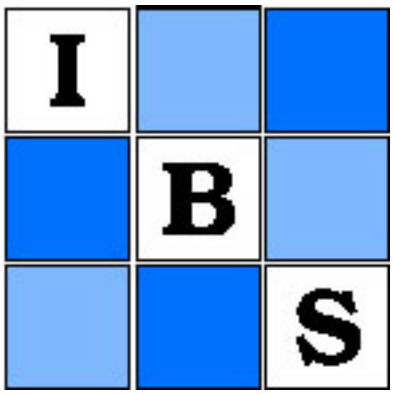

BIOMETRICS 53, 983–997
September 1997

Small Sample Inference for Fixed Effects from Restricted Maximum Likelihood

Michael G. Kenward

Institute of Mathematics and Statistics, The University of Kent at Canterbury, Canterbury, Kent CT2 7NF, U.K.

and

James H. Roger

Live Data Process, Chinnor Hill Lodge, Chinnor, Oxfordshire OX9 4BQ, U.K.

SUMMARY

Restricted maximum likelihood (REML) is now well established as a method for estimating the parameters of the general Gaussian linear model with a structured covariance matrix, in particular for mixed linear models. Conventionally, estimates of precision and inference for fixed effects are based on their asymptotic distribution, which is known to be inadequate for some small-sample problems. In this paper, we present a scaled Wald statistic, together with an F approximation to its sampling distribution, that is shown to perform well in a range of small sample settings. The statistic uses an adjusted estimator of the covariance matrix that has reduced small sample bias. This approach has the advantage that it reproduces both the statistics and F distributions in those settings where the latter is exact, namely for Hotelling T^2 type statistics and for analysis of variance F-ratios. The performance of the modified statistics is assessed through simulation studies of four different REML analyses and the methods are illustrated using three examples.

1. Introduction

Consider the general Gaussian linear model for the n observations $\mathbf{Y}(n \times 1)$,

$$\mathbf{Y} \sim \mathrm{N}(\mathbf{X}\boldsymbol{\beta}; \boldsymbol{\Sigma}),$$

where $\mathbf{X}$ ($n \times p$, rank p) is a matrix of known covariates, $\boldsymbol{\beta}$ ($p \times 1$) is a vector of unknown parameters and $\boldsymbol{\Sigma}$ is an unknown ($n \times n$) variance–covariance matrix. We are principally concerned in this paper with situations in which the data consist of a collection of independent sets, that is, for which $\boldsymbol{\Sigma}$ is block-diagonal. Situations where such a model may be appropriate include nested (or hierarchical), blocked, crossover, and repeated measurements experiments, and $\boldsymbol{\Sigma}$ is assumed to be structured accordingly, for example, to reflect the time series nature of repeated measurements. The methods are not restricted to this setting, however, and might be considered for other covariance structures such as those arising from spatial models. In applications of the Gaussian linear model, either or both of $\boldsymbol{\beta}$ and $\boldsymbol{\Sigma}$ may be of central interest. Here we assume that $\boldsymbol{\beta}$ is the focus of attention and $\boldsymbol{\Sigma}$ is a nuisance to be accommodated in the analysis.

Restricted (or residual) maximum likelihood (REML) has become well established for the estimation of the parameters of $\boldsymbol{\Sigma}$ (Patterson and Thompson, 1971; Harville, 1977; Thompson, 1980; Robinson, 1991) and the corresponding estimator of $\boldsymbol{\beta}$, $\hat{\boldsymbol{\beta}}$ say, has a generalized least squares form that uses the REML estimate of $\boldsymbol{\Sigma}$. Conventionally, the precision of $\hat{\boldsymbol{\beta}}$ is obtained from an estimate of the variance–covariance matrix of its asymptotic distribution. However, this takes no account of the variability in the estimate of $\boldsymbol{\Sigma}$, which, for certain combinations of covariance structure,

Key words: Alpha design; Ante-dependence; Crossover trial; Mixed models; Residual maximum likelihood; Small sample approximation.

design, and sample size, can have a significant impact on the precision of $\hat{\boldsymbol{\beta}}$. In such situations, the conventional asymptotic-based measure of precision can seriously overestimate the true precision. Further, Wald-type test procedures and corresponding confidence intervals that are based on asymptotic chi-squared approximations also ignore the variability in the estimate of $\boldsymbol{\Sigma}$. Our purpose in this paper is to provide more appropriate procedures for small sample inference based on Wald type statistics.

