J. R. Statist. Soc. B (1994)
56, *No.* 1, *pp.* 61-69

Estimation in Generalized Mixed Models

By C. A. McGILCHRIST†

Fred Hutchinson Cancer Research Center, Seattle, USA

[Received November 1991. Final revision July 1992]

SUMMARY

Regression models containing fixed and random effects may have a response variable which is not normally distributed. The generalized mixed model includes both discrete and continuous response variables and is developed here for problems in which the regression variables enter linearly into the model. Best linear unbiased predictor methods are extended to maximum likelihood and residual maximum likelihood estimation procedures. Applications in modelling discrete response variables and in survival analysis are discussed.

Keywords: BEST LINEAR UNBIASED PREDICTION; GENERALIZED MIXED MODEL; RESIDUAL MAXIMUM LIKELIHOOD

1. INTRODUCTION

The advent of generalized linear modelling techniques (McCullagh and Nelder, 1989) has greatly extended the domain of regression analysis into areas where the dependent variable may be discrete. A natural extension of this work is to add to the linear combination of regression variables a linear combination of random components such as random block effects which would account for correlations between observations on the same or associated experimental units. Other uses of such models are to account for overdispersion in binomial and Poisson regression models as in Williams (1982) and Breslow (1984).

There are many publications on inference for such models when the dependent variable is normally distributed and they will not be referenced here except for those papers considered directly relevant to the current development. The Bayesian approach to binary responses is exemplified in Stiratelli *et al.* (1984). In the marginal likelihood approach, the marginal distribution is formed by integrating out with respect to the random effect variables as in Anderson and Aitkin (1985). Often this operation can be done only by using numerical integration techniques, which can be prohibitive for some problems. Moreover, models chosen to have simple properties when conditioned on the random components often lose those properties in the integrated model. Other methods such as in Gilmour *et al.* (1985) have elements in common with the EM analysis of Dempster *et al.* (1977).

The generalized estimating equation approach of Zeger and Liang (1986), Liang and Zeger (1986) and Zeger *et al.* (1988) aims primarily to estimate fixed effects and does not model or estimate the random component terms. However, such estimates are often useful either for prediction of future responses or in model diagnostics.

The methods considered here have been developed from the best linear unbiased

†*Address for correspondence*: Department of Statistics, University of New South Wales, PO Box 1, Kensington, NSW 2033, Australia.

0035–9246/94/56061

prediction (BLUP) approach of Henderson (1963, 1973, 1975), which is reviewed in Robinson (1991). The approach is similar in principle to penalized likelihood approaches and in basic aims has elements in common with Breslow and Clayton (1991). The idea is to use BLUP to obtain approximate residual maximum likelihood (REML) estimates (Patterson and Thompson, 1971). This link between BLUP and REML is outlined in Harville (1977) for the normal case and full details are given in Thompson (1980), Fellner (1986, 1987) and Speed (1991). The current approach is very similar to that of Schall (1991) but uses a different argument for the approximate linearization of the model. By doing so the application of the method is broadened to a wider class of models than those usually included in the generalized linear modelling approach.

2. BEST LINEAR UNBIASED PREDICTION FOR NORMAL ERROR MODELS

The BLUP approach for normal error models is often expressed in terms of the linear model

$$y = \eta + e,$$
$$\eta = X\beta + Zu$$

where y is an observed response vector and e is a random error vector distributed as $N(0, \sigma^2 D)$, where D is a known matrix. The mean response η depends on a fixed component $X\beta$, where X is an $n \times v$ matrix of known constants and β is an unknown regression parameter vector. The component Zu may be partitioned into

$$Z = (Z_1, Z_2, \ldots, Z_k),$$
$$u' = (u_1', u_2', \ldots, u_k')$$

where each u_j has v_j components, is distributed as $N(0, \sigma_j^2 A_j)$ and is independent of the other u components. The u are the random components of the mixed model and the A_j are taken to be known matrices of constants. Let $\sigma_j^2 = \sigma^2\theta_j$ and

$$A = \begin{pmatrix} \theta_1 A_1 & 0 & \ldots & 0 \\ 0 & \theta_2 A_2 & \ldots & 0 \\ . & . & \ldots & . \\ 0 & 0 & \ldots & \theta_k A_k \end{pmatrix}.$$

The BLUP procedure consists of maximizing the sum of two component log-likelihoods. Let

(a) l_1 be the log-likelihood for y given u conditionally fixed,
(b) l_2 be the 'log-likelihood' for u and
(c) $l = l_1 + l_2$.

