Genet. Sel. Evol. 30 (1998) 27–43

Original article

A link function approach to heterogeneous variance components

Jean-Louis Foulley [a*], Richard L. Quaas [b], Caroline Thaon d'Arnoldi [a]

[a] Station de génétique quantitative et appliquée, Institut national de la recherche agronomique, CR de Jouy, 78352 Jouy-en-Josas Cedex, France
[b] Department of Animal Science, Cornell University, Ithaca, NY 14853, USA

(Received 20 June 1997; accepted 4 November 1997)

Abstract – This paper presents techniques of parameter estimation in heteroskedastic mixed models having i) heterogeneous log residual variances which are described by a linear model of explanatory covariates and ii) log residual and log u-components linearly related. This makes the intraclass correlation a monotonic function of the residual variance. Cases of a homogeneous variance ratio and of a homogeneous u-component of variance are also included in this parameterization. Estimation and testing procedures of the corresponding dispersion parameters are based on restricted maximum likelihood procedures. Estimating equations are derived using the standard and gradient EM. The analysis of a small example is outlined to illustrate the theory. © Inra/Elsevier, Paris

heteroskedasticity / mixed model / maximum likelihood / EM algorithm

Résumé – Une approche des composantes de variance hétérogènes par les fonctions de lien. Cet article présente des techniques d'estimation des paramètres intervenant dans des modèles mixtes caractérisés i) par des logvariances résiduelles décrites par un modèle linéaire de covariables explicatives et ii) par des composantes u et e liées par une fonction affine. Cela conduit à un coefficient de corrélation intraclasse qui varie comme une fonction monotone de la variance résiduelle. Le cas d'une corrélation constante et celui d'une composante u constante sont également inclus dans cette paramétrisation. L'estimation et les tests relatifs aux paramètres de dispersion correspondants sont basés sur les méthodes du maximum de vraisemblance restreint (REML). Les équations à résoudre pour obtenir ces estimations sont établies à partir de l'algorithme EM standard et gradient. La théorie est illustrée par l'analyse numérique d'un petit exemple. © Inra/Elsevier, Paris

hétéroscédasticité / modèle mixte / maximum de vraisemblance / algorithme EM

* Correspondence and reprints

1. INTRODUCTION

A previous paper of this series [4], presented an EM–REML (or ML) approach to estimating dispersion parameters for heteroskedastic mixed models. We assumed i) a linear model on log residual (or e) variances, and/or ii) constant u to e variance ratios.

There are different ways to relax this last assumption. The first one is to proceed as with residual variances, i.e. hypothesize that the variation in log u-components or of the u to e-ratio depends on explanatory covariates observed in the experiment, e.g. region, herd, parity, management conditions, etc. This is the so-called structural approach described by San Cristobal et al. [23], and applied by Weigel et al. [28] and De Stefano [2] to milk traits of dairy cattle.

Another procedure consists in assuming that the residual and u-components are directly linked via a relationship which is less restrictive than a constant ratio. A basic motivation for this is that the assumption of homogeneous variance ratios or intra class correlations (e.g. heritability for animal breeders) might be unrealistic [19] although very convenient to set up for theoretical and computational reasons (see the procedure by Meuwissen et al. [16]). As a matter of fact, the power of statistical tests for detecting such heterogeneous heritabilities is expected to be low [25] which may also explain why homogeneity is preferred. The purpose of this second paper is an attempt to describe a procedure of this type which we will call a link function approach referring to its close connection with the parameterization used in GLM theory [3, 14].

The paper will be organized along similar lines as the previous paper [4] including i) an initial section on theory, with a brief summary of the models and a presentation of the estimating equations and testing procedures, and ii) a numerical application based on a small data set with the same structure as the one used in the previous paper [4].

2. THEORY

2.1. Statistical model

It is assumed that the data set can be stratified into several strata indexed by $(i = 1, 2, \ldots, I)$ representing a potential source of heteroskedasticity. For the sake of simplicity, we will consider a standardized one-way random (e.g. sire) model as in Foulley [4] and Foulley and Quaas [5].

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

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 the systematic random part is represented by $\sigma_{u_i}\mathbf{Z}_i\mathbf{u}^*$ 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 $\mathbf{e}_i$, i.e. $\mathbf{u}^* \sim N(\mathbf{0}, \mathbf{A})$, $\mathbf{e}_i \sim N(\mathbf{0}, \sigma^2_{e_i}\mathbf{I}_{n_i})$, and $\mathrm{E}(\mathbf{u}^*\mathbf{e}_i') = \mathbf{0}$.

The influence of factors causing the heteroskedasticity of residual variances is modelled along the lines presented in Leonard [13] and Foulley et al. [6, 7] via a linear regression on log-variances:

$$\ell n\, \sigma^2_{e_i} = \mathbf{p}'_i \boldsymbol{\delta} \tag{2}$$

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.

Residual and u-component parameters are linked via a functional relationship

$$\ln \sigma_{u_i} = a + b \ln \sigma_{e_i} \tag{3a}$$

or equivalently

$$\sigma_{u_i} / \sigma^b_{e_i} = \tau \tag{3b}$$

where the constant τ equals $\exp(a)$.

The differential equation pertaining to [3ab], i.e. $(d\sigma_{u_i}/\sigma_{u_i}) - b(d\sigma_{e_i}/\sigma_{e_i}) = 0$ is a scale-free relationship which shows clearly that the parameter of interest in this model is b. Notice the close connection between the parameterization in equations [2] and [3ab] with that used in the approach of the 'composite link function' proposed by Thompson and Baker [24] whose steps can be summarized as follows: i) $(\sigma_{u_i}, \sigma_{e_i})' = f(a, b, \sigma_{e_i})$; ii) $\sigma_{e_i} = \exp(\eta_i)$, and $\eta_i = (1/2)\mathbf{p}'_i\delta$. As compared to Thompson and Barker, the only difference is that the function f in i) is not linear and involves extra parameters, i.e. a and b.

