat{\gamma} - Z\gamma \end{bmatrix} = \begin{bmatrix} D^*_{11} & D^*_{12} \\ {D^*_{12}}^T & D^*_{22} \end{bmatrix},$$

where

$$\begin{aligned} D^*_{11} &= X(X^TV^{-1}X)^-X^T; \\ D^*_{12} &= -X(X^TV^{-1}X)^-X^TV^{-1}ZGZ^T; \\ D^*_{22} &= ZGZ^TV^{-1}R - (D^*_{12})^TV^{-1}ZGZ^T. \end{aligned}$$

Proof. Notice, that this result resembles closely the form of MSE of combined parameter vector given in standard textbooks, like Searle *et al* (1992). The main difference is, that the proofs leading to those, similar results assume the matrix G to be nonsingular. However, most steps can be easily repeated even if G is singular. Our proof will be based on the results of Harville (1976), which do not make such an assumption. The results of Harville are rewritten using the notation of these thesis in Proposition 2.1. Notice first that the equality

$$V^{-1} = R^{-1} - R^{-1}ZG^{1/2}(I + G^{1/2}Z^TR^{-1}ZG^{1/2})^{-1}G^{1/2}Z^TR^{-1} \tag{5.8}$$

holds. The formula (5.8) can be easily proven by multiplying the right hand side by $(R+ZGZ^T)$. A more general result of this kind, extending (5.8) to the case of singular V and R is also given by Harville (1976). Multiplying (5.8) by GZ^T gives us

$$\begin{aligned} GZ^TV^{-1} &= GZ^TR^{-1} - \\ &\quad -G^{1/2}G^{1/2}Z^TR^{-1}ZG^{1/2}(I + G^{1/2}Z^TR^{-1}ZG^{1/2})^{-1}G^{1/2}Z^TR^{-1}. \end{aligned} \tag{5.9}$$

After adding and subtracting $G^{1/2}(I+G^{1/2}Z^TR^{-1}ZG^{1/2})^{-1}G^{1/2}Z^TR^{-1}$ to (5.9) we get

$$\begin{aligned} GZ^TV^{-1} &= GZ^TR^{-1} \\ &\quad -G^{1/2}(I + G^{1/2}Z^TR^{-1}ZG^{1/2})(I + G^{1/2}Z^TR^{-1}ZG^{1/2})^{-1}G^{1/2}Z^TR^{-1} \\ &\quad +G^{1/2}(I + G^{1/2}Z^TR^{-1}ZG^{1/2})^{-1}G^{1/2}Z^TR^{-1} \\ &= GZ^TR^{-1} - GZ^TR^{-1} + G^{1/2}(I + G^{1/2}Z^TR^{-1}ZG^{1/2})^{-1}G^{1/2}Z^TR^{-1} \\ &= G^{1/2}(I + G^{1/2}Z^TR^{-1}ZG^{1/2})^{-1}G^{1/2}Z^TR^{-1}. \end{aligned} \tag{5.10}$$

Equality (5.10), together with Proposition 2.1, yields the desired result. □

Definition 5.1. *We say that linear combination of fixed and random effects is robustly predictable if it can be expressed in the form $l(X\beta+Z\gamma)$ for some vector l.*

Lemma 5.6. *The MSE of robustly predictable linear combination is*

$$\begin{aligned} &Var(l(X\hat{\beta} + Z\hat{\gamma}) - l(X\beta + Z\gamma)) = \\ &\quad l(P_{X,V}R + ZGZ^TV^{-1}P_{X^\perp,V}R)l^T. \end{aligned} \tag{5.11}$$

Proof.

