Mathematics and Statistics 10(6): 1247-1263, 2022
DOI: 10.13189/ms.2022.100611 http://www.hrpub.org

Nonparametric *REML-like* Estimation in Linear Mixed Models with Uncorrelated Homoscedastic Errors

E.-P. Ndong Nguéma, Betrand Fesuh Nono*

Department of Mathematics, Ecole Polytechnique de Yaoundé, Cameroon

Received June 29, 2022; Revised September 15, 2022; Accepted October 11, 2022

Cite This Paper in the following Citation Styles

***(a)**: [1] E.-P. Ndong Nguéma, Betrand Fesuh Nono, "Nonparametric* REML-like *Estimation in Linear Mixed Models with Uncorrelated Homoscedastic Errors," Mathematics and Statistics, Vol.10, No.6, pp. 1247-1263, 2022. DOI: 10.13189/ms.2022.100611*

***(b)**: E.-P. Ndong Nguéma, Betrand Fesuh Nono (2022). Nonparametric* REML-like *Estimation in Linear Mixed Models with Uncorrelated Homoscedastic Errors. Mathematics and Statistics, 10(6), 1247-1263. DOI: 10.13189/ms.2022.100611*

Abstract Restricted Maximum Likelihood (REML) is the most recommended approach for fitting a Linear Mixed Model (LMM) nowadays. Yet, as ML, REML suffers the drawback that it performs such a fitting by assuming normality for both the random effects and the residual errors, a dubious assumption for many real data sets. Now, there have been several attempts at trying to justify the use of the REML likelihood equations outside of the Gaussian world, with varying degrees of success. Recently, a new fitting methodology, code named 3*S*, was presented for LMMs with only added assumption (to the basic ones) that the residual errors are uncorrelated and homoscedastic. Specifically, the 3*S*-A1 variant was designed and then shown, for Gaussian LMMs, to differ only slightly from ML estimation. In this article, using the same 3*S* framework, we develop another iterative nonparametric estimation methodology, code named 3*S*-A1.RE, for the kind of LMMs just mentioned. However, we show that if the LMM is, indeed, Gaussian with *i.i.d.* residual errors, then the set of estimating equations defining any 3*S*-A1.RE iterative procedure is equivalent to the set of REML equations, but while including the nonnegativity constraints on all variance estimates, as well as positive semi-definiteness on all covariance matrices. In numerical tests on some simulated and real world clustered and longitudinal data sets, our new methods proved to be highly competitive when compared to the traditional REML in the R statistical software.

Keywords Linear Mixed Model, 2-level Model, Uncorrelated Homoscedastic Errors, Random Effects Predictor, Restricted Maximum Likelihood

Mathematics Subject Classification (2000): 62G05, 62J05

1 Introduction

For long now, mixed models have been extremely useful for analyzing data sets with multilevel structure of various kinds (*e.g.* clustered or repeated measures) in a variety of fields, such as medicine [1, 2], agriculture [3], animal breeding [4], small area estimation [5, 6], genetics [7, 8], growth modeling [9, 10], *etc*. For an extensive presentation on LMMs, see [11, 12, 13, 14], with [14] laying more emphasis on the subclass of variance components models (or ANOVA LMMs). Yet, fitting an LMM to a given data set is still no easy task, although there are now, specifically available for that purpose, carefully designed programming routines in the leading statistical software systems. But a question remains: *Are the modeling assumptions upon which such codes are built guaranteed to hold for the data set at hand?* The answer is not a resounding universal "*yes*", because the most common option in these codes is to use the ML or REML method for the fit, thus assuming that the given data set can be fitted by a Gaussian LMM, *i.e.* an LMM in which both random effects and residual errors are assumed to follow Gaussian distributions. The Gaussian assumption, particularly for the random effects, appears dubious for many real world data sets. For instance, [15] present practical cases of non normal random effects. Since random effects are a key component of interest in an LMM, indeed making the difference *w.r.t.* a classical Linear Model (LM), there is a need to either develop new fitting methodologies for LMMs which impose as few assumptions as possible on them, or show that some of the already existing ones are actually valid outside of the Gaussian realm.

As for the latter goal, and considering their leading role as fitting methodology for Gaussian LMMs nowadays, there have been several attempts at trying to justify solving the Gaussian REML equations to fit non Gaussian LMMs as well. But be-

fore briefly expanding on these attempts, we first stress that we are not interested here in asymptotics like the results established, under some additional assumptions by [16] which prove that for ANOVA LMMs, the solutions to the REML equations are consistent estimators of the variance components and asymptotically Gaussian, even when the ANOVA LMM is not Gaussian. We rather focus here on finite distance validation of REML equations for fitting a given data set. In that respect, it is well known that the ANOVA estimating equations (EEs) are equivalent to the REML ones in a balanced ANOVA LMM [14]. Similarly, the iterative variant of Rao's MINQUE (the I-MINQUE in [17]) yields variance components estimates in ANOVA LMMs which satisfy the REML equations. However, these methods apply only to ANOVA LMMs, and, more importantly, they do not guarantee nonnegative estimates of variances, constraint REML estimates *must satisfy*, beyond the REML equations.

On the other hand, the results of [16] mentioned above validate, but only on an asymptotic basis, the *quasi-likelihood approach* of using the Gaussian REML equations to fit an arbitrary ANOVA LMM. Heyde [18] does the same for an arbitrary LMM, but for fixed sample size and in the framework of the relationship between quasi-likelihood and estimating functions. Although the latter framework is more general than that of LMMs and the REML equations are derived in it "painlessly" (to quote Heyde) under the only additional assumption that the LMM unit response has kurtosis 3 (like a Gaussian univariate distribution), that elegant framework [19] is almost a whole alternate mathematical universe to that in which most LMMs aficionados usually live in. Moreover, only the REML equations are derived in that framework, with no allusion to the needed nonnegativity constraints attached to them and alluded to above.

Now, for fitting LMMs, we will take a practical standpoint as in [20] that, in most situations, except when some additional information is available about the data (such as serial correlation in errors for some longitudinal data), one has no other choice than to assume that the residual errors in the LMM are *uncorrelated and homoscedastic* (*u.ho.*). In this paper (as in [20]), it is our objective under that assumption to devise a new iterative approach code named 3*S*-A1.RE for estimating the model parameters without imposing any Gaussian assumptions, particularly for the clusters random effects. We do so, first for the 2-level (or longitudinal) LMMs. As will be remarked later for the set of estimating equations defining our new method *(i)* nonnegativity constraints for variance estimates and positive semi-definiteness for covariance matrices are always guaranteed and *(ii)* when the LMM to fit is Gaussian with *i.i.d.* errors, the set of estimating equations is exactly equivalent to the REML ones. Our approach also turns out to generalize an estimating procedure for REML computation in Gaussian variance components models credited to Henderson by [21] and detailed in [14, pages 278-279]. Finally, we adapt the approach to fitting ANOVA LMMs as well.

The rest of the paper is organized as follows. Section 2 briefly reviews LMMs, up to the random effects prediction issue, emphasizing the famous Best Linear Unbiased Predictor (BLUP), a key tool in our new approach. Section 3 presents some key results on LMMs needing no Gaussian distributional assumption, and which will be the basis of our new estimation methodology designed in Section 4, first for 2-level LMMs with u.ho. errors (and no Gaussian assumption involved). Section 5 shows that the new estimation methodology can be adapted to some other types of LMMs with u.ho. errors, including ANOVA LMMs. In Section 6, the new estimation approach is applied to two data sets, a clustered one and a longitudinal one, and the results compared with those from Gaussian REML, with implementations in the R software [22]. Section 7 gives a conclusion about the work presented and suggests a tantalizing perspective about a possible continuation from this. Finally, the Appendix contains the most lengthy proof of a result presented in the running text.

This paper uses the same notations as in [20]. But here, to gain space, $\mathrm{Rows}(A_1, \cdots, A_m)$ will denote the matrix obtained by vertically stacking matrices $A_1, \cdots, A_m$ having the same number of columns.

2 Linear mixed models

2.1 The general form of an LMM

The most general form of an LMM with observed response $\mathbf{Y} \in \mathbb{R}^n$ is [11]:

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}U + \boldsymbol{\varepsilon}, \tag{2.1}$$

with $\boldsymbol{\beta} \in \mathbb{R}^p$ unknown vector of fixed effects parameters, $U \in \mathbb{R}^q$ vector of unobserved random effects, $\boldsymbol{\varepsilon} \in \mathbb{R}^n$ vector of unobserved residual errors, while $\mathbf{X} \in \mathcal{M}_{n,p}(\mathbb{R})$ and $\mathbf{Z} \in \mathcal{M}_{n,q}(\mathbb{R})$ are given design matrices. The standard assumptions for (2.1) are:

Assumption $\mathcal{A}\mathbf{1}_g$. *U and $\boldsymbol{\varepsilon}$ are two independent and zero mean random vectors with respective positive definite covariance matrices $\mathbb{M}\mathrm{cov}(U) = \mathbf{G} \in \mathcal{M}_q(\mathbb{R})$ and $\mathbb{M}\mathrm{cov}(\boldsymbol{\varepsilon}) = \mathbf{R} \in \mathcal{M}_n(\mathbb{R})$.*

Then the covariance matrix of $\mathbf{Y}$ is $\mathbf{V} = \mathbb{M}\mathrm{cov}(\mathbf{Y}) = \mathbf{Z}\mathbf{G}\mathbf{Z}^T + \mathbf{R}$. To simplify the presentation, we also use:

Assumption $\mathcal{A}\mathbf{2}_g$. *The $n \times p$ design matrix $\mathbf{X}$ in* (2.1) *has full column rank.*

Modern days LMM fitting methodology to a given data set (through estimating the parameters $\boldsymbol{\beta}$, $\mathbf{R}$, $\mathbf{G}$) does so, most often, assuming that the LMM (2.1) is Gaussian. But because that assumption may not hold for many real world data sets (especially as far as the random effects vector U is concerned), the goal here is to devise iterative estimation procedures for LMMs only under the basic Assumptions $\mathcal{A}1_g$-$\mathcal{A}2_g$, and the added assumption that:

Assumption $\mathcal{A}\mathbf{3}_g$. $\mathbf{R} = \sigma^2_{\boldsymbol{\varepsilon}}\mathbf{I}_n$ *($\sigma^2_{\boldsymbol{\varepsilon}} > 0$).*

2.2 The 2-level (or *longitudinal*) LMM

This is one of the types of LMM most used to analyze multilevel data and we will first present our estimation methodology

on that type. For all units in each cluster j it takes the form [23]:

$$\mathbf{Y}_j = \mathbf{X}_j\boldsymbol{\beta} + \mathbf{Z}_j U_j + \boldsymbol{\varepsilon}_j, \quad j = 1, \cdots, m, \tag{2.2}$$

with $\mathbf{Y}_j = (Y_{1j}, \cdots, Y_{n_j,j})^T \in \mathbb{R}^{n_j}$, the response vector; $U_j \in \mathbb{R}^r$, the unobserved vector of random effects for cluster j; $\boldsymbol{\varepsilon}_j = (\varepsilon_{1j}, \cdots, \varepsilon_{n_j,j})^T \in \mathbb{R}^{n_j}$, the vector of the unobserved errors; $\mathbf{X}_j = \text{Rows}(\mathbf{X}_{1j}, \cdots, \mathbf{X}_{n_j,j}) \in \mathcal{M}_{n_j,p}(\mathbb{R})$, the fixed effects design matrix; $\mathbf{Z}_j = \text{Rows}(\mathbf{Z}_{1j}, \cdots, \mathbf{Z}_{n_j,j}) \in \mathcal{M}_{n_j,r}(\mathbb{R})$, the random effects design matrix. Assumption $\mathcal{A}1_g$ translates here to:

Assumption $\mathcal{A}$1. *The U_j's are independent and identically distributed in $\mathbb{R}^r$.*

Assumption $\mathcal{A}$2. *The $\boldsymbol{\varepsilon}_j$'s are independent.*

Assumption $\mathcal{A}$3. *The set of U_j's and set of $\boldsymbol{\varepsilon}_j$'s are independent from each other.*

Assumption $\mathcal{A}$4. $\mathbb{E}(U_j) = \mathbf{0} \in \mathbb{R}^r$, $\mathbb{E}(\boldsymbol{\varepsilon}_j) = \mathbf{0} \in \mathbb{R}^{n_j}$, $\mathbb{M}\text{cov}(U_j) = \mathbf{D} \in \mathcal{M}_r(\mathbb{R})$, $\mathbb{M}\text{cov}(\boldsymbol{\varepsilon}_j) = \mathbf{R}_j \in \mathcal{M}_{n_j}(\mathbb{R})$, *with $\mathbf{D}$ and $\mathbf{R}_j$ symmetric and positive definite (SPD).*

Model (2.2) can be written as (2.1) but with the covariance matrices have the following special diagonal block structures: $\mathbf{G} = \text{diag}(\mathbf{D}, \cdots, \mathbf{D}) \in \mathcal{M}_q(\mathbb{R})$, $\mathbf{R} = \text{diag}(\mathbf{R}_1, \cdots, \mathbf{R}_m) \in \mathcal{M}_n(\mathbb{R})$, $\mathbf{V} = \mathbf{Z}\mathbf{G}\mathbf{Z}^T + \mathbf{R} = \text{diag}(\mathbf{V}_1, \cdots, \mathbf{V}_m) \in \mathcal{M}_n(\mathbb{R})$, with $\mathbf{V}_j = \mathbf{Z}_j\mathbf{D}\mathbf{Z}_j^T + \mathbf{R}_j \in \mathcal{M}_{n_j}(\mathbb{R})$, $j = 1, \cdots, m$.

2.3 Prediction of random effects

In our framework for fitting LMMs, predicting U is, in fact, an embedded key component of the iterative estimation process. In [20], the random effects vector U was estimated (or predicted) through the *Best Linear Predictor* (BLP) of U given $\mathbf{Y}$ [14, Chapter 7]:

$$\text{BLP}(U|\,\mathbf{Y}) = \overline{U} = \mathbf{G}\mathbf{Z}^T\mathbf{V}^{-1}(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}). \tag{2.3}$$

For a Gaussian LMM, it is well known that $\text{BLP}(U|\,\mathbf{Y})$ coincides with $\text{BP}(U|\,\mathbf{Y}) = \mathbb{E}(U|\,\mathbf{Y})$, the *Best Predictor* (BP) of U given $\mathbf{Y}$. Such is not, generally, the case for a non Gaussian LMM.

In the 3*S*-A1.RE approach (to be designed in the next sections), first, as usual, we will aim at estimating $\boldsymbol{\beta}$ in any LMM by

$$\widetilde{\boldsymbol{\beta}}^* = (\mathbf{X}^T\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}^T\mathbf{V}^{-1}\mathbf{Y}, \tag{2.4}$$

its *Generalized Least Squares Estimator* (GLSE), which is also the *Best Linear Unbiased Estimator* (BLUE) of $\boldsymbol{\beta}$. But we shall rather estimate (or predict) the random effects vector U through its *Best Linear Unbiased Predictor* (BLUP) of $U|\,\mathbf{Y}$ given, under Assumptions $\mathcal{A}1_g$-$\mathcal{A}2_g$, by [24]:

$$\text{BLUP}(U|\,\mathbf{Y}) = \widetilde{U}^* = \mathbf{G}\mathbf{Z}^T\mathbf{V}^{-1}(\mathbf{Y} - \mathbf{X}\widetilde{\boldsymbol{\beta}}^*) = \mathbf{G}\mathbf{Z}^T P\mathbf{Y}, \tag{2.5}$$

where P is the n-by-n symmetric matrix

$$P = \mathbf{V}^{-1} - \mathbf{V}^{-1}\mathbf{X}(\mathbf{X}^T\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}^T\mathbf{V}^{-1}. \tag{2.6}$$

See [25]for more detailed discussions about the BLUP.

In practice, both $\widetilde{\boldsymbol{\beta}}^*$ and $\widetilde{U}^*$ cannot directly be computed from the data vector $\mathbf{Y}$ because they each depend on the unknown covariance matrices $\mathbf{G}$ and $\mathbf{R}$ which have to be estimated from the data to get estimates $\widehat{\mathbf{G}}$ and $\widehat{\mathbf{R}}$. Hence, with $\widehat{\mathbf{V}} = \mathbf{Z}\widehat{\mathbf{G}}\mathbf{Z}^T + \widehat{\mathbf{R}}$, practical estimates of $\widetilde{\boldsymbol{\beta}}^*$ and $\widetilde{U}^*$ are:

$$\widehat{\boldsymbol{\beta}}^* = \text{EBLUE}(U|\,\mathbf{Y}, \widehat{\mathbf{G}}, \widehat{\mathbf{R}}) = (\mathbf{X}^T\widehat{\mathbf{V}}^{-1}\mathbf{X})^{-1}\mathbf{X}^T\widehat{\mathbf{V}}^{-1}\mathbf{Y},$$
$$\widehat{U}^* = \text{EBLUP}(U|\,\mathbf{Y}, \widehat{\mathbf{G}}, \widehat{\mathbf{R}}) = \widehat{\mathbf{G}}\mathbf{Z}^T\widehat{\mathbf{V}}^{-1}(\mathbf{Y} - \mathbf{X}\widehat{\boldsymbol{\beta}}^*),$$

which are the so called *Empirical BLUE* (EBLUE) of $\boldsymbol{\beta}$ and *Empirical BLUP* (EBLUP) of $U|\,\mathbf{Y}$.