The intraclass correlation (proportional to heritability for animal breeders)

$$t_i = \sigma^2_{u_i} / (\sigma^2_{u_i} + \sigma^2_{e_i}) = \rho_i/(1 + \rho_i)$$

is an increasing function of the variance ratio $\rho_i = \sigma^2_{u_i}/\sigma^2_{e_i}$. In turn ρ_i increases or decreases with $\sigma^2_{e_i}$ depending on $b > 1$ or $b < 1$, respectively, or remains constant $(b = 1)$ since $d\rho_i/\rho_i = 2(b-1)d\sigma_{e_i}/\sigma_{e_i}$. Consequently the intraclass correlation increases or decreases with the residual variance or remains constant $(b = 1)$. For $b = 0$, the u-component is homogeneous *figure 1.*

$\sigma^2_{e_i}$	b	t	$\sigma^2_{u_i}$
$\uparrow$	>1	$\uparrow$	
$\uparrow$	1	constant	
$\uparrow$	$(0,1)$	$\downarrow$	$\uparrow$
$\uparrow$	0		constant
$\uparrow$	<0		$\downarrow$

Figure 1. t and $\sigma^2_{u_i}$ versus $\sigma^2_{e_i}$ as a function of b.

2.2. EM–REML estimation

The basic EM–REML procedure [1, 18] proposed by Foulley and Quaas (1995) for heterogeneous variances is applied here.

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

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

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

as shown in Foulley and Quaas [5], reduces to

$$Q(\boldsymbol{\gamma}|\boldsymbol{\gamma}^{[t]}) = \text{const} - \frac{1}{2}\sum_{i=1}^{I} n_i \,\ell n\, \sigma^2_{e_i} - \frac{1}{2}\sum_{i=1}^{I} \sigma^{-2}_{e_i} \mathrm{E}^{[t]}_c \left(\mathbf{e}'_i \mathbf{e}_i\right) \tag{4}$$

where $\mathrm{E}^{[t]}_c(.)$ is a condensed notation for a conditional expectation taken with respect to the distribution of $\mathbf{x}$ in Q given the data vector $\mathbf{y}$ and $\boldsymbol{\gamma} = \boldsymbol{\gamma}^{[t]}$.

Given the current estimate $\boldsymbol{\gamma}^{[t]}$ of $\boldsymbol{\gamma}$, the M-step consists in calculating the next value $\boldsymbol{\gamma}^{[t+1]}$ by maximizing $Q(\boldsymbol{\gamma}|\boldsymbol{\gamma}^{[t]})$ in equation (4) with respect to the elements of the vector $\boldsymbol{\gamma}$ of unknowns. This can be accomplished efficiently via the Newton-Raphson algorithm. The system of equations to solve iteratively can be written in matrix form as:

$$\begin{bmatrix} \mathbf{P}'\mathbf{W}_{\delta\delta}\mathbf{P} & \mathbf{P}'\mathbf{W}_{\delta\tau} & \mathbf{P}'\mathbf{W}_{\delta b} \\ \mathbf{W}'_{\delta\tau}\mathbf{P} & w_{\tau\tau} & w_{\tau b} \\ \mathbf{W}'_{\delta b}\mathbf{P} & w_{\tau b} & w_{bb} \end{bmatrix}_{\boldsymbol{\gamma}=\boldsymbol{\gamma}^{[t,\ell]}} \begin{bmatrix} \Delta\boldsymbol{\delta} \\ \Delta\tau \\ \Delta b \end{bmatrix}^{[t,\ell+1]} = \begin{bmatrix} \mathbf{P}'\mathbf{v}_\delta \\ v_\tau \\ v_b \end{bmatrix}_{\boldsymbol{\gamma}=\boldsymbol{\gamma}^{[t,\ell]}} \tag{5}$$

where $\mathbf{P}'_{(r\times I)} = (\mathbf{p}_1, \mathbf{p}_2, \cdots, \mathbf{p}_i, \cdots, \mathbf{p}_I)$; $\mathbf{v}_{\delta[I\times 1]} = \left\{\partial Q/\partial \ell n\, \sigma^2_{e_i}\right\}$, $v_\tau = \{\partial Q/\partial \tau\}$, $v_b = \{\partial Q/\partial b\}$; $W_{\alpha\beta} = \partial Q/\partial\boldsymbol{\alpha}\partial\boldsymbol{\beta}'$, for $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ being components of $\boldsymbol{\gamma} = (\boldsymbol{\delta}', \tau, b)'$.

Note that for this algorithm to be a true EM, one would have to iterate the NR algorithm in equation (5) within an inner cycle (index ℓ) until convergence to the conditional maximizer $\boldsymbol{\gamma}^{[t+1]} = \boldsymbol{\gamma}^{[t,c]}$ at each M step. In practice it may be advantageous to reduce the number of inner iterations, even up to only one. This is an application of the so called 'gradient EM' algorithm the convergence properties of which are almost identical to standard EM [12].

The algebra for the first and second derivatives is given in the Appendix. These derivatives are functions of the current estimates of the parameters $\boldsymbol{\gamma} = \boldsymbol{\gamma}^{[t]}$, and of the components of $\mathrm{E}^{[t]}_c(\mathbf{e}'_i\mathbf{e}_i)$ defined at the E-step.

Let those components be written under a condensed form as:

$$\left.\begin{aligned} S_{i,\varepsilon\varepsilon} &= \boldsymbol{\varepsilon}'_i\boldsymbol{\varepsilon}_i = (\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta})'(\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta}) \\ S_{i,u\varepsilon} &= \mathbf{u}^{*'}\mathbf{Z}'_i(\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta}) \\ S_{i,uu} &= \mathbf{u}^{*'}\mathbf{Z}'_i\mathbf{Z}_i\mathbf{u}^* \end{aligned}\right\} \tag{6}$$

with a cap for their conditional expectations, e.g.

$$\widehat{S}_{i,\varepsilon\varepsilon}^{[t]} = \mathrm{E}_c^{[t]}(S_{i,\varepsilon\varepsilon}) = \mathrm{E}(S_{i,\varepsilon\varepsilon}|\mathbf{y}, \boldsymbol{\gamma} = \boldsymbol{\gamma}^{[t]})$$

