Adilson de Jesus Martins da Silva

Master of Science

Variance Components Estimation In Mixed Linear Models

Dissertação para obtenção do Grau de
Doutor em Estatística e Gestão de Risco

Orientador: Miguel dos Santos Fonseca, Investigador, Universidade Nova de Lisboa

Co-orientadores: Dário Ferreira, Prof. Auxiliar, Universidade de Beira Interior
Francisco Carvalho, Prof. Adjunto, Instituto Politécnico de Tomar

Júri:

Presidente: Prof. Doutor Fernando José Pires Santana
Arguentes: Prof. Doutor Martin Singull
Prof. Doutora Célia Maria Pinto Nunes
Vogais: Prof. Doutor Kamil Feridun Turkman
Prof. Doutora Maria Manuela Costa Neves Figueiredo
Prof. Doutora Isabel Cristina Maciel Natário
Doutor Miguel dos Santos Fonseca

FACULDADE DE
CIÊNCIAS E TECNOLOGIA
UNIVERSIDADE **NOVA** DE LISBOA

February, 2017

Variance Components Estimation In Mixed Linear Models

To my son Evarist, my inspiration.

ACKNOWLEDGEMENTS

ABSTRACT

This work aim to introduce a new method of estimating the variance components in mixed linear models. The approach will be done firstly for models with three variance components and secondly attention will be devoted to general case of models with an arbitrary number of variance components.

In our approach, we construct and apply a finite sequence of orthogonal transformations, here named sub - diagonalizations, to the covariance structure of the mixed linear model producing a set of Gauss-Markov sub-models which will be used to create pooled estimators for the variance components. Indeed, in order to reduce the bias, we apply the sub - diagonalizations to its correspondent restricted model, that is its projection onto the orthogonal subspace generated by the columns of its mean design matrix. Thus, the Gauss - Markov sub-models will be centered. The produced estimator will be called *Sub-D*.

Finally, the numerical behavior of the proposed estimator is examined for the case of models with three variance components, comparing its performance to the ones obtained with the REML and ANOVA estimators. Numerical results show that *Sub-D* produces reasonable and comparable estimates, some times slightly better than those obtained with REML and mostly better than those obtained with ANOVA.

Due to the correlation between the sub-models, the estimated variability of the variability of *Sub-D* will be slightly bigger than the one of the REML estimator. In attempt to solve this problem a new estimator will be introduced.

Keywords: Mixed Linear Models; Sub-diagonalizations; Variance components; Sub-D.

Resumo

Este trabalho pretende introduzir um novo método de estimação das componentes da variância em modelos lineares mistos. Numa primeira instância, aborda-se a estimação em modelos com três componentes da variância. Seguidamente, foca-se no caso geral: estimação em modelos com um número arbitrário de componentes da variância.

Na nossa abordagem, construimos e aplicamos uma sequência finita de transformações ortogonais - aqui denominadas sub-diagonalizadoras - à estrutura da covariância do modelo, produzindo assim um conjunto de sub - modelos de Gauss-Markov que serão usados para criar estimadores agrupados. Na verdade, com o intuito de reduzir o viés, aplicamos as sub-diagonalizadoras ao modelo restrito correspondente, isto é, à projeção do complemento ortogonal no subspaço gerado pelas colunas da sua matriz do delineamento para a esperança (parte dos efeitos fixos), pelo que os sub-modelos de Gauss-Markov acima referidos terão média nula. O estimador resultante será chamado de *Sub-D*.

Finalmente, examina-se o desempenho numérico do estimador proposto para o caso do modelo com três componentes da variância, comparando-o com o dos estimadores REML e ANOVA. Os resultados obtidos mostram que o nosso estimador (*Sub-D*) produz estimativas razovelmente comparáveis, sendo, em alguns casos, ligeiramente melhores que os resultados obtidos com o estimador REML e na maioria dos casos melhores que os obtidos com o estimador ANOVA. Contudo, devido à dependência entre os sub-modelos, a variabilidade estimada será ligeiramente maior que a do estimador REML. Na tentativa de ultrapassar esse problema um novo estimador será introduzido.

Palavras-chave: Modelos lineares mistos; Sub-diagonalizadoras; Componentes da variância; Sub-D.

CONTENTS

LIST OF NOTATIONS AND ABBREVIATIONS

$\mathscr{M}^{n \times m}$	Set of matrices of size $n \times m$
$\mathscr{S}^n$	Set of symmetric matrices with size $n \times n$
$\mathbf{0}_{n,m}$	Null matrix of size $n \times m$
$\mathbf{0}_n$	Null vector of size n
$\mathbf{1}_n$	Vector of 1's of size n
$J_{n,m}$	Matrix of 1's of size $n \times m$
I_n	Identity matrix of size n
X^{-1}	Inverse of matrix X
X^-	Generalized inverse of matrix X
X^+	Moore-Penrose inverse of matrix X
$X^\top$	Transpose of matrix X
$r(X)$	Rank of matrix X
$tr(X)$	Trace of matrix X
$R(X)$	Vector space spanned by the column vectors of matrix X
$N(X)$	Kernel of matrix X
$R(X)^\perp$	Orthogonal complement of $R(X)$
$P_{R(X)}$	Projection Matrix onto $R(X)$
$E(y)$	Expectation of the random vector y
$\Sigma(x,y)$	Cross-covariance between the random vectors x and y
$\Sigma(y)$	Variance-covariance of the random vector y
$Var(y)$	Variance of the random variable y
$y \sim (X\mu, \Sigma)$	y is a vector with expectation $X\mu$ and variance-covariance matrix Σ
$y \sim \mathcal{N}(X\mu, \Sigma)$	y is a normal random vector with mean vector $X\mu$ and variance-covariance matrix Σ
ML	Maximum Likelihood
REML	Restricted Maximum Likelihood
ANOVA	Analysis of Variance
OBS	Orthogonal Block Structure(s)
MLM	Mixed Linear Model(s)

INTRODUCTION

Mixed linear models (MLM) have received much attention recently, namely because they constitute an useful tool for modeling repeated measurement data, and, in particular, small sample and longitudinal data (Wallace and Helms [27] developed procedures providing hypothesis tests and confidence intervals for longitudinal data using MLM).

MLM arise due to the necessity of accessing the amount of variation caused by certain sources in statistical designs with fixed effects (see Khuri [34]) for example the amount of variation that are not controlled by the experimenter and those whose the levels are randomly selected from a large population of levels. The variances of such sources of variation, currently refereed to as variance components, have been widely investigated mainly in the last fifty years of the last century (see Khuri and Sahai [35], Searle ([65], [66]), for example), and thanks to the proliferation of research in applied areas such as genetic, animal and plant breeding, statistical process control and industrial quality improvement (see Anderson ([2], [4], [3]), Anderson and Crump [6], Searle [65] for instance) several techniques of estimation for the variances components have been proposed. Among them we highlight the ANOVA and likelihood based method (see Searle at al. [67] and Casella and Berger [14]), as well as those based on orthogonal block structure (OBS) (see Nelder ([57], [58]). Nevertheless, notwithstanding the ANOVA method adapt readily to mixed models with balanced data and save the unbiasedness, it does not adapt in situation with unbalanced data (mostly because it uses computations derived from fixed effect models rather than mixed models). On its turn, the maximum likelihood - based methods provide estimators with several statistical optimal properties such as consistency and asymptotic normality either for models with balanced data, or for those with unbalanced data. For these optimal properties we recommend, for instance, Miller ([46], [47]) and for some details on applications of such methods we suggest Anderson [4] and Hartley and Rao [25]. The OBS based method plays important role in the theory of randomized block designs (see Calinski and Kageyama ([12], [13])).

This work focuses on developing a new method of estimating variance components in MLM. It will be done firstly for models with 3 variances components, and secondly attention will be devoted

to general case of models with an arbitrary number of variance components. In our approach, in order to reduce the bias of our estimator, we will instead consider the orthogonal projection of the normal MLM onto the subspace generated by the columns of its correspondent mean design matrix, that is the restricted model. We construct and apply a finite sequence of orthogonal matrices to the covariance structure of the restricted model thus producing a set of homoscedasticity sub-models and then use that sub-models structure to developing the above announced estimator. For now on the finite sequence of orthogonal matrices will be refereed to as sub-diagonalizations, and the estimators developed here refereed to as *Sub-D* and *Sub-DI*, where *Sub-DI* is found in an attempt to improve the *Sub-D*. Through this work sometimes we may have the need to refer to the underlying deduction method of *Sub-D* and sometimes to the deducted estimator; when so, for the first case we will refer to as *Sub-D* method, whereas for the second one as *Sub-D* estimator.

The firsts three Chapters of this work, Chapters 1, 2, and 3, are devoted to preliminary notions and the literature review. The development of the estimators *Sub-D* and *Sub-DI* are done in Chapter 4, and Chapter 5 is devoted to the numerical application for *Sub-D* and *Sub-DI*. Finally, Chapter 6 is devoted to final comments and proposals for future works.

2

ALGEBRAIC RESULTS

In this chapter we will review the elements of matrix theory needed in the remainder of the thesis, especially in both chapter 3 and 4. The proofs of the results that seem, somehow, instructive will be included. On the other hand, for the proofs of the remainder results we will always include some references. Among them, we highlight Schott [64], Rao and Rao [61], and Rencher and Schaalje [62].

We begin with the presentation of some basic notions on *matrix theory* in section 2.1 and next we present results on *orthogonal basis* and *projection matrices* (section 2.2), followed by a brief presentation and discussion of *diagonalization* of a *symmetric Matrix*. In section 2.3 the *generalized inverse* matrix notions and some important results for the remainder chapter (see section 2.3). Finally, at the last two sections (sections 2.4 and 2.5) discussion of needed notions and results on *Jordan algebra* and on *Kronecker product* of matrices follow.

2.1 Notation And Preliminary Notions On Matrix Theory

Throughout this work we use the capital letter to represent a matrix and, when needed, the lower case letter to represent a vector. Let $\mathscr{M}^{n \times m}$ stands for the set of matrices with n rows and m columns. Thereby, with $A \in \mathscr{M}^{n \times m}$ we mean a matrix whose the dimension is characterized by n rows and m columns, and the element in the row $0 < i \leq n$ and column $0 < j \leq m$, that is, the $(i,j)th$ element, is a scalars or a variable, usually denoted by a_{ij}.

Let

$$A = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1m} \\ a_{21} & a_{22} & \dots & a_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \dots & a_{nm} \end{bmatrix}, \tag{2.1}$$

be a matrix in the standard form. If $m = 1$, A is said to be a *vector*, and denoted by a lower case letter a instead of A. In this case we will write $\mathbb{R}^n$ in place of $\mathscr{M}^{n \times 1}$.

Interchanging the rows and columns of $A \in \mathscr{M}^{n \times m}$ the resulting matrix is said to be the *transpose matrix* of A and is denoted by $A^\top$; that is, $A^\top$ is a matrix whose the element at row i and column j is exactly the element at row j and column i of matrix A. So,

$$A^\top = \begin{bmatrix} a_{11} & a_{21} & \ldots & a_{n1} \\ a_{12} & a_{22} & \ldots & a_{n2} \\ \vdots & \vdots & \ddots & \vdots \\ a_{1m} & a_{2m} & \ldots & a_{nm} \end{bmatrix}.$$

If $n = m$, A is said to be a *square matrix* and when A is a *square matrix* with $a_{ij} = 0$, for $i \neq j$, A is said to be a *diagonal matrix*. Here we denote it by $D(a_{11}, \ldots, a_{nn})$ or just D when there is no risk of misunderstanding. When $a_{ij} = 0$, for all i and j, A is called a *null matrix*. We will represent a null matrix by $\mathbf{0}_{n,m}$.

Throughout this work we will assume the following notation for some especial matrices (including the vectors):

- $\mathbf{0}_n$ denotes a vector in $\mathbb{R}^n$ whose the entries are all equals to 0;

- $J_{n,m}$ denotes a matrix in $\in \mathscr{M}^{n \times m}$ whose the entries are all equals to 1; when $n = m$ it will be denoted J_n;

- $\mathbf{1}_n$ denotes a vector in $\mathbb{R}^n$ whose the entries are all equals to 1.

A *diagonal matrix* A whose the diagonal elements are all equal to one is called *identity matrix*. We will denote it here by I_n, or just I when there is no risk of misunderstanding.

Definition 2.1.1. Let $A \in \mathscr{M}^{n \times n}$; that is, a *square matrix*. A is said to be a *symmetric matrix* if it holds $A^\top = A$. Here we denote the set of all *symmetric matrices* in $\mathscr{M}^{n \times n}$ by $\mathscr{S}^n$.

For what follows it is assumed that the reader is familiarized with sum and product of matrices (if not, see Lay [38] for instance). We will introduce several functions of matrix and discuss a few of them, mainly those with direct implication on the remainder chapters. For this latter ones we will present the notions and the main results. for the remaining ones we recommend Lay [38] or Schott [64], for instance.

One of the matrix function with no direct implication in this work is $|\ |$, the *determinant* of a *square matrix* (see Horn and Johnson [30] for this topic). Given a *square matrix* A, if $|A| \neq 0$, A is said to be a *non-singular matrix*. See Schott [64] or Lay [38] for more explanation.

Another function defined over a *square matrix* is the *trace* function.

Definition 2.1.2. Let $A \in \mathscr{M}^{n \times n}$. The *trace* of A, denoted by $tr(A)$, is defined by

$$tr(A) = \sum_{s=1}^{n} a_{ss};$$

that is, the sum of the diagonal elements of A.

The *trace* function plays an important role on statistic field, with some emphasis, for example, on the distribution of quadratic forms (see Schott [64] or Rencher and Schaalje [62]). Indeed, given a random vector $y \in \mathbb{R}^n$ with mean vector μ and variance-covariance matrix Σ, and a *symmetric matrix $A \in \mathcal{M}^{n \times n}$*, the expectation of the quadratic form $y^\top A y$, denoted by $E(y^\top A y)$, is given by

$$E(y^\top A y) = tr(A\Sigma) + \mu^\top \Sigma \mu. \tag{2.2}$$

If y has finite fourth moment, we have that the variance-covariance matrix of $y^\top A y$, denoted by $\Sigma(y^\top A y)$, is given by

$$\Sigma(y^\top A y) = 2\left(tr([A\Sigma]^2) + 2\mu^\top A\Sigma A\mu\right). \tag{2.3}$$

The following result summarizes a few useful properties of the *trace* function.

Proposition 2.1.1. *Let $A, B \in \mathcal{M}^{n \times n}$ and $\alpha \in \mathbb{R}$. Then*

(a) $tr(A^\top) = tr(A)$;

(b) $tr(A + \alpha B) = tr(A) + tr(\alpha B) = tr(A) + \alpha tr(B)$;

(c) $tr(AB) = tr(BA)$;

(d) $tr(A^\top A) = \mathbf{0}_{n,n}$ *if and only if $A = 0_{n,n}$.*

Proof. For (a) we only have to note that the diagonal element of A are the same as those of $A^\top$. For (b), with δ_{ii}, $i = 1, \ldots, n$, denoting the diagonal elements of $A + \alpha B$ it follows that $\delta_{ii} = a_{ii} + \alpha b_{ii}$, where a_{ii} and b_{ii} denote the the diagonal elements of A and B, respectively. For (c), let $A_{i\bullet}$ and $A_{\bullet j}$ respectively denote the ith row and the jth column of the matrix A. Thus the element c_{ij} of $C = AB$ will be

$$c_{ij} = A_{i\bullet}B_{\bullet j} = \sum_{k=1}^{n} a_{ik} b_{ik}$$

and the element d_{ij} of $D = BA$ will be

$$d_{ij} = B_{i\bullet}A_{\bullet j} = \sum_{k=1}^{n} b_{ik} a_{ik}.$$

$$\begin{aligned} tr(AB) &= tr(C) = \sum_{i=1}^{n} c_{ii} = \sum_{i=1}^{n} A_{i\bullet}B_{\bullet i} = \sum_{i=1}^{n}\sum_{k=1}^{n} a_{ik}b_{ki} = \sum_{k=1}^{n}\sum_{i=1}^{n} b_{ki}a_{ik} = \sum_{k=1}^{n} B_{k\bullet}A_{\bullet k} = \sum_{k=1}^{n} d_{kk} \\ &= tr(D) = tr(BA). \end{aligned}$$

Finally, for (d), the sufficient condition is obviously once $A = \mathbf{0}_{n,n}$ implies $A^\top A = \mathbf{0}_{n,n}$. Now, for the necessary condition, with $E = A^\top$ and nothing that $E_{i\bullet} = A_{\bullet i}$, we will have

$$tr(A^\top A) = \sum_{i=1}^{n} E_{i\bullet}A_{\bullet i} = \sum_{i=1}^{n}\sum_{k=1}^{n} e_{ik}a_{ki} = \sum_{i=1}^{n}\sum_{k=1}^{n} a_{ki}^2. \tag{2.4}$$

Consequently, $tr(A^\top A) = 0$ holds if and only if $a_{ki} = 0$ for all k and all i which means exactly $A = \mathbf{0}_{n,n}$. □

Definition 2.1.3. Let $A \in \mathscr{M}^{n \times n}$ be a *non-singular matrix*. The unique matrix B such that

$$BA = AB = I_n$$

is called the *inverse matrix* of matrix A, and denoted by A^{-1}.

As seen we gave the above notion presupposing the existence and the uniqueness of the *inverse* of a *non-singular matrix*. The proofs for these facts can be explored at Schottt [64].

Next we summarize a few basic useful properties of the inverse of a matrix in the following proposition. They all can be easily proved using the above definition.

Proposition 2.1.2. *Let $A, B \in \mathscr{M}^{n \times n}$ be non-singular matrices, and α a nonzero scalar.*

(a) $(\alpha A)^{-1} = \frac{1}{\alpha} A^{-1}$;

(b) $(A^{\top})^{-1} = (A^{-1})^{\top}$;

(c) $(A^{-1})^{-1} = A$;

(d) $|A^{-1}| = \frac{1}{|A|}$;

(e) *If A is symmetric, then A^{-1} is symmetric; that is, $A^{-1} = (A^{-1})^{\top}$;*

(f) $(AB)^{-1} = B^{-1}A^{-1}$;

(g) *If $A = D(a_{11}, \ldots, a_{nn})$, then $A^{-1} = D\left(\frac{1}{a_{11}}, \ldots, \frac{1}{a_{nn}}\right)$.*

Proof. See Schott [64], Theorem 1.6. □

Now we turn to what we may call inside structure of a matrix. Specifically, we will reveal a few interesting and useful properties hidden in the matrix columns (rows).

Since each of the m columns of a matrix $A \in \mathscr{M}^{n \times m}$ has n entries they may be identified with vectors in $\mathbb{R}^n$ so that we may write $A = [v_1 \ldots v_m]$, where $v_i = \begin{bmatrix} a_{1i} \\ \vdots \\ a_{ni} \end{bmatrix}$, $i = 1, \ldots, m$. It is easily noted that the linear combination of the column vectors of A can be written as a product of A with a vector $x \in \mathbb{R}^m$: $Ax = x_1 v_1 + \ldots + x_m v_m$, where $x = \begin{bmatrix} x_1 \\ \vdots \\ x_m \end{bmatrix}$.

Definition 2.1.4. The set of all possible linear combination of the column vectors $v_1, \ldots, v_m$ of A is called the *range* of A, and denoted by $R(A)$; that is, with $x = [x_1 \ldots x_m]^{\top}$ representing any vector of reals x_i,

$$R(A) = \{v \in \mathbb{R}^n : Ax = v\} \subset \mathbb{R}^n.$$

Definition 2.1.5. The set of all vector $w \in \mathbb{R}^m$ such that $Aw = 0$ is called the *null space* of A, and denoted by $N(A)$; that is,

$$N(A) = \{w \in \mathbb{R}^m : Aw = \mathbf{0}_n\} \subset \mathbb{R}^m.$$

As it may be noted, $R(A)$ and $N(A)$ are vector subspace of the vector spaces $\mathbb{R}^n$ and $\mathbb{R}^m$, respectively (see Schott [64]). We prove it on the next result.

Proposition 2.1.3. *Let $A \in \mathbb{R}^{n \times m}$. Then*

(a) *$R(A)$ is a vector subspace of $\mathbb{R}^n$;*

(b) *$N(A)$ is a vector subspace of $\mathbb{R}^m$;*

Proof.

(a): $\mathbf{0}_n \in R(A)$, since $A\mathbf{0}_m = \mathbf{0}_n$. If $v, w \in R(A)$, we have that $Ax = v$ and $Ay = w$ for some vectors $x, y \in \mathbb{R}^m$. Thus, $A(x + y) = Ax + Ay = v + w$ so that $v + w \in R(A)$. Finally, let α be an scalar and $v \in R(A)$. Then, since $v = Ax$ for some vector $x \in \mathbb{R}^m$, and $A(\alpha x) = \alpha(Ax) = \alpha v$, we have that $\alpha v \in R(A)$ and so the proof for this part is completed; that is $R(A)$ is a vector subspace.

(b): $\mathbf{0}_m \in N(A)$, since $A\mathbf{0}_m = \mathbf{0}_n$. Letting $v, w \in N(A)$ we have that $Av = \mathbf{0}_n$ and $Aw = \mathbf{0}_n$. Then, $A(v + w) = Av + Aw = \mathbf{0}_n$ and, therefore, $v + w \in N(A)$. Finally, for any scalar α and $v \in N(a)$ we will have that $A(\alpha v) = \alpha(Av) = \mathbf{0}_n$ so that $\alpha v \in N(A)$ and so the proof is completed; that is $N(A)$ is a vector subspace. $\qquad\square$

Definition 2.1.6. The dimension of $R(A)$ is called the *rank* of A. We denote it here by $r(A)$.

The following theorems summarize a set of useful results on *matrix rank*. Some of them will play important role in the chapter 4 (which we may call this work main contributes), since they will have direct implication (either they justify some steps, or taken as consequence) on the proofs for the most important results on that chapter.

Theorem 2.1.4. *Let $A \in \mathcal{M}^{n \times m}$. Then*

(a) *$r(A)$ and the dimension of the row space of A are equal; that is, $r(A) = r(A^\top)$;*

(b) *$r(A) + dimN(A) = m$, where $dimN(A)$ denotes the dimension of $N(A)$.*

(c) *If $n = m$ and A is a diagonal matrix, then $r(A) = r$, where r is the number of its nonzero diagonal elements.*

Proof. For the proof of (a) and (b) see Lay [38]. (c) is a direct consequence of the definition. $\quad\square$

Theorem 2.1.5. *Let $A \in \mathcal{M}^{n \times m}$, $B \in \mathcal{M}^{n \times n}$ and $C \in \mathcal{M}^{m \times m}$ be non-singular matrices. Then*

(a) *$r(A) = r(A^\top) = r(A^\top A) = r(AA^\top)$;*

(b) *$r(BAC) = r(BA) = r(AC) = r(A)$.*

Proof. See Schott [64], Theorems 1.8 and 2.10. $\qquad\square$

Theorem 2.1.6. *Let $A_i \in \mathcal{M}^{n \times m}$, $i = 1, 2$, and $B \in \mathcal{M}^{m \times p}$. Then*

(a) *$r(A_1 + A_2) \leq r(A_1) + r(A_2)$;*

(b) *$r(A_1 B) \leq min\{r(A_1), r(B)\}$.*

Proof. See Schott [64], Theorem 2.10. $\qquad\square$

2.1.1 Diagonalization Of A Symmetric Matrix

In this section we present some brief notions of *eigenvalues* and *eigenvectors* (which are solutions of a specific equation of matrix functions) and then discuss a few important results concerning this subject as well as some others connecting it with *matrix rank*. This concepts are defined over a *square matrix*. The latter part of this section is devoted to a discussion on the diagonalization of symmetric matrix, specifically, the spectral decomposition of a matrix.

Definition 2.1.7. Let $A \in \mathscr{M}^{n \times n}$. Any scalar λ such that

$$(A - \lambda I_n)v = 0, \tag{2.5}$$

for some non-null vector $v \in \mathbb{R}^n$, is called an *eigenvalue* of A. Such a non-null vector v is called the *eigenvector* of A and equation (2.5) the *eigenvalue-eigenvector* equation.

Since $v \neq 0$, it must be noted that the *eigenvalue* λ must satisfy the determinant equation

$$|A - \lambda I_n| = 0 \tag{2.6}$$

which is known as characteristic equation of A since using the definition of determinant function (see Schott [64]) it can be equivalently written as

$$\theta_o - \theta_1 \lambda + \ldots + \theta_{n-1}(-\lambda)^{n-1} + (-\lambda)^n = 0, \tag{2.7}$$

for some scalars θ_i, $i = 0, \ldots, n-1$, that is, as an *nth* degree polynomial in λ.

Theorem 2.1.7. *Let $A \in \mathscr{M}^{n \times n}$. Then*

(a) *λ is eigenvalue of A if and only if λ is eigenvalue of $A^\top$;*

(b) *A is non-singular if and only if A has no null eigenvalues;*

(c) *If $B \in \mathscr{M}^{n \times n}$ is a non-singular matrix, then the eigenvalues of BAB^{-1} are the same as the those of A;*

(d) *$|A|$ is equal to the product of the eigenvalues of A.*

Proof. straightforward using the characteristic equation or the eigenvalue-eigenvector equation. $\square$

It is known that the polynomial in the left side of (2.7) has at most n real roots (and exactly n complex roots); that is, there are at most n scalar, $\lambda_1, \ldots, \lambda_n$ say, satisfying the equation (2.7) if solved in λ, so that A has at most n real eigenvalues.

Theorem 2.1.8. *Let $A \in \mathscr{S}^n$ and $B \in \mathscr{M}^{n \times n}$. Then*

(a) *The set of eigenvectors associated to different eigenvalues of B are linearly independent.*

(b) *Let $\lambda_1, \ldots, \lambda_s$, $s \leq n$, be the eigenvalues of A. Then $\lambda_1, \ldots, \lambda_s$ are all reals, and for each λ_i, $i = 1, \ldots, s$, there is an eigenvector v_i that is a vector of reals;*

(*c*) *It is possible to construct a set of eigenvectors of A such that the set is orthonormal; that is each element in the set has euclidean norm equal to one and they are pairwise orthogonal.*

Proof. (*a*): with $r < n$, suppose $v_1, \ldots, v_r$ are the eigenvectors of A, and let the corresponding eigenvalues $\lambda_1, \ldots, \lambda_r$ be such that $\lambda_i \neq \lambda_j$, whenever $i \neq j$. Now suppose, by contradiction, that $v_1, \ldots, v_r$ are linearly dependent. Let h be the largest integer for which $v_1, \ldots, v_h$ are linearly independent (that is, $v_1, \ldots, v_{h+1}$ must be linearly dependent). Thus, since no eigenvector can be a null vector, there exist scalars $\alpha_1, \ldots, \alpha_{h+1}$ with at least two not equal to zero, such that

$$\alpha_1 v_1 + \ldots + \alpha_{h+1} v_{h+1} = \mathbf{0}_n.$$

Premultiplying both the left-hand and the right-hand side of this equation by $(A - \lambda_{h+1} I_n)$ it is found that

$$\alpha_1 (A - \lambda_{h+1} I_n) v_1 + \ldots + \alpha_{h+1} (A - \lambda_{h+1} I_n) v_{h+1} = \mathbf{0}_n \Leftrightarrow$$
$$\alpha_1 (A v_1 - \lambda_{h+1} v_1) + \ldots + \alpha_{h+1} (A v_{h+1} - \lambda_{h+1} v_{h+1}) = \mathbf{0}_n \Leftrightarrow$$
$$\alpha_1 (\lambda_1 v_1 - \lambda_{h+1} v_1) + \ldots + \alpha_{h+1} (\lambda_{h+1} v_{h+1} - \lambda_{h+1} v_{h+1}) = \mathbf{0}_n \Leftrightarrow$$
$$\alpha_1 (\lambda_1 - \lambda_{h+1}) v_1 + \ldots + \alpha_h (\lambda_h - \lambda_{h+1}) v_h = \mathbf{0}_n. \tag{2.8}$$

Thus, since $v_1, \ldots, v_h$ are linearly independent it follows that

$$\alpha_1 (\lambda_1 - \lambda_{h+1}) = \cdots = \alpha_h (\lambda_h - \lambda_{h+1}) = \mathbf{0}_n.$$

Now, since at least one of the scalars $\alpha_1, \ldots, \alpha_h$ is not equal to zero, for some $i = 1, \ldots, h$ we have that $\lambda_i = \lambda_{h1}$, which contradicts the condition of (*a*).

(*b*): Let $\lambda = \alpha + i\beta$ be an eigenvalue of A and $v = x + iz$ its corresponding eigenvector, where $i = \sqrt{-1}$. Thus, we have that

$$Av = \lambda v \Leftrightarrow A(x + iz) = (\alpha + i\beta)(x + iz), \tag{2.9}$$

and premultiplying by $(x - iz)^\top$, it yields

$$(x - iz)^\top (x + iz) = (\alpha + i\beta)(x - iz)^\top (x + iz). \tag{2.10}$$

Hence A is symmetric equation (2.10) simplifys to

$$x^\top A x + z^\top A z = (\alpha + i\beta) \left(x^\top x + z^\top z \right). \tag{2.11}$$

Since $v \neq \mathbf{0}_n$ (it is an eigenvector) we have that $(x^\top x + z^\top z) > 0$ and, consequently, $\beta = 0$ since the left-hand side of the equation (2.11) is real. Now, replacing β with zero in the eigenvalue-eigenvector equation (2.9) we get

$$A(x + iz) = \alpha(x + iz) \Leftrightarrow$$
$$Ax + iAz = \alpha x + i\alpha z.$$

Then $v = x + iz$ will be an eigenvector of A corresponding to $\lambda = \alpha$ as long as $Ax = \alpha x$, $Az = \alpha z$ and at least one of the vectors x and z must be non-null (once $v \neq \mathbf{0}_n$). Finally, a real eigenvector corresponding to $\lambda = \alpha$ is then constructed by choosing $x \neq \mathbf{0}_n$ such that $Ax = \alpha x$ and $z = 0$.

(c): Let v be any non-null vector orthogonal to each of the eigenvectors in the set $\{v_1, \ldots, v_h\}$, where $1 \leq h < n$. Note that the set $\{v_1, \ldots, v_h\}$ contains at least one eigenvector of A. Note also that for any integer $k \geq 1$, $A^k v$ is also orthogonal to each of the vectors in the set since, if λ_i is the eigenvalue corresponding to v_i, it follows from the symmetry of A and Theorem A.1.1 that

$$v_i^\top A^k v = A^{k^\top} v_i^\top v = (A^k v_i)^\top v = \lambda_i^k v_i^\top v = 0.$$

According with the Theorem A.1.2 we have that the space spanned by the vectors v, Av, ..., $A^{r-1}v$, $r \geq 1$, contains an eigenvector of A. Let it be v^*. Clearly v^* is also orthogonal to the vectors in the set $\{v_1, \ldots, v_h\}$, since it comes from a vector space spanned by a set of vectors which are orthogonal $v_1, \ldots, v_h$. Thus we can take $v_{h+1} = (v^{*\top} v)^{\frac{-1}{2}}$. Then, starting with any eigenvector of A, and proceeding with the same argument $n - 1$ times the theorem follows. $\qquad \square$

It must be noted that if the matrix $A \in \mathscr{M}^{n \times n}$ has $1 < r \leq n$ eigenvalues $\lambda_1, \ldots, \lambda_r$ whose the corresponding eigenvectors will be the non-null vectors $v_1, \ldots, v_r$, i.e. $(A - \lambda_i)v_i = 0$, $i = 1, \ldots, r$, the eigenvalue - eigenvector equation can be written as

$$AV = V\Lambda \text{ or, equivalently, } (AV - V\Lambda) = 0_{n,r}, \tag{2.12}$$

where V is a matrix in $\mathscr{M}^{n \times r}$ whose the columns are $v_1, \ldots, v_r$ and $\Lambda = D(\lambda_1, \ldots, \lambda_r)$.

Thus, if the n (complex) eigenvalues $\lambda_1, \ldots, \lambda_n$ of $A \in \mathscr{M}^{n \times n}$ are all distinct, it follows from the Theorem 2.1.8, part (a), that the matrix V whose the columns are $v_1, \ldots, v_n$, the eigenvectors associated to those eigenvalues, is non-singular. Thereby, in this case, the eigenvalue-eigenvector equation (2.12) can equivalently be written as $V^{-1}AV = \Lambda$ or $A = V\Lambda V^{-1}$, with $\Lambda = D(\lambda_1, \ldots, \lambda_n)$. We may note that by the Theorem 2.1.7, part (c), the eigenvalues of Λ are the same as those of A. Since A can be transformed into a diagonal matrix by post-multiplication by the non-singular matrix V and pre-multiplication by its inverse, A is said to be a *diagonalizable matrix* (see this notion in Schott [64]).

Now, provided A is in $\mathscr{S}^n$, we will have the following result.

Theorem 2.1.9. *Let $A \in \mathscr{S}^n$. Then, the eigenvectors of A associated to different eigenvalues are orthogonal.*

Proof. Let λ_i and λ_j be two different eigenvalues of A whose the corresponding eigenvectors are v_i and v_j, respectively. Since A is symmetric we will have that

$$\lambda_i v_i^\top v_j = (Av_i)^\top v_j = v_i^\top (Av_j) = \lambda_j v_i^\top v_j,$$

and since $\lambda_i \neq \lambda_j$, it must holds $v_i^\top v_j = 0$. $\qquad \square$

Thus, according with Theorem 2.1.8 (c), if $A \in \mathscr{S}^n$, the n columns $v_1, \ldots, v_n$ of the matrix V, can be taken to be orthonormal so that, with out lost in generality, V can be taken to be an orthogonal matrix. Thus, the eigenvalue - eigenvector equation can now be written as

$$V^\top A V = \Lambda \text{ or, equivalently, } A = V \Lambda V^\top, \tag{2.13}$$

which is known as *spectral decomposition* of A (see this notion in Schott [64]).

Theorem 2.1.10. *Let $A \in \mathscr{S}^n$ and suppose A has r nonzero eigenvalues. Then $r(A) = r$;*

Proof. Let $A = V \Lambda V^\top$ be the *spectral decomposition* of A. Then, since the diagonal matrix Λ has r non-null elements (the non-null eigenvalues of A) and the matrix V is an orthogonal matrix, according with Theorem 2.1.7 (c), we have that

$$r(A) = \left(V \Lambda V^\top \right) = r(\Lambda) = r.$$

$r(\Lambda) = r$ follows from the Theorem 2.1.4 (c). $\qquad\square$

2.2 Orthogonal Basis And Projection Matrices

Let S be an vector subspace of $\mathbb{R}^n$ and the set of vectors $\{e_1, \ldots, e_n\}$ be an orthonormal basis for $\mathbb{R}^n$. Let also the set $\{e_1, \ldots, e_r\}$, with $r < n$, be an orthogonal basis for S. The above statements are legitimate since every vector space (except the zero-dimensional one) has an orthogonal basis, as is guaranteed by Schott [64] (See theorem 2.13). $\{e_1, \ldots, e_n\}$ being orthonormal basis for $\mathbb{R}^n$ means $e_i^\top e_j = 0, i \neq j$, and $e_i^\top e_i = 1, i = j$, with $e_i \in \mathbb{R}^n$, and the set $\{e_1, \ldots, e_n\}$ spans $\mathbb{R}^n$.

Definition 2.2.1. The set of vectors in $\mathbb{R}^n$ which are orthogonal to every vector in S is said to be the *orthogonal complement* of S, and is denoted by $S^\perp$; that is, $S^\perp = \{x \in \mathbb{R}^n : x^\top y = 0, \ y \in S\}$.

Every vector $x \in \mathbb{R}^n$ can be written as a sum of a vector $u \in S$ with a vector $v \in S^\perp$, where $S^\perp$ denotes the *orthogonal complement* of the subspace S (see Schott [64], Theorem 2.14, for the proof for such result).

Theorem 2.2.1. *The orthogonal complement of S, $S^\perp$, is also a vector subspace of $\mathbb{R}^n$, i.e., $S^\perp \subset \mathbb{R}^n$.*

Proof. See Schott [64], Theorem 2.15. $\qquad\square$

Let $A_1 = [e_1 \ldots e_r]$, $A_2 = [e_{r+1} \ldots e_n]$, and $A = [A_1 A_2]$, with $e_i \in \mathbb{R}^n$, $i = 1, \ldots, n$, and $\{e_1 \ldots e_n\}$ an orthonormal basis for $\mathbb{R}^n$.

The following results are quickly achieved.

Proposition 2.2.2. *Consider the matrices A_1, A_2 and A defined above. Then*

(a) $A_1^\top A_1 = I_r$;

(b) $A_2^\top A_2 = I_{n-r,n-r}$;

(c) $A_1^\top A_2 = 0_{r,n-r}$;

(d) $A_2^\top A_1 = 0_{n-r,r}$;

(e) $A^\top A = AA^\top = I_n$.

Proof. All points arise due to the fact that the columns vectors $e_1, \ldots, e_n$ are orthonormal. $\square$

Since $\{e_1, \ldots, e_n\}$ is an orthonormal basis for $\mathbb{R}^n$, any vector $x \in \mathbb{R}^n$ can be written as $x = A\alpha$ for some $\alpha = [\alpha_1^* \alpha_2^*]^\top$, with $\alpha_1^* = [\alpha_1 \ldots \alpha_r]$ and $\alpha_2^* = [\alpha_{r+1} \ldots \alpha_n]$.

Definition 2.2.2. A vector u such that

$$A_1 A_1^\top x = A_1 A_1^\top A\alpha = [A_1 \ \mathbf{0}_{r,n-r}][\alpha_1^* \alpha_2^*]^\top = A_1 \alpha_1^* = u \tag{2.14}$$

is said to the *orthogonal projection* of the vector $x \in \mathbb{R}^n$ onto the subspace S, and the matrix $P_S = A_1 A_1^\top$ is said to be the *projection matrix* onto the subspace S. Similarly, a vector v such that

$$A_2 A_2^\top x = A_2 A_2^\top A\alpha = [\mathbf{0}_{n-r,r} \ A_2][\alpha_1^* \alpha_2^*]^\top = A_2 \alpha_2^* = v \tag{2.15}$$

will be the *orthogonal projection* of the vector $x \in \mathbb{R}^n$ onto the subspace $S^\perp$.