Thus l represents the 'log-likelihood' based on the joint distribution of y and u. The log-likelihood is not a likelihood in the conventional sense because it is based on the non-observable u. The BLUP procedure chooses estimates or predictors of β, u, σ^2 and parameters of A to maximize l. Specifically, for normal error models, the BLUP estimators $\tilde{\beta}$ and $\tilde{u}$ are the values which make the derivatives of l with respect to β and u equal to 0. Thus

$$\begin{pmatrix} X'D^{-1}X & X'D^{-1}Z \\ Z'D^{-1}X & Z'D^{-1}Z+A^{-1} \end{pmatrix} \begin{pmatrix} \tilde{\beta} \\ \tilde{u} \end{pmatrix} = \begin{pmatrix} X'D^{-1}y \\ Z'D^{-1}y \end{pmatrix}.$$

BLUP estimators of the variance components are

$$\tilde{\sigma}_j^2 = \tilde{u}_j' A_j^{-1} \tilde{u}_j / v_j,$$
$$\tilde{\sigma}^2 = (y - X\tilde{\beta} - Z\tilde{u})' D^{-1} (y - X\tilde{\beta} - Z\tilde{u})/n.$$

3. FROM BEST LINEAR UNBIASED PREDICTION TO MAXIMUM LIKELIHOOD AND RESIDUAL MAXIMUM LIKELIHOOD FOR NORMAL ERROR MODELS

Both ML and REML estimators of β are the same as the BLUP estimator for given θ. However, estimators of the variance components are different but may be easily derived from the BLUP estimators as shown by Harville (1977), Thompson (1980), Fellner (1986, 1987) and Speed (1991). Adapting expressions slightly and letting

$$(Z'D^{-1}Z + A^{-1})^{-1} = T^* = (T^*_{jj'})$$

be a partition of T^* conformally with the partition of u then ML estimators are

$$\begin{aligned} \hat{\sigma}^2_{\mathrm{ML}} &= y'D^{-1}(y - X\tilde{\beta} - Z\tilde{u})/n, \\ \hat{\sigma}^2_{j(\mathrm{ML})} &= \tilde{u}_j A_j^{-1} \tilde{u}_j/(v_j - r_j^*) \end{aligned} \tag{3.1}$$

where $r_j^* = \theta_j^{-1} \operatorname{tr}(A_j^{-1} T^*_{jj})$. The information matrix for the ML estimators of β, σ^2, θ_1, $\theta_2, \ldots, \theta_k$ is the symmetric matrix

$$\mathscr{I}_{\mathrm{ML}} = \begin{pmatrix} \sigma^{-2}X'\Sigma^{-1}X & 0 & 0 \\ . & n/2\sigma^4 & (v_j - r_j^*)/2\sigma_j^2 \\ . & . & \frac{1}{2}\{\theta_i^{-2}(v_i - 2r_i^*)\delta_{ij} + \theta_i^{-2}\theta_j^{-2} r^*_{ij}\} \end{pmatrix}$$

where $r^*_{ij} = \operatorname{tr}(A_j^{-1} T^*_{ij} A_j^{-1} T^*_{ji})$ and δ_{ij} is the usual Kronecker delta.