These last quantities are just 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 the GLS-BLUP solutions of the Henderson mixed model equations and of their accuracy [11], i.e.

$$\left(\sum_{i=1}^{I} \sigma_{e_i}^{-2}\mathbf{T}_i'\mathbf{T}_i + \boldsymbol{\Sigma}^-\right)\widehat{\boldsymbol{\theta}} = \sum_{i=1}^{I} \sigma_{e_i}^{-2}\mathbf{T}_i'\mathbf{y}_i \tag{7}$$

where $\mathbf{T}_i = (\mathbf{X}_i, \sigma_{u_i}\mathbf{Z}_i)$, $\widehat{\boldsymbol{\theta}} = \left(\widehat{\boldsymbol{\beta}}', \widehat{\mathbf{u}}^{*\prime}\right)'$, and $\boldsymbol{\Sigma}^- = \begin{bmatrix} 0 & 0 \\ 0 & \mathbf{A}^{-1} \end{bmatrix}$.

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

$$\left.\begin{aligned} \widehat{S}_{i,\varepsilon\varepsilon} &= \left(\mathbf{y}_i - \mathbf{X}_i\widehat{\boldsymbol{\beta}}\right)'\left(\mathbf{y}_i - \mathbf{X}_i\widehat{\boldsymbol{\beta}}\right) + \mathrm{tr}\,(\mathbf{X}_i'\mathbf{X}_i\mathbf{C}_{\beta\beta}) \\ \widehat{S}_{i,\varepsilon u} &= \widehat{\mathbf{u}}^{*\prime}\mathbf{Z}_i'\left(y_i - \mathbf{X}_i\widehat{\boldsymbol{\beta}}\right) - \mathrm{tr}\,(\mathbf{Z}_i'\mathbf{X}_i\mathbf{C}_{\beta u}) \\ \widehat{S}_{i,uu} &= \widehat{\mathbf{u}}^{*\prime}\mathbf{Z}_i'\mathbf{Z}_i\widehat{\mathbf{u}}^* + \mathrm{tr}\,(\mathbf{Z}_i'\mathbf{Z}_i\mathbf{C}_{uu}) \end{aligned}\right\} \tag{8}$$

where $\widehat{\boldsymbol{\beta}}$ and $\widehat{\mathbf{u}}^*$ are solutions of the mixed model equations, 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 in equation (7). For grouped data (n_i observations in subclass i with the same incidence matrices $\mathbf{X}_i = \mathbf{1}_{n_i}\mathbf{x}_i'$ and $\mathbf{Z}_i = \mathbf{1}_{n_i}\mathbf{z}_i'$), formulae (8) reduce to:

$$\left.\begin{aligned} \widehat{S}_{i,\varepsilon\varepsilon} &= \sum_{j=1}^{n_i} (y_{ij} - \widehat{\mu}_i)^2 + n_i\,\mathrm{tr}(\mathbf{x}_i\mathbf{x}_i'\mathbf{C}_{\beta\beta}) \\ \widehat{S}_{i,\varepsilon u} &= n_i\left[\widehat{u}_i^*(\overline{y}_{i.} - \widehat{\mu}_i) - \mathrm{tr}(\mathbf{z}_i\mathbf{x}_i'\mathbf{C}_{\beta u})\right] \\ \widehat{S}_{i,uu} &= n_i\left\{\widehat{u}_i^{*2} + \mathrm{tr}(\mathbf{z}_i\mathbf{z}_i'\mathbf{C}_{uu}\right\} \end{aligned}\right\} \tag{9}$$

where $\mu_i = \mathbf{x}_i'\boldsymbol{\beta}$, $u_i^* = \mathbf{z}_i'\mathbf{u}^*$ and $\overline{y}_{i.} = \left(\sum_{j=1}^{n_i} y_{ij}\right)/n_i$.

2.3. Hypothesis testing

Tests of hypotheses about dispersion parameters $\boldsymbol{\gamma} = (\boldsymbol{\delta}', \tau, b)'$ can be carried out via the likelihood ratio statistic (LRS) as proposed by Foulley et al. [6, 7].

Let H_0: $\boldsymbol{\gamma} \in \Omega_0$ be the null hypothesis, and H_1: $\boldsymbol{\gamma} \in \Omega - \Omega_0$ its alternative where Ω_0 and Ω refer to the restricted and unrestricted parameter spaces, respectively, such that $\Omega_0 \subset \Omega$. The LRS is defined as:

$$\lambda = -2L(\widetilde{\boldsymbol{\gamma}}; \mathbf{y}) + 2L(\widehat{\boldsymbol{\gamma}}; \mathbf{y}) \tag{10}$$

where $\widetilde{\boldsymbol{\gamma}}$ and $\widehat{\boldsymbol{\gamma}}$ are the REML estimators of $\boldsymbol{\gamma}$ under the restricted (H_0) and unrestricted ($H_0 \cup H_1$) models. Under standard conditions for H_0 (excluding hypotheses allowing the true parameter to be on the boundary of the parameter space as addressed by Robert et al. [22], λ has an asymptotic chi-square distribution with $r = \dim \Omega - \dim \Omega_0$ degrees of freedom.

Under model (1), an expression of $-2L(\boldsymbol{\gamma}; \mathbf{y})$ is:

$$-2L(\boldsymbol{\gamma}; \mathbf{y}) = \ell n|\mathbf{A}| + \sum_{i=1}^{I} n_i \, \ell n \, \sigma_{e_i}^2 + \ell n \left| \sum_{i=1}^{I} \sigma_{e_i}^{-2} \mathbf{T}_i' \mathbf{T}_i + \boldsymbol{\Sigma}^{-} \right| + \sum_{i=1}^{I} \sigma_{e_i}^{-2} \mathbf{y}_i' (\mathbf{y}_i - \mathbf{T}_i \widehat{\boldsymbol{\theta}}) + \text{const} \quad (11)$$

The theoretical and practical aspects of the hypotheses to be tested about the structural component $\boldsymbol{\delta}$ have been already discussed by Foulley et al. [6, 7], San Cristobal et al. [23] and Foulley [4].