The 3*S*-A1.RE approach will use an iterative procedure based on the Henderson's mixed model equations briefly reviewed next.

2.4 Henderson's mixed model equations

Given $\mathbf{G}$ and $\mathbf{R}$ in a Gaussian LMM (2.1), trying to simultaneously get an estimator for $\boldsymbol{\beta}$ and a predictor for U, Henderson [26, 27] maximized, *w.r.t.* $\boldsymbol{\beta}$ and U (the latter also viewed as a parameter), the joint density of $(\mathbf{Y}, U)$. To achieve that, differentiating that density, successively *w.r.t.* $\boldsymbol{\beta}$ and U, one arrives at what are since then called *Henderson's mixed model equations* (HMMEs):

$$\begin{pmatrix} \mathbf{X}^T\mathbf{R}^{-1}\mathbf{X} & \mathbf{X}^T\mathbf{R}^{-1}\mathbf{Z} \\ \mathbf{Z}^T\mathbf{R}^{-1}\mathbf{X} & \mathbf{Z}^T\mathbf{R}^{-1}\mathbf{Z} + \mathbf{G}^{-1} \end{pmatrix} \begin{pmatrix} \widetilde{\boldsymbol{\beta}} \\ \widetilde{U} \end{pmatrix} = \begin{pmatrix} \mathbf{X}^T\mathbf{R}^{-1}\mathbf{Y} \\ \mathbf{Z}^T\mathbf{R}^{-1}\mathbf{Y} \end{pmatrix}. \tag{2.7}$$

Beyond the initial (and, for some, still lingering) debate about the validity of the procedure, our practical stance is that that debate is closed by the following remarkable result, credited by [21] to Henderson in an unpublished 1963 report:

Theorem 1. *If $\mathbf{X}$ is full column rank while $\mathbf{G}$ and $\mathbf{R}$ are SPD matrices, then the solutions of the linear system* (2.7) *are $\widetilde{\boldsymbol{\beta}} = \widetilde{\boldsymbol{\beta}}^*$ and $\widetilde{U} = \widetilde{U}^*$ given by* (2.4)-(2.5)*, with $\mathbf{V} = \mathbf{Z}\mathbf{G}\mathbf{Z}^T + \mathbf{R}$.*

2.5 REML estimation of parameters in a Gaussian LMM

Nowadays, the REML estimation procedure is the most popular approach for fitting an LMM to a given data set. Traditionally, it assumes that the LMM to fit is Gaussian. In that parametric Gaussian setting, there are two quite distinct leading approaches to derive the REML procedure. One uses a Bayesian viewpoint [28], but the most popular (and the original one) remains the frequentist *error contrasts* approach [29]. A very short, but revealing account on the origins of the REML method is included in [30].

In the error contrasts approach, if one assumes a Gaussian LMM (2.1) also satisfying Assumption $\mathcal{A}2_g$, then to estimate $\mathbf{G}$ and $\mathbf{R}$, one considers A, a maximal rank *matrix of error contrasts* for the LMM, *i.e.* such that:

$$A \in \mathcal{M}_{n,n-p}(\mathbb{R}) \text{ has full column rank and } A^T\mathbf{X} = \mathbf{0}. \tag{2.9}$$

Otherwise stated, the columns of A are orthogonal to those of the fixed effects design matrix $\mathbf{X}$ and, in that category, A has the maximal rank possible ($n - p$ under Assumption $\mathcal{A}2_g$).

Then REML starts by left multiplying the LMM (2.1) by A^T to get:

$$\widetilde{\mathbf{Y}} = \widetilde{\mathbf{Z}}U + \widetilde{\boldsymbol{\varepsilon}}, \tag{2.10a}$$

where

$$\widetilde{\mathbf{Y}} = A^T\mathbf{Y}, \quad \widetilde{\mathbf{Z}} = A^T\mathbf{Z}, \quad \widetilde{\boldsymbol{\varepsilon}} = A^T\boldsymbol{\varepsilon}. \tag{2.10b}$$

So (2.10a) is also an LMM with the same random effects U as (2.1), response $\widetilde{\mathbf{Y}}$, random effects design matrix $\widetilde{\mathbf{Z}}$, residual errors vector $\widetilde{\boldsymbol{\varepsilon}}$, but with no fixed effects. The vector of unknown parameters in that LMM is $\boldsymbol{\theta} = (\boldsymbol{\theta}_U^T, \boldsymbol{\theta}_{\boldsymbol{\varepsilon}}^T)^T$, where $\boldsymbol{\theta}_U$ are the parameters in $\mathbf{G}$ and $\boldsymbol{\theta}_{\boldsymbol{\varepsilon}}$ are those in $\mathbf{R}$. Since (2.10a) is also a Gaussian LMM, the REML applies to it the classical ML procedure to estimate $\boldsymbol{\theta}$. Hence, it can be shown that the log-likelihood for (2.1) is the so called *restricted log-likelihood* for (2.10a) which is:

$$\begin{aligned}\ell_{\mathrm{R}}(\boldsymbol{\theta}|\,\mathbf{y}) =& c - \frac{1}{2}\ln|\mathbf{V}(\boldsymbol{\theta})| \\ &+ \frac{1}{2}(\mathbf{y} - \mathbf{X}\widetilde{\boldsymbol{\beta}}^*(\boldsymbol{\theta}))^T\mathbf{V}(\boldsymbol{\theta})^{-1}(\mathbf{y} - \mathbf{X}\widetilde{\boldsymbol{\beta}}^*(\boldsymbol{\theta})) \\ &+ \frac{1}{2}\ln|\mathbf{X}^T\mathbf{V}(\boldsymbol{\theta})^{-1}\mathbf{X}|, \end{aligned} \tag{2.11}$$

where $\mathbf{V}(\boldsymbol{\theta}) = \mathbb{M}\mathrm{cov}(\mathbf{Y})$, c is a constant and $\widetilde{\boldsymbol{\beta}}^*$ is given by (2.4).

The most remarkable feature about (2.11) is that the matrix A used to derive the LMM (2.10a) appears nowhere in its final expression (2.11), except, maybe, embedded in the constant c. Hence, one of the strengths of the REML methodology is that, to implement it, there is no need to first find a matrix A satisfying (2.9), which would be computationally prohibitive for big sample sizes.

3 Some non Gaussian-related results on LMMs

In the new estimation methodology to present here, the mindset is the same as in [20], hence some of the results needed here have already been proven there. We just recall the ones we will be using subsequently, while adding other results which are vital for the new methodology.

The first useful fact is that, contrary to the widespread belief grounded in the original derivation by Henderson, the system of HMMEs (2.7) is not attached to Gaussian distributional assumptions. Indeed, Theorem 1 is a pure Matrix Algebra result showing that the linear system (2.7) has as unique solutions the BLUE $\widetilde{\boldsymbol{\beta}}^*$ of $\boldsymbol{\beta}$ and the BLUP $\widetilde{U}^*$ of U respectively given by (2.4) and (2.5), and, therefore, characterizes them simultaneously and can be used to compute them efficiently. In effect, neither $\widetilde{\boldsymbol{\beta}}^*$ nor $\widetilde{U}^*$ requires any Gaussian distributional assumption for its derivation [24].

3.1 HMMEs solutions in an LMM with u.ho. errors

Under Assumption $\mathcal{A}3_g$ (in addition to $\mathcal{A}1_g$-$\mathcal{A}2_g$), the HMMEs (2.7) reduce to:

$$\begin{pmatrix}\mathbf{X}^T\mathbf{X} & \mathbf{X}^T\mathbf{Z} \\ \mathbf{Z}^T\mathbf{X} & \mathbf{Z}^T\mathbf{Z} + \sigma_{\boldsymbol{\varepsilon}}^2\mathbf{G}^{-1}\end{pmatrix}\begin{pmatrix}\widetilde{\boldsymbol{\beta}} \\ \widetilde{U}\end{pmatrix} = \begin{pmatrix}\mathbf{X}^T\mathbf{Y} \\ \mathbf{Z}^T\mathbf{Y}\end{pmatrix}, \tag{3.1}$$

whose unique solutions (see Theorem 1) $\widetilde{\boldsymbol{\beta}} = \widetilde{\boldsymbol{\beta}}^*$ of $\boldsymbol{\beta}$ and $\widetilde{U} = \widetilde{U}^*$ of $U|\mathbf{Y}$ yield the following results.

Theorem 2. *In the LMM* (2.1) *with Assumptions $\mathcal{A}1_g$-$\mathcal{A}2_g$, $\mathcal{A}3_g$, and P given by* (2.6)*,*

$$\mathbf{Y} - \mathbf{X}\widetilde{\boldsymbol{\beta}}^* - \mathbf{Z}\widetilde{U}^* = \sigma_{\boldsymbol{\varepsilon}}^2\mathbf{V}^{-1}(\mathbf{Y} - \mathbf{X}\widetilde{\boldsymbol{\beta}}^*) = \sigma_{\boldsymbol{\varepsilon}}^2 P\mathbf{Y}, \tag{3.2a}$$

$$\widetilde{\boldsymbol{\beta}}^* = (\mathbf{X}^T\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}^T\mathbf{V}^{-1}\mathbf{Y} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T(\mathbf{Y} - \mathbf{Z}\widetilde{U}^*). \tag{3.2b}$$

Corollary 3. *If $A = \mathbf{Y}^T(\mathbf{Y} - \mathbf{X}\widetilde{\boldsymbol{\beta}}^* - \mathbf{Z}\widetilde{U}^*)$ and $B = \|\mathbf{Y} - \mathbf{X}\widetilde{\boldsymbol{\beta}}^* - \mathbf{Z}\widetilde{U}^*\|^2$ in the LMM* (2.1) *with Assumptions $\mathcal{A}1_g$-$\mathcal{A}2_g$ and $\mathcal{A}3_g$, then*

$$A - B = \sigma_{\boldsymbol{\varepsilon}}^2 \cdot \|\mathbf{Z}^T\mathbf{V}^{-1}(\mathbf{Y} - \mathbf{X}\widetilde{\boldsymbol{\beta}}^*)\|_{\mathbf{G}}^2 = \sigma_{\boldsymbol{\varepsilon}}^2 \cdot \|\widetilde{U}^*\|_{\mathbf{G}^{-1}}^2. \tag{3.3}$$

Therefore, the real random variable $\mathbf{Y}^T(\mathbf{Y} - \mathbf{X}\widetilde{\boldsymbol{\beta}}^* - \mathbf{Z}\widetilde{U}^*)$ is always nonnegative.

The following theorem produces moment based estimators for $\sigma_{\boldsymbol{\varepsilon}}^2$.

Theorem 4. *In the LMM* (2.1) *with Assumptions $\mathcal{A}1_g$-$\mathcal{A}2_g$ and $\mathcal{A}3_g$, one has:*

$$\mathbb{E}\left[\mathbf{Y}^T(\mathbf{Y} - \mathbf{X}\widetilde{\boldsymbol{\beta}}^* - \mathbf{Z}\widetilde{U}^*)\right] = \sigma_{\boldsymbol{\varepsilon}}^2 \cdot (n - p), \tag{3.4a}$$

$$\mathbb{E}\,\|\mathbf{Y} - \mathbf{X}\widetilde{\boldsymbol{\beta}}^* - \mathbf{Z}\widetilde{U}^*\|^2 = \sigma_{\boldsymbol{\varepsilon}}^4 \cdot \mathrm{tr}(P). \tag{3.4b}$$

We add the following interesting corollary of Theorem 2:

Corollary 5. *In the LMM* (2.1) *with Assumptions $\mathcal{A}1_g$-$\mathcal{A}2_g$ and $\mathcal{A}3_g$,*

$$\mathbf{Y} - \mathbf{X}\widetilde{\boldsymbol{\beta}}^* - \mathbf{Z}\widetilde{U}^* = M(\mathbf{Y} - \mathbf{Z}\widetilde{U}^*), \tag{3.5}$$

where M is the $n \times n$ symmetric matrix

$$M = \mathbf{I}_n - \mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T, \tag{3.6}$$

i.e. *the matrix of the orthogonal projection in $\mathbb{R}^n$,* w.r.t. *the dot product, into $\mathcal{R}(\mathbf{X})^\perp$, the orthogonal complement linear subspace of $\mathcal{R}(\mathbf{X})$, the range space of $\mathbf{X}$, in $\mathbb{R}^n$.*

Proof. Let Assumptions $\mathcal{A}1_g$-$\mathcal{A}2_g$ and $\mathcal{A}3_g$ hold in the LMM (2.1). Then, using (3.2b) and (3.6),

$$\mathbf{Y} - \mathbf{X}\widetilde{\boldsymbol{\beta}}^* - \mathbf{Z}\widetilde{U}^* = M(\mathbf{Y} - \mathbf{Z}\widetilde{U}^*). \qquad \square$$

3.2 Preliminaries for adapting the REML construction to non Gaussian LMMs with u.ho. errors

3.2.1 The Matrix Algebra behind the REML construction

In Section 2.5, we emphasized the classical well known fact that though A is used to construct the REML method in the approach by error contrasts, there is actually no need to compute matrix A satisfying (2.9) in order to be able to fit a Gaussian LMM to a given data using the method. The reason behind the final disappearance of A in the Gaussian REML approach has nothing to do with the Gaussian assumptions. It is just a consequence of the remarkable Matrix Algebra results about any matrix A satisfying (2.9) and generally discovered by pioneers in LMM modeling:

Theorem 6. *If Assumption $\mathcal{A}2_g$ holds, then there exist matrices A satisfying* (2.9)*. For any of them, we have:*

1. $A^TA = \mathbf{I}_{n-p}$ *and* $AA^T = M$*, the $n \times n$ symmetric matrix given by* (3.6)*;*
2. *For any SPD matrix $W \in \mathcal{M}_n(\mathbb{R})$, one has:*
$$A(A^TWA)^{-1}A^T = W^{-1} - W^{-1}\mathbf{X}(\mathbf{X}^TW^{-1}\mathbf{X})^{-1}\mathbf{X}^TW^{-1}$$
3. *Hence, in the LMM* (2.1) *with Assumptions $\mathcal{A}1_g$-$\mathcal{A}2_g$, A and $\mathbf{V} = \mathbb{M}\mathrm{cov}(\mathbf{Y})$ are related by:*
$$A(A^T\mathbf{V}A)^{-1}A^T = P, \tag{3.7}$$
where P is the n-by-n symmetric matrix given by (2.6)*.*

3.2.2 More about the transformed LMM (2.10a) without Gaussian assumptions

As a consequence of the results in Theorem 6, the transformed LMM (2.10a) enjoys some remarkable properties not requiring that any Gaussian distributional assumption be imposed in the initial LMM (2.1). The first three and well known such properties are rather trivial:

$$\begin{aligned}&\mathbb{E}(\widetilde{\mathbf{Y}}) = \mathbf{0}, \quad \widetilde{\mathbf{V}} = \mathbb{M}\mathrm{cov}(\widetilde{\mathbf{Y}}) = A^T\mathbf{V}A,\\ &\widetilde{\mathbf{R}} = \mathbb{M}\mathrm{cov}(\widetilde{\boldsymbol{\varepsilon}}) = A^T\mathbf{R}A,\end{aligned} \tag{3.8}$$

where the notations $\widetilde{\mathbf{Y}}$, $\widetilde{\mathbf{Z}}$, $\widetilde{\boldsymbol{\varepsilon}}$ are as defined in (2.10b). The next property is simple to state, but very striking. To the best of our knowledge, it seems never to have been noticed elsewhere before.

