3

Original article

Inference about multiplicative heteroskedastic components of variance in a mixed linear Gaussian model with an application to beef cattle breeding

M San Cristobal [1*], JL Foulley [1**], E Manfredi [2]

[1] *INRA, Station de Génétique Quantitative et Appliquée,
78352 Jouy-en-Josas Cedex;*
[2] *INRA, Station d'Amélioration Génétique des Animaux, BP,27,
31326 Castanet-Tolosan Cedex, France*

Summary − A statistical method for identifying meaningful sources of heterogeneity of residual and genetic variances in mixed linear Gaussian models is presented. The method is based on a structural linear model for log variances. Inference about dispersion parameters is based on the marginal likelihood after integrating out location parameters. A likelihood ratio test using the marginal likelihood is also proposed to test for hypotheses about sources of variation involved. A Bayesian extension of the estimation procedure of the dispersion parameters is presented which consists of determining the mode of their marginal posterior distribution using log inverted chi-square or Gaussian distributions as priors. Procedures presented in the paper are illustrated with the analysis of muscle development scores at weaning of 8 575 progeny of 142 sires in the Maine-Anjou breed. In this analysis, heteroskedasticity is found, both for the sire and residual components of variance.

heteroskedasticity / mixed linear model / Bayesian technique

Résumé − **Inférence sur une hétérogénéité multiplicative des composantes de la variance dans un modèle linéaire mixte gaussien: application à la sélection des bovins à viande.** *Une méthode statistique est présentée, capable d'identifier les sources significatives d'hétérogénéité de variances résiduelles et génétiques dans un modèle linéaire mixte gaussien. La méthode est fondée sur un modèle structurel de décomposition du logarithme des variances. L'inférence concernant les paramètres de dispersion est basée sur la vraisemblance marginale obtenue après intégration des paramètres de position. Un*

* Correspondence and reprints
** *Adresse actuelle: Laboratoire de génétique cellulaire, BP 27, 31326 Castanet Tolosan Cedex*

test du rapport des vraisemblances utilisant la vraisemblance marginale est aussi proposé afin de tester des hypothèses sur différentes sources de variation. Une extension bayésienne de la procédure d'estimation des paramètres de dispersion est présentée ; elle consiste en la maximisation de leur distribution marginale a posteriori, *pour des distributions* a priori log χ^2 *inverse ou gaussienne. Les procédures présentées dans ce papier sont illustrées par l'analyse de notes de pointages sur le développement musculaire au sevrage de 8 575 jeunes veaux de race Maine-Anjou, issus de 142 pères. Dans cette analyse, une hétéroscédasticité a été trouvée sur les composantes père et résiduelle de la variance.*

hétéroscédasticité / modèles linéaires mixtes / techniques bayésiennes

INTRODUCTION

One of the main concerns of quantitative geneticists lies in evaluation of individuals for selection. The statistical framework to achieve that is nowadays the mixed linear model (Searle, 1971), usually under the assumptions of normality and homogeneity of variances. The estimation of the location parameters is performed with BLUE–BLUP (Best Linear Unbiased Estimation–Prediction), leading to the well-known Mixed Model Equations (MME) of Henderson (1973), and REML (acronym for REstricted -or REsidual- Maximum Likelihood) turns out to be the method of choice for estimating variance components (Patterson and Thompson, 1971).

However, heterogeneous variances are often encountered in practice, *eg* for milk yield in cattle (Hill *et al*, 1983; Meinert *et al*, 1988; Dong and Mao, 1990; Visscher *et al*, 1991; Weigel, 1992) for meat traits in swine (Tholen, 1990) and for growth performance in beef cattle (Garrick *et al*, 1989). This heterogeneity of variances, also called heteroskedasticity (McCullogh, 1985), can be due to many factors, *eg* management level, genotype × environment interactions, segregating major genes, preferential treatments (Visscher *et al*, 1991).

Ignoring heterogeneity of variance may reduce the reliability of ranking and selection procedures although, in cattle for instance, dam evaluation is likely to be more affected than sire evaluation (Hill, 1984; Vinson, 1987; Winkelman and Schaeffer, 1988).

To overcome this problem, 3 main alternatives are possible. First, a transformation of data can be performed in order to match the usual assumption of homogeneity of variance. A log transformation was proposed by several authors in quantitative genetics (see *eg* Everett and Keown, 1984; De Veer and Van Vleck, 1987; Short *et al*, 1990, for milk production traits in cattle). However, while genetic variances tend to stabilize, residual variances of log-transformed records are larger in herds with the lowest production level (De Veer and Van Vleck, 1987; Boldman and Freeman, 1990; Visscher *et al*, 1991).

The second alternative is to develop robust methods which are insensitive to moderate heteroskedasticity (Brown, 1982).

The last choice is to take heteroskedasticity into account. Factors (*eg* region, herd, year, parity, sex) to adjust for heterogeneous variances can be identified. But such a stratification generates a very large number of cells (800 000 levels of herd × year in the French Holstein file) with obvious problems of estimability. Hence, it is logical to handle unequal variances in the same way as unequal means, *ie*

via a modelling (or structural) approach so as to reduce the parameter space, by appropriate identification and testing of meaningful sources of variation of such variances.

The model for the variance components is described in the *Model* section. Model fitting and estimation of parameters based on marginal likelihood procedures are presented in the *Estimation of Parameters*, followed by a test statistic in *Hypothesis Testing*. A Bayesian alternative to maximum marginal likelihood estimation is presented in *A Bayesian Approach to a Mixed Model Structure* In the *Numerical application* section, data on French beef cattle are analyzed to illustrate the procedures given in the paper. Finally, some comments on the methodology are made in the *Discussion and Conclusion*.

MODEL

Following Foulley *et al* (1990, 1992) and Gianola *et al* (1992), the population is assumed to be stratified into I subpopulations, or strata (indexed by $i = 1, 2, \ldots, I$) with an $(n_i \times 1)$ data vector $\mathbf{y}_i$, sampled from a normal distribution having mean $\boldsymbol{\mu}_i$ and variance $\mathbf{R}_i = \sigma_{e_i}^2 \mathbf{I}_{n_i}$. Given $\boldsymbol{\mu}_i$ and $\mathbf{R}_i$

$$\mathbf{y}_i | \boldsymbol{\mu}_i, \mathbf{R}_i \sim N(\boldsymbol{\mu}_i, \mathbf{R}_i) \qquad i = 1, \ldots I. \tag{1}$$

Following Henderson (1973), the vector $\boldsymbol{\mu}_i$ is decomposed according to a linear mixed model structure:

$$\boldsymbol{\mu}_i = \mathbf{X}_i \boldsymbol{\beta} + \mathbf{Z}_i \mathbf{u}_i, \tag{2}$$

with

$$\mathbf{u}_i | \sigma_{u_i}^2 \sim N(\mathbf{0}, \sigma_{u_i}^2 \mathbf{A}_i), \tag{3}$$

where $\mathbf{X}_i$ and $\mathbf{Z}_i$ are $(n_i \times p)$ and $(n_i \times q_i)$ incidence matrices, corresponding to fixed $\boldsymbol{\beta}(p \times 1)$ and random $\mathbf{u}_i(q_i \times 1)$ effects respectively. Fixed effects can be factors or covariates, but it is assumed in the following that, without loss of generality, they represent factors.

In the animal breeding context, $\mathbf{u}_i$ is the vector of genetic merits pertaining to breeding individuals used (sires spread by artificial insemination) or present (males and females) in stratum i. These individuals are related *via* the so-called numerator relationship matrix $\mathbf{A}_i$, which is assumed known and positive definite (of rank q_i).

Elements of $\mathbf{u}_i$ are not usually the same from one stratum to another. A borderline case is the "animal" model (Quaas and Pollak, 1980) where animals with records are completely different from one herd to another. Nevertheless, such individuals are genetically related across herds. Therefore, model [3] has to be refined to take into account covariances among elements of different $\mathbf{u}_i'$s. As proposed by Gianola *et al* (1992), this can be accomplished by relating $\mathbf{u}_i$ to a general $q \times 1$ vector $\mathbf{u}^*$ of standardized genetic merits, *via* the $q_i \times q$ $\mathbf{S}_i$ matrix:

$$\mathbf{u}_i = \sigma_{u_i} \mathbf{S}_i \mathbf{u}^* . \tag{4}$$

Now,

$$\mathbf{u}^* \sim N(\mathbf{0}, \mathbf{A}), \tag{5}$$

with $\mathbf{A}$ being the overall relationship matrix of rank q, relating the q breeding animals involved in the whole population, with $q \leqslant \sum_{i=1}^{I} q_i$.

Thus, $\mathbf{S}_i$ is an incidence matrix with 0 and 1 elements relating the q_i levels of $\mathbf{u}^*$ present in the ith subpopulation to the whole vector $(q \times 1)$ of u elements. For instance, if stratification is made by herd level, the matrices $\mathbf{S}_i$ and $\mathbf{S}_{i'} (i \neq i')$ do not share any non-zero elements in their columns, since animals usually have records only in one herd. On the contrary, in a sire model, a given sire k may have progeny in 2 different herds (i, i') thus resulting in ones in both kth columns of $\mathbf{S}_i$ and $\mathbf{S}'_i$.

Notice that in this model, any genotype $\times$ stratum interaction is due entirely to scaling (Gianola *et al*, 1992).

Formulae [2], [3], [4] and [5] define the model for means; a further step consists in modelling variance components $\{\sigma^2_{e_i}\}_{i=1,\ldots I}$ and $\{\sigma^2_{u_i}\}_{i=1,\ldots I}$ in a similar way, *ie* using a structural model.

The approach taken here comes from the theory of generalized linear models involving the use of a link function so as to express the transformed parameters with a linear predictor (McCullagh and Nelder, 1989). For variances, a common and convenient choice is the log link function (Aitkin, 1987; Box and Meyer, 1986; Leonard, 1975; Nair and Pregibon, 1988):

$$\ln \sigma^2_{e_i} = \mathbf{w}'_{e_i} \boldsymbol{\Upsilon}_e \quad i = 1, \ldots I, \tag{6}$$

$$\ln \sigma^2_{u_i} = \mathbf{w}'_{u_i} \boldsymbol{\Upsilon}_u \quad i = 1, \ldots I, \tag{7}$$

where $\mathbf{w}'_{e_i}$ and $\mathbf{w}'_{u_i}$ are incidence row vectors of size k_e and k_u, respectively, corresponding to dispersion parameters $\boldsymbol{\Upsilon}_e$ and $\boldsymbol{\Upsilon}_u$. These incidence vectors can be a subset of the factors for the mean in [2], but exogeneous information is also allowed. Equations [6] and [7] define the variance component models.

