

**INCLUDING MEAN-VARIANCE RELATIONSHIPS IN
HETEROSKEDASTIC MIXED LINEAR MODELS:
THEORY AND APPLICATION**

Jean-Louis Foulley¹

INRA, Station de Génétique quantitative et appliquée,

78352 Jouy-en-Josas Cedex, France

ABSTRACT

In mixed linear models, it is usually assumed that both residual and random effects have homogeneous components of variance. This paper presents models and corresponding techniques of estimation to relax this restrictive assumption. Models proposed include log link functions linearly relating variance components to explanatory variables that can be either discrete or continuous. Special emphasis is given to two aspects of modelling. First, a structural model for residual variances is considered which incorporates, in addition to classical covariates, a function of the data expectation to take into account mean-variance relationships. Secondly, residual and random effect component of variances are linked via a linear functional relationship. Estimation and testing procedures are based on restricted maximum likelihood procedures (REML) via the expectation-maximization (EM) algorithm. The procedure is illustrated by the analysis of birth weight of rats that were used in a toxicology experiment.

Keywords: Mixed models; Heteroskedasticity; Restricted maximum likelihood; EM algorithm

¹ foulley@jouy.inra.fr

1. INTRODUCTION

Mixed models are tools of choice for analyzing correlated data of different kinds (clusters, repeated measurements in time or space). In univariate mixed models, it is usually assumed that both residual and random effects have homogeneous components of variance. However, this assumption can be unrealistic in many practical applications. For instance, there is now a large amount of experimental evidence of heterogeneous variances for most livestock traits in animal science (Brotherstone and Hill, 1986; Garrick et al, 1989; Visscher et al, 1992). Structural linear models for log components of variance have been proposed as an alternative by Foulley et al, (1990, 1992), San Cristobal et al (1993), Foulley and Quaas (1995) which have since been applied on a large scale in animal production and breeding (Weigel et al, 1993; Robert-Granié et al, 1999; San Cristobal et al, 2002).

These models postulate that log residual and/or random effect components of variance are linear functions of covariates, which are observed in the experiment. Different forms of these models have been already considered (Foulley, 1997; Foulley et al., 1998; San-Cristobal et al, 2001).

The objective of this paper is to extend these models such as to take into account mean-variance relationships while keeping the models as parsimonious as possible.

The paper is organised as follows. First, models are described. Secondly, estimating equations based on the EM algorithms are presented. Thirdly, the models are illustrated by analysing data on the birth weights of rats that were used in a toxicology experiment. The paper ends up with a general discussion of modelling and inferential aspects.

2. THEORY

2.1. Models

It is assumed that data can be structured into strata ($i=1, \dots, I$) representing potential sources of heterogeneity so that components of variance are homogenous for observations within stratum. For the sake of simplicity, we consider a one-way random model which can be written as

$$\mathbf{y}_i = \mathbf{X}_i \boldsymbol{\beta} + \sigma_{u,i} \mathbf{Z}_i \mathbf{u}^* + \sigma_{e,i} \mathbf{e}_i^* \quad (1)$$

where $\mathbf{y}_i = \{y_{ij}\}$ is the $(n_i \times 1)$ data vector for stratum i ; $\boldsymbol{\beta}$ is a $(p \times 1)$ vector of unknown coefficients corresponding to the effects of the covariates in the matrix \mathbf{X}_i . The contribution of random effects is expressed as in Foulley and Quaas (1995) 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 $(n_i \times q)$ incidence matrix, and $\sigma_{u,i}$ the square root of the random effect component of variance the value of which may depend on stratum i . Similarly, the residual vector term \mathbf{e}_i is decomposed as $\sigma_{e,i} \mathbf{e}_i^*$ where \mathbf{e}_i^* is a $(n_i \times 1)$ vector of standardized residuals and $\sigma_{e,i}$ is the residual component of variance (e-component) for stratum i .

Classical assumptions are made for the distributions of \mathbf{u}^* and \mathbf{e}_i^* , i.e., $\mathbf{u}^* \sim \mathcal{N}(\mathbf{0}, \mathbf{A})$ (generally \mathbf{A} is the identity matrix \mathbf{I}_q), $\mathbf{e}_i^* \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_q)$ and $E(\mathbf{u}^* \mathbf{e}_i^{*'}) = 0$ for any i .

In order to take into account all potential sources of heteroskedasticity in a parsimonious way, the influence of these factors is modelled by a structural linear model using a log link function (Leonard, 1975; Foulley et al, 1990, 1992),

$$\ln \sigma_{e,i}^2 = \mathbf{p}_i' \boldsymbol{\delta}, \quad (2)$$

where \mathbf{p}'_i is a $(1 \times r)$ row vector of explanatory variables influencing the log residual variances with corresponding coefficient vector $\boldsymbol{\delta}$.

The same can be done for the re-component of variance, or equivalently for the ratio $\tau_i = \sigma_{u,i} / \sigma_{e,i}$, i.e.,

$$\ln \tau_i = \mathbf{h}'_i \boldsymbol{\lambda} \quad (3)$$

where \mathbf{h}'_i designates the row vector of covariates and $\boldsymbol{\lambda}$ the vector of their coefficients.