As far as the functional relationship between the residual and u-components is concerned, interest focuses primarily on the hypotheses i) a constant variance ratio ($b = 1$), and ii) a constant u-component of variance ($b = 0$) [2, 16, 22, 28].

Note that the structural functional model can be tested against the double structural model: $\ell n \, \sigma_{e_i}^2 = \mathbf{p}_i' \boldsymbol{\delta}_e$, and $\ell n \, \sigma_{u_i}^2 = \mathbf{p}_i' \boldsymbol{\delta}_u$ with the same covariates. The reason for that is as follows. Let $\mathbf{P} = [\mathbf{1}|\mathbf{P}^*]$, $\boldsymbol{\delta}_e = [\eta_e, \boldsymbol{\delta}_e^*]$ and $\boldsymbol{\delta}_u = [\eta_u, \boldsymbol{\delta}_u^*]$ pertaining to a traditional parameterization involving intercepts η_e and η_u for describing the residual and u-components of variance, respectively, of a reference population. The structural functional model reduces to the null hypothesis $\boldsymbol{\delta}_u^* = 2b\boldsymbol{\delta}_e^*$, thus resulting in an asymptotic chi-square distribution of the LRS contrasting the two models with Rank($\mathbf{P}$)–2 degrees of freedom.

2.4. Numerical example

For readers interested in a test example, a numerical illustration is presented based on a small data set obtained by simulation. For pedagogical reasons, this example has the same structure as that presented in Foulley [4], i.e. it includes two crossclassified fixed factors (A and B) and one random factor (sire).

The model used to generate records is:

$$y_{ijklm} = \mu + \alpha_i + \beta_j + \tau \sigma_{eij}^b (s_k^* + 1/2 s_\ell^*) + e_{ijklm} \quad (12)$$

where μ is a general mean, α_i, β_j are the fixed effects of factors A($i = 1, 2$) and B($j = 1, 2, 3$), s_k^* the standardized contribution of male k as a sire and $1/2s_\ell^*$) that of male ℓ as a maternal grand sire.

Except for $\tau = 0.001016$ and $b = 1.75$, the values chosen for the parameters are the same as in Foulley [4]. The data set is listed in *table I.* The issue of model choice for location and log-residual parameters will not discussed again; they are kept the same, i.e. additive as in the previous analysis.

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

No.	A	B	S	T	n	Σy	Σy^2
1	1	1	1	4	21	2 266	251 044
2	1	2	1	4	19	1 789	171 215
3	1	1	1	7	14	1 189	105 173
4	1	3	1	7	7	529	40 995
5	1	2	1	8	6	508	43 628
6	1	3	2	6	12	882	66 634
7	1	1	2	7	7	630	57 608
8	1	2	2	8	18	1 523	133 779
9	2	1	2	8	27	3 149	381 599
10	2	2	3	5	7	778	88 502
11	2	1	3	5	19	2 123	250 801
12	2	2	3	5	10	1 012	107 340
13	2	3	3	2	37	3 066	290 320
14	2	2	3	2	13	1 527 .	181 647
15	2	3	4	7	13	1 478	172 306
16	2	3	4	8	6	939	150 173
17	2	2	4	9	19	2 305	287 059
18	2	2	4	5	12	1 482	187 372

A, B: environmental factors treated as fixed; S = sire and T = maternal grand sire treated as random. Elements of the 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 $(1,2) = (8,9) = 1/4$.

Table II presents $-2L$ values, LR statistics and P-values contrasting the following different models:

1) additive for both $\log \sigma_e^2$ and $\log \sigma_s^2$;
2) additive for $\log \sigma_e^2$ and $\log \sigma_s = a + b \log \sigma_e$;
3) constant variance ratio ($b = 1$);
4) constant sire variance ($b = 0$).

In this example, models (3) and (4) were rejected as expected whatever the alternatives, i.e. models (1) or (2). Model (2) was acceptable when compared to (1) thus illustrating that there is room between the complete structural approach and the constant variance ratio model.

The corresponding estimates of parameters are shown in *table III.* Estimates of the functional relationship are $\tau = 0.001143$ and $b = 3.0121$, this last value being higher than the true one, but – not surprisingly in this small sample – not significantly different ($\lambda = 1.5364$ and P-value $= 0.215$).

3. DISCUSSION AND CONCLUSION

This paper is a further step in the study of heterogeneous variances in mixed models. It provides a technical framework to investigate how the u-component of variance and the intra-class correlation varies with the residual variance.

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

	Model [a]			Likelihood [b]			Test [c]			
No.	Location	Residual	u-or Ratio	par	$-2L$	H_0	Comp	df	Statistic	P-value
(1)	$\mu + A + B$	$\mu^* + A^* + B^*$	$\mu' + A' + B'$, or $\mu'' + A'' + B''$	12	2 360.2722					
(2)	$\mu + A + B$	$\mu^* + A^* + B^*$	$\ln\sigma_{u_i} = a + b\ln\sigma_{e_i}$	10	2 364.0567	see text	2–1	2	3.7845	0.1507
(3)	$\mu + A + B$	$\mu^* + A^* + B^*$	μ'': ratio cst	9	2 368.2891	$A'' + B'' = 0$	3–1	3	8.0169	0.0457
						$b = 1$	3–2	1	4.2324	0.0397
(4)	$\mu + A + B$	$\mu^* + A^* + B^*$	μ': σ_u cst	9	2 373.0454	$A' + B' = 0$	4–1	3	12.7732	0.0051
						$b = 0$	4–2	1	8.9887	0.0027

[a] Covariates: μ: intercept; A: 1st factor (e.g. row); B: 2nd factor (e.g. column); AB: interaction; covariates for log parameters coded as: (e.g. A), residual variance (A^*), u-component (A') and u to residual variance logratio (A''); [b] par: number of parameters: L = maximum of the residual loglikelihood (REML version); [c] likelihood ratio (LR) test based on difference in $-2L$ between the full and reduced models.

Table III. REML estimates of subclass components of variance under different models.

