

## Estimation of heterogeneous variances in nonlinear mixed models via the SAEM-MCMC algorithm with applications to growth curves in poultry

Mylène Duval<sup>1</sup>, Christèle Robert-Granié<sup>1</sup> & Jean-Louis Foulley<sup>2</sup>

### Abstract

The SAEM-MCMC algorithm is a powerful tool for computing maximum likelihood estimators in the wide class of nonlinear mixed effects models. We propose in this article an adaptation of this algorithm to the estimation of heterogeneous variances in such models. Two residual variance models are considered: a linear mixed model on the log-variance, with fixed and random effects, and a mean-variance relationship. As compared to other procedures implemented in R, SAS and Monolix, our algorithm provides more flexibility in modelling variance functions and reliability of the estimates. This algorithm was numerically validated in the case of a heteroskedastic linear mixed model by comparing its results with those of a standard EM algorithm applied to Pothoff and Roy's data. Finally, an application to real data involving a selection experiment on growth in chickens is presented in which that algorithm was compared to results of SAS-Nlmixed, nlme, Monolix and WinBUGS softwares.

**Keywords :** *heteroskedasticity, maximum likelihood estimation, nonlinear mixed models, SAEM-MCMC algorithm*

### Résumé

L'algorithme SAEM-MCMC est un outil puissant en vue de l'estimation des paramètres, par maximum de vraisemblance, dans les modèles non linéaires mixtes. Dans cet article, nous proposons une adaptation de cet algorithme pour estimer les paramètres dans le cadre des modèles linéaires mixtes à variances hétérogènes. Nous considérons, ici, deux modélisations possibles des variances : un modèle structural (incluant des effets fixes et aléatoires) basé sur les log-variances et un modèle s'appuyant sur une liaison moyenne-variance. En comparaison à d'autres procédures mises en œuvre dans les logiciels R, SAS et Monolix, notre algorithme s'avère beaucoup plus flexible pour modéliser des fonctions de variance. L'algorithme que nous proposons a été numériquement validé dans le cas d'un modèle linéaire mixte à variances hétérogènes en comparant les résultats obtenus avec ceux relatifs à un algorithme EM analytique sur le jeu de données standard de Pothoff et Roy. Enfin, nous présentons une application sur des données réelles concernant une expérience de sélection sur la croissance de poulets. Sur ces données, nous avons comparé les résultats de notre algorithme à ceux obtenus avec les procédures suivantes : SAS-NLMIXED, nlme, Monolix et WinBUGS.

**Mots-clés :** *Hétéroscédasticité, Maximum de vraisemblance, modèle non linéaire mixte, algorithme SAEM-MCMC*

**Mathematics Subject Classification:** (62H12, 62P10)

<sup>1</sup>INRA UR631 SAGA, BP 52627, F-31 326 Castanet-Tolosan, France  
Christele.Robert-Granie@toulouse.inra.fr

<sup>2</sup>INRA UMR1313 GABI, domaine de Vilvert, F-78350 Jouy-en-Josas, France

## 1 Introduction

Nonlinear mixed models (NLMM) are the tools of choice for analyzing unbalanced repeated data arising from complex biological mechanisms such as growth curves or pharmacokinetic and pharmacodynamics (PKPD) trials. Moreover, as compared to linear models, parameters of such models provide a better biological interpretation of the mechanisms involved, and the corresponding models are also more parsimonious.

Most studies based on these models assume that variances are homogenous across strata within the sampled population. However, this assumption turns out to be unrealistic in many practical situations encountered in various areas of applied statistics, e.g. biostatistics [34] and economics [18, 11]. In particular, heteroskedasticity arises naturally in within-group or residual variances. There are different ways to model heterogeneous variances including, among others, mean-variance relationships [3] and linear models on log-variances: see Davidian and Carroll [6] for a general discussion. This study will deal with estimation methods of such variance functions in the context of nonlinear models with random effects.

To that effect, several procedures have already been proposed and investigated to estimate parameters of NLMM. The first were based on linearization of the data model such as first-order (FO) approximations [38], or better, first-order conditional expectation (FOCE) approximations [24] or equivalently, Laplace's approximations [42]. Since these methods can produce inconsistent estimates, alternatives based on exact maximum likelihood (ML) have been implemented involving either Gaussian quadratures or stochastic tools. However, integration via Gaussian quadratures can be extremely difficult or inaccurate, in particular in cases with high dimensionality. Due to the complex form of the likelihood function, proper inference for non linear mixed models requires the use of stochastic procedures [30]. Stochastic EM is a powerful alternative to the latter, especially the so-called Stochastic Approximate EM procedure (SAEM) introduced by Delyon et al. [8] and applied initially to NLMM with an additional MCMC step by Kuhn and Lavielle [19, 20]. These authors show that the SAEM algorithm is very efficient for computing the maximum likelihood estimate in homogeneous non linear mixed models. This iterative procedure consists at each iteration, in successively simulating the random effects with the conditional distribution, and updating the unknown parameters of the model. This method has the very nice advantage to converge quickly to a neighborhood of the maximum likelihood estimate. Most software for NLMM already include heterogeneous variances in their syntax, but, either they rely on approximate inference procedures (e.g. nlme) or, if not, they often fail to converge (e.g. SAS-Nlmixed) or offer a limited choice of variance functions (e.g. Monolix).

Thus, the objective of this paper is two fold: (i) to propose a general algorithm involving both a more diversified assortment of variance functions and more reliable estimating procedures for computing ML estimation of parameters in heteroskedastic NLMM via the SAEM-MCMC algorithm ; (ii) to show how incorporating variance functions can improve the efficiency of growth modes with an application to poultry.

The paper is organized as follows. In section 2, we introduce a general class of NLMM with different residual variance functions. Then in section 3, the SAEM-MCMC algorithm is described. In section 4, we present two numerical applications. The first is a validation of our algorithm on a linear mixed model. The second is an application to a real data set of growth curves in chickens, the results of which are compared to those of other methods and softwares.

## 2 The heteroskedastic nonlinear mixed Model

Let us consider the general class of the following nonlinear mixed models (NLMM):

$$y_{ij} = f(\mathbf{z}_{ij}, \boldsymbol{\beta}, \boldsymbol{\phi}_i) + g(\mathbf{w}_{ij}, \boldsymbol{\delta}, \boldsymbol{\psi}_i)\varepsilon_{ij}^* \quad (1)$$

where  $y_{ij}$  is the  $j$ -th observation,  $j \in \{1, \dots, n_i\}$ , of the  $i$ -th individual,  $i \in \{1, \dots, I\}$ .  $f(\mathbf{z}_{ij}, \boldsymbol{\beta}, \boldsymbol{\phi}_i)$  describes the mean response as a function  $f$  of  $\mathbf{z}_{ij}$ , a vector of explanatory variables (for instance, time) and  $\boldsymbol{\phi}_i$  a  $(m \times 1)$  vector of random effects pertaining to individual  $i$ . It is assumed that  $\boldsymbol{\phi}_i$  are i.i.d. Gaussian random variables  $\mathcal{N}(\mathbf{A}_i\boldsymbol{\mu}, \boldsymbol{\Gamma})$  with mean  $\mathbf{A}_i\boldsymbol{\mu}$  and variance  $\boldsymbol{\Gamma}$ .  $\mathbf{A}_i\boldsymbol{\mu}$  specifies the link between the expectation of  $\boldsymbol{\phi}_i$  and  $p$  covariates  $\mathbf{A}_i$  ( $m \times p$ ) (for instance, gender, treatment) with coefficients  $\boldsymbol{\mu}$ . The variance covariance matrix  $\boldsymbol{\Gamma}$  quantifies the inter-individual variation. Vector  $\boldsymbol{\beta}$  corresponds to additional fixed effects involved in some elements of individual coefficients (similar to vector  $\boldsymbol{\phi}_i$ ) but having no random counterpart. In this model, the intra-individual variation is accounted for by the product  $g(\mathbf{w}_{ij}, \boldsymbol{\delta}, \boldsymbol{\psi}_i)\varepsilon_{ij}^*$  where  $\varepsilon_{ij}^*$  are i.i.d.  $\mathcal{N}(0, 1)$  random errors and  $\sigma_{ij} = g(\mathbf{w}_{ij}, \boldsymbol{\delta}, \boldsymbol{\psi}_i)$  stands for the square root of the variance function described with arguments in the same way as in function  $f$ .