These models can be rewritten in a more compact form as follows. Let

$\mathbf{y} = (\mathbf{y}'_1, \ldots, \mathbf{y}'_i, \ldots, \mathbf{y}'_I)'$ be the $n \times 1$ vector of data for the whole population, with $n = \sum_{i=1}^{I} n_i$,

$\boldsymbol{\mu}_i = \mathbf{X}_i \boldsymbol{\beta} + \sigma_{u_i} \mathbf{Z}_i \mathbf{S}_i \mathbf{u}^*$

$\boldsymbol{\mu} = (\boldsymbol{\mu}'_1, \ldots, \boldsymbol{\mu}'_i, \ldots, \boldsymbol{\mu}'_I)'$ be the mean vector of $\mathbf{y}$,

$\mathbf{R} = \bigoplus_{i=1}^{I} \mathbf{R}_i$ be the variance–covariance matrix of $\mathbf{y}$, with $\oplus$ representing the direct sum (Searle, 1982).

Equation [1] can then be rewritten as:

$$\mathbf{y} \mid \boldsymbol{\mu}, \mathbf{R} \sim N(\boldsymbol{\mu}, \mathbf{R}), \tag{8}$$

with $\mathbf{y}, \boldsymbol{\mu}, \mathbf{R}$ defined as previously.

In the same way, [2] becomes:

$$\boldsymbol{\mu} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}^*\mathbf{u}^* = \mathbf{T}\boldsymbol{\theta}, \qquad [9]$$

with $\mathbf{X} = (\mathbf{X}'_1, \ldots, \mathbf{X}'_i, \ldots, \mathbf{X}'_I)'$,

$\mathbf{X}'_i$ the $(n_i \times p)$ incidence matrix defined in [2];

$\mathbf{Z}^* = (\mathbf{Z}_1^{*'}, \ldots, \mathbf{Z}_i^{*'}, \ldots, \mathbf{Z}_I^{*'})'$,

$\mathbf{Z}_i^* = \sigma_{u_i} \mathbf{Z}_i \mathbf{S}_i$ the $(n_i \times q)$ "incidence" matrix pertaining to $\mathbf{u}^*$,

$\mathbf{T} = (\mathbf{X}, \mathbf{Z}^*)$ and

$\boldsymbol{\theta} = (\boldsymbol{\beta}', \mathbf{u}^{*'})'$.

The vector $\boldsymbol{\theta}$ includes $p + q$ location parameters. The matrix $\mathbf{T}$ can be viewed as an "incidence" matrix, but which depends here on the dispersion parameters $\boldsymbol{\gamma}_u$ through the variances $\sigma_{u_i}^2$.

Both variance models can also be compactly written as:

$$\ln \boldsymbol{\sigma}_e^2 = \mathbf{W}_e \boldsymbol{\gamma}_e \qquad [10]$$

and

$$\ln \boldsymbol{\sigma}_u^2 = \mathbf{W}_u \boldsymbol{\gamma}_u, \qquad [11]$$

with

$$\boldsymbol{\sigma}_e^2 = \left\{ \sigma_{e_i}^2 \right\}_{i=1,\ldots I}, \ \boldsymbol{\sigma}_u^2 = \left\{ \sigma_{u_i}^2 \right\}_{i=1,\ldots I}$$

$$\mathbf{W}_e = (\mathbf{w}'_{e_1}, \ldots, \mathbf{w}'_{e_i}, \ldots, \mathbf{w}'_{e_I})', \ \mathbf{W}_u = (\mathbf{w}'_{u_1}, \ldots, \mathbf{w}'_{u_i}, \ldots, \mathbf{w}'_{u_I})'.$$

The $k_e + k_u$ dispersion parameters $\boldsymbol{\gamma}_e$ and $\boldsymbol{\gamma}_u$ can be concatenated into a vector $(\boldsymbol{\gamma} = (\boldsymbol{\gamma}'_e, \boldsymbol{\gamma}'_u)'$ with corresponding incidence matrix $\mathbf{W} = \mathbf{W}_e \oplus \mathbf{W}_u$. The dispersion model then reduces to:

$$\ln \boldsymbol{\sigma}^2 = \mathbf{W}\boldsymbol{\gamma} \qquad [12]$$

where $\boldsymbol{\sigma}^2 = (\boldsymbol{\sigma}_e^{2'}, \boldsymbol{\sigma}_u^{2'})'$ and $\ln \boldsymbol{\sigma}^2$ is a symbolic notation for $(\ln \sigma_{e_1}^2, \ldots, \ln \sigma_{e_I}^2, \ln \sigma_{u_1}^2, \ldots, \ln \sigma_{u_I}^2)'$.

ESTIMATION OF PARAMETERS

In sampling theory, a way to eliminate nuisance parameters is to use the marginal likelihood (Kalbfleisch, 1986). "Roughly speaking, the suggestion is to break the data in two parts, one part whose distribution depends only on the parameter of interest, and another part whose distribution may well depend on the parameter of interest but which will, in addition, depend on the nuisance parameter. [...] This second part will, in general, contain information about the parameter of interest, but in such a way that this information is inextricably mixed up with the nuisance parameter" (Barnard, 1970). Patterson and Thompson (1971) used this approach for estimating variance components in mixed linear Gaussian models. Their derivations were based on error contrasts. The corresponding estimator (the so-called REML) takes into account the loss in degrees of freedom due to the estimation of location parameters.

Alternatively, Harville (1974) proved that REML can be obtained using the non-informative Bayesian paradigm. According to the definition of marginalization in Bayesian inference (Box and Tiao, 1973; Robert, 1992), nuisance parameters are eliminated by integrating them out of the joint posterior density.

Keeping in mind that the sampling and the non-informative Bayesian approaches give rise to the same estimation equations, we have chosen the Bayesian techniques for reasons of coherence and simplicity.

The parameters of interest are here the dispersion parameters γ, and the location parameters θ appear to be nuisance parameters. Inference is hence based on the log marginal likelihood $L(\gamma; y)$ of γ:

$$L(\gamma; y) = \ln p(y \mid \gamma) = \ln \int p(y, \theta \mid \gamma) d\theta. \qquad [13]$$

An estimator $\widehat{\gamma}$ of γ is given by the mode of $L(\gamma; y)$:

$$\widehat{\gamma} = \text{Arg max}_{\gamma \in \Gamma} \, L(\gamma; y), \qquad [14]$$

where Γ is a compact part of $R^{k_e + k_u}$.

This maximization can be performed using a result by Foulley *et al* (1990, 1992) which avoids the integration in [13]. Details can be found in the *Appendix*. This procedure results in an iterative algorithm. Numerically, let $[t]$ denote the iteration t; the current estimate $\widehat{\gamma}^{[t+1]}$ of γ is computed from the following system:

$$\mathbf{W}' \widehat{\mathbf{Q}}^{[t]} \mathbf{W} \left(\widehat{\gamma}^{[t+1]} - \widehat{\gamma}^{[t]} \right) = \mathbf{W}' \widehat{\mathbf{z}}^{[t]}, \qquad [15]$$

where
  $\widehat{\gamma}^{[t]}$ is the current estimate at iteration t,
  $\mathbf{W}$ the incidence matrix defined in [12],

$\widehat{\mathbf{Q}}^{[t]}$ is the weight matrix depending on $\widehat{\theta}^{[t]}$ and on $\widehat{\mathbf{C}}^{[t]}$, which are the solution and the inverse coefficient matrix respectively of the current system in θ (this system is described next),

  $\widehat{\mathbf{z}}^{[t]}$ is the score vector depending on $\widehat{\theta}^{[t]}$ and $\widehat{\mathbf{C}}^{[t]}$.
  Elements of $\widehat{\mathbf{Q}}^{[t]}$ and $\widehat{\mathbf{z}}^{[t]}$ are given in the *Appendix*.
  The second system is:

$$\left(\widehat{\mathbf{T}}^{[t]'} \widehat{\mathbf{R}}^{-1[t]} \widehat{\mathbf{T}}^{[t]} + \Sigma^- \right) \widehat{\theta}^{[t]} = \widehat{\mathbf{T}}^{[t]'} \widehat{\mathbf{R}}^{-1[t]} y, \qquad [16]$$

where $\widehat{\mathbf{T}}^{[t]}$ is the "incidence" matrix $\mathbf{T}$ defined in [9] and evaluated at $\gamma = \widehat{\gamma}^{[t]}$;
  $\widehat{\mathbf{R}}^{-1[t]}$ is the weight matrix evaluated at $\gamma = \widehat{\gamma}^{[t]}$, with $\mathbf{R}$ defined as in [8];
  $\Sigma^- = \begin{pmatrix} 0 & 0 \\ 0 & \mathbf{A}^{-1} \end{pmatrix}$ and takes into account the prior distribution of $\mathbf{u}^*$ in [5].

The system [16] is an iterative modified version of the mixed model equations of Henderson (1984). It provides as a by product an empirical Bayes estimates $\widehat{\theta}$ of the vector θ of location parameters.

Regarding computations involved in [15], 2 types of algorithms can be considered as in San Cristobal (1992). A second order algorithm (Newton–Raphson type) converges rapidly and gives estimates of standard errors of $\widehat{\boldsymbol{\gamma}}$, but computing time can be excessive with the large data sets typical of animal breeding problems. As shown in Foulley et al (1990), a first order algorithm can be easily obtained by approximating the $\mathbf{Q}$ matrix in [15] by its expectation component ($\mathbf{Q}_{\alpha\beta,E}$ in the appendix notations). This EM (Expectation–Maximization; Dempster et al, 1977) algorithm converges more slowly, but needs fewer calculations at each iteration and, on the whole, less total CPU time for large data sets.

HYPOTHESIS TESTING

An adequate modelling of heteroskedasticity in variance components requires a procedure for hypothesis testing. Let $H_0 : \mathbf{H}\boldsymbol{\gamma} = \mathbf{0}$ be the null hypothesis with $\mathbf{H}$ being a full (row) rank matrix with row size equal to the number of linearly independent estimable functions of $\boldsymbol{\gamma}$ defining H_0, and H_1 its alternative. For example, one can be interested in testing the hypothesis of homogeneity of residual variances $H_0 : \sigma_{e_i}^2 = \exp(\gamma_{e_i}) = \text{Const}$ for all i. Letting $\boldsymbol{\gamma}_e = \left\{ \gamma_R, \gamma_{e_2} - \gamma_R, \ldots, \gamma_{e_I} - \gamma_R \right\}'$ with γ_R being the dispersion parameter for the residual variance in the first stratum taken as reference. H_0 can be expressed as $\mathbf{H}_e\boldsymbol{\gamma}_e = \mathbf{0}$, or $(\mathbf{H}_e, \mathbf{0})\boldsymbol{\gamma} = \mathbf{0}$ with $\mathbf{H}_e = (\mathbf{0}_{(I-1)\times 1}, \mathbf{I}_{I-1})$.

