Original article

ECM approaches to heteroskedastic mixed models with constant variance ratios

JL Foulley

Station de génétique quantitative et appliquée,
Institut national de la recherche agronomique,
78352 Jouy-en-Josas cedex, France

Summary – This paper presents techniques of parameter estimation in heteroskedastic mixed models having constant variance ratios and heterogeneous log residual variances that are described by a linear model. Estimation of dispersion parameters is by standard (ML) and residual (REML) maximum likelihood. Estimating equations are derived using the expectation-conditional maximization (ECM) algorithm and simplified versions of it (gradient ECM). Direct and indirect approaches are proposed with the latter allowing hypothesis testing about the variance ratios. The analysis of a small example is outlined to illustrate the theory.

heteroskedasticity / mixed model / maximum likelihood / EM algorithm

Résumé – Approches ECM des modèles mixtes hétéroscédastiques à rapports de variances constants. *Cet article présente des techniques d'estimation des paramètres intervenant dans des modèles mixtes ayant des rapports de variance constants et des variances résiduelles décrites par un modèle linéaire de leurs logarithmes. Les paramètres de dispersion sont estimés par le maximum de vraisemblance classique (ML) et restreint (REML). Les équations à résoudre pour obtenir ces estimations sont établies à partir de l'algorithme d'espérance-maximisation conditionnelle (ECM) et d'une version simplifiée dite du gradient ECM. Des approches directe et indirecte sont proposées, cette dernière conduisant à un test d'hypothèse sur le rapport de variances. La théorie est illustrée par l'analyse numérique d'un petit exemple.*

hétéroscédasticité / modèle mixte / maximum de vraisemblance / algorithme EM

INTRODUCTION

Heteroskedasticity has recently generated much interest in quantitative genetics and animal breeding. To begin with, there is now a large amount of experimental evidence of heterogeneous variances for most important livestock production traits (Garrick et al, 1989; Visscher et al, 1991; Visscher and Hill, 1992). Second, major theoretical and applied work has been carried out for estimating and testing sources

of heterogeneous variances arising in univariate mixed models (Foulley et al, 1990; Gianola et al, 1992; Weigel et al, 1993; DeStefano, 1994; Foulley and Quaas, 1995).

For many reasons (accuracy of estimation, ease of handling large data sets), a major objective in this area lies in making models as parsimonious as possible. This can be accomplished in at least two ways: i) by modelling variances in the case of potentially numerous sources of heteroskedasticity, and ii) by assuming that some functions of those parameters (eg, intra-class correlation or heritability) are constant. The first aspect corresponds to the so-called structural approach in which the heterogeneity of the log components of variances is described via a linear model structure similar to that used for means (Foulley et al; 1990, 1992; San Cristobal, 1993). Restrictions as in ii) were considered by Meuwissen et al (1996) and Robert et al (1995a,b). Meuwissen et al (1996) introduced a multiplicative mixed model to estimate breeding values and heteroskedasticity factors assuming heritability (h^2) constant across herd-years. Robert et al (1995a,b) developed estimation and testing procedures for homogeneity of heritability within and/or genetic correlations across environments. But Meuwissen's study postulates known h^2 and Robert's research applies to only a single classification of heteroskedasticity.

The purpose of this paper is to propose a complete inference approach for parameters having both features i) and ii), ie, for continuous data described by mixed models with constant variance ratios and heteroskedasticity analyzed via a structural approach. For simplicity, the theory will be presented using a one-way random mixed model for data and afterwards it will be generalized to several u-components. Inference is based on likelihood procedures (REML and ML) and estimating equations derived from the expectation-maximization (EM) theory, more precisely the expectation/conditional maximization (ECM) algorithm recently introduced by Meng and Rubin (1993).

THEORY

Statistical model

As usual, it is assumed that the population can be structured into strata ($i = 1, 2, ..., I$) corresponding to potential factors of heterogeneity. Let the one-way random model be written as:

$$\mathbf{y}_i = \mathbf{X}_i\boldsymbol{\beta} + \sigma_{u_i}\mathbf{Z}_i\mathbf{u}^* + \mathbf{e}_i \qquad [1]$$

where $\mathbf{y}_i$ is the ($n_i \times 1$) data vector for stratum i; $\boldsymbol{\beta}$ is a ($p \times 1$) vector of unknown fixed effects with incidence matrix $\mathbf{X}_i$, and $\mathbf{e}_i$ is the ($n_i \times 1$) vector of residuals. The contribution of random effects is expressed as in Foulley and Quaas (1995) as $\sigma_{u_i}\mathbf{Z}_i\mathbf{u}^*$ where $\mathbf{u}^*$ is a ($q \times 1$) vector of standardized deviations, $\mathbf{Z}_i$ is the corresponding incidence matrix and σ_{u_i} is the square root of the u-component of variance the value of which depends on stratum i. Classical assumptions are made for the distributions of $\mathbf{u}^*$ and $\mathbf{e}_i$, ie, $\mathbf{u}^* \sim N(\mathbf{0}, \mathbf{A})$, $\mathbf{e}_i \sim N(\mathbf{0}, \sigma_{e_i}^2\mathbf{I}_{n_i})$, and $E(\mathbf{u}^*\mathbf{e}_i') = \mathbf{0}$.

The notation in [1] is unusual as compared to that used in the statistical literature on mixed effects (eg, Laird et al, 1987). There are practical motivations for such

an expression of the random part especially in animal breeding. For instance the between sire variance may vary according to the environment in which the progeny of the sires are raised. Note also that σ_{u_i} can be viewed as a regression coefficient of any element of $\mathbf{y}_i$ on the corresponding element of $\mathbf{Z}_i\mathbf{u}^*$. Thus, in animal breeding, σ_{u_i} acts as a scaling factor of a vector $\mathbf{u}^*$ of standardized sire values on which, for instance, selection can be based.

A structure is hypothesized on the residual variance so as to model the influence of factors causing heteroskedasticity. This is carried out along the lines presented in Foulley et al (1990, 1992) via a linear regression on log-variances:

$$ln\sigma_{e_i}^2 = \mathbf{p}_i'\boldsymbol{\delta} \qquad [2]$$

where $\boldsymbol{\delta}$ is an unknown $(r \times 1)$ real-valued vector of parameters and $\mathbf{p}_i'$ is the corresponding $(1 \times r)$ row incidence vector of qualitative or continuous covariates.

Furthermore, the assumption of a constant intra-class correlation (or heritability) implies setting

$$\sigma_{u_i} = \tau\sigma_{e_i}, \ \forall i \qquad [3]$$

EM-REML estimation

Use is made here of the EM algorithm of Dempster et al (1977) to compute REML estimates of parameters involved in variance components (Patterson and Thompson, 1971; Searle et al, 1992). The basic procedure proposed by Foulley and Quaas (1995) is applied here after some adjustment of the M-step taking advantage of the ECM algorithm of Meng and Rubin (1993).

Letting $\mathbf{y} = (\mathbf{y}_1', \mathbf{y}_2', \ldots, \mathbf{y}_i', \ldots, \mathbf{y}_I')'$, $\mathbf{e} = (\mathbf{e}_1', \mathbf{e}_2', \ldots, \mathbf{e}_i', \ldots, \mathbf{e}_I')'$, $\boldsymbol{\gamma} = (\boldsymbol{\delta}', \tau)'$, the ECM algorithm is based on a complete data set defined by $\mathbf{x} = (\boldsymbol{\beta}', \mathbf{u}^{*'}, \mathbf{e}')'$ and its log-likelihood $L(\boldsymbol{\gamma}; \mathbf{x})$. The iterative process takes place as follows.

The E-step is defined as usual, ie, at iteration $[t]$, calculate the conditional expectation of $L(\boldsymbol{\gamma}; \mathbf{x})$ given the data $\mathbf{y}$ and $\boldsymbol{\gamma} = \boldsymbol{\gamma}^{[t]}$

$$Q(\boldsymbol{\gamma}|\boldsymbol{\gamma}^{[t]}) = \mathbf{E}\left[L(\boldsymbol{\gamma}; \mathbf{x})|\mathbf{y}, \boldsymbol{\gamma} = \boldsymbol{\gamma}^{[t]}\right]$$

which, as shown in Foulley and Quaas (1995), reduces to

$$Q(\boldsymbol{\gamma}|\boldsymbol{\gamma}^{[t]}) = \text{const} - \frac{1}{2}\sum_{i=1}^{I} n_i ln\sigma_{e_i}^2 - \frac{1}{2}\sum_{i=1}^{I}\sigma_{e_i}^{-2}\, \mathbf{E}_c^{[t]}(\mathbf{e}_i'\mathbf{e}_i) \qquad [4]$$

where $\mathbf{E}_c^{[t]}(.)$ is a condensed notation for a conditional expectation taken with respect to the distribution of $\mathbf{x}|\mathbf{y}, \boldsymbol{\gamma} = \boldsymbol{\gamma}^{[t]}$.

