Genet Sel Evol (1995) 27, 33-42

Original article

Estimating variances and covariances for bivariate animal models using scaling and transformation

R Thompson [1], RE Crump [1,*], J Juga [1,**], PM Visscher [2,†]

[1] *Roslin Institute (Edinburgh) [formerly AFRC Institute of Animal Physiology and Genetics Research (Edinburgh Research Station)], Roslin, Midlothian EH25 9PS;*
[2] *Institute of Cell, Animal and Population Biology, University of Edinburgh, West Mains Road, Edinburgh, EH9 3JT, UK*

(Received 4 August 1993; accepted 1st September 1994)

Summary – The estimation of genetic parameters in bivariate animal models is considered. It is shown that in a variety of models the computation can be reduced by introducing scaled and transformed independent traits. This allows maximization over smaller dimensions of parameter space. In 1 numerical example the procedure reduced the computation by a factor of 8. The advantages of transformed models are outlined.

variance estimation / covariance estimation / animal model / transformation

Résumé – L'estimation des variances et covariances dans un modèle individuel à 2 caractères après standardisation et transformation des variables. *Cet article traite de l'estimation des paramètres génétiques dans un modèle individuel à 2 variables. On montre que, dans beaucoup de situations, le temps de calcul peut être diminué en standardisant les caractères et en les rendant indépendants par une transformation, ce qui permet une maximisation sur un espace de paramètres de moindre dimension. Un exemple numérique particulier montre que le temps de calcul est divisé par 8. Les avantages des différents modèles transformés sont présentés.*

estimation de variance / estimation de covariance / modèle animal / transformation

* Present Address: Genetics and Behavioural Sciences Department, Scottish Agricultural College Edinburgh, West Mains Road, Edinburgh, EH9 3JG, UK.
** Present Address: The Finnish Animal Breeding Association, PO Box 40, SF-01301, Vantaa, Finland.
† Present address: Roslin Institute (Edinburgh), Roslin, Midlothian, EH25 9PS, UK.

INTRODUCTION

There is often the need for estimation of genetic and environmental variances and covariances from animal breeding data. For example, to consider the responses from alternative selection schemes or to efficiently predict the genetic merit of animals. Usually in animal breeding schemes animals are selected on some criteria and so methods of analysis are needed that take account of selection. Maximum likelihood (ML) methods have been shown to take account of selection in univariate and multivariate settings (for example, Henderson *et al*, 1959; Thompson, 1973) if the records on which selection is based are included in the data. If this condition is only partially fulfilled, ML methods are less biased by selection than analysis of variance methods (Meyer and Thompson, 1984). A restricted or residual maximum likelihood (REML) procedure uses the likelihood of residuals and has the advantage that it takes account of the estimation of fixed effects when estimating variance components and corrects for degrees of freedom (Patterson and Thompson, 1971). In general these methods are computationally expensive requiring the solution and inversion of equations of the order of number of animals $\times$ number of traits, but there are simplifications when all the traits are measured on all animals and the same fixed effect model is applied to all traits (Thompson, 1977; Meyer, 1985).

In the past, estimation methods have used equations based on first and second differentials, but recently Graser *et al* (1987) and Meyer (1991) have shown how the likelihood can be calculated recursively in univariate and multivariate settings and advocated the direct maximization of the likelihood. Meyer (1991) also showed that the computational effort can be reduced if, given some of the parameters, it is relatively easy to maximize the likelihood for the rest of the parameters. The maximization then has 2 stages and the dimension of search is reduced.

Meyer (1991) showed the advantage of these techniques for models with equal design matrices and 3 variance components. In the recent analysis of data from a pig nucleus herd (Crump, 1992) we required to estimate genetic correlations between male and female performance when the 2 sexes were reared in different environments and between growth and reproductive traits, and these models do not directly fit into Meyer's algorithm.

In this paper we show how Meyer's method can be extended to fit these and other models, by the introduction of scaling and transformation models. The models considered included those for bivariate traits when different fixed effect models are appropriate to each trait and to models with equal design matrices with more than 2 traits.

ESTIMATION

We will consider in turn estimation for 3 models.

Model 1

The first and simplest model is of the form

$$\mathbf{y}_1 = \mathbf{X}_1\boldsymbol{\beta}_1 + \mathbf{Z}_1\mathbf{u}_1 + \mathbf{e}_1$$
$$\mathbf{y}_2 = \mathbf{X}_2\boldsymbol{\beta}_2 + \mathbf{Z}_2\mathbf{u}_2 + \mathbf{e}_2$$

with

$$\text{var}(\mathbf{u}_1) = \mathbf{A}\sigma^2_{A1}$$
$$\text{var}(\mathbf{u}_2) = \mathbf{A}\sigma^2_{A2}$$
$$\text{cov}(\mathbf{u}_1, \mathbf{u}_2) = \mathbf{A}\sigma_{A12}$$