The construction of Wald-type pivots will be approached through an adjusted estimator of the covariance matrix of the fixed effects. Before pursuing this, we note that it can be argued that, in considering the behavior of pivotal quantities for the purposes of small sample inference, the bias in the estimated covariance matrix of the fixed effects is not of direct relevance. The covariance matrix can be regarded merely as a component of the required pivots rather than as a quantity of intrinsic interest. It follows that the problem of small sample inference could be approached more directly by exploring the small sample distribution of Wald statistics calculated using the estimated asymptotic covariance matrix of $\hat{\boldsymbol{\beta}}$. However, we shall first consider the bias in the estimate of the covariance matrix. There are two reasons for taking this initial step. First, the approximation to the small sample distribution of the Wald statistic is simpler when the adjusted covariance matrix is used. Second, if measures of precision are to be presented, then the adjusted ones are preferable in terms of behaviour and, in addition, if such estimates are used, we are able to maintain the familiar relationship between these and the Wald statistics.

The problem of approximating the small sample precision of $\hat{\boldsymbol{\beta}}$ and related distributional issues has been thoroughly explored in a number of papers by Harville and coworkers (Kackar and Harville, 1984; Harville, 1985; Jeske and Harville, 1988; Hulting and Harville, 1991; Harville and Jeske, 1992; Harville and Carriquiry, 1992). In the following section, we combine results from these papers to provide an adjusted estimator of the covariance matrix of $\hat{\boldsymbol{\beta}}$. In Section 3, a Wald statistic is constructed for linear combinations of fixed effects using this estimator. The statistic is scaled using a quantity calculated from the data, and its small sample distribution is approximated by an F distribution. This leads to a general procedure that reproduces exact distributional results when these exist, such as for Hotelling's T^2 and analysis of variance F-ratios, and allows consistency among distributional approximations for higher-dimensional Wald tests and component t-tests.

Attempts to examine the behaviour of small sample procedures in REML have been confined in the past to rather simple settings, such as one- and two-way random effects models and balanced incomplete block designs (see, e.g., Kackar and Harville, 1984; McLean and Sanders, 1988). To assess the behavior of the proposed methods, we use situations that are rather more varied in their type and complexity and therefore, we hope, reflect a wider range of practical problems. In Section 4, simulation results from four settings are considered: a two-period crossover trial, a random effects model for an incomplete row and column design, a random coefficient regression model, and an ante-dependence analysis for repeated measurements. In Section 5, the approximations are applied in three examples: a small crossover study on the treatment of intermittent claudication, a nested experiment used to investigate the nitrogen fixing ability of rhizobia bacteria, and a repeated measurements experiment on the effect of added nitrogen on the concentration of a soil element.

2. Variance–Covariance Matrix of the REML Fixed Effects

Recall from Section 1 that we have n observations $\mathbf{Y}$ following a multivariate Gaussian distribution,

$$\mathbf{Y} \sim \mathrm{N}(\mathbf{X}\boldsymbol{\beta}; \boldsymbol{\Sigma}).$$

The elements of the variance–covariance matrix $\boldsymbol{\Sigma}$ are assumed to be functions of r parameters, $\boldsymbol{\sigma}$ $(r \times 1)$, that are sufficiently well behaved for the resulting likelihood to be regular in the sense of Cox and Hinkley (1974, p. 281). In particular, we assume that the first two partial derivatives with respect to the elements of $\boldsymbol{\sigma}$ exist. To simplify notation, we show explicit dependence of $\boldsymbol{\Sigma}$ on $\boldsymbol{\sigma}$ only when we wish to emphasise it.

The REML estimator of $\boldsymbol{\sigma}$, which we denote $\hat{\boldsymbol{\sigma}}$, is the maximum likelihood estimator from the marginal likelihood of the variables $\mathbf{Z} = \mathbf{KY}$, where $\mathbf{K}$ is any $(n-p) \times n$ matrix of full rank satisfying $\mathbf{KX} = 0$. The marginal likelihood of $\mathbf{Z}$ does not depend on the particular choice of $\mathbf{K}$ and can be expressed in terms of $\mathbf{Y}$ and $\mathbf{X}$ only as

$$\begin{aligned} 2\log \mathrm{L}(\boldsymbol{\sigma}) = \text{constant} &- \log\{|\boldsymbol{\Sigma}|\} - \log\{|\mathbf{X}^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}\mathbf{X}|\} \\ &- \mathbf{Y}^{\mathrm{T}}\{\boldsymbol{\Sigma}^{-1} - \boldsymbol{\Sigma}^{-1}\mathbf{X}(\mathbf{X}^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}\mathbf{X})^{-1}\mathbf{X}^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}\}\mathbf{Y}. \end{aligned}$$

The REML estimator of $\boldsymbol{\beta}$ is the generalized least squares estimator

$$\hat{\boldsymbol{\beta}} = \Phi(\hat{\boldsymbol{\sigma}})\mathbf{X}^{\mathrm{T}}\boldsymbol{\Sigma}(\hat{\boldsymbol{\sigma}})^{-1}\mathbf{Y}$$

for

$$\boldsymbol{\Phi}(\boldsymbol{\sigma}) = \{\mathbf{X}^{\mathrm{T}}\boldsymbol{\Sigma}(\boldsymbol{\sigma})^{-1}\mathbf{X}\}^{-1}.$$

Kackar and Harville (1984) show that $\hat{\boldsymbol{\beta}}$ is unbiased for $\boldsymbol{\beta}$.