Theorem 7. *If the LMM* (2.1) *satisfies Assumptions $\mathcal{A}1_g$-$\mathcal{A}2_g$, then, in the LMM* (2.10a)*:*

$$\widetilde{U} = \mathrm{BLP}(U|\widetilde{\mathbf{Y}}) = \widetilde{U}^* = \mathrm{BLUP}(U|\mathbf{Y}), \tag{3.9}$$

i.e. *predicting the random effects U through the BLP of $U|\widetilde{\mathbf{Y}}$ in the transformed LMM* (2.10a) *is equivalent to doing the same with the BLUP of $U|\mathbf{Y}$ in the initial LMM* (2.1)*.*

Proof. Let Assumptions $\mathcal{A}1_g$-$\mathcal{A}2_g$ hold in the LMM (2.1). To get (3.9), we apply, to the LMM (2.10a), formula (2.3) giving the BLP of $U|\mathbf{Y}$ in an arbitrary LMM (2.1), then the second equality of (3.8) to get $\widetilde{\mathbf{V}} = \mathbb{M}\mathrm{cov}(\widetilde{\mathbf{Y}})$, followed by (3.7) and (2.5):

$$\begin{aligned}\widetilde{U} &= \mathrm{BLP}(U|\widetilde{\mathbf{Y}}) = \mathbf{G}\widetilde{\mathbf{Z}}^T\widetilde{\mathbf{V}}^{-1}\widetilde{\mathbf{Y}}\\ &= \mathbf{G}\mathbf{Z}^TA(A^T\mathbf{V}A)^{-1}A^T\mathbf{Y} = \mathbf{G}\mathbf{Z}^TP\mathbf{Y} = \widetilde{U}^*.\square\end{aligned}$$

When the initial LMM (2.1) has u.ho. errors, one can even move further than Theorem 7. Indeed, the transformed LMM (2.10a) exhibits the following additional remarkable properties, which intrinsically relate it to the $\mathrm{BLUE}-\mathrm{BLUP}$ couple $(\widetilde{\boldsymbol{\beta}}^*, \widetilde{U}^*)$ in the initial LMM (2.1):

Theorem 8. *If Assumptions $\mathcal{A}1_g$-$\mathcal{A}2_g$ and $\mathcal{A}3_g$ hold in the LMM* (2.1)*, and $\widetilde{U} = \mathrm{BLP}(U|\widetilde{\mathbf{Y}})$, then*

$$\mathbb{M}\mathrm{cov}(\widetilde{\boldsymbol{\varepsilon}}) = \sigma^2_{\boldsymbol{\varepsilon}}\mathbf{I}_{n-p}, \tag{3.10a}$$
$$\widetilde{\mathbf{Y}}^T(\widetilde{\mathbf{Y}} - \widetilde{\mathbf{Z}}\widetilde{U}) = \mathbf{Y}^T(\mathbf{Y} - \mathbf{X}\widetilde{\boldsymbol{\beta}}^* - \mathbf{Z}\widetilde{U}^*), \tag{3.10b}$$
$$\|\widetilde{\mathbf{Y}} - \widetilde{\mathbf{Z}}\widetilde{U}\| = \|\mathbf{Y} - \mathbf{X}\widetilde{\boldsymbol{\beta}}^* - \mathbf{Z}\widetilde{U}^*\|. \tag{3.10c}$$

Proof. Identity (3.10a) is immediate to check, whereas, using (3.9), then Theorem 6, and Corollary 5,

$$\begin{aligned}\widetilde{\mathbf{Y}}^T(\widetilde{\mathbf{Y}} - \widetilde{\mathbf{Z}}\widetilde{U}) &= \mathbf{Y}^TAA^T(\mathbf{Y} - \mathbf{Z}\widetilde{U}^*) = \mathbf{Y}^TM(\mathbf{Y} - \mathbf{Z}\widetilde{U}^*)\\ &= \mathbf{Y}^T(\mathbf{Y} - \mathbf{X}\widetilde{\boldsymbol{\beta}}^* - \mathbf{Z}\widetilde{U}^*).\end{aligned}$$

Finally, using Theorem 6 again and the fact that M is idempotent and symmetric,

$$\begin{aligned}\|\widetilde{\mathbf{Y}} - \widetilde{\mathbf{Z}}\widetilde{U}\|^2 &= \|A^T(\mathbf{Y} - \mathbf{Z}\widetilde{U}^*)\|^2\\ &= (\mathbf{Y} - \mathbf{Z}\widetilde{U}^*)^TM^2(\mathbf{Y} - \mathbf{Z}\widetilde{U}^*)\\ &= \|\mathbf{Y} - \mathbf{X}\widetilde{\boldsymbol{\beta}}^* - \mathbf{Z}\widetilde{U}^*\|^2, \quad \text{from Corollary 5 } .\square\end{aligned}$$

So, if the LMM (2.1) has u.ho. errors, then, from (3.10a), the transformed LMM (2.10a) also has u.ho. errors and with the same residual variance $\sigma^2_{\boldsymbol{\varepsilon}}$, but with an $(n-p)$-response vector. The usefulness of the two last identities (3.10b) and (3.10b) will be explained later on.

3.3 More about the BLUP of $U|\,\mathbf{Y}$ in an LMM

The BLUP $\widetilde{U}^*$, given by (2.5), will play a key role in our new estimation methodology for LMMs with u.ho. errors. That is why we expand on it here. And, again, no Gaussian distributional assumption will be involved.

3.3.1 Covariance matrix and another expression of the BLUP of $U|\,\mathbf{Y}$ in an LMM

The relationship between the covariance matrix of $\widetilde{U}^*$ and that of the true random effects vector U will provide the crux to devise an EE for the latter in our methodology. To derive it, first

$$\mathbb{M}\mathrm{cov}(\widetilde{U}^*) = \mathbf{G}\mathbf{Z}^TP\mathbf{V}P\mathbf{Z}\mathbf{G} = \mathbf{G}\mathbf{Z}^TP\mathbf{Z}\mathbf{G}, \tag{3.11a}$$

thanks to the well known identity $P\mathbf{V}P = P$. Hence, since $\mathbb{E}(\widetilde{U}^*) = \mathbf{0}$.

$$\mathbb{M}\mathrm{cov}(\widetilde{U}^*) = \mathbb{E}(\widetilde{U}^* \widetilde{U}^{*T}) = \mathbf{G} - \mathbf{W}^*, \tag{3.11b}$$

where $\mathbf{W}^*$ is the symmetric matrix:

$$\mathbf{W}^* = \mathbf{G} - \mathbf{G}\mathbf{Z}^T P \mathbf{Z}\mathbf{G}. \tag{3.11c}$$

Note then that (3.11b) implies that we get an unbiased estimator of the random effects covariance matrix $\mathbf{G}$ through the matrix[1]

$$\widetilde{\mathbf{G}}_2 = \mathbf{W}^* + \widetilde{U}^* \widetilde{U}^{*T}. \tag{3.12}$$

But $\widetilde{\mathbf{G}}_2$ is only a preliminary estimator of $\mathbf{G}$ because it still depends on the unknown parameters, including $\mathbf{G}$ itself. Nevertheless, it can serve as basis to design an EE for $\mathbf{G}$, provided there is a simpler formula for the awkward looking $q \times q$ matrix $\mathbf{W}^*$. Indeed, such a formula exists when the LMM has u.ho. residual errors:

Lemma 9. *In the LMM* (2.1) *with Assumptions $\mathcal{A}1_g$-$\mathcal{A}2_g$, $\mathcal{A}3_g$, and M given by* (3.6)*,*

$$\mathbf{W}^* = (\sigma_\varepsilon^{-2}\mathbf{Z}^T M \mathbf{Z} + \mathbf{G}^{-1})^{-1} = \sigma_\varepsilon^2 (\mathbf{Z}^T M \mathbf{Z} + \sigma_\varepsilon^2 \mathbf{G}^{-1})^{-1}, \tag{3.13a}$$

$$\mathbf{G}\mathbf{Z}^T P = \sigma_\varepsilon^{-2} \mathbf{W}^* \mathbf{Z}^T M. \tag{3.13b}$$

Proof. From the LMM (2.1) with Assumptions $\mathcal{A}1_g$-$\mathcal{A}2_g$ and $\mathcal{A}3_g$, let us consider the transformed LMM (2.10a), where A is any matrix satisfying (2.9). Thanks to (3.10a) in Theorem 8, we know that (2.10a) is also an LMM with u.ho. errors of common variance σ_ε^2. Then (3.13a) and (3.13b) result from applying to that LMM the identities (S:1.2c)-(S:1.2d) of Lemma S1.2 in the Online Resource of [20]. Indeed, for that LMM, $\mathbf{V}^*$, $\mathbf{Z}$, $\mathbf{V}$ respectively become $\mathbf{W}^*$ (as a consequence of Theorem 7), $\widetilde{\mathbf{Z}} = A^T\mathbf{Z}$, $\widetilde{\mathbf{V}} = \mathbb{M}\mathrm{cov}(\widetilde{\mathbf{Y}}) = \sigma_\varepsilon^2 \mathbf{I}_{n-p}$, so that those identities yield:

$$\mathbf{W}^* = (\sigma_\varepsilon^{-2}\widetilde{\mathbf{Z}}^T\widetilde{\mathbf{Z}} + \mathbf{G}^{-1})^{-1}, \tag{3.14a}$$

$$\mathbf{G}\widetilde{\mathbf{Z}}^T\widetilde{\mathbf{V}}^{-1} = \sigma_\varepsilon^{-2}\mathbf{W}^*\widetilde{\mathbf{Z}}^T. \tag{3.14b}$$

Now, (3.14a) $\iff$ (3.13a), because $\widetilde{\mathbf{Z}}^T\widetilde{\mathbf{Z}} = \mathbf{Z}^T A A^T \mathbf{Z} = \mathbf{Z}^T M \mathbf{Z}$, the latter thanks to Theorem 6. Similarly, using (3.8), then Theorem 6 again,

$$\begin{aligned}(3.14b) &\iff \mathbf{G}\mathbf{Z}^T A (A^T\mathbf{V}A)^{-1} = \sigma_\varepsilon^{-2}\mathbf{W}^*\mathbf{Z}^T A \\ &\iff \mathbf{G}\mathbf{Z}^T A (A^T\mathbf{V}A)^{-1} A^T = \sigma_\varepsilon^{-2}\mathbf{W}^*\mathbf{Z}^T A A^T \\ &\iff (3.13b).\square\end{aligned}$$

Using arguments similar to those in the above proof, we deduce the following series of results, the first of which shows that the $q \times q$ matrix $\mathbf{W}^*$ allows an alternative useful way of rewriting the BLUP (2.5) in the LMM (2.1) with u.ho. residual errors:

[1] That matrix is denoted $\widetilde{\mathbf{G}}_2$ to distinguish it from the one playing the analogous role in [20] and denoted there $\widetilde{\mathbf{G}}_1$.

Proposition 10. *In the LMM* (2.1) *with Assumptions $\mathcal{A}1_g$-$\mathcal{A}2_g$ and $\mathcal{A}3_g$,*

$$\widetilde{U}^* = \mathrm{BLUP}(U|\,\mathbf{Y}) = \sigma_\varepsilon^{-2}\,\mathbf{W}^*\mathbf{Z}^T M \mathbf{Y}. \tag{3.15}$$

Proposition 11. *In the LMM* (2.1) *with Assumptions $\mathcal{A}1_g$-$\mathcal{A}2_g$, one has:*

$$\mathbb{E}(\|\widetilde{U}^*\|^2_{\mathbf{G}^{-1}}) \;=\; q - \mathrm{tr}(\mathbf{G}^{-1}\mathbf{W}^*). \tag{3.16a}$$

Moreover, if Assumption $\mathcal{A}3_g$ also holds, then we also have:

$$\mathbb{E}(\|\widetilde{U}^*\|^2_{\mathbf{G}^{-1}}) \;=\; n - p - \sigma_\varepsilon^2\,\mathrm{tr}(P) \;=\; \sigma_\varepsilon^{-2}\,\mathrm{tr}(\mathbf{W}^*\mathbf{Z}^T M \mathbf{Z}). \tag{3.16b}$$

Proposition 12. *In the LMM* (2.1) *with Assumptions $\mathcal{A}1_g$-$\mathcal{A}2_g$ and $\mathcal{A}3_g$,*

$$\sigma_\varepsilon^2\,\mathrm{tr}(P) + \sigma_\varepsilon^{-2}\,\mathrm{tr}(\mathbf{W}^*\mathbf{Z}^T M\mathbf{Z}) = n - p, \tag{3.17a}$$

$$\mathrm{tr}(\mathbf{G}^{-1}\mathbf{W}^*) + \sigma_\varepsilon^{-2}\,\mathrm{tr}(\mathbf{W}^*\mathbf{Z}^T M\mathbf{Z}) = q. \tag{3.17b}$$

3.3.2 Cluster partitioning of $\widetilde{U}^*$ in a 2-level LMM

In the 2-level LMM (2.2), to handle $\widetilde{U}^*$ at the cluster level, we use its cluster decomposition:

$$\widetilde{U}^* = \mathrm{Rows}(\widetilde{U}_1^*, \cdots, \widetilde{U}_m^*), \tag{3.18}$$

with $\widetilde{U}_j^*$ a random r-vector ($j = 1, \cdots, m$). In a 2-level LMM with u.ho. errors, to compute $\widetilde{U}_1^*, \cdots, \widetilde{U}_m^*$, we first compute the global vector $\widetilde{U}^*$ through (3.15), then effect the appropriate partitioning of the result to get the $\widetilde{U}_j^*$'s.

We shall also need expressions for the covariance matrices of $\widetilde{U}_1^*, \cdots, \widetilde{U}_m^*$. But an operational difficulty then arises due to the fact that the matrices P and M do not share the same partitioned block-diagonal structure as the covariance matrix $\mathbf{V}$ of $\mathbf{Y}$. Nonetheless, given the partition (3.18) of $\widetilde{U}^*$, there is a conformal decomposition of the $q \times q$ ($q = rm$) matrix $\mathbf{W}^*$ with m blocks of r rows and m blocks of r columns. Let $\mathbf{W}^*_{jk}$ be the $r \times r$ submatrix at position (j, k) in that decomposition of $\mathbf{W}^*$, for $j, k = 1, \cdots, m$. Given (3.11b) and the block-diagonal structure of $\mathbf{G}$, one gets an analogous decomposition for the $q \times q$ matrix $\mathbb{M}\mathrm{cov}(\widetilde{U}^*)$ where the m diagonal $r \times r$ blocks are the respective covariance matrices of $\widetilde{U}_1^*, \cdots, \widetilde{U}_m^*$. So,

$$\mathbb{M}\mathrm{cov}(\widetilde{U}_j^*) = \mathbf{D} - \mathbf{W}^*_{jj} \quad (j = 1, \cdots, m). \tag{3.19}$$

The latter implies the following preliminary unbiased estimator of $\mathbf{D}$:

$$\widetilde{\mathbf{D}}_2 = \frac{1}{m}\sum_{j=1}^{m} \left(\mathbf{W}^*_{jj} + \widetilde{U}_j^* \widetilde{U}_j^{*T}\right), \tag{3.20}$$

which will allow to devise an EE for $\mathbf{D}$ in our new fitting methodology for 2-level LMMs with u.ho. errors.

4 Estimation in 2-level LMMs with u.ho. errors: *a nonparametric REML-like approach*

The aim is to estimate $\boldsymbol{\beta} \in \mathbb{R}^p$, σ_ε^2, $\mathbf{D} \in \mathcal{M}_r(\mathbb{R})$ and predict $U = (U_1^T, \cdots, U_m^T)^T \in \mathbb{R}^q$ in the 2-level LMM (2.2) under Assumptions $\mathcal{A}1$-$\mathcal{A}4$, $\mathcal{A}2_g$ and $\mathcal{A}3_g$. While the BLP of U was used in [20], here, following Theorem 7, we will rather predict U using

$$\begin{aligned}\widetilde{\mathbf{u}}^* &= \mathrm{BLUP}(U|\,\mathbf{Y}=\mathbf{y}) = (\widetilde{\mathbf{u}}_1^{*T}, \cdots, \widetilde{\mathbf{u}}_m^{*T})^T \\ &= \sigma_\varepsilon^{-2}\,\mathbf{W}^*\mathbf{Z}^T M\mathbf{y}, \end{aligned} \tag{4.1}$$

the BLUP of U given $\mathbf{Y} = \mathbf{y}$, where the last equality stems from (3.15) in Proposition 10, and $\mathbf{W}^*$ is the $q \times q$ matrix given by (3.13a) because Assumption $\mathcal{A}3_g$ is assumed to hold. The m r-vectors $\widetilde{\mathbf{u}}_1^*, \cdots, \widetilde{\mathbf{u}}_m^*$ are the realized values of $\widetilde{U}_1^*, \cdots, \widetilde{U}_m^*$ in the cluster block decomposition (3.18) of $\widetilde{U}^*$.

The BLP was used to predict the random effects in [20] because it is simpler to compute than the BLUP, and the same is true when trying to get their respective covariance matrices. However, the construction was based on solving the HMMEs to estimate $\boldsymbol{\beta}$ and predict U. Now, thanks to Theorem 1, the "natural" predictor of the random effects which appears among the solutions of the HMMEs is the BLUP. Thus, using the latter, rather than the BLP, to predict the random effects seems a more logical approach. So, acting in that way, one might expect to get a somewhat more efficient fitting methodology for LMMs than in [20].

To proceed, we are going to devise, using only sound nonparametric estimation principles, estimating equations from which we shall design two iterative procedures to compute estimates for $\boldsymbol{\beta}$, σ_ε^2, $\mathbf{D}$, $\widetilde{\mathbf{u}}^*$. Like in [20], each of the two subvariants of this new approach is deduced from a 3-step construction.

4.1 Two new 3-step sequences for estimation in 2-level LMMs with u.ho. errors

To get our estimating equations, the starting ideas are as follows (we provide 2 different versions). We will let $\widehat{\mathbf{G}} = \mathrm{diag}(\widehat{\mathbf{D}}, \cdots, \widehat{\mathbf{D}})$ when $\widehat{\mathbf{D}}$ is an estimate of $\mathbf{D}$.

4.1.1 Starting ideas: *Version 1*

• *Step 1*: Estimation of $\boldsymbol{\beta}$ and $\widetilde{\mathbf{u}}^*$, given estimates of σ_ε^2 and $\mathbf{D}$

Given some preliminary estimates $\widehat{\sigma}_\varepsilon^2$ and $\widehat{\mathbf{D}}$ of σ_ε^2 and $\mathbf{D}$, we obtain estimates $\widehat{\boldsymbol{\beta}}$ of $\boldsymbol{\beta}$, and $\widehat{\mathbf{u}}$ of $\widetilde{\mathbf{u}}^*$ by solving the system (3.1) using $\sigma_\varepsilon^2 = \widehat{\sigma}_\varepsilon^2$ and $\mathbf{G} = \widehat{\mathbf{G}}$.