and $\text{var}(\mathbf{e}_1) = \mathbf{I}\sigma^2_{e1}$ and $\text{var}(\mathbf{e}_2) = \mathbf{I}\sigma^2_{e2}$ and $\mathbf{e}_1$ and $\mathbf{e}_2$ are uncorrelated, and the fixed effects $\boldsymbol{\beta}_1$ and $\boldsymbol{\beta}_2$ have no elements in common, and the random effects $\mathbf{u}_1$, $\mathbf{u}_2$, $\mathbf{e}_1$ and $\mathbf{e}_2$ normally distributed. The vectors $\mathbf{y}_1$ and $\mathbf{y}_2$ are of length n_1 and n_2 and matrices $\mathbf{X}_1$, $\mathbf{X}_2$, $\mathbf{Z}_1$ and $\mathbf{Z}_2$ are of size $n_1 \times t_1$, $n_2 \times t_2$, $n_1 \times m$ and $n_2 \times m$. Our motivation was a case when a trait $\mathbf{y}_1$ was measured on males and $\mathbf{y}_2$ was measured on females and there was interest in the genetic covariance between traits (σ_{A12}) and there was no environmental covariance between the records. This model was analysed by Schaeffer *et al* (1978) using a method that involved calculation of the second differentials of the likelihood and inversion of a matrix of order $2m$ for each iteration.

If the 2 residual variances are homogeneous then the univariate method used by Meyer (1989) can be used if the model is written in the form

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u} + \mathbf{e}$$

with

$$\mathbf{y} = \begin{pmatrix} \mathbf{y}_1 \\ \mathbf{y}_1 \end{pmatrix}, \ \mathbf{X} = \begin{pmatrix} \mathbf{X}_1 & 0 \\ 0 & \mathbf{X}_2 \end{pmatrix}, \ \mathbf{Z} = \begin{pmatrix} \mathbf{Z}_1 & 0 \\ 0 & \mathbf{Z}_2 \end{pmatrix}$$
$$\boldsymbol{\beta} = \begin{pmatrix} \boldsymbol{\beta}_1 \\ \boldsymbol{\beta}_2 \end{pmatrix}, \ \mathbf{u} = \begin{pmatrix} \mathbf{u}_1 \\ \mathbf{u}_2 \end{pmatrix}, \ \mathbf{e} = \begin{pmatrix} \mathbf{e}_1 \\ \mathbf{e}_2 \end{pmatrix}$$

If, however, the residual variances are not homogeneous the situation is slightly more complicated. If $\sigma^2_{e1} > \sigma^2_{e2}$ the vector of residual can be written as

$$\mathbf{e} = \begin{pmatrix} \mathbf{e}_1 \\ \mathbf{e}_2 \end{pmatrix} = \begin{pmatrix} \mathbf{e}_{12} \\ \mathbf{e}_2 \end{pmatrix} + \begin{pmatrix} \mathbf{e}_1 - \mathbf{e}_{12} \\ 0 \end{pmatrix}$$

and if the elements of the first vector have variance σ^2_{e2} and the non-zero elements of the second vector variance $\sigma^2_{e1} - \sigma^2_{e2}$ then it is seen that an extra component could be introduced (if $\sigma^2_{e1} > \sigma^2_{e2}$) and this component estimated, but this will increase the dimension of the search by 1. We now develop a method that does not increase the dimension of search. It is useful to think of a composite matrix of $\mathbf{y}_1$ and $\mathbf{y}_2$, of the form $\mathbf{Y}_c = [\mathbf{y}^*_1 \mathbf{y}^*_2]$ with

$$\mathbf{y}^*_1 = \begin{bmatrix} \mathbf{y}_1 \\ \mathbf{0} \end{bmatrix} \quad \text{and} \quad \mathbf{y}^*_2 = \begin{bmatrix} \mathbf{0} \\ \mathbf{y}_2 \end{bmatrix}$$

so that the vector of observations is $\mathbf{y} = \mathbf{y}^*_1 + \mathbf{y}^*_2$.

We wish to maximize the log-likelihood of error contrasts (Patterson and Thompson, 1971).

$$\text{Log}L = -1/2[\text{const} + \log|\mathbf{V}| + \log|\mathbf{X}'\mathbf{V}^{-1}\mathbf{X}| + (\mathbf{y} - \mathbf{X}\widehat{\boldsymbol{\beta}})'\mathbf{V}^{-1}(\mathbf{y} - \mathbf{X}\widehat{\boldsymbol{\beta}})] \quad [1]$$

with $\widehat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1}\mathbf{y}$ and $\mathbf{V}$ is of the form $\mathbf{R}+\mathbf{ZGZ}' = \mathbf{R}+\mathbf{Z}(\mathbf{A}\times\mathbf{T}_\mathrm{A})\mathbf{Z}'$ with $\times$ denoting direct product and

$$\mathbf{R} = \begin{bmatrix} \mathbf{I}\sigma^2_{\mathrm{e}1} & 0 \\ 0 & \mathbf{I}\sigma^2_{\mathrm{e}2} \end{bmatrix} = \mathbf{QQ}' \quad \text{with} \quad \mathbf{Q} = \begin{bmatrix} \mathbf{I}\sigma_{\mathrm{e}1} & 0 \\ 0 & \mathbf{I}\sigma_{\mathrm{e}2} \end{bmatrix}$$

and

$$\mathbf{T}_\mathrm{A} = \begin{bmatrix} \sigma^2_{\mathrm{A}1} & \sigma_{\mathrm{A}12} \\ \sigma_{\mathrm{A}12} & \sigma^2_{\mathrm{A}2} \end{bmatrix}$$