The matrix $\boldsymbol{\Phi}$ is the variance–covariance matrix of the asymptotic limiting distribution of $\hat{\boldsymbol{\beta}}$ as $n \to \infty$ and conventionally the estimator of $\boldsymbol{\Phi}$, $\hat{\boldsymbol{\Phi}} = \boldsymbol{\Phi}(\hat{\boldsymbol{\sigma}})$, is used to provide a measure of precision for $\hat{\boldsymbol{\beta}}$. It is possible to identify two main sources of bias in $\hat{\boldsymbol{\Phi}}$ when it is used as an estimator of $V[\hat{\boldsymbol{\beta}}]$ in small samples. First, $\boldsymbol{\Phi}(\boldsymbol{\sigma})$ takes no account of the impact on $V[\hat{\boldsymbol{\beta}}]$ of the variability in $\hat{\boldsymbol{\sigma}}$ and so is not equal to $V[\hat{\boldsymbol{\beta}}]$. Second, $\hat{\boldsymbol{\Phi}}$ is a biased estimator of $\boldsymbol{\Phi}(\boldsymbol{\sigma})$.

We consider first a better approximation than $\boldsymbol{\Phi}$ to the small sample variance–covariance matrix of $\hat{\boldsymbol{\beta}}$. As in shown in Kackar and Harville (1984), the variability in $\hat{\boldsymbol{\beta}}$ can be partitioned into two components,

$$V[\hat{\boldsymbol{\beta}}] = \boldsymbol{\Phi} + \boldsymbol{\Lambda},$$

where the component $\boldsymbol{\Lambda}$ represents the amount to which the asymptotic variance–covariance matrix underestimates (in a matrix sense) $V[\hat{\boldsymbol{\beta}}]$. Using an argument based on a Taylor series expansion around $\boldsymbol{\sigma}$, Kackar and Harville show that $\boldsymbol{\Lambda}$ can be approximated by

$$\boldsymbol{\Lambda} \simeq \boldsymbol{\Phi}\left\{\sum_{i=1}^{r}\sum_{j=1}^{r} W_{ij}(\mathbf{Q}_{ij} - \mathbf{P}_i\boldsymbol{\Phi}\mathbf{P}_j)\right\}\boldsymbol{\Phi}, \tag{1}$$

where

$$\mathbf{P}_i = \mathbf{X}^{\mathrm{T}}\frac{\partial\boldsymbol{\Sigma}^{-1}}{\partial\sigma_i}\mathbf{X} \quad \text{and} \quad \mathbf{Q}_{ij} = \mathbf{X}^{\mathrm{T}}\frac{\partial\boldsymbol{\Sigma}^{-1}}{\partial\sigma_i}\boldsymbol{\Sigma}\frac{\partial\boldsymbol{\Sigma}^{-1}}{\partial\sigma_j}\mathbf{X}$$

and W_{ij} is the (i,j)th element of $\mathbf{W} = V[\hat{\boldsymbol{\sigma}}]$.

We now consider the bias in $\hat{\boldsymbol{\Phi}}$ as an estimator of $\boldsymbol{\Phi}$. Using a Taylor series expansion about $\boldsymbol{\sigma}$, we have

$$\hat{\boldsymbol{\Phi}} \simeq \boldsymbol{\Phi} + \sum_{i=1}^{r}(\hat{\sigma}_i - \sigma_i)\frac{\partial\boldsymbol{\Phi}}{\partial\sigma_i} + \frac{1}{2}\sum_{i=1}^{r}\sum_{j=1}^{r}(\hat{\sigma}_i - \sigma_i)(\hat{\sigma}_j - \sigma_j)\frac{\partial^2\boldsymbol{\Phi}}{\partial\sigma_i\partial\sigma_j},$$

from which we get, ignoring possible bias in $\hat{\boldsymbol{\sigma}}$,

$$\mathrm{E}[\hat{\boldsymbol{\Phi}}] \simeq \boldsymbol{\Phi} + \frac{1}{2}\sum_{i=1}^{r}\sum_{j=1}^{r} W_{ij}\frac{\partial^2\boldsymbol{\Phi}}{\partial\sigma_i\partial\sigma_j}. \tag{2}$$

In terms of the previously defined quantities, this can be written as

$$\frac{\partial^2\boldsymbol{\Phi}}{\partial\sigma_i\partial\sigma_j} = \boldsymbol{\Phi}(\mathbf{P}_i\boldsymbol{\Phi}\mathbf{P}_j + \mathbf{P}_j\boldsymbol{\Phi}\mathbf{P}_i - \mathbf{Q}_{ij} - \mathbf{Q}_{ji} + \mathbf{R}_{ij})\boldsymbol{\Phi},$$

where

$$\mathbf{R}_{ij} = \mathbf{X}^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}\frac{\partial^2\boldsymbol{\Sigma}}{\partial\sigma_i\partial\sigma_j}\boldsymbol{\Sigma}^{-1}\mathbf{X}.$$