Let M_0 and M_1 be the models corresponding to H_0 and H_1, respectively. Since $p(\mathbf{y}|\boldsymbol{\gamma}) = p(\mathbf{y}|\boldsymbol{\sigma}_e^2, \boldsymbol{\sigma}_u^2)$, the marginal likelihood can be interpreted as a likelihood of error contrasts (Harville, 1974), hence the likelihood ratio test based on the marginal likelihood can be applied:

$$\lambda = -2 \left[\mathbf{L}(\widehat{\boldsymbol{\gamma}}_{M_0}; \mathbf{y}) - \mathbf{L}(\widehat{\boldsymbol{\gamma}}_{M_1}; \mathbf{y}) \right]. \qquad [17]$$

Under H_0, λ is asymptotically distributed according to a χ^2 with degrees of freedom equal to the rank of $\mathbf{H}$. In the normal case, explicit calculation of $\mathbf{L}(\boldsymbol{\gamma}; \mathbf{y})$ is analytically feasible:

$$\mathbf{L}(\boldsymbol{\gamma}; \mathbf{y}) = (-1/2)\Big\{ (n-p)\ln 2\pi + \ln|\mathbf{A}| + \ln\left|\mathbf{R}(\boldsymbol{\gamma}_e)\right|$$
$$+ \ln\left|\mathbf{T}'(\boldsymbol{\gamma}_u)\mathbf{R}^{-1}(\boldsymbol{\gamma}_e)\mathbf{T}(\boldsymbol{\gamma}_e) + \boldsymbol{\Sigma}^-\right| + \mathbf{y}'\mathbf{R}^{-1}(\boldsymbol{\gamma}_e)\Big[\mathbf{y} - \mathbf{T}(\boldsymbol{\gamma}_u)\boldsymbol{\theta}(\boldsymbol{\gamma})\Big]\Big\} \qquad [18]$$

A BAYESIAN APPROACH TO A MIXED MODEL STRUCTURE

One can be interested to generalise Henderson's BLUP for subclass means ($\boldsymbol{\mu} = \mathbf{T}\boldsymbol{\theta}$) to dispersion parameters ($\ln\boldsymbol{\sigma}^2 = \mathbf{W}\boldsymbol{\gamma}$) ie proceed as if $\boldsymbol{\gamma}$ had a mixed model structure (Garrick and Van Vleck, 1987). To overcome the difficulty of a realistic interpretation of fixed and random effects for conceptual populations of variances from a frequentist (sampling) perspective, one can alternatively use Bayesian procedures. It is then necessary to place suitable prior distributions on dispersion parameters and follow an informative Bayesian approach.

In linear Gaussian methodology, theoretical considerations regarding conjugate priors or fiducial arguments lead to the use of the inverted gamma distribution as a prior for a variance σ^2 (Cox and Hinkley, 1974; Robert, 1992). Such a density depends on hyperparameters η and s^2. The former conveys the so-called degrees of belief, and the latter is a location parameter. The ideas briefly exposed in the following are similar to those described in Foulley *et al* (1992).

Hence, a prior density for $\gamma = \ln \sigma^2$ can be obtained as a log inverted gamma density. As a matter of fact, it is more interesting to consider the prior distribution of $v = \gamma - \gamma^0$, with $\gamma^0 = \ln s^2$, *ie*

$$p(v|\eta) = (\eta/2)^{\eta/2} \exp\left\{-(\eta/2)\left[v + \exp\left(-v\right)\right]\right\}/\Gamma(\eta/2) \qquad [19]$$

where $\Gamma(.)$ refers to the gamma function.

Let us consider a K-dimensional "random" factor $\mathbf{v}$ such that $v_k|\eta_k (k = 1, \dots K)$ is distributed as a log inverted gamma $\ln G^{-1}(\eta_k)$. Since the levels of each random factor are usually exchangeable, it is assumed that $\eta_k = \eta$ for every k in $\{1, \dots K\}$:

$$v_k|\eta \sim \ln G^{-1}(\eta) \qquad k = 1, \dots K. \qquad [20]$$

For v_k in [20] small enough, the kernel of the product of independent distributions having densities as in [19] can be approximated (using a Taylor expansion of [19] about v equal to 0) by a Gaussian kernel, leading to the following prior for $\mathbf{v}$:

$$\mathbf{v}|\xi \sim N(\mathbf{0}, \xi \mathbf{I}_K) \qquad [21]$$

with $\xi = 2/\eta$.

As explained by Foulley *et al* (1992), this parametrization allows expression of the $\boldsymbol{\gamma}$ vector of dispersion parameters under a mixed model type form. Briefly, from [19] one has $\gamma = \gamma^0 + v$ or $\gamma = \mathbf{p}'\boldsymbol{\delta} + v$ if one writes the location parameter $\gamma^0 = \ln s^2$ as a linear function of some vector $\boldsymbol{\delta}$ of explanatory variables ($\mathbf{p}'$ being a row incidence vector of coefficients). Extending this writing to several classifications in $\mathbf{v}$ leads to the following general expression:

$$\ln \boldsymbol{\sigma}^2 = \mathbf{P}\boldsymbol{\delta} + \mathbf{Q}\mathbf{v}, \qquad [22]$$

where $\mathbf{P}$ and $\mathbf{Q}$ are incidence matrices corresponding to fixed effects $\boldsymbol{\delta}$ and random effects $\mathbf{v}$, respectively, with [20] or [21] as prior distribution for $\mathbf{v}$.

Regarding dispersion parameters $\boldsymbol{\gamma}$, it is then possible to proceed as Henderson (1973) did for location parameters $\boldsymbol{\mu}$, *ie* describe them with a mixed model structure. Again, as illustrated by formula [22], the statistical treatment of this model can be conveniently implemented *via* the Bayesian paradigm.

In fact, equations [22] define a model on residual variances:

$$\ln \boldsymbol{\sigma}_e^2 = \mathbf{P}_e\boldsymbol{\delta}_e + \mathbf{Q}_e\mathbf{v}_e, \qquad [23]$$

and a model on genetic variances as well:

$$\ln \boldsymbol{\sigma}_u^2 = \mathbf{P}_u\boldsymbol{\delta}_u + \mathbf{Q}_u\mathbf{v}_u, \qquad [24]$$

where $\mathbf{P}_e, \mathbf{P}_u, \mathbf{Q}_e, \mathbf{Q}_u$ are incidence matrices corresponding respectively to fixed effects $\boldsymbol{\delta}_e, \boldsymbol{\delta}_u$ and random effects $\mathbf{v}_e = (\mathbf{v}'_{e_1}, \mathbf{v}'_{e_2}, \ldots, \mathbf{v}'_{e_j}, \ldots)', \mathbf{v}_u = (\mathbf{v}'_{u_1}, \mathbf{v}'_{u_2}, \ldots, \mathbf{v}'_{u_k}, \ldots)'$ with, for the jth and kth random classification in $\mathbf{v}_e$ and $\mathbf{v}_u$ respectively,

$$\mathbf{v}_{e_j}|\xi_{e_j} \sim N(\mathbf{0}, \xi_{e_j}\mathbf{I}_{K_{e_j}}) \tag{25}$$

and

$$\mathbf{v}_{u_k}|\xi_{u_k} \sim N(\mathbf{0}, \xi_{u_k}\mathbf{I}_{K_{u_k}}) \tag{26}$$

Let $\boldsymbol{\eta} = (\boldsymbol{\eta}'_e, \boldsymbol{\eta}'_u)'$ with $\boldsymbol{\eta}_e = \{\eta_{e_j}\}$ and $\boldsymbol{\eta}_u = \{\eta_{u_k}\}$ be the vectors of hyperparameters introduced in the variance component models [23], [24], [25] and [26]. An empirical Bayes procedure is chosen to estimate the parameters. The hyperparameters, $\boldsymbol{\eta}$ (or $\boldsymbol{\xi} = (\boldsymbol{\xi}'_e, \boldsymbol{\xi}'_u)'$) are estimated by the mode of the marginal likelihood of these hyperparameters (Berger, 1985; Robert, 1992):

$$\widehat{\boldsymbol{\eta}} = \text{Arg max } \ln p(\mathbf{y}|\boldsymbol{\eta}) \tag{27}$$

or

$$\widehat{\boldsymbol{\xi}} = \text{Arg max } \ln p(\mathbf{y}|\boldsymbol{\xi}). \tag{28}$$

Then, the dispersion parameters are obtained by the mode of the posterior density of $\boldsymbol{\gamma}$ given the hyperparameters equal to their estimates:

$$\widehat{\boldsymbol{\gamma}}(\widehat{\boldsymbol{\eta}}) = \text{Arg max } \ln p(\boldsymbol{\gamma}|\mathbf{y}, \boldsymbol{\eta} = \widehat{\boldsymbol{\eta}}), \tag{29}$$

or similarly for $\widehat{\boldsymbol{\xi}}$.

Maximization in [27] and [28] can be performed with a Newton–Raphson or an EM algorithm, following ideas in the *Estimation of parameters*, Unfortunately, the algorithm derived from [27] is computationally demanding, since it involves digamma and trigamma functions. On the other hand, an EM algorithm derived from [28] has the same form as the EM–REML algorithm for variance components. It just involves the solution and the inverse coefficient matrix of the system in $\boldsymbol{\gamma}$ at iteration $[t]$. This latter system is similar to [15], but it takes into account the informative prior on the dispersion parameters. In the case of a Gaussian prior, this system can be written as

$$\left(\mathbf{W}'\widehat{\mathbf{Q}}^{[t]}\mathbf{W} + \widehat{\boldsymbol{\Gamma}}^{-[t]}\right)\left(\widehat{\boldsymbol{\gamma}}^{[t+1]} - \widehat{\boldsymbol{\gamma}}^{[t]}\right) = \mathbf{W}'\widehat{\mathbf{z}}^{[t]} - \widehat{\boldsymbol{\Gamma}}^{-[t]}\widehat{\boldsymbol{\gamma}}^{[t]}, \tag{30}$$

where $\widehat{\boldsymbol{\Gamma}}^{-[t]}$ is the matrix $\boldsymbol{\Gamma}^-(\boldsymbol{\xi}) = \begin{pmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Lambda}^{-1}(\boldsymbol{\xi}) \end{pmatrix}$ evaluated at the current estimate $\widehat{\boldsymbol{\xi}}^{-[t]}$ of $\boldsymbol{\xi}$, tanking into account the priors *via* $\boldsymbol{\Lambda}(\boldsymbol{\xi}) = \text{Var}(\mathbf{v}'_e, \mathbf{v}'_u) = \boldsymbol{\Lambda}_e \oplus \boldsymbol{\Lambda}_u$ with $\boldsymbol{\Lambda}_e = \bigoplus_j \xi_{e_j}\mathbf{I}_{K_{e_j}}$ and $\boldsymbol{\Lambda}_u = \bigoplus_k \xi_{u_k}\mathbf{I}_{K_{u_k}}$.

Details for the environmental variance part of this development can be found in Foulley *et al* (1992). The extension to the u-part is straightforward.

NUMERICAL APPLICATION

Sires of French beef breeds are routinely evaluated for muscular development (MD)
based on phenotypic performance of their male and female progeny. Qualified
personnel subjectively classify the calves at about 8 months of age, with MD
scores ranging from 0 to 100. Variance components and sire genetic values are
then estimated by applying classical procedures, *ie* REML and BLUP (Henderson,
1973; Thompson, 1979), to a mixed model including the random sire effect and a
set of fixed effects described in table I. The second factor listed in table I, condition
score ("Condsc"), accounts for the previous environmental conditions (*eg* nutrition
via fatness) in which calves have been raised.

Table I. Factors of the Maine–Anjou data.