Models (2) and (3) involve two sets of parameters. In order to make the approach more parsimonious, an alternative form of (3) was proposed by Foulley et al (1998) which links residual to u-component parameters via a functional relationship such as e.g.,

$$\ln \sigma_{u,i} = b_0 + b_1 \ln \sigma_{e,i}, \quad (4)$$

or, equivalently $\sigma_{u,i} / \sigma_{e,i}^{b_1} = \tau$ where $\tau = \exp(b_0)$ is a constant. This parameterization allows a whole range of possibilities in the values of both components of variance retrieving the classical cases of a homogeneous u-component ($b_1 = 0$) and of a constant ratio $\sigma_{u,i} / \sigma_{e,i}$ ($b_1 = 1$).

An extension that may deserve attention consists of incorporating in model (2) an adjustment for the expectation to take into account mean-variance relationships. In its simplest form, assuming now that expectation $\mu_i = E(y_{ij})$ is the same for all the observations of stratum i , the model can be written as

$$\sigma_{e,i}^2 = (\mu_i / \mu_0)^\alpha \sigma_{0,i}^2, \quad (5)$$

where

α is a real real-valued parameter which characterises the form and magnitude of the mean-variance dependency;

. $\sigma_{0,i}^2$ is the adjusted residual variance to a reference basis μ_0 e.g., $\mu_0 = \bar{y} = \left(\sum_{ij} y_{ij} \right) / N$ but any other known value being possible.

The adjusted component $\sigma_{0,i}^2$ is in turn decomposed along the same structural model as in (2) i.e.,

$$\ln \sigma_{0,i}^2 = \mathbf{p}_i' \boldsymbol{\delta}, \quad (6)$$

so that the new model can cope with heteroskedasticity due simultaneously to scale effects and other factors.

2.2. Estimation and computations

Our attention will be restricted here to models defined in (4), (5) and (6), the classical approach including $b_1 = 0$ and 1 having already been presented elsewhere (Foulley et al, 1992; San Cristobal et al, 1993; Foulley, 1997 and Foulley, 2002).

Use is made here of the EM algorithm (Dempster et al, 1977; Foulley, 2002) to compute REML estimates of parameters involved in the variance components.

Letting $\mathbf{y} = (\mathbf{y}'_1, \dots, \mathbf{y}'_i, \dots, \mathbf{y}'_I)'$ and $\boldsymbol{\phi} = (\boldsymbol{\delta}', \alpha, b_0, b_1)'$, we first define the complete data set by $\mathbf{x} = (\mathbf{y}', \mathbf{z}')'$ where the missing part is specified as in Dempster et al, (1977) by $\mathbf{z} = (\boldsymbol{\beta}', \mathbf{u}^{*'})'$ i.e. with $\boldsymbol{\beta}$ treated as a nuisance random variable having an infinite limit distribution. Defining \mathbf{z} that way makes $\boldsymbol{\beta}$ automatically integrated out in the likelihood, providing REML estimates of $\boldsymbol{\phi}$. Using this definition of the complete and incomplete data sets leads to a very simple expression of the complete data likelihood $L(\boldsymbol{\phi}; \mathbf{x}) = \ln p(\mathbf{x} | \boldsymbol{\phi})$ since the density $p(\mathbf{z} | \boldsymbol{\phi})$ does not provide any information about the parameters and data are conditionally independent in the distribution of $\mathbf{y} | \mathbf{z}, \boldsymbol{\phi}$.

At the E step, we determine the conditional expectation of $L(\boldsymbol{\phi}; \mathbf{x})$ given the observed data \mathbf{y} , and the parameters being equal at their current values $\boldsymbol{\phi} = \boldsymbol{\phi}^{[t]}$. This is the $Q(\boldsymbol{\phi}; \boldsymbol{\phi}^{[t]})$ function in the EM terminology which here reduces to:

$$Q(\boldsymbol{\phi}; \boldsymbol{\phi}^{[t]}) = -\frac{1}{2} \left[N \ln 2\pi + \sum_{i=1}^I n_i \ln \sigma_{e,i}^2 + \sum_{i=1}^I E_c^{[t]}(\mathbf{e}_i' \mathbf{e}_i) / \sigma_{e,i}^2 \right], \quad (7)$$

where $E_c^{[t]}(\cdot)$ is a condensed notation for the conditional expectation taken with respect of the distribution of $\mathbf{z} | \mathbf{y}, \boldsymbol{\phi} = \boldsymbol{\phi}^{[t]}$.

At the M step, we update the values of $\boldsymbol{\phi}$ by maximizing $Q(\boldsymbol{\phi}; \boldsymbol{\phi}^{[t]})$ with respect to $\boldsymbol{\phi}$. To make the computation easier, we replace this M step by two sequential maximization steps; this is the so-called Expected Conditional Maximisation algorithm (ECM) as proposed by Meng and Rubin (1993). Letting $\boldsymbol{\phi}_1 = (\alpha, \boldsymbol{\delta}')'$ and $\boldsymbol{\phi}_2 = (b_0, b_1)'$, the two following M steps are applied:

$$\boldsymbol{\phi}_1^{[t+1]} = \arg \max_{\boldsymbol{\phi}_1} Q(\boldsymbol{\phi}_1, \boldsymbol{\phi}_2^{[t]}; \boldsymbol{\phi}^{[t]}), \quad (8)$$

$$\boldsymbol{\phi}_2^{[t+1]} = \arg \max_{\boldsymbol{\phi}_2} Q(\boldsymbol{\phi}_1^{[t+1]}, \boldsymbol{\phi}_2; \boldsymbol{\phi}^{[t]}). \quad (9)$$