$$\begin{aligned}
&\mathrm{Var}(X\hat{\beta} + Z\hat{\gamma} - X\beta - Z\gamma) \\
&\quad = \quad \mathrm{Var}\left([\, X|Z \,] \begin{bmatrix} \hat{\beta} - \beta \\ \hat{\gamma} - \gamma \end{bmatrix} \right) \\
&\quad \overset{Lemma\ 5.5}{=} \quad X(X^T V^{-1} X)^{-1} X^T - \\
&\qquad\qquad -X(X^T V^{-1} X)^{-} X^T V^{-1} Z G Z^T - \\
&\qquad\qquad -ZGZ^T V^{-1} X (X^T V^{-1} X)^{-} X^T + ZGZ^T V^{-1} R + \\
&\qquad\qquad +ZGZ^T V^{-1} X (X^T V^{-1} X)^{-} X^T V^{-1} Z G Z^T. \qquad (5.12)
\end{aligned}$$

But

$$\begin{aligned}
&X(X^T V^{-1} X)^{-1} X^T - X(X^T V^{-1} X)^{-1} X^T V^{-1} Z G Z^T \\
&\quad = \quad X(X^T V^{-1} X)^{-1} X^T (I - V^{-1} Z G Z^T) \\
&\quad = \quad X(X^T V^{-1} X)^{-1} X^T (I - V^{-1} Z G Z^T - V^{-1} R + V^{-1} R) \\
&\quad = \quad X(X^T V^{-1} X)^{-1} X^T (I - V^{-1}(R + ZGZ^T) + V^{-1} R) \\
&\quad = \quad X(X^T V^{-1} X)^{-1} X^T V^{-1} R
\end{aligned}$$

and, hence, (5.12) is equal to

$$\begin{aligned}
&\mathrm{Var}(l(X\hat{\beta} + Z\hat{\gamma}) - l(X\beta + Z\gamma)) = \\
&\quad = \quad X(X^T V^{-1} X)^{-1} X^T V^{-1} R - \\
&\qquad -ZGZ^T V^{-1} X (X^T V^{-1} X)^{-1} X^T V^{-1} R + ZGZ^T V^{-1} R \\
&\quad = \quad X(X^T V^{-1} X)^{-1} X^T V^{-1} R + \\
&\qquad +ZGZ^T V^{-1} (I - X(X^T V^{-1} X)^{-1} X^T V^{-1}) R \\
&\quad = \quad P_{X,V} R + ZGZ^T V^{-1} P_{X^\perp,V} R. \qquad (5.13)
\end{aligned}$$

If

$$\mathrm{Var}(X\hat{\beta} + Z\hat{\gamma} - X\beta - Z\gamma) = P_{X,V} R + ZGZ^T V^{-1} P_{X^\perp,V} R,$$

then

$$\mathrm{Var}(l(X\hat{\beta} + Z\hat{\gamma} - X\beta - Z\gamma)) = l(P_{X,V} R + ZGZ^T V^{-1} P_{X^\perp,V} R) l^T,$$

which completes the proof. □

Theorem 5.2. *Consider two different mixed models differing from each other only by their covariance matrices*

$$\begin{aligned}
V_1 &= ZG_1 Z^T + R, \qquad (5.14) \\
V_2 &= ZG_2 Z^T + R. \qquad (5.15)
\end{aligned}$$

If

$$\mathcal{C}(Z(G_1 - G_2)Z^T) \subset \mathcal{C}(X) \tag{5.16}$$

then for any robustly predictable linear combination its expected MSE,

$$Var(l(X\hat{\beta} + Z\hat{\gamma}) - l(X\beta + Z\gamma)),$$

will be equal regardless of which mixed model, based on covariance matrix V_1 or V_2, is used to calculate the MSE.