The following notions are needed for our next result.

Definition 2.2.3. A matrix $B \in \mathcal{M}^{n \times n}$ such that $B^2 = B$ is said to be *idempotent*.

Definition 2.2.4. Let the columns of the matrix B form a basis for the subspace S. An *symmetric* and *idempotent* matrix P_S such that $r(B) = r(P_S)$ is said to be the *projection matrix* onto $S = R(B)$.

Note 2.2.1. We may sometimes write *symmetric idempotent* matrix in place of *symmetric* and *idempotent* matrix.

The next result ensures that any *symmetric idempotent* matrix is a *projection matrix* for some vector subspace.

Theorem 2.2.3. *Let $Q \in \mathcal{M}^{n \times n}$ be any symmetric idempotent matrix such that $r(Q) = r$. Then Q is a projection matrix of some r-dimensional vector subspace (vector subspace with dimension r).*

Proof. See Schott [64], Theorem 2.19. $\square$

The following theorem establishes that if $S \subseteq \mathbb{R}^n$ is r-dimensional subspace, then $S^\perp$ is $n-r$-dimensional.

Theorem 2.2.4. *Let the columns of the matrix A_1 form an orthonormal basis for S, and the columns of $A = [A_1 A_2]$ an orthonormal basis for $\mathbb{R}^n$. Then, the columns of A_2 will be an orthonormal basis for $S^\perp$.*

Proof. Let T be the vector space spanned by the columns of the matrix A_2, i.e, $T = R(A_2)$. We firstly prove that $T \subseteq S^\perp$, and then that $S^\perp \subseteq T$. Let $u \in S$ and $v \in T$. Then, $u = \alpha_1 e_1 + \ldots \alpha_r e_r$ and $v = \alpha_{r+1} e_{r+1} + \ldots \alpha_n e_n$ for some scalars $\alpha_1, \ldots, \alpha_n$. The orthogonality of the vectors $e_1, \ldots, e_n$ holds $u^\top v = 0$, and therefore $v \in S^\perp$, which means that $T \subseteq S^\perp$.

Now, conversely, suppose $y \in S^{\perp}$. Due to the fact that $S^{\perp} \subseteq \mathbb{R}^n$ (see Theorem 2.2.1), $y = \alpha_1 e_1 + \ldots \alpha_n e_n$ for some scalars $\alpha_1, \ldots, \alpha_n$. Let $u \in S$, so $u = \alpha_1 e_1 + \ldots \alpha_r e_r$, and since $y \in S^{\perp}$ it must hold $y^{\top} u = \alpha_1^2 e_1^{\top} e_1 + \alpha_r^2 e_r^{\top} e_r = \alpha_1^2 + \ldots + \alpha_r^2 = 0$. But this result happens only if $\alpha_1 = \ldots = \alpha_r = 0$. Thus, $y = \alpha_{r+1} e_{r+1} + \ldots + \alpha_n e_n$ which means that $y \in T$, and therefore $S^{\perp} \subseteq T$. The proof was established (see Schott [64], page 52). $\qquad\square$

By now we have material to set the following results, which are immediate consequences of the results stated above in this section, so that sometimes we will not give the proofs.

Theorem 2.2.5. *With $x \in \mathbb{R}^n$, and A, A_1, and A_2 matrices whose the columns form the orthonormal basis for $\mathbb{R}^n$, S, and $S^{\top}$, respectively:*

1. *The orthogonal projection of x onto S is given by $P_S x = A_1 A_1^{\top} x$, for which $A_1 A_1^{\top}$ is the projection matrix for S (see equation (2.14));*

2. *The orthogonal projection of x onto $\mathbb{R}^n$ is given by $P_{\mathbb{R}^n} x = A A^{\top} x = (A_1 A_1^{\top} + A_2 A_2^{\top}) x = x$, for which $P_{\mathbb{R}^n} = A A^{\top} = I_m$ is the projection matrix for $\mathbb{R}^n$.*

3. *The orthogonal projection of x onto $S^{\perp}$ is given by $P_{S^{\perp}} x = A_2 A_2^{\top} x = [0\ A_2][\alpha_1\ \alpha_2]^{\top} = A_2 \alpha_2$, for which $P_{S^{\perp}} = A_2 A_2^{\top} = (I_n - A_1 A_1^{\top})$ is the projection matrix for $S^{\perp}$;*

Although the vector spaces does not have an unique *orthonormal basis*, the projection matrix formed by such basis is unique, as ensured by the following theorem.

Theorem 2.2.6. *Let the columns of a matrix C and D each form an orthonormal basis for the r-dimensional vector subspace S. Then, $CC^{\top} = DD^{\top}$.*

Proof. Each column of D is a linear combination of the columns of the matrix C, since its columns form a basis for S. So, there exists a matrix P such that $D = CP$. Once C and D have orthonormal columns, $C^{\top} C = D^{\top} D = I_r$. Thus, $I_r = D^{\top} D = (CP)^{\top} CP = P^{\top} P$, which means that P is also an orthogonal matrix. Consequently, $PP^{\top} = I_r$. The desired result: $DD^{\top} = CP(CP)^{\top} = CPP^{\top} C^{\top} = CC^{\top}$. $\qquad\square$

The following theorems summarizes the results on the *projection matrix*.

Theorem 2.2.7. *Let $P \in \mathcal{M}^{n \times n}$. Then, the following statements are equivalent.*

(a) *P is a projection matrix;*

(b) *$(I_n - P)$ is a projection matrix;*

(c) *$R(P) = N(I_n - P)$;*

(d) *$N(P) = R(I_n - P)$;*

(e) *$R(P) \cap R(I_n - P) = \{\mathbf{0}_n\}$;*

(f) *$N(P) \cap N(I_n - P) = \{\mathbf{0}_n\}$.*

Proof. $(a) \Rightarrow (b)$: $(I_n - P)^\top (I_n - P) = I_n - P - P^\top + P^\top P = I_n - P$, since $P^\top = P$ and $P^\top P = P$ due to the fact that P is a *projection matrix*. Therefore, $I_n - P$ is a *projection matrix*.

$(b) \Rightarrow (c)$: Let $x \in R(P)$. Then, since P is a *projection matrix*, $Px = x$. thus, $(I - P)x = x - Px = \mathbf{0}_n$ so that $x \in N(I - P)$. Now, conversely, let $x \in N(I_n - P)$. Thus, $(I_n - P)x = x - Px = \mathbf{0}_n \Leftrightarrow Px = x$, which means $x \in R(P)$. Therefore $R(P) = N(I_n - P)$.

$(c) \Rightarrow (d)$: Let $x \in N(P)$. Then $Px = \mathbf{0}_n$. Thus $(I_n - P)x = x - Px = x$, once $Px = \mathbf{0}_n$. So $x \in R(I_n - P)$. Conversely, let $x \in R(I_n - P)$. Then

$$(I_n - P)x = x \Leftrightarrow x - Px = x \to Px = \mathbf{0}_n,$$

that is $x \in N(P)$.

$(d) \Rightarrow (e)$: Let $x \in R(P)$ and $y \in R(I_n - P)$. Then, by point (d) we have that $y \in N(P)$. We have that

$$x^\top y = (Px)^\top (I - P)y = (Px)^\top (y - Py) = x^\top P^\top y = x^\top Py = \mathbf{0}_n,$$

since P is a *projection matrix* and $y \in N(P)$. In other hand

$$y^\top x = [(I_n - P)y]^\top Px = (y - Py)^\top Px = y^\top Px = [P^\top y]^\top x = \mathbf{0}_n,$$

since P is a *projection matrix* and $y \in N(P)$.

$(e) \Rightarrow (f)$: Let $x \in N(P)$ and $y \in N(I_n - P)$. Then, $x \in R(I_n - P)$ and $y \in R(P)$. Thus,

$$x \top y = [(I_n - P)x]^\top Py = x^\top Py = [P^\top x]^\top y = \mathbf{0}_n.$$

In other hand

$$y^\top x = (Py)^\top x = y^\top P \top x = \mathbf{0}_n.$$

$\square$

Theorem 2.2.8. *Let $P_1, \ldots, P_k$ be projections matrices such that $P_i P_j = 0$ for all $i \neq j$. Then*

1. *$P = \sum_{i=1}^k P_i$ is a projection matrix.*

2. *$R(P_i) \cap R(P_j) = \{0\}$ for all $i \neq j$, and $R(P) = R(P_1) \oplus R(P_1) \oplus \ldots \oplus R(P_k)$, with $\oplus$ denoting the direct sum of subspace.*

Proof. See Rao and Rao [61], page 241. $\square$

2.2.1 Application To Statistics

Let $\{x_1, ..., x_r\}$ be a basis for the vector space $S \subseteq \mathbb{R}^n$, i.e., $\{x_1, ..., x_r\}$ are linearly independent and generate S. Let $X \in \mathscr{M}^{n \times r}$, a matrix whose columns are the vectors $x_1, ..., x_r$, i.e, $X = [x_1, \ldots, x_r]$. Then, the columns of the matrix $Z = XA$ will form an *orthonormal basis* for S if $A \in \mathscr{M}^{r \times r}$ is a matrix such that

$$Z^\top Z = A^\top X^\top XA = I_r. \tag{2.16}$$

Thus, A must be a non-singular matrix and $r(A) = r(X) = r$.

$(r(X) = r$, since X has r linearly independent columns.) So, A^{-1} exists.

The equation (2.16) holds if $(X^\top X) = (A^{-1})^\top A^{-1}$ or $(X^\top X)^{-1} = AA^\top$, where A is the square root matrix of $(X^\top X)^{-1}$.

By the Theorem 2.2.5 (see also Definition 2.2.2), the expression for the *projection matrix* onto S, P_S, is

$$P_S = ZZ^\top = XAA^\top X^\top = X(X^\top X)^{-1}X^\top. \tag{2.17}$$

Therefore, the *projection matrix* onto the subspace of $\mathbb{R}^n$ spanned by the columns of the matrix X is $P_S = X(X^\top X)^{-1}X^\top$.

Consider the simple fixed effect linear model

$$y = X\beta + \varepsilon, \tag{2.18}$$

with $X \in M_{n,m}$ a known matrix and $\beta \in \mathbb{R}^m$ a vector of unknown parameters, where $y \in \mathbb{R}^n$ is an observable random vector with expectation $E(y) = X\beta$, and variance-covariance matrix $V = I \in \mathscr{M}^{n \times n}$.

Let $\hat{\beta}$ be an unbiased estimator for β. Hence, an unbiased estimator for y would be $\hat{y} = X\hat{\beta}$, in which $\hat{y}$ is a point in a subspace of $\mathbb{R}^n$, say S, which corresponds exactly to the subspace of $\mathbb{R}^n$ spanned by the linear combinations of the columns of the matrix X.

Remark 2.2.1. According with the paragraph above, every unbiased estimator for y lies onto $S \subseteq \mathbb{R}^n$.

Now, the reader may wonder "what is the best unbiased estimator $\hat{y}$ for y". The answer follows:

Such (best) unbiased estimator must be the point in S which is closest to y, i.e., the *orthogonal projection* of y onto the subspace spanned by the columns of X. That is, $\hat{y} = X\hat{\beta}$ is the best unbiased estimator for y if it is the *orthogonal projection* of the vector y onto S.

So, one must compute the *orthogonal projection* of y onto S. In order to do so, let $r(X) = m$. By the equation (2.17) the projection matrix onto S is $P_S = X(X^\top X)^{-1}X^\top$, so that the *orthogonal projection* of y onto S is given by $P_S y = X(X^\top X)^{-1}X^\top y$. Now, since $\hat{y}$ is the best unbiased estimator for y it holds:

$$\hat{y} = X\hat{\beta} = P_S y = X(X^\top X)^{-1}X^\top y. \tag{2.19}$$

Pre - multiplying each part of the equation above by $(X^\top X)^{-1}X^\top$ it yields:

$$(X^\top X)^{-1}X^\top X\hat{\beta} \quad = \quad (X^\top X)^{-1}X^\top X(X^\top X)^{-1}X^\top y$$
$$\Longleftrightarrow$$
$$\hat{\beta} \quad = \quad (X^\top X)^{-1}X^\top y, \tag{2.20}$$

which corresponds exactly to the *least squares estimator* for β.

Thus, $\hat{\beta}$ is the estimator which minimizes the sum of the *quadratic mean error*, that is, the one which satisfies

$$min_\beta(SSE) = min_\beta(y - X\beta)(y - X\beta),$$

so that

$$
\begin{aligned}
SSE_{\hat{\beta}} &= (y - X\hat{\beta})^{\top}(y - X\hat{\beta}) \\
&= y^{\top}y - y^{\top}X(X^{\top}X)^{-1}X^{\top}y \\
&= y^{\top}\left(I_n - X(X^{\top}X)^{-1}X^{\top}\right)y.
\end{aligned}
\tag{2.21}
$$

According with Theorem 2.2.5, the term $\left(I_n - X(X^{\top}X)^{-1}X^{\top}\right)$ in the equation (2.21) is the *projection matrix* onto the *orthogonal complement* of S, $S^{\perp}$, so that the term $\left(I_n - X(X^{\top}X)^{-1}X^{\top}\right)y$ is the projection of y onto such space. So, the *quadratic mean error* of the estimator $\hat{\beta}$ is the quadratic distance of the projection of y onto $S^{\perp}$.

2.3 Generalized Inverses

The generalized inverse, in short, *g-inverse*, play an important role in linear algebra, as well as in statistics, as we will see throughout this work.

Let consider the following system of linear equation in an unknown vector $x \in \mathbb{R}^n$:

$$
Ax = y,
\tag{2.22}
$$

where $A \in \mathscr{M}^{m \times n}$ and $y \in \mathbb{R}^m$ are unknown. Such system is said to be *consistent* if it admits a solution in x.

If the matrix A is a non-singular, the solution for the equation (2.22) is $x = A^{-1}y$, where A^{-1} denotes the matrix inverse of A.

When A does not admit an inverse matrix (A^{-1}) but the system (2.22) still *consistent*, there still having a simple way to solving the system (2.22) if $r(A) = m$ or $r(A) = n$ (A is a full rank matrix):

- If $r(A) = m$ (the m rows of the matrix A are linearly independent) A admits a *right inverse*, say L, so that the solution for (2.22) is $x = Ly$. Indeed, $A(Ly) = ALAx = Ax = y$.

- If $r(A) = n$ (the n columns of the matrix A are linearly independent) A admits a *left inverse*, say G, so that the solution for the equation (2.22) is $x = Gy$. Indeed, $Ax = A(Gy) = AGAx = AIx = Ax = y$.

The results $P.8.1.1$ and $P.8.1.2$ of Rao and Rao [61] suggest $L = A^{\top}(AA^{\top})^{-1}$ or $L = VA^{\top}(AVA^{\top})^{-1}$, where V is an arbitrary matrix satisfying $r(A) = r(AVA^{\top})$, and

$$
G = (A^{\top}A)^{-1}A^{\top} \text{ or } G = (A^{\top}VA)^{-1}A^{\top}V.
$$

The *g-inverses* arises when one needs to determine solutions for the system (2.22) given it is *consistent*, when $A \in \mathscr{M}^{m \times n}$ has an arbitrary rank. Such inverse, which is denoted by A^{-}, does always exists for any matrix A, as proved by the result $P.8.2.2$ of Rao and Rao [61]. Thus, $x = A^{-}y$ is a solution for the equation (2.22).

Before one set a possible definition of *g-inverse*, one set the following results (whose proofs can be founded in Rao and Rao [61]) which may be useful for that purpose.

Proposition 2.3.1. *Let $A \in \mathcal{M}^{m \times n}$. Then, the following statement are equivalent.*

(a) A^- *is a g-inverse of* A.

(b) AA^- *is an identity on* $R(A)$, *i.e.,* $AA^-A = A$.

(c) AA^- *is idempotent and* $r(A) = r(AA^-)$.

Proof. See Rao and Rao [61], page 267. □

Hereupon, one way to define *g-inverse* may arise.

Definition 2.3.1. *Let $A \in \mathcal{M}^{m \times n}$. De g-inverse of A is a matrix $G \in \mathcal{M}^{n \times m}$ such that $AGA = A$.*

As stated in the Proposition above, given a matrix the *g-inverse* does always exists, but it could may not be unique. Many properties of the g-inverse can be stated including the one concerning the conditions on the uniqueness.

Theorem 2.3.2. *Let $A \in \mathcal{M}^{m \times n}$ and A^- its g-inverse. Then,*

(a) $(A^-)^\top$ *is a g-inverse of* $A^\top$.

(b) $\alpha^{-1}A^-$ *is a g-inverse of* αA, *where α is a scalar.*

(c) *If A is square and non-singular, $A^- = A^{-1}$ and it is unique.*

(d) *If B and C are non-singular, $C^{-1}A^-B^{-1}$ is a g-inverse of BAC.*

(e) $r(A) = r(AA^-) = r(A^-A) \leq r(A^-)$.

(f) $r(A) = m$ *if and only if $AA^- = I_m$.*

(g) $r(A) = n$ *if and only if $A^-A = I_n$.*

Proof. See schott [64], Theorem 5.22. □

An important matrix in the field of linear models is $A(A^\top A)^- A^\top$ as we will see throughout the remaining sections. We see now some properties whose the proofs may be found in schott [64] or Rao and Rao [61].

Proposition 2.3.3. *Let $(A^\top A)^-$ stands for a g-inverse of $A^\top A$. Then*

1. $A(A^\top A)^-(A^\top A) = A$ *and* $(A^\top A)(A^\top A)^- A^\top = A^\top$).

2. $A(A^\top A)^- A^\top$ *is the orthogonal projection matrix of the $R(A)$.*

Proof. See Schott [64] or Rao and Rao [61]. □

Now we turn to maybe the most important *generalized inverse* in statistical application in such a way that we could not talking about g-inverses without mention it: the *Moore-Penrose inverse*. We devote the next subsection for its approach.

2.3.1 Moore-Penrose Inverse

Firstly defined by Moore [9] and later by Penrose [60], the greatest importance of the *Moore-Penrose inverse* is due to the fact that it possesses four properties that the inverse of a square non-singular matrix has (more evident with the Penrose [60] definition) and it is uniquely defined. Such definitions, although from different times, are equivalent as shows the Theorem 5.2 of Schott [64].

The Moore [9] definition follows.

Definition 2.3.2. Let $A \in \mathcal{M}^{n \times m}$. The *Moore-Penrose inverse* of A is the unique matrix, denoted by $A^+ \in \mathcal{M}^{m \times n}$, that satisfies the two following conditions:

(a) $AA^+ = P_{R(A)}$.

(b) $A^+A = P_{R(A^+)}$.

The Penrose [60] definition follows.

Definition 2.3.3. Let $A \in \mathcal{M}^{n \times m}$. The *Moore-Penrose inverse* of A is the unique matrix, denoted by $A^+ \in \mathcal{M}^{m \times n}$, which satisfies the all four following conditions:

(a) $AA^+A = A$.

(b) $A^+AA^+ = A^+$.

(c) $(AA^+)^\top = AA^+$.

(d) $(A^+A)^\top = A^+A$.

Remark 2.3.1. One easily remark that the four conditions of the definition above are satisfied by the inverse, say A^{-1}, of a non-singular matrix A.

The following theorem guarantees the existence and the uniqueness of the *Moore-Penrose inverse*.

Theorem 2.3.4. *For each matrix $A \in \mathcal{M}^{n \times m}$, there exists one and only one matrix, say A^+, satisfying the all four condition of the definition 2.3.3.*

Proof. See Schott [64]. □

Theorem 2.3.5. *Let $A \in \mathcal{M}^{n \times m}$.*

(a) $(\alpha A)^+ = \alpha^{-1}A^+$, $\alpha \neq 0$.

(b) $(A^\top)^+ = (A^+)^\top$.

(c) $(A^+)^+ = A$.

(d) $A^+ = A^{-1}$, *if A is square and non-singular.*

(e) $(A^\top A)^+ = A^+(A^+)^\top$ *and* $(AA^\top)^+ = (A^+)^\top A^+$.

(f) $(AA^+)^+ = AA^+$ and $(A^+A)^+ = A^+A$.

(g) $A^+ = (A^\top A)^+ A^\top = A^\top (AA^\top)^+$.

(h) $A^+ = (A^\top A)^{-1} A^\top$ and $A^\top A = I_n$ if $r(A) = m$.

(i) $A^+ = A^\top (AA^\top)^{-1}$ and $AA^+ = I_m$ if $r(A) = n$.

(j) $A^+ = A^\top$ if the columns of A are orthogonal, that is $A^\top A = I_n$.

Proof. See schott [64], page 174. □

Two results follows: one establishes the relation between the rank of a matrix and its *Moore-Penrose inverse*, and the other one summarizes some special properties of a *Moore-Penrose inverse* of a symmetric matrix.

Theorem 2.3.6. *Let $A \in \mathcal{M}^{n \times m}$ and A^+ its Moore-Penrose inverse. Then,*

$$r(A) = r(A^+) = r(AA^+) = r(A^+A).$$

Proof. Using the conditions (a) and (b) of the definition 2.3.3 together with Theorem 4.2.1 of Rao and Rao [61] it holds

$$r(A) = r(AA^+A) \le r(AA^+) \le r(A^+)$$

(using condition (a)) and similarly

$$r(A^+) = r(A^+AA^+) \le r(A^+A) \le r(A)$$

(using condition (b)), from where the proposed results follows. □

Theorem 2.3.7. *Let $A \in \mathcal{M}^{n \times n}$ be a symmetric matrix and A^+ its Moore-Penrose inverse. Then*

(a) A^+ *is Symmetric;*

(b) $AA^+ = A^+A$;

(c) $A^+ = A$ *if A is idempotent.*

Proof.

Proof of property (a): using the part (b) of the Theorem 2.3.5 and the hypothesis condition $(A = A^\top)$ it follows

$$A^+ = (A^\top)^+ = (A^+)^\top.$$

Proof of property (b): using the condition (c) of the definition 2.3.3 together with the fact that both A and A^+ are symmetric, it holds

$$AA^+ = (AA^+)^\top = (A^+)^\top A^\top = A^+A.$$

Proof of property (c): one proves it proving that, under hypothesis $A^2 = A$, $A^+ = A$ verifies the four properties of the definition 2.3.3. For the condition (a) and (b):

$$AA^+A = AAA = AA = A^2 = A.$$

For the condition (c) and (d):

1. $(AA^+)^\top = (AA)^\top = A^\top A^\top = AA.$

2. $(A^+A)^\top = (AA)^\top = A^\top A^\top = AA.$

For more details concerning the *Moore-Penrose inverse* among others Schott [64] is recommended.

$\square$

2.4 Jordan Algebras

Jordan algebras structures were first introduced by Jordan [32] (the structures name is due to his name), and Jordan et al. [33] in the formalization of an algebraic structure for quantum mechanics. Originally they were called "r-number systems", but later they were renamed *Jordan algebras* by Albert [1] who generalized its notions.

Definition 2.4.1. An algebra $\mathscr{A}$ is a vector space provided with a binary operation * (usually denominated product), in which the following properties hold for all $a, b, c \in \mathscr{A}$:

- $a * (b + c) = a * b + a * c.$

- $(a + b) * c = a * c + b * c.$

- $\alpha(a * b) = (\alpha a) * b = a * (\alpha b), \forall \alpha \in \mathbb{R}(\mathbb{C}).$

$\mathscr{A}$ is a real algebra (complex algebra) whether α is real (complex).

Definition 2.4.2. An algebra $\mathscr{A}$ is said to be *commutative algebra* if, for all $a, b \in \mathscr{A}$, $a * b = b * a$, or *associative algebra* if, for all $a, b, c \in \mathscr{A}$, (a*b)*c = a*(b*c).

Definition 2.4.3. Let $\mathscr{A}$ be an algebra. $\mathscr{S} \subseteq \mathscr{A}$ is a *sub-algebra* if it is a vector space and if

$$\forall a, b \in \mathscr{S} : a * b \in \mathscr{S}.$$

Definition 2.4.4. Let $\mathscr{A}$ be an algebra provided with the binary operation "·" such that the following properties hold for all $a, b \in \mathscr{A}$:

$(J1):$ $a \cdot b = b \cdot a.$

$(J2):$ $a^2 \cdot (b \cdot a) = (a^2 \cdot b) \cdot a$, with $a^2 = a \cdot a$. Holding such conditions, $\mathscr{A}$ is said to be a *Jordan algebra*.

The product "·" defined above here is known as *Jordan product*.

Note 2.4.1.

- Properties *J*1 shows that a *Jordan algebra* is a *commutative algebra*, but, as shows *J*2, is not an *associative* one. In fact, *J*2 shows that $\mathscr{A}$ has a restricted kind of associativity.

- The definition of *Jordan algebra* presented above is not so practical and transparent, especially in statistical models context, so that equivalents and more tractable definitions is presented later in this section.

Next one see an example of a *Jordan algebra*: *the space of real symmetric matrix of order $n \times n$*, $\mathscr{S}^n$. That space (with finite dimension: $\frac{1}{2}n(n+1)$) will accompany us throughout our study in this section.

Note 2.4.2. In what follows, AB means the product matrix in usual sense, that is, the product between the matrices A and B.

Example 2.4.1. Define the product "$\cdot$" on $\mathscr{S}^n$ as

$$A \cdot B = \frac{1}{2}(AB + BA).$$

Provided with such product $\mathscr{S}^n$ is a *Jordan algebra*. Indeed: $A \cdot B = \frac{1}{2}(AB + BA) = \frac{1}{2}(BA + AB) = B \cdot A$, so that the condition $J1$ is proved. To prove the condition $J2$ one easy way is to compute the left side separately and then the right side. After that, one concludes that both results are equal. One should remark the importance of the $\mathscr{S}^n$ due to the fact that the matrix of *variance-covariance* lies on there.

Now one proceed in order to characterizes the *idempotent* and *identity* elements in *Jordan algebra* and $\mathscr{S}^n$.

Definition 2.4.5. Let $\mathscr{A}$ be a *Jordan algebra* and $\mathscr{B}$ a *sub-algebra* of $\mathscr{S}^n$. $\mathscr{A}$ is said to be a *special Jordan algebra* if and only if $\mathscr{A}$ is *algebra-isomorphic* to $\mathscr{B}$, that is, there exists a bijective function $\phi \colon \mathscr{A} \to \mathscr{B}$ such that, for all $\alpha, \beta \in \mathbb{R}$ and $a, b \in \mathscr{A}$:

1. $\phi(\alpha a + \beta b) = \alpha\phi(a) + \beta\phi(b)$.

2. $\phi(a * b) = \phi(a) * \phi(b)$.

Definition 2.4.6. Consider the matrix $E \in \mathscr{S} \subseteq \mathscr{M}^{n \times n}$. E is said to be:

- An *associative identity element* of $\mathscr{S}$ if $ES = SE = S, \forall S \in \mathscr{S}$;

- A *Jordan identity* if $E \cdot S = S, \forall S \in \mathscr{S}$.

Definition 2.4.7. Let $E \in \mathscr{M}^{n \times n}$. E is said to be *idempotent* matrix if $E^2 = E$.

The following theorem (see Malley [43], Lemma 5.1) proves that any *identity element* in a subspace of $\mathscr{S}^n$ is also *identity* on *Jordan algebra*.

Theorem 2.4.1. *Let $\mathscr{S} \subseteq \mathscr{S}^n$ and $E \in \mathscr{S}$ any idempotent element. Then,*

$$\exists S \in \mathscr{S} : E \cdot S = S \Rightarrow ES = SE = S.$$

Proof. See Malley [43], page 9. □

One prove next a more generally result, concern also the orthogonality, That is, orthogonal elements on *Jordan algebra* are also orthogonal on subspaces of $\mathscr{S}^n$, and conversely.

Theorem 2.4.2. *Let $\mathscr{S}$, $\mathscr{T} \subseteq \mathscr{S}^n$, and $E \in \mathscr{S}^n$ an idempotent element. Then,*

1. *$\forall S \in \mathscr{S}$, $E \cdot S = S \Longleftrightarrow \forall S \in \mathscr{S}$, $ES = SE = S$.*

2. *Let E_1 and E_2 idempotent elements. Then, $E_1 \cdot E_2 = 0 \Longleftrightarrow E_1 E_e = 0$.*

3. *$\forall S \in \mathscr{S}, \forall T \in \mathscr{T}$, $S \cdot T = 0 \Longleftrightarrow \forall S \in \mathscr{S}, \forall T \in \mathscr{T}, ST = 0$.*

Proof. We prove only the property 1, for the rest see Malley [43], page 10.

$(1 \Rightarrow)$ E is *idempotent* and such that $E \cdot S = S$ (hypothesis).

Firstly, one may note: $ESE = 2E \cdot (E \cdot S) - E^2 \cdot S = S$. (The equality $2E \cdot (E \cdot S) - E^2 \cdot S = S$ is easily proved provided the equality $A \cdot B = \frac{1}{2}(AB + BA)$.) Indeed, using the equality $2E \cdot (E \cdot S) = \frac{1}{2}(EES + 2ESE + SEE)$ and $E^2 \cdot S = \frac{1}{2}(EES + SEE)$, clearly $2E \cdot (E \cdot S) - E^2 \cdot S = ESE$.

Hence,

$$
\begin{aligned}
ES &= E(ESE) = E(ESE + E^2 \cdot S - E^2 \cdot S) = E(ESE + E^2 \cdot S) - E(E^2 \cdot S) \\
&= EESE + ES - ES = ESE = S,
\end{aligned} \tag{2.23}
$$

using the hypothesis. To prove the case $SE = S$ we proceed identically.

$(1 \Leftarrow)$ E is *idempotent* and such that $ES = SE = S$ (hypothesis).

$$
\begin{aligned}
E \cdot S &= E^2 \cdot S = 2E \cdot (S \cdot E) - ESE = 2\left(\frac{1}{2}(E(S \cdot E) + (S \cdot E)E)\right) - ESE \\
&= \frac{1}{2}ES + \frac{1}{2}SE = S,
\end{aligned} \tag{2.24}
$$

using the hypothesis. $\qquad \square$

The next theorem, whose proof will be given here (see Malley [43]) establish some equivalent conditions for a subspace of S_n with any *identity* element to be a *Jordan algebra*.

Theorem 2.4.3. *Let $\mathscr{S} \subseteq \mathscr{S}^n$, and suppose $\mathscr{S}$ contains an identity element, say E. then, $\mathscr{S}$ is a Jordan algebra if and only if any of the following equivalent conditions hold:*

(i) $\forall A, B \in \mathscr{S}$, $AB + BA \in \mathscr{S}$.

(ii) $\forall A, B \in \mathscr{S}$, $ABA \in \mathscr{S}$.

(iii) $\forall A \in \mathscr{S}$, $A^2 \in \mathscr{S}$.

Proof.

$(i) \Longrightarrow (ii)$ Note that $C = A - E \in \mathscr{S}$. Hence, $ABA = (C + E)B(C + E) = C^2 + B + (CB + BC)$. Now, given (i), $C^2 = \frac{1}{2}(C^2 + C^2) \in \mathscr{S}$, and also $2C \cdot B = CB + BC$. Thus, $ABA \in \mathscr{S}$ proving therefore such implication.

$(ii) \Longrightarrow (iii)$

By (ii), for all $A, B \in \mathscr{S}$, $ABA \in \mathscr{S}$. Therefore, taking $B = E \in \mathscr{S}$, it holds $AEA = A^2 \in \mathscr{S}$, proving therefore such implication.

$(iii) \Longrightarrow (i)$

Consider $C = A + B$. Then, $C^2 = (A + B)^2 = A^2 + B^2 + (AB + BA) \in \mathscr{S}$, so $(AB + BA) \in \mathscr{S}$. Hence, the proof is established.

$\square$

Remark 2.4.1. The condition (ii) together with (iii) implies that if a matrix A belongs to a *Jordan algebra* the $A^n \in \mathscr{S}$, for $n \geq 1$.

In what follows we establish the relationship between arbitrary sets of real symmetric matrices and certain *Jordan algebra* derived from such sets.

Definition 2.4.8. Let $\$$ be a subspace of $\mathscr{S}^n$ spanned by any arbitrary set of matrices $\{M_1, \ldots, M_k\}$, with $M_i \in \mathscr{S}^n$, and suppose $\mathscr{S}^n$ has an *identity element I*. We define the following:

1. $\mathscr{A} = \mathscr{A}(\$)$ stands for the *smallest associative algebra* in $M_{n,n}$ containing $\$$.

2. $\mathscr{B} = \mathscr{B}(\$) = \{B \in \$: SBS \in \$, \forall S \in \$\}$.

3. $\mathscr{L} = \mathscr{L}(\$) \subseteq \mathscr{S}^n$ stands for the *smallest Jordan algebra* in $\mathscr{S}^n$ that containing $\$$.

Theorem 2.4.4.

(a) *Given $\$$, $\mathscr{B}$ is the maximal subspace of $\$$ such that*

$$BSB \in \mathscr{B}, \ \forall S \in \$, \forall S \in B \in \mathscr{B},$$

and is finite dimensional formally real special Jordan algebra.

(b) *Given $\$, \mathscr{A}, \mathscr{B}$, and $\mathscr{L}$, it holds*

$$\mathscr{B} \subseteq \$ \subseteq \mathscr{L} \subseteq \mathscr{A} \cap \mathscr{S}^n \subseteq \mathscr{A}.$$

Proof. See Malley [43]. $\square$

In what follows, one presents some algebraic results established until now and that have application to the study of random quadratic forms, by constructing an unique basis constituted by *mutual orthogonal projection matrices* for *commutative Jordan algebra*.

Definition 2.4.9. Let $\mathscr{S} = \bigoplus_{i=1}^s S_i$ be a subset of $\mathscr{S}^n$. The support of $A \in \mathscr{S}$, with $A = \bigoplus_{i=1}^s A_i$, $A_i \in \mathscr{S}$, is the set $\{i \in \mathbb{N} : A_i \neq 0\}$.

Theorem 2.4.5. *Let $\mathscr{S} \subseteq \mathscr{S}^n$ and $I \in \mathscr{S}$. Then, for any $A, B \in \mathscr{B}$, the following are equivalent:*

(i) $\forall S \in \mathscr{S}$, *it holds* $SASBS = \mathbf{0}_{n,n}$.

(ii) $A \mathscr{S} B = \mathbf{0}_{n,n}$.

(iii) $A\mathscr{B}B = \mathbf{0}_{n,n}$.

(iv) A and B have disjoint support in $\mathscr{B}$.

Theorem 2.4.6. Let $\mathscr{S} \subseteq \mathscr{S}^n$ and $I \in \mathscr{S}$. Then, for any $A, B \in \mathscr{L}$, the following are equivalent:

(i) $\forall S \in \mathscr{L}$, it holds $SASBS = \mathbf{0}_{n,n}$.

(ii) $A\mathscr{S}B = \mathbf{0}_{n,n}$.

(iii) $A\mathscr{L}B = \mathbf{0}_{n,n}$.

(iv) A and B have disjoint support in $\mathscr{L}$.

Theorem 2.4.7. Let $A, B \in \mathscr{S}^n$. Then, $A\mathscr{S}B = \mathbf{0}_{n,n} \Longleftrightarrow A\mathscr{L}B = \mathbf{0}_{n,n}$.

Proof. See Malley [43]. □

Now, we give some results connecting the *commutative Jordan algebra* and the *symmetric matrices spaces* for the purpose of the work developed in latter chapters. The first one (Theorem 2.4.8) is a consequence of the Lemma 1 and Lemma 2 of Seely [68].

Theorem 2.4.8. *For every commutative Jordan algebra, there exists at least one basis*

$$\{Q_1, ..., Q_s\}$$

constituted by projection matrices Q_i, $i \in \{1, \ldots, s\}$ *such that* $Q_iQ_j = \mathbf{0}_{n,n}$, $i \neq j$, $i, j \in \{1, \ldots, s\}$.

Proof. See Seely [68]. □

Theorem 2.4.9. *A subspace* $S \subseteq \mathscr{S}^n$ *is a commutative Jordan algebra if and only if there exists a basis* $\{Q_1, \ldots, Q_s\}$ *formed by orthogonal projection matrices, such that* $Q_iQ_j = \mathbf{0}_{n,n}$, $i \neq j$, $i, j \in \{1, \ldots, s\}$. *Moreover, such a basis is unique.*

Proof. The existence is proved by Theorem 2.4.8. To prove the uniqueness, let $\{P_1, \ldots, P_s\}$ be another basis for S such that P_i, $i \in \{1, \ldots, s\}$, are *orthogonal projection matrices* and $P_iP_j = \mathbf{0}_{n,n}$, $i \neq j$, $i, j \in \{1, \ldots, s\}$. Let the coefficients $\alpha_1, \ldots, \alpha_s$ and $\beta_{n,1}, \ldots, \beta_{n,s}$ be unique such that

$$P_n = \sum_{i=1}^{s} \alpha_i Q_i, \ n \in \{1, \ldots, s\}, \text{ and } Q_t = \sum_{j=1}^{s} \beta_{t,j} P_j, \ t \in \{1, \ldots, s\}.$$

Now one may note that

$$P_nQ_t = \left(\sum_{i=1}^{s} \alpha_i Q_i\right) Q_t = \alpha_t Q_t Q_t = \alpha_t Q_t, \tag{2.25}$$

since $Q_iQ_j = 0$, $i \neq j, i, j \in \{1, \ldots, s\}$, and $Q_1, \ldots, Q_s$ are *orthogonal projection matrices*.