Then $\mathbf{V} = \mathbf{Q}[\mathbf{I} + \mathbf{ZG}_\mathrm{S}\mathbf{Z}']\mathbf{Q}' = \mathbf{Q}[\mathbf{I} + \mathbf{Z}(\mathbf{A}\times\mathbf{T}_\mathrm{AS})\mathbf{Z}']\mathbf{Q}' = \mathbf{QHQ}'$ with

$$\begin{pmatrix} \sigma_{\mathrm{e}1} & \mathbf{0} \\ \mathbf{0} & \sigma_{\mathrm{e}2} \end{pmatrix} \mathbf{T}_\mathrm{AS} \begin{pmatrix} \sigma_{\mathrm{e}1} & \mathbf{0} \\ \mathbf{0} & \sigma_{\mathrm{e}2} \end{pmatrix} = \mathbf{T}_\mathrm{A}$$

so that

$$\mathbf{T}_\mathrm{AS} = \begin{bmatrix} \sigma^2_{\mathrm{A}1}/\sigma^2_{\mathrm{e}1} & \sigma_{\mathrm{A}12}/(\sigma_{\mathrm{e}1}\sigma_{\mathrm{e}2}) \\ \sigma_{\mathrm{A}12}/(\sigma_{\mathrm{e}1}\sigma_{\mathrm{e}2}) & \sigma^2_{\mathrm{A}2}/\sigma^2_{\mathrm{e}2} \end{bmatrix}$$

is a scaled version of $\mathbf{T}_\mathrm{A}$.

The terms in [1] can be written in terms of $\mathbf{H}$, $\sigma^2_{\mathrm{e}1}$, $\sigma^2_{\mathrm{e}2}$ and $\mathbf{Y}_\mathrm{c}$ as follows:

$$\log|\mathbf{V}| = n_1 \log\sigma^2_{\mathrm{e}1} + n_2 \log\sigma^2_{\mathrm{e}2} + \log|\mathbf{H}|$$
$$\log|\mathbf{X}'\mathbf{V}^{-1}\mathbf{X}| = -t_1 \log\sigma^2_{\mathrm{e}1} - t_2 \log\sigma^2_{\mathrm{e}2} - \log|\mathbf{X}'\mathbf{H}^{-1}\mathbf{X}|$$
$$\boldsymbol{\beta} = (\mathbf{X}'\mathbf{H}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{H}^{-1}\mathbf{Y}_\mathrm{c}\mathbf{s} = \boldsymbol{\beta}_\mathrm{c}\mathbf{s}$$

and $(\mathbf{y} - \mathbf{X}\widehat{\boldsymbol{\beta}})'\mathbf{V}^{-1}(\mathbf{y} - \mathbf{X}\widehat{\boldsymbol{\beta}}) = \mathbf{s}'(\mathbf{Y}_\mathrm{c} - \mathbf{X}\widehat{\boldsymbol{\beta}}_\mathrm{c})'\mathbf{H}^{-1}(\mathbf{Y}_\mathrm{c} - \mathbf{X}\widehat{\boldsymbol{\beta}}_\mathrm{c})\mathbf{s}$

with $\mathbf{s} = \begin{pmatrix} 1/\sigma_{\mathrm{e}1} \\ 1/\sigma_{\mathrm{e}2} \end{pmatrix}$

with $\widehat{\boldsymbol{\beta}}_\mathrm{c}$ a matrix of effects for the 2 traits $\mathbf{y}^*_1$, $\mathbf{y}^*_2$ and $(\mathbf{Y}_\mathrm{c} - \mathbf{X}\widehat{\boldsymbol{\beta}}_\mathrm{c})'\mathbf{H}^{-1}(\mathbf{Y}_\mathrm{c} - \mathbf{X}\widehat{\boldsymbol{\beta}}_\mathrm{c})$ a 2×2 matrix of sums of squares and cross-products of residuals for these 2 traits.