The iterative system for computing $\boldsymbol{\phi}_1^{[t+1]}$ can be written as follows

$$\left[\mathbf{P}_1' \mathbf{W}_1 \mathbf{P}_1 \right]^{[t,t]} (\boldsymbol{\phi}_1^{[t,t+1]} - \boldsymbol{\phi}_1^{[t,t]}) = \left[\mathbf{P}_1' \mathbf{v}_1 \right]^{[t,t]}, \quad (10)$$

where

$$\begin{aligned} \cdot \mathbf{P}_{1(I \times 1)} &= (\mathbf{L}_1, \mathbf{P}) \\ \cdot \mathbf{L}_1 &= \{\ln \mu_i - \ln \mu_0\} \text{ and } \mathbf{P} = (\mathbf{p}_1, \dots, \mathbf{p}_i, \dots, \mathbf{p}_I)' \\ \cdot \mathbf{v}_1^{[t,t]} &= \left\{ v_{1,j}^{[t,t]} \right\}, \end{aligned}$$

with the elements $v_{1,j}^{[t,t]}$ of the right hand side being (upper scripts omitted)

$$v_{1,j} = \frac{1}{2} \left\{ \sigma_{e,i}^{-2} \left[E_c(\mathbf{e}_i' \mathbf{e}_i) + b_1 \sigma_{u,i} E_c(\mathbf{u}^* \mathbf{Z}_i' \mathbf{e}_i) \right] - n_i \right\}, \quad (11)$$

$$\cdot \mathbf{W}_{1(I \times I)} = \text{Diag}\{w_{1,i}\},$$

with the elements $w_{1,i}$ of the diagonal matrix

$$w_{1,i} = \frac{1}{2}\sigma_{e,i}^{-2} \left[E_c(\mathbf{e}_i' \mathbf{e}_i) + \frac{1}{2}b_1^2 \sigma_{u,i}^2 E_c(\mathbf{u}^* \mathbf{Z}_i' \mathbf{Z}_i \mathbf{u}^*) + 2b_1(1 - b_1/4) \sigma_{u,i} E_c(\mathbf{u}^* \mathbf{Z}_i' \mathbf{e}_i) \right]. \quad (12)$$

This algorithm to be a true ECM requires iterating the Newton-Raphson procedure within an inner EM cycle until convergence to the conditional maximizer. However, in practice, one can reduce the number of inner iterations to as few as one (Lange, 1995).

If the model does not include heterogeneity in the re-component, one will just solve the system in (10) setting $b_1 = 0$ in (11) and (12).

The iterative system for the second set of parameters $\boldsymbol{\phi}_2 = (b_0, b_1)'$ can be expressed under a similar form as the previous one, i.e.,

$$\left[\mathbf{P}_2' \mathbf{W}_2 \mathbf{P}_2 \right]^{[t,l]} \left(\boldsymbol{\phi}_2^{[t,l+1]} - \boldsymbol{\phi}_2^{[t,l]} \right) = \left[\mathbf{P}_2' \mathbf{v}_2 \right]^{[t,l]}, \quad (13)$$

where

$$\begin{aligned} \cdot \mathbf{P}_{2(I \times 1)} &= (\mathbf{1}_N, \mathbf{L}_2) \text{ with } \mathbf{L}_2 = \{\ln \sigma_{e,i}\}, \\ \cdot \mathbf{v}_2^{[t,l]} &= \{v_{2,i}^{[t,l]}\} \end{aligned}$$

with

$$v_{2,i} = \sigma_{u,i} \sigma_{e,i}^{-2} E_c(\mathbf{u}^* \mathbf{Z}_i' \mathbf{e}_i), \quad (14)$$

$$\cdot \mathbf{W}_{2(I \times I)} = \text{Diag}\{w_{2,i}\}$$

and

$$w_{2,i} = \sigma_{u,i} \sigma_{e,i}^{-2} \left[\sigma_{u,i} E_c(\mathbf{u}^* \mathbf{Z}_i' \mathbf{Z}_i \mathbf{u}^*) - E_c(\mathbf{u}^* \mathbf{Z}_i' \mathbf{e}_i) \right]. \quad (15)$$

The elements of \mathbf{v}_1 , \mathbf{v}_2 , \mathbf{W}_1 and \mathbf{W}_2 can also be written as functions of the conditional expectations of

$$S_{\varepsilon\varepsilon,i} = (\mathbf{y}_i - \mathbf{X}_i \boldsymbol{\beta})' (\mathbf{y}_i - \mathbf{X}_i \boldsymbol{\beta}), \quad S_{\varepsilon u,i} = (\mathbf{y}_i - \mathbf{X}_i \boldsymbol{\beta})' \mathbf{Z}_i \mathbf{u}^* \quad \text{and} \quad S_{uu,i} = \mathbf{u}^* \mathbf{Z}_i' \mathbf{Z}_i \mathbf{u}^*.$$

These quantities themselves can be computed as functions of the statistics $\mathbf{X}_i' \mathbf{y}_i$, $\mathbf{Z}_i' \mathbf{y}_i$, $\mathbf{y}_i' \mathbf{y}_i$ and of elements of Henderson's mixed model equations

$$\left(\sum_{i=1}^I \sigma_{e,i}^{-2} \mathbf{T}_i' \mathbf{T}_i + \Sigma^- \right) \hat{\boldsymbol{\theta}} = \sum_{i=1}^I \sigma_{e,i}^{-2} \mathbf{T}_i' \mathbf{y}_i, \quad (16)$$

where $\mathbf{T}_i = (\mathbf{X}_i, \sigma_{u,i} \mathbf{Z}_i)$, $\hat{\boldsymbol{\theta}} = (\hat{\boldsymbol{\beta}}', \hat{\mathbf{u}}^*)'$ and $\Sigma^- = \begin{pmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{A}^{-1} \end{pmatrix}$.