An adjusted estimator of the small sample variance–covariance matrix of $\hat{\boldsymbol{\beta}}$ can then be obtained by combining (1) and (2),

$$\hat{\boldsymbol{\Phi}}_A = \hat{\boldsymbol{\Phi}} + 2\hat{\boldsymbol{\Phi}}\left\{\sum_{i=1}^{r}\sum_{j=1}^{r} W_{ij}(\mathbf{Q}_{ij} - \mathbf{P}_i\hat{\boldsymbol{\Phi}}\mathbf{P}_j - \frac{1}{4}\mathbf{R}_{ij})\right\}\hat{\boldsymbol{\Phi}}, \tag{3}$$

where $\boldsymbol{\Sigma}(\hat{\boldsymbol{\sigma}})$ is substituted for $\boldsymbol{\Sigma}$ in the quantities within the summation on the righthand side.

An important class of covariance structures is defined by the linear form

$$\boldsymbol{\Sigma} = \sum_{i=1}^{r}\sigma_i\mathbf{G}_i,$$

with the $\{\mathbf{G}_i\}$ known $(n \times n)$ matrices. This class includes structures arising from mixed models.

For such linear covariance structures,

$$\frac{\partial^2 \boldsymbol{\Sigma}}{\partial \sigma_i \partial \sigma_j} = 0$$

for all pairs (i, j) and this implies that $\mathbf{R}_{ij} = \mathbf{0}$. Expression (3) then simplifies to

$$\hat{\boldsymbol{\Phi}}_A = \hat{\boldsymbol{\Phi}} - \sum_{i=1}^{r}\sum_{j=1}^{r} W_{ij} \frac{\partial^2 \boldsymbol{\Phi}}{\partial \sigma_i \partial \sigma_j} \tag{4}$$

$$= \hat{\boldsymbol{\Phi}} + 2\hat{\boldsymbol{\Lambda}} \tag{5}$$

for $\boldsymbol{\Lambda}$ as represented in (1). This factor of two in the bias correction under a linear structure has been noted on a number of occasions (see, e.g., Harville and Jeske, 1992, Section 4.1).

An approximation to the variance–covariance matrix of $\hat{\boldsymbol{\sigma}}$, $\mathbf{W}$, can be obtained from the inverse of the expected information matrix $\mathbf{I}_E$, where

$$2\{\mathbf{I}_E\}_{ij} = \operatorname{tr}\left(\frac{\partial \boldsymbol{\Sigma}^{-1}}{\partial \sigma_i}\boldsymbol{\Sigma}\frac{\partial \boldsymbol{\Sigma}^{-1}}{\partial \sigma_j}\boldsymbol{\Sigma}\right) - \operatorname{tr}(2\boldsymbol{\Phi}\mathbf{Q}_{ij} - \boldsymbol{\Phi}\mathbf{P}_i\boldsymbol{\Phi}\mathbf{P}_j).$$

Alternatively, the expected information matrix can be replaced by the observed or average information matrices. For details, see Gilmour, Thompson, and Cullis (1995).

3. Inference and Degrees of Freedom

Suppose that inferences are to be made simultaneously about the ℓ linear combinations of the elements of $\boldsymbol{\beta}$: $\mathbf{L}\boldsymbol{\beta}$, for $\mathbf{L}$ an $(\ell \times p)$ fixed matrix. It is proposed that the adjusted estimator of the covariance matrix be used in Wald-type pivots of the form

$$F = \frac{1}{\ell}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})^{\mathrm{T}}\mathbf{L}(\mathbf{L}^{\mathrm{T}}\hat{\boldsymbol{\Phi}}_A\mathbf{L})^{-1}\mathbf{L}^{\mathrm{T}}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}). \tag{6}$$

The derivation of an appropriate F approximation for (6) when $\ell = 1$ is comparatively straightforward and essentially recovers Satterthwaite's (1941) approximation. For $\ell > 1$, the situation is more complicated in that it is necessary to take into account the internal random structure of $\mathbf{L}^{\mathrm{T}}\hat{\boldsymbol{\Phi}}\mathbf{L}$, and this may be very different in different settings. Moreover, for the special case of Hotelling's T^2 type statistics (Krzanowski, 1988, Section 8.3), it is known that it is a scaled form of F, F^* say, where

$$F^* = \frac{m}{m + \ell - 1}F,$$

that has an exact $F_{\ell,m}$ distribution. This and some simple studies in which F distributions have been fitted to simulated statistics of the form of F in (6) suggest that, if an F approximation with numerator degrees of freedom equal to ℓ is to be used, then a scaled form $F^* = \lambda F$ will be required, where typically $\lambda \leq 1$. Furthermore, if, as we would like, the chosen approximation is to reproduce the correct F degrees of freedom when the distribution is exact, then such a scale factor will have to be introduced to accommodate Hotelling's T^2 type statistics as well as the more familiar analysis of variance F-ratios.