On the other hand

$$P_nQ_t = P_n\left(\sum_{j=1}^{s} \beta_{t,j} P_j\right) = \beta_{t,n} P_n P_n = \beta_{t,n} P_n, \tag{2.26}$$

since $P_i P_j = 0$, $i \neq j, i, j \in \{1, \ldots, s\}$, and $P_1, \ldots, P_s$ are *orthogonal projection matrices*. Hence,

$$\beta_{t,n} P_n = P_n Q_t = \alpha_t Q_t, \ \ t \in \{1, \ldots, s\}.$$

Thus, since $Q_1, \ldots, Q_s$ are linearly independent and the $\alpha_1, \ldots, \alpha_s$ and $\beta_{n,1}, \ldots, \beta_{n,s}$ must be equal to one or zero, it holds $P_n = Q_t$, for some $t \in \{1, \ldots, s\}$. This result holds for every $n \in \{1, \ldots, s\}$, so that the proof is established. $\square$

2.5 Kronecker Product

Since the matrices involved in analysis of variance related to a statistical model mostly possess a particular type of structure that allows them to be expressed as the Kronecker product of other matrices with well suited structure, this product play an important role in statistics field.

Thus we introduce next the notion of Kronecker product $\otimes$ as well as some of its basic properties (see Schott [64] or Rao and Rao [61]). At the end of this section we generalize the results in the Theorem 7.6. (d, e) and the Theorem 7.7. of Schott [64] (see Proposition 2.5.4).

Definition 2.5.1. Given the matrices $A \in \mathscr{M}^{m \times n}$ and $B \in \mathscr{M}^{p \times q}$, the *Kronecker product* of A and B, denoted by $A \otimes B$, is defined by

$$A \otimes B = \begin{bmatrix} a_{11}B & a_{12}B & \ldots & a_{1n}B \\ a_{21}B & a_{22}B & \ldots & a_{2n}B \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1}B & a_{m2}B & \ldots & a_{mn}B \end{bmatrix} \in \mathscr{M}^{mp \times nq}. \tag{2.27}$$

We list some remarkable properties of this operation in the following three nex theorems. The proofs of some of the properties in the two first ones stem directly from its definition. For the other properties we recommend Rao and Rao [61] or Schott [64].

Theorem 2.5.1. *Let A, B, and C be any matrices; a and b be any two vectors, and α and β be any two scalars. Then*

(a) $ab^\top = a \otimes b^\top = b^\top \otimes a;$

(b) $\alpha \otimes A = A \otimes \alpha = \alpha A;$

(c) $(\alpha A) \otimes (\beta B) = \alpha \beta (A \otimes B).$

Theorem 2.5.2. *Let $A \in \mathscr{M}^{n \times m}$, $B \in \mathscr{M}^{p \times q}$, and $C \in \mathscr{M}^{r \times s}$. Then*

(a) $(A \otimes B)^\top = A^\top \otimes B^\top;$

(b) $(A \otimes B) \otimes C = A \otimes (B \otimes C);$

(c) $(A + B) \otimes C = (A \otimes C) + (B \otimes C)$, *if $n = p$ and $m = q$;*

(d) $A \otimes (B + C) = (A \otimes B) + (A \otimes C)$, *if $p = r$ and $q = s$;*

(e) $(A \otimes B)^+ = A^+ \otimes B^+$.

Theorem 2.5.3. *Let $A \in \mathscr{M}^{n \times m}$, $B \in \mathscr{M}^{p \times q}$, $C \in \mathscr{M}^{m \times s}$, and $D \in \mathscr{M}^{q \times t}$. Then*

$$(A \otimes B)(C \otimes D) = AC \otimes BD.$$

Proof. See Rao and Rao [61], $P6.1.1(4)$, or Schott [64], Theorem 7.7. □

We give the generalization of Theorem 2.5.3 in the next result.

Proposition 2.5.4. *Let $A_j \in \mathscr{M}^{m_j \times h_j}$, $B_j \in \mathscr{M}^{h_j \times q_j}$, $C_a \in \mathscr{M}^{o \times p}$, and $D_b \in \mathscr{M}^{u \times v}$, $j = 1, \ldots, r > 2$, $a = 1, \ldots, s$, $b = 1, \ldots, t$, with s and t positive integers. Then,*

(a)
$$\left(\bigotimes_{j=1}^{r} A_j \right) \left(\bigotimes_{i=1}^{r} B_i \right) = \left(\bigotimes_{j=1}^{r} A_j B_j \right) \in \mathscr{M}^{m \times q},$$

where $m = \prod_{j=1}^{r} m_j$ and $q = \prod_{j=1}^{r} q_j$;

(b)
$$\left(\sum_{a=1}^{s} C_a \right) \otimes \left(\sum_{b=1}^{t} D_b \right) = \sum_{a=1}^{s} \sum_{b=1}^{t} C_a \otimes D_b$$

For the case $r = 2$ see Schott [64], Theorem 7.7.

Proof. We proof part *(a)*.

Due to associativity of the *kronecker product* and according with the Theorem 2.5.3, we will have the following:

$$\left(\bigotimes_{j=1}^{r} A_j \right) \left(\bigotimes_{i=1}^{r} B_i \right) = (A_1 B_1) \otimes (E_1 F_1), \tag{2.28}$$

where $E_i = \bigotimes_{j=i+1}^{r} A_j$ and $F_i = \bigotimes_{j=i+1}^{r} A_j$, with $i = 1, \ldots, r-2$. Now if we repeat the process in (2.28) $r - 2$ times (restarting now with $E_1 F_1$) we will have

$$
\begin{aligned}
\left(\bigotimes_{j=1}^{r} A_j \right) \left(\bigotimes_{i=1}^{r} B_i \right) &= (A_1 B_1) \otimes (E_1 F_1) \\
&= (A_1 B_1) \otimes (A_2 B_2) \otimes (E_2 F_2) \\
&\quad \ldots \\
&= (A_1 B_1) \otimes \ldots \otimes (A_{r-2} B_{r-2}) \otimes E_{r-2} F_{r-2}.
\end{aligned}
$$

Thus, the proof will be complete if we note that

$$
\begin{aligned}
E_{r-2} F_{r-2} &= (A_{r-1} \otimes A_r)(B_{r-1} \otimes B_r) \\
&= (A_{r-1} B_{r-1}) \otimes (A_r B_r).
\end{aligned}
$$

The statement $\left(\bigotimes_{j=1}^{r} A_j B_j \right) \in \mathscr{M}^{m \times q}$ becomes clear if we observe that $A_i B_i \in \mathscr{M}^{m_j \times q_j}$.

The proof for the part *(b)* is straightforward if we use the Theorem 2.5.2. □

The following theorem address the relationship of the eigenvalues of the *Kronecker product* of two matrices and the eigenvalues of each one of them.

Theorem 2.5.5. *Let $\lambda_1,\ldots,\lambda_n$ be the eigenvalues of $A \in \mathcal{M}^{n \times n}$ and $\alpha_1,\ldots,\alpha_m$ be the eigenvalues of $B \in \mathcal{M}^{m \times m}$. Then, eigenvalues of $A \otimes B$ will be $\lambda_i \alpha_j$, $i = 1,\ldots,n$ and $j = 1,\ldots,m$.*

Proof. See Schott [64], Theorem 7.10, or Rao and Rao [61], P.6.1.2. □

The next theorem identifies the relationship between the determinant of the Kronecker product of two matrices and the determinants of each one of them, as well as the relationship between the rank of that product of two matrices and the rank of each matrix.

Theorem 2.5.6. *Let $A \in \mathcal{M}^{n \times m}$ and $B \in \mathcal{M}^{p \times q}$. Then*

(a) $r(A \otimes B) = r(A)r(B)$;

(b) $|A \otimes B| = |A|^p |B|^n$, if $n = m$ and $p = q$.

Proof.

(a): By Theorem 2.1.5 together with Theorems 2.5.2 (c) and 2.5.3 we have that

$$r(A \otimes B) = r((A \otimes B)A \otimes B^\top) = r(AA^\top \otimes BB^\top).$$

Since $AA^\top \otimes BB^\top$ is symmetric, $r(AA^\top \otimes BB^\top)$ is the number of nonzero eigenvalues of $AA^\top \otimes BB^\top$ (see Theorem 2.1.8). Let $\lambda_1,\ldots,\lambda_n$ be the eigenvalues of $AA^\top$ and $\alpha_1,\ldots,\alpha_p$ be the eigenvalues of $BB^\top$. Then, by the Theorem 2.5.5, the eigenvalues of $AA^\top \otimes BB^\top$ will be $\lambda_i \alpha_j$, $i = 1,\ldots,n$ and $j = 1,\ldots,p$. Thus, since $\lambda_i \alpha_j = 0$ if and only if $\lambda_i = 0$ or $\alpha_j = 0$, the number of nonzero eigenvalues of $AA^\top \otimes BB^\top$ will be the number of nonzero eigenvalues of $AA^\top$ times the number of nonzero eigenvalues of $BB^\top$. Now, and finally, since $AA^\top$ and $BB^\top$ are symmetric matrices, the number of nonzero eigenvalues of $AA^\top$ and $BB^\top$ are given by $r(AA^\top)$ and $r(BB^\top)$, respectively.

(b): According with Theorem 2.5.5 (d), we have that

$$|A| = \prod_{i=1}^{n} \lambda_i \text{ and } |B| = \prod_{j=1}^{p} \alpha_j.$$

Now, since the eigenvalues of $A \otimes B$ are $\lambda_i \alpha_j$ (see Theorem 2.5.5), we have that

$$
\begin{aligned}
|A \otimes B| &= \prod_{i=1}^{n}\prod_{j=1}^{p} \lambda_i \alpha_j = \prod_{i=1}^{n} \lambda_i^p \left(\prod_{j=1}^{p} \alpha_j \right) = \prod_{i=1}^{n} \lambda_i^p |B| \\
&= |B|^n \prod_{i=1}^{n} \lambda_i^p = |B|^n \left(\prod_{i=1}^{n} \lambda_i \right)^p \\
&= |B|^n |A|^p,
\end{aligned}
$$

as wished. □

MIXED LINEAR MODELS

Traditionally, statistical models (designs) have been associated with fixed effect models in a given linear model involving one factor with k levels defining groups, referred to as predictor, and n_i independent sampling units in each group i, $i = 1, \ldots, k$, referred to as residual errors, which can be written as the following scalar equation:

$$y_{ij} = \mu + \alpha_i + \varepsilon_{ij}, \ i = 1, \ldots, k; \ j = 1, \ldots, n_i, \tag{3.1}$$

where μ and $\{\alpha_i\}$ are fixed and unknown finite constants which characterized the model means, and $\{\varepsilon_{ij}\}$ the independent random residual errors with mean zero and variance σ_ε^2. It is often assumed that the errors are normal distributed, that is $\varepsilon_{ij} \sim \mathcal{N}\left(0, \ \sigma_\varepsilon^2\right)$.

In matrix notation the model (3.1) can be written as

$$y = X\nu + \varepsilon, \tag{3.2}$$

where

$$
y = \begin{bmatrix} y_{11} \\ \cdots \\ y_{1n_1} \\ y_{21} \\ \cdots \\ y_{2n_2} \\ \cdots \\ \cdots \\ y_{k1} \\ \cdots \\ y_{kn_k} \end{bmatrix}, \ \varepsilon = \begin{bmatrix} \varepsilon_{11} \\ \cdots \\ \varepsilon_{1n_1} \\ \varepsilon_{21} \\ \cdots \\ \varepsilon_{2n_2} \\ \cdots \\ \cdots \\ \varepsilon_{k1} \\ \cdots \\ \varepsilon_{kn_k} \end{bmatrix}, \ X = \begin{bmatrix} 1 & 1 & 0 & \ldots & \ldots & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ 1 & 1 & 0 & \ldots & \ldots & 0 \\ 1 & 0 & 1 & 0 & \ldots & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ 1 & 0 & 1 & 0 & \ldots & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ 1 & 0 & 0 & \ldots & 0 & 1 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ 1 & 0 & 0 & \ldots & 0 & 1 \end{bmatrix} \in \mathcal{M}^{\left(\Sigma_{s=1}^k n_s\right) \times (k+1)},
$$

and $v^\top = \begin{bmatrix} \mu & \alpha_1 & \dots & \alpha_k \end{bmatrix}$, where ε is assumed to have a distribution with mean $\mathbf{0}_{\sum_{s=1}^k n_s}$ and variance-covariance matrix $\sigma_\varepsilon^2 I_{\sum_{s=1}^k n_s}$ or, often, $\varepsilon \sim \mathcal{N}\left(\mathbf{0}_{\sum_{s=1}^k n_s}, \sigma_\varepsilon^2 I_{\sum_{s=1}^k n_s}\right)$. Hence, with $\Sigma(z)$ denoting the variance-covariance matrix of a random vector z, the model (3.2) has a distribution with mean Xv and variance-covariance matrix $\Sigma(y) = \sigma_\varepsilon^2 I_{\sum_{s=1}^k n_s}$.

Due to the necessity of incorporate the amount of variations caused by certain uncontrollable sources in statistical designs with fixed effects, for example, the amount of variations within groups that the experimenter is not able to control and those whose levels must be selected at random, in research fields like as genetic, animal breeding, and quality control and improvement, in early 1960 several designs with both fixed and random effects terms were introduced and widely investigated. Among those designs, nowadays called *mixed linear models* or *linear mixed models*, we highlight the well known and probably most widely discussed *mixed linear model*: "one-way design" (see Khuri [34]), whose algebraic characterization is the one presented in (3.1), but here μ is a fixed and unknown constant characterizing the means, $\{\alpha_i\}$ are the independent effects due the observed y of the i-*th* group, assumed to have a distribution with mean zero and variance σ_α^2, and $\{\varepsilon_{ij}\}$ are the independent random errors, assumed to have a distribution with mean zero and variance σ_ε^2, so that in matrix notation it is written as

$$y = Z\mu + Z_1\alpha + \varepsilon, \tag{3.3}$$

where

$$Z = \mathbf{1}_{\sum_{s=1}^k n_s}, \quad Z_1 = \begin{bmatrix} 1 & 0 & \dots & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots \\ 1 & 0 & \dots & \dots & 0 \\ 0 & 1 & 0 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots \\ 0 & 1 & 0 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & 0 & 1 \\ \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & 0 & 1 \end{bmatrix} \in \mathcal{M}^{\left(\sum_{s=1}^k n_s\right) \times k},$$

$\alpha = \begin{bmatrix} \alpha_1 \\ \dots \\ \alpha_k \end{bmatrix}$ has a distribution with mean $\mathbf{0}_k$ and variance-covariance matrix $\sigma_\alpha^2 I_k$, and y and ε defined as in (3.2). α_i and ε_{ij} are assumed to be mutually independent. Thus, the model (3.3) has a distribution with mean $X\mu$ and variance-covariance matrix given by

$$\Sigma(y) = \sigma_\alpha^2 Z_1 Z_1^\top + \sigma_\varepsilon^2 I_{\sum_{s=1}^k n_s}.$$

The parameters σ_α^2 and σ_ε^2 are refereed to as *variance components*.

The model (3.3) is said to be *balanced* if there is the same number of observations in every groups, that is $n_i = n$ for every groups. Otherwise the model is said to be *unbalanced*. Silva et al. [69] approach the *balanced* "two-way nested model" in the context of tolerance interval studies.

Example 3.0.1. Consider a particular *balanced* design of (3.3) with $k = 3$ and $n_i = 3$, $i = 1, 2, 3$. This model will have mean $\mathbf{1}_9 \otimes \mu$ and variance-covariance

$$\Sigma(y) = I_3 \otimes \left(\sigma_\alpha^2 J_3 + \sigma_\varepsilon^2 I_3 \right).$$

Others widely discussed designs with linear mixed structure are the both "nested (hierarchical)" and "two-way crossed (with interaction or without interaction)" models (see Khuri [34]). We introduce here the "two-way nested models" with mixed linear structure (For the nested design notions, we recommend Anderson and Bancroft [5] and Bainbridge [8], for instance). These models consist of two groups of treatment, A and B say, where the b_i, $i = 1, \ldots, a$, levels of group B are nested within the *i-th* level of group A, so that we write them as

$$
\begin{aligned}
y_{ijk} = \quad & \mu + \alpha_i + \beta_{ij} + \varepsilon_{ijk}, \\
& i = 1, \ldots, a; \ j = 1, \ldots, b_i; \ k = 1, \ldots, n_{ij},
\end{aligned}
\tag{3.4}
$$

where μ is the general mean, $\{\alpha_i\}$ the independent random effects due to the *i-th* level of the group A, $\{\beta_{ij}\}$ the independent random effects due to the *j-th* level of the group B nested within the *i-th* level of the group A, and $\{\varepsilon_{ijk}\}$ the independent residual errors associated to the observed value Y_{ijk}. It is assumed that α_i, β_{ij}, and ε_{ijk} are mutually independent. The effects associated with any group are, clearly, the effects that its levels have on the interest response variable.

In matrix notation the model can be written as

$$y = W\mu + W_1\alpha + W_2\beta + \varepsilon, \tag{3.5}$$

where $W = \mathbf{1}_{\sum_i \sum_j n_{ij}}$, y and ε are vectors whose the entries are, respectively, the observed values $\{y_{ijk}\}$ and the random errors $\{\varepsilon_{ijk}\}$, with ε having a distribution with mean $\mathbf{0}_{\sum_i \sum_j n_{ij}}$ and variance-covariance matrix $\sigma_\varepsilon^2 I_{\sum_i \sum_j n_{ij}}$, and, for instance, for a particular *unbalanced* design with $a = 2$, $b_1 = 3$, $b_2 = 2$, $n_{11} = 2$, $n_{12} = 2$, $n_{13} = 3$, $n_{21} = 3$, and $n_{22} = 2$, we will have that

$$
W_1 = \begin{bmatrix}
1 & 0 \\
1 & 0 \\
1 & 0 \\
1 & 0 \\
1 & 0 \\
1 & 0 \\
0 & 1 \\
0 & 1 \\
0 & 1 \\
0 & 1 \\
0 & 1 \\
0 & 1
\end{bmatrix}, \quad
W_2 = \begin{bmatrix}
1 & 0 & 0 & 0 & 0 \\
1 & 0 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 & 0 \\
0 & 0 & 1 & 0 & 0 \\
0 & 0 & 1 & 0 & 0 \\
0 & 0 & 1 & 0 & 0 \\
0 & 0 & 0 & 1 & 0 \\
0 & 0 & 0 & 1 & 0 \\
0 & 0 & 0 & 1 & 0 \\
0 & 0 & 0 & 0 & 1 \\
0 & 0 & 0 & 0 & 1
\end{bmatrix},
$$

α has a distribution with mean $\mathbf{0}_2$ and variance-covariance matrix $\sigma_\alpha^2 I_2$, and β has a distribution with

mean $\mathbf{0}_5$ and variance-covariance matrix $\sigma_\beta^2 I_5$. It is often assumed that $\beta = \begin{bmatrix} \beta_{11} \\ \beta_{12} \\ \beta_{13} \\ \beta_{21} \\ \beta_{22} \end{bmatrix} \sim \mathcal{N}\left(\mathbf{0}_5, \ \sigma_\beta^2 I_5\right),$

$\alpha = \begin{bmatrix} \alpha_1 \\ \alpha_2 \end{bmatrix} \sim \mathcal{N}\left(\mathbf{0}_2, \ \sigma_\alpha^2 I_2\right)$ and $\varepsilon \sim \mathcal{N}\left(\mathbf{0}_{\sum_i \sum_j n_{ij}}, \ \sigma_\varepsilon^2 I_{\sum_i \sum_j n_{ij}}\right)$.

Thus, we have that this particular design of the model (3.5) has a distribution with mean $\mathbf{1}_{\sum_i \sum_j n_{ij}} \mu = \mathbf{1}_{12} \mu$ and variance-covariance matrix given by

$$\begin{aligned} \Sigma(y) &= \sigma_\alpha^2 W_1 W_2^\top + \sigma_\beta^2 W_2 W_2^\top + \sigma_\varepsilon^2 I_{\sum_i \sum_j n_{ij}}, \\ &= I_2 \otimes \left(\sigma_\alpha^2 J_6 + \sigma_\varepsilon^2 I_6\right) + \sigma_\beta^2 \left(\Pi_1 + \Pi_2 \otimes J_2\right), \end{aligned}$$

where $\Pi_1 = \begin{bmatrix} \mathbf{0}_{4,4} & \mathbf{0}_{4,3} & \mathbf{0}_{4,3} & \mathbf{0}_{4,2} \\ \mathbf{0}_{3,4} & J_3 & \mathbf{0}_{3,3} & \mathbf{0}_{3,2} \\ \mathbf{0}_{3,4} & \mathbf{0}_{3,3} & J_3 & \mathbf{0}_{3,2} \\ \mathbf{0}_{2,4} & \mathbf{0}_{2,3} & \mathbf{0}_{2,3} & \mathbf{0}_{2,2} \end{bmatrix}$ and $\Pi_2 = \begin{bmatrix} I_2 & \mathbf{0}_{2,3} & \mathbf{0}_2 \\ \mathbf{0}_{3,2} & \mathbf{0}_{3,3} & \mathbf{0}_3 \\ \mathbf{0}_2^\top & \mathbf{0}_3^\top & 1 \end{bmatrix}.$

We have thus three *variance components* to be estimated: σ_α^2, σ_β^2, and σ_ε^2.

The "two - way crossed models" with mixed effects structure is defined as

$$\begin{aligned} y_{ijk} &= \mu + \tau_i + \beta_j + \gamma_{ij} + \varepsilon_{ijk}, && (3.6) \\ & i = 1,\ldots,a, \ j = 1,\ldots,b, \ k = 1,\ldots,n_{ij}, \end{aligned}$$

consisting of two groups of treatments, C and D say, where μ is the general mean, $\{\tau_i\}$ the random effects due to the group C, assumed to be normal distributed with mean zero and variance σ_τ^2, $\{\beta_j\}$ the random effects due to the group D, assumed to be normal distributed with variance σ_β^2, $\{\gamma_{ij}\}$ the random effects due to interaction of the *i-th* level of the group C with the *j-th* level of the group D, assumed to be normal distributed with mean zero and variance σ_γ^2, and $\{\varepsilon_{ijk}\}$ the independent residual errors, assumed to be normal distributed with mean zero and variance σ_ε^2. It is assumed that α_i, β_{ij}, and ε_{ijk} are mutually independent. For models with no interaction $\{\gamma_{ij}\}$ are taken to be all nulls.

In matrix notation, the model (3.6) can be written as

$$y = M\mu + M_1 \tau + M_2 \beta + M_3 \gamma + \varepsilon, \qquad (3.7)$$

where $M = \mathbf{1}_{\sum_i \sum_j n_{ij}}$, y and ε are random vectors whose the entries are, respectively, the observed values y_{ijk} and the random errors ε_{ijk} so that ε has a distribution with mean $\mathbf{0}_{\sum_i \sum_j n_{ij}}$ and variance-covariance matrix $\sigma_\varepsilon^2 I_{\sum_i \sum_j n_{ij}}$ (often assumed $\varepsilon \sim \mathcal{N}(\mathbf{0}_{\sum_i \sum_j n_{ij}}, \ \sigma_\varepsilon^2 I_{\sum_i \sum_j n_{ij}})$), and, for instance, for

a particular design with $a = 2$ and $b = 3$ we have that $M_1 = \begin{bmatrix} \Delta_{11}^1 \\ \Delta_{12}^1 \\ \Delta_{13}^1 \\ \Delta_{21}^2 \\ \Delta_{22}^2 \\ \Delta_{23}^2 \end{bmatrix}$, $M_2 = \begin{bmatrix} \Lambda_{11}^1 \\ \Lambda_{12}^2 \\ \Lambda_{13}^3 \\ \Lambda_{21}^1 \\ \Lambda_{22}^2 \\ \Lambda_{23}^3 \end{bmatrix}$, $M_3 = \begin{bmatrix} O_{11}^1 \\ O_{12}^2 \\ O_{13}^3 \\ O_{21}^4 \\ O_{22}^5 \\ O_{23}^6 \end{bmatrix}$,

τ, β and γ have respectively distribution with mean $\mathbf{0}_2$, $\mathbf{0}_3$ and $\mathbf{0}_6$ and variance-covariance matrix $\sigma_\tau^2 I_2$, $\sigma_\beta^2 I_3$ and $\sigma_\gamma^2 I_6$, with Δ_{ij}^k, $k = 1, 2$, an $n_{ij} \times 2$ matrix whose the column k is a vector of 1's and the remain one is a vector of zeros, Λ_{ij}^j an $n_{ij} \times 3$ matrix whose the column j is a vector of 1's and the remains ones are vectors of zeros, and O_{ij}^s, $s = 1, \ldots, 6$, an $n_{ij} \times 6$ matrix whose the column s is a vector of 1's and the remains ones are vectors of zeros. It is oftem assumed

$$\tau = \begin{bmatrix} \tau_1 \\ \tau_2 \end{bmatrix} \sim \mathcal{N}\left(\mathbf{0}_2,\ \sigma_\tau^2 I_2\right),\ \beta = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \end{bmatrix} \sim \mathcal{N}\left(\mathbf{0}_3,\ \sigma_\beta^2 I_3\right) \text{ and } \gamma = \begin{bmatrix} \gamma_1 \\ \ldots \\ \gamma_6 \end{bmatrix} \sim \mathcal{N}\left(\mathbf{0}_6, \sigma_\gamma^2 I_6\right).$$

Thus the model (3.7) has a distribution with mean $M\mu$ and variance-covariance matrix given by

$$\Sigma(y) = \sigma_\tau^2 M_1 M_1^\top + \sigma_\beta^2 M_2 M_2^\top + \sigma_\gamma M_3 M_3^\top + \sigma_\varepsilon^2 I_{\Sigma_i \Sigma_j n_{ij}},$$

where

$$M_1 M_1^\top = \begin{bmatrix} J_{\sum_{j=1}^3 n_{1j}} & \mathbf{0}_{\sum_{j=1}^3 n_{1j}, \sum_{j=1}^3 n_{2j}} \\ \mathbf{0}_{\sum_{j=1}^3 n_{2j}, \sum_{j=1}^3 n_{1j}} & J_{\sum_{j=1}^3 n_{2j}} \end{bmatrix},$$

$$M_2 M_2^\top = \begin{bmatrix} J_{n_{11}} & \mathbf{0}_{n_{11},n_{12}} & \mathbf{0}_{n_{11},n_{13}} & J_{n_{11},n_{21}} & \mathbf{0}_{n_{11},n_{22}} & \mathbf{0}_{n_{11},n_{23}} \\ \mathbf{0}_{n_{12},n_{11}} & J_{n_{12}} & \mathbf{0}_{n_{12},n_{13}} & \mathbf{0}_{n_{12},n_{21}} & J_{n_{12},n_{22}} & \mathbf{0}_{n_{12},n_{23}} \\ \mathbf{0}_{n_{13},n_{11}} & \mathbf{0}_{n_{13},n_{12}} & J_{n_{13}} & \mathbf{0}_{n_{13},n_{21}} & \mathbf{0}_{n_{13},n_{22}} & J_{n_{13},n_{23}} \\ J_{n_{21},n_{11}} & \mathbf{0}_{n_{21},n_{12}} & \mathbf{0}_{n_{21},n_{13}} & J_{n_{21}} & \mathbf{0}_{n_{21},n_{22}} & \mathbf{0}_{n_{21},n_{23}} \\ \mathbf{0}_{n_{22},n_{11}} & J_{n_{22},n_{12}} & \mathbf{0}_{n_{22},n_{13}} & \mathbf{0}_{n_{22},n_{21}} & J_{n_{22}} & \mathbf{0}_{n_{22},n_{23}} \\ \mathbf{0}_{n_{23},n_{11}} & \mathbf{0}_{n_{23},n_{12}} & J_{n_{23},n_{13}} & \mathbf{0}_{n_{23},n_{21}} & \mathbf{0}_{n_{23},n_{22}} & J_{n_{23}} \end{bmatrix}$$

$$M_3 M_3^\top = \begin{bmatrix} J_{n_{11}} & \mathbf{0}_{n_{11},n_{12}} & \mathbf{0}_{n_{11},n_{13}} & \mathbf{0}_{n_{11},n_{21}} & \mathbf{0}_{n_{11},n_{22}} & \mathbf{0}_{n_{11},n_{23}} \\ \mathbf{0}_{n_{12},n_{11}} & J_{n_{12}} & \mathbf{0}_{n_{12},n_{13}} & \mathbf{0}_{n_{12},n_{21}} & \mathbf{0}_{n_{12},n_{22}} & \mathbf{0}_{n_{12},n_{23}} \\ \mathbf{0}_{n_{13},n_{11}} & \mathbf{0}_{n_{13},n_{12}} & J_{n_{13}} & \mathbf{0}_{n_{13},n_{21}} & \mathbf{0}_{n_{13},n_{22}} & \mathbf{0}_{n_{13},n_{23}} \\ \mathbf{0}_{n_{21},n_{11}} & \mathbf{0}_{n_{21},n_{12}} & \mathbf{0}_{n_{21},n_{13}} & J_{n_{21}} & \mathbf{0}_{n_{21},n_{22}} & \mathbf{0}_{n_{21},n_{23}} \\ \mathbf{0}_{n_{22},n_{11}} & \mathbf{0}_{n_{22},n_{12}} & \mathbf{0}_{n_{22},n_{13}} & \mathbf{0}_{n_{22},n_{21}} & J_{n_{22}} & \mathbf{0}_{n_{22},n_{23}} \\ \mathbf{0}_{n_{23},n_{11}} & \mathbf{0}_{n_{23},n_{12}} & \mathbf{0}_{n_{23},n_{13}} & \mathbf{0}_{n_{23},n_{21}} & \mathbf{0}_{n_{23},n_{22}} & J_{n_{23}} \end{bmatrix}.$$

The parameters σ_τ^2, σ_β^2, σ_γ^2 and σ_ε^2 are the *variance components*.

The major innovation here is that the *mixed linear models* with its structures incorporating the expected values and the variance-covariance matrix specified as a function of a finite number of parameters constitutes an useful tool for modeling mistimed or irregularly timed data, and missing observations result in incomplete data, even those who come in small sample or in longitudinal set (Wallace and Helms [27] developed procedures that provide hypothesis tests and confidence intervals for these kind of data using the *mixed linear model*). The fixed effect and random effects models may provide well-behaved estimation but may have difficulty providing completely accurate inference in small samples. See Muller and Stewart [52] for more explanation. Among a selective books covering this mater we could, for instance, suggest Demidenko [16], Rencher and Schaalje [62].

Letting $y \in \mathbb{R}^n$ denotes the vector of responses (observed data), the *mixed linear models* can be expressed as

$$y = X\beta + \sum_{i=1}^{n+1} X_i \beta_i, \tag{3.8}$$

with $X_{n+1} = I_n$, $X_i \in \mathcal{M}^{n_i \times p_i}$ and $\beta_i \in \mathbb{R}^{p_i}$, where $X \in \mathcal{M}^{n \times p}$ is the known (possible non full-rank) design matrix for the fixed effects, $\beta \in \mathbb{R}^p$ the vector of the (unknown) fixed expected values, $X_i, i = 1, \ldots, n+1$, the known and fixed (full-rank) design matrices for the variance-covariance structure, $\beta_i, i = 1, \ldots, n$, the unknown vectors of the unobservable random effects, and β_{n+1} the vector of the residual errors. The model includes the following reasonable assumptions (see McCulloch and Searle [44]): $\beta_i, i = 1, \ldots, n+1$, are mutually independent random vectors such that $E(\beta_i) = \mathbf{0}_{p_i}$ and $\Sigma(\beta_i) = \sigma_i^2 I_{p_i}$, so that

$$E(y) = X\beta \quad \text{and}$$
$$\Sigma(y) = \sum_{i=1}^{n+1} \sigma_i^2 M_i,$$

where $M_i = X_i X_i^\top$, and $\sigma_1^2, \ldots, \sigma_{n+1}^2$ are unknown and fixed positive parameters referred to as *variance components*, verifying $\sigma_i^2 \geq 0$, $i = 1, \ldots, n$, and $\sigma_{n+1}^2 > 0$. The estimation of these parameters is the major goal of this work. The vectors of the unobserved random effects $\beta_i, i = 1, \ldots, n+1$, are often taken to be normal distributed, but in this work we only require them to have second moment.

Thus, since there is no distribution assumed for the model (3.8), we will denote it as

$$y \sim (X\beta, \Sigma), \text{ where } \Sigma = \Sigma(y), \tag{3.9}$$

i.e., y is distributed with expectation $X\beta$ and variance-covariance matrix Σ.

As pointed out earlier, the proliferation of research on *mixed linear models* leaded to the development of several methods of estimation for the *variance components*; highlighting the *ANOVA-based*, *Maximum likelihood-based*, and the *OBS-based* methods (see, for example, Searle et al. [67], Casella and Berger [14], and Calinski and Kageyama ([12], [13]). See Hocking [29] for estimation with *ANOVA-based* and *Maximum likelihood-based* methods, and Nelder ([57], [58]) for OBS - based method.

The next three sections are devoted to the introduction of these methods, starting with the *ANOVA-based* (Section 3.1) followed by the *Maximum likelihood-based* (Section 3.2), and finally the *OBS-based* (Section 3.3).

3.1 Variance Components Estimation - ANOVA Method

The ANOVA - based method is one of the most common procedure for the estimation of *variance components*. Among its many approach we highlight the one suggested by Henderson (see Henderson [28] for explanation) through is three variations known as method 1, method 2, and method 3, especially because of its simplicity in what concern the computational implementation (even on a hand-held calculator), and unbiasedness, properties saved by all ANOVA-based methods. All such methods have the common underlying idea: equating the (observed) quadratic errors for the different sources of variations to their respective expected values (in some case with some readjustment), leading to a system of linear equations, which solved for the *variance components* gives the corresponding estimators.

Let S_i^2, $i = 1, \ldots, n+1$, denote the quadratic error for the ith source of variation in the model (3.8). Then, the quadratic errors in the different sources may be given as $S_i^2 = y^\top P_i y$, where $P_i \in \mathscr{S}^n$ is such that $X^\top P_i X = \mathbf{0}_{p,p}$, and so, since (see Schott [64], Theorems 9.18. and 1.3)

$$
\begin{aligned}
E\left(S_i^2\right) &= tr(P_i \Sigma) + (X\beta)^\top P_i (X\beta) \\
&= tr\left(\sum_{j=1}^{n+1} \gamma_j P_i M_j\right) \\
&= \sum_{j=1}^{n+1} \gamma_j tr\left(X_j^\top P_i X_j\right),
\end{aligned}
\tag{3.10}
$$

the expected value of S_i^2 will depend only on the *variance components*.

With $S = \begin{bmatrix} S_1^2 \\ \ldots \\ S_{n+1}^2 \end{bmatrix}$ and $\gamma = \begin{bmatrix} \gamma_1 \\ \ldots \\ \gamma_{n+1} \end{bmatrix}$, we will have that

$$
E(S) = C\gamma, \text{ where } C = \begin{bmatrix} tr(X_1^\top P_1 X_1) & \ldots & tr(X_{n+1}^\top P_1 X_{n+1}) \\ \vdots & \ldots & \vdots \\ tr(X_1^\top P_{n+1} X_1) & \ldots & tr(X_{n+1}^\top P_{n+1} X_{n+1}) \end{bmatrix}.
$$

Thus, equating S to $E(S)$, i.e., $S = C\gamma$, it holds

$$
\hat{\gamma} = C^{-1} S,
\tag{3.11}
$$

provided C is squared and non-singular.

C is a square matrix once the number of sources of variations equals the number of *variance components*. For the situation in which there is more sources of variations than *variance components*, it might be used one of the variation of the ANOVA-based estimator:

$$
\tilde{\gamma} = (C^\top C)^{-1} C^\top S,
$$

the least square one, provided C is of full-rank. Clearly, both $\hat{\gamma}$ and $\tilde{\gamma}$ are unbiased; indeed,

$$
\begin{aligned}
E(\hat{\gamma}) &= C^{-1} E(S) = C^{-1} C\gamma = \gamma; \\
E(\tilde{\gamma}) &= (C^\top C)^{-1} C^\top E(S) = (C^\top C)^{-1} C^\top C\gamma = \gamma.
\end{aligned}
$$

Example 3.1.1. Lets consider the following unbalanced "one-way design" from the model (3.3):

$$
\begin{aligned}
y_{ij} &= \mu + \alpha_i + e_{ij}, \\
&i = 1,2,3; \ j = 1,\ldots,n_i; \ n_i = i+1.
\end{aligned}
$$

In matrix notation it becomes

$$
y = X\mu + X_1\alpha + X_2 e,
$$

where $X = \mathbf{1}_9$, $X_1 = \begin{bmatrix} \mathbf{1}_2 & \mathbf{0}_2 & \mathbf{0}_2 \\ \mathbf{0}_3 & \mathbf{1}_3 & \mathbf{0}_3 \\ \mathbf{0}_4 & \mathbf{0}_4 & \mathbf{1}_4 \end{bmatrix}$, $X_2 = I_9$, $\alpha^\top = [\alpha_1 \ \alpha_2 \ \alpha_3]$, and $y, e \in \mathbb{R}$. Thus, the variance-covariance matrix of y, $\Sigma(y)$, will be $\Sigma(y) = \gamma_1 M_1 + \gamma_2 I_9$, where

$$M_1 = X_1 X_1^\top = \begin{bmatrix} J_{n_1} & \mathbf{0}_{n_1,n_2} & \mathbf{0}_{n_1,n_3} \\ \mathbf{0}_{n_2,n_1} & J_{n_2} & \mathbf{0}_{n_2,n_3} \\ \mathbf{0}_{n_3,n_1} & \mathbf{0}_{n_3,n_2} & J_{n_3} \end{bmatrix}.$$

Now, for the two source of variation let SS_B and SS_W respectively denotes the between groups sum of squares and the within groups sum of squares, having therefore $S_1^2 = SS_B$ and $S_2^2 = SS_W$. In matrix notation (see Searle et al. [67]) we have that $S_1^2 = y^\top P_1 y$ and $S_2^2 = y^\top P_2 y$, where

$$P_1 = M_1 - J_N \text{ and } P_2 = I_N - M_1, \tag{3.12}$$

with $N = \sum_i^3 n_i = 9$.