Simplifications arise with grouped data which happens when covariates in \mathbf{X}_i and \mathbf{Z}_i are discrete. Then, $\mathbf{X}_i = \mathbf{1}_{n_i} \mathbf{x}_i'$, $\mathbf{Z}_i = \mathbf{1}_{n_i} \mathbf{z}_i'$, and the formulae for the S terms are :

$$S_{\varepsilon\varepsilon,i} = \sum_{j=1}^{n_i} (y_{ij} - \hat{\mu}_i)^2 + n_i \text{tr}(\mathbf{x}_i \mathbf{x}_i' \mathbf{C}_{\beta\beta}), \quad (17a)$$

$$S_{\varepsilon u,i} = n_i \left[\hat{u}_i^* (y_i - \hat{\mu}_i) - \text{tr}(\mathbf{z}_i \mathbf{x}_i' \mathbf{C}_{\beta u}) \right], \quad (17b)$$

$$S_{uu,i} = n_i \left[\hat{u}_i^{*2} + \text{tr}(\mathbf{z}_i \mathbf{z}_i' \mathbf{C}_{uu}) \right], \quad (17c)$$

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

One may also employ a score version of the systems (10) and (13) by replacing \mathbf{W}_1 and \mathbf{W}_2 by their expectations. This means that $w_{1,i}$ and $w_{2,i}$ in (12) and (15) are changed into $\tilde{w}_{1,i} = 1/2 \left[n_i + b_1^2 \sigma_{u,i}^2 \sigma_{e,i}^{-2} \text{tr}(\mathbf{A} \mathbf{Z}_i' \mathbf{Z}_i) / 2 \right]$ and $\tilde{w}_{2,i} = \sigma_{u,i}^2 \sigma_{e,i}^{-2} \text{tr}(\mathbf{A} \mathbf{Z}_i' \mathbf{Z}_i)$ respectively.

Elements of the mixed model equations in (16) are also useful for computing the value of the log residual likelihood (RL). As shown by Foulley and Quaas (1995), the expression for $-2RL$ reduces to:

$$-2RL = \ln |\mathbf{M}| + \sum_{i=1}^I \sigma_{e,i}^{-2} \mathbf{y}_i' (\mathbf{y}_i - \mathbf{T}_i \hat{\boldsymbol{\theta}}) + C, \quad (18)$$

where $\mathbf{M} = \sum_{i=1}^I \sigma_{e,i}^{-2} \mathbf{T}_i' \mathbf{T}_i + \Sigma^-$ is the coefficient matrix in (16), $\hat{\boldsymbol{\theta}}$ is a solution to the mixed model equations and the constant C is equal to $C = [N - r(\mathbf{X})] \ln 2\pi + \ln |\mathbf{A}|$.

3. APPLICATION

3.1. Data set and basic model

The data set contains birth weight records of rat pups whose mothers were used in a toxicology experiment. This experiment involves 27 females allocated at random to 3 treatments: control (C), low (L) and high (H) dose of an experimental component. There were 10, 10 and 7 females in the C, L, and H groups respectively. Litter size in which pups are born and sex of pups are important factors of variation in birth weight so that they were introduced into the model as covariates in addition to treatment. A mixed model was proposed with a random litter (mother) effect to take into account variation between and within litters, and correlation of pups within litters (Dempster et al, 1984).

The model can then be written as

$$y_{ijkl} = \mu + t_i + \beta x_{ij} + u_{ij} + s_k + e_{ikl}, \quad (19)$$

where y_{ijkl} is the birth weight of the l -th pup from the j -th litter allocated to the i -th treatment with the k -th sex; μ represents a mean; t_i is the fixed i -th treatment effect, βx_{ij} measures the effect of litter size x_{ij} of the j -th mother in the i -th treatment; u_{ij} is the random effect of the corresponding female; s_k is the fixed effect of the k -th sex, and e_{ikl} is a residual within treatment, sex and litter component.

Data have been analyzed this way by Dempster et al (1984) and Davidian and Giltinian (1995) assuming random intercepts that were independently and normally distributed with mean zero and variance σ_u^2 constant i.e., $u_{ij} \sim_{\text{iid}} \mathcal{N}(0, \sigma_u^2)$, and similarly for residual terms $e_{ijkl} \sim_{\text{iid}} \mathcal{N}(0, \sigma_e^2)$. More recently, Rosa et al (2003) examined alternative distribution assumptions for the residuals e.g., Student-t or contaminated normal. However, no analysis specifically considered heteroskedastic models.

3.2. Heteroskedastic models

The first step consists of investigating structural models for the residuals as described in (2). Models envisioned include potential effects of treatment (T), sex (S) and litter size (L) discretized into three classes: small ($L \leq 9$), average ($10 \leq L \leq 13$) and large ($L \geq 14$). Models were compared using likelihood ratio statistics and Schartz's information criterion (BIC). Results are given in table I. They showed that sex was not significant source of variation in residual variances unlike treatment and litter size. This results in choosing the model based on these two factors. The choice was made assuming the component of variance among mothers (litters) was constant. In fact, there is no need to introduce heteroskedasticity at that level as shown by comparison of models with different u-components structures (Table II).