Proof. It is sufficient to show that the quantity $\mathrm{Var}(X\hat{\beta}+Z\hat{\gamma}-X\beta-Z\gamma)$ will be the same for two covariance matrices satisfying the requirements of the theorem. Because of Lemma 5.6 (and equivalences $X\hat{\beta} = P_{X,V}Y$ and $Z\hat{\gamma} = ZGZ^TV^{-1}P_{X^\perp,V}Y$) we can write

$$\mathrm{Var}(P_{X,V}Y+ZGZ^TV^{-1}P_{X^\perp,V}Y-X\beta-Z\gamma) = P_{X,V}R+ZGZ^TV^{-1}P_{X^\perp,V}R.$$

If (5.16) holds, then according to the Theorem 5.1, $P_{X,V_1} = P_{X,V_2}$ and $ZGZ^TV_1^{-1}P_{X^\perp,V_1} = ZGZ^TV_2^{-1}P_{X^\perp,V_2}$. Therefore, also

$$P_{X,V_1}R + ZG_1Z^TV_1^{-1}P_{X^\perp,V_1}R = P_{X,V_1}R + ZG_2Z^TV_2^{-1}P_{X^\perp,V_2}R$$

and, hence, the MSE of a robustly predictable linear combination calculated by using V_1 is the same as the one obtained by using V_2. □

5.6 Discussion

Choosing the correct covariance structure for the linear mixed model may sometimes be a difficult task. It may be especially difficult to decide how to handle the interactions between random and fixed factors — different experts have suggested different covariance structures for these interactions (they are sometimes considered to be independent and sometimes restricted to sum zero over the levels of fixed factor). Theory has not offered any reasonable way to make choice between these two approaches. Major statistical software packages like SAS have favored the independence assumption — probably because solutions based on it are easier to implement. However, it is encouraging to know, that there exist relatively general situations were both approaches give equal results — like predicting or making inference on robustly predictable linear combinations. Unfortunately this equality can not be carried over to the wide inference space. Generally the results remain depend on our subjective (and often untestable) choice of the covariance matrix structure of random factors.

Bibliography

[1] Cox, D. R., Reid, N. (1987). Parameter Orthogonality and Approximate Conditional Inference. *Journal of the Royal Statistical Society B*, **49**, 1–39.

[2] Frisk, T., Bilaletdin, Ä., Kaipanen, H., Malve, O., Möls, M. (1999). Modelling phytoplankton dynamics of the eutrophic Lake Võrtsjärv, Estonia. *Hydrobiologia*, **414**, 59–69.

[3] Harville, D. A. (1997). *Matrix algebra from a statistician's perspective.* Springer-Verlag, Berlin, Heidelberg, New York.

[4] Harville, D. A. (1976). Extension of the Gauss-Markov Theorem to Include the Estimation of Random Effects. *The Annals of Statistics*, **4**, 384–395.

[5] Henderson, C. R. (1963). Selection index and expected genetic advance. In: *Statistical Genetics and Plant Breeding* (Hanson, W.D. and Robinson, A.F., eds.) NAS-NRC, Washington, 141–163.

[6] Hinkelmann, K. (2000). Resolving the Mixed Models Controversy, Comments. *The American Statistician*, **54**, 228.

[7] Khuri, A. I., Mathew, T., Sinha, B. K. (1998) *Statistical Tests for Mixed Linear Models*. Wiley, New York.

[8] Lahiri, P., Rao, J. N. K. (1995). Robust Estimation of Mean Squared Error of Small Area Estimators. *Journal of the American Statistical Association*, **90**, 758–766.

[9] McLean, R. A., Sanders, W. L., Stroup, W. W. (1991). A unified approach to mixed linear models. *The American Statistician*, **45**, 54–64.

[10] Möls, M. (2004). On Reparameterization of Random Effects in Linear Mixed Models. *Acta et Commentationes Universitatis Tartuensis de Mathematica*, (accepted).

[11] Möls, M. (2003). Constraints on Random Effects and Mixed Linear Model Predictions. *Acta Applicandae Mathematicae*, **79**, 17–23.

[12] Möls M., Nõges T., Nõges P. (2001). Võrtsjärve ökoloogia modelleerimisest. *Eesti Statistikaseltsi Teabevihik*, **11**, 44–53.

[13] Neter, J., Wasserman, W., Kutner, M. H. (1990). *Applied linear statistical models. Third Edition.*, Irwin, Burr Ridge.