Finally, with $C = \begin{bmatrix} tr(X_1 P_1 X_1) & tr(P_1) \\ tr(X_1 P_2 X_1) & tr(P_2) \end{bmatrix}$ and $S = \begin{bmatrix} S_1^2 \\ S_2^2 \end{bmatrix}$, we find that the ANOVA - based estimator for $\gamma^\top = \begin{bmatrix} \gamma_1 \\ \gamma_2 \end{bmatrix}$ is $\hat{\gamma} = \begin{bmatrix} \hat{\gamma}_1 \\ \hat{\gamma}_2 \end{bmatrix}$, where $\hat{\gamma} = C^{-1} S$, having therefore

$$\hat{\gamma}_2 = \frac{S_2^2}{N-3} \text{ and } \hat{\gamma}_1 = \frac{\frac{S_1^2}{2} - \frac{S_2^2}{N-3}}{\frac{1}{2}\left(N - \frac{1}{N}\sum_{i=1}^3 n_i^2\right)}.$$

The ANOVA - based estimators are useful tools when the estimation process involves repeating the experiments, since been unbiased means among all repetition it is expected that the estimated value equals the true value. The problem arises, for example, when the estimation process requires large amount of data or the data collection process is not so easy. On this situation the repeating process will not be so practical, so that the unbiasedness might be adjudicated in favor of other estimators with better performances in those kind of data. For more details, see, for instance, Searle et al. [67], among many other references.

Nevertheless, notwithstanding the *ANOVA-based* method adapt readily to mixed models with balanced data and save the unbiasedness, it does not adapt in situation with unbalanced data, mostly because it uses computations derived from fixed effect models rather than mixed models.

3.2 Variance Components Estimation - Likelihood Aproach

Adding the Gaussian assumption the model (3.8) (equivalently (3.9)) may be expressed as

$$y \sim \mathcal{N}(X\beta, \Sigma). \tag{3.13}$$

This assumption will allow to carry maximum likelihood estimation from the data. For an overview on likelihood approach we recommend, among other references, Harville [26] for a comprehensive review of the estimation procedures along with computational techniques; Fairclough and Helms [18] and Andrade and Helms [7] which explored the ML estimation procedures for the linear mixed models; and Lair and Ware [37] who discussed the REML estimation relationship to variance components estimation.

3.2.1 ML - Based Method

The likelihood function of the random vector y in model (3.13) is given by

$$L(\beta, \gamma) = (2\pi)^{-\frac{n}{2}} |\Sigma|^{\frac{-1}{2}} e^{\left\{ \frac{-1}{2}(y - X\beta)^\top \Sigma^{-1}(y - X\beta) \right\}}. \tag{3.14}$$

The maximum likelihood estimation of the *variance components* from the available data is achieved by maximizing the logarithm of the likelihood function,

$$
\begin{aligned}
l(\beta, \gamma) &= \log[L(\beta, \gamma)] \\
&= -\frac{n}{2} \log(2\pi) - \frac{1}{2} \log |\Sigma| - \frac{1}{2}(y - X\beta)^\top \Sigma^{-1}(y - X\beta), \tag{3.15}
\end{aligned}
$$

differentiating it with respect to the *variance components* and set to zero. Doing so it yields (see Theorems A.1.3 and A.1.4)

$$
\begin{aligned}
\frac{\partial l(\beta, \gamma)}{\partial \gamma_i} &= -\frac{1}{2} tr\left(\Sigma^{-1} M_i\right) + \frac{1}{2}(y - X\beta)^\top \Sigma^{-1} M_i \Sigma^{-1}(y - X\beta) \\
&= 0, \quad i = 1, \ldots, n+1, \tag{3.16}
\end{aligned}
$$

and thus, equivalently,

$$tr\left(\Sigma^{-1} M_i\right) = (y - X\beta)^\top \Sigma^{-1} M_i \Sigma^{-1}(y - X\beta) \quad i = 1, \ldots, n+1. \tag{3.17}$$

Hence, defining

$$P = \Sigma^{-1} - \Sigma^{-1} X \left(X^\top \Sigma^{-1} X\right)^{-1} X^\top \Sigma^{-1}, \tag{3.18}$$

and noting that $\Sigma^{-1}(y - X\beta^*) = Py$, where β^* is the solution of the general normal equations $X^\top \Sigma^{-1} X\beta = X^\top \Sigma^{-1} y$ in β, we will have

$$tr\left(\Sigma^{-1} M_i\right) = y^\top P M_i P y, \ i = 1, \ldots, n+1. \tag{3.19}$$

$\left(X^\top \Sigma^{-1} X\right)^{-1}$ should be replaced with $\left(X^\top \Sigma^{-1} X\right)^-$ when $\left(X^\top \Sigma^{-1} X\right)$ is a singular matrix.

Noting that

$$
\begin{aligned}
tr\left(\Sigma^{-1} M_i\right) &= tr\left(\Sigma^{-1} M_i \Sigma^{-1} \Sigma\right) \\
&= \sum_{j=1}^{n+1} \gamma_j tr\left(\Sigma^{-1} M_i \Sigma^{-1} M_j\right), \tag{3.20}
\end{aligned}
$$

the system of equations (3.19) becomes (in matrix notation)

$$
\begin{bmatrix}
y^\top P M_1 P y \\
y^\top P M_2 P y \\
\vdots \\
y^\top P M_{n+1} P y
\end{bmatrix}
=
\begin{bmatrix}
tr\left(\Sigma^{-1} M_1 \Sigma^{-1} M_1\right) & \cdots & tr\left(\Sigma^{-1} M_1 \Sigma^{-1} M_{n+1}\right) \\
tr\left(\Sigma^{-1} M_2 \Sigma^{-1} M_1\right) & \cdots & tr\left(\Sigma^{-1} M_2 \Sigma^{-1} M_{n+1}\right) \\
\vdots & \ddots & \vdots \\
tr\left(\Sigma^{-1} M_{n+1} \Sigma^{-1} M_1\right) & \cdots & tr\left(\Sigma^{-1} M_{n+1} \Sigma^{-1} M_{n+1}\right)
\end{bmatrix}
\begin{bmatrix}
\gamma_1 \\
\gamma_2 \\
\vdots \\
\gamma_{n+1}
\end{bmatrix}
\tag{3.21}
$$

which solved in $\gamma^\top = [\gamma_1 \ \gamma_2 \ \ldots \ \gamma_{n+1}]$ gives its desired estimate $\hat{\gamma}^\top = [\hat{\gamma}_1 \ \hat{\gamma}_2 \ \ldots \ \hat{\gamma}_{n+1}]$; $\hat{\gamma}$ is called *ML* estimator for γ.

3.2.2 REML - Based Method

Restricting the distribution of y into the dependence only on the *variance components*, by considering the distribution of $z = Qy$, where Q is a matrix such that $QX = \mathbf{0}_{m,p}$, where m is the number of rows of Q, and Z retain sufficient information needed to estimate the *variance components*, and carrying the ML estimation for the *variance components* in the new model z, we develop the named *restricted maximum likelihood* (REML) estimator for the *variance components*, introduced and explored by Patterson and Thompson [59].

Hence in z the dependence on the fixed effect β is eliminated, z will have less degrees of freedom than y, and consequently the estimators based on it will have less bias. Due to the reduction on the bias the *REML* method is rather preferable than the ML one (see Muller and Pasour [53]).

Recalling the distribution of y in (3.13), and using the Theorem A.1.8, we have that

$$z = Qy \sim \mathcal{N}(\mathbf{0}_m, \Sigma^\circ), \tag{3.22}$$

where $\Sigma^\circ = \sum_{i=1}^{n+1} \gamma_i QM_i Q^\top$. According with Theorem A.1.9, a matrix Q such that $QX = \mathbf{0}_{m,p}$ and with sufficient information needed to estimate the *variance components* must be of full rank with maximal number of rows and an element of $\mathcal{M}^{(n-r) \times n}$, where $r = r(X)$, and of the form $Q = C\left(I - X\left(X^T X\right)^{-1} X^\top\right)$, where C specifies a full rank transformation of the rows of $X(X^T X)^{-1} X^\top$. $\left(X^T X\right)^{-1}$ must be replaced with $\left(X^T X\right)^-$ when $\left(X^T X\right)$ is singular.

The logarithm of the likelihood function of the new model z (model (3.22)) will be

$$l^\circ(\gamma) = -\frac{n-r}{2}log(2\pi) - \frac{1}{2}log|\Sigma^\circ| - \frac{1}{2}z^\top\left(\Sigma^\circ\right)^{-1}z. \tag{3.23}$$

Taking now the partial derivative of l° with respect to the *variance components*, and setting it to zero it holds (See Theorems A.1.3 and A.1.4)

$$
\begin{aligned}
\frac{\partial l^\circ(\gamma)}{\partial \gamma_i} &= -\frac{1}{2}tr\left(\left(\Sigma^\circ\right)^{-1}\frac{\partial \Sigma^\circ}{\partial \gamma_i}\right) + \frac{1}{2}z^\top\left(\left(\Sigma^\circ\right)^{-1}\frac{\partial \Sigma^\circ}{\partial \gamma_i}\left(\Sigma^\circ\right)^{-1}\right)z \\
&= -\frac{1}{2}tr\left(\left(\Sigma^\circ\right)^{-1}\sum_{j=1}^{n+1}\frac{\partial}{\partial \gamma_i}\gamma_j QM_j Q^\top\right) \\
&\quad + \frac{1}{2}z^\top\left(\left(\Sigma^\circ\right)^{-1}\frac{\partial}{\partial \gamma_i}\left(\sum_{j=1}^{n+1}\gamma_j QM_j Q^\top\right)\left(\Sigma^\circ\right)^{-1}\right)z \\
&= -\frac{1}{2}tr\left(\left(\Sigma^\circ\right)^{-1}QM_i Q^\top\right) + \frac{1}{2}z^\top\left(\left(\Sigma^\circ\right)^{-1}QM_i Q^\top\left(\Sigma^\circ\right)^{-1}\right)z \\
&= 0, \quad i = 1, \dots, n+1,
\end{aligned}
$$

so that

$$tr\left(\left(\Sigma^\circ\right)^{-1}QM_i Q^\top\right) = z^\top\left(\left(\Sigma^\circ\right)^{-1}QM_i Q^\top\left(\Sigma^\circ\right)^{-1}\right)z \tag{3.24}$$

Nothing that (see Proposition 2.1.1) the left-hand side of the equation (3.24) is equivalent to

$$tr\left((\Sigma^\circ)^{-1}QM_iQ^\top(\Sigma^\circ)^{-1}\Sigma^\circ\right) \;=\; tr\left((\Sigma^\circ)^{-1}QM_iQ^\top(\Sigma^\circ)^{-1}\sum_{j=1}^{n+1}\gamma_jQM_jQ^\top\right)$$

$$=\; \sum_{j=1}^{n+1}\gamma_j tr\left((\Sigma^\circ)^{-1}QM_iQ^\top(\Sigma^\circ)^{-1}QM_jQ^\top\right)$$

$$=\; \sum_{j=1}^{n+1}\gamma_j tr\left(Q^\top(\Sigma^\circ)^{-1}QM_iQ^\top(\Sigma^\circ)^{-1}QM_j\right). \quad (3.25)$$

Thus, equation (3.24) becomes

$$\chi = M\gamma \quad (3.26)$$

where $\chi = \begin{bmatrix} z^\top\left((\Sigma^\circ)^{-1}QM_1Q^\top(\Sigma^\circ)^{-1}\right)z \\ z^\top\left((\Sigma^\circ)^{-1}QM_2Q^\top(\Sigma^\circ)^{-1}\right)z \\ \vdots \\ z^\top\left((\Sigma^\circ)^{-1}QM_{n+1}Q^\top(\Sigma^\circ)^{-1}\right)z \end{bmatrix}$, $\gamma = \begin{bmatrix} \gamma_1 \\ \gamma_2 \\ \vdots \\ \gamma_{n+1} \end{bmatrix}$, and

$$M = \begin{bmatrix} tr\left(Q^\top\Sigma^{-1}QM_1Q^\top\Sigma^{-1}QM_1\right) & \dots & tr\left(Q^\top\Sigma^{-1}QM_1Q^\top\Sigma^{-1}QM_{n+1}\right) \\ tr\left(Q^\top\Sigma^{-1}QM_2Q^\top\Sigma^{-1}QM_1\right) & \dots & tr\left(Q^\top\Sigma^{-1}QM_2Q^\top\Sigma^{-1}QM_{n+1}\right) \\ \vdots & \ddots & \vdots \\ tr\left(Q^\top\Sigma^{-1}QM_{n+1}Q^\top\Sigma^{-1}QM_1\right) & \dots & tr\left(Q^\top\Sigma^{-1}QM_{n+1}Q^\top\Sigma^{-1}QM_{n+1}\right) \end{bmatrix}.$$

Under certain regularity conditions the likelihood - based estimators have many desirable properties such as consistence, normal asymptotic, and efficiency (see Harville [26], Magnus [42], and Miller([46], [47]), among other references). See Harville [26] for a comprehensive review on estimation procedure along with computational techniques, and Lair and Ware [37] for a comprehensive discussion on REML estimation relationship to *variance components*. For some details on applications of such methods we recommend, for example, Anderson [4] and Hartley and Rao [25].

Meanwhile, as we may see, both system of equations (3.21) and (3.26) cannot be directly managed (in order to produce solutions), since the matrices in the right-hand side ass well as the vectors in the left-hand side are them self dependent on the *variance components* (see the example below), so that, typically, the usual approaches require iterative methods (see McCulloch and Searle [44]).

Example 3.2.1. Lets consider the unbalanced "one-way model" from the Example 3.1.1. Recall that $P = \Sigma^{-1} - \Sigma^{-1}X\left(X^\top\Sigma^{-1}X\right)^{-1}X^\top\Sigma^{-1}$ (see (3.18)), a quantity which depends on γ_1 and γ_2 (through Σ^{-1}).

The desired ML-based estimator $\hat{\gamma} = \begin{bmatrix} \hat{\gamma}_1 \\ \hat{\gamma}_1 \end{bmatrix}$ for γ is achieved solving on γ the system of equation:

$$\begin{bmatrix} y^\top PM_1Py \\ y^\top P^2y \end{bmatrix} = \begin{bmatrix} tr\left(\Sigma^{-1}M_1\Sigma^{-1}M_1\right) & tr\left(\Sigma^{-1}M_1\Sigma^{-1}\right) \\ tr\left(\Sigma^{-1}\Sigma^{-1}M_1\right) & tr\left(\Sigma^{-1}\Sigma^{-1}\right) \end{bmatrix}\begin{bmatrix} \gamma_1 \\ \gamma_2 \end{bmatrix}. \quad (3.27)$$

Clearly the left-hand side and the right-hand side of (3.27) deponds on γ_1 and γ_2.

The most common methods implement a Newton-Rapson type algorithm (Fisher scoring algorithm and average information algorithm, for example). See Gilmour et al. [21]. For the case when some estimate for some of the *variance components* produced thought the iterative method is negative the log likelihood might be reexamined to find values of the *variance components* within the non negative ones that maximize that function.

The practical alternatives to Newton-Rapson type algorithm are the EM algorithm (see Lindstrom et al. [40], for example) and the Parameter expanded (PX) EM algorithm (see Liu et al. [41] or Lewandowski et al. [39], for example), since they both have desirable properties of monotonic convergence and the component updates remains in their parameter space. In addition, the PX algorithm has a rate of convergence that is no slower than the EM algorithm (see Liu et al. [41]), which made it preferred in a practical implementation point of view. However, they both still have gap to be filled once the computational implementation is expensive at each iterate with, as made clear before, relatively slow convergence.

Diffey et al. [17] presented an improved algorithm of PX REML algorithm and EX REML algorithm for the variance components estimation in MLM. In their approach the authors proposed alternative algorithms by consider a new incomplete data specification. Both PX and EM algorithm require specification of the complete data, comprising the incomplete and missing data.

3.3 Variance Components Estimation - Models With OBS

Mixed linear models with orthogonal block structure (*OBS*), introduced and investigated by Nelder ([57], [58]), has playing important role in design experiments (see Houtmam and Speed [31], Mejza [45], for instance) and in nowadays, after more detailed definition and the introduction of orthogonal designs by Houtman and Speed [31], is playing important role in the theory of randomized block designs (see Calinski and Kageyama ([12], [13])).

Definition 3.3.1. The model (3.9) is said to have *OBS* if its variance-covariance matrix, Σ, can be expressed as

$$\Sigma = \sum_{i=1}^{s} \zeta_i Q_i, \tag{3.28}$$

where each Q_i, $i = 1, \ldots, s$, is a projection matrix, and $Q_1, \ldots, Q_s$ are pairwise orthogonal matrix, such that $\sum_{i=1}^{s} Q_i = I$, and ζ_i, $i = 1, \ldots, s$, are non-negative parameters.

Example 3.3.1. As an example, let

$$y_{ijkl} = \mu + \alpha_i + \beta_{ij} + \delta_{ijk} + e_{ijkl},$$

with $i = 1, \ldots, n$, $j = 1, \ldots, n_i$, $k = 1, \ldots, n_{ij}$, and $l = 1, \ldots, n_{ijk}$, be a nested model (see (3.4) for further explanation) where u and the vector $\alpha^\top = [\alpha_1 \ldots \alpha_n]$ are the fixed effect, and $\beta^\top = [\beta_{11} \ldots \beta_{nn_n}]$ and $\delta^\top = [\delta_{111} \ldots \delta_{nn_n n_{n_n}}]$ the random effect vectors. Under the usual assumption for

a mixed linear model, with $X = [\mathbf{1}_{n_{ijk}}, A]$ standing for the fixed effect design matrix and X_1 and X_2 for the random effect design matrices, the model can be represented in matrix notation as

$$y = X\mu^* + X_1\beta + X_2\delta + e, \tag{3.29}$$

with $\mu^* = \begin{bmatrix} \mu \\ \delta \end{bmatrix}$.

Since $R\left(\mathbf{1}_{n_{ijk}}\right) \subset R(A) \subset R(X_1) \subset R(X_2)$, according with VanLeeuwen et al. [72] the set of matrices $\{\mathbf{1}_{n_{ijk}}, A, X_1, X_2\}$ is said to be nested. Supposing the model is completely balanced, that is $n_{ijk} = r$ and $n_{ij} = t$, we will have that $X_1^\top X_1 = D(n_{ij}n_{ijk}) = D(rt, \dots, rt) = rtI$ and $X_2^\top X_2 = D(n_{ijk}) = D(r, \dots, r) = rI$ so that $X_1 X_1^\top = rtP_{X_1}$ and $X_2 X_2^\top = rtP_{X_1}$. Then, according with Theorem A.1.10, y has *OBS*.

With $T = P_{R(X)} = XX^+$, the projection matrix onto the subspace spanned by the columns of the design matrix for the fixed effect X, the model is said to be *COBS* if T commutes with $Q \in \{Q_1, \dots, Q_s\}$ (see Fonseca et al. [19]). In this Section we aim to introduce the estimation of *variance components* in mixed linear models with *OBS*. We introduce the estimation procedure based in likelihood.

Theorem 3.3.1. *Let y have OBS, with variance - covariance matrix $\Sigma = \sum_{i=1}^s \zeta_i Q_i$, and put $r(Q_i) = r_i$. Then*

(a) $|\Sigma| = \prod_{i=1}^s \zeta_i^{r_i}$;

(b) $\Sigma^{-1} = \sum_{i=1}^s \frac{1}{\zeta_i} Q_i$.

Proof. To prove (a) we may note that ζ_i will be the eigenvalue of Σ and r_i its correspondent root. For (b), it follows that

$$\begin{aligned}
\Sigma^{-1}\Sigma &= \sum_{i=1}^s \sum_{j=1}^s \frac{\zeta_j}{\zeta_i} Q_i Q_j = \sum_{i=1}^s Q_i Q_i + \sum_{i \neq j} \frac{\zeta_j}{\zeta_i} Q_j Q_i \\
&= \sum_{i=1}^s Q_i = I. \tag{3.30}
\end{aligned}$$

$\square$

Now, adding the Gaussian assumption to the *OBS* model (3.3.1), with $\zeta^\top = [\zeta_1 \dots \zeta_s]$, the logarithm of the likelihood function will be given by

$$\begin{aligned}
l(\beta, \zeta) &= -\frac{n}{2}log(2\pi) - \frac{1}{2}log|\Sigma| - \frac{1}{2}(y - X\beta)^\top \Sigma^{-1}(y - X\beta) \\
&= -\frac{n}{2}log(2\pi) - \frac{1}{2}\sum_{i=1}^s r_i log(\zeta_i) - \frac{1}{2}\sum_{i=1}^s (y - X\beta)^\top Q_i(y - X\beta). \tag{3.31}
\end{aligned}$$

Noting that $Q_i = \Sigma^{-1}Q_i\Sigma_i^{-1}$, and taking the partial derivative in order to ζ and equating it to zero we will have that

$$\frac{r_i}{\zeta_i} = \frac{1}{\zeta_i^2}(y - X\beta)^\top \Sigma^{-1} Q_i \Sigma^{-1}(y - X\beta), \; i = 1, \dots, s.$$

Therefore

$$\hat{\zeta}_i = \frac{1}{r_i} y^\top P Q_i P y, \tag{3.32}$$

with $\Sigma^{-1}(y - X\beta^*) = Py$ and $P = \Sigma^{-1} - \Sigma^{-1}X\left(X^\top\Sigma^{-1}X\right)^{-}X^\top\Sigma^{-1}$, where β^* is the solution of the general least square equation $\sum_{i=1}^{s} \frac{1}{\zeta_i} X^\top Q_i(y - X\beta) = 0$.

VARIANCE COMPONENTS ESTIMATION - THE SUB-D METHOD

4.1 Introduction And Literature Review

It is not fair discussing *variance components* estimation without discussing it's relation to the underlying design, since the quality of the estimation depends to a large extent on the design used to generate the data (see Khuri [34]). On the matter of search for optimal designs for *variance components* estimation, Anderson ([4], [3]) and many of his co-authors (see Crump [15], Anderson and Crump [6], Bush and Anderson [11], Muse [55], Muse and Anderson [56], etc) are the main contributors. Most of the works on that matter focus on "one-way designs" (see model (3.3)), nested designs (see the "two-way nested design" (3.4) or (3.5), for example), and the crossed designs (see the "two - way crossed designs" (3.6) or (3.7), for example). This designs received much attention due to its application on genetics, animal breeding, process control, and quality control and improvement. We start our review by introducing an overview on (optimal) designs for *variance components* estimation (see Khuri [34]), and secondly the main contributes on estimation procedures for variance components. The main contributers for this last topic are Anderson [2], Anderson and Bancroft [5], Yates [73], Nelder ([57], [58]), among others.

4.1.1 Designs For Variance Components Estimation

According with Khuri [34] works on design aspects of the *variance components* estimation is somewhat limited. Hammesley [24], Crump [15] and Anderson and Crump [6] provided the first works on optimal design for *variance components*.

Recall the "one-way model" (3.3). Considering, in this model, the ANOVA estimators $\hat{\sigma}_\alpha^2$ and $\hat{\sigma}_\varepsilon^2$ for σ_α^2 and σ_ε^2, respectively, and a fixed k, Crump [15] proposed a criterion for the choice of optimal design for the model (3.3), which goes through find the minimum of $\Sigma(\hat{\sigma}_\varepsilon^2)$ or $\Sigma(\frac{\hat{\sigma}_\alpha^2}{\hat{\sigma}_\varepsilon^2})$. For a fixed value of $N = \sum_{i=1}^k n_i$, Hammesley(1949) observed that the minimum of $\Sigma(\hat{\sigma}_\varepsilon^2)$ is achieved

when the model is taken to be balanced and $n_i = n$ is taken to be the closest integer of $\frac{N(\rho++1)+1}{N\rho+1}$, where $\rho = \frac{\sigma_\alpha^2}{\sigma_\varepsilon^2}$. In addition, for fixed k and N, Anderson and Crump [6] showed that the minimum is achieved when $n_i = n = \frac{N}{k}$, and in this case the optimal design is achieved by taking k to be the closest integer of $k^* = \frac{N(N\rho+2)}{N(\rho+1)+1} = \frac{N}{n}$. Consequently, n has to be the closest integer of $\frac{N(\rho+1)+1}{N\rho+2}$.

Anderson and Crump [6] also showed that the optimal design for estimating σ_α^2 is to allocate $p+1$ observations to each of the r groups and p observations to each of the remaining $k-r$ groups, where $N = pk + r$, and $0 \le r < k$. This means that the design is as closest as possible of a balanced model. Further more, Anderson and Crump [6] suggested that the value of k which minimizes $\Sigma(\frac{\hat\sigma_\alpha^2}{\hat\sigma_\varepsilon^2})$ is the closest integer of $k^{**} = \frac{(N-5)(N\rho+1)}{2N\rho+N-3}$. Since, asymptotically, $\frac{k^*}{k^{**}} = \frac{1+2\rho}{1+\rho}$, when N is relatively large it seems that there are needed more groups to estimate σ_α^2 then to estimate ρ (see Khuri [34]).

Reexamining the optimal designs suggested by Anderson and Crump [6], using the restricted ANOVA estimator $\hat\sigma_\alpha^2$, ML and modified ML (MML) estimators (see Klotz et al [36] for the last one),Thonson and Anderson [71] showed that, for a small value of N and $\rho < 1$, a MML estimator of σ_α^2 is superior for certain unbalanced designs.

For fixed k and N, Mukerjee [50] showed that the optimal design is achieved by minimizing uniformly $\Sigma(\tilde\sigma^2)$, where $\tilde\sigma^2 = (\tilde\sigma_\alpha^2, \tilde\sigma_\varepsilon^2)$ is the minimum quadratic unbiased estimator of $\sigma^2 = (\sigma_\alpha^2, \sigma_\varepsilon^2)$. Mukerjee and Huda [51] reached a similar conclusion (see Khuri [34]).

The problem in the search for optimal designs in context of crossed models, was approached for several authors. Gaylor [20], Bush and Anderson [11] and Mostafa [49] are some of the works with a great impact in this matter.

Mostafa [49] proposed two designs, D_1 and D_2, for the crossed model provided the total number of observations $N = \sum_{i=1}^k \sum_{j=1}^b n_{ij}$ is expressed as either $N = r_1(r_1 + 1)$ (for D_1) or $N = r_2(r_2 + 2)$ (for D_2), where r_i denotes the number of observations in row and column of the design D_i. Using Yates [73] methods to obtain unbiased estimators for σ_τ^2, σ_β^2, σ_γ^2 and σ_ε, he showed that designs D_1 and D_2 are more efficient for estimating σ_τ^2, σ_β^2, σ_γ^2 than a balanced model with the same number of observations, particularly in situation that $\frac{\sigma_\tau^2}{\sigma_\varepsilon} > 1$, $\frac{\sigma_\beta^2}{\sigma_\varepsilon} > 1$ and $\frac{\sigma_\gamma^2}{\sigma_\varepsilon} > 1$.

Muse [55] and Muse and Anderson [56] did a notable work comparing several designs for the "two-way crossed model" with no interaction. The authors used the asymptotic variances of the ML estimators and the trace of asymptotic variance-covariance matrix of the vector of the ML estimators of $\sigma = (\sigma_\tau, \sigma_\beta, \sigma_\gamma, \sigma_\varepsilon)$. Further more, the authors provided a report with constructive recommendation (see Khuri [34]) in order to choose the more adequate design. They suggested that prior information is necessary for the selection of a reasonable design; under certain condition, considering the trace criterion, the balanced designs seems to be less efficient than some of the other designs considered in the comparison process. Another notable work is due to Muse et al [54] who extended the comparison to "two-way crossed model" with no interaction, based on asymptotic ML procedures. Haile and webster [23] provided comparison including balanced incomplete block designs, for models without interaction.

Nested (hierarchical) models - useful tool for experiments where the treatment are separated into several groups - was also widely investigated (see Anderson and Bancroft [5], Anderson [2],

Bainbridge [8], for instance). It is known that these designs allocate the degrees of freedom mostly in the last group of treatments, which clearly causes certain unbalanceness in the model. To avoid this problem it requires to increase the size of experiments, which, sometimes, may not be so practical. Bainbridge [8] showed that this problem would be overcome using particular types of unbalanced nested designs.

Goldsmith and Gaylor [22] compared 61 different designs for the random "two-way nested model" (3.4) using the ANOVA estimation procedures for *variance components*. With $\hat{\sigma}_\varepsilon^2$, $\hat{\sigma}_\beta^2$, and $\hat{\sigma}_\alpha^2$ denoting the ANOVA estimators for the variance components σ_ε^2, σ_β^2, and σ_α^2, respectively, in their approach, the authors considered three different criterion based on functions of the variance-covariance matrix of the vector $f = [\hat{\sigma}_\varepsilon^2 \ \hat{\sigma}_\beta^2 \ \hat{\sigma}_\alpha^2]^\top$: the trace criterion - $tr(\Sigma(f)) = \Sigma(\hat{\sigma}_\varepsilon^2) + \Sigma(\hat{\sigma}_\beta^2) + \Sigma(\hat{\sigma}_\alpha^2)$; the determinant criterion - $|f|$; and the adjusted trace criterion - $tr(\Sigma^*(f)) = \Sigma(\hat{\sigma}_\varepsilon^2) + \frac{\Sigma(\hat{\sigma}_\beta^2)}{\rho_1^2} + \frac{\Sigma(\hat{\sigma}_\alpha^2)}{\rho_2^2}$, where $\rho_1 = \frac{\sigma_\beta^2}{\sigma_\varepsilon^2}$, $\rho_2 = \frac{\sigma_\alpha^2}{\sigma_\varepsilon^2}$, and $\Sigma^*(f)$ is simply $\Sigma(f)$ with each of its elements scaled by the size of the variance components involved in the computation of the element. Evidently, a particular design will be optimum in a given class of designs if it has the smallest value of a particular criterion for a given sample size and variance component configuration. The authors reported the following:

(i) The trace criterion (widely used) revealed to be the best one since it tended to concentrate the sampling at the group for which the *variance component* is large relative to the others;

(ii) When the *variance components* for the first and the second group were small, compared to the error variance, the balanced design is found to be optimum since it concentrates the sampling in the third stage. More over, if any stage has large *variance components*, then the highest degrees of freedom for that stage are selected.

4.1.2 Procedures For Variance Components Estimation

A part of the problem of the search for optimal designs, *variance components* have been widely investigated and several methods for its estimation has been suggested. We highlight the ANOVA based methods, the ML based methods, and OBS models. Thanks to its simplicity regarding the implementation, since its underlying idea is to equate the quadratic error for the different sources of variations to their respective expected values and solve for the *variance components*, the ANOVA based methods are common procedures for the *variance components* estimation (see Section 3.1).

The underling idea of the ML based methods goes through assuming the Gauss distribution for the random effects and carrying the maximum likelihood estimation from the data (see Section 3.2). In context of mixed linear models, we highlight the ML estimator and REML estimator (see Harville [26], Fairclough and Helms [18], Andrade and Helms [7] and Lair and Ware [37]).

Finally, the OBS models (see Nelder ([57], [58])) plays important role in design experiments (see Houtmam and Speed [31] and Mejza [45], for instance) and, nowadays, after more detailed definition the introduction of orthogonal designs (see Houtmam and Speed [31]) plays as well an important role in theory of randomized block desings (see Calinski and Kageyama ([12], [13])) so that it constitutes an optimal tool for estimating *variance components* in mixed linear models.

Some of the methods referred here are summarized in Searle et al. [67].

4.2 Sub-Diagonalizing The Variance-Covariance Matrix

Variance components estimation in linear models (with mixed and/or fixed effect) have been widely investigated and consequently several methods for estimation with relevant properties have been derived.

The aim of this work is to provide a new method for estimating the *variance components* in the MLM with properties that may bring some gain relatively to the previous ones. We start our approach introducing a method to "diagonalize" the variance-covariance matrix

$$V = \sum_{d=1}^{r+1} \gamma_d N_d$$

on the mixed linear model

$$z \sim (X\beta, \, V), \tag{4.1}$$

where $\gamma_d \geq 0$, $d = 1, \ldots, r$, $\gamma_{r+1} > 0$, are unknown parameters called *variance components*, $N_d = X_d X_d^\top \in \mathscr{S}^m$, with $X_d \in \mathscr{M}^{m_d \times s_d}$ the known design matrices for the random effects, and $N_{r+1} = I_m$, and develop optimal estimators for the *variance components* $\gamma_1, \ldots, \gamma_{r+1}$. See Silva et al. [70].

Our approach will not assume no underling distribution for the model as do the likelihood based method, we will only require it to have second moment as do the ANOVA - based methods. We will introduce our method firstly for the model with 3 *variance components* and secondly for the model with an arbitrary number of *variance components*.

Since the parameters we want to estimate do not depend on the fixed effect part, it is convenient to us to remove the dependence of the model on the fixed effect part, remarking that this action will cause no loss of information needed to estimate these parameters and will reduce the complexity of the model for the algebraic manipulation, as well as the bias in estimation process. The strategy that we will follow is in all similar to the first phase of REML: we will project the observations vector on the orthogonal complement of X, the subspace spanned by the men vector.

Let $P_o = P_{R(X)}$ denotes the projection matrix onto the subspace spanned by the columns of the matrix X, and $P^* = P_{R(X)^\perp} = I_m - P_o$ the projection matrix onto the orthogonal complement of the columns space of X. There exists a matrix B_o whose columns are the eigenvectors associated to the null eigenvalues of P_o such that

$$B_o^\top B_o = I_{m-r(P_o)} \quad \text{and} \quad B_o B_o^\top = P^*.$$

Thus, instead of the model (4.1) we will consider the restricted model:

$$y = B_o^\top z \sim \left(\mathbf{0}_n, \sum_{d=1}^{r+1} \gamma_d M_d \right), \tag{4.2}$$

where $M_d = B_o^\top N_d B_o$, $n = m - r(P_o)$.

Before proceeding with the method deduction process, we set a needed notion for such a process.

Definition 4.2.1. Let

$$A = \begin{bmatrix} A_{11} & \dots & A_{1n} \\ \vdots & \ddots & \vdots \\ A_{n1} & \dots & A_{nn} \end{bmatrix}$$

be a diagonal blockwise matrix. We say that a matrix T *sub-diagonalizes* A if TA produces a blockwise matrix whose matrices in the diagonal are all diagonal matrices, that is T diagonalizes the matrices $A_{11}, \dots, A_{nn}$ in the diagonal of A.

The two next subsections, 4.2.1 and 4.2.2, approach the diagonalization of the variance - covariance matrix in the mixed linear model (4.2) for the case of models with 3 *variance components*, that is $r = 2$. The third one is devoted to the general case, that is, the diagonalization of the variance - covariance matrix in models with any arbitrary $r \geq 1$ *variance components*.

4.2.1 The Case $r = 2$

In this section we sub-diagonalize the variance - covariance matrix in the mixed linear model (4.2) for $r = 2$, that is

$$y \sim \left(\mathbf{0}_n,\ \gamma_1 M_1 + \gamma_2 M_2 + \gamma_3 I_n\right). \tag{4.3}$$

Since M_1 is a symmetric matrix there exists any orthogonal matrix (see Schott [64])

$$P_1 = \begin{bmatrix} A_{11} \\ \vdots \\ A_{1h_1} \end{bmatrix} \in \mathscr{M}^{\left(\sum_{i=1}^{h_1} g_i\right) \times n}, \tag{4.4}$$

with $A_{1i} \in \mathscr{M}^{g_i \times n}$ ($\sum_{i=1}^{h_1} g_i = n$), such that $M_1 = P_1^T D_1 P_1$, or equivalently $P_1 M_1 P_1^T = D_1$, where

$$D_1 = \begin{bmatrix} \theta_{11} I_{g_1} & 0 & \dots & 0 \\ 0 & \theta_{12} I_{g_2} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \theta_{1h_1} I_{g_{h_1}} \end{bmatrix} \tag{4.5}$$

is a diagonal matrix whose diagonal entries θ_{1i}, $i = 1, \dots, h_1$, are the h_1 different eigenvalues of the matrix M_1 with corresponding roots (multiplicities) $g_i = rank(A_{1i}^\top)$, $i = 1, \dots, h_1$. It must be noted that the set of columns of each matrix $A_{1i}^\top$ forms a set of g_i orthonormal vectors associated to the eigenvalue θ_{1i} of the matrix M_1 (Theorem 2.1.8 guarantees the existence of such matrix $A_{1i}^\top$), so that $A_{1i} A_{1i}^\top = I_{g_i}$ and $A_{1i}^\top A_{1i} = P_{R(A_{1i}^\top)}$. Clearly, we have that $P_1 P_1^\top = I_n$, and (see Theorem 2.2.8)

$$\begin{aligned} P_1^\top P_1 &= A_{11}^\top A_{11} + \dots + A_{1h_1}^\top A_{1h_1} \\ &= P_{R(A_{11}^\top)} + \dots + P_{R(A_{1h_1}^\top)} \\ &= I_n. \end{aligned} \tag{4.6}$$

Putting

$$A_{1i} M_2 A_{1s}^\top = \begin{cases} M_{ii}^2 & i = s \\ W_{is}^2 & i \neq s \end{cases} \tag{4.7}$$

we will have that

$$
\begin{aligned}
\Sigma(P_1 y) &= \gamma_1 P_1 M_1 P_1^\top + \gamma_2 P_1 M_2 P_1^\top + \gamma_3 P_1 P_1^\top \\
&= \gamma_1 \begin{bmatrix} \theta_{11} I_{g_1} & 0 & \dots & 0 \\ 0 & \theta_{12} I_{g_2} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \theta_{1h_1} I_{g_{h_1}} \end{bmatrix} + \gamma_2 \begin{bmatrix} M_{11}^2 & W_{12}^2 & \dots & W_{1h_1}^2 \\ W_{21}^2 & M_{22}^2 & \dots & W_{2h_1}^2 \\ \vdots & \vdots & \ddots & \vdots \\ W_{h_11}^2 & W_{h_12}^2 & \dots & M_{h_1 h_1}^2 \end{bmatrix} \\
&\quad + \gamma_3 \begin{bmatrix} I_{g_1} & 0 & \dots & 0 \\ 0 & I_{g_2} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & I_{g_{h_1}} \end{bmatrix} \\
&= \gamma_1 D(\theta_1 I_{g_1} \dots \theta_{h1} I_{g_{h_1}}) + \gamma_2 \Gamma + \gamma_3 D(I_{g_1} \dots I_{g_{h_1}}),
\end{aligned}
\tag{4.8}
$$

where

$$
\Gamma = \begin{bmatrix} M_{11}^2 & W_{12}^2 & \dots & W_{1h_1}^2 \\ W_{21}^2 & M_{22}^2 & \dots & W_{2h_1}^2 \\ \vdots & \vdots & \ddots & \vdots \\ W_{h_11}^2 & W_{h_12}^2 & \dots & M_{h_1 h_1}^2 \end{bmatrix}.
$$

It is clear that for the three matrix $D(\theta_1 I_{g_1} \dots \theta_{h1} I_{g_{h_1}})$, $D(I_{g_1} \dots I_{g_{h_1}})$ and Γ appearing in (4.8), the blockwise matrix Γ is the only one which is not diagonal.