Since the parameters to be estimated are heterogeneous, the estimating equations are derived at the maximization stage from a slightly different version of the EM algorithm, the so-called ECM algorithm. As explained in detail in Meng and Rubin (1993), a CM stage replaces the M-step by a sequence of several conditional maximization steps. This is basically the same principle as that employed in a cyclic

ascent maximization procedure (Zangwill, 1969). We suggest here the following procedure:

- maximize Q over $\boldsymbol{\gamma}$ to get $\boldsymbol{\delta}^{[t+1]}$ with τ set at its last value $\tau^{[t]}$, ie

$$\boldsymbol{\delta}^{[t+1]} = \text{Argmax}\, Q(\boldsymbol{\delta}, \tau^{[t]}|\boldsymbol{\gamma}^{[t]}) \tag{5}$$

- then, maximize Q over τ to obtain $\tau^{[t+1]}$ with $\boldsymbol{\delta}$ in $\boldsymbol{\gamma}$ of $Q(\boldsymbol{\gamma}|\boldsymbol{\gamma}^{[t]})$ set to $\boldsymbol{\delta}^{[t+1]}$, ie,

$$\tau^{[t+1]} = \text{Argmax}\, Q(\boldsymbol{\delta}^{[t+1]}, \tau|\boldsymbol{\gamma}^{[t]}) \tag{6}$$

Thus, the maximization step consists of two CM-steps within the same E-step in order to reduce the need to compute the conditional expectation of $\mathbf{e}_i'\mathbf{e}_i$, and its components more than once. The algebra of differentiation is given in *Appendix A*. The iterative system for computing formulae $\boldsymbol{\delta}$ can be written as

$$\mathbf{P}'\mathbf{W}_{\delta\delta}^{[t,\ell]}\mathbf{P}.\boldsymbol{\Delta}\boldsymbol{\delta}^{[t,\ell+1]} = \mathbf{P}'\,\mathbf{v}_{\delta}^{[t,\ell]} \tag{7}$$

where

$$\boldsymbol{\Delta}\boldsymbol{\delta}^{[t,\ell+1]} = \boldsymbol{\delta}^{[t,\ell+1]} - \boldsymbol{\delta}^{[t,\ell]}$$

$$\mathbf{P}'_{(r\times I)} = (\mathbf{p}_1, \mathbf{p}_2, \ldots, \mathbf{p}_i, \ldots, \mathbf{p}_I)$$

$$\mathbf{v}_{\delta\,(I\times 1)}^{[t,\ell]} = \left\{\nu_{\delta,i}^{[t,\ell]}\right\}$$

with the elements of the right-hand side being

$$\nu_{\delta,i}^{[t,\ell]} = \frac{\text{E}_c^{[t]}\left[(\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta})'\mathbf{e}_i^{[t,\ell]}\right]}{\sigma_{e_i}^{2[t,\ell]}} - n_i \tag{8a}$$

based on $\mathbf{e}_i^{[t,\ell]} = \mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta} - \tau^{[t]}\sigma_{e_i}^{[t,\ell]}\mathbf{Z}_i\mathbf{u}^*$, and $\sigma_{e_i}^{2[t,\ell]} = \exp(\mathbf{p}_i'\boldsymbol{\delta}^{[t,\ell]})$

$$\mathbf{W}_{\delta\delta\,(r\times r)}^{[t,\ell]} = \text{Diag}\left\{w_{\delta\delta,ii}^{[t,\ell]}\right\}$$

with

$$w_{\delta\delta,ii}^{[t]} = \frac{\text{E}_c^{[t]}\left[(\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta})'(\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta})\right]}{\sigma_{e_i}^{2[t,\ell]}} - \frac{\tau^{[t]}\,\text{E}_c^{[t]}\left[(\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta})'\mathbf{Z}_i\mathbf{u}^*\right]}{2\sigma_{e_i}^{[t,\ell]}} \tag{8b}$$

Note that for this algorithm to be a true ECM, one would have to iterate the NR algorithm in [7] within an inner cycle (index ℓ) until convergence to the conditional maximizer $\boldsymbol{\gamma}^{[t+\ell]} = \boldsymbol{\gamma}^{[t,c]}$ at each M-step $[t]$. In practice it may be advantageous to reduce the number of inner iterations, even up to only one, ie, by solving just once

$$\mathbf{P}'\mathbf{W}_{\delta\delta}^{[t]}\mathbf{P}\cdot\boldsymbol{\Delta}\boldsymbol{\delta}^{[t]} = \mathbf{P}'\mathbf{v}_{\delta}^{[t]} \tag{9}$$

However, caution should be exercised when applying such a hybrid algorithm that no longer guarantees the monotonic convergence in likelihood values (Lange, 1995).

The formula to update τ reduces to

$$\tau^{[t+1]} = \frac{\sum_{i=1}^{I} \mathrm{E}_c^{[t]}\left[\mathbf{u}^{*'}\mathbf{Z}_i'(\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta})\right]/\sigma_{e_i}^{[t+1]}}{\sum_{i=1}^{I} \mathrm{E}_c^{[t]}(\mathbf{u}^{*'}\mathbf{Z}_i'\mathbf{Z}_i\mathbf{u}^*)} \qquad [10]$$

mimicking the form of a scaled regression coefficient pooled over strata.

The elements to compute at the E-step can be expressed as functions of the sums $\mathbf{X}_i'\mathbf{y}_i$, $\mathbf{Z}_i'\mathbf{y}_i$, the sums of squares $\mathbf{y}_i'\mathbf{y}_i$ within strata, and GLS-BLUP solutions of Henderson's mixed model equations and of their accuracy (Henderson, 1984), ie

$$\begin{bmatrix} \sum_{i=1}^{I} \mathbf{X}_i'\mathbf{X}_i/\sigma_{e_i}^{2[t]} & \sum_{i=1}^{I} \mathbf{X}_i'\mathbf{Z}_i\tau^{[t]}/\sigma_{e_i}^{2[t]} \\ \sum_{i=1}^{I} \mathbf{Z}_i'\mathbf{X}_i\tau^{[t]}/\sigma_{e_i}^{2[t]} & \sum_{i=1}^{I} \mathbf{Z}_i'\mathbf{Z}_i\tau^{[t]2} + \mathbf{A}^{-1} \end{bmatrix} \begin{bmatrix} \widehat{\boldsymbol{\beta}}^{[t]} \\ \widehat{\mathbf{u}}^{*[t]} \end{bmatrix} = \begin{bmatrix} \sum_{i=1}^{I} \mathbf{X}_i'\mathbf{y}_i/\sigma_{e_i}^{2[t]} \\ \sum_{i=1}^{I} \mathbf{Z}_i'\mathbf{y}_i/\sigma_{e_i}^{2[t]} \end{bmatrix} \qquad [11]$$

where $\widehat{\mathbf{u}}^{*[t]} = \mathrm{E}(\mathbf{u}^*|\mathbf{y}, \boldsymbol{\delta}^{[t]}, \tau^{[t]})$.

Thus, deleting $[t]$ for the sake of simplicity, one has:

$$\mathrm{E}_c\left[(\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta})'(\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta})\right] = (\mathbf{y}_i - \mathbf{X}_i\widehat{\boldsymbol{\beta}})'(\mathbf{y}_i - \mathbf{X}_i\widehat{\boldsymbol{\beta}}) + \mathrm{tr}(\mathbf{X}_i'\mathbf{X}_i\mathbf{C}_{\beta\beta}) \quad [12a]$$

$$\mathrm{E}\left[\mathbf{u}^{*'}\mathbf{Z}_i'(\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta})\right] = \widehat{\mathbf{u}}^{*'}\mathbf{Z}_i'(\mathbf{y}_i - \mathbf{X}_i\widehat{\boldsymbol{\beta}}) - \mathrm{tr}(\mathbf{Z}_i'\mathbf{X}_i\mathbf{C}_{\beta u}) \qquad [12b]$$

$$\mathrm{E}_c(\mathbf{u}^{*'}\mathbf{Z}_i'\mathbf{Z}_i\mathbf{u}^*) = \widehat{\mathbf{u}}^{*'}\mathbf{Z}_i'\mathbf{Z}_i\widehat{\mathbf{u}}^* + \mathrm{tr}(\mathbf{Z}_i'\mathbf{Z}_i\mathbf{C}_{uu}) \qquad [12c]$$

where $\widehat{\boldsymbol{\beta}}$ and $\widehat{\mathbf{u}}^*$ are mixed model equations for $\boldsymbol{\beta}$ and $\mathbf{u}^*$, and $\mathbf{C} = \begin{bmatrix} \mathbf{C}_{\beta\beta} & \mathbf{C}_{\beta u} \\ \mathbf{C}_{u\beta} & \mathbf{C}_{uu} \end{bmatrix}$ is the partitioned inverse of the coefficient matrix.

Expressions in [12a–c] can easily accommodate grouped data (see *Appendix B*).

The close connection between the system of equations [7] for residual parameters and formula [12] given in Foulley et al (1990) can be observed. There is also a remarkable similarity between formula [9] for the ratio and formula [7] in Foulley and Quaas (1995). This means that the computations can be implemented with very little change in the code used previously. True or gradient EM could also have been applied (see *Appendix A*). The advantage of ECM will be more substantial for the next situations considered, and especially in the case of the indirect approach.

Extension to several u-components

Formulae [7], [8ab] and [9] can easily be generalized to a mixed model including several $(k = 1, 2, \ldots, K)$ independent u-components

$$\mathbf{y}_i = \mathbf{X}_i\boldsymbol{\beta} + \sum_{k=1}^{K} \tau_k \sigma_{e_i} \mathbf{Z}_{ik}\mathbf{u}_k^* + \mathbf{e}_i \qquad [13]$$

with $\tau_k = \sigma_{u_{ik}}/\sigma_{e_i}$ constant over strata i.