Next, we investigated the potential need for introducing a mean-variance relationship to explain heterogeneity of residual variances. Several possibilities were tested starting from the simplest ones. One interesting combination comprises this relationship plus a sex effect. The corresponding model (symbolized by μ^α, μ^*+S^*) requires 4 dispersion parameters (three for the residual plus one for the mother component σ_u^2) with a BIC value of 3265 versus 6 parameters for the previously selected structural model : "Treatment" + "Litter size" (in short $\mu^*+T^*+L^*$) (five for the residual part plus one for σ_u^2) and a BIC value of 3269. It is even possible to improve upon this model by inserting a functional relationship between random effect and residual components of variance as in (4), the null hypothesis ($b_1 = 0$) being rejected as shown in table III and BIC being now equal to 3260.

The analysis clearly identifies a strong mean-variance relationship ($\alpha = 8.66$) as anticipated graphically (figure 1). The negative sex effect

observed (male –female=-0.85 on the log-scale of residual variance or ratio of male to female residual standard deviation being equal to $0.65 = \exp(-\frac{1}{2} \times 0.85)$) should be understood everything being equal i.e., for pup rats having the same subclass mean but with different genders. Moreover, a negative relationship is observed between the variation in the residual and mother components of variance so that $\frac{\partial \sigma_{u,i}}{\sigma_{u,i}} = -1.78 \frac{\partial \sigma_{e,i}}{\sigma_{e,i}}$.

This negative relationship tends to reduce at the overall data level the large amount of heterogeneity seen on the residual variance. Nevertheless, the introduction of heteroskedasticity in the model has a clear impact on the estimation of some of fixed effects as seen in table IV. The sex effect is slightly decreased as compared to its estimation in the completely homogeneous model ($s_M - s_F=35.91$ vs 33.05 in the original and final models respectively) while the effect of litter size has slightly increased ($|\beta|=12.9$ vs 13.6). However, the main changes occur on the treatment effects. For instance, the effect of the “Low dose” as compared to the “Control” varied from -42.9 to -59.6 in the original and final models respectively i.e., a difference of 16.7 dcg or about one third ($16.7/51$) of the overall standard deviation.

Diagnostic tools such as Quantile-Quantile plots of residuals also clearly showed that heteroskedastic models fit better the data than homoskedastic ones (figure 2ab).

4. DISCUSSION

The approach described in this paper represents a further step in analyzing heterogeneous variances within the framework of linear mixed models.

Including a data variance proportional to a power of the data expectation had been envisioned by several authors: see e.g., textbooks by Judge et al (1985) (chapter 11) and Seber and Wild (2003), chapter 2.8. Box and Hill

(1974) have shown the relationship existing between the coefficient of this power and the parameter of a Box-Cox transformation for stabilizing variance. The interest of our approach is to combine this modelling device with others such as the structural approach based on exogenous explanatory variables and heterogeneity of both residual and random effect components of variance to make models more realistic. At this stage, models are specified very simply to make them parsimonious. This is especially important as far as dispersion parameters are concerned, since there is generally less information in the data about them as there is for location parameters. However, other mean-variance dependency functions can be considered (Davidian and Carroll, 1987) as well as more complicated relationships between random and residual components of variance. The procedure described here is flexible enough to cope with all these modelling variants.

Regarding inferential aspects, our choice was directed towards EM-REML type procedures although other options could have been envisioned. In particular, ML was discarded as it is known to be very sensitive regarding fixed effects with respect to misspecification in the functional form of the variance contrarily to GLS (Carroll and Ruppert, 1982a). Our approach has close links with the structural inference analysis developed by Levenbach (1973) and the iterative procedures developed by Box and Hill (1974), Pritchard et al (1977) for purely fixed linear models and variance functions involving only a mean-variance relationship. Connections also exist with the different weighted least squares procedures investigated by Beal and Sheiner (1988). All these methods rely on iterative GLS and some approximation is made when integrating out fixed effects to get the marginal or posterior density of dispersion parameters (Seber and Wild, 2003). Intensive stochastic procedures would be required to overcome this difficulty.

This method might be also extended to take into account other distributions than the normal, such as the lognormal and the gamma. This again will imply implementation of MCMC algorithms (Robert and Casella, 1999; Sorensen and Gianola, 2002). Finally, as results are known to be sensitive to distribution assumptions and to outliers, robust procedures should be examined along the lines of current works in this area (see e.g., Carroll and Ruppert, 1982b).