Factors	Levels	Meanings
Classi	12	Classifiers, with more than 20 scores
Condsc	4	Condition score, from very bad to very good
Sex	2	Sex of calf
Age	7	Age groups, from less than 50 days to more than 300 days
Region	5	Region of birth: Pays de la Loire, Poitou-Charente, Bretagne, Nord, Centre-et-Sud-Ouest
Parity	10	Parity of dam; parities greater than 10 were grouped together
Year	7	Year of birth, from 1983 to 1989
Month	12	Month of birth
Sire	142	Sire of calf: sires with more than 30 calves present in the data set, plus grand-sires (with or without data) for including the relationship matrix **A**

Some factors among those described in table I may induce heterogeneous
variances. In particular, different classifiers are expected to generate not only
different MD means, but different MD variances as well. Thus, the usual sire model
with assumption of homogeneous variances may be inadequate. This hypothesis was
tested on the Maine-Anjou breed. After elimination of twins and further editing
described in table I, the Maine-Anjou file included performance records on 8 575
progeny out of 142 sires ("Sire") recorded in 5 regions ("Region") and 7 years
("Year"). Other factors taken into account were: sex of calves ("Sex"), age at
scoring ("Age"), claving parity ("Parity"), month of birth ("Month") and classifier
("Classi"). In most strata defined as combinations of levels of the previous factors,
only one observation was present.

Preliminary analysis

A histogram of the MD variable can be found in figure 1. The distribution of MD
seems close to normality, with a fair PP-plot (although the use of this procedure is
somewhat controversial), and skewness and kurtosis coefficients were estimated as
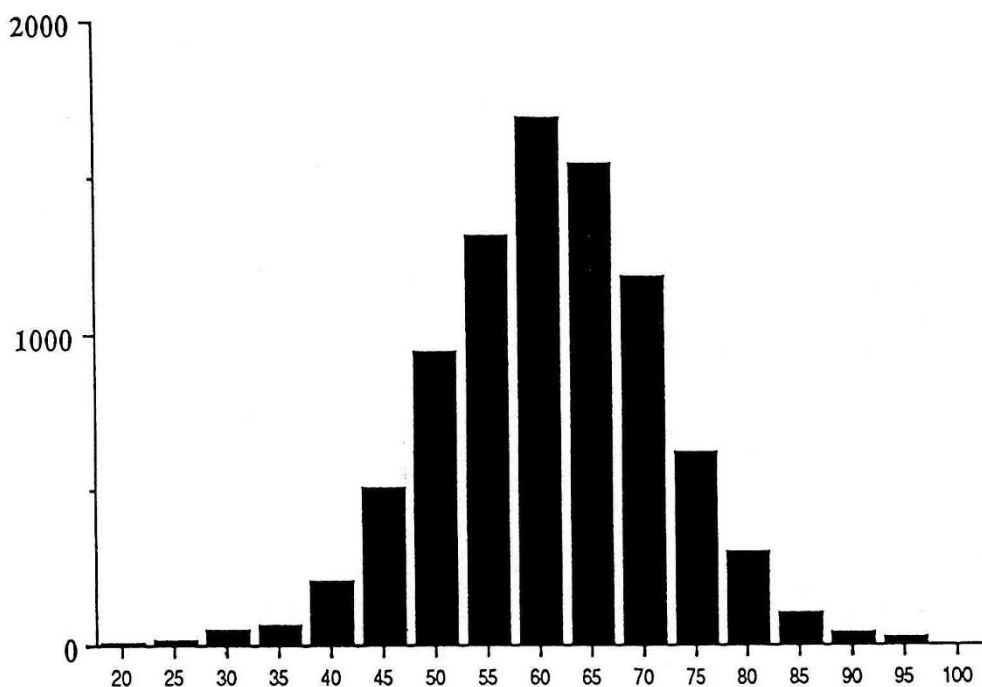−0.09 and 0.37 respectively. Some commonly used tests for normality rejected the

Fig 1. Histogram of the MD variable.

null hypothesis, while others did not reject it, namely Geary's u, Pearson's tests for skewness and kurtosis (Morice, 1972) at the 1% level.

Bartlett's test for homogeneity of variances was computed for each of the first 8 factors described in table I. Results in table IIa indicate strong evidence for heteroskedastic variances among subclasses of each factor considered in this data set.

The usual sire model with all factors from table I in the mean model, and variance components estimated by EM–REML, was fitted, leading to estimates $\widehat{\sigma}_e^2 = 70.11, \widehat{\sigma}_u^2 = 6.91$, and $\widehat{h}^2 = 4\widehat{\sigma}_u^2/(\widehat{\sigma}_e^2 + \widehat{\sigma}_u^2) = 0.36$. Note that this model is equivalent, in our notation, to the homogeneous model in Υ_e and Υ_u.

Search for a model for the variances

The following additive mean model M_θ was considered as true throughout the whole analysis

$$M_\theta : \text{Classi} + \text{Condsc} + \text{Sex} + \text{Age} + \text{Region}$$
$$+\text{Parity} + \text{Year} + \text{Month} + \text{Sire (random)} \qquad [31]$$

This model was chosen in agreement with technicians of the Maine–Anjou breed and is used routinely for genetic evaluation of Maine–Anjou sires.

Table IIa. Bartlett's test on observed MD values.

γ_e-*Models*	*Bartlett's test*	*P-Value*	$s_{i,min}^2 \cdot s_{i,max}^2$
Classi	426.56	0	0.56
Year	232.00	0	0.67
Condsc	2 275.69	0	0.63
Sex	219.40	0	0.88
Age	75.49	0	0.93
Month	52.98	2.10^{-7}	0.78
Region	20.51	4.10^{-4}	0.61
Parity	83.47	0	0.64

IIb. Search for a 1-factor model on residual variances.

γ_e-*Model*	λ^a	*df*	*P-Value*	$\widehat{\sigma}_{i,min}^2/\widehat{\sigma}_{i,max}^2$	$\widehat{\sigma}_u^2$
Homoskedastic	0			1	6.91
Classi	93.95	11	0	0.46	7.13
Year	79.69	6	0	0.71	7.03
Condsc	49.85	3	0	0.60	6.93
Sex	45.25	1	0.015	0.85	6.74
Age	22.12	6	10^{-3}	0.80	6.92
Month	37.36	11	10^{-4}	0.77	6.93
Region	19.14	4	7.10^{-4}	0.64	6.89
Parity	21.38	9	0.01	0.74	6.93

[a] Marginal likelihood ratio test.

A forward selection of factors strategy was chosen to find a good variance model M_γ but in 2 stages; a backward selection strategy would have been difficult to implement because of the large number of models to compare and the small amount of information in some strata generated by those models.

(i) since σ_e^2 represents $> 90\%$ of the total variation, it was decided to model that component first, assuming the γ_u- part homogenous;

(ii) the "best" γ_u-model was thereafter chosen while keeping unchanged the "best" γ_e-model.

The different nested models were fitted using the maximum marginal likelihood ratio test (MLRT) λ described in [17].

During the first stage (i), the homogeneous sire variance was estimated, for computational ease, with an EM–REML algorithm, and the γ_e parameter estimates were calculated as in Foulley *et al* (1992). This strategy leads, of course, to the same results as those obtained with the algorithm described in the *Estimation of parameters*.

The first step consisted of choosing the best one-factor variance model from results presented in table IIb. The next steps, *ie* the choice of an adequate 2-factor model, and then of a 3-factor model, etc, are summarised in table III. Finally, the following additive model was chosen:

$$M_{\gamma_e} : \text{Classi} + \text{Condsc} + \text{Year} + \text{Month} \qquad [32]$$

The model can also be simplified after comparing estimates of factor levels, and then collapsing these levels if there are not significantly different.

For the (ii) stage, the "best" γ_u-model was the model (see table IV):

$$M_{\gamma_u} : \text{Sex} \qquad [33]$$

Table III. Main steps of the search for an adequate residual variance model.

γ_e-Models compared	λ^a	df	P-Value
Classi + Condsc vs Classi	36.03	3	7.10^{-8}
Classi + Condsc + Year vs Classi + Condsc	25.73	6	2.10^{-4}
Classi + Condsc + Year + Month vs Classi + Condsc + Year	24.67	11	10^{-2}

[a] Marginal likelihood ratio test.

Table IV. Main steps of the search for an adequate genetic variance model.

γ_u-Models compared	λ^a	df	P-Value
Sex vs homoskedastic	12.62	1	4.10^{-4}

[a] Marginal likelihood ratio test.

We were not able to reach convergence of the iterative procedure for the models $(M_\theta, M_{\gamma_e}, \text{Classi})$ and $(M_\theta, M_{\gamma_e}, \text{Region})$, although some levels of the Classi factor were collapsed. This phenomenon is related to a strong unbalance of the design: for instance, one classifier noted the calves of only 4 sires, making quite impossible a coherent estimation of Classi-heterogeneous sire variances. The other factors (except Year) had no significant effect on the variation of the sire variances. Because of imbalance, the model

$$M'_{\gamma_u} : \text{Sex} + \text{Year} \qquad [34]$$

gave unsatisfactory results eg heritability estimates greater than one.

Results

Estimates of the dispersion parameters for the selected model designated here as $(M_\theta, M_{\gamma_e}, M_{\gamma_e})$ are shown in table Va. As expected, the γ_e-estimates of the $(M_\theta, M_{\gamma_e}, \text{homogeneity})$ model, ie of the best γ_e-model with only one genetic variance, are quite similar to the γ_e-estimates of the $(M_\theta, M_{\gamma_e}, M_{\gamma_u})$ model (table Va). In contrast, γ_e-estimates of the $(M_\theta, M_{\gamma_e}, \text{homogeneity})$ model, with:

$$M'_{\gamma_e} : \text{Classi (random)} + \text{Condsc} + \text{Year (random)} + \text{Month (random)} \qquad [35]$$

are different for the "random" factors (see table Vb). Estimated hyperparameters for variances of the Classi, Year and Month factors, are $\hat{\xi}_{e,\text{Classi}} = 0.021$,

$\widehat{\xi}_{e,\text{Year}} = 0.009$ and $\widehat{\xi}_{e,\text{Month}} = 0.0024$ respectively, or alternatively using % values of the coefficient of variation for σ_e^2, $(\xi_e = CV_e^2)$ $CV_{e,\text{Classi}} = 14.5\%$, $CV_{e,\text{Year}} = 9.5\%$ and $CV_{e,\text{Month}} = 4.9\%$ respectively. In fact, the smaller the cell size (n_i), and the smaller CV, the greater the shrinkage of the sample estimates $(\widehat{\sigma}_i^2)$ toward the mean variance $(\overline{\sigma}^2)$ since the regression coefficient toward this mean in the equation $\widetilde{\sigma}_i^2 = \overline{\sigma}^2 + b(\widehat{\sigma}_i^2 - \overline{\sigma}^2)$ is approximately $b = n_i/[n_i + (2/CV^2)]$ with $\eta = 2/CV^2$: see also Visscher and Hill (1992).