Letting $\boldsymbol{\gamma} = (\boldsymbol{\delta}', \boldsymbol{\tau}')'$ as previously but now with $\boldsymbol{\tau} = \{\tau_k\}$ being a vector of ratios of standard deviations, the Q function to be maximized has the same form as in [4] with $\mathbf{e}_i$ expressed from [13]. One can perform the CM-steps using either i) the sequence $\boldsymbol{\delta}, \tau_1, \tau_2, \ldots, \tau_k, \ldots, \tau_K$, ie, each τ_k one by one, the remaining ones being held constant, or ii) the sequence $\boldsymbol{\delta}$, and $\boldsymbol{\tau}$ as a whole with all the τ_ks maximized jointly. In both cases, the algorithm for computing $\boldsymbol{\delta}$ is formally the same as in [7] with only a slight change in the definition of the elements of $\mathbf{W}_{\delta\delta}$, $\mathbf{v}_\delta$ being unchanged

$$w_{\delta\delta,ii}^{[t,\ell]} = \frac{\mathrm{E}_c^{[t]}\left[(\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta})'(\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta})\right]}{\sigma_{e_i}^{2[t,\ell]}} - \frac{\displaystyle\sum_{k=1}^{K} \tau_k^{[t]} \mathrm{E}_c^{[t]}\left[(\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta})'\mathbf{Z}_i\mathbf{u}_k^*\right]}{2\sigma_{e_i}^{[t,\ell]}} \qquad [14]$$

If the conditional maximization of the τ_ks takes place one by one (case i), formula [9] still applies for each of them. Otherwise (case ii), one has to solve the following system:

$$\mathbf{F}^{[t]}\boldsymbol{\tau}^{[t+1]} = \mathbf{g}^{[t]} \qquad [15]$$

where

$$\mathbf{F}_{(K \times K)}^{[t]} = \left\{f_{k\ell}^{[t]}\right\} = \left\{\sum_{i=1}^{I} \mathrm{E}_c^{[t]}(\mathbf{u}_k^{*'} \mathbf{Z}_{ik}' \mathbf{Z}_{i\ell} \mathbf{u}_\ell^*)\right\} \qquad [16a]$$

and

$$\mathbf{g}_{(K \times 1)}^{[t]} = \left\{g_k^{[t]}\right\} = \left\{\sum_{i=1}^{I} (\sigma_{e_i}^{[t]})^{-1} \mathrm{E}_c^{[t]}\left[\mathbf{u}_k^{*'} \mathbf{Z}_{ik}'(\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta})\right]\right\} \qquad [16b]$$

An indirect approach

The original model with a constant τ ratio specified in [1–3] can be viewed as a special case of a more general model

$$\mathbf{y}_i = \mathbf{X}_i\boldsymbol{\beta} + \tau_i \sigma_{e_i} \mathbf{Z}_i\mathbf{u}^* + \mathbf{e}_i \qquad [17]$$

with, as previously, $ln\sigma_{e_i}^2 = \mathbf{p}_i'\boldsymbol{\delta}$, but also with a linear structure on log-ratios

$$ln\,\tau_i = \mathbf{h}_i'\boldsymbol{\lambda} \qquad [18]$$

involving either the same $(\mathbf{h}_i = \mathbf{p}_i)$ or possibly different covariates.

Letting $\gamma = (\delta', \lambda')'$ here, the sequence of the CM-steps are

$$\delta^{[t+1]} = \text{Argmax} Q(\delta, \lambda^{[t]} | \gamma^{[t]}) \tag{19a}$$

and

$$\lambda^{[t+1]} = \text{Argmax} Q(\delta^{[t+1]}, \lambda | \gamma^{[t]}) \tag{19b}$$

The algorithm for δ is the same as in [7]. The algebra for λ is shown in the *Appendix*, and leads to a system that can be written under a similar form as that of δ

$$\mathbf{H}' \mathbf{W}_{\lambda\lambda}^{[t,\ell]} \mathbf{H} \cdot \Delta\lambda^{[t,\ell+1]} = \mathbf{H}' \mathbf{v}_{\lambda}^{[t,\ell]} \tag{20}$$

where

$$\Delta\lambda^{[t,\ell+1]} = \lambda^{[t,\ell+1]} - \lambda^{[t,\ell]}$$

$$\mathbf{H}'_{(s \times I)} = (\mathbf{h}_1, \mathbf{h}_2, \ldots, \mathbf{h}_i, \ldots, \mathbf{h}_I)$$

$$\mathbf{v}_{\lambda}^{[t,\ell]}{}_{(I \times 1)} = \left\{ \nu_{\lambda,i}^{[t,\ell]} \right\} = \left\{ \tau_i^{[t,\ell]} \left(\sigma_{e_i}^{[t]} \right)^{-1} \mathbf{E}_c^{[t]} \left(\mathbf{u}^{*'} \mathbf{Z}_i' \mathbf{e}_i^{[t,\ell]} \right) \right\} \tag{21}$$

based on $\mathbf{e}_i^{[t,\ell]} = \mathbf{y}_i - \mathbf{X}_i\beta - \tau_i^{[t,\ell]}\sigma_{e_i}^{[t]}\mathbf{Z}_i\mathbf{u}^*$, and $\tau_i^{[t,\ell]} = \exp(\mathbf{h}_i'\lambda^{[t,\ell]})$

$$\mathbf{W}_{\lambda\lambda}^{[t,\ell]}{}_{(s \times s)} = \text{diag} \left\{ \tau_i^{[t,\ell]} \left[\tau_i^{[t,\ell]} \mathbf{E}_c^{[t]}(\mathbf{u}^{*'} \mathbf{Z}_i' \mathbf{Z}_i \mathbf{u}^*) - \left(\sigma_{e_i}^{[t]} \right)^{-1} \mathbf{E}_c^{[t]} \left(\mathbf{u}^{*'} \mathbf{Z}_i' \mathbf{e}_i^{[t,\ell]} \right) \right] \right\} \tag{22}$$

For practical reasons, one may also wish to limit the number of inner iterations (index ℓ) even to only one in order to reduce the volume of computation but the application of this ECM gradient algorithm should be performed carefully. Further empirical simplifications for the elements of [22] can be proposed along the same lines as in Foulley et al (1990).

Again, these results can be extended to a model with several random independent factors ($k = 1, 2, \ldots, K$) by setting

$$\ell n \tau_{k,i} = \mathbf{h}_{ik}' \lambda_k$$

Actually, if the CM-steps are performed for each vector λ_k separately, the same formulae as in [20], [21] and [22] apply: just replace $\tau_i, \mathbf{Z}_i, \mathbf{u}^*$ by $\tau_{i,k}, \mathbf{Z}_{ik}, \mathbf{u}_k^*$ and set $\mathbf{e}_i$ equal to $\mathbf{y}_i - \mathbf{X}_i\beta - \sum_{k=1}^{K} \tau_{i,k}\sigma_{e_i}\mathbf{Z}_{ik}\mathbf{u}_k^*$.

ML estimation

It may be interesting in some instances to use ML rather than REML for estimating variance components (see *Discussion*). The ECM procedure developed in this paper can be easily adapted to obtain ML parameter estimates. β is now part of the parameter vector instead of being a vector of random effects with infinite variance included in missing data. The Q function to be maximized has the same formal expression as in [4] but here at the E-step, expectations have to be taken with

respect to the distribution of $\mathbf{u}^*$ given $\mathbf{y}, \boldsymbol{\gamma} = \boldsymbol{\gamma}^{[t]}$, and $\boldsymbol{\beta} = \boldsymbol{\beta}^{[t]}$. Maximization with respect to $\boldsymbol{\beta}$ can be based on the equation $\partial Q/\partial \boldsymbol{\beta} = \mathbf{0}$, ie

$$\sum_{i=1}^{I} \sigma_{e_i}^{-2} \mathbf{X}_i' \left[\mathbf{y}_i - \mathbf{X}_i \boldsymbol{\beta} - \tau \sigma_{e_i} \mathbf{Z}_i \, \mathrm{E}(\mathbf{u}^*|\mathbf{y}, \boldsymbol{\beta}, \boldsymbol{\gamma}) \right] = \mathbf{0} \qquad [23]$$

One can proceed as previously, ie, run two CM-steps for the dispersion parameters based on the same E-step so as to obtain $\boldsymbol{\delta}^{[t+1]}$ and $\tau^{[t+1]}$ (or $\boldsymbol{\lambda}^{[t+1]}$), and then perform an additional CM-step for computing $\boldsymbol{\beta}^{[t+1]}$ based on [23], ie

$$\sum_{i=1}^{I} \sigma_{e_i}^{-2[t+1]} \mathbf{X}_i' \mathbf{X}_i \boldsymbol{\beta}^{[t+1]} = \sum_{i=1}^{I} \sigma_{e_i}^{-2[t+1]} \mathbf{X}_i' \left[\mathbf{y}_i - \tau^{[t+1]} \sigma_{e_i}^{[t+1]} \mathbf{Z}_i \, \mathrm{E}(\mathbf{u}^*|\mathbf{y}, \boldsymbol{\beta}^{[t]}, \boldsymbol{\delta}^{[t]}, \tau^{[t]}) \right]$$

$$[24]$$

Alternatively, it may be advantageous to perform the CM-step for $\boldsymbol{\beta}$ and the next E-step jointly by solving Henderson's mixed model equations in $\boldsymbol{\beta}^{[t+1]}$ and $\widehat{\mathbf{u}}^{*[t+1]} = \mathrm{E}(\mathbf{u}^*|\mathbf{y}, \boldsymbol{\delta}^{[t+1]}, \tau^{[t+1]})$ based on $\boldsymbol{\delta}^{[t+1]}$ and $\tau^{[t+1]}$.