Model	Component	AB Subclass 11	12	13	21	22	23	Functional parameters[b]
	σ_u	9.676	4.274	18.201	11.895	5.255	22.376	
1	σ_e	17.068	13.478	17.929	25.875	20.432	27.181	
	t^{a}	0.243	0.091	0.507	0.174	0.062	0.404	
	σ_u	7.082	3.101	9.378	19.141	8.381	25.347	$\tau = 0.001143$
2	σ_e	18.152	13.800	19.926	25.251	19.196	27.718	$b = 3.0121^{c}$
	t	0.132	0.048	0.181	0.365	0.160	0.455	
	σ_u	8.879	6.768	9.989	13.343	10.171	15.011	$\tau = 0.511269$
3	σ_e	17.366	13.237	19.537	26.099	19.894	29.361	$b = 1$
	t	0.207	0.207	0.207	0.207	0.207	0.207	
	σ_u	10.382	10.382	10.382	10.382	10.382	10.382	$\tau = 10.382233$
4	σ_e	16.775	13.459	18.803	26.252	21.063	29.426	$b = 0$
	t	0.277	0.373	0.234	0.135	0.195	0.110	

[a] $t = \sigma_u^2/(\sigma_u^2 + \sigma_e^2)$; [b] $\ln\sigma_u = a + b + \ln\sigma$; [c] LRS for the null hypothesis $b = 1.75$ against its alternative (b unspecified) is 1.5364 for one degree of freedom (P-value = 0.2151).

This has been an issue for many years in the animal breeding community. For instance for milk yield, the assumption of a constant heritability across levels of environmental factors (e.g. countries, regions, herds, years, management conditions) has generated considerable controversy: see Garrick and Van Vleck [8], Wiggans and VanRaden [29]; Visscher and Hill [26], Weigel et al. [28] and DeStefano [2]. Maximum likelihood computations are based, here, on the EM algorithm and different simplified versions of it (gradient EM, ECM). This is a powerful tool for addressing problems of variance component estimation, in particular those of heterogeneous variances [4, 5, 7, 20, 21]. It is not only an easy procedure to implement but also a flexible one. For instance, ML rather than REML estimators can be obtained after a slight modification of the E-step resulting for grouped data in

$$\widehat{\widehat{S}}_{i,\varepsilon\varepsilon} = \sum_{j=1}^{n_i}(y_{ij} - \widehat{\mu}_i)^2$$

$$\widehat{\widehat{S}}_{i,\varepsilon u} = n_i\widehat{u}_i^*(\overline{y}_{i.} - \widehat{\mu}_i)$$

$$\widehat{\widehat{S}}_{i,uu} = n_i\left\{\widehat{u}_i^{*2} + \mathrm{tr}(\mathbf{z}_i\mathbf{z}_i'\mathbf{M}_{uu}^{-1})\right\} \tag{13}$$

where $\mathbf{M}_{uu}$ is the $u \times u$ block of the coefficient matrix of the Henderson mixed model equations.

Posterior mode estimators can also be derived using EM [5, 9, 27].

Moreover the procedure can be extended to models with several($k = 1, 2, \ldots, K$) uncorrelated u random factors, e.g.

$$\mathbf{y}_i = \mathbf{X}_i\boldsymbol{\beta} + \sum_{k=1}^{K} \tau_k \sigma_{e_i}^{b_k} \mathbf{Z}_{ik}\mathbf{u}_k^* + \mathbf{e}_i \tag{14}$$

Such an extension will be easy to make via the ECM (expectation conditional maximization) algorithm [15] in its standard or gradient version along the same lines as those described in Foulley [4]. However caution should be exercised in applying the gradient ECM, for this algorithm no longer guarantees convergence in likelihood values. Other alternatives might be considered as well such as the average information-REML procedure [10, 17].

In conclusion, the likelihood framework provides a powerful tool both for estimation and hypothesis testing of different competing models regarding those problems. However, additional research work is still needed to study some properties of these procedures especially from a practical point of view, for example the power of testing such assumptions as $b = 1$.

ACKNOWLEDGEMENTS

This work was conducted while Caroline Thaon was on a 'stage de fin d'études' at the Station de génétique quantitative et appliquée (SGQA), Inra, Jouy-en-Josas as a student from the Ecole Polytechnique, Palaiseau. She greatly acknowledges the support of both institutions in making this stay feasible. Thanks are expressed to Elinor Thompson (Inra-Jouy en Josas) for the English revision of the manuscript.

REFERENCES

[1] Dempster A.P., Laird N.M., Rubin D.B., Maximum likelihood from incomplete data via the EM algorithm, J. R. Statist. Soc. B 39 (1977) 1–38.

[2] DeStefano A.L., Identifying and quantifying sources of heterogeneous residual and sire variances in dairy production data, Ph.D. thesis, Cornell University, Ithaca, New York.

[3] Fahrmeir L., Tutz G., Multivariate Statistical Modelling Based on Generalized Linear Models. Springer Verlag, Berlin, 1994.

[4] Foulley J.L., ECM approaches to heteroskedastic mixed models with constant variance ratios, Genet. Sel. Evol. 29 (1997) 297–318.

[5] Foulley J.L., Quaas R.L., Heterogeneous variances in Gaussian linear mixed models, Genet. Sel. Evol. 27 (1995) 211–228.

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

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

[8] Garrick D.J., Van Vleck L.D., Aspects of selection for performance in several environments with heterogeneous variances, J. Anim. Sci. 65 (1987) 409–421.

[9] Gianola D., Foulley J.L., Fernando R.L., Henderson C.R., Weigel K.A., Estimation of heterogeneous variances using empirical Bayes methods: theoretical considerations, J. Dairy Sci. 75 (1992) 2805–2823.

[10] Gilmour A.R., Thompson R., Cullis B.R., Average information REML: an efficient algorithm for variance parameter estimation in linear mixed models, Biometrics 51 (1995) 1440–1450.

[11] Henderson C.R., Applications of Linear Models in Animal Breeding, University of Guelph, Guelph, Ontario, Canada, 1984.

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

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

[14] McCullagh P., Nelder J., Generalized Linear Models, 2nd ed., Chapman and Hall, London, 1989.