More specifically, we are considering the two following variance functions to describe heteroskedasticity in residual variances. The first approach consists of modelling residual variances as a structural linear mixed model involving explanatory covariates  $\mathbf{w}_{ij}$ ,  $\mathbf{q}_{ij}$  and the corresponding fixed and random effects  $\boldsymbol{\delta}$ ,  $\mathbf{v}$  via a log-link function:

$$\log(\sigma_{ij}^2) = \mathbf{w}'_{ij}\boldsymbol{\delta} + \mathbf{q}'_{ij}\mathbf{v} \quad (2)$$

where  $\mathbf{v} \sim \mathcal{N}(0, \boldsymbol{\Lambda})$ .

This model was proposed initially by Foulley et al. [12] as a mixed model extension of the classical linear model for log-variances. It can be viewed as part of the multilevel mixed linear models developed by Goldstein [15] and was also investigated later on by Lin et al. [23], Brown et al. [4], Rigby and Stasinopoulos [36], Lu et al. [28] and Lee and Nelder [22] in the context of linear and generalized linear mixed models. It provides great flexibility in modelling potential sources of variation in the residual variances with the key idea of parsimony due to the introduction of random effects. The second approach assumes that the variance is proportional to a power of the (conditional) mean [3, 6]:

$$\sigma_{ij}^2 = \delta_1 f(\mathbf{z}_{ij}, \boldsymbol{\beta}, \boldsymbol{\phi}_i)^{\delta_2} \quad (3)$$

This model is very popular for taking into account, in an easy way, a scale effect linking the variance to the mean, e.g., with growth curve data [2], household expenditures [1] and in PKPD analysis [7]. Notice that for  $\delta_2 = 2$ , it reduces to the case of a constant coefficient of variation which in turn is usually handled via a log-transformation of the data. The power transformation may be a very reasonable assumption but in most cases, the coefficient  $\delta_2$  remains unknown and has to be estimated.

### 3 A monitored SAEM-MCMC algorithm

#### 3.1 Computation of the ML estimations

Let  $\theta = (\beta, \mu, \Gamma, \delta, \Lambda)$  denote the vector of parameters of model (1). We have to compute the maximum likelihood (ML) estimates of these parameters by maximizing the log-likelihood of the data vector considered as a function of  $\theta$ . This can be done conveniently in mixed models via the EM algorithm [9] using the random effects as missing data  $\mathbf{u}$  and proceeding according to the following two steps. The E-step evaluates  $Q(\theta, \theta^*) = \mathbb{E}[\log p(\theta; \mathbf{y}, \mathbf{u}) | \mathbf{y}, \theta = \theta^*]$  defined as the conditional expectation of the log-likelihood of the complete data vector  $(\mathbf{y}, \mathbf{u})$  given the observed data  $\mathbf{y}$  and the current value  $\theta^*$  of the parameter vector whereas the M-step maximises it with respect to the parameters so as to update the values of  $\theta^*$ .

More precisely, as far as the variance function (2) is concerned, using  $\mathbf{u} = (\phi, \mathbf{v})$ ,

$$\begin{aligned} \log p(\theta; \mathbf{y}, \mathbf{u}) = & \text{const} - \sum_{ij} \log g(\mathbf{w}_{ij}, \delta, \psi_i) - \frac{1}{2} \sum_{ij} \frac{(y_{ij} - f(\mathbf{z}_{ij}, \beta, \phi_i))^2}{g(\mathbf{w}_{ij}, \delta, \psi_i)^2} \\ & - \frac{I}{2} \log(|\Gamma|) - \frac{1}{2} \sum_{i=1}^I (\phi_i - \mathbf{A}_i \mu)' \Gamma^{-1} (\phi_i - \mathbf{A}_i \mu) \\ & - \frac{1}{2} \log(|\Lambda|) - \frac{1}{2} \mathbf{v}' \Lambda^{-1} \mathbf{v} \end{aligned}$$

and similarly for variance function (3), using just  $\mathbf{u} = \phi$

$$\begin{aligned} \log p(\theta; \mathbf{y}, \mathbf{u}) = & \text{const} - \sum_{ij} \log g(\mathbf{w}_{ij}, \delta, \psi_i) - \frac{1}{2} \sum_{ij} \frac{(y_{ij} - f(\mathbf{z}_{ij}, \beta, \phi_i))^2}{g(\mathbf{w}_{ij}, \delta, \psi_i)^2} \\ & - \frac{I}{2} \log(|\Gamma|) - \frac{1}{2} \sum_{i=1}^I (\phi_i - \mathbf{A}_i \mu)' \Gamma^{-1} (\phi_i - \mathbf{A}_i \mu) \end{aligned}$$

Unfortunately in NLMM's,  $Q(\theta, \theta^*)$  is not in closed form and we have to rely upon a stochastic version of it. Namely, this refers to the SAEM algorithm, defined by Delyon et al. [8] which proved to be more efficient than a classical Monte Carlo EM due to a recycling of simulations from one iteration to the next. When simulation cannot be performed directly, it is replaced by a MCMC procedure (e.g. Metropolis-Hastings algorithm) as proposed by Kuhn and Lavielle [19]. In practice the main problem of this SAEM-MCMC algorithm is to adequately calibrate its parameters. Here, we proceed along the same lines as defined by Duval and Robert-Granié [10] for homoskedastic NLMM's (see details in appendix).

In addition, we achieve the M step via a Levenberg-Marquardt algorithm instead of a classical Newton-Raphson algorithm by introducing an adaptive sequence  $(\lambda_k)_k$  of real numbers [29].

This algorithm interpolates between the Gauss-Newton algorithm and the method of gradient descent. It is more robust than the Gauss-Newton algorithm, which means that in many cases it finds a solution even if it starts very far off the final solution. The non-negative damping factor ( $\lambda_k$ ) is adjusted at each iteration. A similar damping factor appears in Tikhonov regularization [39] which is used to solve linear ill-posed problems.