The REML estimators of the variance components are developed similarly. Let

$$\begin{pmatrix} X'D^{-1}X & X'D^{-1}Z \\ Z'D^{-1}X & Z'D^{-1}Z+A^{-1} \end{pmatrix}^{-1} = \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix}$$

and $A_{22} = (T_{jj'})$ be a partition of A_{22} conformally to the partition of u; then

$$\begin{aligned} \hat{\sigma}^2_{\mathrm{REML}} &= y'D^{-1}(y - X\tilde{\beta} - Z\tilde{u})/(n - v), \\ \hat{\sigma}^2_{j(\mathrm{REML})} &= \tilde{u}_j' A_j^{-1} \tilde{u}_j/(v_j - r_j) \end{aligned} \tag{3.2}$$

where $r_j = \theta_j^{-1} \operatorname{tr}(A_j^{-1} T_{jj})$. The REML information matrix for the estimators of σ^2, θ_1, $\theta_2, \ldots, \theta_k$ is the symmetric matrix

$$\mathscr{I}_{\mathrm{REML}} = \begin{pmatrix} (n-v)/2\sigma^4 & (v_j - r_j)/2\sigma_j^2 \\ . & \frac{1}{2}\{\theta_i^{-2}(v_i - 2r_i)\delta_{ij} + \theta_i^{-2}\theta_j^{-2} r_{ij}\} \end{pmatrix}$$

where $r_{ij} = \operatorname{tr}(A_i^{-1} T_{ij} A_j^{-1} T_{ji})$.

4. GENERALIZED LINEAR MIXED MODELS

For a generalized linear mixed model the distribution of the response vector y depends on a vector quantity η which is related to vector regression variables through the equation

$$\eta = X\beta + Zu$$

as in the previous section. Let $f(y; \beta|u)$ be the probability (density) function of y conditional on fixed u. The log-likelihood of the observation vector $y = (Y_1, Y_2, \ldots, Y_n)'$ conditional on fixed u is

$$l_1 = \ln f(y; \beta|u).$$

The likelihood of the random component vector u is

$$l_2 = \text{constant} - \frac{1}{2}\sum_{j=1}^{k} \{v_j \ln(2\pi\sigma_j^2) + \sigma_j^{-2} u_j' A_j^{-1} u_j\}$$

and the joint log-likelihood of y and u is $l = l_1 + l_2$. The derivatives of l are

$$\partial l/\partial\beta = \partial l_1/\partial\beta,$$

$$\partial l/\partial u_j = \partial l_1/\partial u_j - \sigma_j^{-2} A_j^{-1} u_j, \qquad j = 1, 2, \ldots, k.$$

The second-order derivatives of l which involve at least one β are the same as the second-order derivatives of l_1 while

$$\partial^2 l/\partial u_j \partial u_j' = \partial^2 l_1/\partial u_j \partial u_j' - \sigma_j^{-2} A_j^{-1},$$

$$\partial^2 l/\partial u_j \partial u_l' = \partial^2 l_1/\partial u_j \partial u_l'.$$

The Newton–Raphson iterative procedure for estimating β and u is

$$\begin{pmatrix} \tilde\beta \\ \tilde u \end{pmatrix} = \begin{pmatrix} \beta_0 \\ u_0 \end{pmatrix} + V^{-1}\begin{pmatrix} \partial l_1/\partial\beta_0 \\ \partial l_1/\partial u_0 \end{pmatrix} - V^{-1}\begin{pmatrix} 0 \\ \sigma^{-2}A^{-1}u_0 \end{pmatrix}. \tag{4.1}$$

The matrix V is the matrix of second-order derivatives

$$V = \begin{pmatrix} -\partial^2 l_1/\partial\beta\,\partial\beta' & -\partial^2 l_1/\partial\beta\,\partial u' \\ -\partial^2 l_1/\partial u\,\partial\beta' & -\partial^2 l_1/\partial u\,\partial u' + \sigma^{-2}A^{-1} \end{pmatrix}.$$

If V is replaced by $\mathscr{E}(V)$ then the iterative procedure becomes the method of scoring.

The likelihood surface l_1 is a function of β and u, and in many problems the number of components k of u increases as the number of components n of y increases. For this reason l_1 will often have ridges in the u-space with little variation of l_1 along the ridge. In contrast l_2 is flat only in the β-directions. A consequence is that the surface l usually has a well-defined maximum and if the surface is quadratic in the region of the maximum then

$$l \simeq \text{constant} + \frac{1}{2}\begin{pmatrix} \beta - \tilde\beta \\ u - \tilde u \end{pmatrix}' V \begin{pmatrix} \beta - \tilde\beta \\ u - \tilde u \end{pmatrix}.$$

In that case $\tilde\beta$ and $\tilde u$ have approximately a joint normal distribution with means β and u and variance matrix V^-.