• *Step 2*: Improved estimate of σ_ε^2, given $\widehat{\boldsymbol{\beta}}$ and $\widehat{\mathbf{u}}$ from Step 1

Given $\widehat{\boldsymbol{\beta}}$ and $\widehat{\mathbf{u}}$ from Step 1 above, (3.4a) in Theorem 4 suggests to take as hopefully improved estimate of σ_ε^2:

$$\widehat{\sigma}_\varepsilon^2 = \frac{\mathbf{y}^T(\mathbf{y} - \mathbf{X}\widehat{\boldsymbol{\beta}} - \mathbf{Z}\widehat{\mathbf{u}}\,)}{n-p}. \tag{4.2}$$

Thanks to Corollary 3 and how we got $\widehat{\boldsymbol{\beta}}$, $\widehat{\mathbf{u}}$ in Step 1, $\widehat{\sigma}_\varepsilon^2$ is nonnegative.

• *Step 3*: Improved estimate of $\mathbf{D}$ given preliminary estimates $\widehat{\sigma}_\varepsilon^2$, $\widetilde{\mathbf{D}}$, $\widehat{\mathbf{u}}$ of σ_ε^2, $\mathbf{D}$, $\widetilde{\mathbf{u}}^*$

Let $\widehat{\sigma}_\varepsilon^2$, $\widetilde{\mathbf{D}}$, $\widehat{\mathbf{u}}$ be respective preliminary estimates of σ_ε^2, $\mathbf{D}$, $\widetilde{\mathbf{u}}^*$. Then to get an improved estimate $\widehat{\mathbf{D}}$ of $\mathbf{D}$, we start with the unbiased estimator $\widetilde{\mathbf{D}}_2$ of $\mathbf{D}$ given by (3.20). Since it still depends on the unknown parameters, including $\mathbf{D}$ itself, using the preliminary estimates $\widetilde{\mathbf{D}}$ of $\mathbf{D}$, and $\widehat{\sigma}_\varepsilon^2$ of σ_ε^2, (3.13a) suggests estimating $\mathbf{W}^*$ by

$$\widehat{\mathbf{W}}^* = (\widehat{\sigma}_\varepsilon^{-2}\mathbf{Z}^T M\mathbf{Z} + \widetilde{\mathbf{G}}^{-1})^{-1} = \widehat{\sigma}_\varepsilon^2(\mathbf{Z}^T M\mathbf{Z} + \widehat{\sigma}_\varepsilon^2\widetilde{\mathbf{G}}^{-1})^{-1}, \tag{4.3}$$

where $\widetilde{\mathbf{G}} = \mathrm{diag}(\widetilde{\mathbf{D}}, \cdots, \widetilde{\mathbf{D}})$ (m times). Then, for $j = 1, \cdots, m$, given the available estimate $\widehat{\mathbf{u}}_j$ of $\widetilde{\mathbf{u}}_j^*$, a hopefully improved computable estimate of $\mathbf{D}$ will be:

$$\widehat{\mathbf{D}} = \frac{1}{m}\sum_{j=1}^{m}\left(\widehat{\mathbf{W}}_{jj}^* + \widehat{\mathbf{u}}_j\widehat{\mathbf{u}}_j^T\right), \tag{4.4}$$

where $\widehat{\mathbf{W}}_{jj}^*$ is the j^{th} $r \times r$ diagonal block of $\widehat{\mathbf{W}}^*$ when that $rm \times rm$ matrix is partitioned through m blocks of r rows and m blocks of r columns.

4.1.2 Starting ideas: *Version 2*

Here, Step 3 is the same as in Version 1, while Steps 1 and 2 are changed to:

• *Step 1*: Estimation of $\widetilde{\mathbf{u}}^*$ and $\boldsymbol{\beta}$, given preliminary estimates of $\mathbf{D}$, σ_ε^2

With respective preliminary estimates $\widehat{\mathbf{D}}$, $\widehat{\sigma}_\varepsilon^2$ of $\mathbf{D}$, σ_ε^2, we estimate $\widetilde{\mathbf{u}}^*$ by:

$$\widehat{\mathbf{u}} = \widehat{\sigma}_\varepsilon^{-2}\,\widehat{\mathbf{W}}^*\mathbf{Z}^T M\mathbf{y}, \quad \text{with } \widehat{\mathbf{W}}^* = \widehat{\sigma}_\varepsilon^2(\mathbf{Z}^T M\mathbf{Z} + \widehat{\sigma}_\varepsilon^2\widehat{\mathbf{G}}^{-1})^{-1}. \tag{4.5}$$

Then, thanks to Theorem 2, we hope to obtain an improved estimate of $\boldsymbol{\beta}$ through:

$$\widehat{\boldsymbol{\beta}} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T(\mathbf{y} - \mathbf{Z}\widehat{\mathbf{u}}\,). \tag{4.6}$$

• *Step 2*: Improved estimate of σ_ε^2, given $\widehat{\boldsymbol{\beta}}$, $\widehat{\mathbf{u}}$, from Step 1, and preliminary estimates of σ_ε^2, $\mathbf{D}$

Ensuring a nonnegative estimate of σ_ε^2, we combine Corollary 3 and Theorem 4 to obtain, from the preliminary estimates

$\widetilde{\sigma}_{\boldsymbol{\varepsilon}}^2$ and $\widehat{\mathbf{D}}$ of $\sigma_{\boldsymbol{\varepsilon}}^2$ and $\mathbf{D}$, the following hopefully improved estimate of $\sigma_{\boldsymbol{\varepsilon}}^2$, with $\widehat{\boldsymbol{\beta}}$, $\widehat{\mathbf{u}}$ from Step 1:

$$\widehat{\sigma}_{\boldsymbol{\varepsilon}}^2 = \frac{\|\mathbf{y} - \mathbf{X}\widehat{\boldsymbol{\beta}} - \mathbf{Z}\widehat{\mathbf{u}}\|^2 + \widetilde{\sigma}_{\boldsymbol{\varepsilon}}^2 \|\widehat{\mathbf{u}}\|^2_{\widehat{\mathbf{G}}^{-1}}}{n-p}. \tag{4.7}$$

4.2 Targeted estimates for $\boldsymbol{\beta}$, $\sigma_{\boldsymbol{\varepsilon}}^2$, $\mathbf{D}$ and predictions of U

4.2.1 The estimating equations.

The construction of Steps 1-2-3 in Version 1 above suggests to seek, given the data vector $\mathbf{Y} = \mathbf{y}$, respective estimates $\widehat{\boldsymbol{\beta}}$, $\widehat{\sigma}_{\boldsymbol{\varepsilon}}^2$, $\widehat{\mathbf{D}}$, of parameters $\boldsymbol{\beta}$, $\sigma_{\boldsymbol{\varepsilon}}^2$, $\mathbf{D}$, and prediction $\widehat{\mathbf{u}} = (\widehat{\mathbf{u}}_1^T, \cdots, \widehat{\mathbf{u}}_m^T)^T$ of the random effects vector U satisfying the system of equations:

$$\begin{pmatrix} \mathbf{X}^T\mathbf{X} & \mathbf{X}^T\mathbf{Z} \\ \mathbf{Z}^T\mathbf{X} & \mathbf{Z}^T\mathbf{Z} + \widehat{\sigma}_{\boldsymbol{\varepsilon}}^2 \widehat{\mathbf{G}}^{-1} \end{pmatrix} \begin{pmatrix} \widehat{\boldsymbol{\beta}} \\ \widehat{\mathbf{u}} \end{pmatrix} = \begin{pmatrix} \mathbf{X}^T\mathbf{y} \\ \mathbf{Z}^T\mathbf{y} \end{pmatrix}, \tag{4.8a}$$

$$\widehat{\sigma}_{\boldsymbol{\varepsilon}}^2 = \frac{\mathbf{y}^T(\mathbf{y} - \mathbf{X}\widehat{\boldsymbol{\beta}} - \mathbf{Z}\widehat{\mathbf{u}})}{n-p}, \tag{4.8b}$$

$$\widehat{\mathbf{D}} \text{ as in (4.4), with } \widehat{\mathbf{W}}^* \text{ as in (4.5).} \tag{4.8c}$$

From Steps 1-2-3 of Version 2, we instead get the following system of equations:

$$\widehat{\mathbf{u}} = \widehat{\sigma}_{\boldsymbol{\varepsilon}}^{-2} \widehat{\mathbf{W}}^* \mathbf{Z}^T M \mathbf{y}, \tag{4.9a}$$

$$\widehat{\boldsymbol{\beta}} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T(\mathbf{y} - \mathbf{Z}\widehat{\mathbf{u}}), \tag{4.9b}$$

$$\widehat{\sigma}_{\boldsymbol{\varepsilon}}^2 = \frac{\|\mathbf{y} - \mathbf{X}\widehat{\boldsymbol{\beta}} - \mathbf{Z}\widehat{\mathbf{u}}\|^2 + \widehat{\sigma}_{\boldsymbol{\varepsilon}}^2 \|\widehat{\mathbf{u}}\|^2_{\widehat{\mathbf{G}}^{-1}}}{n-p}, \tag{4.9c}$$

$$\widehat{\mathbf{D}} \text{ as in (4.4), with } \widehat{\mathbf{W}}^* \text{ as in (4.5).} \tag{4.9d}$$

From Theorem 2, Corollary 3 and Proposition 10, if $\widehat{\mathbf{D}}$ is SPD, $\widehat{\sigma}_{\boldsymbol{\varepsilon}}^2 > 0$ and Assumption $\mathcal{A}2_g$ holds, then one has the following logical equivalence:

$$\text{(4.8a)} - \text{(4.8b)} \iff \text{(4.9a)} - \text{(4.9b)} - \text{(4.9c)}.$$

Hence the above two systems of equations are equivalent, and, so, have the same solutions $\widehat{\boldsymbol{\beta}}$, $\widehat{\sigma}_{\boldsymbol{\varepsilon}}^2$, $\widehat{\mathbf{u}}$, $\widehat{\mathbf{D}}$, if any. However, they suggest two different successive approximations algorithms to try to calculate the 4 unknowns $\widehat{\boldsymbol{\beta}}$, $\widehat{\sigma}_{\boldsymbol{\varepsilon}}^2$, $\widehat{\mathbf{u}}$, $\widehat{\mathbf{D}}$.

4.2.2 3*S*-A1.RE: *a nonparametric REML-like fitting methodology for 2-level LMMs?*

Given the two systems of EEs derived above, our estimation approach here for the 2-level LMM (2.2) under Assumptions $\mathcal{A}1$-$\mathcal{A}4$ and $\mathcal{A}2_g$, $\mathcal{A}3_g$ proceeds by solving either of them for the 4 unknowns $\widehat{\boldsymbol{\beta}}$, $\widehat{\sigma}_{\boldsymbol{\varepsilon}}^2$, $\widehat{\mathbf{u}}$, $\widehat{\mathbf{D}}$, using a successive approximations iterative procedure. The approach is code named 3*S*-A1.RE, the "RE" part standing for "REML without ML".

But, at first sight, it is not obvious that this has any relationship with the traditional REML method for Gaussian LMMs. However, there is a linkage. Indeed, comparing with how the 3*S*-A1 approach was designed in [20], we deduce, from Theorems 7 and 8, that 3*S*-A1.RE is essentially equivalent to applying 3*S*-A1 to the transformed LMM (2.10a) from which the error contrasts construction of the Gaussian REML method is built. Furthermore, we will state a theorem below (Theorem 14 in Section 4.4) showing that when the 2-level LMM (2.2) is Gaussian with Assumption $\mathcal{A}3_g$ (*i.e.*, in this case, *i.i.d.* Gaussian errors), any of our 2 systems of EEs is equivalent to the system of REML equations.

4.2.3 Two new iterative estimating procedures for 2-level LMMs with u.ho. errors

With the response vector $\mathbf{y} = (\mathbf{y}_1^T, \cdots, \mathbf{y}_m^T)^T \in \mathbb{R}^n$ as data, and the system of EEs (4.8a)-(4.8c), our first nonparametric REML-like iterative estimating algorithm for a 2-level LMM with u.ho. errors is:

Algorithm 3*S*-A1.RE-V1. *: Estimation of $\boldsymbol{\beta}$, $\sigma_{\boldsymbol{\varepsilon}}^2$, $\mathbf{D}$ and prediction of $U_1, \cdots, U_m$ in a 2-level LMM with u.ho. errors:* **Version 1**

1. ***Initialization****: Same as in Algorithm 3S-A1-V1 of [20].*
2. ***The iterative process****: Given $\widehat{\boldsymbol{\beta}}^{(t)}$, $\widehat{\sigma}_{\boldsymbol{\varepsilon}}^{2(t)}$, $\widehat{\mathbf{u}}^{(t)} = (\widehat{\mathbf{u}}_1^{(t)T}, \cdots, \widehat{\mathbf{u}}_m^{(t)T})^T$ and $\widehat{\mathbf{D}}^{(t)}$ from iteration t, compute estimates at iteration $t+1$ as follows:*
   - 2.1. *Solve for $\widehat{\boldsymbol{\beta}}^{(t+1)}$ and $\widehat{\mathbf{u}}^{(t+1)}$ in the linear system, with $\widehat{\mathbf{G}}^{(t)} = \mathrm{diag}(\widehat{\mathbf{D}}^{(t)}, \cdots, \widehat{\mathbf{D}}^{(t)})$:*

$$\begin{pmatrix} \mathbf{X}^T\mathbf{X} & \mathbf{X}^T\mathbf{Z} \\ \mathbf{Z}^T\mathbf{X} & \mathbf{Z}^T\mathbf{Z} + \widehat{\sigma}_{\boldsymbol{\varepsilon}}^{2(t)} \widehat{\mathbf{G}}^{(t)^{-1}} \end{pmatrix} \begin{pmatrix} \widehat{\boldsymbol{\beta}}^{(t+1)} \\ \widehat{\mathbf{u}}^{(t+1)} \end{pmatrix} = \begin{pmatrix} \mathbf{X}^T\mathbf{y} \\ \mathbf{Z}^T\mathbf{y} \end{pmatrix}; \tag{4.10}$$

   - 2.2. $\widehat{\sigma}_{\boldsymbol{\varepsilon}}^{2(t+1)} = \mathbf{y}^T(\mathbf{y} - \mathbf{X}\widehat{\boldsymbol{\beta}}^{(t+1)} - \mathbf{Z}\widehat{\mathbf{u}}^{(t+1)})/(n-p)$;
   - 2.3. $\widehat{\mathbf{W}}^{*(t+1)} = \widehat{\sigma}_{\boldsymbol{\varepsilon}}^{2(t+1)}(\mathbf{Z}^T M \mathbf{Z} + \widehat{\sigma}_{\boldsymbol{\varepsilon}}^{2(t+1)} \widehat{\mathbf{G}}^{(t)^{-1}})^{-1}$;
   - 2.4. $\widehat{\mathbf{D}}^{(t+1)} = \frac{1}{m}\sum_{j=1}^m \left[\widehat{\mathbf{W}}_{jj}^{*(t+1)} + \widehat{\mathbf{u}}_j^{(t+1)}(\widehat{\mathbf{u}}_j^{(t+1)})^T\right]$;
3. ***Stopping criterion****: Same as in Algorithm 3S-A1-V1;*
4. ***Extracting estimates****: $\widehat{\boldsymbol{\beta}} = \widehat{\boldsymbol{\beta}}^{(t+1)}$, $\widehat{\sigma}_{\boldsymbol{\varepsilon}}^2 = \widehat{\sigma}_{\boldsymbol{\varepsilon}}^{2(t+1)}$, $\widehat{\mathbf{D}} = \widehat{\mathbf{D}}^{(t+1)}$, $\widehat{\mathbf{u}} = \widehat{\mathbf{u}}^{(t+1)}$.*

Based on the second system of estimating equations (4.9a)-(4.9d), our second nonparametric REML-like iterative estimating algorithm for 2-level LMMs with u.ho. errors proceeds as follows:

Algorithm 3*S*-A1.RE-V2. *: Estimation of $\boldsymbol{\beta}$, $\sigma_{\boldsymbol{\varepsilon}}^2$, $\mathbf{D}$ and prediction of $U_1, \cdots, U_m$ in a 2-level LMM with u.ho. errors:* **Version 2**

1. ***Initialization****: Same as in Algorithm 3S-A1-V1;*
2. ***The iterative process****: Given $\widehat{\sigma}_{\boldsymbol{\varepsilon}}^{2(t)}$, $\widehat{\mathbf{u}}_1^{(t)}, \cdots, \widehat{\mathbf{u}}_m^{(t)}$, hence $\widehat{\mathbf{u}}^{(t)} = (\widehat{\mathbf{u}}_1^{(t)T}, \cdots, \widehat{\mathbf{u}}_m^{(t)T})^T$, and $\widehat{\mathbf{D}}^{(t)}$ from iteration t, we obtain estimates at iteration $t+1$ as follows:*
   - 2.1. $\widehat{\boldsymbol{\beta}}^{(t+1)} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T(\mathbf{y} - \mathbf{Z}\widehat{\mathbf{u}}^{(t)})$;
   - 2.2. $\widehat{\sigma}_{\boldsymbol{\varepsilon}}^{2(t+1)} = \frac{\|\mathbf{y} - \mathbf{Z}\widehat{\mathbf{u}}^{(t)} - \mathbf{X}\widehat{\boldsymbol{\beta}}^{(t+1)}\|^2 + \widehat{\sigma}_{\boldsymbol{\varepsilon}}^{2(t)} \|\widehat{\mathbf{u}}^{(t)}\|^2_{\widehat{\mathbf{G}}^{(t)-1}}}{n-p}$, *with* $\widehat{\mathbf{G}}^{(t)} = \mathrm{diag}(\widehat{\mathbf{D}}^{(t)}, \cdots, \widehat{\mathbf{D}}^{(t)})$;

2.3. $A^{(t+1)} = (\mathbf{Z}^T M \mathbf{Z} + \widehat{\sigma}_{\boldsymbol{\varepsilon}}^{2(t+1)} \widehat{\mathbf{G}}^{(t)^{-1}})^{-1}$;

2.4. $\widehat{\mathbf{u}}^{(t+1)} = A^{(t+1)} \mathbf{Z}^T M \mathbf{y}$ *and* $\widehat{\mathbf{W}}^{*(t+1)} = \widehat{\sigma}_{\boldsymbol{\varepsilon}}^{2(t+1)} A^{(t+1)}$;

2.5. $\widehat{\mathbf{D}}^{(t+1)} = \frac{1}{m} \sum_{j=1}^{m} \big[\widehat{\mathbf{W}}_{jj}^{*(t+1)} + \widehat{\mathbf{u}}_j^{(t+1)} (\widehat{\mathbf{u}}_j^{(t+1)})^T \big]$;

3. ***Stopping criterion****: As in Algorithm 3S-A1-V1.*

4. ***Extracting estimates****:* $\widehat{\boldsymbol{\beta}} = \widehat{\boldsymbol{\beta}}^{(t+1)}$, $\widehat{\sigma}_{\boldsymbol{\varepsilon}}^2 = \widehat{\sigma}_{\boldsymbol{\varepsilon}}^{2(t+1)}$, $\widehat{\mathbf{D}} = \widehat{\mathbf{D}}^{(t+1)}$, $\widehat{\mathbf{u}} = \widehat{\mathbf{u}}^{(t+1)}$.