The genetic variation in heifers turns out to be less than one half what it is in bulls even though the phenotypic variance was virtually the same. This may be due to the fact that classifiers do not score exactly the same trait in males (muscling) as in females (size and/or fatness). It may also suggest that the regime of male calves is supplemented with concentrate.

Location parameters are compared in figures 2a–d under different dispersion models, through scatter plots of estimates of standardized sire merits $(\widehat{\mathbf{u}}^*)$. Indexes based on "subclass means" $(\widehat{\mu}_i = \overline{y}_i, i = 1, \dots I$, with homogeneous variances) and those based on the "sire model" under the homogeneity of variance assumption are far away from each other (see fig 2a). Figure 2a is just a reference of discrepancy, which illustrates the impact of the BLUP methodology. When heterogeneity is introduced among residual variances, sires' genetic values do not vary too much, as shown in figure 2b. Modelling of the genetic variances has a larger impact on the sire genetic values (see figure 2c) than modelling of residual variances. Finally, the Bayesian treatment of $\mathbf{\gamma}_e$-parameters by introducing random effects in the model (M_θ, M_{γ_e}) does not have any influence on the sire genetic merits (fig 3d).

Evaluation of sires can be biased if true heterogeneity of variance is not taken into account. As shown in table VI, sire number 13 went down from the 16th to the 24th position because his calves were scored mostly by classifier no 1 who uses a large scale of notation (see $\mathbf{\gamma}$-estimates in table V). On the other hand, sire 103 went up from the 25th to the 14th place since the corresponding Classi and Condsc levels have low residual variance (for the other factor levels represented, the variances were at the average). For the same reason, the sire genetic merits were also affected by modelling $\ln \mathbf{\sigma}_e^2$. The difference in genetic merit for sire 56 (1.40 *vs* 1.74 under the homoskedastic and the residual heteroskedastic models respectively) is also explained by the fact that the calves of this sire were scored exclusively by classifier no 12 and in 1983 (Year = 1). Due to modelling σ_u^2, this sire went down again (from 1.74 to 1.63 under the full heteroskedastic model) because all its progeny are females with a lower σ_u^2 component than in males. Other things being equal, a reduction in the σ_u^2 variance results in a larger ratio, or equivalently a smaller heritability and consequently in a higher shrinkage of the estimated breeding value toward the mean. In other words, if a decrease in genetic variance is ignored, sires above the mean are overevaluated and sires below the mean are underevaluated.

Hypothesis checking

Normality assumptions made in [1] and [5] were checked at each step of the analysis.

Table Va. Final estimates of dispersion parameters for the $(M_\theta, M_{\gamma_e}, M_{\gamma_u})$ model.

Parameter		$\widehat{\gamma}$	$\widehat{\sigma}^2 = exp\,\widehat{\gamma}$
e-part			
Baseline	1	4.75	115.86
Classi	1.5[a]	0	1
	2	−0.07	0.93
	3.10[a]	−0.11	0.89
	4	−0.24	0.79
	6	−0.22	0.80
	7	−0.51	0.60
	8	0.03	1.03
	9	−0.21	0.81
	11	−0.58	0.56
	12	−0.80	0.44
Condsc	1.4[a]	0	1
	2	−0.29	0.75
	3	−0.42	0.66
Year	1	0	1
	2	0.08	1.08
	3	−0.14	0.87
	4.5[a]	−0.21	0.81
	6	−0.09	0.91
	7	−0.01	0.99
Month	1	0	1
	2.4[a]	0.15	1.16
	3	0.12	1.13
	5	0.17	1.19
	6	0.19	1.21
	7	0.05	1.05
	8	0.03	1.03
	9	0.16	1.18
	10	0.01	1.01
	11	−0.01	0.99
	12	0.17	1.18
u-Part			
Baseline	1	2.32	10.22
Sex	1	0	1
	2	−0.76	0.47

[a] Levels with the same estimates on the log scale (to 10^{-3}).
Example: for a cell with factors Classi = 5, Condsc = 2, Year = 2, Month = 12, Sex = 2, the estimated residual variance will be $\widehat{\sigma}_e^2 = exp\,(4.75 + 0 − 0.29 + 0.08 + 0.17) = 115.86 \times 1 \times 0.75 \times 1.08 \times 1.18$ and the estimated genetic variance will be $\widehat{\sigma}_u^2 = 10.22 \times 0.47$.

Table Vb. Estimates of dispersion parameters under the residual heteroskedastic model with some factors treated as random.

Parameter		$\widehat{\gamma}$	$\widehat{\sigma}^2 = exp\,\widehat{\gamma}$
Baseline	1	4.58	97.57
Classi	1	0.16	1.17
	2	0.07	1.07
	3	0.06	1.06
	4	−0.05	0.95
	5	0.18	1.20
	6	−0.03	0.97
	7	−0.09	0.92
	8	0.05	1.05
	9	−0.02	0.98
	10	0.06	1.06
	11	−0.17	0.84
	12	−0.21	0.81
Condsc	1	0	1
	2	−0.30	0.74
	3	−0.43	0.65
	4	0.02	1.02
Year	1	−0.02	0.98
	2	0.13	1.14
	3	−0.02	0.98
	4	−0.09	0.91
	5	−0.09	0.91
	6	0.01	1.01
	7	0.08	1.08
Month	1	−0.06	0.94
	2	0.02	1.02
	3	0.00	1.00
	4	0.02	1.02
	5	0.03	1.03
	6	0.03	1.03
	7	−0.01	0.99
	8	−0.02	0.98
	9	0.03	1.03
	10	−0.04	0.96
	11	−0.04	0.96
	12	0.03	1.03

The estimated sire variance was 7.08. Variances of the Classi, Year and Month factors are, respectively, $\widehat{\xi}_{e,\text{Classi}} = 0.021$ and $\widehat{\xi}_{e,\text{Year}} = 0.009$ and $\widehat{\xi}_{e,\text{Month}} = 0.0024$.

(a)

(b)

(c)

(d)
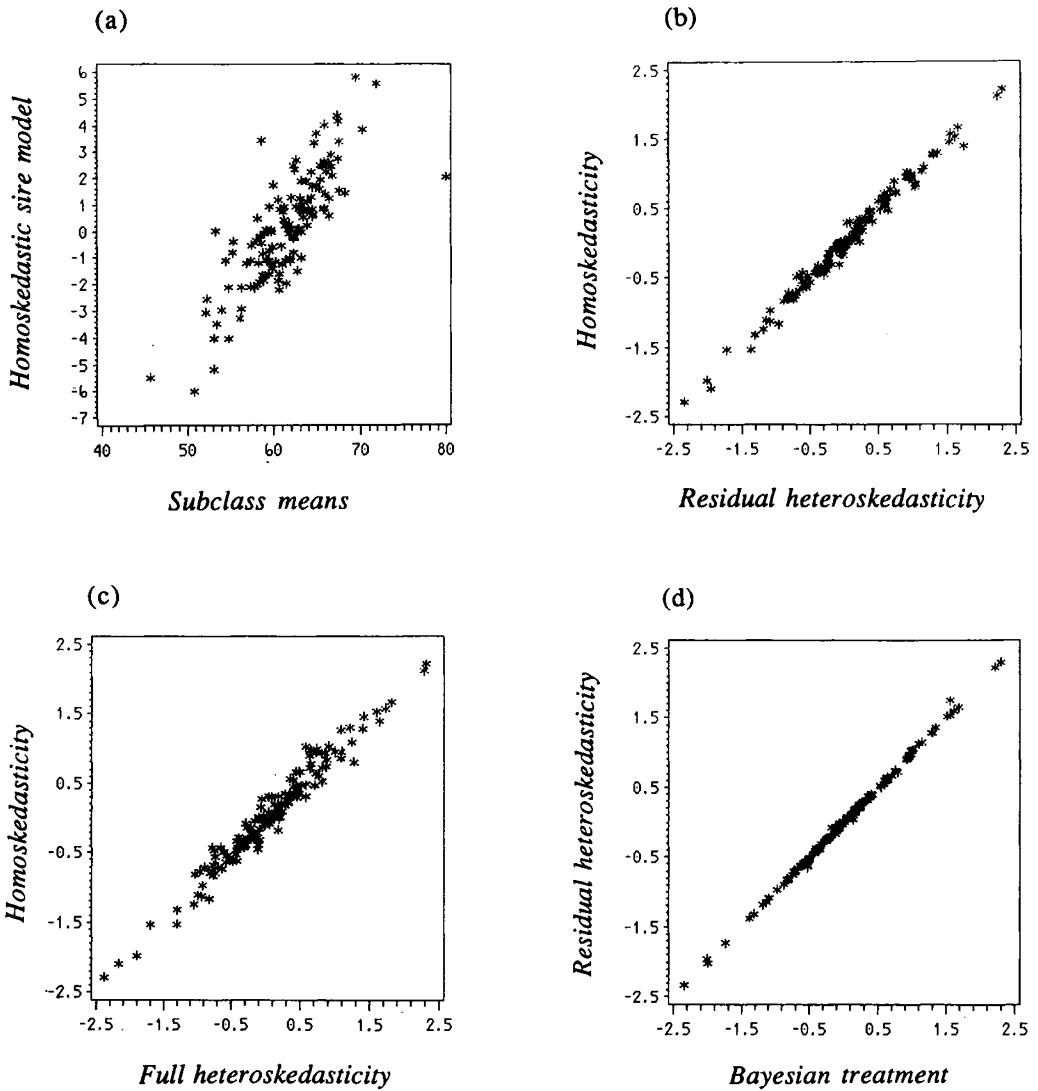
Fig 2. Comparison of sire genetic values under different variance models. a. Homoskedastic "sire model" *vs* evaluation with subclass means per sire. b. Homogeneous variance model *vs* heterogeneous residual variance model. c. Homoskedastic variance model *vs* heterogeneous variance component model. d. Residual heteroskedasticity: marginal likelihood approach *vs* Bayesian treatment.

Table VI. Some estimates of standardized genetic values and ranks of sires under several variance models.

Sire No	Homoskedasticity		Residual heteroskedasticity		Full heteroskedasticity	
	$\widehat{u}^*$	Rank	$\widehat{u}^*$	Rank	$\widehat{u}^*$	Rank
13	0.96	16	0.90	24	1.08	12
103	0.80	25	1.02	14	1.27	9
56	1.40	7	1.74	3	1.63	5

After modelling residual variances, the distribution of standardized residuals became closer to normality, in terms of skewness and especially kurtosis. This phenomenon was observed in the whole sample and also in the subsamples defined by the levels of the factor considered in γ_e. On the other hand, normality of the residuals was stable in the subsamples defined by the factors absent from the γ_e-model.

Normality of the distribution of the standardized sire values in terms of kurtosis and PP-plot was continuously damaged at each step of the variance modelling: estimated kurtosis was 0.61, 0.72 and 0.90, for the homoskedastic, residual heteroskedastic and fully heteroskedastic models respectively. Moreover, skewness for the 142 sire genetic merits improved slightly during that process: -0.09, -0.003 and -0.03 for the same models respectively.