Given  $\boldsymbol{\theta}^{(k)}$ , the value of  $\boldsymbol{\theta}$  at iteration  $k$ , the SAEM algorithm is implemented as follows:

- Simulation step: draw  $L$  random vectors ( $\mathbf{u}^{(k,l)}$ ;  $l = 1, \dots, L$ ) from the conditional distribution  $p(\mathbf{u}|\mathbf{y}, \boldsymbol{\theta}^{(k,l)})$  using  $L$  independent Metropolis-Hastings chains (see details in appendix).
- Stochastic approximation step:  
Let us define  $\mathbf{S}_{1,i}^{(k)} = \mathbb{E}[\phi_i|\mathbf{y}, \boldsymbol{\theta}^{(k)}]$ ,  $\mathbf{S}_{2,i}^{(k)} = \mathbb{E}[\phi_i \phi_i'|\mathbf{y}, \boldsymbol{\theta}^{(k)}]$  and  $\mathbf{S}_{3,i}^{(k)} = \mathbb{E}[\mathbf{v}\mathbf{v}'|\mathbf{y}, \boldsymbol{\theta}^{(k)}]$  with  $\tilde{\mathbf{S}}_{1,i}^{(k)}$ ,  $\tilde{\mathbf{S}}_{2,i}^{(k)}$  and  $\tilde{\mathbf{S}}_{3,i}^{(k)}$  being their corresponding stochastic estimations, these quantities are updated as follows:

$$\tilde{\mathbf{S}}_{1,i}^{(k)} = \tilde{\mathbf{S}}_{1,i}^{(k-1)} + \gamma_k \left[ \sum_{l=1}^L \phi_i^{(k,l)} / L - \tilde{\mathbf{S}}_{1,i}^{(k-1)} \right] \quad (4)$$

$$\tilde{\mathbf{S}}_{2,i}^{(k)} = \tilde{\mathbf{S}}_{2,i}^{(k-1)} + \gamma_k \left[ \sum_{l=1}^L \phi_i^{(k,l)} \phi_i^{(k,l)'} / L - \tilde{\mathbf{S}}_{2,i}^{(k-1)} \right] \quad (5)$$

$$\tilde{\mathbf{S}}_{3,i}^{(k)} = \tilde{\mathbf{S}}_{3,i}^{(k-1)} + \gamma_k \left[ \sum_{l=1}^L \mathbf{v}^{(k,l)} \mathbf{v}^{(k,l)'} / L - \tilde{\mathbf{S}}_{3,i}^{(k-1)} \right] \quad (6)$$

where  $\gamma_k$  is a decreasing sequence of positive real scalars such that  $\gamma_k = 1$  if  $k < K$  and  $\gamma_k = (k - K)^{-1}$  otherwise.

The three previous formulae pertain to parameters  $(\boldsymbol{\mu}, \boldsymbol{\Gamma}, \boldsymbol{\Lambda})$ . Regarding  $\boldsymbol{\eta} = (\boldsymbol{\beta}, \boldsymbol{\delta})$ , we evaluate via stochastic approximation the following quantities:

$$\tilde{\mathbf{D}}^{(k)} = \sum_{l=1}^L \partial_{\boldsymbol{\eta}} \log p(\mathbf{y}, \mathbf{u}^{(k,l)}; \boldsymbol{\theta}^{(k)}) / L \quad (7)$$

$$\tilde{\mathbf{B}}^{(k)} = \sum_{l=1}^L \partial_{\boldsymbol{\eta}}^2 \log p(\mathbf{y}, \mathbf{u}^{(k,l)}; \boldsymbol{\theta}^{(k)}) / L \quad (8)$$

- Estimation step:  
Maximisation of the  $Q$  function with respect to the parameters gives

$$\boldsymbol{\mu}^{(k)} = \left( \sum_{i=1}^I \mathbf{A}_i' (\boldsymbol{\Gamma}^{(k-1)})^{-1} \mathbf{A}_i \right)^{-1} \sum_{i=1}^I \mathbf{A}_i' (\boldsymbol{\Gamma}^{(k-1)})^{-1} \mathbf{S}_{1,i}^{(k)} \quad (9)$$

$$\boldsymbol{\Gamma}^{(k)} = \frac{1}{I} \sum_{i=1}^I \left( \mathbf{S}_{2,i}^{(k)} - \mathbf{A}_i \boldsymbol{\mu}^{(k)} \mathbf{S}_{1,i}^{(k)'} - \mathbf{S}_{1,i}^{(k)} \boldsymbol{\mu}^{(k)'} \mathbf{A}_i' + \mathbf{A}_i \boldsymbol{\mu}^{(k)} \boldsymbol{\mu}^{(k)'} \mathbf{A}_i' \right) \quad (10)$$

$$\boldsymbol{\Lambda}^{(k)} = \mathbf{S}_{3,i}^{(k)} \quad (11)$$

Regarding  $\boldsymbol{\eta}^{(k)}$ , maximisation is performed according to the Gradient-EM procedure based on a second order algorithm [21]. In this case, we use a Levenberg-Marquardt algorithm defined as:

$$\boldsymbol{\eta}^{(k)} = \boldsymbol{\eta}^{(k-1)} - ((\tilde{\mathbf{M}}^{(k-1)})^{-1} \tilde{\mathbf{D}}^{(k-1)}) \quad (12)$$

where  $\tilde{\mathbf{M}}^{(k)} = \tilde{\mathbf{B}}^{(k)} + \lambda_k \text{diag}(\tilde{\mathbf{B}}^{(k)})$  and  $\lambda_k$  is an adaptative sequence ( $\lambda_k ; k = 1, \dots, K$ ) of real numbers which is increasing if  $Q(\boldsymbol{\theta}^{(k)}; \boldsymbol{\theta}^{(k-1)})$  is itself increasing from one iteration to the next, and decreasing otherwise. More precisely, if  $Q(\boldsymbol{\theta}^{(k)}; \boldsymbol{\theta}^{(k-1)}) > Q(\boldsymbol{\theta}^{(k-1)}; \boldsymbol{\theta}^{(k-2)})$  then  $\lambda_k = \lambda_{k-1}/10$ , otherwise  $\lambda_k = 10\lambda_{k-1}$ .

These are the formulae for the variance function defined in (2). As far as the second variance function is concerned (3), the same formulae apply ignoring equations (6) and (11) pertaining  $\mathbf{S}_{3,i}^{(k)}$  and  $\Lambda^{(k)}$ .

### 3.2 Estimations of the log-likelihood and standard errors

The integration with respect to the random effects of the model involved in the likelihood of the observed data  $p(\boldsymbol{\theta}; \mathbf{y})$  can be estimated conveniently via an importance sampling scheme

$$\hat{p}(\boldsymbol{\theta}; \mathbf{y}) = \frac{1}{S} \sum_{s=1}^S p(\mathbf{y} | \mathbf{u}_s, \boldsymbol{\theta}) \frac{p(\mathbf{u}_s, \boldsymbol{\theta})}{\tilde{p}(\mathbf{u}_s, \boldsymbol{\theta})}$$

where  $p(\mathbf{u}, \boldsymbol{\theta})$  is the distribution of the missing data and  $\mathbf{u}_1, \dots, \mathbf{u}_s, \dots, \mathbf{u}_S$  are drawn from the importance distribution  $\tilde{p}(\mathbf{u}_s, \boldsymbol{\theta})$ .

Here we choose for  $\tilde{p}(\mathbf{u}_s, \boldsymbol{\theta})$ , a Gaussian distribution with mean  $\mathbb{E}(\mathbf{u}_s | \mathbf{y}, \boldsymbol{\theta})$  and variance  $\text{Var}(\mathbf{u}_s | \mathbf{y}, \boldsymbol{\theta})$  calculated from draws of  $\mathbf{u}_s$  obtained at the simulation step of the SAEM algorithm when convergence is reached for  $\boldsymbol{\theta}$ .

This choice should be efficient as this distribution is expected to be close to the true conditional distribution  $p(\mathbf{u}_s | \mathbf{y}, \boldsymbol{\theta})$  which is known to be the one providing a minimum variance estimator of  $p(\boldsymbol{\theta}; \mathbf{y})$  [37].