[15] Meng X.L., Rubin D.B., Maximum likelihood estimation via the ECM algorithm: A general framework, Biometrika 80 (1993) 267–278.

[16] Meuwissen T.H.E., De Jong G., Engel B., Joint estimation of breeding values and heterogenous variances of large data files, J. Dairy Sci. 79 (1996) 310–316.

[17] Meyer K., An average information restricted maximum likelihood algorithm for estimating reduced rank genetic covariance matrices or covariance functions for animal models with equal design matrices, Genet. Sel. Evol. 29 (1997) 97–116.

[18] Patterson H.D., Thompson R., Recovery of interblock information when block sizes are unequal, Biometrika 58 (1971) 545–554.

[19] Robert C., Etude de quelques problèmes liés à la mise en oeuvre du REML en génétique quantitative, thèse de Doctorat, Université Paul Sabatier, Toulouse, France.

[20] Robert C., Foulley J.L., Ducrocq V., Inference on homogeneity of intra-class correlations among environments using heteroskedastic models, Genet. Sel. Evol. 27 (1995) 51–65.

[21] Robert C., Foulley J.L., Ducrocq V., Estimation and testing of constant genetic and intra-class correlation coefficients among environments, Genet. Sel. Evol. 27 (1995) 125–134.

[22] Robert C., Ducrocq V., Foulley J.L., Heterogeneity of variance for type traits in the Montbéliarde cattle, Genet. Sel Evol 29, (1997) 545–570.

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

[24] Thompson R., Baker R.J., Composite link functions in generalized linear models. Appl. Stat. 30 (1981) 125–131.

[25] Visscher P.M., 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 (1992) 1320–1330.

[26] Visscher P.M., Hill W.G., Heterogeneity of variance and dairy cattle breeding, Anim. Prod. 55 (1992) 321–329.

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

[28] Weigel K.A., Gianola D., Yandel B.S., Keown J.F., Identification of factors causing heterogeneous within-herd variance components using a structural model for variances, J. Dairy Sci. 76 (1993) 1466–1478.

[29] Wiggans G.R., VanRaden P.M., Method and effect of adjustment for heterogeneous variance, J. Dairy Sci. 74 (1991) 4350–4357.

A1. APPENDIX: Derivatives for the EM algorithm

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

$$Q(\boldsymbol{\gamma}) = \text{const} - \frac{1}{2}\sum_{i=1}^{I} n_i \, \ell n \, \sigma_{e_i}^2 - \frac{1}{2}\sum_{i=1}^{I} \sigma_{e_i}^{-2} \mathrm{E}_c(\mathbf{e}_i'\mathbf{e}_i) \tag{A1}$$

with

$$\ell n \, \sigma_{e_i}^2 = \mathbf{p}_i'\boldsymbol{\delta} \tag{A2}$$

and,

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

A1.1. Derivative with respect to $\boldsymbol{\delta}$ (log residual parameters)

According to the chain rule, one has

$$\frac{\partial Q}{\partial \boldsymbol{\delta}} = \sum_{i=1}^{I} \frac{\partial Q}{\partial \ell n \, \sigma_{e_i}^2} \frac{\partial \ell n \, \sigma_{e_i}^2}{\partial \boldsymbol{\delta}}$$

Now

$$\frac{\partial Q}{\partial \ell n \, \sigma_{e_i}^2} = \sigma_{e_i}^2 \frac{\partial Q}{\partial \sigma_{e_i}^2}$$

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

That is

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

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

and

$$\frac{\partial \mathbf{e}_i}{\partial \sigma_{e_i}} = -\tau b \sigma_{e_i}^{b-1} \mathbf{Z}_i \mathbf{u}^*$$

Thus,

$$\frac{\partial Q}{\partial \sigma_{e_i}^2} = -\frac{1}{2}\left[\frac{n_i}{\sigma_{e_i}^2} - \frac{\mathrm{E}_c(\mathbf{e}_i'\mathbf{e}_i)}{\sigma_{e_i}^4} - \frac{\tau b \sigma_{e_i}^{b-2} \mathrm{E}_c(\mathbf{u}^{*\prime}\mathbf{Z}_i'\mathbf{e}_i)}{\sigma_{e_i}^2}\right]$$

Letting $v_{\delta,i} = \partial Q / \partial \ell n \, \sigma_{e_i}^2$ so that $\dfrac{\partial Q}{\partial \boldsymbol{\delta}} = \sum_{i=1}^{I} v_{\delta,i} \mathbf{p}_i = \mathbf{P}' \mathbf{v}_\delta$, then

$$v_{\delta,i} = \frac{1}{2} \left\{ \frac{\mathrm{E}_c \left\{ \left[\mathbf{y}_i - \mathbf{X}_i \boldsymbol{\beta} - (1-b) \tau \sigma_{e_i}^b \mathbf{Z}_i \mathbf{u}^* \right]' \mathbf{e}_i \right\}}{\sigma_{e_i}^2} - n_i \right\} \tag{A4}$$

Let us define

$$\begin{aligned} S_{i,\varepsilon\varepsilon} &= \boldsymbol{\varepsilon}_i' \boldsymbol{\varepsilon}_i = (\mathbf{y}_i - \mathbf{X}_i \boldsymbol{\beta})'(\mathbf{y}_i - \mathbf{X}_i \boldsymbol{\beta}), \\ S_{i,u\varepsilon} &= \mathbf{u}^{*'} \mathbf{Z}_i' (\mathbf{y} - \mathbf{X}_i \boldsymbol{\beta}), \\ S_{i,uu} &= \mathbf{u}^{*'} \mathbf{Z}_i' \mathbf{Z}_i \mathbf{u}^* \end{aligned}$$

and, the same symbols with a hat for their conditional expectations, i.e.

$$\widehat{S}_{i,..} = \mathrm{E}_c(S_{i,..}) = \mathrm{E}(S_{i,..} | \mathbf{y}, \boldsymbol{\gamma} = \boldsymbol{\gamma}^{[t]})$$

an alternative expression for computing (A4) is

$$\boxed{v_{\delta,i} = \frac{1}{2} \left[\sigma_{e_i}^{-2} \widehat{S}_{i,\varepsilon\varepsilon} - (2-b) \tau \sigma_{e_i}^{b-2} \widehat{S}_{i,u\varepsilon} + (1-b) \tau^2 \sigma_{e_i}^{2(b-1)} \widehat{S}_{i,uu} - n_i \right]} \tag{A5}$$