An alternative formulation of the problem is given in McGilchrist and Aisbett (1991a) in which the component l_1 of the BLUP procedure is replaced by the log-likelihood of $\hat\beta$ and $\hat u$ as given by its approximate asymptotic distribution, i.e. normal with mean β and u and variance matrix inverse given by the information matrix for $\hat\beta$ and $\hat u$. Corresponding to the ridges in the surface of l_1, the asymptotic variance matrix of $\hat\beta$ and $\hat u$ is ill conditioned. However, the resulting BLUP log-likelihood obtained by using such approximate asymptotics is identical with the quadratic expression given above. The estimating equations for β and u are exactly those given by the method of scoring. Hence, provided that l is approximately quadratic in β and u then we may consider the BLUP estimation as having been derived from the very approximate asymptotic distribution of $\hat{\beta}$ and $\hat{u}$.

If $\mathscr{I}(\beta, u)$ is the information matrix for β and u derived from l_1 then

$$\mathscr{I}(\beta, u) = \begin{pmatrix} X' \\ Z' \end{pmatrix} B(X, Z),$$

where $B = -\mathscr{E}(\partial^2 l_1/\partial\eta\,\partial\eta')$. Replacing l_1 by l_1^*, the log-likelihood based on the approximate asymptotic distribution of $\hat{\beta}$ and $\hat{u}$ gives

(a) $l_1^* = \text{constant} - \dfrac{1}{2}\begin{pmatrix} \hat{\beta}-\beta \\ \hat{u}-u \end{pmatrix} \begin{pmatrix} X' \\ Z' \end{pmatrix} B(X, Z) \begin{pmatrix} \hat{\beta}-\beta \\ \hat{u}-u \end{pmatrix}$

$= \text{constant} - \tfrac{1}{2}(y^* - X\beta - Zu)'B(y^* - X\beta - Zu)$

where $y^* = X\hat{\beta} + Z\hat{u}$,

(b) $l_2 = \text{constant} - \dfrac{1}{2}\sum_{j=1}^{k} \{v_j \ln(2\pi\sigma_j^2) + \sigma_j^{-2} u_j' A_j^{-1} u_j\}$,

(c) $l^* = l_1^* + l_2$

as the three components of the BLUP procedure.

The formulation of the problem is now exactly as described for normal theory models with y^* replacing y, B in place of D^{-1} and $\sigma^2 = 1$ implying $\theta_j = \sigma_j^2$. It follows that BLUP estimators $\tilde{\beta}$ and $\tilde{u}$ may be used to find ML and REML estimators.

The linearization argument presented here is analogous to the argument given by Schall (1991) when applied to generalized linear models. However, the current argument aims at a much broader class of problem. The method applies when the distribution of the response variable conditional on fixed random components can be expressed in terms of the quantity η which is a linear combination of fixed and random components. The method can be extended to non-linear models but only to the extent that a linearized version of the model is appropriate. The likelihood function for the response variable conditional on fixed η need not be obtained from the exponential family of distributions. Examples are given in Section 6.

5. MODELS WITH SCALE PARAMETER

It is possible to generalize the procedure by allowing the distribution of y conditional on u to depend on other parameters contained in a vector ϕ. In some cases ϕ may be a single parameter such as σ^2 in normal error models. The theory requires some minor adjustments to cover this case.

The expression for l_1^* is modified by incorporating the scale parameter as

$$l_1^* = \text{constant} - \frac{1}{2\phi}\begin{pmatrix} \hat{\beta}-\beta \\ \hat{u}-u \end{pmatrix}' \begin{pmatrix} X' \\ Z' \end{pmatrix} B(X, Z) \begin{pmatrix} \hat{\beta}-\beta \\ \hat{u}-u \end{pmatrix}.$$

The formulation then parallels the normal theory case with B in place of D^{-1}, $\sigma^2 = \phi$ and $\sigma_j^2 = \phi\theta_j$.