4.2.4 The issue with using the $n \times n$ matrix M in the 3*S*-A1.RE fitting of LMMs

The practical implementation details discussed at length in [20] are also valid for both 3*S*-A1.RE algorithms presented above. But an apparent additional hurdle in 3*S*-A1.RE is the seemingly need to compute, beforehand, the dense $n \times n$ matrix M given by (3.6). The reason is that it seems a prerequisite to calculate the product matrix $\mathbf{Z}^T M \mathbf{Z}$ and vector $\mathbf{Z}^T M \mathbf{y}$. Now, if the sample size n is big, then M can be really huge, both for storage and computational demands. But, as a matter of fact, the two products $\mathbf{Z}^T M \mathbf{Z}$ and $\mathbf{Z}^T M \mathbf{y}$ can be calculated without first computing M, or manipulating any $n \times n$ matrix. We see this from:

$$\mathbf{Z}^T M \mathbf{Z} = \mathbf{Z}^T \mathbf{Z} - \mathbf{Z}^T \mathbf{X} (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{Z} = \mathbf{Z}^T \mathbf{Z} - A_{21} A_{11}^{-1} A_{12},$$
$$\mathbf{Z}^T M \mathbf{y} = \mathbf{Z}^T \mathbf{y} - \mathbf{Z}^T \mathbf{X} (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{y} = Y_2 - A_{21} A_{11}^{-1} Y_1,$$

where $A_{11} = \mathbf{X}^T \mathbf{X} \in \mathcal{M}_p(\mathbb{R})$, $A_{12} = \mathbf{X}^T \mathbf{Z} \in \mathcal{M}_{p,q}(\mathbb{R})$, $A_{21} = \mathbf{Z}^T \mathbf{X} = A_{12}^T$, $Y_1 = \mathbf{X}^T \mathbf{y} \in \mathbb{R}^p$, $Y_2 = \mathbf{Z}^T \mathbf{y} \in \mathbb{R}^q$ are all blocks in the coefficient matrix and *r.h.s.* partitions of the HMMEs (4.10) to solve during the iterations of Algorithm 3*S*-A1.RE-V1. Moreover, we stress that the product matrices $\mathbf{Z}^T M \mathbf{Z}$ and $\mathbf{Z}^T M \mathbf{y}$ are iteration-independent, so they can be computed before any iteration starts.

4.3 First properties of the estimates $\widehat{\boldsymbol{\beta}}$, $\widehat{\sigma}_{\boldsymbol{\varepsilon}}^2$, $\widehat{\mathbf{D}}$ and $\widehat{\mathbf{u}}$

4.3.1 3*S*-A1.RE fitting: *Relationship with the HMMEs*

As for the 3*S*-A1 iterative algorithms, using Theorem 1, Theorem 2, Corollary 3 and Proposition 10, we get the following first key observation about $\widehat{\boldsymbol{\beta}}$, $\widehat{\sigma}_{\boldsymbol{\varepsilon}}^2$, $\widehat{\mathbf{D}}$, $\widehat{\mathbf{u}}$ obtained at convergence (if any) of each of the two 3*S*-A1.RE algorithms designed above:

Theorem 13. *If one takes* $\mathbf{G} = \mathrm{diag}(\widehat{\mathbf{D}}, \cdots, \widehat{\mathbf{D}})$ *(m times) and* $\sigma_{\boldsymbol{\varepsilon}}^2 = \widehat{\sigma}_{\boldsymbol{\varepsilon}}^2$ *in the HMMEs* (3.1)*, then* $\widetilde{\boldsymbol{\beta}} = \widehat{\boldsymbol{\beta}}$ *and* $\widetilde{U} = \widehat{\mathbf{u}}$ *are the solutions to these HMMEs. Therefore, the estimates* $\widehat{\boldsymbol{\beta}}$, $\widehat{\sigma}_{\boldsymbol{\varepsilon}}^2$, $\widehat{\mathbf{D}}$ *and prediction* $\widehat{\mathbf{u}}$ *satisfy, with* $\widehat{\mathbf{V}} = \mathbf{Z} \widehat{\mathbf{G}} \mathbf{Z}^T + \widehat{\sigma}_{\boldsymbol{\varepsilon}}^2 \mathbf{I}_n$*:*

$$\widehat{\boldsymbol{\beta}} = \mathrm{EBLUE}(\boldsymbol{\beta} | \, \mathbf{y}, \widehat{\mathbf{G}}, \widehat{\sigma}_{\boldsymbol{\varepsilon}}^2) = (\mathbf{X}^T \widehat{\mathbf{V}}^{-1} \mathbf{X})^{-1} \mathbf{X}^T \widehat{\mathbf{V}}^{-1} \mathbf{y}, \tag{4.11}$$

$$\widehat{\mathbf{u}} = \mathrm{EBLUP}(U | \, \mathbf{y}, \widehat{\mathbf{G}}, \widehat{\sigma}_{\boldsymbol{\varepsilon}}^2) = \widehat{\mathbf{G}} \mathbf{Z}^T \widehat{\mathbf{V}}^{-1} (\mathbf{y} - \mathbf{X} \widehat{\boldsymbol{\beta}}), \tag{4.12}$$

That is $\widehat{\boldsymbol{\beta}}$ *and* $\widehat{\mathbf{u}}$ *are, respectively, an empirical BLUE of* $\boldsymbol{\beta}$ *and an empirical BLUP of* U *when* $\mathbf{G}$ *is estimated by* $\widehat{\mathbf{G}}$ *and* $\sigma_{\boldsymbol{\varepsilon}}^2$ *by* $\widehat{\sigma}_{\boldsymbol{\varepsilon}}^2$.

4.4 3*S*-A1.RE fitting: *Relationship with Gaussian REML with* i.i.d. *errors*

Consider a 2-level LMM satisfying Assumptions $\mathcal{A}1$-$\mathcal{A}4$ and the Gaussian assumptions that:

Assumption $\mathcal{A}5$. $U_j \overset{\mathcal{D}}{\sim} \mathcal{N}_r(\mathbf{0}, \mathbf{D})$ *and* $\boldsymbol{\varepsilon}_j \overset{\mathcal{D}}{\sim} \mathcal{N}_{n_j}(\mathbf{0}, \sigma_{\boldsymbol{\varepsilon}}^2 \mathbf{I}_{n_j})$, $j = 1, \cdots, m$.

As stated below, each of the two sets of EEs in this 3*S*-A1.RE is equivalent to the system of restricted likelihood equations (see Proof in the Appendix):

Theorem 14. *Let* $\widetilde{\boldsymbol{\beta}}$ *and* $\widetilde{U}$ *satisfy the HMMEs* (3.1) *given* $\sigma_{\boldsymbol{\varepsilon}}^2, \mathbf{D}$ *for the 2-level LMM* (2.2)*, with Assumptions $\mathcal{A}1$-$\mathcal{A}4$ and $\mathcal{A}5$, and* $\ell_{\mathrm{R}}(\sigma_{\boldsymbol{\varepsilon}}^2, \mathbf{D} \,|\, \mathbf{y})$ *be the restricted log-likelihood of the LMM* (2.1) *given by* (2.11)*. One has:*

$$\frac{\partial}{\partial \mathbf{D}} \ell_{\mathrm{R}}(\sigma_{\boldsymbol{\varepsilon}}^2, \mathbf{D} \,|\, \mathbf{y}) = \mathbf{0} \iff \mathbf{D} = \frac{1}{m} \sum_{j=1}^{m} (\mathbf{W}_{jj}^* + \widetilde{U}_j \widetilde{U}_j^T). \tag{4.13}$$

Moreover, if the left hand side of (4.13) *holds, then:*

$$\frac{\partial}{\partial \sigma_{\boldsymbol{\varepsilon}}^2} \ell_{\mathrm{R}}(\sigma_{\boldsymbol{\varepsilon}}^2, \mathbf{D} \,|\, \mathbf{y}) = 0 \iff \sigma_{\boldsymbol{\varepsilon}}^2 = \frac{\mathbf{y}^T (\mathbf{y} - \mathbf{X} \widetilde{\boldsymbol{\beta}} - \mathbf{Z} \widetilde{U})}{n - p}. \tag{4.14}$$

5 3*S*-A1.RE iterative fitting of LMMs: *beyond the 2-level ones*

As for the 3*S*-A1 approach before, one observes that in both versions of 3*S*-A1.RE constructed in Section 4, the 2-level aspect of the LMM intervenes only in Step 3, when deriving an EE for the random effects covariance matrix. Hence the approach can be adapted to fit any LMM with u.ho. errors, provided one can devise a reliable EE for its random effects covariance matrix $\mathbf{G}$. We examine 2 such situations hereafter and suggest a somewhat tantalizing prospect.

5.1 3*S*-A1.RE fitting of 2-level LMMs with u.ho. errors and cluster random effects diagonal covariance matrix D, and ANOVA LMMs

Proceeding as it was done in the 3*S*-A1 approach, one can easily adapt Algorithms 3*S*-A1.RE-V1 and 3*S*-A1.RE-V2 for the 2-level LMM with u.ho. errors and a cluster random effects diagonal covariance matrix $\mathbf{D}$, to get, say, Algorithms 3*S*-A1.RE-V1-diag and 3*S*-A1.RE-V2-diag. The same is true for fitting an ANOVA LMM to get, say, Algorithms 3*S*-A1.RE-V1-ANOVA and 3*S*-A1.RE-V2-ANOVA. In both cases, the two main changes *w.r.t.* to the corresponding 3*S*-A1 algorithm are to replace the estimate of the matrix $\mathbf{V}^*$ by that of $\mathbf{W}^*$, and the computing formulas of the BLP by those of the BLUP.

Remark 5.1. *Algorithm 3S-A1.RE-V1-ANOVA is the same as one of those derived by Henderson for Gaussian ANOVA LMMs [21, 14].*

5.2 Estimation in an LMM with u.ho. errors: *the one cluster case?*

It is obvious that the 2-level LMM (2.2) is a particular instance of the arbitrary LMM (2.1). But, though less apparent, the converse is also true. Indeed, (2.1) is the particular case of (2.2) where $m = 1$, *i.e.* there is just one (big) cluster in the population. In particular then, $\mathbf{G} = \mathbf{D}$. This implies that any estimation method developed for 2-level LMMs can, at least theoretically, be applied to any LMM, as long as no sample variation between clusters is explicitly used during the estimation process. Our two basic 3*S*-A1.RE algorithms designed in Section 4 satisfy that latter requirement. This was also true about the 3*S*-A1 algorithms.

Hence, on a theoretical basis, we could take any of those algorithms to estimate the vector $\boldsymbol{\beta} \in \mathbb{R}^p$, the variance $\sigma^2_{\boldsymbol{\varepsilon}} > 0$, the random effects covariance matrix $\mathbf{G} \in \mathcal{M}_q(\mathbb{R})$ and predict the random effects vector U in the general LMM (2.1) with u.ho. errors. The only change is that the two respective estimating equations for $\mathbf{G}$ then become:

$$\widehat{\mathbf{G}} = \widehat{\mathbf{V}}^* + \widehat{\mathbf{u}}\widehat{\mathbf{u}}^T, \quad \text{with } \widehat{\mathbf{V}}^* = \widehat{\sigma}^2_{\boldsymbol{\varepsilon}}(\mathbf{Z}^T\mathbf{Z} + \widehat{\sigma}^2_{\boldsymbol{\varepsilon}}\widehat{\mathbf{G}}^{-1})^{-1}, \tag{5.1a}$$

$$\text{or } \widehat{\mathbf{G}} = \widehat{\mathbf{W}}^* + \widehat{\mathbf{u}}\widehat{\mathbf{u}}^T, \quad \text{with } \widehat{\mathbf{W}}^* = \widehat{\sigma}^2_{\boldsymbol{\varepsilon}}(\mathbf{Z}^T M\mathbf{Z} + \widehat{\sigma}^2_{\boldsymbol{\varepsilon}}\widehat{\mathbf{G}}^{-1})^{-1}, \tag{5.1b}$$

the first form for the 3*S*-A1 algorithms, and the second for the 3*S*-A1.RE ones. The main technical difficulty in acting so then stems from the following question, based on comparing these expressions of $\widehat{\mathbf{G}}$ to that of $\widehat{\mathbf{D}}$ in (4.4) (or the analogous one in the 3*S*-A1 construction):

> *Since no averaging is involved in its estimation through* (5.1a) *or* (5.1b)*, how precise can one expect to estimate the matrix* $\mathbf{G}$ *when using Algorithms 3S-A1-V1, 3S-A1-V2, 3S-A1.RE-V1 or 3S-A1.RE-V2 in that way to fit an arbitrary LMM with u.ho. errors?*

But before that, even more basically: *can any convergence of the algorithms even be possible in that scenario?* These are lingering issues that still require investigation, both theoretical and empirical.

6 Numerical examples

We use simulated data sets and both the classical `cake` [31] and `Blackmore` (from the `car` R package) data sets to illustrate one of the iterative procedures for LMM fitting, Algorithm 3*S*-A1.RE-V1, presented in Section 4.2.3 (no results presented for Algorithm 3*S*-A1.RE-V2 because they are almost identical). In all situations, we compared Algorithm 3*S*-A1.RE-V1 vs. Gaussian REML implemented by the `lmer` function in the reference `lme4` package [32] of the R software. Hereafter, we denote the latter by `lmer`-REML. We do not here recall the set up and the variables in the two data sets since they are the same as in [20].

6.1 A simulation study

To investigate the performance of Algorithm 3*S*-A1.RE-V1 vs. `lmer`-REML, we fitted both to 500 simulated data sets, each of size $n = 200$, under the same simulation settings as in [20]. So, we considered three options for the distribution of the residual errors: a Gaussian, a mixture of 4 Gaussians with given mixing proportions, and a discrete distribution with 4 mass points. Each of these options was associated to each of the following five options for the distribution of the U_j's in $\mathbb{R}^3$: Gaussian with $\mathbf{D}$ SPD, Gaussian with $\mathrm{rank}(\mathbf{D}) = 1$, Gaussian with $\mathrm{rank}(\mathbf{D}) = 2$, mixture of 3 Gaussians with given mixing proportions, and a discrete distribution with 3 mass points. Thus, we have $3 \times 5 = 15$ possible scenarios.

The simulation results from both Algorithm 3*S*-A1.RE-V1 and Gaussian REML given in Table 1 are comparable as expected. Nevertheless, Gaussian REML dominated Algorithm 3*S*-A1.RE-V1 only in 5 out of the 15 scenarios with respect to the total number of lower bias and MSE/MSPE in the estimation of $\boldsymbol{\beta}$, $\mathbf{D}$, $\sigma^2_{\boldsymbol{\varepsilon}}$ and prediction of random effects and responses. The 5 scenarios are: Gaussian errors and respectively Gaussian random effects with $\mathbf{D}$ SPD (Gaussian REML = 7/10), Gaussian random effects with $\mathrm{rank}(\mathbf{D}) = 2$ (Gaussian REML = 6/10) and Gaussian mixture random effects (Gaussian REML = 7/10), and Gaussian mixture errors and respectively Gaussian random effects with $\mathbf{D}$ SPD (Gaussian REML = 6/10) and Gaussian mixture random effects (Gaussian REML = 6/10). Gaussian REML did not dominate Algorithm 3*S*-A1.RE-V1 in all scenarios involving a discrete distribution for the residual errors.