By using these relationships, and the formulae for log-likelihood of a model with variance matrix $\mathbf{H}$ developed by Graser *et al* (1987), it can be shown that $\log L$ can be written using

$$-2\log L = df_1 \log\sigma^2_{\mathrm{e}1} + df_2 \log\sigma^2_{\mathrm{e}2} + \log|\mathbf{G}_\mathrm{s}| + \log|\mathbf{C}| + \mathbf{s}'\mathbf{Y}'_\mathrm{c}\mathbf{P}\mathbf{Y}_\mathrm{c}\mathbf{s} \qquad [2]$$

where $df_i = n_i - t_i$ $(i = 1, 2)$, $\mathbf{C}$ is the coefficient matrix of mixed-model equations with variance matrix $\mathbf{I} + \mathbf{ZG}_\mathrm{s}\mathbf{Z}'$ and $\mathbf{P} = \mathbf{H}^{-1} - \mathbf{H}^{-1}\mathbf{X}(\mathbf{X}'\mathbf{H}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{H}^{-1}$. The terms in $\mathbf{C}$ and $\mathbf{P}$ can be calculated in an analogous way to Graser *et al* (1987) by the formation of $\mathbf{M}$, an augmented matrix of mixed-model coefficients and right-hand sides of the form:

$$\mathbf{M} = \begin{bmatrix} \mathbf{Y}'_\mathrm{c}\mathbf{Y}_\mathrm{c} & \mathbf{Y}'_\mathrm{c}\mathbf{X} & \mathbf{Y}'_\mathrm{c}\mathbf{Z} \\ \mathbf{X}'\mathbf{Y}_\mathrm{c} & \mathbf{X}'\mathbf{X} & \mathbf{X}'\mathbf{Z} \\ \mathbf{Z}'\mathbf{Y}_\mathrm{c} & \mathbf{Z}'\mathbf{X} & \mathbf{Z}'\mathbf{Z} + \mathbf{G}_\mathrm{s}^{-1} \end{bmatrix}$$

with $\log|\mathbf{C}|$ associated with the pivots involved in the Gaussian elimination of the terms associated with $\mathbf{X}$ and $\mathbf{Z}$, and

$$\mathbf{Y}_c'\mathbf{P}\mathbf{Y}_c = \begin{bmatrix} U_{11} & U_{12} \\ U_{21} & U_{22} \end{bmatrix}$$

is the (2×2) residual matrix after elimination of the term associated with $\mathbf{X}$ and $\mathbf{Z}$. The term $\log|\mathbf{G}_s|$ can be written, using properties of direct products (Searle, 1982), as $m \log|\mathbf{T}_{AS}| + 2 \log|\mathbf{A}|$, with the last term independent of the variance parameters.

For given $\mathbf{T}_{AS}$, $\log L$ can be written as a function of terms of $\mathbf{M}$, σ^2_{e1} and σ^2_{e2} using

$$-2 \log L = df_1 \log\sigma^2_{e1} + df_2 \log\sigma^2_{e2} + m \log|\mathbf{T}_{AS}| + 2 \log|\mathbf{A}| + \log|\mathbf{C}| + (1/\sigma_{e1}\ 1/\sigma_{e2}) \begin{bmatrix} U_{11} & U_{12} \\ U_{21} & U_{22} \end{bmatrix} \begin{pmatrix} 1/\sigma_{e1} \\ 1/\sigma_{e2} \end{pmatrix} \quad [3]$$

Differentiating this log-likelihood with respect to σ^2_{e1} and σ^2_{e2}, noting that $\mathbf{G}_s$ and $\mathbf{C}$ are independent of σ^2_{e1} and σ^2_{e2}, and equating to zero gives

$$df_1\sigma^2_{e1} = U_{11} + rU_{12} \quad [4]$$

$$df_2\sigma^2_{e2} = (1/r)U_{12} + U_{22} \quad [5]$$

with the ratio $r = \sigma_{e1}/\sigma_{e2}$ satisfying the equation

$$r^2U_{22}/df_2 + rU_{12}df^* - U_{11}/df_1 = 0 \quad [6]$$

with $df^* = 1/df_2 - 1/df_1$ and so

$$r = \left\{-(df^*)U_{12} + \left[(df^*)^2U_{12}^2 - (4U_{22}U_{11})/(df_1df_2)\right]^{1/2}\right\}/[2U_{22}/df_2] \quad [7]$$

and hence σ^2_{e1} and σ^2_{e2} can be found from [4] and [5] given the 3 parameters in $\mathbf{T}_{AS}$. Substitution of the values for σ^2_{e1} and σ^2_{e2} in [3] gives $\log L$ for given $\mathbf{T}_{AS}$.

Hence the ML estimates could be found with maximization over the 3 parameters in $\mathbf{T}_{AS}$. Essentially the structure of the model has allowed the scaling of $\mathbf{y}_1$ and $\mathbf{y}_2$ by σ_{e1} and σ_{e2} to be carried out independently of $\mathbf{T}_{AS}$.