6. APPLICATIONS

An advantage of this procedure is that it applies to a wide class of problems. To change the method from one problem to another requires only the reprogramming of the first and second derivatives of l_1. BLUP estimates of β and u are obtained from equation (4.1) where

$$V = \begin{pmatrix} X' \\ Z' \end{pmatrix} B(X, Z) + \begin{pmatrix} 0 & 0 \\ 0 & \sigma^{-2}A^{-1} \end{pmatrix},$$

$B = -\partial^2 l_1/\partial\eta\,\partial\eta'$ (or expectation). Both ML and REML estimates of β are the same as the BLUP estimate for given θ. ML and BLUP estimates of the variance components are obtained from equations (3.1) and (3.2) respectively. Standard errors are obtained from the associated information matrices.

6.1. *Generalized Linear Models*

The usual generalized linear model takes the probability (density) function of observed Y as

$$f(y) = \exp\left\{\frac{y\alpha - b(\alpha)}{a(\phi)} + c(y, \phi)\right\}.$$

It is well known that

$$\mathscr{E}(Y) = b'(\alpha) = \mu,$$
$$\operatorname{var}(Y) = a(\phi)\, b''(\alpha)$$

and the mean is related to the regression vector x by the inverse link function

$$\mu = h(\eta),$$
$$\eta = x'\beta + z'u.$$

The log-likelihood is given by $l_1 = \Sigma \ln f(y_i)$ from which

$$\partial l_1/\partial\eta = [\operatorname{diag}\{c_i h'(\eta_i)\}]^{-1}(y - \mu),$$
$$B = \operatorname{diag}(c_i^{-1})$$

where $c_i = \operatorname{var}(Y_i)\, h'(\eta_i)^{-2}$. Application of the BLUP method to generalized linear models, but not its extension to REML, is given in McGilchrist and Aisbett (1991a). The full application to ML and REML estimation is given by Schall (1991). An illustration using simulated data is given in the next section, where the results of the ML and REML extension are compared with the BLUP.

6.2. *Frailty Models in Survival Analysis*

Multivariate failure time data have often proved difficult to analyse. When models for such data are based on a proportional hazards function including a frailty component, the marginal failure time distribution formed by integrating out the random frailty component loses the simple properties of the original hazard function formulation. No such integration is required here and the method preserves all the simple properties, particularly the cancellation of the hazard 'shape' function in the partial likelihood approach.

Suppose that patient i has several recurrence times which may result in failure or censoring with the jth such time denoted by T_{ij}. The associated risk variable vector is denoted by x_{ij} and the frailty for patient i is denoted by U_i which are considered to be independent selections from $N(0, \sigma_1^2)$. The proportional hazards model is

$$h(t_{ij};\ x_{ij}, U_i) = \lambda(t_{ij})\, g(\eta_{ij}),$$
$$\eta_{ij} = x_{ij}'\beta + U_i.$$

Let the failure or censoring times be ordered and t_n be the nth such time with

$$D_{ijn} = \begin{cases} 1, & \text{if patient } i \text{ with } j\text{th recurrence time fails at } t_n, \\ 0, & \text{otherwise;} \end{cases}$$

then the partial log-likelihood conditional on fixed U-values gives

$$l_1 = \sum_n \sum_i \sum_j D_{ijn} \ln p_{ijn}$$

where p_{ijn} is the probability that patient i (failure time j) is the event that occurs at t_n. Note that

$$p_{ijn} = g(\eta_{ij}) \Big/ \sum_{t_{k1} \geqslant t_{ij}} g(\eta_{k1})$$

and the base-line hazard function $\lambda(t)$ cancels from the expression. Following differentiation of l_1 with respect to the parameters in the appropriate way, the general procedure for ML and REML estimation of the parameters proceeds as in the general formulation. A partial implementation of the BLUP procedure is given in McGilchrist and Aisbett (1991b).