We can also obtain an estimate of the Fisher information matrix using Louis formula [27]. Let us define the following quantities obtained by simulation:

$$\mathbf{R}_1 = \sum_{l=1}^L \partial_{\boldsymbol{\theta}} \log p(\mathbf{y}, \mathbf{u}^{(k,l)}; \boldsymbol{\theta}^{(k)}) / L$$

$$\mathbf{R}_2 = \sum_{l=1}^L \partial_{\boldsymbol{\theta}}^2 \log p(\mathbf{y}, \mathbf{u}^{(k,l)}; \boldsymbol{\theta}^{(k)}) / L$$

$$\mathbf{R}_3 = \sum_{l=1}^L (\partial_{\boldsymbol{\theta}} \log p(\mathbf{y}, \mathbf{u}^{(k,l)}; \boldsymbol{\theta}^{(k)})) (\partial_{\boldsymbol{\theta}} \log p(\mathbf{y}, \mathbf{u}^{(k,l)}; \boldsymbol{\theta}^{(k)}))' / L$$

Then we calculate at the stochastic approximation step

$$\Delta^{(k)} = \Delta^{(k-1)} + \gamma_k [\mathbf{R}_1 - \Delta^{(k-1)}]$$

$$\begin{aligned} \mathbf{G}^{(k)} &= \mathbf{G}^{(k-1)} + \gamma_k[\mathbf{R}_2 + \mathbf{R}_3 - \mathbf{G}^{(k-1)}] \\ \mathbf{H}^{(k)} &= \mathbf{G}^{(k)} - \Delta^{(k)}(\Delta^{(k)})' \end{aligned}$$

and we approximate the Fisher information matrix by the inverse of  $-\mathbf{H}^{(k)}$  at convergence of  $\hat{\theta}$ .

## 4 Numerical applications

The proposed SAEM-MCMC algorithm was applied to two data sets. The first one described by a linear model allows us to validate the algorithm in comparison with an analytical EM. The second one deals with growth data in chickens modelled via a Gompertz function and illustrates the potential of the algorithm for analyzing real data sets via NLMM's.

### 4.1 Verification on a linear model: Pothoff and Roy's data

A validation of our algorithm was first carried out on the growth data presented originally by Pothoff and Roy [35] and analyzed later on in more detail by Jenrich and Schuller [17] and Verbeke and Molenberghs [41].

Here, the data set analyzed is the incomplete data set conceived by Little and Rubin [26] after deleting nine observations at age 10. Among all the possible models envisioned with respect to mean and covariance structures, we restricted our attention to the random intercept model (or compound symmetry) which turned out to be the simplest one consistent with the data according to Verbeke and Molenberghs [41]. This model can be written as:

$$y_{hij} = A_{hi} + B_{hi}t_j + \sigma_{hij}\varepsilon_{hij}^*$$

where  $y_{hij}$  is the  $j$ -th measurement (distance from the center of the pituitary to the pterygomaxillary fissure in 1/10 mm) made on the  $i$ -th individual nested within the  $h$ -th gender ( $h=1,2$  for males and females respectively) at age  $t_j$  (8, 10, 12 and 14 years);  $A_{hi}$  represents the individual intercept assumed i.i.d.  $\mathcal{N}(\mu_h, \tau^2)$  and  $B_{hi}$  stands for the slope. In a compound symmetry model,  $B_{hi}$  reduces to its fixed components term  $B_{hi} = \beta_h$ , the regression coefficient of response on age for gender  $h$ .  $\sigma_{hij}$  designates the residual standard deviation for observation  $hij$  and the corresponding error term  $\varepsilon_{hij}^*$  are supposed i.i.d.  $\mathcal{N}(0, 1)$ .

Attention here was restricted to variance functions comprising purely fixed effects since only such functions are analytically tractable in EM theory applied to LMM's [12]. We consider a general variance function involving variation with respect to age and gender as follows:

$$\log \sigma_{hij}^2 = \delta_h + \delta_h^* t_j$$

This allows us to describe several sub-models of interest e.g.:

- M0: "Homogeneity":  $\delta_1 = \delta_2, \delta_1^* = \delta_2^* = 0$ ;
- M1: "Age effect":  $\delta_1 = \delta_2, \delta_1^* = \delta_2^*$ ;
- M2: "Gender effect":  $\delta_1 \neq \delta_2, \delta_1^* = \delta_2^* = 0$ ;
- M3: "Age + Gender effects":  $\delta_1 \neq \delta_2, \delta_1^* = \delta_2^*$ ;

- M4: “Age + Gender+ Age\*Gender effects”:  $\delta_1 \neq \delta_2, \delta_1^* \neq \delta_2^*$ ;

The SAEM-MCMC algorithm corresponding to this model was implemented according to Duval and Robert-Granié [10]: see details in the appendix about the criteria values used to calibrate the algorithm. Results for the different submodels are shown in table 1. Likelihood ratio tests clearly highlight the significant effect of gender on the residual variance. Results of the SAEM-MCMC algorithm for Model M4 were compared with those of models M2 and M3 based on an analytical EM algorithm derived along the same lines as in Foulley et al. [12] (table 2). As can be expected, the two algorithms gave very similar results in terms of parameter estimates and their SE (i.e., standard error estimates). In addition, we checked that these values also agreed with the results obtained via SAS-Proc Mixed (results not presented here) which is based not on EM but on a second order algorithm (Newton-Raphson or Fisher scoring).

## 4.2 Application to a non linear mixed model : growth curves in poultry

Data come from a selection experiment on growth in broiler chickens carried out at INRA, Nouzilly station, France, by F Ricard. A divergent selection, described in Mignon-Grasteau et al. [32], was applied to weights at 8 (W8) and 36 (W36) weeks of age and resulted into the following 5 lines: +- (high at W8, low at W36); -+ (low at W8, high at W36); ++ (high at both W8 and W36), -- (low at both W8 and W36) and a control (C) unselected. The data set analyzed concerns a subsample of 10 females per line born at the last generation and recorded at 12 different times (0,4,6,8,12,16,20,24,28,32 and 40 weeks of age).

Following Mignon-Grasteau et al. [32] and Meza et al. [31], these data were analyzed using a Gompertz function corresponding to the following model:

$$y_{hij} = A_{hi} \exp \left( - B_{hi} \exp(-C_{hi}t_j/100) \right) + \sigma_{hij}\varepsilon_{hij}^* \quad (13)$$

where  $y_{hij}$  is the weight performance of the  $i$ -th ( $i=1,\dots,10$ ) hen within the  $h$ -th ( $h=1,\dots,5$ ) line at age  $t_j$ .  $A_{hi}$  represents the asymptotic mature weight of animal  $hi$ ;  $B_{hi}$  is a parameter linking the adult weight to the birth weight and  $C_{hi}$  is the so called “maturation rate”. We assumed that the  $A_{hi}$ ’s are i.i.d.  $A_{hi} \sim \mathcal{N}(a_h, \tau_a^2)$ , the  $C_{hi}$ ’s are i.i.d.  $C_{hi} \sim \mathcal{N}(c_h, \tau_c^2)$ , and correlated to the  $A_{hi}$ ’s according to  $\tau_{ac} = Cov(A_{hi}, C_{hi})$ . Here, due to convergence problems encountered with some softwares,  $B_{hi}$  was restricted to its fixed line effect  $\beta_h$ .  $\sigma_{hij}$  is the residual standard deviation for observation  $hij$  and the error terms  $\varepsilon_{hij}^*$  are assumed i.i.d. with  $\varepsilon_{hij}^* \sim \mathcal{N}(0, 1)$  independent of the  $A_{hi}$ ’s and the  $C_{hi}$ ’s.