Model 2

A natural extension of *Model 1* is to allow a non-zero covariance matrix between the residuals $\mathbf{e}_1$ and $\mathbf{e}_2$, $\mathbf{B}\sigma_{e12}$ with, for simplicity, the $(n_1 \times n_2)$ matrix

$$\mathbf{B} = \begin{pmatrix} \mathbf{I}_2 \\ 0 \end{pmatrix}$$

so that the first n_2 animals are measured on $\mathbf{y}_1$ and $\mathbf{y}_2$ and this will be denoted *Model 2*. There are obviously several ways of writing this model. We give below

a form of this model that has 2 properties. Firstly, this form allows univariate algorithms to be used to calculate likelihoods. This is achieved by introducing uncorrelated effects, $\mathbf{u}_b$, to help model covariance effects. Secondly, in order that scaled versions of $\mathbf{y}_1$ and $\mathbf{y}_2$ have homogeneous contributions for $\mathbf{e}_1$ and $\mathbf{e}_2$, as in model 1, but also for $\mathbf{u}_b$, scaling factors, a and b for the contributions of $\mathbf{u}_b$ to $\mathbf{y}_1$ and $\mathbf{y}_2$ are introduced. This model has the general form:

$$\begin{aligned} \mathbf{y}_1 &= \mathbf{X}_1\boldsymbol{\beta}_1 + \mathbf{Z}_1\mathbf{u}_1 + a\mathbf{Z}_{b11}\mathbf{u}_b + \mathbf{e}_1^* \\ \mathbf{y}_2 &= \mathbf{X}_2\boldsymbol{\beta}_2 + \mathbf{Z}_2\mathbf{u}_2 + b\mathbf{Z}_{b21}\mathbf{u}_b + \mathbf{e}_2^* \end{aligned} \qquad [8]$$

with $\mathbf{Z}_{b11} = \mathbf{I}, \mathbf{Z}_{b21} = \mathbf{B}'$

It should be noted that the range of maximization of σ^2_{b1} is from minus infinity to infinity rather than 0 to infinity in order to allow negative environmental variances. This only needs minor changes to the algorithm. The term $\log|\mathbf{G}| + \log|\mathbf{C}|$ is normally found from, say, $\Sigma \log g_i + \log c_i$ where the terms g_i and c_i are positive. If σ^2_{b1} is negative then the term $\log|\mathbf{G}| + \log|\mathbf{C}|$ is still positive definite and therefore there are an even number of negative terms in g_i and c_i. Therefore $\log|\mathbf{G}| + \log|\mathbf{C}|$ can be calculated as $\Sigma \log|g_i| + \Sigma \log|c_i|$. The equivalent residual variance structure has 2 equivalent formulations

$$\begin{pmatrix} \mathbf{I}_1\sigma^2_{e1} & \mathbf{B}\sigma_{e12} \\ \mathbf{B}'\sigma_{e12} & \mathbf{I}_2\sigma^2_{e2} \end{pmatrix} \text{ and } \begin{pmatrix} \mathbf{I}_1\sigma^{*2}_{e1} + a^2\mathbf{I}_1\sigma^2_{b1} & ab\mathbf{B}\sigma^2_{b1} \\ ab\mathbf{B}'\sigma^2_{b1} & \mathbf{I}_2\sigma^{*2}_{e2} + b^2\mathbf{I}_2\sigma^2_{b1} \end{pmatrix}$$

so the relationships between the parameters are:

$$\begin{aligned} \sigma^2_{e1} &= \sigma^{*2}_{e1} + a^2\sigma^2_{b1} \\ \sigma_{e12} &= ab\sigma^2_{b1} \\ \sigma^2_{e2} &= \sigma^{*2}_{e2} + b^2\sigma^2_{b1} \end{aligned}$$

Any non-zero value for a and b can be used but if $a = \sigma^*_{e1}$ and $b = \sigma^*_{e2}$ then the log-likelihood can be expressed in a form analogous to [2] using a matrix $\mathbf{M}$ given by

$$\mathbf{M} = \begin{bmatrix} \mathbf{Y}_c'\mathbf{Y}_c & \mathbf{Y}_c'\mathbf{X} & \mathbf{Y}'\mathbf{Z} & \mathbf{Y}_c'\mathbf{Z}_b \\ \mathbf{X}'\mathbf{Y}_c & \mathbf{X}'\mathbf{X} & \mathbf{X}'\mathbf{Z} & \mathbf{X}'\mathbf{Z}_b \\ \mathbf{Z}'\mathbf{Y}_c & \mathbf{Z}'\mathbf{X} & \mathbf{Z}'\mathbf{Z} + \mathbf{G}_s^{-1} & \mathbf{Z}'\mathbf{Z}_b \\ \mathbf{Z}_c'\mathbf{Y}_c & \mathbf{Z}_c'\mathbf{X} & \mathbf{Z}_b'\mathbf{Z} & \mathbf{Z}_b'\mathbf{Z}_b + \mathbf{I}(1/\sigma^2_{b1}) \end{bmatrix} \qquad [9]$$

with $\mathbf{Z}_b = \begin{bmatrix} \mathbf{Z}_{b11} \\ \mathbf{Z}_{b21} \end{bmatrix}$ and $\mathbf{G}_S = \mathbf{T}^*_{AS}$ and $\mathbf{T}_A = \begin{pmatrix} \sigma^*_{e1} & 0 \\ 0 & \sigma^*_{e2} \end{pmatrix} \mathbf{T}^*_{AS} \begin{pmatrix} \sigma^*_{e1} & 0 \\ 0 & \sigma^*_{e2} \end{pmatrix}$