[14] Patterson, H. D., Thompson, R. (1971). Recovery of inter-block information when block sizes are unequal. *Biometrika*, **58**, 545–555.

[15] Prasad, N. G. N., Rao, J. N. K. (1990). The Estimation of Mean Squared Error of Small-Area Estimators. *Journal of the American Statistical Association*, **85**, 163–171.

[16] Puntanen, S., Styan, P. H. (1987). The equality of the ordinary least squares estimator and the best linear unbiased estimator. *The American Statistican*, **43**, 153–164.

[17] Rao, C. R., Rao, M. B. (1998). *Matrix Algebra and Its Applications to Statistics and Econometrics.* Singapore, World Scientic.

[18] Robinson, G. K. (1991). That BLUP is a good thing: The estimation of random efects. *Statistical Science*, **6**, 15–51.

[19] SAS Institute Inc. (1999). *SAS OnlineDoc, Version 8.* Cary, NC, SAS Institute Inc.

[20] Scheffe., M. (1959). *The Analysis of Variance.* Wiley, New York.

[21] Searle, S. R. (1971). *Linear Models.* Wiley, New York.

[22] Searle, S. R., Casella, G., McCulloch, C. E. (1992). *Variance Components.* Wiley, New York.

[23] Searle, S. R. (1994). Extending some results and proofs for the singular linear model. *Linear Algebra and its Applications*, **210**, 139–151.

[24] Searle, S. R. (1997a). The Matrix Handling of BLUE and BLUP in the Mixed Linear Model. *Linear Algebra and its Applications*, **264**, 291–311.

[25] Searle, S. R. (1997b). Built-In Restrictions on Best Linear Unbiased Predictors (BLUP) of Random Effects in Mixed Models. *The American Statistician*, **51**, 19–21.

[26] Smyth, G. K., Verbyla, A. P. (1996). A conditional likelihood approach to REML in Generalized Linear Models *Journal of the Royal Statistical Society B*, **58**, 565–572.

[27] Verbeke, G., Molenberghs, G. (2000). *Linear mixed models for longitudinal data.* Springer, New York.

[28] Voss, D. T. (1999) Resolving the Mixed Models Controversy. *The American Statistican*, **53**, 352–356.

[29] Wolfinger, R., Stroup, W.W. (2000). Resolving the Mixed Models Controversy, Comments. *The American Statistician*, **54**, 228.

ACKNOWLEDGEMENTS

I would like to express my deep gratitude to my supervisor Docent Tõnu Möls for sharing with me his enthusiasm and support.

I am equally indebted to my colleagues and friends from the Institute of Mathematical Statistics for providing a stimulating and pleasant environment in which to learn, work and grow. I am especially grateful to Prof. Tõnu Kollo and Prof. Kalev Pärna for helpful comments and suggestions.

It would have been difficult to complete the work without financial support from the Estonian Ministry of Education and Research (research project TMTMS1776).

I am also grateful to my family for their love, patience and understanding.

Ekvivalentsete prediktoritega lineaarsed segamudelid

Kokkuvõte

Lineaarseid segamudeleid kasutatakse paljudel erialadel — nagu näiteks bioloogias, meditsiinis, sotsioloogias jne. Segamudelite rakendamisel osutub sageli kõige raskemaks ülesandeks kasutatava kovariatsioonistruktuuri valik. Teoreetilised käsitlused on teemat harva puudutanud ja antud vähesed soovitused on olnud mõneti vastuolulised. Praktilisema kallakuga töödes on sageli piirdutud kasutatud kovariatsioonistruktuuri võimalikkuse põhjendamisega, selle ainuõigsus on aga enamasti jäänud tõestamata. Selle tõttu oleks hädasti vaja tulemusi, mis võimaldaks kirjeldada ja paremini mõista kovariatsioonistruktuuri valikul tehtud vigade mõju analüüsi tulemustele. Selles vallas on avaldatud üksikuid töid (näiteks Puntanen & Styan 1989, Harville 1997), kuid mitmed probleemid, nagu näiteks mõju lineaarse segamudeli prognoosidele, on jäänud käsitlemata. Antud töö sisaldab uusi tulemusi, mis võimaldavad paremini mõista kovariatsioonistruktuuri valiku mõju lineaarse segamudeli prognoositulemustele.