Computational aspects

Programmes were written in Fortran 77 on an IBM 3090 by implementing an EM algorithm corresponding to [15]. The convergence was fast: 15–20 cycles for heteroskedastic γ_e-models with σ_u^2 estimated by EM–REML ((i) stage), and 15–40 cycles for fully heteroskedastic γ-models or heteroskedastic γ_e-models with random effects. CPU time was between 2–5 min per model fit (estimation of parameters and computation of the log marginal likelihood.

DISCUSSION AND CONCLUSION

This paper is an extension to u-components of variances of the approach developed by Foulley *et al* (1992) to consider heterogeneity in residual variances using a structural model to describe dispersion parameters, in a similar way as usually done on subclass means.

In that respect, our main concern focuses on ways to render models as parsimonious as possible so as to reduce the number of parameters needed to assess heteroskedasticity of variances. An interesting feature of this procedure is to assess, through a kind of analysis of variance, the effects of factors marginally or jointly. For instance, one can test heterogeneity of sire variances among breeds of dams after adjusting for possible sources of variation such as management level. In the same way, differences among group of sires in within-sire variances (which might

be related to a segregating major gene) can be tested while taking into account the influence of other nuisance factors (season, nutrition...). However, the power of the likelihood ratio test for detecting heterogeneity of variance can be a real issue in many practical instances.

From the genetic point of view, the approach is quite general since it can deal with heterogeneity among within and between family components of variances, or among genetic and environmental variances. Factors involved for u and e components of variance may be different or the same, making the method especially flexible. Our modelling allows one to assume (or even test) whether the ratios of variances or heritabilities are constant over levels of some single factor or combination of factors (Visscher and Hill, 1992). If a constant heritability or ratio of variances $\alpha = \sigma_{e_i}^2/\sigma_{u_i}^2$ among strata is assumed, the model involves the parameters $\boldsymbol{\gamma}_e$ and α only, and reduces to $\ln \sigma_{e_i}^2 = \mathbf{w}_{e_i}' \boldsymbol{\gamma}_e$ with $\sigma_{u_i}^2$ replaced by $\sigma_{e_i}^2/\alpha$ in the likelihood function.

The shrinkage estimator for the variances proposed by eg, Gianola et al (1992), follows the same idea of the Bayesian estimator described in the *Bayesian approach* section. When a Gaussian prior density is employed for the dispersion parameters $\boldsymbol{\gamma}$, the hyperparameter ξ acts as a shrinker. But the Bayesian approach for a direct shrinkage of variance components assumes that heterogeneity in such components (residual and u components) is due only to one factor. The approach presented in this paper is more general since it can cope with more complex structures of stratification which may differ from one component to the other. Moreover, its mixed model structure allows great flexibility to adjust variances in relation to the amount of information for factors in the model; eg when data provide little information for some factors (or levels) or considerable for others, our procedure behaves like BLUP (or James–Stein) ie shrink estimates of dispersion parameters toward zero if there is little information; only with sufficient information can the estimate deviate. For instance, our methodology provides a simple and rational procedure to shrink herd variances (whatever they are, genetic, residual or phenotypic) towards different population values (eg regions, as proposed by Wiggans and VanRaden, 1991) due to poor accuracy of within herd or herd-year variances (Brotherstone and Hill, 1986). It then suffices to use a hierarchical (linear) mixed model for herd log-variances and take the population factor (eg region) as fixed and herd as random within that factor. An illustration of the flexibility and feasibility of our procedure was recently given by Weigel (1992) in analyzing sources of heterogeneous variances for milk and fat yield in US Holsteins.

Coming back to the case of a unique factor of variation for the sire variances, one can think of a simpler model, such as $y_{ijk} = \mu_i + u_{ij} + e_{ijk} (i = 1, \ldots I; \ j = 1, \ldots J; \ k = 1, \ldots n_{ij})$, where μ_i is the mean effect of environment i, u_{ij} is the (random) effect of the jth sire in the ith environment, such that $\mathbf{u}_i = \{u_{ij}\}_j \sim N(\mathbf{0}, \sigma_{u_i}^2 \mathbf{A})$, and e_{ijk} is the residual effect pertaining to the kth calf of the jth sire in the ith environment. Usually in such hierarchical models, it is assumed that $\mathrm{Cov}(\mathbf{u}_i, \mathbf{u}_{i'}) = \mathbf{0}$ if $i \neq i'$. On the contrary, our modelling procedure via the change in variables $u_j^* = \sigma_{u_i} u_{ij}$ (see [4]) takes into account covariances among the same (or genetically related) sires used in different herds, ie $\mathrm{Cov}(\mathbf{u}_i, \mathbf{u}_{i'}) = \sigma_{u_i} \sigma_{u_{i'}} \mathbf{A}_{ii'} (\mathbf{A}_{ii'} = $ relationship matrix pertaining to $\mathbf{u}_i$ and $\mathbf{u}_{i'}$) and so recovers the inter-block information. The loss in power in hypothesis testing due to ignoring that kind of information was recently investigated by Visscher (1992).

Although this presentation is restricted to a single random factor $\mathbf{u}^*$, it can be generalized to a multiple random factor situation. If such factors are uncorrelated, the extension is straightforward. When covariances exist, one may simply assume, as proposed by Quaas *et al* (1989), that heterogeneity in covariances is due to scaling. This means, for instance, that in a sire (s_i)-maternal grand sire (t_j) model $y_{ijk} = \mathbf{x}'_{ijk}\boldsymbol{\beta} + s_i + t_j + e_{ijk}$, one will model $\sigma^2_{s_h}, \sigma^2_{t_h}$ as previously, and assume that the covariance is $\sigma_{st_h} = \rho\sigma_{s_h}\sigma_{t_h}$ for stratum h. If the model is parameterized in terms of direct a_0 and maternal a_m effects as follows through the transformation $s_i = a_{0i}/2$ and $t_j = a_{0j}/4 + a_{mj}/2$, one can set the genetic correlation ρ_a to a constant, *ie* $\sigma_{a_{0m_h}} = \rho_a\sigma_{a_{0_h}}\sigma_{a_{m_h}}$. Notice that this condition is not equivalent to the previous one, except if $\sigma_{a_{0_h}}/\sigma_{a_{m_h}}$ does not depend on h.

Although the methodology is appealing, attention must be drawn to the feasibility of the method. The first problem is the inversion of the coefficient matrix in [16] required for the computation of the variance system [15]. In animal breeding applications, this matrix is usually very large. This limiting factor is already becoming less important due to constant progress in computing software and hardware. The technique of absorption is usually used to reduce the size of matrices to invert. Another approach is to approximate the inverse. One can, for instance, use a Taylor series expansion of order N for a square invertible matrix $\mathbf{A}$

$$\mathbf{A}^{-1} = \left\{ \sum_{n=0}^{N} (-1)^n \left[\mathbf{A}_0^{-1}(\mathbf{A} - \mathbf{A}_0) \right]^n \right\} \mathbf{A}_0^{-1} + o\left(\parallel \mathbf{A} - \mathbf{A}_0 \parallel^N \right)$$

where the square matrix $\mathbf{A}_0$ is a matrix close to $\mathbf{A}$ and is, of course, easy to invert, and where $\parallel \cdot \parallel$ denotes some norm on the space of invertible matrices. Methods viewed in Boichard *et al* (1992) can also help to approximate $\mathbf{A}^{-1}$ in particular cases such as sparse matrices, "animal model", etc.

Statistical power for likelihood ratio tests was investigated for detection of heterogeneous variances in the usual designs of quantitative genetics and animal breeding. Results given by Visscher (1992) and Shaw (1991) indicate generally low power values for detecting heterogeneity in genetic variance. According to Shaw, a nested design of 900 individuals out of 100 sire families provides a power of 0.5 for genetic variances differing by a factor of 2.5. This clearly indicates the minimum requirements in sample size and family numbers which should be met before carrying out such an analysis and the limits therein. Therefore it seems unrealistic to model genetic variances in practice according to more than 1 or 2 factors, and it might be wise to consider some of them as random if little information is provided by the data in each level of such factors.

Although statistical constraints are satisfied for estimates of the parameters of the model (positive variance estimates, intra-class correlations within $[-1, +1]$), some genetic constraints such as about the heritability estimate $(h^2 = 4\sigma_u^2/(\sigma_u^2 + \sigma_e^2)$ for a sire model) ranging within $[0,1]$ are not imposed by our model. This can be dealt with by choosing appropriate prior distributions on the dispersion parameters that would take this constraint into account, but this procedure appears to be extremely complicated. Fortunately, the unconstrained solutions are the constrained solutions if they are in the parameter space (here $\widehat{h}_i^2 \in [0,1], i = 1, \ldots I$). If not,

maximization procedure under constraints must be performed or the posterior distribution under the constraint can be obtained from the unconstrained posterior distribution multiplied by a corrective factor (Box and Tiao, 1973). This problem does not occur with an "animal model", but can arise when a "sire model" is used, and is not specifically related to heteroskedasticity.

From a statistical point of view, the procedure uses the concept of variance function (Davidian and Carroll, 1987) as an extension to dispersion parameters of the link function. Our presentation focuses on the log link function which is the most common choice in this field (see for instance San Cristobal, 1992, for a review of variance models) "for physical and numerical reasons" (Nair and Pregibon, 1988).

Following Davidian and Carroll (1987) or Duby *et al* (1975), the question can be asked whether or not variances vary according to means or location parameters. In the Maine–Anjou data, however, it does not seem to be the case, thus validating our choice in [10].

It would be interesting to extend our method to a fully generalized linear mixed model on means and on variances with or without common parameters between the mean model and the variance model. Numerical integration or Gibbs sampling procedures would then be required although approximate methods of inference can also be used for such models (Breslow and Clayton, 1992; Firth, 1992). Statistical problems arising with common parameters are already highlighted by van Houwelingen (1988).

With a fully fixed effect variance model, techniques of estimation and hypothesis testing for dispersion parameters presented here are those of the classical theory of likelihood inference (likelihood and likelihood ratio test), except that the marginal likelihood function $L(\boldsymbol{\gamma}; \mathbf{y})$ was preferred to the usual likelihood $L(\boldsymbol{\beta}, \boldsymbol{\gamma}; \mathbf{y})$, in the light of ideas behind REML estimators of variance components. This test reduces to Bartlett's test (Bartlett, 1937) for a one classification model in variances and under a saturated fixed model on the means (*ie* $\hat{\mu}_i = \bar{y}_i, i = 1, \ldots I$). Unfortunately, Bartlett's test is known to be sensitive to departure from normality (Box, 1953). Simulations are needed to study the robustness of this test and other competing tests. From a Bayesian perspective, the Bayes factor is usually applied for hypothesis testing (see Robert, 1992, for a discussion). The posterior Bayes factor (Aitkin, 1991) could also be used to compare dispersion models, but numerical integration would then be required (see the expression of the likelihood in [18]).

In this paper, focus was on an appropriate way to model heterogeneous variances, but the initial motivation was a best fitting of location parameters (animal evaluation for animal breeders). This difficult problem of feedback, also related to the Behrens–Fisher problem, has to be solved in our particular approach.

Moreover, a great research perspective is open on the important and complicated question of the joint modelling of means and variances (Aitkin, 1987; Nelder, 1991; Helder and Lee, 1991).