Equations [4] – [6] allow estimation of σ^{*2}_{e1} and σ^{*2}_{e2} given $\mathbf{T}^*_{AS}$ and σ^2_{b1}. Hence a 6 parameter problem has been converted to a search over 4 parameters. Crump (1992) has considered extensions of this model to allow estimation of genetic covariances between growth and reproductive traits.

Model 3

Models 1 and *2* are models that allow different fixed effect models with 2 traits, but there are interesting implications if a similar approach is applied to models with p traits, where all traits are measured on all animals and the same fixed effect structure is applied to all p traits. When there are 2 variance component matrices to be estimated a canonical transformation to make the traits independent can be useful in reducing p trait equations into p sets of univariate calculations (Thompson, 1977; Meyer, 1985; Taylor *et al*, 1985). Meyer (1991), for the case of additive and residual matrices, has recently emphasized a 2-stage maximization procedure, using $\mathbf{S}$, a $p \times p$ transformation matrix, and $\boldsymbol{\lambda}$, a $p \times p$ diagonal matrix of canonical heritabilities. For a given value of $\boldsymbol{\lambda}$, the log likelihood can be written in a form analogous to [2], with the use of p matrices of the form of $\mathbf{M}$ and with $\mathbf{Y}_c$ an $n \times p$ matrix $\mathbf{Y}$ with the ith column of $\mathbf{Y}$ the ith variate $\mathbf{y}_i$. Given $\boldsymbol{\lambda}$, the log-likelihood maximization in terms of $\mathbf{S}$ is computationally easier, and in fact when $p = 2$ there is an explicit estimate of $\mathbf{S}$ in terms of the residual matrices (Juga and Thompson, 1992). On a small numerical example Meyer (1991) has reduced computation to a half by such a technique and one would expect larger savings as p increased.

For more than 2 sets of components, there is a natural extension of Meyer's method and the method used for *Models 1* and *2*. To illustrate the method suppose 3 symmetric (2×2) matrices $\mathbf{E}$, $\mathbf{T}_A$ and $\mathbf{T}_B$ require estimation and the variance matrix is

$$\mathbf{Z}_a(\mathbf{T}_A \times \mathbf{A})\mathbf{Z}_a' + \mathbf{Z}_b(\mathbf{T}_B \times \mathbf{B})\mathbf{Z}_b' + \mathbf{E} \times \mathbf{I}.$$

With 2 components a transformation to simultaneously diagonalize the variance matrices is available, but not generally for more than 2 components. However, there is a transformation $\mathbf{S}(= \mathbf{Q}^{-1})$ such that $\mathbf{SES}' = \mathbf{I}$, $\mathbf{ST}_B\mathbf{S}' = \mathbf{T}_{BS}$ and $\mathbf{ST}_A\mathbf{S}' = \mathbf{T}_{AS}$ with $\mathbf{T}_{BS}$ a diagonal matrix. When $p = 2$, the $3 \times 3 = 9$ parameters in $\mathbf{E}$, $\mathbf{T}_A$ and $\mathbf{T}_B$ can therefore be written in terms of the 4 parameters in $\mathbf{S}$, 2 in $\mathbf{T}_{BS}$ and 3 in $\mathbf{T}_{AS}$. The calculation of the log likelihood is now based on a composite $p \times p^2$ matrix $\mathbf{Y}_c$. This matrix has p^2 variates formed from the direct product of the $p \times p$ identity matrix and $\mathbf{Y}$, the $n \times p$ matrix of observations with each column representing a trait. The log likelihood can be calculated using a formula similar to [2], and it can be seen that $\mathbf{Y}_c$ includes the variates used in *Models 1* and *2* and expands the matrix $\mathbf{Y}$ used by Meyer (1991) when estimating 2 components.

The log likelihood in this case is similar to [2] of the form:

$$-2 \log L = -2df \log|\mathbf{S}| + \log|\mathbf{G}_s| + \log|\mathbf{B}_s| + \log|\mathbf{C}| + \mathbf{s}'\mathbf{Y}_c'\mathbf{P}\mathbf{Y}_c\mathbf{s} \qquad [10]$$

with $\mathbf{G}_s = \mathbf{A} \times \mathbf{T}_{AS}\mathbf{B}_s = \mathbf{B} \times \mathbf{T}_{BS}$ and $|\mathbf{C}|$ found from [7] with $\mathbf{I}(1/\sigma_b^2)$ replaced by $\mathbf{B}_s^{-1}$. The $(1 \times p^2)$ vector $\mathbf{s}'$ is found by stacking the rows of $\mathbf{S}$ into a vector, *ie* $s_{(i-1)p+j} = S_{ij}$. The term $\mathbf{Y}_c'\mathbf{P}\mathbf{Y}_c = \mathbf{U}$ can be found from the residual sum of squares and cross-product residual matrix for the p^2 variates in $\mathbf{Y}_c$ after adjusting for all the effects.