Notice that (A4) applied with $b = 0$ and $b = 1$ retrieves the classical formulae

$$v_{\delta,i} = \frac{1}{2} \left\{ \frac{\mathrm{E}_c(\mathbf{e}_i' \mathbf{e}_i)}{\sigma_{e_i}^2} - n_i \right\}, \quad \text{and} \quad v_{\delta,i} = \frac{1}{2} \left\{ \frac{\mathrm{E}_c \left[y_i - X_i \beta \right]' \mathbf{e}_i}{\sigma_{e_i}^2} - n_i \right\}$$

already reported by Foulley et al. [6] and Foulley [4] for models with a homogeneous u-component of variance, and a constant u to e variance ratio, respectively.

A1.2. Derivative with respect to τ

$$\frac{\partial Q}{\partial \tau} = -\frac{1}{2} \sum_{i=1}^{I} \sigma_{e_i}^{-2} \frac{\partial \mathrm{E}_c(\mathbf{e}_i' \mathbf{e}_i)}{\partial \tau}$$

with

$$\frac{\partial \mathrm{E}_c(\mathbf{e}_i' \mathbf{e}_i)}{\partial \tau} = 2 \, \mathrm{E}_c \left[\left(\frac{\partial \mathbf{e}_i'}{\partial \tau} \right) \mathbf{e}_i \right]$$

and

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

so that

$$v_\tau = \frac{\partial Q}{\partial \tau} = \sum_{i=1}^{I} \sigma_{e_i}^{b-2} \mathrm{E}_c(\mathbf{u}^{*'} \mathbf{Z}_i' \mathbf{e}_i) \tag{A6}$$

or, more explicitly

$$\boxed{v_\tau = \sum_{i=1}^{I} \sigma_{e_i}^{b-2} \widehat{S}_{i,u\varepsilon} - \tau \sum_{i=1}^{I} \sigma_{e_i}^{2(b-1)} \widehat{S}_{i,uu}} \tag{A7}$$

A1.3. Derivative with respect to b

Similarly

$$\frac{\partial Q}{\partial b} = -\sum_{i=1}^{I} \sigma_{e_i}^{-2} \mathrm{E}_c \left(\frac{\partial \mathbf{e}_i'}{\partial b} \mathbf{e}_i \right)$$

with

$$\frac{\partial \mathbf{e}_i}{\partial b} = -\tau \sigma_{e_i}^{b} \ell n\, \sigma_{e_i} \mathbf{Z}_i \mathbf{u}^*$$

so that

$$v_b = \frac{\partial Q}{\partial b} = \tau \sum_{i=1}^{I} \sigma_{e_i}^{b-2} \ell n\, \sigma_{e_i} \mathrm{E}_c \left(\mathbf{u}^{*'} \mathbf{Z}_i' \mathbf{e}_i \right) \tag{A8}$$

or alternatively,

$$\boxed{v_b = \tau \left[\sum_{i=1}^{I} \sigma_{e_i}^{b-2} \ell n\, \sigma_{e_i} \widehat{S}_{i,u\varepsilon} - \tau \sum_{i=1}^{I} \sigma_{e_i}^{2(b-1)} \ell n\, \sigma_{e_i} \widehat{S}_{i,uu} \right]} \tag{A9}$$

A1.4. $\boldsymbol{\delta} - \boldsymbol{\delta}$ derivatives

Let us define

$$-\frac{\partial^2 Q}{\partial \boldsymbol{\delta} \partial \boldsymbol{\delta}'} = \sum_{i=1}^{I} w_{\delta\delta,ii} \mathbf{p}_i \mathbf{p}_i' = \mathbf{P}' \mathbf{W}_{\delta\delta} \mathbf{P} \tag{A10}$$

where

$$w_{\delta\delta,ii} = -\frac{\partial v_{\delta,i}}{\partial \ell n\, \sigma_{e_i}^2} = -\sigma_{e_i}^2 \frac{\partial v_{\delta,i}}{\partial \sigma_{e_i}^2}$$

Now

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

and

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

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

After developing and rearranging, one obtains

$$\boxed{w_{\delta\delta,ii} = \frac{1}{2}\left[\sigma_{e_i}^{-2}\widehat{S}_{i,\varepsilon\varepsilon} - \frac{(2-b)^2}{2}\tau\sigma_{e_i}^{b-2}\widehat{S}_{i,u\varepsilon} + (1-b)^2\tau^2\sigma_{e_i}^{2(b-1)}\widehat{S}_{i,uu}\right]} \qquad \text{(A11)}$$

Letting $b = 0$ and $b = 1$ in (A11) leads to

$$w_{\delta\delta,ii} = \frac{\mathrm{E}_c(\mathbf{e}_i'\mathbf{e})}{2\sigma_{e_i}^2}$$

and

$$w_{\delta\delta,ii} = \frac{1}{2\sigma_{e_i}^2}\left(\widehat{S}_{i,\varepsilon\varepsilon} - \frac{\tau}{2}\sigma_{e_i}\widehat{S}_{i,u\varepsilon}\right)$$

Again these are the same expressions as those given by Foulley et al. [6] and Foulley [4] for a constant u-component of variance and a constant variance ratio, respectively.