ACKNOWLEDGMENTS

REFERENCES

Aitkin M (1987) Modelling variance heterogeneity in normal regression using GLIM. *Appl Stat* 36, 332-339

Aitkin M (1991) Posterior Bayes factor. *JR Stat Soc B* 53, 111-142

Barnard GA (1970) Discussion on paper by Dr Kalbfleisch and Dr Sprott. *JR Stat Soc B* 32, 194-195

Bartlett MS (1937) Properties of sufficiency and statistical tests. *Proc R Soc Ser A* 160, 268-282

Bechhofer RE (1960) A multiplicative model for analyzing variances which are affected by several factors. *J Am Stat Assoc* 55, 245-264

Berger J (1985) *Statistical Decision Theory and Bayesian Analysis*. Springer-Verlag, New York, 2nd edn

Boichard D, Schaeffer RL, Lee AJ (1992) Approximate restricted maximum likelihood and approximate prediction error variance of the Mendelian sampling effect. *Genet Sel Evol* 24, 331-343

Boldman KG, Freeman AE (1990) Adjustment for heterogeneity of variances by herd production level in dairy cow and sire evaluation. *J Dairy Sci* 73, 503-512

Box GEP (1953) Non-normality and tests on variances. *Biometrika* 40, 318-335

Box GEP, Meyer RD (1986) Dispersion effects from fractional designs. *Technometrics* 28, 19-27

Box GEP, Tiao GC (1973) *Bayesian Inference in Statistical Analysis*. Addison-Wesley Publ Co Inc, Reading

Breslow NE, Clayton DG (1992) *Approximate Inference in Generalized Linear Mixed Models*. Tech Rep No 106, Univ Washington, Seattle, WA

Brotherstone S, Hill WG (1986) Heterogeneity of variance among herds for milk production. *Anim Prod* 42, 297-303

Brown BM (1982) Robustness against inequality of variances. *Aust J Stat* 24, 283-295

Cox DR, Hinkley DV (1974) *Theoretical Statistics*. Chapman and Hall, London

Davidian M, Carroll RJ (1987) Variance function estimation. *J Am Stat Assoc* 82, 1079-1091

De Veer JC, Van Vleck LD (1987) Genetic parameters for first lactation milk yields at three levels of herd production. *J Dairy Sci* 70 1434-1441

Dempster AP, Laird NM, Rubin DB (1977) Maximum likelihood from incomplete data *via* the EM algorithm. *JR Stat Soc B* 39, 1-38

Dong MC, Mao IL (1990) Heterogeneity of (Co)variance and heritability in different levels of intraherd milk production variance and of herd average. *J Dairy Sci* 73, 843-851

Duby C, Mougey Y, Ulmo J (1975) Analyse d'expériences à deux facteurs contrôlés quand l'écart type est une fonction connue de la moyenne. *Rev Stat Appl* 23, 5-33

Everett RW, Keown JF (1984) Mixed model sire evaluation with dairy cattle. Experience and genetic gain. *J Anim Sci* 59, 529-541

Firth D (1992) Quasi-Likelihood and Pseudo-Likelihood for Inference About a Variance Function. Preprint Ser No 223, Univ Southampton

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

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

Garrick DJ, Van Velck LD (1987) Aspects of selection for performance in several environments with heterogeneous variances. *J Anim Sci* 65, 409-421

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

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

Harville DA (1974) Bayesian inference for variance components using only error contrasts. *Biometrika* 61, 383-385

Henderson CR (1973) Sire evaluation and genetic trends. *In: Proc Anim Breeding Genet Symp in Honor of Dr JL Lush*. Am Soc Anim Sci Am Dairy Sci Assoc, Champaign, IL, 10-41

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

Hill WG (1984) On selection among groups with heterogeneous variance. *Anim Prod* 39, 473-477

Hill WG, Edwards MR, Ahmed MKA, Thompson R (1983) Heritability of milk yield and composition at different levels and variability of production. *Anim Prod* 36, 59-68

Houwelingen (van) JC (1988) Use and abuse of variance models in regression. *Biometrics* 44, 1073-1081

Kalbfleisch JD (1986) Pseudo-likelihood. *In: Encyclopedia of Statistical Science* (Kotz S, Johnson NL, eds) John Wiley and Sons, New York, vol 7, 324-327

Layard MWJ (1973) Robust large-sample tests for homogeneity of variances. *J Am Stat Assoc* 68, 195-198

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

McCullagh P, Nelder JA (1989) *Generalized Linear Models*. Chapman and Hall, London, 2nd edn

McCullogh JH (1985) On heteros*edasticity. *Econometrica* 53, 483

Meinert TR, Pearson RE, Vinson WE, Cassel BG (1988) Prediction of daughter's performance from dam's cow index adjusted for within-herd variance. *J Dairy Sci* 71, 220-2231

Morice E (1972) Tests de normalité d'une distribution observée. *Rev Stat Appl* 20, 5-25

Nair VN, Pregibon D (1988) Analysing dispersion effects from replicated factorial experiments. *Technometrics* 30, 247-257

Nelder JA (1991) Joint modelling of mean and dispersion. *In: 6th Proc Int Workshop Stat Modelling: Invited Papers* (Jansen W, van der Heijden PGM, eds) Utrecht, 15–19 July, 1991, *Methods Ser* MS-91-3, 45-52

Nelder JA, Lee Y (1991) Generalized linear models for the analysis of Taguchi-type experiments. *Appl Stochastic Models Data Anal* 7, 107-120

Nelder JA, Pregibon D (1987) An extended quasi-likelihood function. *Biometrika* 74, 221-232

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

Quaas RL, Pollak EJ (1980) Mixed model methodology for farm and ranch beef cattle testing programs. *J Anim Sci* 51, 1277-1287

Quaas RL, Garrick DJ, McElhenney WH (1989) Multitrait prediction for a type of model with heterogeneous genetic and residual covariance structures. *J Anim Sci* 67, 2529-2535

Robert C (1992) *L'Analyse Statistique Bayésienne.* Economica, Paris

San Cristobal M (1992) Méthodes d'inférence statistique en modélisation de la variance. Application en génétique quantitative. Thesis, Univ Paul Sabatier, Toulouse, France

Searle SR (1971) *Linear Models.* J Wiley and Sons, New York

Searle SR (1982) *Matrix Algebra Useful to Statistics.* J Wiley and Sons, New York

Shaw RG (1991) The comparison of quantitative genetic parameters between populations. *Evolution* 45, 143-151

Short TH, Blake RW, Quaas RL, Van Vleck LD (1990) Heterogeneous within-herd variance. 1. Genetic parameters for first and second lactation milk yields of grade Holstein cows. *J Dairy Sci* 73, 3312-3320

Tholen E (1990) *Untersuchung von Ursachen und Auswirkungen heterogener Varianzen der Indexmerkmale in der deutschen Schweineherdbuchzucht.* Wissenschaftliche Mitteillunger der Bundesforschungsanstalt für Landwirtschaft Braunschweig-Völkenrode (FAL) 111

Thompson R (1979) Sire evaluation. *Biometrics* 35, 339-353

Vinson WE (1987) Potential bias in genetic evaluations from differences in variation within herds. *J Dairy Sci* 70, 2450-2455

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

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

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

Weigel KA (1992) Estimation of heterogeneous components of variances in mixed linear models with an application to dairy cattle breeding. PhD thesis, Univ Wisconsin, Madison, WI

Wiggans GR, Van Raden PM (1991) Method and effect of adjustment for heterogeneous variance. *J Dairy Sci* 74, 4350-4357

Winkelman A, Schaeffer LR (1988) Effect of heterogeneity of variance on dairy sire evaluation. *J Dairy Sci* 71, 3033-3039

APPENDIX

Derivation of the Newton–Raphson algorithm

Maximizing $L(\gamma;y)$ with respect to γ leads to a non linear system which can be solved iteratively using a second order algorithm such as the Newton–Raphson procedure

$$-\ddot{L}(\gamma;y^{[t]}) \cdot (\gamma^{[t+1]} - \gamma^{[t]}) = \dot{L}(\gamma;y)^{[t]} \qquad [A.1]$$

where $\dot{L}(\gamma;y) = \dfrac{\partial L}{\partial \gamma}$ and $\ddot{L}(\gamma;y) = \dfrac{\partial^2 L}{\partial \gamma \partial \gamma'}$ are the gradient vector and Hessian matrix of $L(\gamma;y)$ respectively. They can be calculated using the following lemma:

Lemma

Let y be a rv depending on parameters θ_1 and θ_2 with prior joint density $p(\theta_1, \theta_2) = p(\theta_2|\theta_1) \cdot p(\theta_1)$. Let us note $L(\theta_1;y) = \ln p(\theta_1|y)$, $\dot{L}(\theta_1;y)$ its first derivative and $\ddot{L}(\theta_1;y)$ its second derivative with respect to θ_1. Then, following Foulley et al (1990, 1992).

$$\dot{L}(\theta_1;y) = E_c\left[\frac{\partial}{\partial\theta_1} \ln p(y, \theta_2|\theta_1)\right] + \frac{\partial}{\partial\theta_1} \ln p(\theta_1)$$

$$\ddot{L}(\theta_1;y) = E_c\left[\frac{\partial^2}{\partial\theta_1\partial\theta_1'} \ln p(y, \theta_2|\theta_1)\right] + \mathrm{Var}_c\left[\frac{\partial}{\partial\theta_1} \ln p(y, \theta_2|\theta_1)\right]$$

$$+ \frac{\partial^2}{\partial\theta_1\partial\theta_1'} \ln p(\theta_1)$$

where E_c and Var_c stand for expectation and variance, respectively, to be taken with respect to $p(\theta_2|\theta_1, y)$.

Notice that if θ_1 and θ_2 are a priori independent, then

$$\frac{\partial}{\partial\theta_1} \ln p(y, \theta_2|\theta_1) = \frac{\partial}{\partial\theta_1} \ln p(y|\theta_2, \theta_1)$$

This lemma applies to $\theta_1 = \gamma$ and $\theta_2 = \theta$. Thus

$$\dot{L}(\gamma;y) = E_c\left[\frac{\partial}{\partial\gamma} \ln p(y|\theta, \gamma)\right] \qquad [A.2]$$

and

$$\ddot{L}(\gamma;y) = E_c\left[\frac{\partial^2}{\partial\gamma\partial\gamma'} \ln p(y|\theta, \gamma)\right] + \mathrm{Var}_c\left[\frac{\partial}{\partial\gamma} \ln p(y|\theta,\gamma)\right] \qquad [A.3]$$

where E_c and Var_c stand for expectation and variance respectively to be taken with respect to $p(\theta|y,\gamma)$.

Terms related to the prior distribution of γ are ignored (eg [A.2, A.3] or added (see [30]) depending on whether diffuse or informative priors are considered for that vector.