Formulae for the two CM-steps do not change. The only additional modification results from taking the conditional expectation of components of $\mathbf{e}_i' \mathbf{e}_i$ given $\mathbf{y}, \boldsymbol{\gamma} = \boldsymbol{\gamma}^{[t]}, \boldsymbol{\beta} = \boldsymbol{\beta}^{[t]}$ instead of $\mathbf{y}, \boldsymbol{\gamma} = \boldsymbol{\gamma}^{[t]}$. Formulae in [12] reduce to

$$\mathrm{E}_{c'} \left[(\mathbf{y}_i - \mathbf{X}_i \boldsymbol{\beta})'(\mathbf{y}_i - \mathbf{X}_i \boldsymbol{\beta}) \right] = (\mathbf{y}_i - \mathbf{X}_i \widehat{\boldsymbol{\beta}})'(\mathbf{y}_i - \mathbf{X}_i \widehat{\boldsymbol{\beta}}) \qquad [25a]$$

$$\mathrm{E}_{c'} \left[\mathbf{u}^{*'} \mathbf{Z}_i'(\mathbf{y}_i - \mathbf{X}_i \boldsymbol{\beta}) \right] = \widehat{\mathbf{u}}^{*'} \mathbf{Z}_i'(\mathbf{y}_i - \mathbf{X}_i \widehat{\boldsymbol{\beta}}) \qquad [25b]$$

$$\mathrm{E}_{c'} (\mathbf{u}^{*'} \mathbf{Z}_i' \mathbf{Z}_i \mathbf{u}^*) = \widehat{\mathbf{u}}^{*'} \mathbf{Z}_i' \mathbf{Z}_i \widehat{\mathbf{u}}^* + \mathrm{tr}\,(\mathbf{Z}_i' \mathbf{Z}_i \mathbf{M}_{uu}^{-1}) \qquad [25c]$$

where $\mathbf{M}_{uu}$ is the u by u block of the coefficient matrix [11].

Note that the trace terms inside those formulae have disappeared or have been greatly simplified owing to conditioning with respect to $\boldsymbol{\beta} = \boldsymbol{\beta}^{[t]}$. More generally, for models [13] involving several u-components, [25c] becomes

$$\mathrm{E}_{c'} (\mathbf{u}_k^{*'} \mathbf{Z}_{ik}' \mathbf{Z}_{i\ell} \mathbf{u}_k^*) = \widehat{\mathbf{u}}_k^{*'} \mathbf{Z}_{ik}' \mathbf{Z}_{i\ell} \widehat{\mathbf{u}}_k^* + \mathrm{tr} \left[\mathbf{Z}_{ik}' \mathbf{Z}_{i\ell} (\mathbf{M}_{uu}^{-1})_{k\ell} \right] \qquad [26]$$

where $(\mathbf{M}_{uu}^{-1})_{k\ell}$ is the block pertaining to random factors k and ℓ in the inverse of the random part of the coefficient matrix.

Numerical example

The procedures presented in this paper are illustrated with a small data set obtained from simulation. Data were generated according to a cross-classified model having two (environmental) fixed factors ($A = 2$ levels; $B = 3$ levels) and one (genetic) random factor ($S = 9$ levels). The genetic contribution consists of sire and maternal grand sire effects, the latter being assumed to have half the value of the first one. The model to generate the records was

$$y_{ijklm} = \mu + a_i + b_j + \tau \sigma_{eij}(s_k^* + 1/2 s_\ell^*) + e_{ijklm} \qquad [27]$$

where μ is a general mean, a_i the effect of environmental factor A ($i = 1, 2$), b_j the effect of environmental factor B ($j = 1, 2, 3$), s_k^* the standardized contribution of male k as a sire, and $1/2s_\ell^*$ the standardized contribution of male ℓ as a maternal grand sire, and e_{ijklm} the residual term.

Values chosen for the fixed effects were (using a full-rank parameterization): $\mu + a_1 + b_1 = 100$; $a_2 - a_1 = 20$; $b_2 - b_1 = -10$; $b_3 - b_1 = -20$. The vector $\mathbf{s}^* = \{s_k^*\}$ of sire effects is assumed to be $N(\mathbf{0}, \mathbf{A})$ with elements of the relationship matrix $\mathbf{A}$ shown at the bottom of table I.

Table I. Structure of the data set, number (n), sum of observations ($\sum y$) and sum of squares ($\sum y^2$) per cell.

No	A	B	S	T	n	$\sum y$.	$\sum y^2$
1	1	1	1	4	21	2287	258771
2	1	2	1	4	19	1739	162655
3	1	1	1	7	14	1536	174610
4	1	3	1	7	7	628	57224
5	1	2	1	8	6	598	60398
6	1	3	2	6	12	1045	95517
7	1	1	2	7	7	770	86554
8	1	2	2	8	18	1726	169736
9	2	1	2	8	27	3356	437890
10	2	2	3	5	7	786	92030
11	2	1	3	5	19	2523	350873
12	2	2	3	5	10	1280	168166
13	2	3	3	2	37	4305	536409
14	2	2	3	2	13	1460	168160
15	2	3	4	7	13	1231	137599
16	2	3	4	8	6	243	14887
17	2	2	4	9	19	1930	204578
18	2	2	4	5	12	1141	111575

A, B: environmental factors treated as fixed; S = sire and T = maternal grand sire treated as random. Elements of the $\mathbf{A}$ matrix are the following $\forall i$, $(i, i) = 1$; $\forall i \neq j$, $(i, j) = (j, i) = 0$ except for $(1, 5) = (2, 5) = (3, 7) = (4, 6) = 1/2$, and $(1, 2) = (8, 9) = 1/4$.

Residual variances were obtained from

$$\sigma_{e_{ij}}^2 = \exp(\mu^* + a_i^* + b_j^*) \qquad [28]$$

with a base line value $\sigma_{e_{11}}^2 = \exp(\mu^* + a_1^* + b_1^*) = 400$, and multiplicative adjustment factors: $\exp(a_2^* - a_1^*) = 2$; $\exp(b_2^* - b_1^*) = 1/2$ and $\exp(b_3^* - b_1^*) = 3/2$. The ratio $\tau = \sigma_{s_{ij}}/\sigma_{e_{ij}}$ of the square root of the sire to the residual variance was taken as constant over A × B cells and set to $8.75^{-1/2}$ (heritability equal to 0.41).

There were 267 observations distributed among 18 different AB × sire× maternal grand sire subclasses. The data structure is displayed in table I as well as cell size (n), sum ($\sum y$) and sum of squares ($\sum y^2$) in each suclass.

Tests of hypotheses about the location parameters $\boldsymbol{\beta}$, the residual dispersion parameters $\boldsymbol{\delta}$ and the ratios τ_{ij} were carried out via the likelihood ratio statistic as described in previous studies (Foulley et al, 1990 1992; San Cristobal et al, 1993; Meyer et al, 1993; Foulley and Quaas, 1995). Formulae by Quaas (1992) were used to compute maximized likelihood functions $(L_{\max})$.

Results can be arranged as an analysis of variance (or deviance) table: see table II for hypothesis testing about $\boldsymbol{\beta}$, and table III for residual ($\boldsymbol{\delta}$) and ratio ($\boldsymbol{\lambda}$) parameters. Note also that the test statistic for $\boldsymbol{\beta}$ relies on $-2L_{\max}$ evaluated from the ML estimates of all parameters, whereas a maximized residual likelihood can be better employed for $\boldsymbol{\delta}$ and $\boldsymbol{\lambda}$.

Interaction effects on location parameters are constantly rejected under different assumptions for the other parameters. The hypothesis of residual variance homogeneity is strongly rejected as well as single factor descriptions of heterogenity. The assumption of a constant ratio τ turns out to be a reasonable one. The test results eventually agree with the simulation model; they support the practical conclusion that the $\mu + A + B$ model is the most appropriate to account for variation both in location and in log-residual variances, the ratio τ being constant.

The estimation procedure for $\boldsymbol{\delta}$ and τ (or $\boldsymbol{\lambda}$) is illustrated in table IV for this model and an alternative one using both standard and residual maximum likelihood methods of estimation. ML and REML estimates of residual variances do not differ very much; on the contrary, the ML estimates of the ratio τ turns out to be, as expected, lower than the REML ones, the values of the latter being close to the true value.

DISCUSSION AND CONCLUSION

The main purpose of this paper was to extend the general structural approach to heteroskedasticity in mixed models proposed by Foulley et al (1990, 1992) to the case of homogeneous ratios of u to e variance components.

In a sire by environment interaction, this is equivalent to postulating homogeneous intra-class correlations or heritabilities. This seems to be a reasonable assumption in practice, or at least serves as a suitable compromise between the existence of heteroskedasticity and parsimony of models. Less restrictive assumptions might also be investigated (Quaas, 1995, pers comm). This paper also provides a generalization of LR tests of this assumption to unbalanced data and complex model structures: see the previous work of Visscher (1992) on a one-way random balanced design, and that by Robert et al (1995a,b) for heterogeneous variances due to a single classification.