We diagonalize the symmetric matrices M_{ii}^2, $i = 1, \dots, h_1$, that appear in the diagonal of the matrix Γ; that is, we sub-diagonalize the matrix Γ.

Since M_{ii}^2 is symmetric there exists (see Schott [64]) an orthogonal matrix

$$
P_{2i} = \begin{bmatrix} A_{2i1} \\ \vdots \\ A_{2ih_{2i}} \end{bmatrix} \in \mathscr{M}^{\left(\Sigma_{j=1}^{h_{2i}} g_{ij} \right) \times g_i},
$$

where $A_{2ij} \in \mathscr{M}^{g_{ij} \times g_i}$ ($\Sigma_{j=1}^{h_{2i}} g_{ij} = g_i$), such that

$$
D_{ii}^2 = P_{2i} M_{ii}^2 P_{2i}^\top = \begin{bmatrix} \theta_{2i1} I_{g_{i1}} & 0 & \dots & 0 \\ 0 & \theta_{2i2} I_{g_{i2}} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \theta_{2ih_{2i}} I_{g_{ih_{2i}}} \end{bmatrix}, \quad i = 1, \dots, h_1.
\tag{4.9}
$$

It must be noted that the matrix $A_{2ij}^\top$, $i = 1, \dots, h_1$, $j = 1, \dots, h_{2i}$, is an orthogonal matrix whose columns form a set of $g_{ij} = rank(A_{2ij}^\top)$ orthonormal eigenvectors associated to the different eigenvalues θ_{2ij} of the matrix M_{ii}^2; that is, g_{ij} is the multiplicity of the eigenvalues θ_{2ij}, and $A_{2ij}^\top A_{2ij} = P_{R(A_{2ij}^\top)}$ and $A_{2ij} A_{2ij}^\top = I_{g_{ij}}$.

Let

$$
P_2 = \begin{bmatrix} P_{21} & 0 & \dots & 0 \\ 0 & P_{22} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & P_{2h_1} \end{bmatrix} \in \mathscr{M}^{\left(\Sigma_{i=1}^{h_1} \Sigma_{j=1}^{h_{2i}} g_{ij} \right) \times \left(\Sigma_{i=1}^{h_1} g_i \right)}.
\tag{4.10}
$$

Proposition 4.2.1. *P_2 is an orthogonal matrix.*

Proof.

$$P_2 P_2^\top = \begin{bmatrix} P_{21}P_{21}^\top & 0 & \ldots & 0 \\ 0 & P_{22}P_{22}^\top & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & P_{2h_1}P_{2h_1}^\top \end{bmatrix},$$

where

$$P_{2i}P_{2i}^\top = \begin{bmatrix} A_{2i_1}A_{2i_1}^\top & 0 & \ldots & 0 \\ 0 & A_{2i_2}A_{2i_2}^\top & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & A_{2i_{h_{2i}}}A_{2i_{h_{2i}}}^\top \end{bmatrix}.$$

Noting that $A_{2i_1}A_{2i_1}^\top = I_{g_{ij}}$, we see that $P_2 P_2^\top = I_{\sum_{i=1}^{h_1} g_i}$.

Now,

$$P_2^\top P_2 = \begin{bmatrix} P_{21}^\top P_{21} & 0 & \ldots & 0 \\ 0 & P_{22}^\top P_{22} & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & P_{2h_1}^\top P_{2h_1} \end{bmatrix},$$

with (see Theorem 2.2.8)

$$\begin{aligned} P_{2i}^\top P_{2i} &= A_{2i_1}^\top A_{2i_1} + A_{2i_2}^\top A_{2i_2} + \ldots + A_{2i_{h_{2i}}}^\top A_{2i_{h_{2i}}} \\ &= P_{R(A_{2i_1}^\top)} + P_{R(A_{2i_2}^\top)} + \ldots + P_{R(A_{2i_{h_{2i}}}^\top)} \\ &= I_{g_i}, \end{aligned} \tag{4.11}$$

which completes the proof. $\qquad\qquad\square$

Thus, the new model $P_2 P_1 Y$ will have variance-covariance matrix given by

$$\begin{aligned} \Sigma(P_2 P_1 y) &= \gamma_1 P_2 D(\theta_1 I_{g_1} \ldots \theta_{h1} I_{g_{h_1}}) P_2^T + \gamma_2 P_2 \Gamma P_2^T + \gamma_3 P_2 D(I_{g_1} \ldots I_{g_{h_1}}) P_2^T \\ &= \gamma_1 \begin{bmatrix} \theta_{11}P_{21}P_{21}^T & 0 & \ldots & 0 \\ 0 & \theta_{12}P_{22}P_{22}^T & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & \theta_{1h_1}P_{2h_1}P_{2h_1}^T \end{bmatrix} \\ &\quad + \gamma_2 \begin{bmatrix} D_{11}^2 & P_{21}W_{12}^2 P_{22}^T & \ldots & P_{21}W_{1h_1}^2 P_{2h_1}^T \\ P_{22}W_{21}^2 P_{21}^T & D_{22}^2 & \ldots & P_{22}W_{2h_1}^2 P_{2h_1}^T \\ \vdots & \vdots & \ddots & \vdots \\ P_{2h_1}W_{h_11}^2 P_{21}^T & P_{2h_1}W_{h_12}^2 P_{22}^T & \ldots & D_{h_1h_1}^2 \end{bmatrix} \\ &\quad + \gamma_3 \begin{bmatrix} P_{21}P_{21}^T & 0 & \ldots & 0 \\ 0 & P_{22}P_{22}^T & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & P_{2h_1}P_{2h_1}^T \end{bmatrix}, \end{aligned} \tag{4.12}$$

where

$$P_{2i}P_{2i}^{\top} = \begin{bmatrix} A_{2i1}A_{2i1}^{\top} & 0 & \dots & 0 \\ 0 & A_{2i2}A_{2i2}^{\top} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & A_{2ih_{2i}}A_{2ih_{2i}}^{\top} \end{bmatrix} = \begin{bmatrix} I_{g_{i1}} & 0 & \dots & 0 \\ 0 & I_{g_{i2}} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & I_{g_{ih_{2i}}} \end{bmatrix},$$

and, with $i \neq s$,

$$P_{2i}W_{is}^{2}P_{2s}^{\top} = \begin{bmatrix} A_{2i1}W_{is}^{2}A_{2s1}^{\top} & A_{2i1}W_{is}^{2}A_{2s2}^{\top} & \dots & A_{2i1}W_{is}^{2}A_{2sh_{2s}}^{\top} \\ A_{2i2}W_{is}^{2}A_{2s1}^{\top} & A_{2i2}W_{is}^{2}A_{2s2}^{\top} & \dots & A_{2i2}W_{is}^{2}A_{2sh_{2s}}^{\top} \\ \vdots & \vdots & \ddots & \vdots \\ A_{2ih_{2i}}W_{is}^{2}A_{2s1}^{\top} & A_{2ih_{2i}}W_{is}^{2}A_{2s2}^{\top} & \dots & A_{2ih_{2i}}W_{is}^{2}A_{2sh_{2s}}^{\top} \end{bmatrix}.$$

The Matrix $D_{ii}^{2} = P_{2i}M_{ii}^{2}P_{2i}^{T}$, $i = 1,\dots,h_{1}$, appearing in the diagonal in the right side of (4.12) is defined in (4.9).

Definition 4.2.2. The orthogonal matrices P_{1} and P_{2} respectively defined in (4.4) and (4.10) will be called *sub-diagonalization* matrices; the matrix $P_{2}P_{1}$ *sub-diagonalizes* the variance-covariance matrix, $\sum_{d=1}^{3}\gamma_{d}M_{d}$, where $M_{3} = I_{n}$.

Note that

$$P_{2}P_{1}y = \begin{bmatrix} A_{211}A_{11}y \\ \vdots \\ A_{21h_{21}}A_{11}y \\ A_{221}A_{12}y \\ \vdots \\ A_{22h_{22}}A_{12}y \\ \vdots \\ A_{2h_{1}1}A_{1h_{1}}y \\ \vdots \\ A_{2h_{1}h_{2h_{1}}}A_{1h_{1}}y \end{bmatrix}.$$

The distribution of the sub-models

$$y_{ij} = A_{2ij}A_{1i}y, \ i = 1,\dots,h_{1}, \ j = 1,\dots,h_{2i},$$

as well as the cross-covariance between the sub-models, y_{ij} and y_{sk} say, are summarized in the following results.

Proposition 4.2.2.

$$y_{ij} \sim \left(\mathbf{0}_{g_{ij}}, \lambda_{ij}I_{g_{ij}}\right), \ i = 1,\dots,h_{1}; \ j = 1,\dots,h_{2i},$$

where $\lambda_{ij} = \gamma_{1}\theta_{1i} + \gamma_{2}\theta_{2ij} + \gamma_{3}$.

Proof.

Recalling that $A_{2ij}A_{1i} \in \mathcal{M}^{g_{ij} \times n}$ and $g_{ij} \leq n$, according with Theorem A.1.8(c) we will have that

$$y_{ij} \sim \left(\mathbf{0}_{g_{ij}}, \sum_{d=1}^{2} \gamma_d A_{2ij}A_{1i}M_d A_{1i}^\top A_{2ij}^\top + \gamma_3 A_{2ij}A_{1i}A_{1i}^\top A_{2ij}^\top \right).$$

The portions $\sum_{d=1}^{2} \gamma_d A_{2ij}A_{1i}M_d A_{1i}^T A_{2ij}^T$ and $\gamma_3 A_{2ij}A_{1i}A_{1i}^T A_{2ij}^T$ in the variance-covariance matrix yield:

$$\sum_{d=1}^{2} \gamma_d A_{2ij}A_{1i}M_d A_{1i}^T A_{2ij}^T = \gamma_1 A_{2ij}\left(\theta_{1i}I_{g_i}\right)A_{2ij}^T + \gamma_2 A_{2ij}M_{ii}^2 A_{2ij}^T$$

$$= \gamma_1 \theta_{1i}I_{g_{ij}} + \gamma_2 \theta_{2ij}I_{g_{ij}};$$

and

$$\gamma_3 A_{2ij}A_{1i}A_{1i}^T A_{2ij}^T = \gamma_3 A_{2ij}I_{g_i}A_{2ij}^T = \gamma_3 I_{g_{ij}}$$

which, clearly, complete the proof. $\qquad\square$

Proposition 4.2.3. *With $i \leq s$ and $j \leq k$ (symmetry applies)*

$$\Sigma(y_{ij}, y_{sk}) = \begin{cases} \mathbf{0}_{g_{ij}, g_{ik}} & i = s; j \neq k \\ \lambda_{ij}I_{g_{ij}} & i = s; j = k \\ \gamma_2 A_{2ij}A_{1i}M_2 A_{1s}^\top A_{2sk}^\top & i \neq s. \end{cases} \qquad (4.13)$$

Clearly the sub-models y_{ij} and y_{sk} are correlated for $i \neq s$, and not correlated for $i = s$.

Proof.

$$\begin{aligned} \Sigma(y_{ij}, y_{sk}) &= A_{2ij}A_{1i}\Sigma(y)A_{1s}^\top A_{2sk}^\top \qquad\qquad (4.14) \\ &= A_{2ij}A_{1i}(\gamma_1 M_1 + \gamma_2 M_2 + \gamma_3 I_n)A_{1s}^\top A_{2sk}^\top \\ &= \gamma_1 U_1 + \gamma_2 U_2 + \gamma_3 U_3, \end{aligned}$$

where $U_d = A_{2ij}A_{1i}M_d A_{1s}^\top A_{2sk}^\top$, with $M_3 = I_n$.

When $i = s$ and $j = k$ it holds $\Sigma(y_{ij}, y_{sk}) = \Sigma(y_{ij}) = \lambda_{ij}$, as we may remark from the previous proposition. When $i = s$ and $j \neq k$ it holds the following:

$$\begin{cases} U_1 &= \theta_{1i}A_{2ij}A_{2ik}^\top = \mathbf{0}_{g_{ij}, g_{ik}}; \\ U_2 &= A_{2ij}M_{ii}^2 A_{2ik}^\top = \mathbf{0}_{g_{ij}, g_{ik}}; \\ U_3 &= A_{2ij}(I_{g_{ij}, g_{ik}})A_{2ik} = \mathbf{0}_{g_{ij}, g_{ik}}. \end{cases}$$

Finally, when $i \neq s$ we found that

$$\begin{cases} U_1 &= A_{2ij}(\mathbf{0}_{g_i, g_s})A_{2sk}^\top = \mathbf{0}_{g_{ij}, g_{sk}}; \\ U_2 &= A_{2ij}W_{is}^2 A_{2sk}; \\ U_3 &= A_{2ij}(\mathbf{0}_{g_i, g_s})A_{2sk} = \mathbf{0}_{g_{ij}, g_{sk}}. \end{cases}$$

$\qquad\square$

4.2.2 Estimation For $r = 2$

From the subsection 4.2.1 we see that $P_2 P_1 y$ produces (with i and j respectively replaced by i_1 and i_2, for convenience) the fixed linear sub-models

$$y_{i_1 i_2} \sim \left(\mathbf{0}_{g_{i_1 i_2}}, \lambda_{i_1 i_2} I_{g_{i_1 i_2}} \right), \; i_1 = 1, \ldots, h_1, \; i_2 = 1, \ldots, h_{2 i_1}, \tag{4.15}$$

with $\lambda_{i_1 i_2} = \gamma_1 \theta_{1 i_1} + \gamma_2 \theta_{2 i_1 i_2} + \gamma_3$, the model $y \sim (\mathbf{0}_n, \gamma_1 M_1 + \gamma_2 M_2 + \gamma_3 I_n)$.

An estimator for $\lambda_{i_1 i_2}$ in the model (4.15) is

$$S_{i_1 i_2}^2 = \frac{y_{i_1 i_2}^\top y_{i_1 i_2}}{g_{i_1 i_2}},$$
$$i_1 = 1, \ldots, h_1, \; i_2 = 1, \ldots, h_{2 i_1}.$$

Indeed (see Theorem A.1.8),

$$\begin{aligned} E(S_{i_1 i_2}^2) &= \frac{1}{g_{i_1 i_2}} tr \left\{ \lambda_{i_1 i_2} I_{g_{i_1 i_2}} \right\} \\ &= \lambda_{i_1 i_2}. \end{aligned} \tag{4.16}$$

Thus

$$E(S_{i_1 i_2}^2) = \lambda_{i_1 i_2} = \gamma_1 \theta_{1 i_1} + \gamma_2 \theta_{2 i_1 i_2} + \gamma_3, \; i_1 = 1, \ldots, h_1, \; i_2 = 1, \ldots, h_{2 i_1}$$

so that , with $S = \begin{bmatrix} S_{11}^2 \\ \cdots \\ S_{1 h_{21}}^2 \\ S_{21}^2 \\ \cdots \\ S_{2 h_{22}}^2 \\ \cdots \\ S_{h_1 1}^2 \\ \cdots \\ S_{h_1 h_{2 h_1}}^2 \end{bmatrix}$, $\Theta = \begin{bmatrix} \theta_{11} & \theta_{211} & 1 \\ \cdots & \cdots & \cdots \\ \theta_{11} & \theta_{21 h_{21}} & 1 \\ \theta_{12} & \theta_{221} & 1 \\ \cdots & \cdots & \cdots \\ \theta_{12} & \theta_{22 h_{22}} & 1 \\ \cdots & \cdots & \cdots \\ \theta_{1 h_1} & \theta_{2 h_1 1} & 1 \\ \cdots & \cdots & \cdots \\ \theta_{1 h_1} & \theta_{2 h_1 h_{2 h_1}} & 1 \end{bmatrix}$, and $\gamma = \begin{bmatrix} \gamma_1 \\ \gamma_2 \\ \gamma_3 \end{bmatrix}$, we will have

$$E(S) = \Theta \gamma. \tag{4.17}$$

Thus, for $i_1 = 1, \ldots, h_1$, $i_2 = 1, \ldots, h_{2 i_1}$, equalizing the variances $\lambda_{i_1 i_2}$ to the correspondent estimators of $S_{i_1 i_2}^2$ it yields the following system of equations:

$$
\begin{aligned}
S_{11}^2 &= \gamma_1\theta_{11} + \gamma_2\theta_{211} + \gamma_3; \\
\cdots\;\; & \quad \ldots\ldots\ldots\ldots; \\
S_{1h_{21}}^2 &= \gamma_1\theta_{11} + \gamma_2\theta_{21h_{21}} + \gamma_3; \\
S_{21}^2 &= \gamma_1\theta_{12} + \gamma_2\theta_{221} + \gamma_3; \\
\cdots\;\; & \quad \ldots\ldots\ldots\ldots \\
S_{2h_{22}}^2 &= \gamma_1\theta_{12} + \gamma_2\theta_{22h_{22}} + \gamma_3; \\
\cdots\;\; & \quad \ldots\ldots\ldots\ldots; \\
S_{h_11}^2 &= \gamma_1\theta_{1h_1} + \gamma_2\theta_{2h_11} + \gamma_3; \\
\cdots\;\; & \quad \ldots\ldots\ldots\ldots; \\
S_{h_1h_{2h_1}}^2 &= \gamma_1\theta_{1h_1} + \gamma_2\theta_{2h_1h_{2h_1}} + \gamma_3;
\end{aligned}
$$

which in matrix notation becomes

$$S = \Theta\gamma. \tag{4.18}$$

Proposition 4.2.4. Θ *in equation* (4.18) *is a full-rank matrix.*

Proof. Let $\Theta = [c_1 c_2 c_3]$, where c_i, $i = 1,\ldots,3$, denotes its ith column. We thus have that: (a) the entries of c_1 are $\theta_{11},\ldots,\theta_{1h_1}$, the different eigenvalues of the symmetric matrix M_1; (b) the entries of c_2 are $\theta_{211},\ldots,\theta_{2h_1h_{2h_1}}$, where θ_{2i_11}, $\ldots$, $\theta_{2i_1h_{2i_1}}$ are the different eigenvalues of the symmetric matrix M_{ii}^2; (c) $c_3 = \mathbf{1}_{\sum_{i_1=1}^{h_1} h_{2i_1}}$, i.e., is a vector of 1's.

Let a and b be any scalars. Then,

$$
ac_1 + bc_3 = 0 \Leftrightarrow
\begin{cases}
\theta_{11} = -\frac{b}{a} \\
\quad\vdots \\
\theta_{1h_1} = -\frac{b}{a}
\end{cases}
\Leftrightarrow \theta_{11} = \ldots = \theta_{1h_1}
$$

which can not be truth (by construction $\theta_{1i_1} \neq \theta_{1i_1'}$, $i_1 \neq i_1'$), unless y is an 1×1 vector.

$$
ac_2 + bc_3 = 0 \Leftrightarrow
\begin{cases}
\theta_{2i_11} = -\frac{b}{a} \\
\quad\vdots \\
\theta_{2i_1h_{2i_1}} = -\frac{b}{a}
\end{cases}
\Leftrightarrow \theta_{2i_11} = \ldots = \theta_{2i_1h_{2i_1}}
$$

which cannot be truth once by construction $\theta_{2i_1i_2} \neq \theta_{2i_1i_2'}$, $i_2 \neq i_2'$, and ($M_{i_1i_1} \neq M_{i_1'i_1'}$ for $i_1 \neq i_1'$) $\theta_{2i_1i_2} \neq \theta_{2i_1'i_2}$ for $i_1 \neq i_1'$. Finally,

$$
ac_1 + bc_2 = 0 \Leftrightarrow
\begin{cases}
a\theta_{1i_1} + b\theta_{2i_11} = 0 \\
\quad\vdots \\
a\theta_{1i_1} + b\theta_{2i_1h_{2i_1}} = 0
\end{cases}
\Leftrightarrow
\begin{cases}
\theta_{2i_11} = -\frac{a}{b}\theta_{1i_1} \\
\quad\vdots \\
\theta_{2i_1h_{2i_1}} = -\frac{a}{b}\theta_{1i_1}
\end{cases}
\Leftrightarrow \theta_{2i_11} = \ldots = \theta_{2i_1h_{2i_1}}
$$

which cannot be truth (by construction $\theta_{2i_1i_2} \neq \theta_{2i_1i_2'}$, $i_2 \neq i_2'$). We must note as well that $\theta_{2i_1i_2} \neq \theta_{2i_1'i_2}$, $i_1 \neq i_1'$, once the $M_{i_1i_1} \neq M_{i_1'i_1'}$. Therefore, $r(\Theta) = 3$. $\qquad\square$

By the Theorem A.1.5 the matrix

$$
\Theta^\top \Theta =
\begin{bmatrix}
\sum_{i_1}^{h_1} \sum_{i_2}^{h_{2i_1}} \theta_{1i_1}^2 & \sum_{i_1}^{h_1} \sum_{i_2}^{h_{2i_1}} \theta_{1i_1} \theta_{2i_1 i_2} & \sum_{i_1}^{h_1} \sum_{i_2}^{h_{2i_1}} \theta_{1i_1} \\[2em]
\sum_{i_1}^{h_1} \sum_{i_2}^{h_{2i_1}} \theta_{1i_1} \theta_{2i_1 i_2} & \sum_{i_1}^{h_1} \sum_{i_2}^{h_{2i_1}} \theta_{2i_1 i_2}^2 & \sum_{i_1}^{h_1} \sum_{i_2}^{h_{2i_1}} \theta_{2i_1 i_2} \\[2em]
\sum_{i_1}^{h_1} \sum_{i_2}^{h_{2i_1}} \theta_{1i_1} & \sum_{i_1}^{h_1} \sum_{i_2}^{h_{2i_1}} \theta_{2i_1 i_2} & \sum_{i_1}^{h_1} \sum_{i_2}^{h_{2i_1}}
\end{bmatrix}
$$

is positive - definite, and by Theorem A.1.6 it follows that $\Theta^\top \Theta$ is a non-singular matrix. We, thus, take its inverse to be $(\Theta^\top \Theta)^{-1}$.

Now, Pre-multiplying the system (4.18) in both side by $\Theta^\top$ the resulting system of equations will be

$$\Theta^\top S = \Theta^\top \Theta \gamma, \tag{4.19}$$

whose unique solution (and therefore an estimator for γ) is

$$\hat{\gamma} = (\Theta^\top \Theta)^{-1} \Theta^\top S. \tag{4.20}$$

We call it *Sub-D* estimator.

Proposition 4.2.5. *The Sub-D estimator, $\hat{\gamma} = (\Theta^\top \Theta)^{-1} \Theta^\top S$, is an unbiased estimator of γ.*

Proof.

Indeed, $E(\hat{\gamma}) = E\left((\Theta^\top \Theta)^{-1} \Theta^\top S\right) = (\Theta^\top \Theta)^{-1} \Theta^\top E(S) = (\Theta^\top \Theta)^{-1} \Theta^\top \Theta \gamma = \gamma.$ $\qquad\square$

We may now be interested in find out the distribution of *Sub-D* estimator. In order to do it, we consider the next results.

Proposition 4.2.6.

$$
\Sigma \left(S_{ij}^2, S_{i^* j^*}^2 \right) =
\begin{cases}
(\boldsymbol{a})\ i = i^*; j \neq j^*: & 0, \\[0.5em]
(\boldsymbol{b})\ i = i^*; j = j^*: & 2\dfrac{\lambda_{ij}^2}{g_{ij}}, \\[0.5em]
(\boldsymbol{c})\ i \neq i^*: & 2\gamma_2^2 tr(\Omega M_2),
\end{cases}
$$

where $\Omega = \nabla_{ij} M_2 \nabla_{i^* j^*}$, with $\nabla_{ij} = \frac{A_{1i}^\top A_{2ij}^\top A_{2ij} A_{1i}}{g_{ij}}$.

Proof.

We have

$$
\begin{aligned}
\Sigma\left(S_{ij}^2, S_{i^*j^*}^2\right) &= \Sigma\left(\frac{y_{ij}^\top y_{ij}}{g_{ij}}; \frac{y_{i^*j^*}^\top y_{i^*j^*}}{g_{i^*j^*}}\right) \\
&= \Sigma\left(y^\top\left(\frac{A_{1i}^\top A_{2ij}^\top A_{2ij}A_{1i}}{g_{ij}}\right)y; \; y^\top\left(\frac{A_{1i^*}^\top A_{2i^*j^*}^\top A_{2i^*j^*}A_{1i^*}}{g_{i^*j^*}}\right)y\right) \\
&= \Sigma\left(y^\top \nabla_{ij} y; \; y^\top \nabla_{i^*j^*} y\right) \\
&= 2tr\left(\nabla_{ij} V \nabla_{i^*j^*} V\right) \\
&= 2\gamma_1^2 tr(\nabla_{ij}M_1\nabla_{i^*j^*}M_1) + 2\gamma_1\gamma_2 tr(\nabla_{ij}M_1\nabla_{i^*j^*}M_2) + 2\gamma_1\gamma_3 tr(\nabla_{ij}M_1\nabla_{i^*j^*}) \\
&\quad + 2\gamma_2\gamma_1 tr(\nabla_{ij}M_2\nabla_{i^*j^*}M_1) + 2\gamma_2^2 tr(\nabla_{ij}M_2\nabla_{i^*j^*}M_2) + 2\gamma_2\gamma_3 tr(\nabla_{ij}M_2\nabla_{i^*j^*}) \\
&\quad + 2\gamma_3\gamma_1 tr(\nabla_{ij}\nabla_{i^*j^*}M_1) + 2\gamma_3\gamma_2 tr(\nabla_{ij}\nabla_{i^*j^*}M_2) + 2\gamma_3^2 tr(\nabla_{ij}\nabla_{i^*j^*}) \\
&= \begin{cases}
i = i^*; j \neq j^* : & 0, \\
i = i^*; j = j^* : & 2\frac{\lambda_{ij}^2}{g_{ij}}, \\
\quad i \neq i^* : & 2\gamma_2^2 tr(\nabla_{ij}M_2\nabla_{i^*j^*}M_2).
\end{cases}
\end{aligned}
$$

For the case (**a**), that is $i = i^*$ and $j \neq j^*$, we have that

$$
\begin{aligned}
\nabla_{ij}M_1\nabla_{ij^*} &= \frac{1}{g_{ij}g_{ij^*}}A_{1i}^\top A_{2ij}^\top A_{2ij}A_{1i}M_1 A_{1i}^\top A_{2ij^*}^\top A_{2ij^*}A_{1i} \\
&= \frac{1}{g_{ij}g_{ij^*}}A_{1i}^\top A_{2ij}^\top A_{2ij}\left(\theta_{1i}I_{g_i}\right)A_{2ij^*}^\top A_{2ij^*}A_{1i} \\
&= \mathbf{0}_{g_i \times g_i} \; (\text{see (4.5) });
\end{aligned}
\tag{4.21}
$$

$$
\begin{aligned}
\nabla_{ij}M_2\nabla_{ij^*} &= \frac{1}{g_{ij}g_{ij^*}}A_{1i}^\top A_{2ij}^\top A_{2ij}A_{1i}M_2 A_{1i}^\top A_{2ij^*}^\top A_{2ij^*}A_{1i} \\
&= \frac{1}{g_{ij}g_{ij^*}}A_{1i}^\top A_{2ij}^\top A_{2ij}\left(M_{ii}^2\right)A_{2ij^*}^\top A_{2ij^*}A_{1i} \\
&= \mathbf{0}_{g_i \times g_i} \; (\text{see (4.9)});
\end{aligned}
\tag{4.22}
$$

$$
\begin{aligned}
\nabla_{ij}\nabla_{ij^*} &= \frac{1}{g_{ij}g_{ij^*}}A_{1i}^\top A_{2ij}^\top\left(\mathbf{0}_{g_{ii}\times g_{ij^*}}\right)A_{2ij^*}A_{1i} \\
&= \mathbf{0}_{g_i \times g_i}.
\end{aligned}
\tag{4.23}
$$

(4.21), (4.22) and (4.23) together with Proposition 2.1.1 (c) proofs the case (**a**).

For the case (**c**), that is $i \neq i^*$, the desired result becomes clear if we note that

$$
A_{1i}M_1A_{1i^*} = A_{1i}A_{1i^*} = \mathbf{0}_{g_i \times g_{i^*}},
$$

$tr(\nabla_{ij}M_2\nabla_{i^*j^*}M_1) = tr(\nabla_{i^*j^*}M_1\nabla_{ij}M_2) = 0$, and $tr(\nabla_{ij}M_2\nabla_{i^*j^*}) = tr(\nabla_{i^*j^*}\nabla_{ij}M_2) = 0$.

Finally, for the case (**b**), that is $i = i^*; j = j^*$, recalling $y_{ij} \sim \left(\mathbf{0}_{g_{ij}}, \lambda_{ij}I_{g_{ij}}\right)$, it holds

$$
\begin{aligned}
\Sigma\left(S_{ij}^2\right) &= \Sigma\left(\frac{y_{ij}^\top y_{ij}}{g_{ij}}; \frac{y_{ij}^\top y_{ij}}{g_{ij}}\right) = 2tr\left\{\frac{\lambda_{ij}}{g_{ij}}I_{g_{ij}}\frac{\lambda_{ij}}{g_{ij}}I_{g_{ij}}\right\} = 2\frac{\lambda_{ij}^2}{g_{ij}^2}tr\{I_{g_{ij}}\} \\
&= 2\frac{\lambda_{ij}^2}{g_{ij}},
\end{aligned}
\tag{4.24}
$$

and therefore the proof is complete. □

The next result introduce the variance-covariance matrix of the *Sub-D* estimator:

$$\hat{\gamma} = (\Theta^\top \Theta)^{-1} \Theta^\top S.$$

Proposition 4.2.7. *Let* $\Sigma_{S_{ij}S_{kl}}$ *denotes* $\Sigma(S_{ij}^2, S_{kl}^2)$. *Then,*

$$\Sigma(\hat{\gamma}) = (\Theta^\top \Theta)^{-1} \Theta^\top \Sigma(S) \Theta (\Theta^\top \Theta)^{-1}, \tag{4.25}$$

where $\Sigma(S) = \begin{bmatrix} D_1 & \Lambda_{12} & \Lambda_{13} & \dots & \Lambda_{1h_1} \\ \Lambda_{21} & D_2 & \Lambda_{23} & \dots & \Lambda_{2h_1} \\ \Lambda_{31} & \Lambda_{32} & D_3 & \dots & \Lambda_{3h_1} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \Lambda_{h_11} & \Lambda_{h_12} & \Lambda_{h_13} & \dots & D_{h_1} \end{bmatrix}$, *with* $D_i = 2 \begin{bmatrix} \frac{\lambda_{i1}^2}{g_{i1}} & 0 & \dots & 0 \\ 0 & \frac{\lambda_{i2}^2}{g_{i2}} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \frac{\lambda_{ih_{2i}}^2}{g_{ih_{2i}}} \end{bmatrix}$ *and* $\Lambda_{ks} =$

$$\begin{bmatrix} \Sigma_{S_{k1}S_{s1}} & \Sigma_{S_{k1}S_{s2}} & \dots & \Sigma_{S_{k1}S_{sh_{2s}}} \\ \Sigma_{S_{k2}S_{s1}} & \Sigma_{S_{k2}S_{s2}} & \dots & \Sigma_{S_{k2}S_{sh_{2s}}} \\ \vdots & \vdots & \ddots & \vdots \\ \Sigma_{S_{kh_{2k}}S_{s1}} & \Sigma_{S_{kh_{2k}}S_{s2}} & \dots & \Sigma_{S_{kh_{2k}}S_{sh_{2s}}} \end{bmatrix}.$$

The next section intends to generalize the method introduced here, that is, introducing the *Sub-D* estimator for a MLM with an arbitrary number of *variance components*.

4.2.3 The General Case: $r \geq 1$

The general *mixed linear models* may be expressed as (see (3.8))

$$z = X\beta + \sum_{i=1}^{r+1} X_i \beta_i, \tag{4.26}$$

with $X_{r+1} = I_m$, $X_i \in \mathcal{M}^{m_i \times p_i}$ and $\beta_i \in \mathbb{R}^{p_i}$, where $X \in \mathcal{M}^{m \times p}$ is the known (possible non full-rank) design matrix for the fixed effects, $\beta \in \mathbb{R}^p$ is the vector of the (unknown) fixed effects values, X_i, $i = 1, \dots, r+1$, the known and fixed (full-rank) design matrices for the variance-covariance structure, β_i, $i = 1, \dots, r$, the unknown vectors of the unobservable random effects, and β_{r+1} the vector of the residual errors, where β_i, $i = 1, \dots, r+1$, are mutually independent, such that $E(\beta_i) = \mathbf{0}_{p_i}$ and $\Sigma(\beta_i) = \gamma_i I_{p_i}$. Thus,

$$z \sim \left(X\beta, \sum_{i=1}^{r+1} \gamma_i N_i \right),$$

where $N_i = X_i X_i^\top$, and the unknown fixed parameters $\gamma_i \geq 0$, $i = 1 \dots, r$, $\gamma_{r+1} > 0$ denote the variance components.

In order to reduce the complexity of the algebraic manipulation in the estimator development process, and since the bias will be reduced with no loss of information needed to estimate the variance components (see Section 4.2), we will approach the model

$$y = B_o^\top z \sim \left(\mathbf{0}_n, \sum_{d=1}^{r+1} \gamma_d M_d \right), \tag{4.27}$$

with $M_d = B_o^\top N_d B_o \in \mathscr{S}^n$, $n = m - r(P_o)$, where B_o is a matrix whose columns are the eigenvectors associated to the null eigenvalues of $P_o = P_{R(X)}$, such that $B_o^\top B_o = I_{m-r(P_o)}$ and $B_o B_o^\top = P^*$, with $P^* = P_{R(X)^\perp} = I_m - P_o$.