Töös kasutatavad abitulemused maatriksalgebrast ja lineaarsete segamudelite teooriast on esitatud peatükkides 1–2.

Peatükis 3 näidatakse, et lineaarse segamudeli parameetrite hindamisel kasutatav populaarne tehnika — REML — on vaadeldav kui vale kovariatsioonistruktuuri kasutamine suurima tõepära meetodi rakendamisel.

Peatükis 4 näidatakse, et tüüpilise prognoosiülesande lahendamiseks pole kovariatsioonistruktuuri ühene määratlemine hädavajalik. Ühe alternatiivina pakutakse välja võimalus osad kovariatsioonistruktuuri kohta käivad eeldused asendada juhuslike faktorite reparametriseerimistingimustega.

Ühe kovariatsioonistruktuuri asendamine teisega võib, kuid ei pruugi, muuta lineaarse segamudeli prognoositulemusi. Peatükis 5 on antud suhteliselt kergesti rakendatav tingimus kontrollimaks, kas kaks kovariatsioonistruktuuri viivad samade prognoositulemusteni.

CURRICULUM VITAE
Märt Möls

Born:	February 04, 1972 in Tartu, Estonia
Martial status:	married, 2 children
Mailing address:	Institute of Mathematical Statistics, J. Liivi 2, 50409 Tartu, Estonia
Contacts:	tel: +372 7 376 400, fax: +372 7 375 898, e-mail: Mart.Mols@ut.ee

Education:

1997–2002	doctoral student at the Institute of Mathematical Statistics, Faculty of Mathematics, University of Tartu
1995–1996	MSc studies in Biostatistics, Limburgs Universitair Centrum, Belgium. MSc degree conferred in October, 1996.
1990–1995	BSc studies in Mathematical Statistics, University of Tartu, diploma issued in 1996

Professional employment:

1991–1992	assistant in Institute of Zoology and Botany
1992–1993	programmer in Tartu City Council
1993–1996	assistant in Tartu University
1996–2000	statistical consultant in Resta
1997–1998	researcher in Institute of Zoology and Botany
2000–2002	researcher in Tartu University
2002–2004	lecturer in Tartu University

Scientific work:

Main research field: linear mixed models and their applications in life sciences.

ELULOOKIRJELDUS
Märt Möls

Sünniaeg ja koht:	04. veebruar 1972, Tartu, Eesti.
Perekonnaseis:	abielus, 2 last
Aadress:	Matemaatilise Statistika Instituut, J. Liivi 2, 50409 Tartu.
Kontaktandmed:	tel: +372 7 376 400, fax: +372 7 375 898, e-post: Mart.Mols@ut.ee

Haridus:

1997–2002	doktoriõpe Tartu Ülikooli Matemaatikateaduskonnas, matemaatilise statistika erialal.
1995–1996	magistriõpe Limburgi Ülikoolis, Belgias. Teadusmagistri kraad biostatistika alal omistatud oktoobris 1996.
1990–1995	bakalaureuseõpe Tartu Ülikooli Matemaatikateaduskonnas. Diplom välja antud 1996.a.

Teenistuskäik:

1991–1992	laborant Zooloogia ja Botaanika Instituudis
1992–1993	programmeerija Tartu Linnavalitsuses
1993–1996	laborant Tartu Ülikoolis
1996–2000	statistikakonsultant AS Restas
1997–1998	teadur Zooloogia ja Botaanika Instituudis
2000–2002	teadur Tartu Ülikoolis
2002–2004	lektor Tartu Ülikoolis

Teadustegevus:

Segamudelid ja nende rakendamine eluteadustes.