The EM algorithm turns out to be a convenient and powerful tool for solving variance component estimation problems. The ECM algorithm allows us to simplify the estimating equations, in particular the ECM gradient version. The advantage of this algorithm was especially clear here in the case of the indirect approach. A few examples of this for the mixed model have been already mentioned (Meng and Rubin, 1993 example 1; Walker, 1996). It offers great flexibility in defining the sequence of the conditional maximization steps, all the alternatives of which have not been investigated here. In the case addressed in this paper, the basic statistics

Table III. Likelihood statistics for testing location and dispersion parameters in heteroskedastic mixed models: residual components and ratios of variance

No	Location	Model[a] Residual	Ratio	Likelihood[b] par	-2L	Covariate	Test[c] Comp	df	Statistic	P-value
1	$\mu+A+B$	$\mu^*+A^*+B^*+AB^*$	μ''	11	2420.9841					
2	$\mu+A+B$	$\mu^*+A^*+B^*$	μ''	9	2424.5359	AB^*	2–1	2	3.5517	0.169
3	$\mu+A+B$	μ^*+B^*	μ''	8	2444.0881	A^*	3–2	1	19.5522	10^{-5}
4	$\mu+A+B$	$\mu+A^*$	μ''	7	2446.1860	B^*	4–2	2	21.6501	2.10^{-5}
5	$\mu+A+B$	μ^*	μ''	6	2475.4890	$A^*+B^*+AB^*$	5–1	5	54.5049	0
1	$\mu+A+B$	$\mu^*+A^*+B^*$	$\mu''+A''+B''+AB''$	14	2418.1126					
2	$\mu+A+B$	$\mu^*+A^*+B^*$	$\mu''+A''+B''$	12	2418.1783	AB''	2–1	2	0.0657	0.968
3	$\mu+A+B$	$\mu^*+A^*+B^*$	$\mu''+B''$	11	2421.5106	A''	3–2	1	3.3323	0.068
4	$\mu+A+B$	$\mu^*+A^*+B^*$	$\mu''+A''$	10	2421.9895	B''	4–2	2	3.8112	0.149
5	$\mu+A+B$	$\mu^*+A^*+B^*$	μ''	9	2424.5359	$A''+B''+AB''$	5–1	5	6.4233	0.267

[a] Covariates: μ: intercept; A: 1st factor (eg, row); B: 2nd factor (eg, column); AB: interaction. Covariates coded as: location (eg, A), residual (A^*), u to residual variance log-ratio (A''). [b] par: number of parameters: L = maximum of the log-likelihood (REML version). [c] Likelihood ratio (LR) test based on difference in $-2L$ between the full and reduced models; P-value based on the asymptotic chi-square distribution of the LR statistic under the null hypothesis.

Table II. Likelihood statistics for testing location and dispersion parameters in heteroskedastic mixed models: location parameters

No	Location	Model[a] Residual	u- or Ratio	Likelihood[b] par	−2L	Covariate	Test[c] Comp	df	Statistic	P-value
1	$\mu+A+B+AB$	$\mu^*+A^*+B^*+AB^*$	μ'	13	2431.7779					
2	$\mu+A+B$	$\mu^*+A^*+B^*+AB^*$	μ'	11	2435.5389	AB	2-1	2	3.7610	0.153
3	$\mu+B$	$\mu^*+A^*+B^*+AB^*$	μ'	10	2445.0835	A	3-2	1	9.5446	0.0020
4	$\mu+A$	$\mu^*+A^*+B^*+AB^*$	μ'	9	2467.1045	B	4-2	2	31.5656	$1.4.10^{-7}$
1	$\mu+A+B+AB$	$\mu^*+A^*+B^*+AB^*$	μ''	13	2427.8476					
2	$\mu+A+B$	$\mu^*+A^*+B^*+AB^*$	μ''	11	2431.6372	AB	2-1	2	3.7896	0.150
3	$\mu+B$	$\mu^*+A^*+B^*+AB^*$	μ''	10	2442.2131	A	3-2	1	10.5759	0.0011
4	$\mu+A$	$\mu^*+A^*+B^*+AB^*$	μ''	9	2457.5191	B	4-2	2	25.8819	$2.4.10^{-6}$
1	$\mu+A+B+AB$	$\mu^*+A^*+B^*$	μ''	11	2431.8649					
2	$\mu+A+B$	$\mu^*+A^*+B^*$	μ''	9	2435.5829	AB	2-1	2	3.7180	0.156
3	$\mu+B$	$\mu^*+A^*+B^*$	μ''	8	2445.5092	A	3-2	1	9.9263	0.0016
4	$\mu+A$	$\mu^*+A^*+B^*$	μ''	7	2462.1999	B	4-2	2	26.6170	$1.7.10^{-6}$

[a] Covariates: μ: intercept; A: 1st factor (eg, row); B: 2nd factor (eg, column); AB: interaction. Covariates coded as: (eg, A), residual variance (A*), u-component (A$'$) and u to residual variance log-ratio (A$''$). [b] par: number of parameters; L = maximum of the loglikelihood (ML version). [c] Likelihood ratio (LR) test based on difference in −2L between the full and reduced models; P-value based on the asymptotic chi-square distribution of the LR statistic under the null hypothesis.

Table IV. ML and REML estimates of dispersion parameters (a) and subclass variances (b) under two different models.

(a)

No	Model	Component	Method of estimation	Parameter			
				Intercept	$a_2 - a_1$	$b_2 - b_1$	$b_3 - b_1$
I	Residual	$\mu^* + A^* + B^*$	REML	5.955404	0.82921	−0.67086	0.27739
			ML	5.93191	0.85011	−0.67577	0.28717
	Ratio[a]	$\mu'' + A''$	REML	−2.76809	2.05948		
			ML	−3.43898	2.49856		
II	Residual	$\mu^* + A^* + B^*$	REML	5.94316	0.85746	−0.67391	0.30203
			ML	5.92066	0.87583	−0.67257	0.30706
	Ratio	μ''	REML	−1.11978			
			ML	−1.31294			

[a] $\tau_i = \sigma_{u_i}/\sigma_{e_i}$: estimates of parameters are expressed on a log-scale.

(b)

No	Model	Component	Method	Subclass					
				11	12	13	21	22	23
I	Resid	$\mu^* + A^* + B^*$	REML	385.305	196.995	508.480	882.932	441.516	1165.187
			ML	376.874	191.740	502.242	881.848	448.652	1175.196
	Ratio[b]	$\mu'' + A''$	REML	0.0628	0.0628	0.0628	0.4923	0.4923	0.4923
			ML	0.0321	0.0321	0.0321	0.3905	0.3905	0.3905
II	Resid	$\mu^* + A^* + B^*$	REML	381.135	194.269	515.521	898.398	457.923	1215.168
			ML	372.657	190.201	506.600	894.701	456.649	1216.282
	Ratio[b]	μ''	REML	0.3264	0.3264	0.3264	0.3264	0.3264	0.3264
			ML	0.2690	0.2690	0.2690	0.2690	0.2690	0.2690

In both instances I and II, mode for location parameters is $\mu + A + B$. [b] True value = 0.338.

generated by the EM algorithm are strikingly natural (see *Appendix B*) thus giving a flavour of simplicity to the whole procedure.

It also makes it possible to switch from REML to ML or vice versa with very little change in implementing computations (Foulley et al, 1994). Some authors such as Leonard (1975), Denis (1983) and Anderson (1984) in statistics and Shaw (1987) in quantitative genetics have advocated the use of ML rather than REML procedures to estimate variance components. Although the interest of ML versus REML in that case remains questionable, ML estimates remain mandatory for hypothesis testing about β via the LR statistic (see the numerical example). Bayesian point estimators can also be envisioned via EM (Foulley et al, 1992; Gianola et al, 1992; San Cristobal et al, 1993; Weigel and Gianola, 1993; Foulley and Quaas, 1995).

In addition, as already pointed out by Denis et al (1996), a LR test about β requires δ and τ (or λ) being the same for the null hypothesis and its alternative; the same rule holds in hypothesis testing about δ (or λ) by keeping the other parameters the same over the models to be compared. This is part of the general

and difficult problem of joint modelling of means and variances, which is related to such issues as the Behrens–Fisher problem and multi-stage hypothesis testing, and which needs further consideration.

Another area that deserves caution and further development is that of estimability. Difficulties are expected when all the cells contributing to an element j of δ or λ (or to a linear combination of them) have a weight tending towards zero. This may arise due to i) purely overparameterization problems, or due to ii) parameter values becoming extreme (eg, ratios τ_i tending to zero implying elements of λ being infinite). This last phenomenon is similar to what happens in the analysis of binary and ordinal data with latent variable models (Misztal et al, 1989; Fahrmeir and Tutz, 1994). Such difficulties can be avoided by reparameterization (i), or by setting lower bounds to the diagonal elements of the system of equations to solve (ii).

Finally, asymptotic accuracy can also be worked out numerically within the EM framework via the so-called SEM algorithm (Meng and Rubin, 1991). Although it only requires the code of the complete data variance covariance matrix and of the EM or ECM outputs, the burden of calculations is then heavier, so that we suggest that it is restricted to the simplest models. Other computing alternatives should also be considered, eg, the average information-restricted maximum likelihood algorithm (AI-REML) of Gilmour et al (1995).