One may note that $y = B_o^\top z = \sum_{d=1}^{r+1} B_o^\top X_d \beta_d$, where

$$\beta_d \sim (\mathbf{0}_{s_d},\ \gamma_d I_{s_d}),\ d = 1,\ldots,r,\ \beta_{r+1} \sim (\mathbf{0}_n,\ \gamma_d I_n).$$

With $i_1 = 1,\ldots,h_1$, $i_j = 1,\ldots,h_{j,i_1,\ldots,i_{j-1}}$, consider the finite sequence of r matrices $P_1, P_2, \ldots, P_r$ defined as follow:

$$P_1 = \begin{bmatrix} A_{11} \\ A_{12} \\ \vdots \\ A_{1h_1} \end{bmatrix} \in \mathscr{M}^{\left(\Sigma_{i_1}^{h_1} g_{i_1}\right) \times n}, \text{ with } A_{1i_1} \in \mathscr{M}^{(g_{i_1}) \times n} \text{ (note that } \sum_{i_1}^{h_1} g_{i_1} = n);$$

$$P_2 = \begin{bmatrix} P_{21} & 0 & \ldots & 0 \\ 0 & P_{22} & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & P_{2h_1} \end{bmatrix} \in \mathscr{M}^{\left(\Sigma_{i_1}^{h_1} \Sigma_{i_2}^{h_{2,i_1}} g_{i_1 i_2}\right) \times \left(\Sigma_{i_1}^{h_1} g_{i_1}\right)}, \text{ where}$$

$$P_{2i_1} = \begin{bmatrix} A_{2i_1 1} \\ A_{2i_1 2} \\ \vdots \\ A_{2i_1 h_{2i_1}} \end{bmatrix} \in \mathscr{M}^{\left(\Sigma_{i_2}^{h_{2,i_1}} g_{i_1 i_2}\right) \times g_{i_1}}, \text{ with } \sum_{i_2}^{h_{2,i_1}} g_{i_1 i_2} = g_{i_1} \text{ and } A_{2i_1 i_2} \in \mathscr{M}^{g_{i_1 i_2} \times g_{i_1}};$$

$$P_3 = \begin{bmatrix} P_{31} & 0 & \ldots & 0 \\ 0 & P_{32} & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & P_{3h_1} \end{bmatrix} \in \mathscr{M}^{\left(\Sigma_{i_1}^{h_1} \Sigma_{i_2}^{h_{2,i_1}} \Sigma_{i_3}^{h_{3,i_1,i_2}} g_{i_1 i_2 i_3}\right) \times \left(\Sigma_{i_1}^{h_1} \Sigma_{i_2}^{h_{2,i_1}} g_{i_1 i_2}\right)},$$

$$\text{where } P_{3i_1} = \begin{bmatrix} P_{3i_1 1} & 0 & \ldots & 0 \\ 0 & P_{3i_1 2} & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & P_{3i_1 h_{2,i_1}} \end{bmatrix} \in \mathscr{M}^{\left(\Sigma_{i_2}^{h_{2,i_1}} \Sigma_{i_3}^{h_{3,i_1,i_2}} g_{i_1 i_2 i_3}\right) \times \left(\Sigma_{i_2}^{h_{2,i_1}} g_{i_1 i_2}\right)} \text{ and}$$

$$P_{3i_1 i_2} = \begin{bmatrix} A_{3i_1 i_2 1} \\ A_{3i_1 i_2 2} \\ \vdots \\ A_{3i_1 i_2 h_{3,i_1,i_2}} \end{bmatrix} \in \mathscr{M}^{\left(\Sigma_{i_3}^{h_{3,i_1,i_2}} g_{i_1 i_2 i_3}\right) \times g_{i_1 i_2}}, \text{ with } \sum_{i_3}^{h_{3,i_1,i_2}} g_{i_1 i_2 i_3} = g_{i_1 i_2} \text{ and}$$

$$A_{3i_1 i_2 i_3} \in \mathscr{M}^{g_{i_1 i_2 i_3} \times g_{i_1 i_2}};$$

Thus, for $r \geq 2$, each matrix P_r will be given by (P_1 is given in (4.28)):

$$P_r \quad = \quad \begin{bmatrix} P_{r1} & 0 & \ldots & 0 \\ 0 & P_{r2} & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & P_{rh_1} \end{bmatrix} \tag{4.28}$$

$$\in \mathcal{M}\left(\Sigma_{i_1}^{h_1} \ldots \Sigma_{i_r}^{h_{r,i_1,\ldots,i_{r-1}}} g_{i_1 \ldots i_r} \right) \times \left(\Sigma_{i_1}^{h_1} \ldots \Sigma_{i_{(r-1)}}^{h_{(r-1),i_1,\ldots,i_{r-2}}} g_{i_1 \ldots i_{(r-1)}} \right),$$

where

$$P_{ri_1} \quad = \quad \begin{bmatrix} P_{ri_1 1} & 0 & \ldots & 0 \\ 0 & P_{ri_1 2} & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & P_{ri_1 h_{2,i_1}} \end{bmatrix}$$

$$\in \mathcal{M}\left(\Sigma_{i_2}^{h_{2,i_1}} \ldots \Sigma_{i_r}^{h_{r,i_1,\ldots,i_{r-1}}} g_{i_1 \ldots i_r} \right) \times \left(\Sigma_{i_2}^{h_{2,i_1}} \ldots \Sigma_{i_{(r-1)}}^{h_{(r-1),i_1,\ldots,i_{r-2}}} g_{i_1 \ldots i_{(r-1)}} \right),$$

$$\ldots \ldots \ldots \ldots \ldots$$

$$P_{ri_1 \ldots i_{(r-2)}} \quad = \quad \begin{bmatrix} P_{ri_1 \ldots i_{(r-2)} 1} & 0 & \ldots & 0 \\ 0 & P_{ri_1 \ldots i_{(r-2)} 2} & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & P_{ri_1 \ldots i_{(r-2)} h_{r-1,i_1,\ldots,i_{r-2}}} \end{bmatrix}$$

$$\in \mathcal{M}\left(\Sigma_{i_{(r-1)}}^{h_{(r-1),i_1,\ldots,i_{r-2}}} \Sigma_{i_r}^{h_{r,i_1,\ldots,i_{r-1}}} g_{i_1 \ldots i_r} \right) \times \left(\Sigma_{i_{(r-1)}}^{h_{(r-1),i_1,\ldots,i_{r-2}}} g_{i_1 \ldots i_{(r-1)}} \right),$$

$$\text{and} \quad P_{ri_1 \ldots i_{(r-1)}} = \begin{bmatrix} A_{ri_1 \ldots i_{(r-1)} 1} \\ A_{ri_1 \ldots i_{(r-1)} 2} \\ \vdots \\ A_{ri_1 \ldots i_{(r-1)} h_{r,i_1,\ldots,i_{r-1}}} \end{bmatrix} \in \mathcal{M}\left(\Sigma_{i_r}^{h_{r,i_1,\ldots,i_{r-1}}} g_{i_1 \ldots i_r} \right) \times g_{i_1 \ldots i_{(r-1)}},$$

$$\text{with} \quad \sum_{i_r}^{h_{r,i_1,\ldots,i_{r-1}}} g_{i_1 \ldots i_r} = g_{i_1 \ldots i_{(r-1)}}, \sum_{i_1}^{h_1} g_{i_1} = n, \, A_{ri_1 \ldots i_r} \in \mathcal{M}^{g_{i_1 \ldots i_r} \times g_{i_1 \ldots i_{(r-1)}}};$$

Theorem 4.2.8. *Let the matrices $P_1, P_2, \ldots, P_r$ defined above be such that:*

(c_1) *The columns of $A_{1i_1}^\top$, $i_1 = 1, \ldots, h_1$, form a set of $g_{i_1} = r(A_{1i_1}^\top)$ orthonormal eigenvectors associated to the different eigenvalues θ_{1i_1} of the matrix M_1 (θ_{1i_1} has multiplicity g_{i_1});*

(c_2) *The columns of $A_{2i_1i_2}^\top$, $i_2 = 1,\ldots,h_{2,i_1}$, form a set of $g_{i_1i_2} = r(A_{2i_1i_2}^\top)$ orthonormal eigenvectors associated to the different eigenvalues $\theta_{2i_1i_2}$ of the matrix $M_{i_1i_1}^2 = A_{1i_1}M_2A_{1i_1}^\top$ ($\theta_{2i_1i_2}$ has multiplicity $g_{i_1i_2}$);*

(c_3) *The columns of $A_{3i_1i_2i_3}^\top$, $i_3 = 1,\ldots,h_{3,i_1,i_2}$, form a set of $g_{i_1i_2i_3} = r(A_{3i_1i_2i_3}^\top)$ orthonormal eigenvectors associated to the different eigenvalues $\theta_{3i_1i_2i_3}$ of the matrix*

$$A_{2i_1i_2}M_{i_1i_1}^3A_{2i_1i_2}^\top = A_{2i_1i_2}A_{1i_1}M_3A_{1i_1}^\top A_{2i_1i_2}$$

($\theta_{3i_1i_2i_3}$ has multiplicity $g_{i_1i_2i_3}$);

$$\ldots\ldots \ldots \ldots\ldots$$

(c_r) *The columns of $A_{ri_1\ldots i_r}^\top$, $i_r = 1,\ldots,h_{r,i_1,\ldots,i_{r-1}}$, form a set of $g_{i_1\ldots i_r} = r(A_{ri_1\ldots i_r}^\top)$ orthonormal eigenvectors associated to the different eigenvalues $\theta_{ri_1\ldots i_r}$ of the matrix*

$$A_{(r-1)i_1\ldots i_{(r-1)}}\ldots A_{1i_1}M_rA_{1i_1}^\top \ldots A_{(r-1)i_1\ldots i_{(r-1)}}^\top \tag{4.29}$$

($\theta_{ri_1\ldots i_r}$ has multiplicity $g_{i_1\ldots i_r}$).

Then each matrix P_d, $d = 1,\ldots,r$, in the finite sequence of matrices $P_1,P_2,\ldots,Pr$ will be orthogonal matrices.

Proof.

According with the way P_d is defined (see (4.28)), since

$$P_{di_1\ldots i_{(d-1)}} = \begin{bmatrix} A_{di_1\ldots i_{(d-1)}1} \\ A_{di_1\ldots i_{(d-1)}2} \\ \vdots \\ A_{di_1\ldots i_{(d-1)}h_{d,i_1,\ldots,i_{d-1}}} \end{bmatrix}, \, i_{(d-1)} = 1,\ldots,h_{(d-1),i_1,\ldots,i_{d-2}},$$

and according with condition c_d we see that the matrices $P_{di_1\ldots i_{(d-1)}}$ are orthogonal. Thus, the desired result comes if we see that $P_d^\top P_d$ will be a diagonal blockwise matrix whose diagonal entries are $P_{di_1}^\top P_{di_1}$, $i_1 = 1,\ldots,h_1$. The diagonal entries $P_{di_1}^\top P_{di_1}$ will be diagonal blockwise matrices whose diagonal entries will be $P_{di_1i_2}^\top P_{di_1i_2}$, $i_2 = 1,\ldots,h_{2,i_1}$. Proceeding this way $d-2$ times, we will find that the diagonal entries of the blockwise matrices $P_{di_1\ldots i_{(d-2)}}^\top P_{di_1\ldots i_{(d-2)}}$, $i_{(d-2)} = 1,\ldots,h_{(d-2),i_1,\ldots,i_{d-3}}$, will be (see Theorem 2.2.8)

$$\begin{aligned} P_{di_1\ldots i_{(d-1)}}^\top P_{di_1\ldots i_{(d-1)}} &= A_{di_1\ldots i_{(d-1)}1}^\top A_{di_1\ldots i_{(d-1)}1} + \ldots + A_{di_1\ldots i_{(d-1)}h_{d,i_1,\ldots,i_{d-1}}}^\top A_{di_1\ldots i_{(d-1)}h_{d,i_1,\ldots,i_{d-1}}} \\ &= I_{g_{i_1\ldots i_{(d-1)}}}, \end{aligned}$$

reaching, therefore, the desired result. Proceeding in same way we would also see that

$$P_{di_1\ldots i_{(d-1)}}P_{di_1\ldots i_{(d-1)}}^\top$$

is a Blockwise diagonal matrix whose diagonal entries are

$$A_{di_1\ldots i_{(d-1)}1}A_{di_1\ldots i_{(d-1)}j}^\top, \, j = 1,\ldots,h_{d,i_1,\ldots,i_{d-1}},$$

so that $P_dP_d^\top$ is an identity matrix. $\qquad\square$

The model $P_r \ldots P_2 P_1 y$ will produces the following sub - models:

$$y_{i_1 \ldots i_r} = A_{r i_1 \ldots i_r} A_{(r-1) i_1 \ldots i_{(r-1)}} \ldots A_{2 i_1 i_2} A_{1 i_1} y, \tag{4.30}$$

$$i_1 = 1, \ldots, h_1, i_j = 1, \ldots, h_{j, i_1, \ldots, i_{j-1}}.$$

Definition 4.2.3. The orthogonal matrices P_r, $r \geq 1$, defined in (4.28) will be called *sub-diagonalization* matrices; here it make sense since the matrix $P_r \ldots P_2 P_1$ *sub-diagonalizes* the variance-covariance matrix, $\sum_{i=1}^{r+1} \gamma_d M_d$, where $M_{r+1} = I_n$.

At this point we are able to summarize the distribution of the sub-models in (4.30) as well as the cross-covariance between the sub-models, $y_{i_1 \ldots i_r}$ and $y_{i_1^* \ldots i_r^*}$ say, in the following results.

Proposition 4.2.9.

$$y_{i_1 \ldots i_r} \sim \left(\mathbf{0}_{g_{i_1 \ldots i_r}}, \ \lambda_{i_1 \ldots i_r} I_{g_{i_1 \ldots i_r}} \right),$$

where $\lambda_{i_1 \ldots i_r} = \sum_{d=1}^{r} \gamma_d \theta_{d i_1 \ldots i_d} + \gamma_{r+1}$.

Proof.

The proof becomes clear after looking to the proof of the proposition 4.2.2. □

Proposition 4.2.10. *With* $i_j = 1, \ldots, h_{j, i_1, \ldots, i_{j-1}}$, $j = 1, \ldots, r$, *and* $i_j \leq i_j^*$ *(symmetry applies) we have*

$$\Sigma(y_{i_1 \ldots i_r}, y_{i_1^* \ldots i_r^*}) = \begin{cases} \mathbf{0}_{g_{i_1 \ldots i_r}, g_{i_1^* \ldots i_r^*}} & i_1 = i_1^*, \ldots, i_{r-1} = i_{r-1}^*, \ i_r \neq i_r^* \\ \lambda_{i_1 \ldots i_r} I_{g_{i_1 \ldots i_r}} & i_j = i_j^*, \ j = 1, \ldots, r \\ \sum_{d=2}^{r} \gamma_d V_d & i_1 \neq i_1^* \\ \sum_{d=s+1}^{r} \gamma_d V_d & i_j = i_j^*, j = 1, \ldots, s-1, \ i_s \neq i_s^*, \ 1 < s < r-1 \end{cases} \tag{4.31}$$

where $V_d = A_{r i_1 \ldots i_r} \ldots A_{1 i} M_d A_{1 i_1}^\top \ldots A_{r i_1^* \ldots i_r^*}^\top$. *This result ensure that the sub-models* $y_{i_1 \ldots i_r}$ *and* $y_{i_1^* \ldots i_r^*}$ *are correlated for* $i_s \neq i_s^*$, $1 \leq s < r-1$, *and not correlated for* $i_s = i_s^*$, $s \leq r-1$, $i_r \neq i_r^*$.

Proof.

We proceed as at the case of three *variance components*.

Starting with case $i_1 = i_1^*, \ldots, i_{r-1} = i_{r-1}, \ i_r \neq i_r^*$, we have the following:

$$\begin{cases} V_1 &= A_{r i_1 \ldots i_r} (\theta_{1 i} I_{g_{i_1 \ldots i_{r-1}}}) A_{r i_1^* \ldots i_r^*}^\top = \mathbf{0}_{g_{i_1 \ldots i_r}, g_{i_1^* \ldots i_r^*}}; \\ V_2 &= A_{r i_1 \ldots i_r} (\theta_{2 i_1 i_2} I_{g_{i_1 \ldots i_{r-1}}}) A_{r i_1^* \ldots i_r^*}^\top = \mathbf{0}_{g_{i_1 \ldots i_r}, g_{i_1^* \ldots i_r^*}} \\ & \ldots \\ V_{r-1} &= A_{r i_1 \ldots i_r} (\theta_{(r-1) i_1 \ldots i_{r-1}} I_{g_{i_1 \ldots i_{r-1}}}) A_{r i_1^* \ldots i_r^*}^\top = \mathbf{0}_{g_{i_1 \ldots i_r}, g_{i_1^* \ldots i_r^*}} \\ V_r &= \mathbf{0}_{g_{i_1 \ldots i_r}, g_{i_1^* \ldots i_r^*}} \\ V_{r+1} &= \mathbf{0}_{g_{i_1 \ldots i_r}, g_{i_1^* \ldots i_r^*}} \end{cases}$$

The result for V_r is due to the fact that the columns of $A_{r i_1 \ldots i_r}^\top$ form a set of orthonormal eigenvectors associated to the different eigenvalues $\theta_{r i_1 \ldots i_r}$ of $A_{r i_1 \ldots i_r} \ldots A_{1 i_1} M_r A_{1 i_1}^\top \ldots A_{(r-1) i_1 \ldots i_{r-1}}^\top$.

When $i_j = i_j^*$, $j = 1, \ldots, r$, we have

$$\Sigma(y_{i_1 \ldots i_r}, y_{i_1^* \ldots i_r^*}) = \Sigma(y_{i_1 \ldots i_r}) = \lambda_{i_1 \ldots i_r},$$

as seen at the previous proposition.

For the case when $i_j = i_j^*, j = 1, \ldots, s-1, i_s \neq i_s^*, 1 \leq s < r-1$ we have the following:

$$
\begin{cases}
V_1 & = \ A_{ri_1 \ldots i_r}(\theta_{1i} I_{g_{i_1 \ldots i_{r-1}}}) A_{ri_1^* \ldots i_r^*}^\top = \mathbf{0}_{g_{i_1 \ldots i_r}, g_{i_1^* \ldots i_r^*}}; \\
& \cdots \\
V_{s-1} & = \ A_{ri_1 \ldots i_r} l \ldots A_{si_1 \ldots i_s}(\theta_{(s-1)i_1 \ldots i_{s-1}} I_{g_{i_1 \ldots i_{s-1}}}) A_{si_1^* \ldots i_s^*}^\top \ldots A_{ri_1^* \ldots i_r^*}^\top = \mathbf{0}_{g_{i_1 \ldots i_r}, g_{i_1^* \ldots i_r^*}},
\end{cases}
$$

but with $s \leq d \leq r$, we have $V_d = A_{ri_1 \ldots i_r} \ldots A_{si_1 \ldots i_s} M_d^{(s-1)} A_{si_1^* \ldots i_s^*} \ldots A_{ri_1^* \ldots i_r^*}^\top$, where $M_d^{(s-1)} = A_{(s-1)i_1 \ldots i_{(s-1)}} \ldots A_{1i} M_d A_{1i_1}^\top \ldots A_{(s-1)i_1 \ldots i_{(s-1)}}^\top$. Now, since $M_{r+1} = I_n$, it is straightforward verified that $V_{r+1} = \mathbf{0}_{g_{i_1 \ldots i_r}, g_{i_1^* \ldots i_r^*}}$.

Finally, the case $i_1 \neq i_1^*$ refer to the previous case when $s = 1$. $\qquad \square$

4.2.4 The General case: Estimation For $r \geq 1$

Recalling that for the mixed linear model (4.2), $P_r \ldots P_2 P_1 y$ produces the sub-models

$$
\begin{aligned}
y_{i_1 i_2 \ldots i_r} & \sim \ (\mathbf{0}_{g_{i_1 \ldots i_r}}, \lambda_{i_1 i_2 \ldots i_r} I_{g_{i_1 i_2 \ldots i_r}}), \\
& i_1 = 1, \ldots, h_1, i_j = 1, \ldots, h_{j, i_1, \ldots, i_{j-1}}
\end{aligned} \tag{4.32}
$$

where

$$
\lambda_{i_1 i_2 \ldots i_r} = \sum_{d=1}^{r} \gamma_d \theta_{di_1 \ldots i_d} + \gamma_{r+1}.
$$

The matrices P_d, $d = 1, \ldots, r$, are defined in the subsection (4.2.3).

An unbiased estimator of $\lambda_{i_1 i_2 \ldots i_r}$ in the model (4.32) is (the one based on least squares)

$$
S_{i_1 i_2 \ldots i_r}^2 \ = \ \frac{1}{g_{i_1 i_2 \ldots i_r}} y_{i_1 i_2 \ldots i_r}^\top y_{i_1 i_2 \ldots i_r}
$$

Indeed (see Theorem A.1.8 (b) and the explanation for (4.16)),

$$
\begin{aligned}
E\left(S_{i_1 i_2 \ldots i_r}^2\right) & = \ \frac{\lambda_{i_1 i_2 \ldots i_r}}{g_{i_1 i_2 \ldots i_r}} tr\left[I_{g_{i_1 i_2 \ldots i_r}}\right] \\
& = \ \lambda_{i_1 i_2 \ldots i_r}.
\end{aligned} \tag{4.33}
$$

For convenience, instead of $S_{i_1 i_2 \ldots i_r}^2$, we may sometimes use the notation $S_{i_1 i_2 \ldots i_{(r-1)} i_r}^2$ in what follows.

Thus

$$
\begin{aligned}
E(S_{i_1 i_2 \ldots i_r}^2) & = \ \sum_{d=1}^{r} \gamma_d \theta_{di_1 \ldots i_d} + \gamma_{r+1} \\
& = \ \gamma_1 \theta_{1i_1} + \gamma_2 \theta_{2i_1 i_2} + \ldots + \gamma_r \theta_{ri_1 i_2 \ldots i_r} + \gamma_{r+1},
\end{aligned}
$$

$$
i_1 = 1, \ldots, h_1; i_j = 1, \ldots, h_{j, i_1, \ldots, i_{j-1}}
$$

61

so that , with

$$S^* = \begin{bmatrix} S^2_{11\ldots11} \\ S^2_{11\ldots12} \\ \ldots \\ S^2_{11\ldots1h_{r,1,\ldots,1}} \\ S^2_{11\ldots21} \\ \ldots \\ S^2_{11\ldots2h_{r,1,\ldots,2}} \\ \ldots \\ S^2_{h_11\ldots11} \\ \ldots \\ S^2_{h_1h_{2,h_1}\ldots h_{r,h_1,\ldots,h_{r-1}}} \end{bmatrix},$$

$$\Theta^* = \begin{bmatrix} \theta_{11} & \theta_{211} & \theta_{3111} & \ldots & \theta_{r11\ldots11} & 1 \\ \theta_{11} & \theta_{211} & \theta_{3111} & \ldots & \theta_{r11\ldots12} & 1 \\ \ldots & \ldots & \ldots & \ldots & \ldots & \ldots \\ \theta_{11} & \theta_{211} & \theta_{3111} & \ldots & \theta_{r11\ldots1h_{r,1,\ldots,1,h_{r-1}}} & 1 \\ \theta_{11} & \theta_{211} & \theta_{3111} & \ldots & \theta_{r11\ldots21} & 1 \\ \ldots & \ldots & \ldots & \ldots & \ldots & \ldots \\ \theta_{11} & \theta_{211} & \theta_{3111} & \ldots & \theta_{r11\ldots2h_{r,1,\ldots,2,h_{r-1}}} & 1 \\ \ldots & \ldots & \ldots & \ldots & \ldots & \ldots \\ \theta_{1h_1} & \theta_{2h_11} & \theta_{3h_111} & \ldots & \theta_{rh_11\ldots11} & 1 \\ \ldots & \ldots & \ldots & \ldots & \ldots & \ldots \\ \theta_{1h_1} & \theta_{2h_1h_{2,h_1}} & \theta_{3h_1h_{2,h_1}h_{3,h_1,h_2}} & \ldots & \theta_{rh_1h_{2,h_1}\ldots h_{(r-1),h_1,\ldots,h_{r-2}}h_{r,h_1,\ldots,h_{r-1}}} & 1 \end{bmatrix},$$

and $\gamma^* = \begin{bmatrix} \gamma_1 \\ \gamma_2 \\ \gamma_3 \\ \ldots \\ \gamma_r \\ \gamma_{(r+1)} \end{bmatrix}$, we will have

$$E(S^*) = \Theta^* \gamma^*. \tag{4.34}$$

Thus, for $i_1 = 1, \ldots, h_1$, $i_j = 1, \ldots, h_{j,i_1,\ldots,i_{j-1}}$, $j > 1$, equalizing the variances $\lambda_{i_1 i_2 \ldots i_r}$ to the correspondent estimators $S^2_{i_1 i_2 \ldots i_r}$ yields the following system of equations (in matrix notation)

$$S^* = \Theta^* \gamma^*. \tag{4.35}$$

Proposition 4.2.11. Θ^* *in equation* (4.35) *is a full-rank matrix.*

Proof. The proof is done in same fashion as for the Proposition 4.2.4. Indeed, by construction $\theta_{1i_1} \neq \theta_{1i_1'}$ they are the different eigenvalues of M_1, $\theta_{2i_1i_2} \neq \theta_{2i_1i_2'}$ the distinct eigenvalues of $M_{ii}^2 =$

$A_{1i_1}M_2A_{1i_1}^\top$, $\theta_{3i_1i_2i_3} \neq \theta_{3i_1i_2i_3'}$ the distinct eigenvalues of $A_{2i_1i_2}A_{1i_1}M_2A_{1i_1}^\top A_{2i_1i_2}^\top$, $\ldots$, $\theta_{ri_1i_2\ldots i_{(r-1)}i_r} \neq \theta_{ri_1i_2\ldots i_{(r-1)}i_r'}$ the distinct eigenvalues of

$$A_{(r-1)i_1i_2\ldots i_{(r-1)}}\ldots A_{1i_1}M_rA_{1i_1}^\top \ldots A_{(r-1)i_1i_2\ldots i_{(r-1)}}^\top$$

where $i_j \neq i_j'$, $j = 1,\ldots,r$. Thus we have that $r(\Theta^*) = r+1$. $\square$

By the Theorem A.1.5, with $\sum$ denoting $\sum_{i_1}^{h_1}\sum_{i_2}^{h_2}\ldots\sum_{i_r}^{h_r}$, the matrix

$$(\Theta^*)^\top \Theta^* = \begin{bmatrix} \sum\theta_{1i_1}^2 & \sum\theta_{1i_1}\theta_{2i_1i_2} & \sum\theta_{1i_1}\theta_{3i_1i_2i_3} & \cdots & \sum\theta_{1i_1}\theta_{ri_1\ldots ir} & \sum\theta_{1i_1} \\ \\ \sum\theta_{1i_1}\theta_{2i_1i_2} & \sum\theta_{2i_1i_2}^2 & \theta_{2i_1i_2}\theta_{3i_1i_2i_3} & \cdots & \sum\theta_{2i_1i_2}\theta_{ri_1\ldots ir} & \sum\theta_{2i_1i_2} \\ \\ \sum\theta_{1i_1}\theta_{3i_1i_2i_3} & \sum\theta_{2i_1i_2}\theta_{3i_1i_2i_3} & \sum\theta_{3i_1i_2i_3}^2 & \cdots & \sum\theta_{3i_1i_2i_3}\theta_{ri_1\ldots ir} & \sum\theta_{3i_1i_2i_3} \\ \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ \\ \sum\theta_{1i_1}\theta_{ri_1\ldots ir} & \sum\theta_{2i_1i_2}\theta_{ri_1\ldots ir} & \sum\theta_{3i_1i_2i_3}\theta_{ri_1\ldots ir} & \cdots & \sum\theta_{ri_1\ldots ir}^2 & \sum\theta_{ri_1\ldots ir} \\ \\ \sum\theta_{1i_1} & \sum\theta_{2i_1i_2} & \sum\theta_{3i_1i_2i_3} & \cdots & \sum\theta_{ri_1\ldots ir} & \sum \end{bmatrix}$$

is positive - definite, and by Theorem A.1.6 it follows that $(\Theta^*)^\top\Theta^*$ is non-singular; that is, it is invertible. We take its inverse to be $\left((\Theta^*)^\top\Theta^*\right)^{-1}$.

Now, premultiplying the system (4.35) in both side by Θ^T the resulting system of equations will be

$$(\Theta^*)^\top S = (\Theta^*)^\top \Theta^* \gamma, \tag{4.36}$$

whose unique solution (and therefore any estimator of γ) is

$$\hat{\gamma} = \left((\Theta^*)^\top\Theta^*\right)^{-1}(\Theta^*)^\top S. \tag{4.37}$$

Proposition 4.2.12. $\hat{\gamma} = ((\Theta^*)^\top\Theta^*)^{-1}(\Theta^*)^\top S$ *is an unbiased estimator of* $\gamma = \begin{bmatrix} \gamma_1 \\ \gamma_2 \\ \gamma_3 \\ \ldots \\ \ldots \\ \gamma_r \\ \gamma_{(r+1)} \end{bmatrix}$, *where*

$$\begin{bmatrix} \hat{\gamma_1} \\ \hat{\gamma_2} \\ \hat{\gamma_3} \\ \ldots \\ \ldots \\ \hat{\gamma_r} \\ \hat{\gamma_{(r+1)}} \end{bmatrix}.$$

Indeed,

$$
\begin{aligned}
E(\hat{\gamma}) &= E\left(\left((\Theta^*)^\top \Theta^*\right)^{-1}(\Theta^*)^\top S\right) = \left((\Theta^*)^\top \Theta^*\right)^{-1}(\Theta^*)^\top E(S) \\
&= \left((\Theta^*)^\top \Theta^*\right)^{-1}(\Theta^*)^\top \Theta^* \gamma = \gamma.
\end{aligned}
\tag{4.38}
$$

4.3 Improving The Sub-D Estimator

As demonstrated in previous section through theoretical results, and corroborated with numerical simulations (see chapter 5), the Sub-D estimator provides unbiased estimates whatever the mixed linear design we choose, having overcame the performance of ANOVA estimator in crossed and nested designs with unbalanced data, and the one of REML estimator in nested design; indeed, as may be seen in chapter 5, when applied to nested design, REML provides low accurate estimates for some parameters, whereas when applied to crossed and nested designs ANOVA provides unrealistic estimates for some parameters. It must be point out that all designs referenced here have some empty cells.

As we may remark from the chapter 5, despite its great performance, the numerical test reveals that the *Sub-D* estimator produces estimates with a higher variability comparing to the variability of the estimates produced with REML estimator. This problem seems to be due to the non null correlation between the sub-models (it is null for models with one or two variance components and orthogonal models). For example, for the case of models with three *variance components* (see model (4.3)), the sub-models $y_{ij} = A_{2ij}A_{1i}y$ and $y_{sk} = A_{2sk}A_{1s}y$, with $i, s = 1, \ldots, h_1$, $j, k = 1, \ldots, h_{2i}, h_{2s}$, are correlated as seen before. From (4.13) we see that the variance-covariance matrix of the new model $P_2 P_1 y$ is a blockwise matrix whose diagonal matrices are $D_1, \ldots, D_{h_1}$, where $D_i = diag(\lambda_{i1} \ldots \lambda_{ih_{2i}})$, corresponding to $\Sigma(y_{ij}, y_{sk})$ for $i = s$, $j = k$, and the off diagonal matrices are the non-null matrices $\gamma_2 A_{2ij}A_{1i}M_2 A_{1s}^\top A_{2sk}^\top$, corresponding to $\Sigma(y_{ij}, y_{sk})$ for $i \neq s$; this last one, i.e the cross-covariance between y_{ij} and y_{sk} for $i \neq s$, is not considered by the *Sub-D* estimator on its deduction process as we may have seen.

In attempt to reduce the variability of the estimated values produced with *Sub-D* estimator, we introduce now an improved estimator for variance components; the improvement is achieved by incorporating the structure of the covariance in the *Sub-D* estimator deduction process. This new estimator will be referred to as *Sub-DI* estimator.

Recall (from the previous section) that the *Sub-D* estimator is a solution (in γ) for the system of equations

$$
S^* = \Theta^* \gamma^*
\tag{4.39}
$$

(for the case of three *variance components*: $S = \Theta \gamma$), which consists in equating the sum of square errors for each sub-model to its respective expectation; clearly, as previously remarked, it does not take in account the correlation between the sub-models $y_{i_1,\ldots,i_r}$ and $y_{i_1^*,\ldots,i_r^*}$ for $i_j = i_j^*, j = 1, \ldots, s-1$, $i_s \neq i_s^*$, $1 \leq s < r-1$ (for the case of three *variance components*: y_{2ij} and y_{2sk}, $i \neq s$).

Considering that in data collecting process for some experiment a large amount of data is required, the repeating process may not be so practical, so that the unbiasedness of a particular

estimator might be neglected in favor of others estimators with better performance such as consistency, asymptoticity, among others. Hence, considering the idea of improving the *Sub-D* estimator in that sense, we may be interested in incorporate the system of equations which take into account the cross-covariance between the sub-models $y_{i_1,\ldots,i_r}$ and $y_{i_1^*,\ldots,i_r^*}$ for $i_j = i_j^*, j = 1,\ldots,s-1$, $i_s \neq i_s^*$, $1 \leq s < r-1$ to the system of equations (4.39), been aware that the unbiasedness might not be preserved.

Given that the expectations for all sub-models $y_{i_1,\ldots,i_r}$ are null vectors, i.e. $E(y_{i_1,\ldots,i_r}) = \mathbf{0}_{g_{i_1,\ldots,i_r}}$, it holds (see Theorem A.1.7 and Proposition 4.2.10)

$$
\begin{aligned}
E(y_{i_1,\ldots,i_r} y_{i_1^*,\ldots,i_r^*}^\top) &= \Sigma(y_{i_1,\ldots,i_r},\, y_{i_1^*,\ldots,i_r^*}) \\
&= \sum_{d=s+1}^{r} \gamma_d V^d_{i_1,\ldots,i_r,i_1^*,\ldots,i_r^*}
\end{aligned}
\tag{4.40}
$$

for $i_j = i_j^*, j = 1,\ldots,s-1$, $i_s \neq i_s^*$, $1 < s < r-1$, where

$$
V^d_{i_1,\ldots,i_r,i_1^*,\ldots,i_r^*} = A_{r i_1 \ldots i_r} \ldots A_{1 i} M_d A_{1 i_1^*}^\top \ldots A_{r i_1^* \ldots i_r^*}^\top.
$$

The result in (4.40) is equal to $\sum_{d=2}^{r} \gamma_d V^d_{i_1,\ldots,i_r,i_1^*,\ldots,i_r^*}$ when $s = 1$ (see (4.31)).

Recall the system of equations that produce the *Sub-D* estimator,

$$
\begin{aligned}
E(S^2_{i_1 i_2 \ldots i_{(r-1)} i_r}) &= E\left(\frac{y_{i_1,\ldots,i_r}^\top y_{i_1,\ldots,i_r}}{g_{i_1,\ldots,i_r}} \right) \\
&= \sum_{d=1}^{r} \gamma_d \theta_{d i_1 \ldots i_d} + \gamma_{r+1} \\
& \quad i_1 = 1,\ldots,h_1; i_j = 1,\ldots,h_{j,i_1,\ldots,i_{j-1}}
\end{aligned}
\tag{4.41}
$$

(see system (4.39) for matrix notation), system which, as stated before, doesn't take into account the fact that the sub-models are correlated. Now, noting that with $y_{i_1,\ldots,i_r}^{(k)}$, $k = 1,\ldots,g_{i_1,\ldots,i_r}$, denoting the kth element of the sub-model $y_{i_1,\ldots,i_r}$, we find that

$$
E(y_{i_1,\ldots,i_r}^{(k)} y_{i_1^*,\ldots,i_r^*}^{(l)}) = \sum_{d=s+1}^{r} \gamma_d v^{d(kl)}_{i_1,\ldots,i_r,i_1^*,\ldots,i_r^*},
\tag{4.42}
$$

where $v^{d(kl)}_{i_1,\ldots,i_r,i_1^*,\ldots,i_r^*}$ is the entry at row k and column l of matrix $V^d_{i_1,\ldots,i_r,i_1^*,\ldots,i_r^*}$.

Finally, equating $\frac{y_{i_1,\ldots,i_r}^\top y_{i_1,\ldots,i_r}}{g_{i_1,\ldots,i_r}}$ to its expectation $\sum_{d=1}^{r} \gamma_d \theta_{d i_1 \ldots i_d} + \gamma_{r+1}$, and $y_{i_1,\ldots,i_r} y_{i_1^*,\ldots,i_r^*}^\top$ to its expectation $\sum_{d=s+1}^{r} \gamma_d V^d_{i_1,\ldots,i_r,i_1^*,\ldots,i_r^*}$, and putting together both the equations in one single system of equations (see (4.43) below), the new estimator for the *variance components*, which we shall call *Sub-D* improved, denoted as *Sub-DI*, will be the solution in $\tilde{\gamma}^\top = [\tilde{\gamma}_1 \ldots \tilde{\gamma}_{r+1}]$ for the system

$$
\begin{cases}
\dfrac{y_{i_1,\ldots,i_r}^\top y_{i_1,\ldots,i_r}}{g_{i_1,\ldots,i_r}} = \sum_{d=1}^{r} \tilde{\gamma}_d \theta_{d i_1 \ldots i_d} + \tilde{\gamma}_{r+1} \\
\qquad \text{for } i_j = i_j^*, j = 1,\ldots,r \\
y_{i_1,\ldots,i_r}^{(k)} y_{i_1^*,\ldots,i_r^*}^{(l)} = \sum_{d=s+1}^{r} \tilde{\gamma}_d v^{d(kl)}_{i_1,\ldots,i_r,i_1^*,\ldots,i_r^*} \\
\qquad \text{for } i_j = i_j^*, j = 1,\ldots,s-1, i_s \neq i_s^*, 1 < s < r-1,
\end{cases}
\tag{4.43}
$$

with $i_1 = 1,\ldots,h_1; i_j = 1,\ldots,h_{j,i_1,\ldots,i_{j-1}}$; it must be noted that right hand-side of the second equation in the system (4.43) is equal to $\sum_{d=s}^{r} \tilde{\gamma}_d v^{d(kl)}_{i_1,\ldots,i_r,i_1^*,\ldots,i_r^*}$ when $s = 1$.

Remark 4.3.1. As it may be seen, for models with one or two variance components (fixed effect or "one-way" models) the sub-models $y_i = A_{1i}y$ will not be correlated so that, for these models, the *Sub-DI* estimator is equivalent to *Sub-D*, since the right-hand side of the second system of equations in (4.43) will be null.

NUMERICAL RESULTS

In order to test the performance of *Sub-D* estimator, as well as its improved version, the *Sub-DI* estimator, we carry out numerical tests for several types of designs for MLM. More precisely, we test their performance using a balanced and an unbalanced "one-way designs", an unbalanced "two-way crossed design" and an unbalanced "two-way nested design". The test will be done comparing its performance with the performance of REML and ANOVA estimators. For the REML estimator we will use the lme4 package (for R software), which covers approximately the same ground as the earlier nlme package, providing also functions for fitting and analyzing mixed models, but with some additional advantages in what concerning the MLM, namely (see Bates et al. [10]):

(1) It uses modern and efficient linear algebra methods and reference classes to avoid undue copying of large objects; it is therefore likely to be faster and more memory-efficient than nlme;

(2) It includes generalized linear mixed model (GLMM) capabilities (via the glmer function);

(3) It offers built-in facilities for likelihood profiling and parametric bootstrapping;

(4) Notwithstanding it is not (yet) as well-documented as nlme, it is designed to be more modular than nlme, making it easier for end-users to re-use its components for extensions of the basic mixed model framework;

(5) It also allows more flexibility for specifying different functions for optimizing over the random-effects variance-covariance parameters.

For the computational implementation of ANOVA method we will follow Sahai and Ojeda [63].

REML is the preferred method for estimating the variance components in MLM (Diffey et al. [17]); it is therefore likely that for the *Sub-D* and *Sub-DI* estimators a reasonable way to prove their values goes through producing results which can be compared with those of the REML estimator.

5.1 The Choice Of Designs

Due to their widely application, with special emphasis on genetics, process control and quality control and improvement, most of works on "search for optimal designs for variance components estimation" and "estimation procedures for variance components" focus on one-way designs (see model in (3.3)), nested designs (see the "two-way nested design" in (3.4) or (3.5)) and crossed designs (see the "two - way crossed design" in (3.6) or (3.7)); it is therefore likely that on the matter of search for its place within estimators for variance components on MLM the Sub-D and Sub-DI estimators must prove their values facing these designs.

On this purpose, and since the smaller the sample, more difficulty to provide accurate estimates for either fixed and random linear models or MLM, we will test the performance of Sub-D and Sub-DI using reasonably small samples. The test will be done using an *unbalanced* "two-way crossed design" and an *unbalanced* "two-way nested design" with 12 observations each, constituting, reasonably, a small sample. We also used two "one-way designs", one with *balanced* data and the other with *unbalanced* data. Both the "one-way designs" will include 21 observations.

For the "two-way designs" (crossed and nested) some cells will be taken to be empty in order to take the methods to the extreme.

The test will be done proceeding as follows: for the same 10000 observations of the underlying model, the three estimators will be simultaneously applied and, in order to favor the comparison, the results will be organized in different tables.

We will use the R software for all the simulations in this work, and the results will be rounded to four decimal places.