References

- Beal S.L., Sheiner L.B. (1988), Heteroscedastic nonlinear regression, *Technometrics*, 30, 327-338
- Box G.E.P., Hill W.J. (1974), Correcting inhomogeneity of variance with power transformation weighting, *Technometrics*, 16, 385-389.
- Brotherstone S., Hill W.G. (1986), Heterogeneity of variance amongst herds for milk production, *Animal Production*, **42**, 297-303.
- Carroll R.J., Ruppert D. (1982a), A comparison between maximum likelihood and generalized least squares in a heteroscedastic model, *Journal of the American Statistical Association*, 77, 878-882.
- Carroll R.J., Ruppert D. (1982b), Robust estimation in heteroscedastic linear models, *The Annals of Statistics*, 10, 429-441.
- Davidian M., Carroll R.J. (1987), Variance function estimation, *Journal of the American Statistical Association*, 82, 1079-1091.
- Davidian M., Giltinian D.M. (1995), *Nonlinear models for repeated measurement data*, Chapman & Hall, New-York.
- Dempster A., Laird N., Rubin R. (1977), Maximum likelihood estimation from incomplete data via the EM algorithm, *Journal of the Royal Statistical Society B*, 39,1-38.
- Dempster A., Selwyn M.R., Patel C.M., Roth A.J. (1984) Statistical and computational aspects of mixed model analysis, *Applied Statistics*, 33, 203-214.
- Foulley J.L. (1997), ECM approaches to heteroskedastic mixed models with constant variance ratios, *Genetics Selection Evolution*, 29, 297-318.
- Foulley J.L. (2002), Algorithme EM: théorie at application au modèle mixte, *Journal de la Société Française de Statistique*, 143, 57-109.
- Foulley J.L., Gianola D., San Cristobal M., Im S. (1990), A method for assessing extent and sources of heterogeneity of residual variances in mixed linear models, *Journal of Dairy Science*, **73**, 1612-1624.
- Foulley J.L., San Cristobal M., Gianola D., Im S. (1992), Marginal likelihood and Bayesian approaches to the analysis of heterogeneous residual variances in mixed linear models, *Computational Statistics and Data Analysis*, **13**, 291-305.
- Foulley J.L., Quaas R.L. (1995), Heterogeneous variances in Gaussian linear mixed models, *Genetics Selection Evolution*, **27**, 211-228.

- Foulley J.L., Quaas R.L., Thaon d'Arnoldi C. (1998), A link function approach to heterogeneous variance components, *Genetics Selection Evolution*, **30**, 27-43.
- Garrick D.J., Pollak E.J., Quaas R.L., Van Vleck L.D. (1989), Variance heterogeneity in direct and maternal weight traits by sex and percent purebred for Simmental sired calves, *Journal of Animal Science*, **67**, 2515-2528.
- Lange K. (1995), A gradient algorithm locally equivalent to the EM algorithm. *Journal of the Royal Statistical Society B*, **57**, 425-437.
- Leonard T. (1975), Bayesian approach to the linear model with unequal variance, *Technometrics*, **17**, 95-102.
- Levenbach H. (1973), The estimation of heteroscedasticity from a marginal likelihood function, *Journal of the American Statistical Association*, **68**, 436-439.
- Meng X.L., Rubin D.B. (1993), Maximum likelihood estimation via the ECM algorithm: a general framework, *Biometrika*, **80**, 267-278.
- Pritchard D.J., Downie J. Bacon D.W. (1977), Further consideration of heteroskedasticity in fitting kinetic models. *Technometrics*, **19**, 109-115
- Robert C.P., Casella G. (1999), *Monte Carlo Statistical Methods*, Springer, Berlin.
- Robert-Granié C., Bonaiti B., Boichard D., Barbat A. (1999), Accounting for variance heterogeneity in French dairy cattle genetic evaluation, *Livestock Production Science*, **60**, 343-357
- Rosa G.J.M., Padovani C.R., Gianola D. (2003), Robust linear mixed models with normal/independent distributions and Bayesian MCMC implementation, *Biometrical Journal*, **45**, 573-590.
- San Cristobal M., Foulley J.L., Manfredi E. (1993), Inference about multiplicative heteroskedastic components of variance in a mixed linear Gaussian model with an application to beef cattle, *Genetics, Selection, Evolution*, **25**, 3-30
- San Cristobal M., Elsen J.M., Bodin L., Chevalet C. (2001), Prediction of the response to a selection for canalisation of a continuous trait in animal breeding, *Genetics Selection Evolution*, **33**, 249-271
- San Cristobal M., Robert-Granié C., Foulley J.L. (2002), Hétéroscédasticité et modèles linéaires mixtes: théorie et applications en génétique quantitative, *Journal de la Société Française de Statistique*, **143**, 155-165.
- Seber G.A.F., Wild C.J. (2003), *Nonlinear regression*, Wiley and Sons, New-York
- Sorensen D., Gianola D. (2002), *Likelihood, Bayesian, and MCMC Methods in Quantitative Genetics*, Springer Verlag.

- Visscher P.M., Thompson R., Hill W.G. (1991), Estimation of genetic and environmental variances for fat yield in individual herds and an investigation into heterogeneity of variance between herds. *Livestock Production Science*, 28, 273-290
- Weigel K.A., Gianola D., Yandel B.S., Keown J.F. (1993) Identification of factors causing heterogeneous within herd variance components using a structural model for variances. *Journal of Dairy Science*, 76, 1466-1478