A1.5. $\delta - \tau$ derivatives

$$-\frac{\partial^2 Q}{\partial\delta\partial\tau} = \sum_{i=1}^{I} w_{\delta\tau,i}\mathbf{p}_i = \mathbf{P}'\mathbf{W}_{\delta\tau} \qquad \text{(A12)}$$

where

$$w_{\delta\tau,i} = -\frac{\partial v_{\delta,i}}{\partial\tau}$$

Now

$$\frac{\partial v_{\delta,i}}{\partial\tau} = -\frac{1}{2\sigma_{e_i}^2}(1-b)\sigma_{e_i}^b\mathrm{E}_c(\mathbf{u}^{*\prime}\mathbf{Z}_i'\mathbf{e}_i) + \frac{1}{2\sigma_{e_i}^2}\mathrm{E}_c\left\{\left[\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta} - (1-b)\tau\sigma_{e_i}^b\mathbf{Z}_i\mathbf{u}^*\right]'\frac{\partial\mathbf{e}_i}{\partial\tau}\right\}$$

Finally

$$\boxed{w_{\delta\tau,i} = \frac{(2-b)}{2}\sigma_{e_i}^{b-2}\widehat{S}_{i,u\varepsilon} - \tau(1-b)\sigma_{e_i}^{2(b-1)}\widehat{S}_{i,uu}} \qquad \text{(A13)}$$

A1.6. $\delta - b$ *derivatives*

$$-\frac{\partial^2 Q}{\partial\delta\partial b} = \sum_{i=1}^{I} w_{\delta b,i}\mathbf{p}_i = \mathbf{P}'\mathbf{W}_{\delta b} \quad \text{(A14)}$$

where

$$w_{\delta b,i} = -\frac{\partial v_{\delta,i}}{\partial b}$$

Now

$$\frac{\partial v_{\delta,i}}{\partial b} = -\frac{\tau}{2\sigma_{e_i}^2}\frac{\partial\left[(1-b)\sigma_{e_i}^b\right]}{\partial b}\mathrm{E}_c(\mathbf{u}^{*\prime}\mathbf{Z}_i'\mathbf{e}_i) + \frac{1}{2\sigma_{e_i}^2}\mathrm{E}_c\left\{\left[\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta} - (1-b)\tau\sigma_{e_i}^b\mathbf{Z}_i\mathbf{u}^*\right]'\frac{\partial\mathbf{e}_i}{\partial b}\right\}$$

and

$$\frac{\partial\left[(1-b)\sigma_{e_i}^b\right]}{\partial b} = \sigma_{e_i}^b\left[(1-b)\ell n\,\sigma_{e_i} - 1\right]$$

so that

$$\boxed{w_{\delta b,i} = \frac{\tau}{2}\sigma_{e_i}^{b-2}\left\{\left[(2-b)\ell n\,\sigma_{e_i} - 1\right]\widehat{S}_{i,u\varepsilon} + \tau\sigma_{e_i}^b\left[1 + 2(b-1)\ell n\,\sigma_{e_i}\right]\widehat{S}_{i,uu}\right\}} \quad \text{(A15)}$$

A1.7. $\tau - \tau$ *derivatives*

Differentiating (A7) once again with respect to τ leads to

$$\boxed{w_{\tau\tau} = -\frac{\partial^2 Q}{\partial\tau^2} = \sum_{i=1}^{I}\sigma_{e_i}^{2(b-1)}\widehat{S}_{i,uu}} \quad \text{(A16)}$$

A1.8. $\tau - b$ *derivatives*

From (A7), one has

$$\boxed{w_{\tau b} = -\frac{\partial^2 Q}{\partial\tau\partial b} = 2\tau\sum_{i=1}^{I}\sigma_{e_i}^{2(b-1)}\ell n\,\sigma_{e_i}\widehat{S}_{i,uu} - \sum_{i=1}^{I}\sigma_{e_i}^{b-2}\ell n\,\sigma_{e_i}\widehat{S}_{i,u\varepsilon}} \quad \text{(A17)}$$

A1.9. $b - b$ *derivatives*

From (A9), one gets

$$\boxed{w_{bb} = -\frac{\partial^2 Q}{\partial b^2} = \tau\left[2\tau\sum_{i=1}^{I}\sigma_{e_i}^{2(b-1)}(\ell n\,\sigma_{e_i})^2\widehat{S}_{i,uu} - \sum_{i=1}^{I}\sigma_{e_i}^{b-2}(\ell n\,\sigma_{e_i})^2\widehat{S}_{i,u\varepsilon}\right]} \quad \text{(A18)}$$

Finally, the Newton-Raphson algorithm to implement for the M-step of the EM algorithm can be written in condensed form as:

$$\begin{bmatrix} \mathbf{P}'\mathbf{W}_{\delta\delta}\mathbf{P} & \mathbf{P}'\mathbf{W}_{\delta\tau} & \mathbf{P}'\mathbf{W}_{\delta b} \\ \mathbf{W}'_{\delta\tau}\mathbf{P} & w_{\tau\tau} & w_{\tau b} \\ \mathbf{W}'_{\delta b}\mathbf{P} & w_{\tau b} & w_{bb} \end{bmatrix}^{[n-1]} \begin{bmatrix} \boldsymbol{\Delta\delta} \\ \Delta\tau \\ \Delta b \end{bmatrix}^{[n]} = \begin{bmatrix} \mathbf{P}'\mathbf{v}_{\delta} \\ v_{\tau} \\ v_b \end{bmatrix}^{[n-1]} \tag{A19}$$

where at iteration $[n]$, $\boldsymbol{\Delta\delta}^{[n]} = \boldsymbol{\delta}^{[n]} - \boldsymbol{\delta}^{[n-1]}$, and $\Delta\tau^{[n]} = \tau^{[n]} - \tau^{[n-1]}$ and $\Delta b^{[n]} = b^{[n]} - b^{[n-1]}$.

A gradient EM version would be to solve:

$$\left\{ \begin{aligned} [\mathbf{P}'\mathbf{W}_{\delta\delta}\mathbf{P}]^{[n-1]}\,\boldsymbol{\Delta\delta}^{[n]} &= \mathbf{P}'\mathbf{v}_{\delta}^{[n-1]} \\ \Delta b^{[n]} &= v_b^{[n-1]}/w_{bb}^{[n-1]} \\ \tau^{[n]} &= \left[\sum_{i=1}^{I} \sigma_{e_i}^{b-2}\widehat{S}_{i,u\varepsilon}\right]^{[n-1]} \Bigg/ \left[\sum_{i=1}^{I} \sigma_{e_i}^{2(b-1)}\widehat{S}_{i,uu}\right]^{[n-1]} \end{aligned} \right. \tag{A20}$$