Comment

Terry Speed

Geoff Robinson is to be congratulated for writing this paper. It is lucidly written, it bridges a number of gulfs that have developed in our subject, and it is provocative. That he wrote it is clearly a Good Thing! I welcome the opportunity to say this and to make a few remarks that he might have made. I believe that these remarks will strengthen his already strong case for a much more explicit recognition of the role of BLUPs in our subject.

1. THE BAYESIAN DERIVATION

In Section 4.2 Robinson describes a Bayesian derivation, stating that the posterior mode is given by the BLUP estimates when β is regarded "as a parameter with a uniform, improper prior distribution and u as a parameter which has a prior distribution which has mean zero and variance $G\sigma^2$, independent of β." All this is certainly true, but it may be helpful to add that if β is given a proper prior (normal) distribution with mean zero and variance $B\sigma^2$, say, with u as before, then all of the results one could possibly want (posterior means, posterior variances, etc.) can be derived straightforwardly by the standard Bayesian formulae. Then all one has to do to derive the corresponding BLUP formulae is let $B^{-1} \to 0$. An identity which I have found useful, perhaps even indispensible, for carrying out this last step, is discussed in de Hoog, Speed and Williams (1990). Note that the approach just described is essentially that adopted in Dempster, Rubin and Tsutakawa (1981).

2. FORMULAE FOR $\hat{u}$

The only actual formulae given in the paper for $\hat{u}$ in the general case is the rather complicated one in Section 4.3. This is a pity, because there is an obvious "plug-in" expression, namely

$$(1)\qquad \hat{u} = GZ^T V^{-1}(y - X\hat{\beta}),$$

where $V = ZGZ^T + R$. This may be viewed as the result of regressing u on y, with the mean $X\beta$ of y replaced by its obvious linear estimator.

A variant of (1) is

$$(1')\qquad \hat{u} = (Z^T R^{-1} Z + G^{-1})^{-1} Z^T R^{-1}(y - X\hat{\beta}).$$

Terry Speed is Professor, Department of Statistics, University of California, Berkeley, California 94720.

The simpler formulae (5.3) and (5.4) arising when there are no fixed effects also have more general analogues, namely

$$(2)\qquad (Z^T A Z + G^{-1})\hat{u} = Z^T A y,$$

where $A = R^{-1}(I - S)$, $S = P^R_{\mathscr{R}(X)}$ being the projector onto $\mathscr{R}(X)$ orthogonal with respect to $\langle a, b\rangle = a^T R^{-1} b$, and for the variance-covariance matrix of $\hat{u}$:

$$(3)\qquad \left\{G^{-1} - \left(Z^T A Z + G^{-1}\right)^{-1}\right\}\sigma^2.$$

These expressions can be derived readily using the Bayesian approach outlined in (1) above, together with the matrix identity already referred to. I note in passing that Robinson's formulae (5.4) is in fact the variance-covariance matrix of $\hat{u} - u$, not, as stated, of $\hat{u}$.

3. SOLVING THE BLUP EQUATIONS

Perhaps in order to avoid messy algebra, Robinson has said little about the actual solution of the BLUP equations. I know that he has worked on this problem with some enormous data sets, and so I am hesitant to comment here. However, it does seem worthwhile to make one easy point, in order to connect this topic with another, closely related one. The obvious rearrangement of the first equation in (1.2),

$$(4)\qquad X^T R^{-1} X\hat{\beta} = X^T R^{-1}(y - Z\hat{u}),$$

can be combined with either (1) or (1′) above, to form the basis of an iterative solution of the BLUP equations, provided, of course, that the separate problems are readily solved. Just such a strategy is recommended more generally in Green (1985) in the context of smoothing, a topic to which I shall return.

It is also worth pointing out that (1′) or (2) is to be preferred when G^{-1} has simple structure, whereas if G is simple and V is readily inverted, (1) is more useful. In many animal breeding problems it is G^{-1} which has the simpler structure, as it also does in the Kalman filter case.

4. REML AND BLUP

In Section 5.4 Robinson states that "REML is the method of estimating variance components that seems to have the best credentials from a Classical