ACKNOWLEDGMENTS

REFERENCES

Anderson TW (1984) An Introduction to Multivariate Statistical Analysis. *J Wiley and Sons*, New York

Dempster AP, Laird NM, Rubin DB (1977) Maximum likelihood from incomplete data via the EM algorithm. *J R Statist Soc B* 39, 1–38

Denis JB (1983) Extension du modèle additif d'analyse de variance par modélisation multiplicative des variances. *Biometrics* 39, 849–856

Denis JB, Piepho HP, van Eeuwijk A (1996) Mixed models for genotype by environment tables with an emphasis on heteroskedasticity. Technical Report, Laboratoire de Biométrie

DeStefano AL (1994) Identifying and quantifying sources of heterogeneous residual and sire variances in dairy production data. PhD thesis, Cornell University, Ithaca, New York

Fahrmeir L, Tutz G (1994) *Multivariate Statistical Modelling Based on Generalized Linear Models*. Springer Verlag, Berlin

Foulley JL, Gianola D, San Cristobal M, Im S (1990) A method for assessing extent and sources of heterogeneity of residual variances in mixed linear models. *J Dairy Sci* 73, 1612–1624

Foulley JL, San Cristobal M, Gianola D, Im S (1992) Marginal likelihood and Bayesian approaches to the analysis of heterogeneous residual variances in mixed linear Gaussian models. *Comput Stat Data Anal* 13, 291–305

Foulley JL, Hébert D, Quaas RL (1994) Inferences on homogeneity of between family components of variance and covariance among environments in balanced cross-classified designs. *Genet Sel Evol* 26, 117-136

Foulley JL, Quaas RL (1995) Heterogeneous variances in Gaussian linear mixed models. *Genet Sel Evol* 27, 211–228

Garrick DJ, Pollak EJ, Quaas RL, Van Vleck LD (1989) Variance heterogeneity in direct and maternal weight traits by sex and percent purebred for Simmental sired calves. *J Anim Sci* 67, 2515–2528

Gianola D, Foulley JL, Fernando RL, Henderson CR, Weigel KA (1992) Estimation of heterogeneous variances using empirical Bayes methods: theoretical considerations. *J Dairy Sci* 75, 2805–2823

Gilmour AR, Thompson R, Cullis BR (1995) Average information REML: an efficient algorithm for variance parameter estimation in linear mixed models. *Biometrics* 51, 1440–1450

Henderson CR (1984) *Applications of Linear Models in Animal Breeding*. University of Guelph, Guelph, Ontario, Canada.

Lange K (1995) A gradient algorithm locally equivalent to the EM algorithm. *J R Statist Soc B* 57, 425–437

Laird NM, Lange N, Stram D (1987) Maximum likelihood computations with repeated measures: application to the EM algorithm. *J Am Statist Assoc* 82, 97–105

Leonard T (1975) A Bayesian approach to the linear model with unequal variances. *Technometrics* 17, 95–102

Meuwissen THE, De Jong G, Engel B (1996) Joint estimation of breeding values and heterogeneous variances of large data files. *J Dairy Sci* 79, 310–316

Misztal I, Gianola D, Foulley JL (1989) Computing aspects of a non linear method of sire evaluation for categorical data. *J Dairy Sci* 72, 1557–1568

Meng XL., Rubin DB (1991) Using EM to obtain asymptotic variance-covariance matrices: the SEM algorithm. *J Am Stat Assoc* 86, 899–909

Meng XL, Rubin DB (1993) Maximum likelihood estimation via the ECM algorithm: A general framework. *Biometrika* 80, 267–278

Meyer K, Carrick J, Donnelly BJP (1993) Genetic parameters for growth traits of Australian beef cattle from a multibreed selection experiment. *J Anim Sci* 71, 2614–2622

Patterson HD, Thompson R (1971) Recovery of interblock information when block sizes are unequal. *Biometrika* 58, 545–554

Quaas RL (1992) RL. REML Note book, Mimeo, Cornell University, Ithaca, New York

Robert C, Foulley JL, Ducrocq V (1995a) Inference on homogeneity of intra-class correlations among environments using heteroskedastic models. *Genet Sel Evol* 27, 51–65

Robert C, Foulley JL, Ducrocq V (1995b) Estimation and testing of constant genetic and intra-class correlation coefficients among environments. *Genet Sel Evol* 27, 125–134

San Cristobal M, Foulley JL, Manfredi E (1993) Inference about multiplicative heteroskedastic components of variance in a mixed linear Gaussian model with an application to beef cattle breeding. *Genet Sel Evol* 25, 3–30

Searle SR, Casella G, McCulloch CE (1992) *Variance Components*. J Wiley and Sons, New-York

Shaw RG (1987) Maximum likelihood approaches applied to quantitative genetics of natural populations. *Evolution* 45, 143–151

Visscher PM (1992) On the power of likelihood ratio tests for detecting heterogeneity of intra-class correlations and variances in balanced half-sib designs. *J Dairy Sci* 75, 1320–1330

Visscher PM, Thompson R, Hill WG (1991) Estimation of genetic and environmental variances for fat yield in individual herds and an investigation into heterogeneity of variance between herds. *Livest Prod Sci* 28, 273–290

Visscher PM, Hill WG (1992) Heterogeneity of variance and dairy cattle breeding. *Anim Prod* 55, 321–329

Walker S (1996) An EM algorithm for non linear random effects model. *Biometrics* 52, 934–944

Weigel KA, Gianola D (1993) A computationally simple Bayesian method for estimation of heterogeneous within-herd phenotypic variances. *J Dairy Sci* 76, 1455–1465

Weigel KA, Gianola D, Yandel BS, Keown JF (1993) Identification of factors causing heterogeneous within-herd variance components using a structural model for variances. *J Dairy Sci* 76, 1466–1478

Zangwill W (1969) *Non Linear Programming: A Unified Approach.* Prentice Hall, Englewood Cliffs

APPENDIX A: Algebra for the estimating equations

The Q function to be maximized is (in condensed notation)

$$Q(\gamma) = \text{const} - \frac{1}{2}\sum_{i=1}^{I} n_i \ln \sigma_{e_i}^2 - \frac{1}{2}\sum_{i=1}^{I} \sigma_{e_i}^{-2}\, \mathrm{E}_c(\mathbf{e}_i'\mathbf{e}_i) \qquad [A1]$$

with

$$\ln \sigma_{e_i}^2 = \mathbf{p}_i'\boldsymbol{\delta} \qquad [A2]$$

and

$$\mathbf{e}_i = \mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta} - \tau\sigma_{e_i}\mathbf{Z}_i\mathbf{u}^* \qquad [A3]$$

Derivatives with respect to δ (residual dispersion parameters)

First derivatives: according to the chaining rule, one has

$$\frac{\partial(2Q)}{\partial\boldsymbol{\delta}} = \sum_{i=1}^{I} \frac{\partial(2Q)}{\partial \ln \sigma_{e_i}^2}\,\frac{\partial \ln \sigma_{e_i}^2}{\partial\boldsymbol{\delta}}$$

Now

$$\frac{\partial(2Q)}{\partial \ln \sigma_{e_i}^2} = \sigma_{e_i}^2\,\frac{\partial(2Q)}{\partial\sigma_{e_i}^2}$$

$$\frac{\partial \ln \sigma_{e_i}^2}{\partial\boldsymbol{\delta}} = \mathbf{p}_i$$

That is

$$\frac{\partial(2Q)}{\partial\sigma_{e_i}^2} = -\frac{n_i}{\sigma_{e_i}^2} - \frac{\partial\left[\sigma_{e_i}^{-2}\,\mathrm{E}_c(\mathbf{e}_i'\mathbf{e}_i)\right]}{\partial\sigma_{e_i}^2}$$

$$\frac{\partial\left[\sigma_{e_i}^{-2}\,\mathrm{E}_c(\mathbf{e}_i'\mathbf{e}_i)\right]}{\partial\sigma_{e_i}^2} = -\frac{\mathrm{E}_c(\mathbf{e}_i'\mathbf{e}_i)}{\sigma_{e_i}^4} + \frac{1}{\sigma_{e_i}^2}\frac{\partial\mathrm{E}_c(\mathbf{e}_i'\mathbf{e}_i)}{\partial\sigma_{e_i}^2}$$

$$\frac{\partial\mathrm{E}_c(\mathbf{e}_i'\mathbf{e}_i)}{\partial\sigma_{e_i}^2} = \frac{1}{2\sigma_{e_i}}\frac{\partial\mathrm{E}_c(\mathbf{e}_i'\mathbf{e}_i)}{\sigma_{e_i}} = \frac{1}{\sigma_{e_i}}\mathrm{E}_c\left[\left(\frac{\partial\mathbf{e}_i'}{\partial\sigma_{e_i}}\right)\mathbf{e}_i\right]$$

and

$$\frac{\partial\mathbf{e}_i'}{\partial\sigma_{e_i}} = -\tau\mathbf{u}^{*'}\mathbf{Z}_i'$$

Thus

$$\frac{\partial\mathrm{E}_c(\mathbf{e}_i'\mathbf{e}_i)}{\partial\sigma_{e_i}^2} = -\frac{\tau\mathrm{E}_c(\mathbf{u}^{*'}\mathbf{Z}_i'\mathbf{e}_i)}{\sigma_{e_i}}$$

and

$$\frac{\partial(2Q)}{\partial\sigma_{e_i}^2} = -\frac{n_i}{\sigma_{e_i}^2} + \frac{\mathrm{E}_c(\mathbf{e}_i'\mathbf{e}_i)}{\sigma_{e_i}^4} - \frac{\tau\mathrm{E}_c(\mathbf{u}^{*'}\mathbf{Z}_i'\mathbf{e}_i)}{\sigma_{e_i}^3}$$