6.2 Application to the `cake` data

We modeled the response variable `angle` by fitting to the `cake` data the two LMMs, in the R software notations [22]:

```
angle ~ fe(-1+temp) + re(1)
        + Gr(recipe:replicate)        (6.1a)
angle ~ fe(-1+temperature) + re(1)
        + Gr(recipe:replicate)        (6.1b)
```

The results of Algorithm 3*S*-A1.RE-V1 for models (6.1a) and (6.1b) are given in Tables 2 and 3, respectively, with Figure 1 showing convergence of their respective parameters iterates. When compared to the results of Algorithm 3*S*-A1-V1 in [20], the three main observations are that in both models, and as would be expected, fixed effects estimates are almost unchanged, the residual variance estimates marginally decreased, whereas the estimates of variance of the lone random effects, the *random intercept*, increased somewhat more substantially. As a consequence of the latter fact, the random effects ratio estimates also increased. Their p–values are less than $5 \cdot 10^{-4}$, confirming the presence of random effects on the intercept *w.r.t.* the `recipe:replicate` clustering considered for the baked cakes. Figure 1 shows a smooth convergence for all parameters in model (6.1a), and an almost constant sequence of fixed effects parameters iterates in model (6.1b) for Algorithm 3*S*-A1.RE-V1, with each algorithm converging after 10 iterations.

For comparison, given Theorem 14, we also fitted each of the two LMMs (6.1a) and (6.1b) by Gaussian REML. Table 4 presents the results and, as expected, the parameters estimates are almost identical to those in the two previous tables.

Table 1. Estimates of the bias and MSE of estimates of parameters, predicted random effects and responses in the same simulation model as in [20] applied to simulated data sets with different distributions for the random effects and residual errors.

Distribution of random effects and errors	$b_R(\widehat{\boldsymbol{\beta}})$	$\sqrt{\mathrm{RMSE}(\widehat{\boldsymbol{\beta}})}$	$b_R(\widehat{\sigma}^2_{\boldsymbol{\varepsilon}})$	$\sqrt{\mathrm{RMSE}(\widehat{\sigma}^2_{\boldsymbol{\varepsilon}})}$	$b_R(\widehat{\mathbf{D}})$	$\sqrt{\mathrm{RMSE}(\widehat{\mathbf{D}})}$	$b(\widehat{U})$	$\sqrt{\mathrm{MSE}(\widehat{U})}$	$b(\widehat{Y})$	$\sqrt{\mathrm{MSE}(\widehat{Y})}$
Gaussian errors										
Gaussian RE ($\mathbf{D}$ *SPD*)	0.00353	0.0688	0.00351	0.112	0.531	0.958	0.0165	0.661	$3.80e-05$	$5.96e-03$
	(0.00228)	*(0.0663)*	*(0.000530)*	*(0.110)*	*(0.532)*	*(0.937)*	*(0.00495)*	*(0.674)*	*(1.14e − 04)*	*(5.88e − 03)*
Gaussian RE ($\mathrm{R}(\mathbf{D})=1$)	0.00332	0.0542	0.0222	0.104	0.878	1.02	0.00762	0.365	$4.56e-06$	$5.17e-03$
	(0.00295)	*(0.0559)*	*(0.0326)*	*(0.106)*	*(0.851)*	*(1.91)*	*(0.00497)*	*(0.387)*	*(7.81e − 05)*	*(5.19e − 03)*
Gaussian RE ($\mathrm{R}(\mathbf{D})=2$)	0.00452	0.0626	0.0182	0.108	0.740	1.26	0.00338	0.535	$4.20e-05$	$5.76e-03$
	(0.00164)	*(0.0630)*	*(0.0207)*	*(0.106)*	*(0.554)*	*(1.21)*	*(0.00740)*	*(0.546)*	*(3.03e − 05)*	*(5.72e − 03)*
Gaussian mixture RE	0.00220	0.0676	0.00658	0.109	0.428	1.16	0.0116	0.684	$1.45e-04$	$4.43e-03$
	(0.00183)	*(0.0696)*	*(0.000862)*	*(0.116)*	*(0.364)*	*(0.820)*	*(0.00940)*	*(0.681)*	*(1.27e − 05)*	*(4.46e − 03)*
Discrete RE	0.00224	0.0607	0.0254	0.106	0.823	1.84	0.00414	0.513	$2.91e-05$	$3.57e-03$
	(0.00271)	*(0.0613)*	*(0.0281)*	*(0.106)*	*(0.596)*	*(1.01)*	*(0.00859)*	*(0.522)*	*(6.54e − 05)*	*(3.62e − 03)*
Gaussian mix. errors										
Gaussian RE ($\mathbf{D}$ *SPD*)	0.00111	0.0641	0.365	0.406	0.724	1.44	0.00813	0.597	$9.23e-05$	$5.84e-03$
	(0.00136)	*(0.0629)*	*(0.396)*	*(0.437)*	*(0.686)*	*(1.09)*	*(0.00461)*	*(0.596)*	*(1.20e − 04)*	*(5.80e − 03)*
Gaussian RE ($\mathrm{R}(\mathbf{D})=1$)	0.00154	0.0617	0.408	0.444	0.935	2.20	0.00365	0.440	$2.21e-05$	$5.42e-03$
	(0.00253)	*(0.0615)*	*(0.457)*	*(0.485)*	*(4.50)*	*(10.4)*	*(0.00644)*	*(0.458)*	*(2.94e − 05)*	*(5.15e − 03)*
Gaussian RE ($\mathrm{R}(\mathbf{D})=2$)	0.00365	0.0625	0.404	0.442	0.670	0.991	0.00204	0.520	$2.80e-05$	$5.72e-03$
	(0.00367)	*(0.0644)*	*(0.426)*	*(0.459)*	*(0.803)*	*(1.41)*	*(0.00546)*	*(0.527)*	*(5.71e − 05)*	*(5.66e − 03)*
Gaussian mixture RE	0.00374	0.0626	0.395	0.433	0.511	0.899	0.00824	0.608	$1.13e-06$	$4.45e-03$
	(0.00403)	*(0.0618)*	*(0.389)*	*(0.432)*	*(0.620)*	*(0.773)*	*(0.00475)*	*(0.602)*	*(9.73e − 05)*	*(4.52e − 03)*
Discrete RE	0.00329	0.0637	0.386	0.426	0.728	1.18	0.00729	0.511	$1.86e-05$	$3.69e-03$
	(0.00278)	*(0.0624)*	*(0.437)*	*(0.466)*	*(0.873)*	*(0.936)*	*(0.00277)*	*(0.514)*	*(1.96e − 05)*	*(3.70e − 03)*
Discrete errors										
Gaussian RE ($\mathbf{D}$ *SPD*)	0.00134	0.0577	0.774	0.779	0.644	1.57	0.00339	0.507	$6.21e-05$	$5.99e-03$
	(0.00210)	*(0.0570)*	*(0.777)*	*(0.781)*	*(0.660)*	*(1.07)*	*(0.00217)*	*(0.511)*	*(3.62e − 05)*	*(5.75e − 03)*
Gaussian RE ($\mathrm{R}(\mathbf{D})=1$)	0.00130	0.0594	0.778	0.782	0.722	1.43	0.00369	0.471	$1.22e-05$	$5.20e-03$
	(0.00198)	*(0.0606)*	*(0.787)*	*(0.791)*	*(0.715)*	*(1.54)*	*(0.00344)*	*(0.466)*	*(3.52e − 05)*	*(5.19e − 03)*
Gaussian RE ($\mathrm{R}(\mathbf{D})=2$)	0.00218	0.0585	0.774	0.778	1.10	2.55	0.00723	0.486	$5.35e-07$	$5.73e-03$
	(0.00231)	*(0.0590)*	*(0.782)*	*(0.787)*	*(0.639)*	*(1.40)*	*(0.00444)*	*(0.492)*	*(7.70e − 06)*	*(5.71e − 03)*
Gaussian mixture RE	0.00174	0.0570	0.778	0.783	0.385	0.955	0.00119	0.500	$7.66e-05$	$4.53e-03$
	(0.00211)	*(0.0560)*	*(0.773)*	*(0.777)*	*(0.441)*	*(0.857)*	*(0.00638)*	*(0.503)*	*(1.24e − 05)*	*(4.55e − 03)*
Discrete RE	0.00191	0.0597	0.773	0.777	0.848	1.40	0.00777	0.484	$2.97e-05$	$3.69e-03$
	(0.00189)	*(0.0590)*	*(0.780)*	*(0.784)*	*(1.73)*	*(3.34)*	*(0.00359)*	*(0.489)*	*(2.03e − 05)*	*(3.75e − 03)*

- $\mathbf{D}$ is the cluster random effects (RE) 3×3 covariance matrix; $\mathrm{R}(\mathbf{D}) = \mathrm{rank}(\mathbf{D})$ is the rank of $\mathbf{D}$.
- In each cell of two numbers in the table, the one on top is the result from the Algorithm 3*S*-A1.RE-V1, while the one below it, in parentheses, is the result from Gaussian REML.

6.3 Application to the `Blackmore` longitudinal data

For the `Blackmore` data, we fit a LMM using, as fixed effects, an intercept, `age` minus 8 (denoted `age8`), `group` and the interaction of the two latter. With the clustering variable being `subject`, a random intercept and a random slope for `age8` are also included. The model to fit is therefore:

$$\log_2\texttt{exercise} \sim \texttt{fe(1 + age8 + group + age8*group)} + \texttt{re(1 + age8)} + \texttt{Gr(subject)} \quad (6.2)$$

The results using Algorithm 3*S*-A1.RE-V1 and REML are presented respectively in Tables 5 and 6, while convergence results for Algorithm 3*S*-A1.RE-V1 are given in Figure 2. Parameters estimates from Algorithm 3*S*-A1.RE-V1 are almost identical to those from REML. In both, the interaction of age with group is highly significant at a 5% level, unlike the fixed intercept and group effects. A similar trend was also observed between Algorithm 3*S*-A1-V1 and ML. Estimates from 3*S*-A1.RE-V1 compared to those from 3*S*-A1-V1 in [20] show that fixed effects estimates are almost unchanged and residual variance estimate marginally increased. Conversely, estimates of the random intercept variance $\sigma^2_{\texttt{intercept}}$ and the random slope variance $\sigma^2_{\texttt{age8}}$ decreased, while significantly greater than zero in both approaches. From 3*S*-A1.RE-V1 results, a significant negative correlation between those two random effects is also probable. The p–value for the random effects ratio estimate is again less than 0.001. A smooth convergence for all parameters is observed for Algorithm 3*S*-A1.RE-V1 and convergence was archived after 34 iterations.

7 Conclusion

Validating the use of the Gaussian REML equations beyond Gaussian LMMs has been an untold objective of many researchers in the field of LMM fitting. Till today, the output of that endeavor has been mixed, successful only either for the subclass of variance components models (or ANOVA LMMs) through the ANOVA (for balanced data) or I-MINQUE methods, or for a broader class of LMMs in the somewhat involved global framework of estimating functions. But the constant deficiency in all those approaches is that they provide, each, a set of estimating equations (EEs) for the LMM parameters not ensuring satisfaction of the required nonnegativity of variances

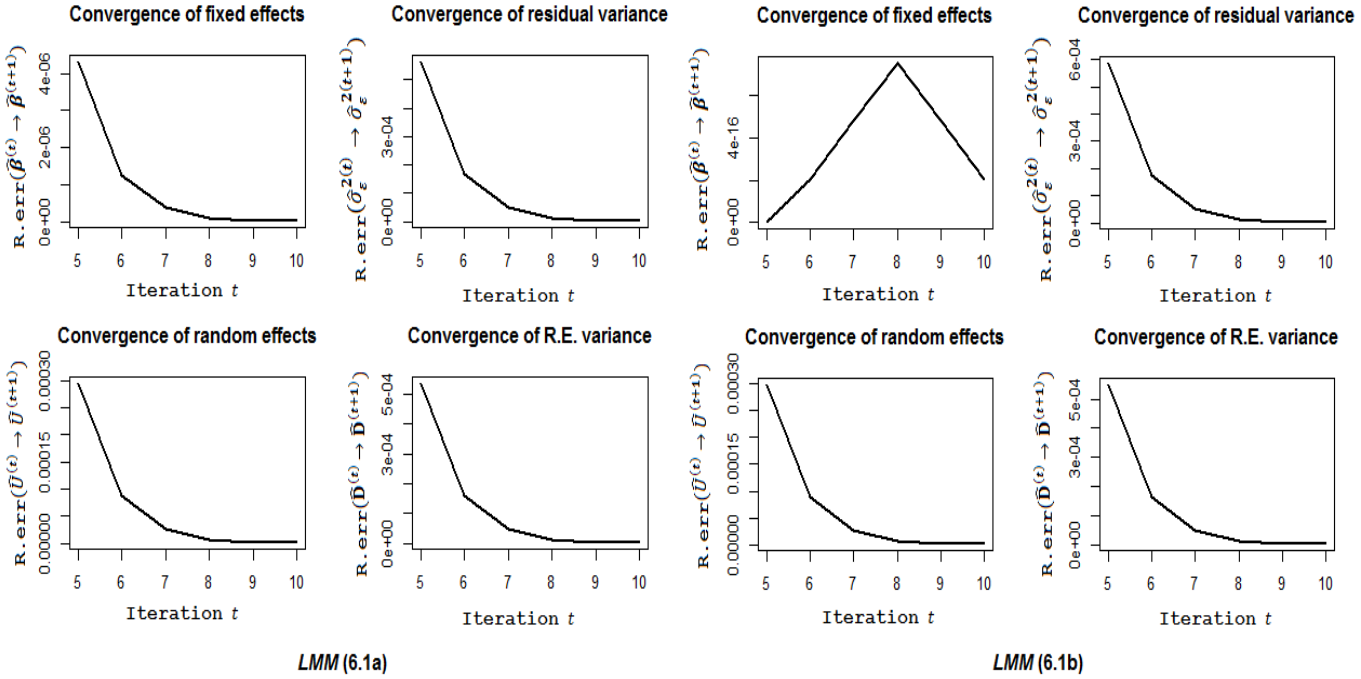

Figure 1. Convergence of parameters in 3*S*-A1.RE-V1 for the `cake` data models (6.1a) and (6.1b). The y-axis in each case represents the relative variation of the estimate from previous iteration. Particularly, $\text{R.err}(\widehat{\boldsymbol{\beta}}^{(t)} \to \widehat{\boldsymbol{\beta}}^{(t+1)}) = \|\widehat{\boldsymbol{\beta}}^{(t+1)} - \widehat{\boldsymbol{\beta}}^{(t)}\|/\|\widehat{\boldsymbol{\beta}}^{(t+1)}\|$, $\text{R.err}(\widehat{\sigma}_{\boldsymbol{\varepsilon}}^{2(t)} \to \widehat{\sigma}_{\boldsymbol{\varepsilon}}^{2(t+1)}) = |\widehat{\sigma}_{\boldsymbol{\varepsilon}}^{2(t+1)} - \widehat{\sigma}_{\boldsymbol{\varepsilon}}^{2(t)}|/\widehat{\sigma}_{\boldsymbol{\varepsilon}}^{2(t+1)}$, $\text{R.err}(\widehat{\mathbf{D}}^{(t)} \to \widehat{\mathbf{D}}^{(t+1)}) = \|\widehat{\mathbf{D}}^{(t+1)} - \widehat{\mathbf{D}}^{(t)}\|^{(M)}/\|\widehat{\mathbf{D}}^{(t+1)}\|^{(M)}$ with $\widehat{\mathbf{D}} = \widehat{\sigma}_r^2$, $\text{R.err}(\widehat{U}^{(t)} \to \widehat{U}^{(t+1)}) = \|\widehat{U}^{(t+1)} - \widehat{U}^{(t)}\|/\|\widehat{U}^{(t+1)}\|$.

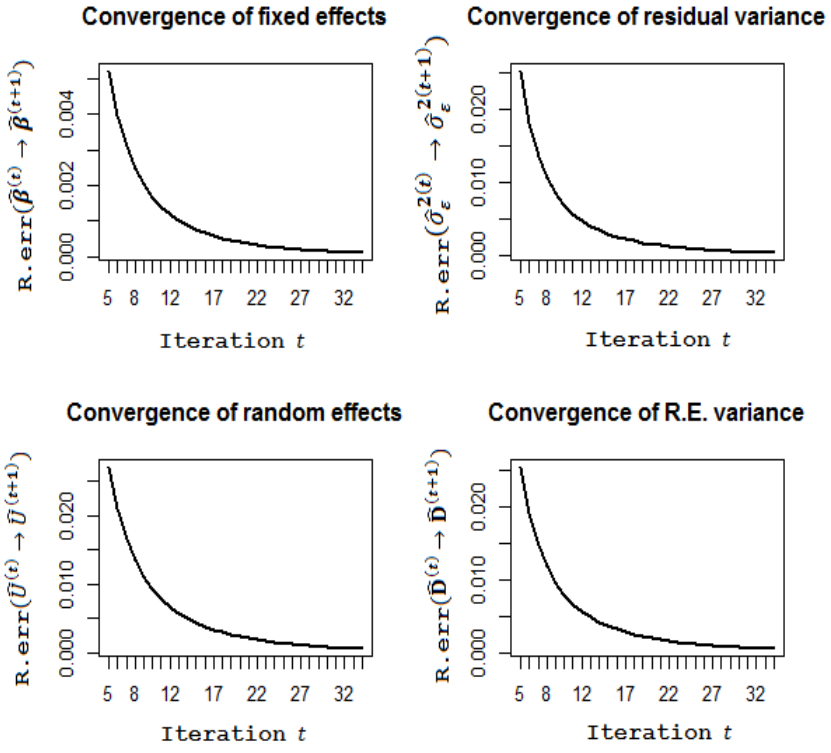

Figure 2. Convergence of parameters in 3*S*-A1.RE-V1 for the LMM (6.2) to the `Blackmore` data. The y-axis is as in Figure 1.