5.2 The Performance I: "One-Way Design"

Recall the "one-way design" (see Section 3)

$$z = Z\mu + Z_1\alpha + e,$$

where

$$Z = \mathbf{1}_{\sum_{s=1}^{k} n_s}, \ Z_1 = \begin{bmatrix} 1 & 0 & \ldots & \ldots & 0 \\ \ldots & \ldots & \ldots & \ldots & \ldots \\ 1 & 0 & \ldots & \ldots & 0 \\ 0 & 1 & 0 & \ldots & 0 \\ \ldots & \ldots & \ldots & \ldots & \ldots \\ 0 & 1 & 0 & \ldots & 0 \\ \ldots & \ldots & \ldots & \ldots & \ldots \\ \ldots & \ldots & \ldots & \ldots & \ldots \\ 0 & 0 & \ldots & 0 & 1 \\ \ldots & \ldots & \ldots & \ldots & \ldots \\ 0 & 0 & \ldots & 0 & 1 \end{bmatrix} \in \mathscr{M}^{\left(\Sigma_{s=1}^{k} n_s\right) \times k}, \text{ and } \alpha = \begin{bmatrix} \alpha_1 \\ \ldots \\ \alpha_k \end{bmatrix}.$$

5.2.1 Balanced "One Way Design"

Lets consider a particular *balanced* design of the "one-way design" with $k = 3$ and $n_i = 7$, $i = 1, 2, 3$. Thus we will have that

$$z \sim (Z\mu, V), \text{ where } V = \gamma_1 Z_1 Z_1^\top + \gamma_2 I_{\sum_{i=1}^3 n_i}, \tag{5.1}$$

with $Z = 1_{21}$ and $Z_1 Z_1^\top = \begin{bmatrix} J_7 & 0_{7,7} & 0_{7,7} \\ 0_{7,7} & J_7 & 0_{7,7} \\ 0_{7,7} & 0_{7,7} & J_7 \end{bmatrix}$. Let B_o be a matrix whose columns are the eigenvectors associated to the null eigenvalues of $\frac{1}{21} J_{21}$. Then $B_o B_o^\top = I_{21} - \frac{1}{21} J_{21}$ and $B_o^\top B_o = I_{20}$, and so the new model will be

$$y = B_o^\top z \sim (0_{20}, \ \gamma_1 M + \gamma_2 I_{20}),$$

where $M = B_o^\top Z_1 Z_1^\top B_o$.

Since $r(Z_1 Z_1^\top) = 3$ it follows that (see Theorem 2.1.5) $r(M) = r(B_o^\top Z_1 Z_1^\top B_o) = 3$. The eigenvalues of M are $\theta_{11} = 7$, with multiplicity equal to 2, and $\theta_{12} = 0$ with multiplicity equal to 18. Thus $\Theta = \begin{bmatrix} 7 & 1 \\ 0 & 1 \end{bmatrix}$.

Now, assuming $\alpha \sim \mathcal{N}(0_3, \ \gamma_1 I_3)$ and $e \sim \mathcal{N}(0_{20}, \ \gamma_2 I_{20})$, with $\gamma_2 = 1$ fixed, the particular design can be rewritten as

$$y = B_o^\top Z_1 \alpha + B_o^\top e. \tag{5.2}$$

For each $\gamma_1 \in \{0.1, 0.25, 0.5, 0.75, 1, 2, 5\}$, we simulated 10000 observations of the model stated in (5.1) and for each observation the Sub-D is applied and the variance components γ_1 and γ_2 (error) are estimated. In order to compare the performance of Sud-D with the ones of REML and ANOVA, for the same 10000 observation of y, REML and ANOVA methods were applied and the average of the estimated values presented in Tables 5.1, 5.2, 5.3, and 5.4.

Table 5.1: Estimates for γ_1 using Sub-D, REML and ANOVA.

γ_1	0.1	0.25	0.5	0.75	1	2	5
Sub-D	0.0999	0.2454	0.5036	0.7529	1.0083	1.9966	5.0378
REML	0.1379	0.2715	0.5201	0.7646	1.0175	2.0014	5.0402
ANOVA	0.0999	0.2454	0.5036	0.7529	1.0083	1.9966	5.0378

Table 5.2: Mean Square Error of estimated γ_1 using Sub-D, REML and ANOVA.

γ_1	0.1	0.25	0.5	0.75	1	2	5
Sub-D	0.2474	0.3910	0.6607	0.9040	1.1624	2.1466	5.2111
REML	0.2213	0.3698	0.6466	0.8935	1.1541	2.1421	5.2088
ANOVA	0.2474	0.3910	0.6607	0.9040	1.1624	2.1466	5.2111

As it may be seen from the Tables 5.1, 5.2, 5.3, and 5.4, the average estimates for variance components as well as their respective standard deviation using *Sub-D* estimator are exactly the

Table 5.3: Estimates for γ_2 (error) using Sub-D, REML and ANOVA.

γ_1	0.1	0.25	0.5	0.75	1	2	5
Sub-D	1.0018	1.0020	0.9983	1.0041	1.0019	0.9990	1.0001
REML	0.9752	0.9838	0.9867	0.9959	0.9955	0.9926	0.9973
ANOVA	1.0018	1.0020	0.9983	1.0041	1.0019	0.9990	1.0001

Table 5.4: Mean Square Error of estimated γ_2 (error) using Sub-D, REML and ANOVA.

γ_1	0.1	0.25	0.5	0.75	1	2	5
Sub-D	0.3320	0.3324	0.3303	0.3357	0.3339	0.3328	0.3332
REML	0.3190	0.3233	0.3248	0.3314	0.3306	0.3312	0.3321
ANOVA	0.3320	0.3324	0.3303	0.3357	0.3339	0.3328	0.3332

same as those obtained using ANOVA estimator. As seen, the *Sub-D* and ANOVA estimates are extremely unbiased unlike the REML ones. Indeed, despite its slightly smaller variation than *Sub-D* and ANOVA (see Tables 5.2 and 5.4), REML estimator provided low accurate estimates for small values (see the REML estimates for $\gamma_1 = 0.1, 0.25, 0.5$); therefore, Tables 5.1 and 5.2 suggest that *Sub-D* and ANOVA estimators are preferred, particularly when the variance components are small values.

5.2.2 Unbalanced "One-Way Design"

Now we consider a particular *unbalanced* design of the one-way design with $k = 3$, $n_1 = 2$, $n_2 = 12$ and $n3 = 7$, having therefore that $y \sim (Z\mu, V)$, where

$$V = \gamma_1 Z_1 Z_1^\top + \gamma_2 I_{\sum_{i=1}^3 n_i},$$

with $Z = 1_{21}$ and $Z_1 Z_1^\top = \begin{bmatrix} J_2 & \mathbf{0}_{2,12} & \mathbf{0}_{2,7} \\ \mathbf{0}_{12,2} & J_{12} & \mathbf{0}_{12,7} \\ \mathbf{0}_{7,2} & \mathbf{0}_{7,12} & J_7 \end{bmatrix}$.

With B_o a matrix whose columns are the eigenvectors associated to the null eigenvalues of $\frac{1}{21} J_{21}$, yielding so $B_o B_o^\top = I_{21} - \frac{1}{21} J_{21}$ and $B_o^\top B_o = I_{20}$, we have that the eigenvalues of $M = B_o^\top Z_1 Z_1^\top B_o$ will be $\theta_{11} = 8.9321$, $\theta_{12} = 2.6869$ and $\theta_{13} = 0$, with θ_{13} having root equal 18.

Once again, for each $\gamma_1 \in \{0.1, 0.25, 0.5, 0.75, 1, 2, 5\}$, we simulate 10000 observations of the model stated in (5.1), and for each observation the three methods (Sub-D, REML and ANOVA) are applied and the variance components γ_1 and γ_2 are estimated. The average of the estimated values are available in Tables 5.5, 5.6, 5.7, and 5.8.

From Tables 5.5 and 5.7 we may see that the *Sub-D* steel providing unbiased estimates although with larger dispersion (see Tables 5.6 and 5.8). For γ_2 REML provides accurate estimates, although not so accurate as those provides by *Sub-D*. But for values 0.1, 0.25, 0.5 and 0.75 of γ_1 the estimates are not so accurate as those when the values for γ_1 are 1, 2 and 5. Although no accurate, ANOVA provides acceptable estimates for γ_1, but for γ_2 it produces unrealistic estimates.

Table 5.5: Estimates for γ_1 using Sub-D, REML and ANOVA.

γ_1	0.1	0.25	0.5	0.75	1	2	5
Sub-D	0.1059	0.2457	0.4950	0.7458	1.0063	2.0045	4.9568
REML	0.1584	0.2931	0.5322	0.7757	1.0276	2.0221	4.9622
ANOVA	0.1979	0.3410	0.5915	0.8410	1.1009	2.1018	5.0516

Table 5.6: Mean Square Error of estimated γ_1 using Sub-D, REML and ANOVA.

γ_1	0.1	0.25	0.5	0.75	1	2	5
Sub-D	0.3234	0.5506	0.9107	1.3175	1.7040	3.2762	7.8761
REML	0.2928	0.4456	0.6941	0.9617	1.2195	2.2725	5.2650
ANOVA	0.3012	0.4500	0.7139	1.0063	1.2943	2.4726	5.9077

Table 5.7: Estimates for γ_2 using Sub-D, REML and ANOVA.

γ_1	0.1	0.25	0.5	0.75	1	2	5
Sub-D	0.9954	1.0057	1.0069	1.0019	1.0018	1.0095	1.0002
REML	0.9735	0.9809	0.9823	0.9876	0.9946	0.9917	0.9951
ANOVA	0.4585	0.4543	0.4521	0.4507	0.4517	0.4511	0.4514

Table 5.8: Mean Square Error of estimated γ_2 using Sub-D, REML and ANOVA.

γ_1	0.1	0.25	0.5	0.75	1	2	5
Sub-D	0.9137	1.2392	1.7770	2.4312	3.0030	5.3988	12.333
REML	0.3212	0.3240	0.3248	0.3261	0.3323	0.3328	0.3312
ANOVA	0.5854	0.5878	0.5888	0.5902	0.5898	0.5912	0.5895

5.3 The Performance II: "Two-Way Crossed Design"

In this section we approach the test for the performance of *Sub-D* and *Sub-DI* in an *unbalanced* "two-way crossed design" with no interaction (a MLM with three variance components), comparing it to the ones of REML and ANOVA. Consider the the "two-way crossed design" (with no interaction)

$$z = X\mu + X_1\beta_1 + X_2\beta_2 + e, \tag{5.3}$$

where $z \sim (X\mu,\ \gamma_1 N_1 + \gamma_2 N_2 + \gamma_3 I_{12})$, with $N_j = X_j X_j^\top$, $j = 1, 2$, whose design matrices are

$$X = 1_{12},\ X_1 = \begin{bmatrix} 1_3 & 0_3 & 0_3 \\ 0_5 & 1_5 & 0_5 \\ 0_4 & 0_4 & 1_4 \end{bmatrix},\ \text{and } X_2 = \begin{bmatrix} 1_2 & 0_2 & 0_2 \\ 0_3 & 1_3 & 0_3 \\ 0_4 & 0_4 & 1_4 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{bmatrix}.$$

Let B_o be a matrix whose columns are the eigenvectors associated to the null eigenvalues of

$\frac{1}{12}J_{12}$. Then, since $B_o B_o^\top = I_{12} - \frac{1}{12}J_{12}$ and $B_o^\top B_o = I_{11}$, the new model to be approached will be

$$y = B_o^\top z \sim \left(\mathbf{0}_{11}, \; \gamma_1 M_1 + \gamma_2 M_2 + \gamma_3 I_{11}\right),$$

where $M_d = B_o^\top N_d B_o$, $d = 1, 2$.

The eigenvalues of $M_1 = B_o^\top N_d B_o$ are $\theta_{11} = 4.5000$, $\theta_{12} = 3.3333$ and $\theta_{13} = 0$. θ_{13} has root equal to 9. Recalling that A_{11}, A_{12} and A_{13} are matrices whose columns are the eigenvectors associated to the eigenvalues θ_{11}, θ_{12} and θ_{13}, respectively, we have that $M_{11}^2 = A_{11} M_2 A_{11}^\top = 1.23809$ and $M_{22}^2 = A_{12} M_2 A_{12}^\top = 0.52857$ are 1×1 matrices, and $M_{33}^2 = A_{13} M_2 A_{13}^\top$ is a 9×9 matrix.

For the matrices M_{11}^2, M_{22}^2 and M_{33}^2 we have the following: M_{11}^2 has eigenvalue $\theta_{211} = 1.23809$; M_{22}^2 has eigenvalue $\theta_{221} = 0.52857$; M_{33}^2 has 3 eigenvalues: $\theta_{231} = 3.96142$; $\theta_{232} = 2.27191$; $\theta_{233} = 0$. θ_{233} has multiplicity equal to 7.

Finally we found that $\Theta = \begin{bmatrix} 4.5000 & 1.23809 & 1 \\ 3.33333 & 0.52857 & 1 \\ 0 & 3.96142 & 1 \\ 0 & 2.27191 & 1 \\ 0 & 0 & 1 \end{bmatrix}$.

Assuming $\beta_i \sim \mathcal{N}(\mathbf{0}_3, \; \gamma_i I_3)$, $i = 1, 2$, and $e \sim \mathcal{N}(\mathbf{0}_{12}, I_{12})$, for each pair of γ_1 and γ_2 taking values in $\{0.1, 0.25, 0.5, 1, 2, 5, 10\}$ and $\gamma_3 = 1$ fixed, the model in (5.3) is observed 1000 times and for each observation the four methods Sub-D, Sub-DI, REML and ANOVA are applied and the variance components γ_1, γ_2, and γ_3 (error) estimated. See Tables 5.9, 5.11, and 5.13 for the respective average of the estimated values of γ_1, γ_2, and γ_3. For the mean square error of the respective estimated values see Tables 5.10, 5.12, and 5.14.

As it may be pointed out, *Sub-D* and *Sub-DI* estimators provided accurate estimates for all the parameters γ_1, γ_2 and γ_3 while, similarly to what happened in the "one-way designs", REML estimator provided accurate estimates for γ_1, $\gamma_2, \gamma_3 \in \{1, 2, 5\}$ but however not so accurate as those provided by *Sub-D* and *Sub-DI*. For γ_1, γ_2, $\gamma_3 \in \{0.1, 0.25, 0.5\}$ REML produced estimates with low accuracy. All the tree estimators, *Sub-D*, *Sub-DI*, and REML, produces comparable and accurate estimates for γ_3. It must be pointed out, however, that the estimates produced with *Sub-D* and *Sub-DI* have a slightly higher standard deviation than the ones produced with REML. Despite their accuracy, as seen from the Tables 5.9, 5.11, and 5.13, the estimates produced with *Sub-DI* have in general smaller mean square error than those produced with *Sub-D*, as it was expected; indeed, for γ_2 and γ_3 it is clearly that *Sub-DI* produces estimates with smaller mean square error than *Sub-D*, whereas for γ_1 the mean square error are somewhat comparable.

The ANOVA estimator provided acceptable estimates for γ_1 (although with low accuracy) but for γ_2 and γ_3 the estimates provided are extremely unrealistic in such a way that we may not be interested in apply such a method in any study for which there may have empty cells in the model.

5.4 The Performance III: "Two-Way Nested Design"

In this section we test for the performances of *Sub-D* and *Sub-D* in an *unbalanced* "two-way nested design" (which is a MLM with three variance components) with some empty cells, comparing them to the ones of REML and ANOVA.

Supposing that the data come from the "two-way nested design"

$$z = Z\mu + Z_1\beta_1 + Z_2\beta_2 + e, \tag{5.4}$$

where $z \sim (Z\mu, \; \gamma_1 N_1 + \gamma_2 N_2 + \gamma_3 I_{12})$, with $N_j = Z_j Z_j^\top$, $j = 1, 2$, whose design matrices are

$$Z = 1_{12}, \; Z_1 = \begin{bmatrix} 1_4 & 0_4 & 0_4 \\ 0_5 & 1_5 & 0_5 \\ 0_3 & 0_3 & 1_3 \end{bmatrix}, \; \text{and } Z_2 = \begin{bmatrix} 1_2 & 0_2 & 0_2 & 0_2 \\ 0_4 & 1_4 & 0_4 & 0_4 \\ 0_3 & 0_3 & 1_3 & 0_3 \\ 0_3 & 0_3 & 0_3 & 1_3 \end{bmatrix}.$$

Letting B_o be the matrix defined in Section 5.4, holding therefore $B_o B_o^\top = I_{12} - \frac{1}{12}J_{12}$ and $B_o^\top B_o = I_{11}$, the new model to be approached will be

$$y = B_o^\top z \sim (0_{11}, \; \gamma_1 M_1 + \gamma_2 M_2 + \gamma_3 I_{11}),$$

where $M_d = B_o^\top N_d B_o$, $d = 1, 2$.

The eigenvalues of M_1 are $\theta_{11} = 4.5000$, $\theta_{12} = 3.3333$ and $\theta_{13} = 0$. θ_{13} has root equal to 9. We have that $M_{11}^2 = A_{11}M_2A_{11}^\top = 1.5$ and $M_{22}^2 = A_{12}M_2A_{12}^\top = 2.9333$ are 1×1 matrices, and $M_{33}^2 = A_{13}M_2A_{13}^\top$ is a 9×9 matrix.

For the matrices M_{11}^2, M_{22}^2 and M_{33}^2 we have the following: M_{11}^2 has eigenvalue $\theta_{211} = 1.5000$; M_{22}^2 has eigenvalue $\theta_{221} = 2.9333$; M_{33}^2 has 4 eigenvalues: $\theta_{231} = 3.3135$; $\theta_{232} = 1.0864$; $\theta_{233} = 0$. θ_{233} has multiplicity equal to 7.

The matrix Θ is given by $\Theta = \begin{bmatrix} 4.5000 & 1.5000 & 1 \\ 3.33333 & 2.9333 & 1 \\ 0 & 3.3136 & 1 \\ 0 & 1.0864 & 1 \\ 0 & 0 & 1 \end{bmatrix}$. Assuming $\beta_1 \sim \mathcal{N}(0_3, \; \gamma_1 I_3)$, $\beta_2 \sim \mathcal{N}(0_4, \; \gamma_2 I_4)$, and $e \sim \mathcal{N}(0_{12}, I_{12})$, for each pair of γ_1 and γ_2 taking values in $\{0.25, 0.5, 0.75, 1, 2, 5\}$ and $\gamma_3 = 1$ fixed, the model in (5.4) is observed 10000 times. For each observation the four methods *Sub-D*, *Sub-DI*, REML and ANOVA are applied and the variance components γ_1, γ_2, and γ_3 (error) are estimated. See Tables 5.15, 5.17, and 5.19 for the average of the estimated values of γ_1, γ_2, and γ_3. For the standard deviation of the respective estimated values see Tables 5.16, 5.18, and 5.20.

Taking a look at Tables 5.15, 5.17, and 5.19, and comparing the averages of the estimated values from the *Sub-D* and *Sub-DI* methods to the ones of the REML and ANOVA, the reader may easily reaches the conclusion that the only ones accurate estimates are the ones provided by *Sub-D* and *Sub-DI*. More over, both REML and ANOVA methods provided estimates with low accuracy, being that ANOVA produces unrealistic estimates, as we may see by looking to those of

γ_1; therefore, as suggested by the results, the *Sub-D* and *Sub-DI* estimators are the preferred ones. Once again, Tables 5.16, 5.18, and 5.20, evidence that the estimates produced by *Sub-DI* have in general smaller mean square error than those produced by *Sub-D*, as it was suspected; indeed, it is clearly that *Sub-DI* produces estimates with smaller mean square errors than *Sub-D* for all parameters γ_1, γ_2 and γ_3.

In general, we may point out that *Sub-D* and *Sub-DI* kept a constant and accurate performance towards all designs approached here.

Table 5.9: Estimation on unbalanced "two-way crossed design": estimate for γ_1.

$\gamma_1 \setminus \gamma_2$		0.1	0.25	0.5	0.75	1	2	5
0.1	Sub-D	0.1020	0.0966	0.1025	0.1022	0.1020	0.0933	0.0985
	Sub-DI	0.1029	0.0953	0.1034	0.1022	0.1007	0.0967	0.0964
	REML	0.1954	0.1979	0.2081	0.2076	0.2145	0.2128	0.2185
	ANOVA	0.1049	0.1006	0.1124	0.1176	0.1243	0.1438	0.2073
0.25	Sub-D	0.2561	0.2523	0.2528	0.2474	0.2542	0.2583	0.2589
	Sub-DI	0.2568	0.2526	0.2529	0.2458	0.2550	0.2589	0.2650
	REML	0.3248	0.3318	0.3288	0.3381	0.3372	0.3494	0.3536
	ANOVA	0.2588	0.2591	0.2595	0.2653	0.2721	0.3029	0.3770
0.5	Sub-D	0.5089	0.4920	0.4909	0.4983	0.4879	0.4974	0.4971
	Sub-DI	0.5072	0.4916	0.4878	0.4951	0.4863	0.5006	0.5037
	REML	0.5599	0.5498	0.5594	0.5739	0.5608	0.5822	0.5763
	ANOVA	0.5078	0.4956	0.5019	0.5178	0.5095	0.5538	0.6147
0.75	Sub-D	0.7548	0.7598	0.7544	0.7583	0.7512	0.7406	0.7630
	Sub-DI	0.7570	0.7578	0.7552	0.7616	0.7507	0.7403	0.7565
	REML	0.7946	0.8055	0.8015	0.8044	0.8136	0.8131	0.8140
	ANOVA	0.7583	0.7632	0.7661	0.7715	0.7781	0.7974	0.8658
1	Sub-D	1.0148	0.9784	1.0245	1.0153	1.0191	1.0330	0.9971
	Sub-DI	1.0185	0.97985	1.0238	1.0182	1.0280	1.0345	0.9840
	REML	1.0424	1.0167	1.0487	1.0580	1.0473	1.0479	1.0682
	ANOVA	1.0178	0.9852	1.0258	1.0376	1.0441	1.0665	1.1033
2	Sub-D	2.0089	1.9906	2.0344	1.9651	1.9832	1.9749	1.9980
	Sub-DI	2.0113	1.9906	2.0335	1.9697	1.9844	1.9823	2.0032
	REML	2.0369	2.0153	2.0459	1.9837	2.0303	1.9997	2.0259
	ANOVA	2.0209	2.0003	2.0396	1.9877	2.0202	2.0301	2.1150
5	Sub-D	4.9369	5.0755	4.9950	4.9530	4.9347	4.9842	5.011
	Sub-DI	4.9358	5.0829	4.9925	4.9561	4.9337	4.9833	5.0243
	REML	4.9537	5.0854	4.9756	4.9554	4.9729	4.9930	5.0284
	ANOVA	4.9476	5.0915	4.9851	4.9653	4.9725	5.0262	5.1425

Table 5.10: Estimation on unbalanced "two-way crossed design": mean square error of γ_1.

$\gamma_1 \setminus \gamma_2$		0.1	0.25	0.5	0.75	1	2	5
0.1	Sub-D	0.4265	0.4653	0.5451	0.6043	0.6766	0.8785	1.4192
	Sub-DI	0.4384	0.4947	0.6005	0.6866	0.7888	1.1236	2.0992
	REML	0.3588	0.3646	0.3829	0.3808	0.3993	0.3946	0.4053
	ANOVA	0.3981	0.4287	0.4976	0.5516	0.6242	0.8707	1.6423
0.25	Sub-D	0.5885	0.6425	0.6946	0.7692	0.8473	1.0855	1.6419
	Sub-DI	0.5979	0.6684	0.7474	0.801	0.9610	1.3219	2.2891
	REML	0.5047	0.5318	0.5215	0.5380	0.5388	0.5617	0.5611
	ANOVA	0.5454	0.5908	0.6284	0.6967	0.7663	1.0306	1.7816
0.5	Sub-D	0.8737	0.8952	0.9688	1.0374	1.1045	1.3920	2.0089
	Sub-DI	0.8773	0.9133	1.0161	1.1112	1.2080	1.6215	2.6554
	REML	0.7730	0.7619	0.7809	0.7920	0.7943	0.8305	0.8204
	ANOVA	0.7974	0.8067	0.8751	0.9316	0.9935	1.2856	2.0502
0.75	Sub-D	1.1637	1.2015	1.2532	1.3624	1.4050	1.6609	2.3786
	Sub-DI	1.1646	1.2206	1.2951	1.4427	1.5141	1.8935	2.9986
	REML	1.0399	1.0308	1.0412	1.0481	1.0644	1.0664	1.0884
	ANOVA	1.0660	1.0858	1.1282	1.2137	1.2711	1.5166	2.3151
1	Sub-D	1.4102	1.4306	1.5823	1.5741	1.6999	1.9768	2.602
	Sub-DI	1.4091	1.4447	1.6264	1.6434	1.8053	2.2082	3.1934
	REML	1.2619	1.2679	1.3061	1.3037	1.3290	1.3277	1.3394
	ANOVA	1.2879	1.2987	1.4086	1.4221	1.5291	1.7784	2.4834
2	Sub-D	2.5588	2.5064	2.7236	2.7200	2.7397	3.0712	3.9100
	Sub-DI	2.5395	2.4985	2.7428	2.7724	2.8164	3.2845	4.4813
	REML	2.3262	2.2226	2.3279	2.3147	2.3379	2.3153	2.3499
	ANOVA	2.3408	2.2589	2.4200	2.4490	2.4808	2.7471	3.5315
5	Sub-D	5.8325	6.0874	6.1362	6.0655	6.1062	6.4789	7.2090
	Sub-DI	5.7587	6.0489	6.0841	6.0547	6.1438	6.6297	7.7095
	REML	5.2549	5.4043	5.2836	5.2692	5.3028	5.4262	5.3899
	ANOVA	5.2852	5.4842	5.4189	5.4136	5.4887	5.8101	6.4425

Table 5.11: Estimation on unbalanced "two-way crossed design": estimates for γ_2.

$\gamma_1 \setminus \gamma_2$		0.1	0.25	0.5	0.75	1	2	5
0.1	Sub-D	0.0924	0.2588	0.5071	0.7379	1.0085	2.0307	4.9809
	Sub-DI	0.0956	0.2539	0.5107	0.7380	1.0036	2.0434	4.9732
	REML	0.1881	0.3224	0.5575	0.7758	1.0277	2.0654	4.9705
	ANOVA	0.0514	0.1420	0.2843	0.3979	0.5417	1.1056	2.7132
0.25	Sub-D	0.1006	0.2566	0.5152	0.7577	0.9986	2.0215	4.9437
	Sub-DI	0.1034	0.2576	0.5154	0.7520	1.0014	2.0237	4.9660
	REML	0.1995	0.3355	0.5644	0.7967	1.0303	2.0357	4.9766
	ANOVA	0.0540	0.1398	0.2856	0.3970	0.5487	1.0844	2.7067
0.5	Sub-D	0.1020	0.2546	0.5117	0.7528	0.9874	1.9948	4.9851
	Sub-DI	0.0969	0.2533	0.5005	0.7410	0.9815	2.0068	5.0096
	REML	0.1978	0.3375	0.5669	0.8003	1.0284	2.0320	5.0399
	ANOVA	0.0517	0.1336	0.2677	0.4025	0.5341	1.0760	2.7609
0.75	Sub-D	0.1051	0.2540	0.4884	0.7529	0.9888	2.0423	5.1110
	Sub-DI	0.1129	0.2469	0.4913	0.7650	0.9869	2.0411	5.0871
	REML	0.2173	0.3330	0.5588	0.7988	1.0323	2.0682	5.0763
	ANOVA	0.0616	0.1337	0.2584	0.4105	0.5353	1.0855	2.7212
1	Sub-D	0.0893	0.2433	0.5175	0.7388	0.9771	2.0575	4.9926
	Sub-DI	0.1029	0.2487	0.5149	0.7494	1.0096	2.0633	4.9445
	REML	0.2029	0.3349	0.5642	0.8046	1.0395	2.0483	4.9404
	ANOVA	0.0609	0.1380	0.2752	0.4042	0.5536	1.1185	2.6696
2	Sub-D	0.0927	0.2475	0.5193	0.7497	0.9790	1.9829	5.0865
	Sub-DI	0.1015	0.2475	0.5160	0.7665	0.9833	2.0101	5.1056
	REML	0.2189	0.3417	0.5783	0.8136	1.0361	2.0231	5.1289
	ANOVA	0.0557	0.1343	0.2752	0.4220	0.5321	1.0912	2.7774
5	Sub-D	0.0922	0.2319	0.5240	0.7410	0.9817	2.0399	4.9576
	Sub-DI	0.0879	0.2591	0.5150	0.7525	0.9778	2.0365	5.0062
	REML	0.2143	0.3387	0.5661	0.7996	1.0304	2.0571	5.0435
	ANOVA	0.0421	0.1390	0.2734	0.4205	0.5286	1.1144	2.7291

Table 5.12: Estimation on unbalanced "two-way crossed design": mean square error of γ_2.

$\gamma_1 \setminus \gamma_2$		0.1	0.25	0.5	0.75	1	2	5
0.1	Sub-D	0.5008	0.7061	1.0061	1.3086	1.6618	2.9127	6.8080
	Sub-DI	0.4142	0.5659	0.8098	1.0468	1.3235	2.3450	5.3876
	REML	0.3424	0.5043	0.7603	0.9840	1.2595	2.3142	5.2795
	ANOVA	0.3620	0.4709	0.6759	0.8699	1.0744	1.9233	4.4902
0.25	Sub-D	0.5350	0.7178	1.0312	1.3329	1.6886	2.9682	6.7244
	Sub-DI	0.4660	0.6112	0.8576	1.1750	1.3794	2.3622	5.3736
	REML	0.3615	0.5258	0.7803	1.0312	1.2999	2.2881	5.2680
	ANOVA	0.3978	0.5040	0.7058	0.8980	1.1144	1.8964	4.5355
0.5	Sub-D	0.6005	0.7483	1.0783	1.3944	1.6789	2.9889	6.8052
	Sub-DI	0.5388	0.6698	0.9255	1.1862	1.3942	2.4615	5.5450
	REML	0.3600	0.5283	0.7967	1.0581	1.2773	2.3164	5.4290
	ANOVA	0.4560	0.5449	0.7536	0.9562	1.1415	1.9774	4.5846
0.75	Sub-D	0.6936	0.8316	1.0757	1.3991	1.6769	3.0952	6.9255
	Sub-DI	0.6436	0.7701	0.9785	1.2404	1.4622	2.5848	5.6195
	REML	0.4020	0.5319	0.7876	1.0421	1.2745	2.4075	5.3686
	ANOVA	0.5295	0.6219	0.7901	0.9963	1.1892	2.0226	4.5250
1	Sub-D	0.7509	0.8828	1.1666	1.4571	1.7414	3.0626	6.8400
	Sub-DI	0.7139	0.8383	1.0946	1.3171	1.5755	2.6295	5.5654
	REML	0.3744	0.5316	0.7815	1.0618	1.3003	2.3582	5.2727
	ANOVA	0.5922	0.6761	0.8661	1.0470	1.2547	2.1240	4.5393
2	Sub-D	1.1915	1.2631	1.5131	1.7181	1.9761	3.1884	7.1184
	Sub-DI	1.1034	1.1835	1.4475	1.6422	1.8429	2.9196	6.0685
	REML	0.4044	0.5505	0.8176	1.0615	1.2938	5.5311	5.5311
	ANOVA	0.8395	0.9103	1.0754	1.2759	1.4361	4.8245	4.8245
5	Sub-D	2.4776	2.5930	2.6931	2.8484	3.0409	3.8888	7.3530
	Sub-DI	2.1245	2.3748	2.4603	2.6703	2.8962	3.8043	6.7677
	REML	0.3972	0.5339	0.8068	1.0770	1.2965	2.3694	5.4310
	ANOVA	1.5199	1.7097	1.7869	1.9590	2.1322	2.9358	5.3588

Table 5.13: Estimation on unbalanced "two-way crossed design": estimates for γ_3.

$\gamma_1 \setminus \gamma_2$		0.1	0.25	0.5	0.75	1	2	5
0.1	Sub-D	1.0003	0.9809	0.9931	0.9953	0.9823	1.0235	0.9783
	Sub-DI	0.9936	0.9909	0.9859	0.9951	0.9923	0.9978	0.9939
	REML	0.8743	0.8820	0.8963	0.9071	0.9017	0.9111	0.9130
	ANOVA	1.0269	1.0704	1.1515	1.2355	1.3111	1.6450	2.5577
0.25	Sub-D	0.9958	0.9979	0.9869	0.9942	0.9884	0.9953	1.0459
	Sub-DI	0.9903	0.9958	0.9864	1.0058	0.9827	0.9907	1.0006
	REML	0.8841	0.8934	0.9149	0.9151	0.9232	0.9300	0.9356
	ANOVA	1.0271	1.0750	1.1578	1.2441	1.3100	1.6468	2.5643
0.5	Sub-D	0.9818	0.9909	0.9936	1.0054	1.0037	1.0508	1.0629
	Sub-DI	0.9941	0.9936	1.0163	1.0293	1.0159	1.0264	1.0133
	REML	0.8971	0.9074	0.9241	0.9287	0.9361	0.9478	0.9523
	ANOVA	1.0287	1.0787	1.1663	1.2403	1.3161	1.6528	2.5637
0.75	Sub-D	1.0076	0.9913	1.0133	0.9953	1.0205	1.0277	0.9388
	Sub-DI	0.9918	1.0056	1.0074	0.9706	1.0242	1.0303	0.9872
	REML	0.9047	0.9207	0.9373	0.9440	0.9441	0.9518	0.9663
	ANOVA	1.0333	1.0825	1.1680	1.2411	1.3190	1.6661	2.6409
1	Sub-D	1.0097	1.0119	0.9637	1.0189	1.0232	0.9327	0.9450
	Sub-DI	0.9821	1.0009	0.9688	0.9975	0.9573	0.9210	1.0425
	REML	0.9112	0.9280	0.9411	0.9376	0.9470	0.9573	0.9725
	ANOVA	1.0239	1.0813	1.1611	1.2311	1.3000	1.6266	2.5923
2	Sub-D	1.0289	1.0108	0.9801	1.0254	1.0462	1.0443	1.0425
	Sub-DI	1.0110	1.0108	0.9868	0.9913	1.0376	0.9891	1.0040
	REML	0.9099	0.9321	0.9511	0.9564	0.9550	0.9714	0.9810
	ANOVA	1.0284	1.0819	1.1673	1.2352	1.3151	1.6297	2.6166
5	Sub-D	1.0208	1.0414	0.9286	0.9929	1.0531	0.9930	1.1287
	Sub-DI	1.0295	0.9863	0.9469	0.9698	1.0608	1.0000	1.0302
	REML	0.9165	0.9378	0.9563	0.9582	0.9766	0.9851	0.9949
	ANOVA	1.0349	1.0757	1.1596	1.2189	1.3304	1.6437	2.5896

Table 5.14: Estimation on unbalanced "two-way crossed design": mean square error of γ_3.

$\gamma_1 \setminus \gamma_2$		0.1	0.25	0.5	0.75	1	2	5
0.1	Sub-D	0.8624	0.9828	1.239	1.4553	1.7420	2.826	5.937
	Sub-DI	0.8218	0.9053	1.0368	1.1662	1.2860	1.6812	2.5478
	REML	0.4491	0.4583	0.4644	0.4620	0.4657	0.4680	0.4744
	ANOVA	0.5334	0.5907	0.6873	0.7884	0.9241	1.5014	3.1818
0.25	Sub-D	1.0037	1.1400	1.3434	1.5773	1.8297	2.8760	5.9856
	Sub-DI	0.9729	1.0854	1.2086	1.3110	1.4999	1.9449	2.9199
	REML	0.4575	0.4631	0.4707	0.4757	0.4709	0.4856	0.4824
	ANOVA	0.5500	0.5915	0.7033	0.8204	0.9572	1.5076	3.1751
0.5	Sub-D	1.2462	1.3516	1.5495	1.8026	2.0044	3.0443	6.1719
	Sub-DI	1.2250	1.3353	1.4732	1.6623	1.7873	2.3800	3.5615
	REML	0.4578	0.4678	0.4789	0.4819	0.4838	0.4903	0.5002
	ANOVA	0.5669	0.6233	0.7260	0.8464	0.9610	1.5279	3.2260
0.75	Sub-D	1.5570	1.6491	1.8059	2.0345	2.2366	3.2449	6.3159
	Sub-DI	1.4990	1.6254	1.7820	1.9825	2.1356	2.7443	4.0825
	REML	0.4645	0.4739	0.4866	0.4868	0.4911	0.4999	0.5051
	ANOVA	0.6051	0.6566	0.7555	0.8823	0.9918	1.6313	3.3963
1	Sub-D	1.8128	1.8726	2.0869	2.3070	2.4787	3.3826	6.2802
	Sub-DI	1.7232	1.8410	2.1071	2.2386	2.4644	3.0640	4.4107
	REML	0.4772	0.4804	0.4878	0.4924	0.4960	0.5038	0.5131
	ANOVA	0.6363	0.6888	0.7912	0.8945	1.0267	1.5860	3.3086
2	Sub-D	3.1543	3.1371	3.3679	3.4201	3.5698	4.3222	7.2249
	Sub-DI	2.8608	2.8984	3.2357	3.3407	3.5079	4.3117	6.1946
	REML	0.4728	0.4821	0.4926	0.5029	0.5059	0.5181	0.5269
	ANOVA	0.7463	0.7937	0.9030	1.0189	1.1494	1.7062	3.5398
5	Sub-D	6.9049	7.1216	7.1263	7.2161	7.3574	7.8234	10.051
	Sub-DI	5.8670	6.3473	6.3308	6.5306	6.8999	7.6876	9.9917
	REML	0.4761	0.4928	0.4985	0.5054	0.5282	0.5276	0.5354
	ANOVA	1.1629	1.2871	1.3445	1.4426	1.6024	2.1386	3.7606

Table 5.15: Estimation on unbalanced "two-way nested design": estimate for γ_1.