Letting $\nu_{\delta,i} = \partial(2Q)/\partial\ln\sigma_{e_i}^2$

$$\nu_{\delta,i} = \frac{\mathrm{E}_c(\mathbf{e}_i'\mathbf{e}_i)}{\sigma_{e_i}^2} - n_i + \frac{\tau}{\sigma_{e_i}}\mathrm{E}_c(\mathbf{u}^{*'}\mathbf{Z}_i'\mathbf{e}_i)$$

$$= \frac{1}{\sigma_{e_i}^2}\mathrm{E}_c\left[(\mathbf{e}_i + \tau\sigma_{e_i}\mathbf{Z}_i\mathbf{u}^*)'\,\mathbf{e}_i\right] - n_i$$

so that

$$\frac{\partial(2Q)}{\partial\boldsymbol{\delta}} = \sum_{i=1}^{I}\nu_{\delta,i}\mathbf{p}_i = \mathbf{P}'\mathbf{v}_\delta \qquad\qquad [\text{A4a}]$$

with

$$\mathbf{v}_\delta = \left\{\frac{\mathrm{E}_c\left[(\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta})'\,\mathbf{e}_i\right]}{\sigma_{e_i}^2} - n_i\right\} \qquad\qquad [\text{A4b}]$$

Second derivatives: from [A4a], one has

$$\frac{\partial(2Q)}{\partial\boldsymbol{\delta}\partial\boldsymbol{\delta}'} = \sum_{i=1}^{I}\mathbf{p}_i'\frac{\partial\nu_{\delta,i}}{\partial\ln\sigma_{e_i}^2}\mathbf{p}_i$$

Now, using [A4b]

$$\frac{\partial\nu_{\delta,i}}{\partial\sigma_{e_i}^2} = -\frac{1}{\sigma_{e_i}^4}\mathrm{E}_c\left[(\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta})'\,\mathbf{e}_i\right] + \frac{1}{\sigma_{e_i}^2}\frac{\partial\mathrm{E}_c\left[(\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta})'\,\mathbf{e}_i\right]}{\partial\sigma_{e_i}^2}$$

$$\frac{\partial E_c \left[(\mathbf{y}_i - \mathbf{X}_i \boldsymbol{\beta})' \mathbf{e}_i \right]}{\partial \sigma_{e_i}^2} = \frac{1}{2\sigma_{e_i}} E_c \left[(\mathbf{y}_i - \mathbf{X}_i \boldsymbol{\beta})' \frac{\partial \mathbf{e}_i}{\partial \sigma_{e_i}} \right] = -\frac{\tau}{2\sigma_{e_i}} E_c \left[(\mathbf{y}_i - \mathbf{X}_i \boldsymbol{\beta})' \mathbf{Z}_i \mathbf{u}^* \right]$$

Thus, letting

$$\frac{\partial (2Q)}{\partial \boldsymbol{\delta} \partial \boldsymbol{\delta}'} = -\sum_{i=1}^{I} w_{\delta\delta,ii} \mathbf{p}_i' \mathbf{p}_i = -\mathbf{P}' \mathbf{W} \mathbf{P} \qquad \text{[A5a]}$$

$$w_{\delta\delta,ii} = \frac{1}{\sigma_{e_i}^2} E_c \left[(\mathbf{y}_i - \mathbf{X}_i \boldsymbol{\beta})' \left(\mathbf{e}_i + \frac{1}{2} \tau \sigma_{e_i} \mathbf{Z}_i \mathbf{u}^* \right) \right] \qquad \text{[A5b]}$$

or, alternatively

$$w_{\delta\delta,ii} = \frac{1}{\sigma_{e_i}^2} E_c \left[(\mathbf{y}_i - \mathbf{X}_i \boldsymbol{\beta})' (\mathbf{y}_i - \mathbf{X}_i \boldsymbol{\beta}) \right] - \frac{\tau}{2\sigma_{e_i}} E_c \left[(\mathbf{y}_i - \mathbf{X}_i \boldsymbol{\beta})' \mathbf{Z}_i \mathbf{u}^* \right] \qquad \text{[A5c]}$$

Finally, the non-linear system to solve reduces to

$$\mathbf{P}' \mathbf{W}_{\delta\delta} \mathbf{P} \cdot \boldsymbol{\Delta\delta} = \mathbf{P}' \mathbf{v}_{\delta} \qquad \text{[A6]}$$

Derivatives with respect to τ (ratio)

$$\frac{\partial Q}{\partial \tau} = -\frac{1}{2} \sum_{i=1}^{I} \frac{1}{\sigma_{e_i}^2} \frac{\partial \left[E_c \left(\mathbf{e}_i' \mathbf{e}_i \right) \right]}{\partial \tau}$$

Now

$$\frac{\partial E_c (\mathbf{e}_i' \mathbf{e}_i)}{\partial \tau} = 2 E_c \left[\left(\frac{\partial \mathbf{e}_i'}{\partial \sigma_{e_i}} \right) \mathbf{e}_i \right]$$

and

$$\frac{\partial \mathbf{e}_i'}{\partial \tau} = -\sigma_{e_i} \mathbf{u}^{*'} \mathbf{Z}_i'$$

Thus

$$\frac{\partial Q}{\partial \tau} = \sum_{i=1}^{I} \frac{1}{\sigma_{e_i}} E_c (\mathbf{u}^{*'} \mathbf{Z}_i' \mathbf{e}_i)$$

$$= \sum_{i=1}^{I} \frac{1}{\sigma_{e_i}} E_c \left[\mathbf{u}^{*'} \mathbf{Z}_i' (\mathbf{y}_i - \mathbf{X}_i \boldsymbol{\beta}) \right] - \tau \sum_{i=1}^{I} E_c (\mathbf{u}^{*'} \mathbf{Z}_i' \mathbf{Z}_i \mathbf{u}^*) \qquad \text{[A7]}$$

The equation $\partial Q / \partial \tau = 0$ results in

$$\tau = \frac{\displaystyle\sum_{i=1}^{I} \sigma_{e_i}^{-1} E_c \left[\mathbf{u}^{*'} \mathbf{Z}_i' (\mathbf{y}_i - \mathbf{X}_i \boldsymbol{\beta}) \right]}{\displaystyle\sum_{i=1}^{I} E_c (\mathbf{u}^{*'} \mathbf{Z}_i' \mathbf{Z}_i \mathbf{u}^*)} \qquad \text{[A8]}$$

Additional derivatives for the exact EM

From the second expression in [A7] it follows immediately

$$W_{\tau\tau} = -\frac{\partial^2(2Q)}{\partial\tau^2} = 2\sum_{i=1}^{I} E_c(\mathbf{u}^{*'}\mathbf{Z}_i'\mathbf{Z}_i\mathbf{u}^*) \qquad [A9]$$

One has also to express

$$\frac{\partial^2(2Q)}{\partial\tau\partial\delta} = \sum_{i=1}^{I} \frac{\partial^2 Q}{\partial\tau\partial\sigma_{e_i}}\sigma_{e_i}\mathbf{P}_i$$

Now, from [A7]

$$\frac{\partial^2 Q}{\partial\tau\partial\sigma_{e_i}} = -\frac{1}{\sigma_{e_i}^2} E_c\left[\mathbf{u}^{*'}\mathbf{Z}_i'(\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta})\right]$$

so that

$$-\frac{\partial^2(2Q)}{\partial\tau\partial\delta} = \sum_{i=1}^{I} w_{\tau\delta,i}\mathbf{P}_i = \mathbf{P}'\mathbf{W}_{\tau\delta} \qquad [A10a]$$

where $\mathbf{W}_{\tau\delta}$ is a $(I \times 1)$ defined by

$$\mathbf{W}_{\tau\delta} = \{w_{\tau\delta,i}\} = \left\{\frac{1}{\sigma_{e_i}} E_c\left[\mathbf{u}^{*'}\mathbf{Z}_i'(\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta})\right]\right\} \qquad [A10b]$$

The system for true EM (or gradient EM without inner iteractions) is then

$$\begin{bmatrix} \mathbf{P}'\mathbf{W}_{\delta\delta}\mathbf{P} & \mathbf{P}'\mathbf{W}_{\tau\delta} \\ \mathbf{W}_{\tau\delta}'\mathbf{P} & W_{\tau\tau} \end{bmatrix}\begin{bmatrix} \boldsymbol{\Delta}\boldsymbol{\delta} \\ \boldsymbol{\Delta}\boldsymbol{\tau} \end{bmatrix} = \begin{bmatrix} \mathbf{P}'\mathbf{v}_\delta \\ \nu_\tau \end{bmatrix} \qquad [A11]$$

where $\nu_\tau = 2\dfrac{\partial Q}{\partial\tau}$, see [A7]

Extension to K independent random factors

Q in [A1] remains formally unchanged with $\mathbf{e}_i$ in [A3] being now

$$\mathbf{e}_i = \mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta} - \sigma_{e_i}\sum_{k=1}^{I}\tau_k\mathbf{Z}_{ik}\mathbf{u}_k^* \qquad [A12]$$

Based on this expression of the residual, formulae [A4ab] still hold.
Similarly, the expression of $w_{\delta\delta,ii}$ in [A5b] becomes

$$w_{\delta\delta,ii} = \frac{1}{\sigma_{e_i}^2} E_c\left[(\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta})'\left(\mathbf{e}_i + \frac{\sigma_{e_i}}{2}\sum_{k=1}^{K}\tau_k\mathbf{Z}_{ik}\mathbf{u}_k^*\right)\right] \qquad [A13a]$$

or, alternatively

$$w_{\delta\delta,ii} = \frac{1}{\sigma_{e_i}^2} E_c\left[(\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta})'(\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta})\right] - \frac{1}{2\sigma_{e_i}} \sum_{k=1}^{K} \tau_k E_c\left[(\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta})'\mathbf{Z}_{ik}\mathbf{u}_k^*\right]$$