Among all possible variance functions involving the effects of age, line and individual, we considered the five ones defined as follows:

V0:  $\log(\sigma_{hij}^2) = \delta$  for the homogeneous case;

V1:  $\log(\sigma_{hij}^2) = \delta_1 + \delta_2 t_j^* + \delta_3 t_j^{*2}$ , i.e. a quadratic adjustment of log-variances according to age defined here for computing convenience as  $t_j^* = (t_j - 20)/100$ ;

V2:  $\log(\sigma_{hij}^2) = \delta_{1h} + \delta_{2h} t_j^* + \delta_{3h} t_j^{*2}$ , the same as V1 but with line specific adjustments;

V3:  $\log(\sigma_{hij}^2) = \delta_{1hi} + \delta_{2h} t_j^* + \delta_{3h} t_j^{*2}$ , the same as V2 but with a random intercept component  $\delta_{1hi} \sim \mathcal{N}(\delta_{1h}, \tau_d^2)$ ;



$$V4: \sigma_{hij}^2 = r f_{hij}^p, \text{ with } f_{hij} = A_{hi} \exp(-B_{hi} \exp(-C_{hi} t_j / 100)).$$

Functions V0, V1, V2 and V3 are special cases of linear models on log-variances defined in (2) while V4 corresponds to the power of the mean model defined in (3).

Global performance of the five variance functions are reported on table 3. As shown by all the criteria displayed (deviance, AIC and BIC) modelling residual variances using any of the four functions mentioned considerably improves the efficiency of the analysis. The addition of the line effect over that of age is especially critical making models more effective as indicated by the very small P-value of the log-likelihood ratio test, and the large change in the AIC (-134.7) and BIC (-111.8) values when contrasting models V1 and V2 respectively. An additional improvement of the AIC (-10.5) and BIC (-8.6), when models V2 and V3 are compared, can also be obtained by considering the intercept  $\delta_{1h}$  of this function as random (V3) which clearly illustrates the potential interest of using random effects in the variance functions. On the other hand, the power of the mean function (V4), which is the most common function used to take heteroskedasticity into account in commercial softwares does not perform better than the Age  $\times$  Line fixed function V2.

An important aspect of this study consisted in contrasting the use of different softwares on this real data application. Softwares listed for the output comparison in addition to our SAEM-MCMC algorithm were the Nlmixed procedure of SAS [25], Monolix running on Matlab (<http://www.monolix.org/>), nlme for Splus and R [34] and finally WinBUGS (<http://www.mrc-bsu.cam.ac.uk/bugs/>). SAS-Nlmixed and Monolix are based on “exact” ML procedures via Gaussian quadratures and Stochastic EM algorithms respectively. nlme is a FOCE approximation method while WinBUGS performs Bayesian Posterior Inference via Gibbs sampling.

All these softwares do not provide the same flexibility in variance modelling, so it was not possible to make a complete comparison in their outputs for all variance functions. V4 can be used in all these softwares, but V1 and V2 cannot be set up in Monolix while V3 -which in this example was the best model- is only available in our SAEM-MCMC algorithm and WinBUGS. The estimates of parameters for this model are shown in table 4 under these two procedures. ML and Bayes estimates are in good agreement, especially for fixed effects of both location and dispersion parameters. Regarding variance components, as expected Bayes estimates are somewhat higher than ML estimates due to the well known bias of the latter [31].

Table 5 provides a comparison of outputs for the V2 variance function using SAEM-MCMC, Nlmixed, nlme and WinBUGS. We observe that the SAEM-MCMC and Nlmixed estimates are very consistent with one another, and this provides a verification of our algorithm in the nonlinear case. nlme outputs are also very close to the previous estimates which indicates a good behaviour of this FOCE algorithm in such an example. The largest comparison between softwares is shown on table 6 for the power variance function. Again, the SAEM-MCMC and Nlmixed estimates coincide perfectly whereas some differences occur with nlme and Monolix results. Regarding standard errors problems arise with Nlmixed, Monolix and nlme due to either non positive definite Hessian matrices (Nlmixed, nlme) or the way some parameters are handled ( $\beta_h$  treated as random with a small variance in Monolix). As in the former table, WinBUGS produces higher variance components estimations for that model V4 than the other algorithms.

Finally, SAEM-MCMC and WinBUGS turn out to be the most flexible and secure softwares among those compared to estimate parameters using either ML or Bayes inference.

## 5 Discussion - Conclusion

This paper represents a further step in the development of nonlinear mixed models by enriching them with a large class of variance functions. The key idea is that heterogeneity of residual variance is not an exception but should be part of model construction as it is for subclass means. Doing so turns out to be very effective for improving the adequacy of the statistical analysis (as well illustrated in the chicken example shown here) as in other practical instances.

In addition to the classical “power of the mean” function, we consider a mixed model extension of the classical linear model on log-variances which provides much flexibility in modelling potential sources of variation in residual variances. This approach is especially interesting as it takes into account major sources of variation in situations with little information per subclass or experimental unit in a very parsimonious way. Treating the source of variation as random is equivalent to constructing a shrinkage estimator of variances combining a population and subject (or subclass) specific estimators: see e.g., Jaffrézic et al. [16] for variance estimation in differentially expressed gene studies. Nevertheless, other variance functions could be envisioned especially in the case of repeated data structures such as those based on stochastic processes (ARCH models of Engle [11]) or on semi-parametric techniques (e.g. B-splines reported by Torres [40]). An important issue that was not covered here consists of incorporating heterogeneity in components of variance and covariance of random effects. This could be done along the same lines as described in linear mixed models by Foulley and Quaas [13]).

Regarding inferential aspects, here the choice was made to rely on maximum likelihood in contrast to other approximation techniques. In fact, we strongly believe that exact statistical procedures are always to be preferred to approximations even though some of them, such as FOCE (via here nlme) can work quite well in some examples. Unfortunately, this is not always the case and it is difficult to know a priori in which conditions and with which models. Here, our attention was focused on ML estimation of all the parameters of the model including dispersion parameters. Correction of bias induced by ML on such parameters could be achieved using REML estimations by integrating fixed effects out the  $f(\cdot)$  function. Again, there is conceptually no difficulty to do that via the EM algorithm as shown by Meza et al. [31] by including those fixed effects as part of the missing data vector.

Finally, ML estimation was obtained with the SAEM-MCMC algorithm. This is a very simple algorithm which works quite well provided its parameters are well calibrated. It is also clearly more reliable than Gaussian quadratures in this kind of model as illustrated here on table 6. This does not mean that the story is over especially concerning the stochastic procedure chosen for the integration step. To that goal, procedures based either on Population Monte Carlo [5] on quasi Monte-Carlo integration [33] might deserve future testing.

**Availability** -The codes of SAEM-MCMC, SAS and Winbugs programmes can be available upon request to the corresponding author.

**Acknowledgements** - The first author is indebted to Ecole Doctorale 'ABIES' for supporting her PhD work. Thanks are also expressed to the INRA Poultry Research Unit of Tours for kindly providing the data on growth measurements.