$\gamma_1 \setminus \gamma_2$		0.1	0.25	0.5	0.75	1	2	5
0.1	Sub-D	0.1005	0.1050	0.0994	0.1025	0.0865	0.1032	0.1128
	Sub-DI	0.1002	0.1044	0.0994	0.1022	0.0861	0.1054	0.1112
	REML	0.1878	0.2307	0.2766	0.3363	0.3655	0.5556	0.8387
	ANOVA	0.0955	0.0823	0.0512	0.0427	-0.001	-0.064	-0.311
0.25	Sub-D	0.2440	0.2465	0.2527	0.2464	0.2355	0.2642	0.2287
	Sub-DI	0.2439	0.2468	0.2535	0.2458	0.2364	0.2652	0.2314
	REML	0.3022	0.3441	0.4087	0.4545	0.4928	0.6744	0.9984
	ANOVA	0.2380	0.2224	0.2142	0.1754	0.1499	0.0975	-0.193
0.5	Sub-D	0.5041	0.4916	0.4961	0.4965	0.5024	0.4861	0.4419
	Sub-DI	0.5049	0.4924	0.4953	0.4957	0.5033	0.4861	0.4417
	REML	0.5309	0.5474	0.6022	0.6565	0.7088	0.8705	1.2005
	ANOVA	0.5010	0.4741	0.4574	0.4303	0.4127	0.3081	-0.005
0.75	Sub-D	0.7471	0.7419	0.7532	0.7403	0.7357	0.7411	0.7291
	Sub-DI	0.7470	0.7428	0.7543	0.7411	0.7356	0.7410	0.7336
	REML	0.7380	0.7769	0.8407	0.8720	0.9159	1.0833	1.4658
	ANOVA	0.7455	0.7269	0.7143	0.6787	0.6389	0.5777	0.3002
1	Sub-D	1.0014	1.0022	1.0040	0.9797	1.0106	0.9931	1.0169
	Sub-DI	1.0019	1.0024	1.0038	0.9791	1.0116	0.9908	1.0099
	REML	0.9833	1.0145	1.0595	1.0856	1.1626	1.3046	1.6756
	ANOVA	0.9933	0.9790	0.9590	0.9267	0.9301	0.8230	0.5597
2	Sub-D	1.9613	2.0174	1.9894	1.9769	2.0024	1.9596	2.0267
	Sub-DI	1.9614	2.0176	1.9898	1.9775	2.0009	1.9595	2.0305
	REML	1.8956	1.9745	1.9749	2.0145	2.0465	2.1666	2.6138
	ANOVA	1.9580	1.9990	1.9537	1.9204	1.9087	1.7868	1.6325
5	Sub-D	4.9632	4.9570	5.0111	4.9951	4.9934	5.0276	5.0872
	Sub-DI	4.9647	4.9605	5.0112	4.9953	4.9972	5.0244	5.0877
	REML	4.8168	4.8326	4.8870	4.9012	4.9277	4.9684	5.3729
	ANOVA	4.9539	4.9463	4.9743	4.9586	4.9002	4.8368	4.6882

Table 5.16: Estimation on unbalanced "two-way nested design": mean square error of γ_1.

$\gamma_1 \setminus \gamma_2$		0.1	0.25	0.5	0.75	1	2	5
0.1	Sub-D	0.4823	0.6032	0.7565	0.9087	1.0671	1.7920	3.8020
	Sub-DI	0.4732	0.5874	0.7312	0.8767	1.0255	1.7112	3.6133
	REML	0.3784	0.4774	0.5694	0.7079	0.7979	1.2920	2.3357
	ANOVA	0.5474	0.6965	0.8976	1.1088	1.3164	2.2438	4.8475
0.25	Sub-D	0.6414	0.7373	0.8689	1.0575	1.2085	1.9021	3.8404
	Sub-DI	0.6318	0.7213	0.8451	1.0220	1.1658	1.8201	3.6498
	REML	0.5469	0.5980	0.7259	0.8357	0.9493	1.4138	2.4919
	ANOVA	0.6821	0.7998	0.9944	1.2292	1.4417	2.3408	4.8712
0.5	Sub-D	0.8836	0.9841	1.1247	1.2802	1.4522	2.0756	4.0664
	Sub-DI	0.8726	0.9668	1.0975	1.2437	1.4039	1.9944	3.8780
	REML	0.7898	0.8508	0.9616	1.0747	1.1706	1.6070	2.7156
	ANOVA	0.8871	1.0061	1.2032	1.4175	1.6217	2.4858	5.1019
0.75	Sub-D	1.1585	1.2509	1.3920	1.5241	1.6893	2.3507	4.2620
	Sub-DI	1.1443	1.2300	1.3629	1.4829	1.6385	2.2604	4.0597
	REML	1.0213	1.1172	1.2452	1.3262	1.4209	1.8737	2.8391
	ANOVA	1.1167	1.2337	1.4401	1.6081	1.8307	2.7058	5.2088
1	Sub-D	1.4420	1.5296	1.6994	1.7580	1.9792	2.5305	4.5685
	Sub-DI	1.4269	1.5052	1.6664	1.7167	1.9240	2.4373	4.3645
	REML	1.3255	1.3706	1.5002	1.5519	1.7390	2.0728	3.1299
	ANOVA	1.3842	1.4725	1.6990	1.8102	2.0701	2.8370	5.4976
2	Sub-D	2.5265	2.7061	2.7276	2.8952	3.0893	3.5652	5.4697
	Sub-DI	2.4966	2.6699	2.6825	2.8426	3.0259	3.4619	5.2523
	REML	2.2974	2.4571	2.4979	2.6624	2.7361	3.06338	4.2182
	ANOVA	2.3289	2.5237	2.5932	2.8225	3.0326	3.7374	6.2409
5	Sub-D	5.9223	5.8785	6.1256	6.1631	6.2676	6.8900	8.5391
	Sub-DI	5.8617	5.8185	6.0471	6.0748	6.1719	6.7518	8.2930
	REML	5.2967	5.3841	5.5423	5.6354	5.6455	6.0925	7.0896
	ANOVA	5.3516	5.4092	5.6219	5.7504	5.8311	6.5798	8.9226

Table 5.17: Estimation on unbalanced "two-way nested design": estimates for γ_2.

$\gamma_1 \setminus \gamma_2$		0.1	0.25	0.5	0.75	1	2	5
0.1	Sub-D	0.1033	0.2434	0.5013	0.7667	1.0028	2.0369	4.9549
	Sub-DI	0.1049	0.2477	0.5015	0.7688	1.0051	2.0216	4.9668
	REML	0.1744	0.2841	0.4852	0.6858	0.8812	1.7468	4.4893
	ANOVA	0.0926	0.2494	0.5169	0.7568	1.0020	2.0115	4.9748
0.25	Sub-D	0.1103	0.2482	0.5032	0.7483	1.0059	2.0040	5.0332
	Sub-DI	0.1110	0.2456	0.4978	0.7523	0.9993	1.9968	5.0141
	REML	0.2108	0.3067	0.4926	0.7055	0.8970	1.7792	4.5139
	ANOVA	0.1074	0.2534	0.4943	0.7642	1.0002	1.9995	5.0265
0.5	Sub-D	0.1131	0.2551	0.4845	0.7581	0.9961	2.0121	5.0047
	Sub-DI	0.1074	0.2498	0.4899	0.7635	0.9894	2.0121	5.0060
	REML	0.2218	0.3348	0.5254	0.7446	0.9376	1.8137	4.5676
	ANOVA	0.0958	0.2441	0.48175	0.7608	0.9998	2.0276	5.0654
0.75	Sub-D	0.1044	0.2697	0.5024	0.7501	0.9897	2.0073	4.9985
	Sub-DI	0.1050	0.2635	0.4952	0.7446	0.9901	2.0078	4.9666
	REML	0.2497	0.3609	0.5462	0.7547	0.9694	1.8297	4.5394
	ANOVA	0.0903	0.2530	0.4956	0.7431	1.0164	1.9975	5.0074
1	Sub-D	0.1112	0.2459	0.5147	0.7478	0.9962	2.0080	4.9622
	Sub-DI	0.1079	0.2447	0.5166	0.7520	0.9892	2.0242	5.0120
	REML	0.2638	0.3724	0.5774	0.7672	0.9799	1.8819	4.6375
	ANOVA	0.1099	0.2504	0.5110	0.7280	0.9867	2.0164	5.0636
2	Sub-D	0.0982	0.2663	0.5000	0.7608	0.9843	2.0343	4.9964
	Sub-DI	0.0973	0.2643	0.4975	0.7560	0.9952	2.0348	4.9695
	REML	0.2867	0.4228	0.6346	0.8400	1.0740	1.9810	4.6380
	ANOVA	0.0875	0.2603	0.4896	0.7446	1.0077	2.0368	4.9383
5	Sub-D	0.0995	0.2831	0.5184	0.7700	0.9763	1.9345	5.0321
	Sub-DI	0.0883	0.2586	0.5175	0.7682	0.9497	1.9576	5.0287
	REML	0.3433	0.4941	0.7109	0.9229	1.1655	2.1271	4.9520
	ANOVA	0.0977	0.2584	0.5014	0.7160	0.9969	1.9866	4.9750

Table 5.18: Estimation on unbalanced "two-way nested design": mean square error of γ_2.

$\gamma_1 \setminus \gamma_2$		0.1	0.25	0.5	0.75	1	2	5
0.1	Sub-D	0.6236	0.8331	1.1512	1.4745	1.8100	3.1769	7.2599
	Sub-DI	0.5196	0.6807	0.9466	1.2086	1.4882	2.5949	5.9306
	REML	0.3528	0.4768	0.7113	0.9196	1.1388	1.9727	4.5617
	ANOVA	0.6054	0.7635	1.0393	1.2931	1.5764	2.7044	6.0615
0.25	Sub-D	0.7007	0.8978	1.1924	1.5133	1.8494	3.1996	7.1163
	Sub-DI	0.5990	0.741	0.9837	1.2330	1.5144	2.5994	5.7695
	REML	0.4161	0.5328	0.7241	0.9642	1.1644	2.0598	4.6022
	ANOVA	0.6167	0.7668	1.0017	1.2860	1.5576	2.6786	6.0029
0.5	Sub-D	0.8782	1.0389	1.3402	1.6433	2.0217	3.2509	7.2862
	Sub-DI	0.7422	0.8656	1.0983	1.3409	1.6405	2.6526	5.9341
	REML	0.4470	0.5804	0.7882	1.0147	1.2180	2.0655	4.6149
	ANOVA	0.6114	0.7499	1.0136	1.2850	1.5805	2.6401	6.0117
0.75	Sub-D	1.0658	1.2341	1.5108	1.7953	2.1108	3.4252	7.2936
	Sub-DI	0.8984	1.0425	1.2595	1.4571	1.7122	2.7794	5.8989
	REML	0.5248	0.6354	0.8280	1.0636	1.2605	2.1567	4.6937
	ANOVA	0.6055	0.7572	0.9951	1.2793	1.5701	2.6446	5.9878
1	Sub-D	1.2554	1.4179	1.6895	1.9329	2.2688	3.5614	7.5266
	Sub-DI	1.0666	1.1751	1.4013	1.5766	1.8329	2.8857	6.1418
	REML	0.5368	0.6513	0.8979	1.0801	1.3315	2.2063	4.8250
	ANOVA	0.6101	0.7594	1.0497	1.2529	1.5476	2.6593	6.1121
2	Sub-D	2.0739	2.2694	2.4371	2.6854	2.9262	4.1564	7.8640
	Sub-DI	1.7178	1.8725	2.0036	2.1914	2.3979	3.3339	6.3360
	REML	0.6266	0.8035	1.0649	1.2402	1.4892	2.4098	4.9291
	ANOVA	0.6082	0.7518	1.0119	1.2731	1.5731	2.7114	5.8598
5	Sub-D	4.5957	4.7497	4.9588	5.1901	5.3053	6.2507	9.8571
	Sub-DI	3.8065	3.9031	4.0163	4.2035	4.2878	4.9624	7.9472
	REML	0.9769	1.1772	1.3347	1.6247	1.8047	2.8375	5.6901
	ANOVA	0.6181	0.7661	1.0261	1.2456	1.5493	2.6080	6.0478

Table 5.19: Estimation on unbalanced "two-way nested design": estimates for γ_3.

$\gamma_1 \setminus \gamma_2$		0.1	0.25	0.5	0.75	1	2	5
0.1	Sub-D	0.9815	1.0177	1.0219	0.9874	1.0086	0.9553	1.0438
	Sub-DI	0.9791	1.0111	1.0216	0.9843	1.0051	0.9789	1.0254
	REML	0.8826	0.9095	0.9195	0.9207	0.9391	0.9328	0.9527
	ANOVA	0.9959	1.0095	1.0011	0.9978	1.0119	0.9937	1.0090
0.25	Sub-D	1.0039	1.0021	0.9884	1.0163	0.9953	1.0021	1.0009
	Sub-DI	1.0028	1.0061	0.9969	1.0100	1.0055	1.0131	1.0303
	REML	0.8993	0.9047	0.9197	0.9276	0.9410	0.9496	0.9539
	ANOVA	1.0041	0.9969	0.9988	0.9974	1.0060	1.0046	1.0035
0.5	Sub-D	0.9774	0.9996	1.0029	1.0024	1.0026	1.0110	1.0859
	Sub-DI	0.9862	1.0079	0.9946	0.9940	1.0129	1.0109	1.0839
	REML	0.9021	0.9255	0.9284	0.9327	0.9429	0.9455	0.9634
	ANOVA	0.9999	1.0139	1.0041	0.9999	1.0033	0.9923	1.0035
0.75	Sub-D	0.9982	0.9856	0.9924	0.9819	1.0218	1.0028	1.0037
	Sub-DI	0.9972	0.9952	1.0035	0.9905	1.0213	1.0020	1.0529
	REML	0.9157	0.9230	0.9322	0.9304	0.9380	0.9665	0.9559
	ANOVA	1.0134	1.0049	0.9995	0.9916	0.9907	1.0099	0.9923
1	Sub-D	0.9978	1.0037	0.9807	0.9813	0.9895	1.0231	1.1419
	Sub-DI	1.0030	1.0055	0.9777	0.9749	1.0002	0.9981	1.0650
	REML	0.9087	0.9184	0.9224	0.9418	0.9488	0.9653	0.9712
	ANOVA	0.9983	0.9986	0.9897	1.0024	1.0008	1.0060	1.0039
2	Sub-D	0.9974	0.9951	0.9888	0.9738	1.0314	1.0010	0.9424
	Sub-DI	0.9987	0.9983	0.9927	0.9813	1.0145	1.0004	0.9840
	REML	0.9193	0.9269	0.9339	0.9383	0.9562	0.9662	0.9818
	ANOVA	1.0097	0.9997	0.9979	0.9915	1.0026	0.9957	1.0044
5	Sub-D	0.9958	0.9704	0.9861	0.9414	1.0225	1.0565	0.9437
	Sub-DI	1.0131	1.0082	0.9876	0.9442	1.0637	1.0208	0.9489
	REML	0.9157	0.9256	0.9479	0.9501	0.9574	0.9706	0.9934
	ANOVA	1.0003	0.9964	1.0068	0.9986	1.0055	0.9985	1.0062

Table 5.20: Estimation on unbalanced "two-way nested design": mean square error of γ_3.

$\gamma_1 \setminus \gamma_2$		0.1	0.25	0.5	0.75	1	2	5
0.1	Sub-D	0.8656	1.0355	1.2806	1.4894	1.7167	2.7652	6.1466
	Sub-DI	0.7868	0.8909	1.0543	1.1441	1.2998	1.9424	4.1094
	REML	0.4445	0.4620	0.4779	0.4741	0.4806	0.4825	0.5332
	ANOVA	0.5277	0.5402	0.5420	0.5336	0.5365	0.5273	0.4935
0.25	Sub-D	0.9667	1.0859	1.3251	1.5956	1.8118	2.9562	6.1851
	Sub-DI	0.8847	0.9527	1.0892	1.2460	1.3850	2.0989	4.1439
	REML	0.4582	0.4614	0.4723	0.4776	0.4895	0.4993	0.5059
	ANOVA	0.5359	0.5319	0.5328	0.5276	0.5376	0.5386	0.5391
0.5	Sub-D	1.1154	1.2573	1.4866	1.7132	1.9917	3.0555	6.3437
	Sub-DI	1.0054	1.0816	1.2242	1.3387	1.5103	2.2089	4.3396
	REML	0.4629	0.4683	0.4752	0.4856	0.4910	0.5020	0.5072
	ANOVA	0.5356	0.5361	0.5341	0.5352	0.5340	0.5322	0.5327
0.75	Sub-D	1.3044	1.4681	1.6800	1.9058	2.1322	3.2666	6.423
	Sub-DI	1.1359	1.2640	1.3995	1.4955	1.6309	2.3706	4.3260
	REML	0.4679	0.4716	0.4819	0.4805	0.4855	0.5097	0.5101
	ANOVA	0.5421	0.5328	0.5373	0.5260	0.5237	0.5418	0.5336
1	Sub-D	1.5138	1.6392	1.8546	2.0969	2.3808	3.4153	6.6557
	Sub-DI	1.3212	1.3926	1.5215	1.6517	1.8257	2.4512	4.5444
	REML	0.4704	0.4692	0.4805	0.4953	0.4971	0.5053	0.5154
	ANOVA	0.5347	0.5317	0.5275	0.5423	0.5376	0.5336	0.5327
2	Sub-D	2.3981	2.5889	2.7542	2.9721	3.1300	4.1260	7.3119
	Sub-DI	1.9709	2.1133	2.2134	2.3410	2.4474	3.0435	5.0833
	REML	0.4698	0.4761	0.4882	0.4817	0.4967	0.5117	0.5328
	ANOVA	0.5332	0.5337	0.5360	0.5201	0.5323	0.5283	0.5385
5	Sub-D	5.1947	5.3592	5.5787	5.8248	5.8859	6.7661	9.7886
	Sub-DI	4.1480	4.2305	4.3155	4.5098	4.5694	5.0368	7.1337
	REML	0.4759	0.4796	0.4937	0.5029	0.4996	0.557	0.5420
	ANOVA	0.5356	0.5320	0.5394	0.5392	0.5309	0.5326	0.5313

6

FINAL COMENTS

As pointed out before, REML is the preferred method for estimating variance components in MLM (Diffey et al. [17]). In addition to its simple and fast computational implementation, once it depends only on the information retained by the eigenvalues of the design matrices and the quadratic errors of the sub-models, *Sub-D* provides unbiased estimates whether the data is *balanced* or *unbalanced* and in both crossed and nested designs, even having empty cells. This is not the case of ANOVA and REML estimators as seen through the tests carried out at the previous chapter.

As seen at Chapter 5, *Sub-D* estimator provides a slightly more accurate estimates (due to its unbiasedness) than REML estimator in all the designs approached, having, in some case, a little more dispersion (mostly in unbalanced models, but steel comparable; when the model is balanced they have a little bit more comparable dispersion). This problem is attenuated with the introduction of the *Sub-DI* estimator, which also produces unbiased estimates but with less dispersion than *Sub-D*.

REML estimator does not look to have a good performance in nested designs (see section 5.4) with low accurate estimates unlike *Sub-D* and *Sub-DI* which, as previously seen, provides accurate estimates. The ANOVA estimator provides low accurat estimates in all the crossed and nested designs as seen in previous chapter, it just seem to provide accurate estimates in "one-way designs"; this is rightful since ANOVA uses fixed effect techniques. The reader must be aware that, despite the samples considered are reasonably small, both the crossed and nested designs considered in the numerical simulation have some empty cells, so that the estimators were taking to the extreme.

As a complement, we may remark that *Sub-D* and *Sub-DI* keep a somewhat constant performance for all the models in which it was applied, providing always accurate estimates whereas REML does not show a constant performance (for example, in a particular unbalanced "two - way nested design" (see Section 5.4) REML provided non centered estimates). It also seems it have better performance for variance components with values bigger than 1. For the ANOVA estimator, the scenario is even worse, since it does provide non centered estimates in both nested and crossed

design (see the both nested and crossed design approached in the previous chapter).

Since in any computational program (source code) when we are interested in share the code, create package or use it repeatedly, we might consider its efficiency and, for this matter, the code run-time constitutes a good start point. Doing so, to compute the estimates and the corresponding standard deviation in each two-way designs considered here, with γ_1, γ_2 and γ_3 taking values in $\{0.25, 0.5, 0.75, 1, 2, 5\}$, for 10000 observations we found that the ANOVA, *Sub-D* and *Sub-DI* run-times are about 1.2471, 2.06 and 4.3338 seconds respectively, while the REML estimator run-time is about 6.2618 minutes, which means that the code for ANOVA and *Sub-D* are more than 187 times faster than the one for REML.

The process of sub-diagonalizing the variance-covariance matrix in different orders will be considered in future works; more over, the following topics will also be considered in future works:

- Improving the variability of the estimated values obtained with *Sub-D* and *Sub-DI*;

- Confidence region and tests of hypothesis for the variance components.

APPENDIX

A.1 Useful Results

A.1.1 Algebraic Results

Theorem A.1.1. *Let $A \in \mathcal{M}^{n \times n}$ and let λ be its eigenvalue with correspondent eigenvector v. Then*

(a) *λ^m is an eigenvalue of A^m corresponding to the eigenvector v, where m is an integer that $m \geq 1$.*

(b) *λ^{-1} is an eigenvalue of A^{-1} corresponding to the eigenvector v, providing A is non-singular.*

Proof. See Theorem 3.4 of Schott [64]. □

Theorem A.1.2. *Let $A \in \mathcal{S}^n$ and let $v \in \mathbb{R}^n$ be any nonzero vector. Then the vector space spanned (generated) by the vectors v, Av, ..., $A^{r-1}v$ contains an eigenvector of A, for some $r \geq 1$.*

Proof. See Theorem 3.9 of Schott [64]. □

Theorem A.1.3. *Let $H \in \mathcal{M}^{n \times n}$ be a non-singular matrix with derivative $\partial H / \partial x$. Then*

$$\partial \frac{H^{-1}}{\partial x} = H^{-1} \partial \frac{H}{\partial x} H^{-1}.$$

Proof. Since A is non-singular, we have $H^{-1}H = \mathbf{0}_{n \times n}$. Thus, $\frac{\partial H^{-1}}{\partial x}H + H^{-1}\frac{\partial H}{\partial x} = \mathbf{0}_{n \times n}$, so that $\frac{\partial H^{-1}}{\partial x}H = -H^{-1}\frac{\partial H}{\partial x}$, leading to $\frac{\partial H^{-1}}{\partial x} = -H^{-1}\frac{\partial H}{\partial x}H^{-1}$. □

Theorem A.1.4. *Let $H \in \mathcal{M}^{n \times n}$ be a positive definite matrix. Then*

$$\frac{\partial log |H|}{\partial x} = tr\left(H^{-1}\frac{\partial H}{\partial x} \right).$$

Proof. □

Theorem A.1.5. *Let $A \in \mathcal{M}^{n \times p}$. Then*

(a) $A^\top A$ *is positive definite, if* $r(A) = p$.

(b) $A^\top A$ *is positive semi-definite, if* $r(A) = n$.

Proof. □

Theorem A.1.6. *A positive definite matrix is non-singular.*

Proof. See the Theorem 2.6c and its corollary in Rencher and Schaalje [62]. □

A.1.2 Statistical Results

Theorem A.1.7. *Let $y \in \mathbb{R}^n$ and $z \in \mathbb{R}^m$ be random vectors with second moment such that $E(y) = \mathbf{0}_n$ and $E(z) = \mathbf{0}_m$. Then,*

$$\Sigma(y,z) = E(yz^\top).$$

Proof.

$$\begin{aligned}
\Sigma(y,z) &= E\left([y - E(y)][z - E(z)]^\top\right) \\
&= E(yz^\top) - yE(z)^\top - E(y)z^\top + E(y)E(z)^\top \\
&= E(yz^\top)
\end{aligned} \tag{A.1}$$

□

Theorem A.1.8. *Let $y \sim \mathcal{N}(\mu, \Sigma)$, where $\mu \in \mathbb{R}^n$, and consider $A \in \mathcal{M}^{m \times n}$ a matrix of constants and $B \in \mathscr{S}^n$. Then*

(a) $z = Ay \sim \mathcal{N}(A\mu, A\Sigma A^\top)$;

(b) $E(yBy) = tr(B\Sigma) + \mu^\top B\mu$.

(c) $z = Ay \sim (A\mu, A\Sigma A^\top)$, *provided* $y \sim (A\mu, \Sigma)$.

Proof. See Theorem 2.1.2 and Theorem 1.3.1 of Moser [48] for (a) and (c), respectively, and Theorem 5.2a and 3.6d of Rencher and Schaalje [62] for (b). Typically it is assumed $m \leq n$. □

Theorem A.1.9. *Let $y \sim \mathcal{N}(X\beta, \Sigma)$, with $\Sigma = \sum_{i=1}^{r} \gamma_i Z_i Z_i^\top + \gamma_{r+1} I$, where $X \in \mathcal{M}^{n \times p}$ is of rank $r \leq p$, and $\Sigma \in \mathcal{M}^{n \times n}$ is a positive definite matrix.*

Then a full-rank matrix K with maximal number of rows such that $KX = 0$, is in $\mathcal{M}^{(n-r) \times n}$. Furthermore, K must be of the form

$$K = C(I - H) = C\left(I - X(X^\top X)^\top\right),$$

where C specifies a full-rank transformation of the rows of the matrix $I - H$.

Proof. See Theorem 17.4a. of Rencher and Schaalje [62]. □

Theorem A.1.10. *Let $y \sim (X\beta, \Sigma)$, where $\Sigma = \sum_{i=1}^{r} \gamma_i Z_i Z_i^\top + \gamma_{r+1} I$, and consider the set $\mathscr{B} = \{Z_1, \ldots, Zr\}$ of known matrices $Z_1, \ldots, Zr$. Then, if $\mathscr{B}$ is balanced and nested, y has OBS.*

Proof. See Proposition 3.3. of VanLeeuwen at all [72]. □

BIBLIOGRAPHY

[1] A. Albert. "On Jordan algebras of linear transformations." In: *Transactions of the American Mathematical Society.* 59.3 (1946), pp. 524–555.

[2] R. Anderson. "Use of variance component analysis in the interpretation of biological experiments." In: *Bulletin of the International Statistical Institute.* 37 (1960), pp. 1–22.

[3] R. Anderson. "Recent developments in designs and estimators for variance components." In: *Statisitcs and Related Topics, North-Holland, Amsterdam.* (1981), pp. 3–22.

[4] R. L. Anderson. "Designs and estimators for variance components." In: *Statisitcal Design and linear Model, North-Holland, Amsterdam.* (1975), pp. 1–30.

[5] R. L. Anderson and T. A. Bancroft. *Statistical Theory in Research.* McGraw-Hill Book Company, Inc.; New York., 1952.

[6] R. L. Anderson and P. P. Crump. "Comparisons of designs and estimation procedures for estimating parameters in a two-stages nested process." In: *Technometrics.* 9.4 (1967), pp. 499–516.

[7] D. F. Andrade and R. W. Helms. "ML estimation and LR tests for the multivariate normal distribution with general linear model mean and linearstructure covariance matrix". In: *Communications in Statistics-Theory and Methods.* 15.1 (1986), pp. 89–107.

[8] T. R. Bainbridge. "Staggered Nested Designs for Estimating Variance Components." In: *Industrial Quality Control.* 22.1 (1965), pp. 12–20.

[9] R. W. Barnard and E. H. Moore. "General Analysis. Part 1." In: *Memoirs Of The American Philosophical Society* (1935).

[10] D. Bates, M. Maechler, B. Bolker, S. Walker, R. H. B. Christensen, H. Singmam, B. Dai, and G. Grothendieck. "Package lme4". In: *R Foundation for Statistical Computing, Vienna* 12 (2014).

[11] N. Bush and R. Anderson. "A comparison of three different procedures for estimating variance components." In: *Technometrics.* 5.4 (1963), pp. 421–440.

[12] T. Calinski and S. Kageyama. "Block designs: A Randomization Approach. Vol. I: Analysis." In: *Lecture notes in statistics.* 150 (2000).

[13] T. Calinski and S. Kageyama. "Block Designs: A Randomization Approach. Vol. II: Design". In: *Lecture Notes in Statistics, Springer.* 170 (2003).

[14]   G. Casella and R. Berger. *Statistical Inference.* Vol. 2. Duxbury Pacific Grove, CA, 2002.

[15]   P. P. Crump. "Optimal designs to estimate the parameters of a variance component model". PhD thesis. North Carolina State University, 1954.

[16]   E. Demidenko. *Mixed Models: Theory and Applications (Wiley Series in Probability and Statistics).* Wiley-Interscience, 2004.

[17]   S. Diffey, A. Welsh, A. Smith, and B. R. Cullis. "A faster and computationally more efficient REML (PX) EM algorithm for linear mixed models". In: (2013).

[18]   D. L. Fairclough and R. H Helms. "A mixed linear model with linear covariance structure: a sensitivity analysis of maximum likelihood estimators". In: *Journal of Statistical Computation and Simulation* 25.3-4 (1986), pp. 205–236.

[19]   M. Fonseca, J. T. Mexia, and R. Zmyslony. "Inference in normal models with commutative orthogonal block structure". In: *Act Et Commentationes Universitates tartuensis de Mathematica* 12 (2008).

[20]   D. Gaylor. "The construction and evaluation of some designs for the estimation of parameters in random models". PhD thesis. North Carolina State University, 1960.

[21]   A. R. Gilmour, R. Thompson, and B. R. Cullis. "Average information REML: an efficient algorithm for variance parameter estimation in linear mixed models". In: *Biometrics* (1995), pp. 1440–1450.

[22]   C. Goldsmith and D. Gaylor. "Three stage nested designs for estimating variance components". In: *Technometrics* 12.3 (1970), pp. 487–498.

[23]   A. Haile and J. Webster. "A comparison of four designs for estimating the variances of random effects in a two-way classification." In: *Technometrics* 17.4 (1975), pp. 495–500.

[24]   J. Hammersley. "The unbiased estimate and standard error of the interclass variance." In: *Metron* 15 (1949), pp. 189–205.

[25]   H. O. Hartley and J. K. Rao. "Maximum likelihood estimation for the mixed analysis of variance model". In: *Biometrika* 54.1-2 (1967), pp. 93–108.

[26]   D. A. Harville. "Maximum likelihood approaches to variance component estimation and to related problems." In: *Journal of the American Statistical Association* 72.358 (1977), pp. 320–338.

[27]   R. D. Helms and R. W. Wallace. "An approximate F statistic for testing population effects in longitudinal studies via mixed models". PhD thesis. University of North Carolina at Chapel Hill, 1993.

[28]   C. Henderson. "Estimation of variance and covariance components". In: *Biometrics* 9.2 (1953), pp. 226–252.

[29]   R. Hocking. *The analysis of linear models*. Brooks/Cole Pub Co., 1985.

[30]   R. Horn and C. Johnson. *Matrix analysis. Corrected reprint of the 1985 original.* Cambridge University Press, Cambridge., 1985.

[31] A. Houtman and T. Speed. "Balance in designed experiments with orthogonal block structure". In: *The Annals of Statistics*. 1069-1085 (1983).

[32] P. Jordan. "Ueber Verallgemeinerungsmglichkeiten des Formalismus der Quantenmechanik". In: *Nachr. Akad. Wiss. Gttingen. Math. Phys* (1933).

[33] P Jordan, V. Neumann, and E Wigner. "On an Algebraic Generalization of the Quantum Mechanical Formalism". In: *The Collected Works of Eugene Paul Wigner*. Springer, 1993, pp. 298–333.

[34] A. Khuri. "Design for Variance Components Estimation: Past and Present". In: *International Statistical Review* 68.3 (2000), pp. 311–322.

[35] A. Khuri and H. Sahai. "Variance components analysis: a selective literature survey". In: *International Statistical Review* (1985), pp. 279–300.

[36] J. Klotz, R. Milton, and S. Zacks. "Mean square efficiency of estimators of variance component." In: *Journal of the American Statistical Association* 64.328 (1969), pp. 1383–1402.

[37] N. M. Laird and J. Ware. "Random-effects models for longitudinal data". In: *Biometrics* (1982), pp. 963–974.

[38] D. Lay. *Linear Algebra and Its Applications*. Addison-Wesley, 1994.

[39] A. Lewandowski, C. Liu, and S. Wiel. "Parameter Expansion and Efficient Inference". In: *Statistical Science* 25.4 (2010), pp. 533 –544.

[40] M. Lindstrom and D. Bates. "Newton-Raphson and EM Algorithms for Linear Mixed-Effects Models for Repeated-Measures Data". In: *Journal of the American Statistical Association* 83.404 (1988), pp. 1014 –1022.

[41] C Liu, D. Rubin, and Y. Wu. "Parameter Expansion to Accelerate EM: The PX-EM Algorithm". In: *Biometrika* 85.755 - 770 (1998).

[42] J. R. Magnus and H. Neudecker. *Matrix Differential Calculus With Applications in Statistics and Econometrics*. John Wiley & Sons, 1995.

[43] J. D. Malley. *Statistical Applications Of Jordan Algebras*. Vol. 91. Springer, 1994.

[44] C. E. McCulloch and S. R. Searle. *Generalized, Linear, and Mixed Models*. New York: Wiley., 2001.

[45] S. Mejza. "On some aspects of general balance in designed experiments". In: *Statistica* 52.2 (1992), pp. 263–278.

[46] J. J. Miller. *Asymptotic properties and computation of maximum likelihood estimates in the mixed model of the analysis of variance*. Tech. rep. DTIC Document, 1973.

[47] J. J. Miller. "Asymptotic properties of maximum likelihood estimates in the mixed model of the analysis of variance". In: *The Annals of Statistics* (1977), pp. 746–762.

[48] B. Moser. *Linear Models: A Mean Model Approach*. Elsevier, 1996.

[49]   M. G. Mostafa. "Designs for the simultaneous estimation of functions of variance components from two-way crossed classifications". In: *Biometrika* 54.1-2 (1967), pp. 127–131.

[50]   R. Mukerjee. "Optimality of balanced designs for minimum norm quadratic unbiased estimation of variance components". In: *Metrika* 37.1 (1990), pp. 181–188.

[51]   R. Mukerjee and S. Huda. "Optimal design for the estimation of variance components". In: *Biometrika* 75.1 (1988), pp. 75–80.

[52]   K. Muller and P. Stewart. *Linear Model Theory : Univariate, Multivariate, and Mixed Models*. John Wiley & Sons, 2006.

[53]   K. E. Muller and V. B. Pasour. "Bias in linear model power and sample size due to estimating variance". In: *Communications in Statistics - Theory and Methods* 26.4 (1997), pp. 839–851.

[54]   H. Muse, R. Anderson, and B. Thitakamol. "Additional comparisons of designs to estimate variance components in a two-way classification model". In: *Communications in Statistics, Theory and Methods* (1982).

[55]   H. D. Muse. "Comparison of designs to estimate variance components in a two-way classification model". PhD thesis. University of Kentucky, 1974.

[56]   H. Muse and R. Anderson. "Comparison of designs to estimate variance components in a two-way classification model". In: *Technometrics* 20.2 (1978), pp. 159–166.

[57]   J. Nelder. "The analysis of randomized experiments with orthogonal block structure. I. Block structure and the null analysis of variance". In: *Mathematical and Physical Sciences* (1964).

[58]   J. A. Nelder. "The analysis of randomized experiments with orthogonal block structure. II. Treatment structure and general analysis of Variance". In: *Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*. Vol. 283. 1393. The Royal Society. 1965, pp. 147–162.

[59]   H. D. Patterson and R. Thompson. "Recovery of inter-block information when block sizes are unequal." In: *Biometrika* 58.3 (1971), pp. 545–554.

[60]   R. Penrose. "A Generalized Inverse For Matrices." In: vol. 51. 3. Cambridge Univ Press. 1955, pp. 406–413.

[61]   C. R. Rao and M. B. Rao. *Matrix Algebra And Its Applications To Statistics And Econometrics*. World Scientific, 1998.

[62]   A. C. Rencher and G. B. Schaalje. *Linear Models in Statisitcs*. John Wiley & Sons., 2008.

[63]   H. Sahai and M. M. Ojeda. *Analysis of Variance for Random Models, Volume 2: Unbalanced Data*. Vol. 2. Springer Science & Business Media, 2004.

[64]   J. R. Schott. *Matrix Analysis for Statistics*. John Wiley & Sons, 1997.

[65]   S. Searle. "A Biometrics Invited Paper. Topics in variance component estimation". In: *Biometrics* 27.1 (1971), pp. 1–76.

[66]   S. Searle. "An overview of variance component estimation". In: *Metrika* 42.1 (1995), pp. 215–230.

[67]   S. Searle, G. Casella, and C McCulloch. *Variance Components*. John Wiley & Sons, 1992.

[68]   J. Seely. "Quadratic Subspaces and Completeness". In: *The Annals of Mathematic Statistics* 42 (1971), pp. 710–721.

[69]   A. Silva, F. Carvalho, J. Mexia, and M. Fonseca. "One-sided tolerance interval in a two-way balanced nested model with mixed effects". In: *11th International Conference of Numerical Analysis and Applied Mathematics 2013: ICNAAM 2013*. Vol. 1558. 1. AIP Publishing. 2013, pp. 818–820.

[70]   A. Silva, J. Mexia, and M. Fonseca. "Contribute on Mixed Linear Model - Simultaneous Diagonalization of the Covariance Matrices". In: *Proceedings of Mathematics & Statistics*. 2015.

[71]   W. Thompson and R. Anderson. "A comparison of designs and estiqtors for the two-stage nested random model". In: *Technometrics.* 17.1 (1975), pp. 37–44.

[72]   D. M. VanLeeuwen, J. F. Seely, and D. Birkes. "Sufficient Conditions for Orthogonal Designs in Mixed Linear Models". In: *Journal of Statitistical Planning and Inference* 73.1 (1998), pp. 373–389.

[73]   F. Yates. "The analysis of multiple classifications with unequal numbers in the different classes". In: *Journal of the American Statistical Association* 29.185 (1934), pp. 51–66.

2017

Adilson da Silva

Variance Components Estimation In Mixed Linear Models