Table 2. Results of Algorithm 3*S*-A1.RE-V1 fitting the LMM (6.1a) to the `cake` data.

parameters	*est.*	$[2.5\% - 97.5\%]$	*bias*	*std.dev*	$\sqrt{\text{MSE}}$	*t-stat*	*p–value*
$\beta_{\texttt{temp}}$	0.1604	$[0.1519, 0.1706]$	0.0002298	0.004843	0.004848	33.08	$5.482e-240$
σ_1^2	40.18	$[21.81, 62.69]$	0.7765	10.25	10.28	3.910	$9.223e-05$
σ_ε^2	20.63	$[16.87, 24.43]$	-0.1885	1.929	1.938	10.64	$1.863e-26$
ρ	1.948	$[1.059, 3.135]$	0.07597	0.5480	0.5532	3.522	$4.285e-04$

- $\beta_{\texttt{temp}}$: coeff. of fixed effects `temp`; σ_1^2 : variance of the random intercept; σ_ε^2 : residual variance; ρ : random effects ratio;
- *est.* : estimate; $[2.5\% - 97.5\%]$, *bias*, *std.dev*, $\sqrt{\text{MSE}}$, *t-stat*, *p–value*: respectively estimated two-side 95% percentile confidence interval, bias, standard deviation, square root of the Mean Squared Error, t -statistic and p–value of the t -test for a zero value, all computed by a bootstrap simulation of 1000 replicates of the model fit using the nonparametric residuals bootstrap approach described in [33]. These variability estimates are computed from the 1000 bootstrap replicates using the standard formulas introduced by Efron [34].

Table 3. Results of Algorithm 3*S*-A1.RE-V1 fitting the LMM (6.1b) to the `cake` data.

parameters	*est.*	$[2.5\% - 97.5\%]$	*bias*	*std.dev*	$\sqrt{\text{MSE}}$	*t-stat*	*p–value*
$\beta_{\texttt{temperature.175}}$	27.98	$[25.87, 30.29]$	0.01378	1.150	1.150	24.33	$8.771e-131$
$\beta_{\texttt{temperature.185}}$	29.96	$[27.80, 32.38]$	-0.001025	1.185	1.185	25.29	$4.348e-141$
$\beta_{\texttt{temperature.195}}$	31.42	$[29.29, 33.94]$	0.03470	1.200	1.201	26.17	$5.962e-151$
$\beta_{\texttt{temperature.205}}$	32.18	$[29.89, 34.52]$	0.01014	1.203	1.203	26.75	$1.409e-157$
$\beta_{\texttt{temperature.215}}$	35.84	$[33.75, 38.43]$	0.04873	1.197	1.198	29.91	$1.504e-196$
$\beta_{\texttt{temperature.225}}$	35.36	$[33.16, 37.84]$	0.001267	1.170	1.170	30.22	$1.313e-200$
σ_1^2	40.29	$[21.87, 62.99]$	0.7400	10.26	10.29	3.917	$8.963e-05$
σ_ε^2	20.48	$[16.52, 24.43]$	-0.2039	2.022	2.032	10.08	$7.026e-24$
ρ	1.968	$[1.071, 3.218]$	0.07841	0.5574	0.5629	3.496	$4.730e-04$

- Meaning of parameters as in Table 2

estimates and positive semi-definiteness of covariance matrices ones. In this paper, we designed an estimation methodology for LMM fitting, code named 3*S*-A1.RE. The approach first applies to 2-level (or longitudinal) LMMs with the only added assumption (to the basic ones) that the residual errors are uncorrelated and homoscedastic. Each method in this new approach is based on iteratively solving a set of EEs always embedding the aforementioned nonnegativity constraints. Even though no Gaussian assumption was used to devise these EEs, we, however, showed that if the 2-level LMM is, *indeed*, Gaussian with *i.i.d.* errors, then these EEs are equivalent to the REML ones. Furthermore, the same ideas can be used to fit some other classes of LMMs as well, as we showed for variance components (or ANOVA) LMMs, generalizing an old method of Henderson for REML estimation in such models of Gaussian type. Finally, we initiated a discussion about the tantalizing prospect that the 3*S*-A1.RE (and 3*S*-A1) methods could actually be used for fitting any LMM with u.ho. errors. But that discussion is still open. Furthermore, all the fitting methods rest on the assumption that the fixed effects design matrix has full column rank and that residual errors are uncorrelated and homoscedastic. Therefore, in the future, it will be interesting to examine how to adapt the methods to eliminate those assumptions.

Acknowledgments

We started this work with our former colleague, Pr. Jean Henri Gwét who, unfortunately, passed away on February 18th, 2022. We are much indebted to his collaboration and constant encouragement.

A Proof of Theorem 14

We shall use the following lemmas:

Lemma A.15. *Let* $S \in \mathcal{M}_n(\mathbb{R})$. *Then one has:* $2S - \text{diag}(S) = 0 \iff S = 0$.

Lemma A.16. *In the LMM* (2.1)*, if* $\mathbf{G} = \text{diag}(\mathbf{D}, \cdots, \mathbf{D})$ *(m times), with* $\mathbf{D} \in \mathcal{M}_r(\mathbb{R})$*, then:*

$$\mathbf{Z}\mathbf{G}\mathbf{Z}^T = \widetilde{Z}_1\mathbf{D}\widetilde{Z}_1^T + \cdots + \widetilde{Z}_m\mathbf{D}\widetilde{Z}_m^T, \qquad \text{(a.1)}$$

Table 4. LMM fitting of the `cake` data by Gaussian REML.

LMM (6.1a)			
parameters	*est.*	*std.dev*	*t-stat*
$\beta_{\texttt{temp}}$	0.1604	0.00471	34.05
σ_1^2	40.18		
σ_ε^2	20.63		
LMM (6.1b)			
parameters	*est.*	*std.dev*	*t-stat*
$\beta_{\texttt{temperature.175}}$	27.98	1.162	24.08
$\beta_{\texttt{temperature.185}}$	29.96	1.162	25.78
$\beta_{\texttt{temperature.195}}$	31.42	1.162	27.04
$\beta_{\texttt{temperature.205}}$	32.18	1.162	27.69
$\beta_{\texttt{temperature.215}}$	35.84	1.162	30.84
$\beta_{\texttt{temperature.225}}$	35.36	1.162	30.43
σ_1^2	40.29		
σ_ε^2	20.48		

with $\widetilde{Z}_1, \cdots, \widetilde{Z}_m \in \mathcal{M}_{n,r}(\mathbb{R})$*, partitioning the matrix* $\mathbf{Z}$ *columnwise,* i.e. $\mathbf{Z} = (\widetilde{Z}_1 \ \cdots \ \widetilde{Z}_m)$.

Proof. $\mathbf{Z}\mathbf{G}\mathbf{Z}^T = (\widetilde{Z}_1 \ \cdots \ \widetilde{Z}_m) \cdot \mathrm{diag}(\mathbf{D}, \cdots, \mathbf{D}) \cdot \mathrm{Rows}(\widetilde{Z}_1^T, \cdots, \widetilde{Z}_m^T)$, hence (a.1). □

Lemma A.17. *In the 2-level LMM* (2.2) *with Assumption* $\mathcal{A}2_g$*, let* A *be a matrix satisfying* (2.9)*. If we partition* A *row-wise, conformally to the block-diagonal structure of* $\mathbf{Z} = \mathrm{diag}(\mathbf{Z}_1, \cdots, \mathbf{Z}_m) \in \mathcal{M}_{n,\,q}(\mathbb{R})$*, with* $\mathbf{Z}_j \in \mathcal{M}_{n_j,\,r}(\mathbb{R})$ *(*$j = 1, \cdots, m$*),* i.e.

$$A = \mathrm{Rows}(A_1, \cdots, A_m), \quad \textit{with } A_j \in \mathcal{M}_{n_j,\,n-p}(\mathbb{R}), \tag{a.2a}$$

then, for $\widetilde{\mathbf{Y}} = A^T\mathbf{Y} \in \mathbb{R}^{n-p}$*,* $\widetilde{\mathbf{Z}} = A^T\mathbf{Z} \in \mathcal{M}_{n-p,\,q}(\mathbb{R})$*,* $\widetilde{\mathbf{V}} = A^T\mathbf{V}A \in \mathcal{M}_{n-p}(\mathbb{R})$*, one has:*

$$\widetilde{\mathbf{Z}} = (\widetilde{Z}_1 \ \cdots \ \widetilde{Z}_m), \quad \textit{with } \widetilde{Z}_j = A_j^T\mathbf{Z}_j \in \mathcal{M}_{n-p,\,r}(\mathbb{R}) \ , \tag{a.2b}$$

while the random subvectors $\widetilde{U}_j^*$ *in* (3.18) *and the submatrices* $\mathbf{W}_{jj}^*$ *of* $\mathbf{W}^*$ *in* (3.19) *satisfy, for* $j = 1, \cdots, m$*:*

$$\widetilde{U}_j^* = \mathbf{D}\widetilde{Z}_j^T\widetilde{\mathbf{V}}^{-1}\widetilde{\mathbf{Y}}, \tag{a.2c}$$

$$\mathbf{W}_{jj}^* = \mathbf{D} - \mathbf{D}\widetilde{Z}_j^T\widetilde{\mathbf{V}}^{-1}\widetilde{Z}_j\mathbf{D}. \tag{a.2d}$$

Also, for $j, k = 1, \cdots, m$*,*

$$\mathbf{W}_{jk}^* = -\mathbf{D}\widetilde{Z}_j^T\widetilde{\mathbf{V}}^{-1}\widetilde{Z}_k\mathbf{D} = -\mathbf{D}\widetilde{Z}_k^T\widetilde{\mathbf{V}}^{-1}\widetilde{Z}_j\mathbf{D}. \tag{a.2e}$$

Proof. $\widetilde{\mathbf{Z}} = A^T\mathbf{Z} = (A_1^T \ \cdots \ A_m^T) \cdot \mathrm{diag}(\mathbf{Z}_1, \cdots, \mathbf{Z}_m) = (A_1^T\mathbf{Z}_1 \ \cdots \ A_m^T\mathbf{Z}_m)$, hence (a.2b). On the other hand, from (3.8), Theorem 7 and formula (2.3) giving the BLP of $U|\mathbf{Y}$ in the general LMM (2.1), $\widetilde{U}^* = \mathbf{G}\widetilde{\mathbf{Z}}^T\widetilde{\mathbf{V}}^{-1}\widetilde{\mathbf{Y}}$. Now, using the fact that $\mathbf{G} = \mathrm{diag}(\mathbf{D}, \cdots, \mathbf{D})$,

$$\begin{aligned}\mathbf{G}\widetilde{\mathbf{Z}}^T &= \mathrm{diag}(\mathbf{D}, \cdots, \mathbf{D}) \cdot \mathrm{Rows}(\widetilde{Z}_1^T, \cdots, \widetilde{Z}_m^T)\\ &= \mathrm{Rows}(\mathbf{D}\widetilde{Z}_1^T, \cdots, \mathbf{D}\widetilde{Z}_m^T),\\ \widetilde{U}^* &= \mathrm{Rows}(\mathbf{D}\widetilde{Z}_1^T\widetilde{\mathbf{V}}^{-1}\widetilde{\mathbf{Y}}, \cdots, \mathbf{D}\widetilde{Z}_m^T\widetilde{\mathbf{V}}^{-1}\widetilde{\mathbf{Y}})\\ &= \mathrm{Rows}(\widetilde{U}_1^*, \cdots, \widetilde{U}_m^*).\end{aligned}$$

The latter entails (a.2c) because for $j = 1, \cdots, m$, each $\mathbf{D}\widetilde{Z}_j^T\widetilde{\mathbf{V}}^{-1}\widetilde{\mathbf{Y}}$ is a vector of same length r as vector $\widetilde{U}_j^*$. Lastly, from (3.11c) and (3.7), we have: $\mathbf{W}^* = \mathbf{G} - \mathbf{G}\widetilde{\mathbf{Z}}^T\widetilde{\mathbf{V}}^{-1}\widetilde{\mathbf{Z}}\mathbf{G}$. Now,

$$\begin{aligned}\widetilde{\mathbf{Z}}^T\widetilde{\mathbf{V}}^{-1}\widetilde{\mathbf{Z}} &= \mathrm{Rows}(\widetilde{Z}_1^T\widetilde{\mathbf{V}}^{-1}, \cdots, \widetilde{Z}_m^T\widetilde{\mathbf{V}}^{-1}) \cdot (\widetilde{Z}_1 \ \cdots \ \widetilde{Z}_m)\\ &= \left(\widetilde{Z}_j^T\widetilde{\mathbf{V}}^{-1}\widetilde{Z}_k\right)_{1\,\leq\,j,\,k\,\leq\,m},\end{aligned}$$

which at once implies (a.2d) and (a.2e), given the definition of $\mathbf{W}_{jk}^*$ in Section 3.3.2. □

We also need to generalize a lemma used in [20], with the proof omitted and using tools found in [35]:

Lemma A.18. *Let* $\mathbf{x} \in \mathbb{R}^n$ *and* $N \in \mathcal{M}_n^{\mathrm{spd}}(\mathbb{R})$ *such that*

$$N = Z_1QZ_1^T + \cdots + Z_mQZ_m^T + \alpha\mathbf{I}_n, \tag{a.3a}$$

where $Q \in \mathcal{M}_r^{\mathrm{sym}}(\mathbb{R})$*,* $\alpha \in \mathbb{R}$*, and* $Z_1, \cdots, Z_m \in \mathcal{M}_{n,r}(\mathbb{R})$*.*

If the upper triangular part of Q *has functionally independent elements and which are independent from* α*, whereas* $\mathbf{x}$*,* $Z_1, \cdots, Z_m$ *are independent from* Q *and* α*, then:*

$$\frac{\partial\left(\mathbf{x}^TN^{-1}\mathbf{x}\right)}{\partial Q} = \mathrm{diag}(W) - 2W, \tag{a.3b}$$

$$\frac{\partial \log|N|}{\partial Q} = 2B - \mathrm{diag}(B), \tag{a.3c}$$

$$\frac{\partial\left(\mathbf{x}^TN^{-1}\mathbf{x}\right)}{\partial\alpha} = -\|N^{-1}\mathbf{x}\|^2, \tag{a.3d}$$

$$\frac{\partial \log|N|}{\partial\alpha} = \mathrm{tr}\left(N^{-1}\right), \tag{a.3e}$$

where $W = W_1 + \cdots + W_m$ *and* $B = B_1 + \cdots + B_m$*, with, for* $j = 1, \cdots, m$*:*

$$W_j = Z_j^TN^{-1}\mathbf{x}\mathbf{x}^TN^{-1}Z_j \quad \textit{and} \quad B_j = Z_j^TN^{-1}Z_j.$$

We shall also need the following lemma:

Lemma A.19. *Let* $\widetilde{U}_1, \cdots, \widetilde{U}_m \in \mathbb{R}^r$*. In the 2-level LMM* (2.2)*, if the clusters random effects covariance matrix* $\mathbf{D}$ *satisfies:*

$$\mathbf{D} = \frac{1}{m}\sum_{j\,=\,1}^{m}(\mathbf{W}_{jj}^* + \widetilde{U}_j\widetilde{U}_j^T), \tag{a.4}$$

where $\mathbf{W}_{11}^*, \cdots, \mathbf{W}_{mm}^*$ *are the* $r \times r$ *submatrices* $\mathbf{W}_{jj}^*$ *of* $\mathbf{W}^*$ *in* (3.19)*, then*

$$\mathrm{tr}(\widetilde{\mathbf{Z}}\mathbf{G}\widetilde{\mathbf{Z}}^T\widetilde{\mathbf{V}}^{-1}) = \|\widetilde{U}\|^2_{\mathbf{G}^{-1}}, \quad \textit{with } \widetilde{U} = (\widetilde{U}_1^T, \cdots, \widetilde{U}_m^T)^T \in \mathbb{R}^q. \tag{a.5}$$

Table 5. Results of Algorithm 3*S*-A1.RE-V1 fitting the LMM (6.2) to the `Blackmore` data.

parameters	*est.*	$[2.5\% - 97.5\%]$	*bias*	*std.dev*	$\sqrt{\text{MSE}}$	*t-stat*	*p–value*
$\beta_{\texttt{intercept}}$	-0.2756	$[-0.6389, 0.09318]$	0.0007576	0.1859	0.1859	-1.483	$1.381e-01$
$\beta_{\texttt{age8}}$	0.06376	$[0.006223, 0.1245]$	0.001173	0.03060	0.03062	2.082	$3.732e-02$
$\beta_{\texttt{group:patient}}$	-0.3549	$[-0.8587, 0.1017]$	-0.0008712	0.2411	0.2411	-1.472	$1.410e-01$
$\beta_{\texttt{age8*[group:patient]}}$	0.2404	$[0.1613, 0.3193]$	-0.001331	0.03945	0.03947	6.089	$1.133e-09$
$\sigma^2_{\texttt{intercept}}$	2.106	$[1.670, 2.648]$	0.05703	0.2477	0.2542	8.285	$1.183e-16$
$\sigma^2_{\texttt{age8}}$	0.02889	$[0.02046, 0.04823]$	0.00451	0.007292	0.008574	3.369	$7.545e-04$
$\text{corr}_{\texttt{intercept, age8}}$	-0.2938	$[-0.5071, -0.08937]$	-0.02702	0.3277	0.3327	-2.654	$7.960e-03$
$\sigma^2_{\boldsymbol{\varepsilon}}$	1.535	$[1.301, 1.689]$	-0.03503	0.09844	0.1045	14.69	$7.137e-49$
ρ	2.287	$[1.650, 3.410]$	0.2055	0.4654	0.5088	4.495	$6.960e-06$

• $\text{corr}_{\texttt{intercept, age8}}$: correlation coefficient between random effects variables `intercept` and `age8`.