[A13b]

As far as τ is concerned, [A6] becomes

$$\frac{\partial Q}{\partial \tau_k} = \sum_{i=1}^{I} \frac{1}{\sigma_{e_i}} E_c(\mathbf{u}_k^{*'}\mathbf{Z}_{ik}'\mathbf{e}_i)$$

[A14]

leading to the following system

$$\tau_k \sum_{i=1}^{I} E_c(\mathbf{u}_k^{*'}\mathbf{Z}_{ik}'\mathbf{Z}_{ik}\mathbf{u}_k^*) + \sum_{\ell \neq k} \tau_\ell \sum_{i=1}^{I} E_c(\mathbf{u}_k^{*'}\mathbf{Z}_{ik}'\mathbf{Z}_{i\ell}\mathbf{u}_\ell^*)$$

$$= \sum_{i=1}^{I} \sigma_{e_i}^{-1} E_c(\mathbf{u}_k^{*'}\mathbf{Z}_{ik}'\mathbf{Z}_{i\ell}\mathbf{u}_\ell^*)$$

[A15]

for $k = 1, 2, \ldots, K$

Derivatives with respect to λ (parameters of the log-ratio)

Q and the model for $\ln \sigma_{e_i}^2$ are the same as in [A1] and [A2] but the vector of residuals is defined as

$$\mathbf{e}_i = \mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta} - \tau_i \sigma_{e_i} \mathbf{Z}_i \mathbf{u}^*$$

[A16]

with

$$\ln \tau_i = \mathbf{h}_i'\boldsymbol{\lambda}$$

[A17]

Using condensed notation, the iterative system to solve can be written in the same was as previously, ie

$$\mathbf{H}'\mathbf{W}_{\lambda\lambda}\mathbf{H} \cdot \Delta\boldsymbol{\lambda} = \mathbf{H}'\mathbf{v}_\lambda$$

[A18]

where

$$\mathbf{H}'_{(s \times 1)} = (\mathbf{h}_1, \mathbf{h}_2, \ldots, \mathbf{h}_i, \ldots, \mathbf{h}_I)$$

$$\mathbf{v}_\lambda = \{\nu_{\lambda,i} = \partial Q / \partial \ln \tau_i\}$$

[A19]

$$\mathbf{W}_{\lambda\lambda} = \text{Diag}\left\{w_{\lambda\lambda,ii} = -\frac{\partial \nu_{\lambda,i}}{\partial \ln \tau_i}\right\}$$

[A20]

Now

$$\nu_{\lambda,i} = \tau_i \frac{\partial Q}{\partial \tau_i} = \frac{\tau_i}{\sigma_{e_i}} E_c(\mathbf{u}^{*'}\mathbf{Z}_i'\mathbf{e}_i)$$

[A21a]

or, after some algebra

$$\nu_{\lambda,i} = \frac{\tau_i}{\sigma_{e_i}} E_c\left[\mathbf{u}^{*'}\mathbf{Z}_i'(\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta})\right] - \tau_i^2 E_c(\mathbf{u}^{*'}\mathbf{Z}_i'\mathbf{Z}_i\mathbf{u}^*) \qquad \text{[A21b]}$$

Furthermore,

$$\frac{\partial \nu_{\lambda,i}}{\partial \tau_i} = \frac{1}{\sigma_{e_i}} E_c(\mathbf{u}^{*'}\mathbf{Z}_i'\mathbf{e}_i) + \frac{\tau_i}{\sigma_{e_i}} E_c\left(\mathbf{u}^{*'}\mathbf{Z}_i'\frac{\partial \mathbf{e}_i}{\partial \tau_i}\right)$$

$$\frac{\partial \mathbf{e}_i}{\partial \tau_i} = -\sigma_{e_i}\mathbf{Z}_i\mathbf{u}^*$$

Then

$$w_{\lambda\lambda,ii} = -\tau_i\frac{\partial \nu_{\lambda,i}}{\partial \tau_i}$$

$$w_{\lambda\lambda,ii} = -\frac{\tau_i}{\sigma_{e_i}} E_c(\mathbf{u}^{*'}\mathbf{Z}_i'\mathbf{e}_i) + \tau_i^2 E_c(\mathbf{u}^{*'}\mathbf{Z}_i'\mathbf{Z}_i\mathbf{u}^*) \qquad \text{[A22a]}$$

or

$$w_{\lambda\lambda,ii} = 2\tau_i^2 E_c(\mathbf{u}^{*'}\mathbf{Z}_i'\mathbf{Z}_i\mathbf{u}^*) - \frac{\tau_i}{\sigma_{e_i}} E_c\left[\mathbf{u}^{*'}\mathbf{Z}_i'(\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta})\right] \qquad \text{[A22b]}$$

This can be easily generalized to several independent random factors

$$\ln \tau_{ik} = \mathbf{h}_{ik}'\boldsymbol{\lambda}_k \qquad \text{for} \qquad k = 1, 2, \ldots, K$$

if conditional maximization is performed factor by factor. The system [A18] applies to each factor k with

$$\nu_{\lambda,i}^{(k)} = \frac{\tau_{ik}}{\sigma_{e_i}} E_c(\mathbf{u}_k^{*'}\mathbf{Z}_{ik}'\mathbf{e}_i) \qquad \text{[A23a]}$$

$$w_{\lambda\lambda,ii}^{(k)} = -\frac{\tau_{ik}}{\sigma_{e_i}} E_c(\mathbf{u}_k^{*'}\mathbf{Z}_{ik}'\mathbf{e}_i) + \tau_{ik}^2 E_c(\mathbf{u}_k^{*'}\mathbf{Z}_{ik}'\mathbf{Z}_{ik}\mathbf{u}_k^*) \qquad \text{[A23b]}$$

APPENDIX B: Formulae [12] for grouped data

In some instances (see, eg, the example in table I) data can be grouped so that the n_i observations within a stratum i share the same covariates

$$\begin{cases} \mathbf{X}_i = \mathbf{1}_{n_i}\mathbf{x}_i' \\ \mathbf{Z}_i = \mathbf{1}_{n_i}\mathbf{z}_i' \end{cases} \qquad \text{[B1]}$$

where $\mathbf{x}_i'$ and $\mathbf{z}_i'$ are the common row vectors $(1 \times p)$ and $(1 \times q)$ pertaining to fixed and random effects, respectively.

$$\mu_i = \mathbf{x}_i'\boldsymbol{\beta}, \ u_i^* = \mathbf{z}_i'\mathbf{u}^* \qquad \text{[B2]}$$

and

$$
\begin{bmatrix}
\widehat{S}_{\varepsilon\varepsilon,i} = \mathrm{E}_c\left[(\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta})'(\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta})\right] \\
\widehat{S}_{\varepsilon u,i} = \mathrm{E}_c\left[\mathbf{u}^{*'}\mathbf{Z}_i'(\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta})\right] \\
\widehat{S}_{uu,i} = \mathrm{E}_c(\mathbf{u}^{*'}\mathbf{Z}_i'\mathbf{Z}_i\mathbf{u}^*)
\end{bmatrix}
\tag{B3}
$$

one has

$$
\widehat{S}_{\varepsilon\varepsilon,i} = \mathbf{y}_i'\mathbf{y}_i - 2\widehat{\boldsymbol{\beta}}'\mathbf{X}_i'\mathbf{y}_i + \widehat{\boldsymbol{\beta}}'\mathbf{X}_i'\mathbf{X}_i\widehat{\boldsymbol{\beta}} + \mathrm{tr}(\mathbf{X}_i'\mathbf{X}_i\mathbf{C}_{\beta\beta})
$$

Substituting $\mathbf{X}_i$ by its expression in [B1] gives

$$
\begin{aligned}
\widehat{S}_{\varepsilon\varepsilon,i} &= \sum_{j=1}^{n_i}(y_{ij} - \widehat{\mu}_i)^2 + n_i\,\mathrm{tr}(\mathbf{x}_i\mathbf{x}_i'\mathbf{C}_{\beta\beta}) \\
&= \sum_{j=1}^{n_i} y_{ij}^2 + n_i\widehat{\mu}_i^2 - 2\widehat{\mu}_i y_{i+} + n_i\,\mathrm{tr}(\mathbf{x}_i\mathbf{x}_i'\mathbf{C}_{\beta\beta})
\end{aligned}
\tag{B4}
$$

where y_{ij} is the jth element of $\mathbf{y}_i$, and $y_{i+} = \sum_{j=1}^{n_i} y_{ij}$

Similarly

$$
u_i^* = \mathbf{z}_i'\mathbf{u}^*
$$
$$
\widehat{S}_{\varepsilon u,i} = \widehat{u}_i^*(y_{i+} - n_i\widehat{\mu}_i) - n_i\,\mathrm{tr}(\mathbf{z}_i\mathbf{x}_i'\mathbf{C}_{\beta u})
\tag{B5}
$$

and

$$
\widehat{S}_{uu,i} = n_i\left\{\widehat{u}_i^{*2} + \mathrm{tr}(\mathbf{z}_i\mathbf{z}_i'\mathbf{C}_{uu})\right\}
\tag{B6}
$$