We derive such an approximation as follows. We need to calculate two quantities from the data, the scale factor λ and the denominator degrees of freedom m.

First, using a Taylor series expansion for $\{L^{\mathrm{T}}\hat{\boldsymbol{\Phi}}_A L\}^{-1}$, we have

$$\{\mathbf{L}^{\mathrm{T}}\hat{\boldsymbol{\Phi}}_A\mathbf{L}\}^{-1} \simeq \{\mathbf{L}^{\mathrm{T}}\boldsymbol{\Phi}_A\mathbf{L}\}^{-1} + \sum_{i=1}^{r}(\hat{\sigma}_i - \sigma_i)\frac{\partial\{\mathbf{L}^{\mathrm{T}}\hat{\boldsymbol{\Phi}}_A\mathbf{L}\}^{-1}}{\partial \sigma_i}$$
$$+ \frac{1}{2}\sum_{i=1}^{r}\sum_{j=1}^{r}(\hat{\sigma}_i - \sigma_i)(\hat{\sigma}_j - \sigma_j)\frac{\partial^2\{\mathbf{L}^{\mathrm{T}}\hat{\boldsymbol{\Phi}}_A\mathbf{L}\}^{-1}}{\partial \sigma_i \partial \sigma_j}.$$

Ignoring possible bias in $\hat{\boldsymbol{\sigma}}$ and possible statistical dependence between $\hat{\boldsymbol{\beta}}$ and $\hat{\boldsymbol{\Phi}}$ and using the relationships

$$\mathrm{E}[F] = \mathrm{E}_{\hat{\boldsymbol{\Phi}}_A}[\mathrm{E}[F \mid \hat{\boldsymbol{\Phi}}_A]]$$

and

$$\mathrm{V}[F] = \mathrm{E}_{\hat{\boldsymbol{\Phi}}_A}[\mathrm{V}[F \mid \hat{\boldsymbol{\Phi}}_A]] + \mathrm{V}_{\hat{\boldsymbol{\Phi}}_A}[\mathrm{E}[F \mid \hat{\boldsymbol{\Phi}}_A]],$$

we can write, approximately, after taking appropriate expectations and eliminating higher-order terms in the Taylor series expansion,

$$\mathrm{E}[F] \simeq 1 + \frac{1}{\ell}(A_2)$$

and

$$\mathrm{V}[F] \simeq \frac{2}{\ell}(1 + B),$$

where

$$B = \frac{1}{2\ell}(A_1 + 6A_2)$$

and

$$A_1 = \sum_{i=1}^{r}\sum_{j=1}^{r} W_{ij}\mathrm{tr}(\boldsymbol{\Theta\Phi}\mathbf{P}_i\boldsymbol{\Phi})\mathrm{tr}(\boldsymbol{\Theta\Phi}\mathbf{P}_j\boldsymbol{\Phi}),$$

$$A_2 = \sum_{i=1}^{r}\sum_{j=1}^{r} W_{ij}\mathrm{tr}(\boldsymbol{\Theta\Phi}\mathbf{P}_i\boldsymbol{\Phi\Theta\Phi}\mathbf{P}_j\boldsymbol{\Phi})$$

for

$$\boldsymbol{\Theta} = \mathbf{L}(\mathbf{L}^{\mathrm{T}}\boldsymbol{\Phi}\mathbf{L})^{-1}\mathbf{L}^{\mathrm{T}}.$$

To the order of the Taylor series approximation, by matching moments of F^* with those of the approximating F distribution, we get

$$m = 4 + \frac{\ell + 2}{\ell\rho - 1}, \qquad \text{where } \rho = \frac{\mathrm{V}[F]}{2\mathrm{E}[F]^2}, \tag{7}$$

and

$$\lambda = \frac{m}{\mathrm{E}[F](m-2)}. \tag{8}$$

These quantities can be estimated by substituting $\hat{\boldsymbol{\sigma}}$ for $\boldsymbol{\sigma}$, and an estimate of $\mathbf{W}$ can be obtained as in Section 2.