Figure 1. Mean variance relationship

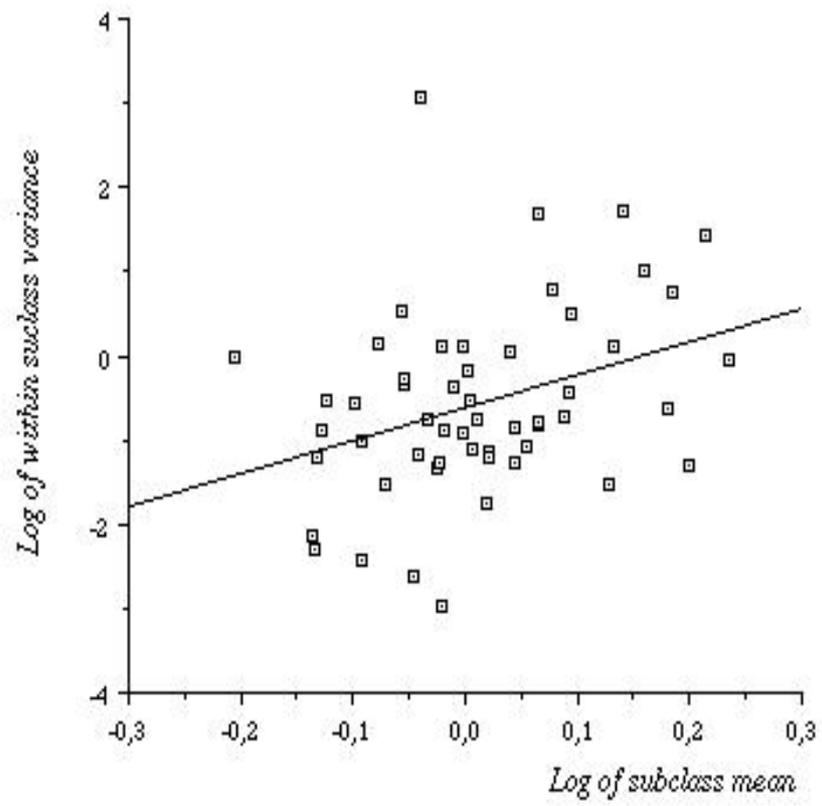


Figure 2. Q-Q plots of residuals:

a) completely homogeneous model (No 1); b) heterogeneous model (No 4)

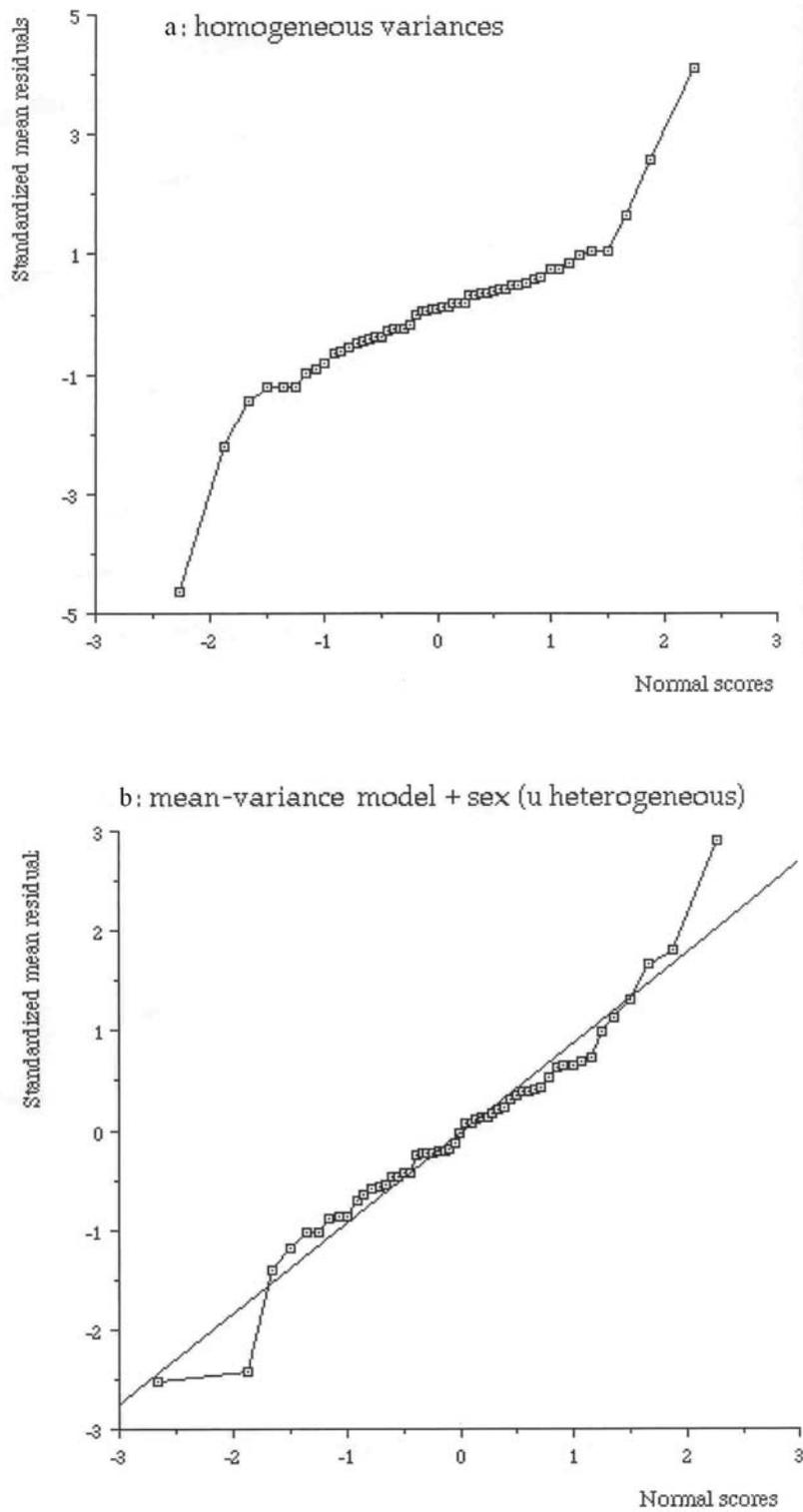


Table I: Likelihood statistics for testing dispersion parameters in heteroskedastic mixed models: choice of a structural model for residual variances