Table 6. Results of Gaussian REML fitting of the LMM (6.2) to the `Blackmore` data.

parameters	*est.*	*std.dev*	*t-stat*	*parameters*	*est.*
$\beta_{\texttt{intercept}}$	-0.2760	0.1824	-1.514	$\sigma^2_{\texttt{intercept}}$	2.084
$\beta_{\texttt{age8}}$	0.0640	0.03136	2.041	$\sigma^2_{\texttt{age8}}$	0.02716
$\beta_{\texttt{group:patient}}$	-0.3540	0.2353	-1.504	$\text{corr}_{\texttt{intercept, age8}}$	-0.28
$\beta_{\texttt{age8*[group:patient]}}$	0.2399	0.03941	6.087	$\sigma^2_{\boldsymbol{\varepsilon}}$	1.548

Proof. First, given (a.2b) in Lemma A.17, we can apply Lemma A.16 to the transformed LMM (2.10a) to get:

$$\text{tr}(\widetilde{\mathbf{Z}}\mathbf{G}\widetilde{\mathbf{Z}}^T\widetilde{\mathbf{V}}^{-1}) = \sum_{j=1}^{m} \text{tr}(\mathbf{D}\widetilde{Z}_j^T\widetilde{\mathbf{V}}^{-1}\widetilde{Z}_j\mathbf{D}\mathbf{D}^{-1}).$$

Thus, from (a.2d) in Lemma A.17,

$$\text{tr}(\widetilde{\mathbf{Z}}\mathbf{G}\widetilde{\mathbf{Z}}^T\widetilde{\mathbf{V}}^{-1}) = \text{tr}\left[\sum_{j=1}^{m}(\mathbf{D} - \mathbf{W}^*_{jj})\cdot\mathbf{D}^{-1}\right].$$

The end of the proof proceeds as in that of the analogous lemma in [20]. □

• We now proceed to the Proof of Theorem 14:

Proof. Assume observed $\mathbf{Y} = \mathbf{y} \in \mathbb{R}^n$ in the 2-level LMM (2.2), with Assumptions $\mathcal{A}1$-$\mathcal{A}4$ and $\mathcal{A}5$. We start by left multiplying the corresponding LMM (2.1) by A^T, where A is a matrix satisfying (2.9), to get the transformed LMM (2.10a), given (2.10b). Now, with $\mathcal{A}5$,

$$\widetilde{\mathbf{Y}} \overset{\mathcal{D}}{\sim} \mathcal{N}_{n-p}(\mathbf{0}, \widetilde{\mathbf{V}}), \text{ with } \widetilde{\mathbf{V}} = A^T\mathbf{V}A,$$

the latter identity coming from (3.10a) in Theorem 8. Hence, the likelihood of the LMM (2.10a), given data $\widetilde{\mathbf{Y}} = \widetilde{\mathbf{y}} = A^T\mathbf{y}$, is:

$$\begin{aligned}\mathcal{L}_{\text{R}}(\sigma^2_{\boldsymbol{\varepsilon}}, \mathbf{D}|\,\mathbf{y}) &= f_{\widetilde{\mathbf{Y}}}(\widetilde{\mathbf{y}}|\,\sigma^2_{\boldsymbol{\varepsilon}}, \mathbf{D}) \\ &= (2\pi)^{-(n-p)/2}\cdot|\widetilde{\mathbf{V}}|^{-1/2}e^{-\widetilde{\mathbf{y}}^T\widetilde{\mathbf{V}}^{-1}\widetilde{\mathbf{y}}/2},\end{aligned}$$

with log-likelihood:

$$\ell_{\text{R}}(\sigma^2_{\boldsymbol{\varepsilon}}, \mathbf{D}|\,\mathbf{y}) = c - \big(\log|\widetilde{\mathbf{V}}| + \widetilde{\mathbf{y}}^T\widetilde{\mathbf{V}}^{-1}\widetilde{\mathbf{y}}\big)/2. \tag{a.6a}$$

As a preliminary, note that, using (3.10a),

$$\widetilde{\mathbf{V}} = \widetilde{\mathbf{Z}}\mathbf{G}\widetilde{\mathbf{Z}}^T + \sigma^2_{\boldsymbol{\varepsilon}}\mathbf{I}_{n-p}\,. \tag{a.6b}$$

But, since $\mathbf{G} = \text{diag}(\mathbf{D}, \cdots, \mathbf{D})$ (m times), then, thanks to Lemma A.16,

$$\widetilde{\mathbf{Z}}\mathbf{G}\widetilde{\mathbf{Z}}^T = \widetilde{Z}_1\mathbf{D}\widetilde{Z}_1^T + \cdots + \widetilde{Z}_m\mathbf{D}\widetilde{Z}_m^T, \tag{a.6c}$$

with $\widetilde{Z}_1, \cdots, \widetilde{Z}_m \in \mathcal{M}_{n-p,r}(\mathbb{R})$ partitioning $\widetilde{\mathbf{Z}} = A^T\mathbf{Z}$ columnwise. The submatrices $\widetilde{Z}_1, \cdots, \widetilde{Z}_m$ are given by (a.2b) in Lemma A.17. On the other hand, let $\widetilde{\boldsymbol{\beta}}$ and $\widetilde{U}$ satisfy the HMMEs (3.1) given $\sigma^2_{\boldsymbol{\varepsilon}}$, $\mathbf{D}$. Then, from Theorem 1, $\widetilde{U} = \widetilde{U}^*$, which, thanks to (a.2c) in Lemma A.17, is equivalent here to:

$$\widetilde{U}_j = \mathbf{D}\widetilde{Z}_j^T\widetilde{\mathbf{V}}^{-1}\widetilde{\mathbf{y}}, \quad \text{for } j = 1, \cdots, m. \tag{a.6d}$$

1. Given (a.6b)-(a.6c), we use (a.3b) and (a.3c) in Lemma A.18 to differentiate (a.6a) *w.r.t.* $\mathbf{D}$:

$$-2\frac{\partial\ell_{\text{R}}}{\partial\mathbf{D}} = \frac{\partial\log|\widetilde{\mathbf{V}}|}{\partial\mathbf{D}} + \frac{\partial\big(\widetilde{\mathbf{y}}^T\widetilde{\mathbf{V}}^{-1}\widetilde{\mathbf{y}}\big)}{\partial\mathbf{D}} = \text{diag}(S) - 2S,$$

where $S = \sum_{j=1}^{m} S_j$ and $S_j = W_j - B_j$ ($j = 1, \cdots, m$), with

$$W_j = \widetilde{Z}_j^T\widetilde{\mathbf{V}}^{-1}\widetilde{\mathbf{y}}\widetilde{\mathbf{y}}^T\widetilde{\mathbf{V}}^{-1}\widetilde{Z}_j \quad \text{and} \quad B_j = \widetilde{Z}_j^T\widetilde{\mathbf{V}}^{-1}\widetilde{Z}_j.$$

Thanks to Lemma A.15, the invertibility of $\mathbf{D}$, then Lemma A.17 and (a.6d),

$$\begin{aligned}\frac{\partial \ell_{\mathrm{R}}}{\partial \mathbf{D}} = 0 \iff S = 0 &\iff \sum_{j=1}^{m} W_j = \sum_{j=1}^{m} B_j \\ &\iff \mathbf{D} \cdot \sum_{j=1}^{m} W_j \cdot \mathbf{D} = \mathbf{D} \cdot \sum_{j=1}^{m} B_j \cdot \mathbf{D} \\ &\iff \sum_{j=1}^{m} \widetilde{U}_j \widetilde{U}_j^T = \sum_{j=1}^{m} (\mathbf{D} - \mathbf{W}_{jj}^*) \\ &\iff \mathbf{D} = \frac{1}{m} \sum_{j=1}^{m} (\mathbf{W}_{jj}^* + \widetilde{U}_j \widetilde{U}_j^T) = \widetilde{\mathbf{D}},\end{aligned} \tag{a.7a}$$

which proves (4.13).

2. Assume that the *l.h.s.* of (4.13) holds. Hence, (a.7a) is true. Moreover, (a.7a) implies that

$$\mathbf{G} = \widetilde{\mathbf{G}} = \mathrm{diag}(\widetilde{\mathbf{D}}, \cdots, \widetilde{\mathbf{D}}) \quad (m \text{ times}).$$

Using (a.3d) and (a.3e) in Lemma A.18 to differentiate (a.6a) w.r.t. $\sigma_{\boldsymbol{\varepsilon}}^2$,

$$-2\frac{\partial \ell_{\mathrm{R}}}{\partial \sigma_{\boldsymbol{\varepsilon}}^2} \;=\; \mathrm{tr}(\widetilde{\mathbf{V}}^{-1}) - \|\widetilde{\mathbf{V}}^{-1}\widetilde{\mathbf{y}}\|^2. \tag{a.7b}$$

Now, since the LMM (2.1) has u.ho. errors, so does the LMM (2.10a), thanks to (3.10a) in Theorem 8. Then using (3.10c) in Theorem 8, the remaining of the proof proceeds similarly to the end of the analogous theorem in [20], replacing ℓ by ℓ_{R} and n by $n - p$. □

REFERENCES

[1] Gibbons, R. D., Hedeker D., "Applications of Mixed-Effects Models in Biostatistics," Sankhyä: The Indian Journal of Statistics, Series B (1960-2002), vol. 62, no. 1, pp. 70-103, 2000, JSTOR, http://www.jstor.org/stable/25053120

[2] Yarkiner Z., Hunter G., O'Neil R., de Lusignan S., "Applications of Mixed Models for Investigating Progression of Chronic Disease in a Longitudinal Dataset of Patient Records from General Practice," J Biomet Biostat S9: 001, 2013, doi:10.4172/2155-6180.S9-001

[3] van der Merwe A. J., Pretorius, A. L., "An application of the mixed linear model and the dirichlet process prior in veterinary medicine research," Journal of Agricultural, Biological, and Environmental Statistics, vol. 8, no. 3, pp. 328-43, 2003, doi: 10.1198/1085711032192

[4] Milkevych V., Madsen P., Gao H., Jensen J., "The relative effect of genomic information on efficiency of Bayesian analysis of the mixed linear model with unknown variance," Journal of Animal Breeding and Genetics, vol. 138, no. 1 pp. 14-22, 2020, doi: 10.1111/jbg.12497

[5] Torkashvand E., Jozani M. J., Torabi M., "Clustering in Small Area Estimation with Area Level Linear Mixed Models," Journal of the Royal Statistical Society. Series A (Statistics in Society), vol. 180, no. 4, pp. 1253-79, 2017, doi: 10.1111/rssa.12308

[6] Maiti T., "Robust generalized linear mixed models for small area estimation," Journal of Statistical Planning and Inference, vol. 98, no. 1-2, pp. 225-238, 2001, doi: 10.1016/S0378-3758(00)00302-5

[7] Kizilkaya K., Garrick D. J., Fernando R. L., Mestav B., Yildiz M. A., "Use of linear mixed models for genetic evaluation of gestation length and birth weight allowing for heavy-tailed residual effects," Genetics Selection Evolution, vol. 42 no. 1, pp. 26, 2010, doi:10.1186/1297-9686-42-26

[8] Dandine-Roulland C., Perdry H., "The Use of the Linear Mixed Model in Human Genetics," Human Heredity, vol. 80, no. 4, pp. 196-206, 2015, doi: 10.1159/000447634

[9] Lussetti D., Kuljus K., Ranneby B., Ilstedt U., Falck J., Karlsson A., "Using linear mixed models to evaluate stand level growth rates for dipterocarps and Macaranga species following two selective logging methods in Sabah, Borneo," Forest Ecology and Management, vol. 437, pp. 372-379, 2019, doi: 10.1016/j.foreco.2019.01.044

[10] Zamudio F., Yanez M., Guerra F., Fuentes D., Gonzalez A., "Comparative Analysis of Snp Data and Hybrid Taxa Information by Using a Classificatory Linear Mixed Model to Study the Genetic Variation and Heritability of Initial Height Growth In Selected Poplar Hybrids," Tree genetics & genomes, v. 16 , no. 5, pp. 69, 2020, doi: 10.1007/s11295-020-01435-1

[11] Jiang J., "Linear and Generalized Linear Mixed Models and Their Applications," Springer, New York, NY, USA, 2007

[12] Demidenko E., "Mixed Models: Theory and Applications with R (Sec. Ed.)," John Wiley & Sons, Inc., Hoboken, New Jersey, 2013

[13] West B. T., Welch K. B., Galecki A. T., "Linear Mixed Models: A Practical Guide Using Statistical Software," Chapman & Hall/CRC, London, England, 2007

[14] Searle S. R., Casella G., McCulloch C. E., "Variance Components," Wiley, NY, 2006

[15] Lange N., Ryan L., "Assessing normality in random effects models," The Annals of Statistics, vol. 17, no. 2, pp. 624-42, 1989

[16] Jiang J., "REML estimation: Asymptotic behavior and related topics," The Annals of Statistics, vol. 24, no. 1, pp. 255-86, 1996

[17] Rao C. R., Kleffe J., "Estimation of Variance Components and Applications," North-Holland, Amsterdam, 1988

[18] Heyde C. C., "A Quasi-likelihood approach to the REML estimating equations," Statistics & Probability Letters, vol. 21, no. 5, pp. 381-384, 1994, doi: 10.1016/0167-7152(94)00035-2

[19] Heyde C. C., "Quasi-Likelihood And Its Application: A General Approach to Optimal Parameter Estimation," Springer, New York, NY, USA, 1997

[20] Ndong Nguéma E.-P., Fesuh Nono B., Gwét H., "Nonparametric Estimation in Linear Mixed Models with Uncorrelated Homoscedastic Errors," Open Journal of Statistics, vol. 11, no. 4, pp. 558-605, 2021, doi: 10.4236/ojs.2021.114035

[21] Harville D. A., "Maximum Likelihood Approaches to Variance Component Estimation and to Related Problems," Journal of the American Statistical Association, vol. 72, no. 358, pp. 320-38 72:320–338, 1977, doi: 10.2307/2286796

[22] R Core Team, "R: A language and environment for statistical computing," R Foundation for Statistical Computing, Vienna, Austria, 2019, URL: https://www.R-project.org/

[23] Laird N. M., Ware J. M., "Random effects models for longitudinal data," Biometrics, vol. 38, no. 4, 1982, pp. 963-974, 1982, doi: 10.2307/2529876

[24] Searle, S. R., "The Matrix Handling of BLUE and BLUP in the Mixed Linear Model," BU-1275-MA, Biometrics Unit, Cornell University, Ithaca, N.Y., 1995

[25] Robinson G. K., "That BLUP is a Good Thing: The Estimation of Random Effects," Statistical Science vol. 6, no. 1, pp. 15-32, 1991, doi: 10.1214/ss/1177011926

[26] Henderson C. R., "Estimation of genetic parameters (abstract)," Annals of Mathematical Statistics, vol. 21, pp. 309-310, 1950

[27] Henderson C. R., Kempthorne O., Searle S. R., von Krosigk C. N., "The Estimation of environmental and genetic trends from records subject to culling," Biometrics, vol. 15, no. 2, pp. 192-218, 1959, doi: 10.2307/2527669

[28] Harville D. A., "Bayesian inference for variance components using only error contrasts," Biometrika, vol. 61, no. 2, pp. 383-85, 1974, doi: 10.2307/2334370

[29] Patterson H. D., Thompson R., "Recovery of inter-block information when block sizes are unequal," Biometrika, vol. 58, no. 3, pp. 545-54, 1971, doi: 10.2307/2334389

[30] Thompson R., "Estimation of quantitative genetic parameters," Proceedings: Biological Sciences, vol. 275, no. 1635, pp. 679-86, 2008

[31] Cook F. E., "Chocolate cake, I. Optimum baking temperature," MasterÂ's Thesis, Iowa State College, Iowa, 1938

[32] Bates D., Maechler M., Bolker B., Walker S., "Fitting Linear Mixed-Effects Models Using **lme4**," Journal of Statistical Software, vol. 67, no. 1, pp. 1-48, 2015, doi:10.18637/jss.v067.i01

[33] Carpenter J. R., Goldstein H., Rasbash J., "A novel bootstrap procedure for assessing the relationship between class size and achievement," Journal of the Royal Statistical Society, Series C, vol. 52, no. 4, pp. 431-443, 2003, doi: 10.1111/1467-9876.00415

[34] Efron B., Tibsharani J., "An Introduction to the Bootstrap," Chapman and Hall, London, 1993

[35] Graham A., "Kronecker Products and Matrix Calculus: with Applications," Ellis Horwood, Chichester, West Sussex, England, 1981