Differentiation of [10] with respect to $\mathbf{S}$ shows that the estimates of $\mathbf{S}$ that maximize [8] for given values of $\mathbf{T}_{AS}$ and $\mathbf{T}_{BS}$ satisfy

$$df(S^{ij}) = (\mathbf{US})_{(i-1)p+j} \qquad [11]$$

with $S^{ij} = (S^{-1})_{ij}$.

The appendix shows that $\mathbf{S}$ can be found as the solution of an eigenvalue problem if $p = 2$. Hence if $p = 2$ maximization can be reduced from considering 9 parameters to a search over the 5 dimensional space of $\mathbf{T}_{AS}$ and $\mathbf{T}_{BS}$.

Meyer (1991) illustrated her methods with data from a selection experiment of Sharp *et al* (1984) and fitted a 3-component model to bivariate data. The likelihood was maximised over a 9 dimensional space and required 722 iterations to reach convergence. Using the same starting values and convergence criteria the 5 dimensional strategy outlined in this section reached convergence in 89 evaluations.

DISCUSSION

It has been shown, for a variety of models, how scaling and transformation can reduce the considerable effort in finding maximum likelihood estimates, especially for multivariate models. Another advantage is that the transformation can suggest more parsimonious models. For example, one could consider a constrained model of the form:

$$\mathbf{SES}' = \mathbf{I}, \mathbf{ST}_{BS}\mathbf{S}' = \mathbf{T}_B \text{ and } \mathbf{ST}_{AS}\mathbf{S}' = \mathbf{T}_A \text{ with } \mathbf{T}_{AS} \text{ and } \mathbf{T}_{BS} \text{ diagonal,}$$

that is fitting a model with underlying uncorrelated traits that are transformed using $\mathbf{Q} = \mathbf{S}^{-1}$ to form the p measured traits. This model has the advantage of having fewer parameters ($p(p + 2)$) and that the likelihood, for given $\mathbf{T}_{BS}$ and $\mathbf{T}_{AS}$, can be calculated in about $(1/p^2)$ of the time of the unconstrained model because the likelihood of each underlying trait can be calculated separately. For Meyer's example an underlying uncorrelated model converged in 50 iterations. The difference in $2 \log L$ was 0.94 suggesting that an underlying independent model would adequately fit the data. Lin and Smith (1990) have pointed out that by transforming to these approximately uncorrelated traits, simpler best linear unbiased predictors can be obtained. Villanueva *et al* (1993) have given examples where this procedure has very high efficiency. The strategy outlined gives a logical method for choosing the transformation $\mathbf{S}$.

In the multivariate case $\mathbf{S}$ can be found, for given $\mathbf{T}_{AS}$ and $\mathbf{T}_{BS}$, by derivative-free methods but the explicit solution for $p = 2$ has advantages. In fact for the numerical example above at the maximum likelihood estimates for $\mathbf{T}_{AS}$ and $\mathbf{T}_{BS}$, 2 of the solutions of [8] for $\mathbf{S}$ correspond to local maxima and 2 to saddlepoints. Explicit solutions for $\mathbf{S}$ if $p > 2$ were not obtained and the best computational strategy in terms of derivative-free methods or iterative use of [8] or solutions for $p = 2$ deserves further investigation.

The motivation has been to reduce the computation in derivative-free estimation procedure, but the idea of scaling and transformation carries over to other methods of estimation, for example, using first and/or second differentials of likelihoods.

Formulae for derivatives of the scaled parameters $\mathbf{T}_{AS}$, $\mathbf{T}_{BS}$, are easily derived if not easily calculated, for a given transformation matrix $\mathbf{S}$. The arguments in this paper show how to calculate derivatives for any $\mathbf{S}$. This allows for any $\mathbf{T}_{AS}$ and $\mathbf{T}_{BS}$, $\mathbf{S}$ to be found using the derivatives calculated at this optimal $\mathbf{S}$. The efficiency of this technique will depend on the structure of the data and the correlation of the parameters and deserves further investigation.

If, after fitting this underlying model, there is interest in getting some information on covariances between the underlying traits but full p trait evaluation is impractical, then use of the Thompson and Hill (1990) procedure to estimate covariance parameters from analysis of sums of approximately independent traits is an attractive option.