In the normal, case, [A.2] and [A.3] are easy to obtain since

$$\ln p(\mathbf{y}|\boldsymbol{\theta}, \boldsymbol{\gamma}) = \text{const} - \sum_i (n_i/2)\mathbf{w}'_{e_i}\boldsymbol{\gamma}_e - (1/2)\sum_i \mathbf{e}'_i\mathbf{e}_i \exp\left(-\mathbf{w}'_{e_i}\boldsymbol{\gamma}_e\right) \qquad [\text{A.4}]$$

where $\mathbf{e}_i = \mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta} - \mathbf{Z}^*_i\mathbf{u}^* \exp\left(\mathbf{w}'_{u_i}\boldsymbol{\gamma}_u/2\right)$.
After some manipulations, first derivatives can be expressed as

$$\frac{\partial}{\partial \boldsymbol{\gamma}_e} \ln p(\mathbf{y}|\boldsymbol{\theta}, \boldsymbol{\gamma}) = \sum_i \mathbf{w}_{e_i} z_{e_i} = \mathbf{W}_e \mathbf{z}_e \qquad [\text{A.5}]$$

with $z_{e_i} = [(\mathbf{e}'_i\mathbf{e}_i/\sigma^2_{e_i}) - n_i]/2$,

$$\frac{\partial}{\partial \boldsymbol{\gamma}_u} \ln p(\mathbf{y}|\boldsymbol{\theta}, \boldsymbol{\gamma}) = \sum_i \mathbf{w}_{u_i} z_{u_i} = \mathbf{W}_u \mathbf{z}_u \qquad [\text{A.6}]$$

with $z_{u_i} = \sigma_{u_i}\mathbf{u}^{*'}\mathbf{Z}^{*'}_i\mathbf{e}_i/2\sigma^2_{e_i}$.
In order to get [A.2], one must take the conditional expectation of [A.5] and [A.6] with respect to the distribution of $\boldsymbol{\theta}|\mathbf{y}, \boldsymbol{\gamma}$

$$\dot{\mathbf{L}}(\boldsymbol{\gamma}; \mathbf{y}) = \mathbf{W}\hat{\mathbf{z}} \qquad [\text{A.7}]$$

with $\mathbf{W} = \begin{pmatrix} \mathbf{W}_e & \mathbf{0} \\ \mathbf{0} & \mathbf{W}_u \end{pmatrix}$ and $\hat{\mathbf{z}} = (\hat{\mathbf{z}}'_e, \hat{\mathbf{z}}'_u)'$.
Elements of $\hat{\mathbf{z}}$ can be easily expressed analytically using *eg* Searle (1971), p 65:

$$E_c(z_{e_i}) = \left\{ [\hat{\mathbf{e}}'_i\hat{\mathbf{e}}_i + \text{tr}(\mathbf{T}'_i\hat{\mathbf{C}}\mathbf{T}_i)]/(2\sigma^2_{e_i}) \right\} - n_i/2 \qquad [\text{A.8}]$$

$$E_c(z_{u_i}) = \left\{ \sigma_{u_i}\hat{\mathbf{u}}^{*'}\mathbf{Z}^{*'}_i\hat{\mathbf{e}}_i - \text{tr}[\sigma_{u_i}\mathbf{Z}^{*'}_i(\mathbf{X}_i\hat{\mathbf{C}}_{\beta u} + \sigma_{u_i}\mathbf{Z}^*_i\hat{\mathbf{C}}_{uu})] \right\} /(2\sigma^2_{e_i}) \qquad [\text{A.9}]$$

where $\hat{\mathbf{e}}_i = \mathbf{y}_i - \mathbf{X}_i\hat{\boldsymbol{\beta}} - \sigma_{u_i}\mathbf{Z}^*_i\hat{\mathbf{u}}^*, \hat{\boldsymbol{\beta}} = E_c(\boldsymbol{\beta}), \hat{\mathbf{u}}^* = E_c(\mathbf{u}^*), \hat{\mathbf{C}} = \text{Var}_c(\boldsymbol{\theta}), \hat{\mathbf{C}}_{\beta u}$ and $\hat{\mathbf{C}}_{uu}$ are submatrices of $\hat{\mathbf{C}}$ pertaining to the $(\boldsymbol{\beta}, \mathbf{u}^*)$ and the $(\mathbf{u}^*, \mathbf{u}^*)$ blocks of $\mathbf{C}$ respectively.
Expressions of the second derivatives are obtained in the same way:

$$\frac{\partial^2}{\partial \boldsymbol{\gamma}_e \partial \boldsymbol{\gamma}'_e} \ln p(\mathbf{y}|\boldsymbol{\theta}, \boldsymbol{\gamma}) = -\mathbf{W}'_e \mathbf{Q}_{ee,E} \mathbf{W}_e \qquad [\text{A.10}]$$

with $\mathbf{Q}_{ee,E} = \text{diag}\{\mathbf{e}'_i\mathbf{e}_i/(2\sigma^2_{e_i})\}$, giving $\hat{\mathbf{Q}}_{ee,E} = E_c(\mathbf{Q}_{ee,E})$,

$$\frac{\partial^2}{\partial \boldsymbol{\gamma}_e \partial \boldsymbol{\gamma}'_u} \ln p(\mathbf{y}|\boldsymbol{\theta}, \boldsymbol{\gamma}) = -\mathbf{W}'_e \mathbf{Q}_{eu,E} \mathbf{W}_u \qquad [\text{A.11}]$$

with $\mathbf{Q}_{eu,E} = \text{diag}\{\sigma_{u_i}\mathbf{u}^*\mathbf{Z}^*_i\mathbf{e}_i/(2\sigma^2_{e_i})\}$, giving $\hat{\mathbf{Q}}_{eu,E} = E_c(\mathbf{Q}_{eu,E})$,

$$\frac{\partial^2}{\partial \boldsymbol{\gamma}_u \partial \boldsymbol{\gamma}'_u} \ln p(\mathbf{y}|\boldsymbol{\theta}, \boldsymbol{\gamma}) = -\mathbf{W}'_u \mathbf{Q}_{uu,E} \mathbf{W}_u \qquad [\text{A.12}]$$

with $\mathbf{Q}_{uu,E} = \text{diag}\{[\sigma_{u_i}\mathbf{u}^*\mathbf{Z}_i^{*\prime}\mathbf{e}_i - \sigma_{u_i}^2\mathbf{u}^{*\prime}\mathbf{Z}_i^{*\prime}\mathbf{Z}_i^*\mathbf{u}^*]/(4\sigma_{e_i}^2)\}$, giving $\widehat{\mathbf{Q}}_{uu,E} = E_c(\mathbf{Q}_{uu,E})$.

Formulae [A.10], [A.11], [A.12] provide the first term of $\ddot{\mathbf{L}}(\boldsymbol{\gamma};\mathbf{y})$ in [A.3]. The second term of $\ddot{\mathbf{L}}(\boldsymbol{\gamma};\mathbf{y})$ involves the variance of [A.5] and [A.6]:

$$\text{Var}_c\left[\frac{\partial}{\partial\boldsymbol{\gamma}}\ln p(\mathbf{y}|\boldsymbol{\theta},\boldsymbol{\gamma})\right] = \mathbf{W}'\text{Var}_c(\mathbf{z})\mathbf{W} \qquad [\text{A.13}]$$

Elements of $\text{Var}_c(\mathbf{z}) = [\text{Cov}_c(\mathbf{z}_\alpha, \mathbf{z}_\beta)]_{\alpha,\beta=e,u}$ are obtained after some tedious but straightforward calculations, using, *eg* Searle (1971). Finally, maximization of the marginal likelihood can be performed *via* the iterative system

$$\mathbf{W}'\widehat{\mathbf{Q}}^{[t]}\mathbf{W}(\widehat{\boldsymbol{\gamma}}^{[t+1]} - \widehat{\boldsymbol{\gamma}}^{[t]}) = \mathbf{W}'\widehat{\mathbf{z}}^{[t]} \qquad [\text{A.14}]$$

at the current iteration $[t]$, with $\mathbf{W}$ and $\widehat{\mathbf{z}}$ as in [A.7], and $\widehat{\mathbf{Q}} = \begin{pmatrix} \widehat{\mathbf{Q}}_{ee} & \widehat{\mathbf{Q}}_{eu} \\ \widehat{\mathbf{Q}}_{eu}' & \widehat{\mathbf{Q}}_{uu} \end{pmatrix}$, with

$\widehat{\mathbf{Q}}_{\alpha\beta} = \widehat{\mathbf{Q}}_{\alpha\beta,E} - \text{Cov}_c(\mathbf{z}_\alpha, \mathbf{z}_\beta), \alpha, \beta = e, u$.

$\widehat{\boldsymbol{\theta}}$ and $\widehat{\mathbf{C}}$ defined in [A.9] and involved also in [A.14] can be computed as follows. From Bayes theorem:

$$p(\boldsymbol{\theta}|\mathbf{y},\boldsymbol{\gamma} = \widehat{\boldsymbol{\gamma}}) \propto p(\mathbf{y}|\boldsymbol{\theta},\boldsymbol{\gamma} = \widehat{\boldsymbol{\gamma}})p(\boldsymbol{\theta}) \propto p(\mathbf{y}|\boldsymbol{\theta},\boldsymbol{\gamma} = \widehat{\boldsymbol{\gamma}})p(\mathbf{u}^*) \qquad [\text{A.15}]$$

Now,

$$p(\mathbf{u}^*) = (2\pi)^{-q/2}\exp\left(-\mathbf{u}^{*\prime}\mathbf{A}^{-1}\mathbf{u}^*/2\right) \qquad [\text{A.16}]$$

and $p(\mathbf{y}|\boldsymbol{\theta},\boldsymbol{\gamma} = \widehat{\boldsymbol{\gamma}})$ is obtained from [A.4], the maximization of [A.15]

$$\frac{\partial}{\partial\boldsymbol{\theta}}\ln p(\boldsymbol{\theta}|\mathbf{y},\boldsymbol{\gamma} = \widehat{\boldsymbol{\gamma}}) = 0 \qquad [\text{A.17}]$$

leads to the system

$$(\widehat{\mathbf{T}}'\widehat{\mathbf{R}}^{-1}\widehat{\mathbf{T}} + \boldsymbol{\Sigma}^-)\widehat{\boldsymbol{\theta}} = \widehat{\mathbf{T}}'\widehat{\mathbf{R}}^{-1}\mathbf{y} \qquad [\text{A.18}]$$

where $\widehat{\mathbf{T}} = \mathbf{T}(\widehat{\boldsymbol{\gamma}}_u), \widehat{\mathbf{R}} = \mathbf{R}(\widehat{\boldsymbol{\gamma}}_e), \boldsymbol{\Sigma}^-$ are as defined in the text (see [16]). Hence $\widehat{\boldsymbol{\theta}} = E_c(\boldsymbol{\theta})$ can be viewed as the solution of [A.18] and $\widehat{\mathbf{C}} = \text{Var}_c(\boldsymbol{\theta})$ as the inverse of the coefficient matrix. At convergence, $\widehat{\boldsymbol{\theta}}$ is an empirical Bayes estimator of $\boldsymbol{\theta}$, since it maximizes $p(\boldsymbol{\theta}|\mathbf{y},\boldsymbol{\gamma} = \widehat{\boldsymbol{\gamma}})$.