## Bibliography

- [1] G.E. Battese and B.P. Bonyhady. Estimation of household expenditure functions: an application of a class of heteroscedastic regression models. *The Economic Record*, 57, 80-85, 1981.
- [2] A. Blasco, M. Piles and L. Varona. A bayesian analysis of the effect of selection for growth rate on growth curves in rabbits. *Genetics Selection Evolution*, 35, 21-41, 2003.
- [3] G.E.P Box and W.J. Hill. Correcting inhomogeneity of variances with power transformation wighting. *Technometrics*, 16, 385-389, 1974.
- [4] W.J. Brown, D. Draper, H. Golstein and J. Rasbash. Bayesian and likelihood methods for fitting multilevel models with complex level-1 variation. *Computational Statistics and Data Analysis*, 39, 203-225, 2002.
- [5] O. Cappé, A. Guillin, J.M. Marin and C.P. Robert. Population Monte Carlo. *Journal of Computational and Graphical Statistics*, 13, 907-929, 2004.
- [6] M. Davidian and J. Carroll. Variance Function Estimation. *Journal of the American Statistical Association*, 82, 1079-1091, 1987.
- [7] M. Davidian and D.M. Giltinan. Nonlinear mixed models for repeated Measures Data. New York: Chapman and Hall, 1995.
- [8] B. Delyon, M. Lavielle and E. Moulines. Convergence of a stochastic approximation version of the EM algorithm. *Annals of Statistics*, 27, 94-128, 1999.
- [9] A.P. Dempster, N.M. Laird and D.B. Rubin. Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society: Series B*, 39, 1-38, 1977.
- [10] M. Duval and C. Robert-Granié. SAEM-MCMC: some criteria, 2007. <http://hal.archives-ouvertes.fr/hal-00189580/fr/>
- [11] R.F. Engle. Autoregressive conditional heteroscedasticity with estimates of variance of united Kingdom inflations. *Econometrica* 50, 987-1008, 1982.
- [12] J.L. Foulley, M. San Cristobal, D. Gianola and S. Im. Marginal likelihood and Bayesian approaches to the analysis of heterogeneous residual variances in mixed linear Gaussian models. *Computational Statistics and Data Analysis*, 13, 291-305, 1992.

- [13] J.L. Foulley and R.L. Quaas. Heterogeneous variances in Gaussian linear mixed models. *Genetic Selection Evolution*, 27, 211-228, 1995.
- [14] A. Gelman and D.B. Rubin. Inference from iterative simulation using multiple sequences. *Statistical Science*, 7, 457-511, 1992.
- [15] H. Goldstein. Multilevel models in educational and social research. New York: Oxford University Press, 1987.
- [16] F. Jaffrézic, G. Marot, S. Degrelle, I. Hue and J.L. Foulley. A structural mixed model for variances in differential gene expression studies. *Genetical Research*, 89, 19-25, 2007.
- [17] R.I. Jenrich and M.D. Schuller. Unbalanced repeated-measures models with structured covariance matrices. *Biometrics*, 42, 805-820, 1986.
- [18] G.G. Judge, W.E. Griffiths, R. Carter Hill, H. Lutkepohl and T.C. Lee. The theory and practice of econometrics. New-York: J Wiley and Sons, 1985.
- [19] E. Kuhn and M. Lavielle. Coupling a stochastic approximation version of EM with a MCMC procedure. *ESAIM Probability and Statistics*, 8, 115-131, 2004.
- [20] E. Kuhn and M. Lavielle. Maximum likelihood estimation in nonlinear mixed effects models. *Computational Statistics and Data Analysis*, 49, 1020-1038, 2005.
- [21] K. Lange. A gradient algorithm locally equivalent to the EM algorithm. *Journal of the Royal Statistical Society, Serie B*, 57(2), 425-437, 1995.
- [22] Y. Lee and J.A. Nelder. Double hierarchical generalized linear models (with discussion). *Journal of the Royal Statistical Society: Series C*, 55, 139-180, 2006.
- [23] X. Lin, J. Ray and S.D. Harlow. Linear mixed models with heterogeneous within-cluster variances. *Biometrics*, 53, 910-923, 1997.
- [24] M. Lindstrom and D. Bates. Nonlinear mixed effects models for repeated measures data. *Biometrics*, 44, 673-687, 1990.
- [25] R. Littell, G. Milliken, W. Stroup, R. Wolfinger and O. Schabenberger. SAS for Mixed Models. New York: SAS Institute, 2006.
- [26] R.J.A Little and D.B. Rubin. Statistical analysis with missing data. New-York: Wiley and Sons, 1977.
- [27] T.A. Louis. Finding the observed information matrix when using the EM algorithm. *Journal of the Royal Statistical Society: Series B*, 44, 226-233, 1982.
- [28] J.C. Lu, D. Chen and W. Zhou. Quasi-likelihood estimation for GLM with random scales. *Journal of Statistical Planning and Inference*, 136, 401-429, 2006.
- [29] D. Marquardt. An Algorithm for Least-Squares Estimation of Nonlinear Parameters. *SIAM Journal on Applied Mathematics*. 11, 431-441, 1963.

- [30] C.E. Mc Culloch. Maximum likelihood algorithms for generalized linear mixed models. *Journal of the American Statistical Association*, 92, 162-170, 1987.
- [31] C. Meza, F. Jaffrézic and J.L. Foulley. REML estimation of variance parameters in nonlinear mixed effects models using SAEM algorithm. *The Biometrical Journal*, 49, 876-888, 2007.
- [32] S. Mignon-Grasteau, M. Piles, L. Varona, J.P. Poivey and A. Blasco. Genetic analysis of growth curve parameters for male and female chickens resulting from selection on shape of growth curve. *Journal of Animal Science*, 78, 2532-2531, 2000.
- [33] J. Pan and R. Thompson. Quasi Monte Carlo estimation in generalized linear mixed models. *Computational Statistics and Data analysis*, 51, 5765-5775, 2006.
- [34] J.C. Pinheiro and D.M. Bates. *Mixed-effects Models in S and S-PLUS*. New York: Springer, 2000.
- [35] R.F. Pothoff and S.N. Roy. A generalized multivariate analysis of variance function useful especially for growth curve problems. *Biometrika*, 51, 313-326, 1964.
- [36] R.A. Rigby and D.M. Stasinopoulos. Generalized additive models for location, scale and shape (with discussion). *Applied Statistics*, 54, 507-554, 2005.
- [37] C.P. Robert and G. Casella. *Monte Carlo Statistical methods*. New York: Springer, 2004.
- [38] L.B. Sheiner and S.L. Beal. Evaluation of methods for estimating population pharmacokinetic parameters. I. Michaelis-Menten model: routine clinical pharmacokinetic data. *Journal of pharmacokinetics and Biopharmaceutics*, 8, 553-571, 1980.
- [39] A.N. Tikhonov and V.A Arsenin. *Solution of Ill-posed Problems*, Winston & Sons, Washington, 1977.
- [40] R.A. Torres. Markov chain Monte Carlo methods for estimating the covariance structure of longitudinal data - an application to dairy cattle. PhD thesis, Cornell University, 2001.
- [41] G. Verbeke and G. Molenberghs. *Linear models for longitudinal data*. New York: Springer Verlag, 1999.
- [42] R.D. Wolfinger. Laplacian's approximation for nonlinear mixed models. *Biometrika*, 80, 791-795, 1993.

## 6 Appendix

### 6.1 The Metropolis-Hastings algorithm

The simulation step of the SAEM-MCMC algorithm is implemented via a Metropolis-Hastings algorithm [37]. This MCMC procedure produces Markov chains with stationary distribution being the conditional distribution of the missing data  $\mathbf{u}$  given the observations  $\mathbf{y}$  and the vector of parameters  $\boldsymbol{\theta}$ .

Let us define  $p^{(k)}(\mathbf{u})$  the density of  $\mathbf{u}$ ,  $\mathbf{V}^{(k)}$  the variance of  $\mathbf{u}$  at iteration  $k$ , and  $p(\mathbf{y}|\boldsymbol{\theta}, \mathbf{u})$  the density of the observations given the vector of the parameters  $\boldsymbol{\theta}$  and the missing values  $\mathbf{u}$ .