No	Model				Likelihood		Test				
	Location	Residual	Dam	par	-2L	-2BIC	H ₀	Comp	df	Statistic	P-value
(1)	$\mu+T+S+L$	$\mu^*+T^*+S^*+L^*$	μ' : σ_u cst	7	3232.0704	3272.3828					
(2)	$\mu+T+S+L$	$\mu^*+S^*+L^*$	μ' : σ_u cst	5	3259.2976	3288.0921	T*=0	2-1	2	27.2272	10E-6
(3)	$\mu+T+S+L$	$\mu^*+T^*+L^*$	μ' : σ_u cst	6	3233.9926	3268.5460	S*=0	3-1	1	1.9222	0.1656
(4)	$\mu+T+S+L$	$\mu^*+T^*+S^*$	μ' : σ_u cst	5	3271.3391	3300.1336	L*=0	4-1	2	39.2687	3E-9

a: Covariates: μ : intercept; T: treatment; S:sex ; L: litter size

Covariates for log parameters coded as: (eg for Treatment T), location (T), residual variance (T*), T-component (T')

b: par: number of dispersion parameters: ; L=maximum of the loglikelihood (REML version)

c: Likelihood ratio (LR) test based on difference in -2L between the full and reduced models;

Table II: Likelihood statistics for testing dispersion parameters in heteroskedastic mixed models: choice of a model for the u or ratio components

No	Location	Model			Likelihood		Test				
		Residual	Dam	par	-2L	-2BIC	H ₀	Comp	df	Statistic	P-value
(1)	$\mu+T+S+L$	$\mu^*+T^*+L^*$	$\mu'+T'+L'$, or $\mu''+T''+L''$	10	3230.86	3288.45					
(2)	$\mu+T+S+L$	$\mu^*+T^*+L^*$	$\ln \sigma_{u_i} = b_0 + b_1 \ln \sigma_{e_i}$	7	3232.36	3272.67	Model 2	2-1	3	1.5046	0.6812
(3)	$\mu+T+S+L$	$\mu^*+T^*+L^*$	μ'' : ratio cst	6	3239.31	3273.86	$T''=L''=0$ $b_1 = 1$	3-1 3-2	4 1	8.4495 6.9448	0.0764 0.0084
(4)	$\mu+T+S+L$	$\mu^*+T^*+L^*$	μ' : σ_u cst	6	3233.99	3268.55	$T'=L'=0$ $b_1 = 0$	4-1 4-2	4 1	3.1362 1.6315	0.5353 0.2015

a: Covariates: μ : intercept; T: treatment; S:sex ; L: litter size

Covariates for log parameters coded as: (eg for Treatment T), location (T), residual variance (T*), T-component (T'), and u to residual variance log-ratio (T'')

b: par: number of dispersion parameters: L=maximum of the loglikelihood (REML version)

c: Likelihood ratio (LR) test based on difference in -2L between the full and reduced models;

Table III: Likelihood statistics for testing dispersion parameters in heteroskedastic mixed models: choice of a model for residual variances including a mean-variance relationship

No	Model ^a			par	Likelihood ^b		Test ^c				
	Location	Residual	Dam		-2L	-2BIC	H ₀	Comp	df	Statistic	P-value
(1)	$\mu+T+S+L$	μ^*	μ' : σ_u cst	2	3316.68	3328.20					
(2)	$\mu+T+S+L$	μ^α, μ^*	μ' : σ_u cst	3	3259.16	3276.44	$\alpha=0$	2-1	1	57.51	1^E-14
(3)	$\mu+T+S+L$	μ^α, μ^*+S^*	μ' : σ_u cst	4	3241.88	3264.92	$S^*=0$	3-2	1	17.28	3^E-5
(4)	$\mu+T+S+L$	μ^α, μ^*+S^*	(b_0, b_1)	5	3231.26	3260.06	$b_1 = 0$	4-3	1	10.62	0.001

a: Covariates: μ : intercept; T: treatment; S:sex ; L: litter size

Covariates for log parameters coded as: (eg for Treatment T), location (T), residual variance (T*), T-component (T')

b: par: number of dispersion parameters; ; L=maximum of the loglikelihood (REML version); -2BIC=-2L+par log(N-#fixed effects)

c: Likelihood ratio (LR) test based on difference in -2L between the full and reduced models;

Table IV : Estimation of parameters in different models for birth weight of pup rats

Parameter	(1)	(2)	(3)	(4)
Intercept	640.28±11.19	643.21±12.35	644.76±10.16	650.29±8.80
Treatment	-42.85±15.04	-44.77±15.85	-48.44±15.23	-59.56±13.70
	-85.87±18.18	-90.30±19.85	-82.49±18.45	-92.23±17.18
Sex	35.91±4.75	34.18±3.91	34.03±3.89	33.05±4.68
Litter size	-12.90±1.88	-13.19±2.20	-13.20±2.01	-13.56±2.17
-2BIC	3328.28	3268.55	3272.67	3260.06

Intercept : control, female, litter size equal to 12

Treatment : 1st row : Low dose-Control ; 2nd row: High dose-Control

Sex : male-female ; Litter size : coefficient of regression

(1) Homogeneous ; (2) Residual= μ +treatment+litter size

(3) Residual=idem ; $\sigma_{u_i} / \sigma_{e_i}^{h_i} = cst$; (4) Residual= μ^{α} , sex ; $\sigma_{u_i} / \sigma_{e_i}^{h_i} = cst$