REFERENCES

Crump RE (1992) Quantitative genetic analysis of a commercial pig population undergoing selection. PhD Thesis, University of Edinburgh, UK

Graser HU, Smith SP, Tier B (1987) A derivative-free approach for estimating variance components in animal models by restricted maximum likelihood. *J Anim Sci* 64, 1362-1370

Henderson CR, Kempthorne O, Searle SR, Von Korsegk CM (1959) Estimation of environmental and genetic trends from records subject to culling. *Biometrics* 13, 192-218

Juga J, Thompson R (1992) A derivative-free algorithm to estimate bivariate (co)variance components using canonical transformations and estimated rotations. *Acta Agric Scand A* 42, 191-197

Lin CY, Smith SP (1990) Transformation of multitrait to unitrait mixed model analysis of data with multiple random effects. *J Dairy Sci* 73, 2494-2502

Meyer K (1985) Maximum likelihood estimation of variance components for a multivariate mixed model with equal design matrices. *Biometrics* 41, 153-165

Meyer K (1989) Restricted maximum likelihood to estimate variance components for animal models with several random effects using a derivative-free algorithm. *Genet Sel Evol* 21, 317-340

Meyer K (1991) Estimating variance and covariances for multivariate animal models by restricted maximum likelihood. *Genet Sel Evol* 23, 67-83

Meyer K, Thompson R (1984) Bias in variance and covariance component estimation due to selection on a correlated trait. *Z Tierz Zuchtungsbiol* 101, 33-50

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

Schaeffer LR, Wilton JW, Thompson R (1978) Simultaneous estimation of variance and covariance components from multitrait mixed model equations. *Biometrics* 34, 199-208

Searle SR (1982) *Matrix Algebra Useful for Statistics.* John Wiley and Sons, NY, USA

Sharp GL, Hill WG, Robertson A (1984) Effects of selection on growth, body composition, and food intake in mice. I Response in selected traits. *Genet Res* 43, 75-93

Taylor JF, Bean B, Marshall CE, Sullivan JJ (1985) Genetic and environmental components of semen production traits of artificial insemination Holstein bulls. *J Dairy Sci* 68, 2703-2722

Thompson R (1973) The estimation of variance and covariance components with an application when records are subject to culling. *Biometrics* 29, 527-550

Thompson R (1977) Estimation of quantitative genetic parameters. *In: Proc Int Conf Quant Genet* 639-657, Iowa State Press, Ames, IA, USA

Thompson R, Hill WG (1990) Univariate REML analyses for multivariate data with the animal model. *In: Proc Fourth World Congr Genet Appl Livest Prod* (WG Hill, R Thompson, JA Woolliams, eds) 13, 484-487

Villanueva B, Wray NR, Thompson R (1993) Prediction of asymptotic rates of response from selection on multiple traits using univariate and multivariate best linear unbiased predictors. *Anim Prod* 57, 1-13

APPENDIX

Solution of equation [11]

When $p = 2$ equation [11] can be written as

$$df \begin{pmatrix} S^{11} \\ S^{12} \\ S^{21} \\ S^{22} \end{pmatrix} = \mathbf{US}$$

or

$$df \begin{pmatrix} 0 & 0 & 0 & 1 \\ 0 & 0 & -1 & 0 \\ 0 & -1 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{pmatrix} \mathbf{S} = (S_{11}S_{22} - S_{12}S_{21})\mathbf{US}$$

or

$$\mathbf{FS} = (S_{11}S_{22} - S_{12}S_{21})\mathbf{US} \qquad [A1]$$

This equation is similar to equations relating the ith eigenvector x_i and the ith eigenvalue λ_i of $\mathbf{F}$ and $\mathbf{U}$, *ie* $\mathbf{F}x_i = \lambda_i \mathbf{U}x_i$.

For a given eigenvector x_i with eigenvalue λ_i a scaled vector $k_i x_i$ will satisfy [A1] if

$$k_i \mathbf{F}x_i = k_i^3 (x_{i1}x_{i4} - x_{i2}x_{i3})\mathbf{U}x_i$$

so $(x_{i1}x_{i4} - x_{i2}x_{i3})k_i^2 = \lambda_i$, and so a scaled vector of x_i can be found to satisfy [A1] as a function of x_i and λ_i.

Hence 4 vectors $\mathbf{S}_i$ can be calculated to satisfy [11]. Each vector can be substituted into [11] in order to find $\mathbf{S}$ to maximize [10]. This result can be thought of as a generalization of result [3]–[5] for *Model 1* and the result of Juga and Thompson (1992) for 2 components. In the first case $\mathbf{U}$ is of the form:

$$\mathbf{U} = \begin{pmatrix} U_{11} & 0 & 0 & U_{12} \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ U_{21} & 0 & 0 & U_{22} \end{pmatrix}$$

and in the second:

$$\mathbf{U} = \begin{pmatrix} U_{11} & U_{12} & 0 & 0 \\ U_{21} & U_{22} & 0 & 0 \\ 0 & 0 & U_{33} & U_{34} \\ 0 & 0 & U_{43} & U_{44} \end{pmatrix}$$