Let  $m_1 \in \mathbb{N}$ ,  $\rho_1 \in \mathbb{R}^+$ ,  $\rho_2 \in \mathbb{R}^+$ ,  $\rho_2 > \rho_1$  and  $\boldsymbol{\theta}^{(k)}$  the value of  $\boldsymbol{\theta}$  at iteration  $k$  of the SAEM-MCMC algorithm.

Let  $\mathbf{x}^{(t)} = \mathbf{u}^{(k,l,t-1)}$  obtained at the  $(t-1)$ th step of the  $l$ th chain of the Metropolis-Hastings algorithm performed at the  $k$ th step of the SAEM-MCMC algorithm. The next step is implemented as follows.

- If  $t < m_1$ , then generate  $\mathbf{W}_t$  from  $p^{(k-1)}(\mathbf{u})$  with acceptance rate  $\alpha(\mathbf{x}^{(t)}, \mathbf{w}_t) = \min\left(\frac{p(\mathbf{y}|\mathbf{w}_t, \boldsymbol{\theta}^{(k)})}{p(\mathbf{y}|\mathbf{x}^{(t)}, \boldsymbol{\theta}^{(k)})}, 1\right)$
- If  $t \geq m_1$ , then generate  $\rho \sim \mathcal{U}_{[\rho_1, \rho_2]}$  and  $\mathbf{W}_t$  from  $\mathcal{N}(\mathbf{x}^{(t)}, \rho \mathbf{V}^{(k)})$  with acceptance rate  $\alpha(\mathbf{x}^{(t)}, \mathbf{w}_t) = \min\left(\frac{p(\mathbf{y}|\mathbf{w}_t, \boldsymbol{\theta}^{(k)}) p^{(k)}(\mathbf{w}_t)}{p(\mathbf{y}|\mathbf{x}^{(t)}, \boldsymbol{\theta}^{(k)}) p^{(k)}(\mathbf{x}^{(t)})}, 1\right)$
- Finally  $\mathbf{u}^{(k,l,t)} = \begin{cases} \mathbf{w}_t & \text{with probability } \alpha(\mathbf{x}^{(t)}, \mathbf{w}_t) \\ \mathbf{x}^{(t)} & \text{with probability } 1 - \alpha(\mathbf{x}^{(t)}, \mathbf{w}_t) \end{cases}$

Then for the sake of simplicity, we note  $\mathbf{u}^{(k,l)} = \mathbf{u}^{(k,l,T)}$ , where  $T$  is the length of the chain.

### 6.2 Some criteria to calibrate the parameters of the SAEM-MCMC algorithm

Duval and Robert-Granié [10] presented criteria to choose the parameter values of the SAEM-MCMC algorithm. These parameters were calibrated as follows: we run a Markov chain using the Metropolis-Hastings algorithm for one representative individual taking the initial value  $\boldsymbol{\theta}_0$  of the parameters and then the following criteria are used. Finally these parameters are used to run the Metropolis-Hastings algorithm for all the individuals.

- **The iteration at which we switch to another instrumental distribution in the Metropolis-Hastings algorithm, noted  $m_1$ :**

The different behaviors of the chain under the two instrumental distributions lead us to choose

$m_1 = T/10$ , according to Duval and Robert-Granié [10].

- **The second instrumental distribution of the Metropolis-Hastings algorithm:**

$\rho_1$  and  $\rho_2$  are chosen such that the acceptance rate is between 30% and 40 %, according to Robert and Casella [37].

- **The length of each chain in the Metropolis-Hastings algorithm, noted  $T$ :**

We run long Markov chains in the Metropolis-Hasting, such as they are under the stationary distribution. Convergence was tested according to the Gelman and Rubin criteria [14], by comparing within-chain and between-chain variances.

- **The number of independent chains simulated in the Metropolis-Hastings algorithm, noted  $L$ :**

In practice we run several chains and compare the behavior of each of them. If their properties (mean, variance) are significantly different, we may choose  $L = 5$  or  $L = 10$  chains, else,  $L = 1$  chain can be enough [20].

- **Choice of the parameter  $K$  for the sequence  $(\gamma_k)_k$ :**

Let

$$e^{(k)} = \max_j \left( \frac{|\theta_j^{(k)} - \theta_j^{(k-1)}|}{|\theta_j^{(k)}|} \right) \quad (14)$$

We choose  $K$  as the first iteration  $k$  such that  $e^{(k)} < 10^{-2}$ .

- **Stopping rule:**

The SAEM-MCMC algorithm is stopped when  $e^{(k)} < 10^{-4}$ .

### 6.3 Parameters of the SAEM-MCMC algorithm used in Pothoff and Roy's data analysis

Criteria provided the following parameters:  $T = 300$ ,  $m_1 = 30$ ,  $L = 5$ ,  $\rho_1 = 0.5$ ,  $\rho_2 = 1.5$  and  $K = 222$ . An acceptance rate around 30% and a Gelman and Rubin criterion at 1.07 were obtained. Our algorithm stopped at 362 iterations and the EM algorithm at 53 iterations.

$\theta$	$\mu_1$	$\mu_2$	$\beta_1$	$\beta_2$	$\tau^2$	$\delta_1$	$\delta_2$	$\delta_1^*$	$\delta_2^*$
$\theta_0^{(a)}$	200	200	5	12	300	2	7	-2	-2
M0									
$\hat{\theta}$	211.29	225.97	9.78	15.74	311.38	5.30			
SE	6.41	5.51	1.87	1.67	99.92	0.33			
M1									
$\hat{\theta}$	211.47	225.77	9.74	15.91	313.50	5.51		-0.14	
SE	6.70	5.62	1.94	1.67	101.66	0.33		0.20	
M2									
$\hat{\theta}$	211.05	226.04	9.85	15.72	328.00	4.08	5.68		
SE	5.84	5.88	1.06	1.93	101.64	0.26	0.21		
M3									
$\hat{\theta}$	210.97	225.65	9.91	16.00	334.89	4.34	5.96	-0.19	
SE	5.96	6.33	1.06	2.05	103.25	0.39	0.38	0.21	

<sup>(a)</sup> $\theta_0$ =initial value

M0: Homogeneous ( $\delta_1 = \delta_2$  and  $\delta_1^* = \delta_2^* = 0$ ), M1: Age ( $\delta_1 = \delta_2$  and  $\delta_1^* = \delta_2^*$ )

M2: Gender ( $\delta_1^* = \delta_2^* = 0$ ), M3: Age+Gender ( $\delta_1^* = \delta_2^*$ )

	$-2L^{(a)}$	$\Delta(-2L)^{(b)}$
M0	857.23	0.00
M1	856.73	0.50 (P=0.48)
M2	838.87	18.36 (P=2E-5)
M3	838.13	19.10 (P=2E-5)

<sup>(a)</sup>  $-2L = -2 \log \hat{p}(\hat{\theta}; \mathbf{y})$

<sup>(b)</sup>  $\Delta(-2L)$  : variation of the value of  $-2L$  between Homogeneous Model and the corresponding Model M0, M1, M2 and M3 ; and their corresponding p-values

Table 1 – SAEM-MCMC estimates ( $\hat{\theta}$ ) and standard errors (SE) for the analysis of Pothoff and Roy's data for variance functions M0 to M3.