These estimates of m and λ in (7) and (8) will match the known values from those special cases in which the F approximation is exact up to the order of the Taylor series approximation. To obtain estimates that are exact in these special cases, we modify the estimates of m and λ in such a way that we leave unchanged the approximation up to the order of the Taylor series expansion; that is, we add selected terms of higher degree. Omitting details of the derivation of these terms, we arrive at the following modified estimates of m and λ. Essentially, we just modify the approximation of the expectation and variance of F to produce E^* and V^*, i.e.,

$$\mathrm{E}^* = \{1 - A_2/\ell\}^{-1}$$

$$\mathrm{V}^* = \frac{2}{\ell}\left\{\frac{1 + c_1 B}{(1 - c_2 B)^2(1 - c_3 B)}\right\},$$

where

$$c_1 = \frac{g}{3\ell + 2(1-g)}$$

$$c_2 = \frac{l - g}{3\ell + 2(1-g)}$$

$$c_3 = \frac{l + 2 - g}{3\ell + 2(1-g)}$$

for

$$g = \frac{(\ell+1)A_1 - (\ell+4)A_2}{(\ell+2)A_2}.$$

The two quantities E^* and V^* replace $\mathrm{E}[F]$ and $\mathrm{V}[F]$ in (7) and (8) to produce the modified estimates of m and λ.

It can be checked that these match the previous values to the order of the Taylor series expansion and also that they reproduce the correct values when λF has an exact F distribution.

For the special case $\ell = 1$, λ is exactly equal to one and, after taking the square root, the estimate of m turns out to be a simple modification of Satterthwaites (1941) degree of freedom estimator for the t-test. Giesbrecht and Burns (1985) apply this method in the current setting for t statistics with a mixed model covariance structure. The advantage of the approximation developed here is that it is consistent across all values of ℓ. For example, ignoring the scale factor when $\ell > 1$ can lead to values of m that are considerably smaller than Satterthwaite's approximation for any one-dimensional component $\mathbf{a}^{\mathrm{T}}\mathbf{L}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})$, where $\mathbf{a}$ is a fixed $(\ell \times 1)$ vector.

4. Simulation Results

To explore the behavior of the statistics developed in the previous two sections, we have conducted simulation studies for four different settings. We present the results from these in turn. Comparisons are made with unadjusted tests; that is, with tests that use the asymptotic χ^2_ℓ approximation to the distribution of

$$(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})^{\mathrm{T}}\mathbf{L}(\mathbf{L}^{\mathrm{T}}\hat{\boldsymbol{\Phi}}\mathbf{L})^{-1}\mathbf{L}^{\mathrm{T}}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})$$

or, when $\ell = 1$, the standard normal approximation to its square root.

4.1 *A Four-Treatment Two-Period Crossover Trial*

In a crossover trial, each unit receives a sequence of treatments and, to avoid confounding treatment and period (or time) effects, such trials use several different sequences (see, e.g., Jones and Kenward, 1989). Conventionally, data from crossover trials are analyzed using within-unit information only; that is, with fixed unit effects. It has been suggested, however, that for those designs in which there is information on treatment effect in the between-unit stratum, this information should be recovered through the use of random unit effects (Chi, 1991). Many crossover trials as used in practice have few units, and it is with such examples that we might expect the small sample effects considered in this paper to be particularly apparent.

We choose for the simulation study a 12-sequence design with 4 treatments. The sequences consist of all pairs of treatments in all orders and one unit is assumed to be randomized to each sequence, making a total of 12 units. The fixed-effects component of the model incorporates the period difference and three treatment effects. The random component of the model, and generated data, consists of between-unit and within-unit variance components, with their ratio denoted by ρ. Five settings for ρ have been used: 0.25, 0.5, 1, 2, and 4.

For each setting of ρ, 10,000 sets of data have been simulated from the corresponding multivariate Gaussian distribution, with zero treatment effects, and for each set the REML estimates of variance components and fixed effects have been calculated. A Fisher scoring algorithm was used employing the average information matrix as specified in Gilmour, Thompson, and Cullis (1996). If this failed to converge for any particular set of data, a more time-consuming simplex search was used. The between-subject variance component was not constrained to be positive. The variance adjustments and degrees of freedom were calculated using all three forms of information: expected, observed, and averaged. There was very little difference between these and so only results from the expected information are presented.

In Table 1, we present the percentage relative bias in the estimates of variance of the fixed effects. This is defined for the fixed effect $\ell^{\mathrm{T}}\boldsymbol{\beta}$ as

$$100\left(\frac{\mathrm{E}_S[\ell^{\mathrm{T}}\hat{\boldsymbol{\Phi}}_A\ell]}{\mathrm{V}_S[\ell^{\mathrm{T}}\hat{\boldsymbol{\beta}}]} - 1\right),$$

where the subscript S indicates moments taken over the simulated sets of data. The bias of the unadjusted estimator is very large and negative, as expected. This bias is reduced to an acceptable level by the adjustment, although not totally eliminated for relatively small between-subject variability. The bias in all the estimators decreases with increasing between-subject variability.