6.3. *Threshold Models*

If a response variable Y is discrete with a finite number of classes or is an ordered response variable, then the categories may be labelled with the integers 0, 1, . . ., b and the distribution specified by the cumulative distribution function $F(y, \eta)$ where $\eta = x'\beta + z'u$. The interpretation of η and its component terms is the same as in previous illustrations. For several models such as

(a) the threshold model, $F(y, \eta) = \Phi(\alpha_y - \eta)$,
(b) the proportional odds model, $F(y, \eta) = \exp(\alpha_y - \eta)/\{1 + \exp(\alpha_y - \eta)\}$, and
(c) the proportional hazards model, $F(y, \eta) = 1 - \exp\{-\exp(\alpha_y - \eta)\}$,

the model is of the form $G(\alpha_y - \eta)$ with the α_y threshold parameters.

If r indexes the whole set of observations and, in general, y is the observation at r while x and z are the regression vectors then

$$l_1 = \sum_r \ln\{G(\alpha_y - \eta) - G(\alpha_{y-1} - \eta)\}$$

and the estimation proceeds along standard lines. Here the α-parameters are fixed and are very similar in nature to the β-parameters. A description of the BLUP application is given in Zhaorong *et al.* (1992).

7. SIMULATION OF BINOMIAL-LOGIT MODEL

The intention of this section is to demonstrate the improvement obtained by the REML extension over the ML and BLUP procedures for a particular problem rather than to carry out extensive simulations. The particular problem taken is that reported in McGilchrist and Aisbett (1991a) in which 30 response variables Y_i are generated having distribution $B(6, \pi_i)$ and π_i given by

$$\pi_i = \exp \eta_i/(1+\exp \eta_i), \qquad \eta' = (\eta_1, \eta_2, \ldots, \eta_{30}) = X\beta + Zu$$

$$X = \begin{pmatrix} 1 & 1 & \ldots & 1 \\ -14 & -13 & & 15 \end{pmatrix}, \qquad Z = \begin{pmatrix} 1 & 1 & & & & & \\ & & 1 & 1 & & & \\ & & & & \ldots & & \\ & & & & & 1 & 1 \end{pmatrix},$$

$$u' = (U_1, U_2, \ldots, U_{15})$$

where the U_i are independently and identically distributed $N(0, \sigma_1^2)$. This corresponds to binomial observations in blocks of two, corresponding to six trials and probabilities given by a logit link function related to two fixed components.

Results of the simulation are given in Table 1 where BLUP, ML and REML estimates are obtained for each simulated data set. Parameters β_0 and β_1 are estimated in the same way by each method, differing only as a result of the different estimates of σ_1^2 produced. The negative bias of the BLUP estimator of σ_1^2 is reduced in the ML estimator but is much smaller again in the REML estimator. This bias is not significant for the REML estimator for the first set of simulations when $\sigma_1^2 = 1$. However, when the variance of the random components is increased to 2 in the second

TABLE 1

Results of BLUP, ML and REML estimation in 100 simulated data sets, each set consisting of 30 binomially distributed observations

Parameter	*True value*	*Average bias†*			*Average of SE‡ of estimates*			*SE of estimates § over simulations*		
		BLUP	*ML*	*REML*	*BLUP*	*ML*	*REML*	*BLUP*	*ML*	*REML*
Simulation 1										
β_0	0.2	0.036 (0.029)	0.049 (0.031)	0.053 (0.031)	0.178	0.276	0.299	0.290	0.309	0.314
β_1	0.1	−0.006 (0.004)	0.000 (0.004)	0.001 (0.004)	0.022	0.033	0.035	0.036	0.037	0.038
σ_1^2	1	−0.931 (0.021)	−0.295 (0.044)	−0.093 (0.054)	—	0.412	0.532	0.207	0.439	0.540
Simulation 2										
β_0	0.2	−0.098 (0.032)	−0.091 (0.034)	−0.089 (0.035)	0.228	0.348	0.381	0.322	0.343	0.350
β_1	0.1	−0.011 (0.004)	−0.003 (0.005)	−0.001 (0.005)	0.027	0.041	0.045	0.044	0.048	0.049
σ_1^2	2	−1.607 (0.051)	−0.664 (0.064)	−0.319 (0.080)	—	0.673	0.871	0.512	0.640	0.795

†Standard errors (SEs) are given in parentheses.
‡Each simulation gives an SE of the estimate. The tabular value is the average of these SEs.
§The tabular value is the SE of the 100 simulated estimates.

set of simulations, the REML estimator has a small but significant negative bias. The bias is much less than for the BLUP and ML estimators.