$\theta$	$\mu_1$	$\mu_2$	$\beta_1$	$\beta_2$	$\tau^2$	$\delta_1$	$\delta_2$	$\delta_1^*$	$\delta_2^*$	$-2L^{(a)}$
initial value	200	200	5	12	300	2	7	-2	-2	-
$\hat{\theta}_{EM}$	210.89	225.71	9.94	15.97	336.37	4.37	5.94	-0.21	-0.18	838.13
SE	5.99	6.18	1.07	1.98	104.70	0.63	0.42	0.42	0.25	-
$\hat{\theta}_{SAEM-MCMC}$	210.90	225.66	9.94	15.99	336.24	4.36	5.96	-0.20	-0.18	838.13
SE	6.01	6.22	1.08	1.99	104.74	0.63	0.42	0.43	0.25	-

<sup>(a)</sup>  $-2L = -2 \log \hat{p}(\hat{\theta}; \mathbf{y})$

M4: Age+Gender+Age\*Gender

Table 2 – Comparison of EM and SAEM-MCMC estimates ( $\hat{\theta}$ ) and standard errors (SE) for the analysis of Pothoff and Roy’s set with the variance function M4.

Model	Parameters	$-2L^{(a)}$	Comparison	$\Delta(-2L)$	P-value	AIC	BIC
V0	19	7525.6	-	-	-	7563.6	7599.9
V1	21	7161.6	V0 vs V1	364.0	< 0.0001	7203.6	7243.8
V2	33	7002.9	V1 vs V2	159.6	< 0.0001	7068.9	7132.0
V3	34	6990.4	V2 vs V3	12.5	0.0008	7058.4	7123.4
V4	20	7053.1	V0 vs V4	472.5	< 0.0001	7093.1	7131.4

<sup>(a)</sup>  $-2L = -2 \log \hat{p}(\hat{\theta}; \mathbf{y})$

V0: “Homogeneous”, V1: “Age”, V2: “Age×Line”, V3: “Age × Line+Ind”, V4: “Power”

Table 3 – Comparison of models for residual variance functions based on ML estimation via the SAEM-MCMC algorithm.

parameters	initial value	WinBUGS		SAEM-MCMC	
$\theta$	$\theta_0$	$\hat{\theta}$	SE	$\hat{\theta}$	SE
$a_0$	2500	3011.00	93.02	3025.50	93.61
$\beta_0$	4	4.38	0.04	4.37	0.04
$c_0$	15	11.96	0.52	11.90	0.48
$\tau_a^2$	50000	59960.00	13860.00	51972.00	11357.00
$\tau_c^2$	5	2.43	0.56	1.98	0.44
$\tau_{ac}$	-100	-160.40	67.05	-160.77	55.33
$\tau_d^2$	10	0.14	0.01	0.10	0.10
$\delta_{10}$	5	10.53	0.14	10.96	0.60
$\delta_{20}$	10	16.36	1.31	16.53	1.35
$\delta_{30}$	-100	-114.60	20.65	-117.81	32.63
$-2L^{(a)}$		6558.00 <sup>(b)</sup>		6990.40	
AIC		6814.00 <sup>(c)</sup>		7058.40	

<sup>(a)</sup>  $-2L = -2 \log \hat{p}(\hat{\theta}; \mathbf{y})$

<sup>(b)</sup> “Dhat” in WinBUGS terminology i.e. posterior expectation of deviance

<sup>(c)</sup> DIC criterion

Table 4 – Parameter estimates ( $\hat{\theta}$ ) with their standard errors (SE) for the chicken data with residual variance functions V3 using WinBUGS and SAEM-MCMC.

Parameter	initial value	SAEM-MCMC		Nlmixed		nlme		Winbugs	
$\theta$	$\theta_0$	$\hat{\theta}$	SE	$\hat{\theta}$	SE	$\hat{\theta}$	SE	$\hat{\theta}$	SE
$a_0$	2500	3025.20	85.66	3027.06	86.07	3026.10	81.53	3018.00	93.47
$\beta_0$	4	4.37	0.03	4.37	0.03	4.37	0.03	4.34	0.04
$c_0$	15	11.88	0.48	11.88	0.48	11.88	0.47	11.93	0.53
$\tau_a^2$	50000	52230.00	11247.00	52167.00	11265.00	52168.00		60310.00	13870.00
$\tau_c^2$	5	1.96	0.43	1.96	0.43	1.95		2.42	0.56
$\tau_{ac}$	-100	-159.21	54.69	-158.59	54.85	-158.31		-157.80	66.65
$\delta_{10}$	5	11.02	0.33	11.01	0.33	11.00		10.95	0.34
$\delta_{20}$	10	16.48	1.24	16.49	1.24	16.47		16.34	1.26
$\delta_{30}$	-100	-119.67	19.45	-118.85	16.68	-118.20		-113.10	20.39
$-2L^{(a)}$		7002.88		7003.00		7002.88		6613.70 <sup>(b)</sup>	
AIC		7068.90		7069.00		7068.90		6834.95 <sup>(c)</sup>	
BIC		7132.00		7132.10		7132.00			

<sup>(a)</sup>  $-2L = -2 \log \hat{p}(\hat{\theta}; \mathbf{y})$

<sup>(b)</sup> "Dhat" in WinBUGS terminology i.e. posterior expectation of deviance

<sup>(c)</sup> DIC criterion

Table 5 – Parameter estimates ( $\hat{\theta}$ ) with their standard errors (SE) for residual variance function  $V_2$  for SAEM-MCMC, Nlmixed, nlme procedures and WinBUGS software on chicken data.

Parameter	initial value	SAEM-MCMC		Nlmixed		Monolix	
$\theta$	$\theta_0$	$\hat{\theta}$	SE	$\hat{\theta}$	SE	$\hat{\theta}$	SE
$a_0$	2500	3164.70	83.98	3164.60	83.66	3160.00	80.90
$\beta_0$	5	4.43	0.03	4.43	0.03	4.41	0.03
$c_0$	15	11.76	0.43	11.76	0.43	11.70	0.42
$\tau_a^2$	50000	54694.00	12941.00	54674.00	0.11 <sup>(d)</sup>	48400.00	11100.00
$\tau_c^2$	5	1.58	0.40	1.58	0.39	1.55	0.38
$\tau_{ac}$	-100	-137.15	53.30	-137.17	41.56	-120.51	
$r$	exp(-1)	0.04	0.06	0.03	0.01	0.01	
$p$	1	1.75	0.12	1.75	0.06	1.90	
$-2L^{(a)}$		7053.10		7054.10		7061.04	
AIC		7093.10		7094.40		7101.00	
BIC		7131.34		7132.70		7139.30	

Parameter	initial value	Winbugs		nlme	
$\theta$	$\theta_0$	$\hat{\theta}$	SE	$\hat{\theta}$	SE
$a_0$	2500	3139.00	88.11	3153.10	84.36
$\beta_0$	5	4.43	0.03	4.42	0.03
$c_0$	15	11.83	0.49	11.75	0.43
$\tau_a^2$	50000	64200.00	15740.00	54251.00	
$\tau_c^2$	5	2.16	0.54	1.53	
$\tau_{ac}$	-100	-150.10	67.43	-130.83	
$r$	exp(-1)	0.04	0.02	0.03	
$p$	1	1.72	0.06	1.74	
$-2L^{(a)}$		6719.20 <sup>(b)</sup>		7055.20	
AIC		6911.95 <sup>(c)</sup>		7095.20	
BIC				7133.40	

<sup>(a)</sup>  $-2L = -2 \log \hat{p}(\hat{\theta}; \mathbf{y})$

<sup>(b)</sup> “Dhat” in WinBUGS terminology i.e. posterior expectation of deviance

<sup>(c)</sup> DIC criterion

<sup>(d)</sup> Warning of SAS: hessian not positive definite

Table 6 – Parameter estimates ( $\hat{\theta}$ ) with their standard errors (SE) for residual variance function V4 for SAEM-MCMC and Nlmixed procedures, Monolix and WinBUGS softwares, and nlme procedure on chicken data.