The observed size of nominal 5% t- and F-tests are also presented in Table 1 together with the average over the simulations of the estimated effective degrees of freedom. The t test is for a single treatment comparison, the particular choice is immaterial given the balance in the design. The F-test is for an overall treatment difference on 3 degrees of freedom and, because the design is balanced with respect to treatment, the scale factor associated with this statistic is equal to one irrespective of the estimated variance components, and so is not presented in the tables. The behaviour of the unadjusted tests is unacceptably poor, noting that with 10,000 simulations the standard devi-

Table 1

Two-period crossover trial: percentage relative bias in the variance estimates, estimated effective degrees of freedom, and actual size of nominal 5% Wald t- and F-tests from the simulation study

	% Relative bias, variance estimates		Mean effective	Observed size Wald t-tests		Observed size Wald F-tests	
ρ	Asy.	Adj.	d.f.	Asy.	Adj.	Asy.	Adj.
0.25	−18	−2	13.9	9.8	5.5	13.5	5.7
0.5	−16	−2	12.8	9.7	5.5	13.6	5.7
1.0	−12	0	11.4	9.4	5.4	13.2	5.4
2.0	−8	0	10.1	8.9	5.2	12.4	5.3
4.0	−5	1	9.1	8.7	5.2	12.2	5.1

ation of the observed size is approximately 0.2%, but the combination of adjusted estimate of precision and estimated degrees of freedom produces the required significance level. Note that the estimated degrees of freedom decrease with increasing between-subject variance. This reflects the decreasing contribution of the between-subject information to the estimate of the fixed effects and their covariance matrix.

4.2 *A Row–Column α-Design*

We next consider an incomplete block design with two blocking factors using random row and column effects. We use a row–column α-design as defined by John and Eccleston (1986), generated using the algorithm of Nguyen and Williams (1993). The design has 3 replicates of 12 treatments (labelled 1–12) in 3×4 arrays:

replicate												
1				2				3				
10	4	8	1	12	4	2	6	12	7	4	10
11	5	2	7	9	11	7	10	11	8	6	1
3	9	12	6	1	8	3	5	5	9	3	2

Row and column effects are assumed to be independent across replicates and there is assumed to be no replicate effect. In the simulations, the variances of the row and column effects have been set equal (to σ_B^2, say) and we denote by ρ the ratio of σ_B^2 to the residual mean square. A range of values of ρ from 0.25 to 4 has been used. The fixed effects (treatment) part of the model was parameterized in terms of the 11 differences from treatment 1.

For each setting of ρ, 5000 sets of data were simulated from the corresponding multivariate Gaussian distribution with zero treatment effects. The same numerical methods were used as for the crossover simulation above and the between-row and between-column variance components were not constrained to be positive. The relative percentage bias in the variance estimates are presented in Table 2 and the observed test sizes in Table 3. These are presented for adjustments calculated using all three types of information matrix. For the single contrast statistics (variance bias, degrees of freedom, and t-test size), the minimum, maximum (in absolute size), and mean of the 11 treatment differences are given. The design is not balanced and so we expect some variation in the behavior of these statistics.

The pattern of bias in the estimates of precision of the fixed effects is similar to the previous example. The relative bias of the unadjusted estimator ranges between 7 and 32% and is unacceptably large. This bias decreases as the ratio of between-block to within-block variability increases. The adjusted estimators remove much of this bias, with the expected and adjusted information-based estimators performing somehat better that that based on the observed information matrix. The latter tends to overcorrect for the bias, producing slightly conservative estimates.

The actual sizes of nominal 5% Wald tests for each treatment difference are summarized in Table 3 (with an approximate standard deviation of 0.3%) for the same range of settings as in Table 2. The observed size of nominal 5% tests for overall treatment effect (on 11 degrees of freedom), the average calculated scale factor, and the average estimated denominator degrees of freedom were as follows:

ρ	0.25	0.5	1	2	4
Size, asymptotic test	35.7	32.8	27.9	26.9	21.7
Size, adjusted test	4.9	5.3	5.3	5.4	5.1
Scale factor	0.93	0.96	0.99	1.00	1.00
Denominator d.f.	11.4	11.3	10.5	9.4	8.3

As with the crossover design above, there was little to choose between the results from the three types of information, and so the results from the expected information only are presented. An approximate 95% probability interval for the observed size based on an actual 5% level is ±0.6 percentage points. The sizes of the adjusted tests are quite acceptable, in this example correcting the grossly inflated sizes of the asymptotic tests. Again, the behavior of the degrees of freedom and in addition the scale factor, with increasing between-row and between-column variability, is as expected. The limiting case is provided by an analysis with fixed row and column effects from which the value of the denominator degrees of freedom and scale factor are equal to 5 and 1, respectively.