Using the information matrix, the standard errors of estimation are obtained for each simulation. The average of such values may be compared with the standard error of the simulations. It is clear that good agreement is obtained for both ML and REML approaches.

REFERENCES

Anderson, D. A. and Aitkin, M. (1985) Variance component models with binary response: interviewer variability. *J. R. Statist. Soc.* B, **47**, 203–210.

Breslow, N. E. (1984) Extra-Poisson variation in log-linear models. *Appl. Statist.*, **33**, 38–44.

Breslow, N. E. and Clayton, D. G. (1991) Approximate inference in generalised linear mixed models. *Technical Report 106*. Biostatistics Department, University of Washington, Seattle.

Dempster, A. P., Laird, N. M. and Rubin, D. B. (1977) Maximum likelihood from incomplete data via the EM algorithm. *J. R. Statist. Soc.* B, **39**, 1–22.

Fellner, W. H. (1986) Robust estimation of variance components. *Technometrics*, **28**, 51–60.

——— (1987) Sparse matrices, and the estimation of variance components by likelihood methods. *Communs Statist. Simuln*, **16**, 439–463.

Gilmour, A. R., Anderson, R. D. and Rae, A. L. (1985) The analysis of binomial data by a generalised linear mixed model. *Biometrika*, **72**, 593–599.

Harville, D. A. (1977) Maximum likelihood approaches to variance component estimation and to related problems. *J. Am. Statist. Ass.*, **72**, 320–340.

Henderson, C. R. (1963) Selection index and expected genetic advance. In *Statistical Genetics and Plant Breeding* (eds W. D. Hanson and H. F. Robinson), pp. 141–163. Washington DC: National Academy of Science.

——— (1973) Sire evaluation and genetic trends. In *Proc. Animal Breeding and Genetics Symp. in Honor of Dr Jay Lush*, pp. 10–41. Champaign: American Society of Animal Science.

——— (1975) Best linear unbiased estimation and prediction under a selection model. *Biometrics*, **31**, 423–447.

Liang, K. Y. and Zeger, S. L. (1986) Longitudinal data analysis using generalised linear models. *Biometrika*, **73**, 13–22.

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

McGilchrist, C. A. and Aisbett, C. W. (1991a) Restricted BLUP for mixed linear models. *Biometr. J.*, **33**, 131–141.

——— (1991b) Regression with frailty in survival analysis. *Biometrics*, **47**, 461–466.

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

Robinson, G. K. (1991) The estimation of random effects. *Statist. Sci.*, **6**, 15–51.

Schall, R. (1991) Estimation in generalised linear models with random effects. *Biometrika*, **78**, 719–727.

Speed, T. (1991) Comment on Robinson: Estimation of random effects. *Statist. Sci.*, **6**, 42–44.

Stiratelli, R., Laird, N. and Ware, J. (1984) Random effects models for serial observations with binary responses. *Biometrics*, **40**, 961–971.

Thompson, R. (1980) Maximum likelihood estimation of variance components. *Math. Operforsch. Statist. Ser. Statist.*, **11**, 545–561.

Williams, D. A. (1982) Extra-binomial variation in logistic linear models. *Appl. Statist.*, **31**, 144–148.

Zeger, S. L. and Liang, K. Y. (1986) Longitudinal data analysis for discrete and continuous outcomes. *Biometrics*, **42**, 121–130.

Zeger, S. L., Liang, K. Y. and Albert, P. S. (1988) Models for longitudinal data: a generalised estimating equation approach. *Biometrics*, **44**, 1049–1060.

Zhaorong, J., McGilchrist, C. A. and Jorgensen, M. A. (1992) Mixed model discrete regression. *Biometr. J.*, to be published.