4.3 *A Random Coefficient Regression Model*

For the third simulation study, we consider a simple random coefficient regression model as used in the analysis of longitudinal data (see, e.g., Laird and Ware, 1982). We assume that the repeated measurements from each of n units can be represented by a simple linear regression over time with parameters varying among units. Suppose that the ith unit is observed at q_i times $t_{i1}, \ldots t_{iq_i}$, with associated observations $\mathbf{y}_i = (Y_{i1}, \ldots Y_{iq_i})^{\mathrm{T}}$. A simple linear regression relationship is used for the first stage of the random coefficient model, i.e.,

$$\mathbf{Y} \mid \mathbf{b}_i \sim \mathrm{N}(\mathbf{X}_i \mathbf{b}_i; \sigma^2 \mathbf{I}_{q_i}),$$

where $\mathbf{b}_i = (b_{i0}, b_{i1})^{\mathrm{T}}$, the intercept and slope parameter for this unit, and the jth row of $\mathbf{X}_i$ is equal to $(1, t_{ij})$. The regression parameters are then assumed to follow a multivariate Gaussian distribution $\mathbf{b}_i \sim \mathrm{N}(\boldsymbol{\beta}, \boldsymbol{\Omega})$ with $\boldsymbol{\Omega}$ unstructured, so that marginally

$$\mathbf{Y}_i \sim \mathrm{N}(\mathbf{X}_i \boldsymbol{\beta}; \sigma^2 \mathbf{X}_i \mathbf{X}_i^{\mathrm{T}} + \boldsymbol{\Omega}),$$

for $\boldsymbol{\beta} = (\beta_0, \beta_1)^{\mathrm{T}}$.

If the repeated measurements were balanced, that is if all units were observed on the same occasions, the GLS estimate of $\boldsymbol{\beta}$ would be independent of σ^2 and $\boldsymbol{\Omega}$. In such a situation, the adjustment term in $\hat{\boldsymbol{\Phi}}_A$ vanishes. We therefore choose for the simulations an extremely unbalanced set. It is assumed that there is a notional set of 9 equally spaced times of measurement: $0, 1, \ldots, 8$,

Table 2

Row and column α-design: percentage relative bias in the variance estimates. Minimum and maximum refer to absolute size over all pairwise treatment differences.

			Adjusted		
ρ		Asymptotic	Expected	Observed	Average
0.25	Minimum	−24	−3	9	1
	Mean	−31	−6	4	−4
	Maximum	−34	−9	−3	−8
0.5	Minimum	−24	−3	3	−2
	Mean	−32	−8	0	−7
	Maximum	−35	−10	−3	−9
1	Minimum	−15	0	4	0
	Mean	−24	−5	−1	−5
	Maximum	−27	−8	−3	−7
2	Minimum	−15	−3	9	−2
	Mean	−21	−6	3	−5
	Maximum	−24	−8	−2	−7
4	Minimum	−7	2	2	2
	Mean	−10	0	1	0
	Maximum	−12	−2	−2	−2

Table 3

Row and column α-design: actual size of nominal 5% Wald tests. The minimum, maximum, and mean refer to t-tests for all pairwise treatment differences.

			Adjusted		
ρ		Asymptotic	Expected	Observed	Average
0.25	Maximum	14.3	6.3	5.8	6.2
	Mean	12.9	5.7	5.4	5.7
	Minimum	11.8	5.1	5.0	5.0
0.5	Maximum	13.8	5.8	4.8	6.2
	Mean	12.3	5.8	5.9	5.7
	Minimum	11.4	5.2	5.4	5.1
1	Maximum	11.8	6.1	6.1	6.1
	Mean	11.3	5.7	5.7	5.9
	Minimum	10.3	5.0	5.0	6.5
2	Maximum	11.9	6.5	6.3	6.5
	Mean	11.1	6.0	5.8	5.9
	Minimum	10.4	5.7	5.3	5.6
4	Maximum	10.3	5.6	5.6	5.7
	Mean	9.8	5.4	5.3	5.4
	Minimum	9.4	5.0	5.0	5.0

with 24 subjects split into 3 groups of 8, the members of each group being observed on the same 3 occasions. The three sets of occasions are chosen to be nonoverlapping: $(0, 1, 2)$, $(3, 4, 5)$, and $(6, 7, 8)$.

For $\boldsymbol{\Omega}$, we define

$$\sigma^{-2}\boldsymbol{\Omega} = \rho \begin{bmatrix} 1.00 & -0.53 \\ -0.53 & 1.00 \end{bmatrix},$$

with ρ taking the values 1.0 and 0.5 in the simulations, and the residual variance was set equal to 0.25. Each simulation run consisted of 20,000 samples. An accelerated EM algorithm (Dempster, Laird, and Rubin, 1977) was used to fit the covariance model, and this was replaced by a simplex search when convergence was particularly slow. The estimate of $\boldsymbol{\Omega}$ was constrained to be positive definite.

The results from the simulations are summarized in Table 4. As before, we present the percentage relative bias of the unadjusted and adjusted variance estimates and the size of nominal 5% Wald tests for the two parameters. No overall F-test was calculated; it was thought to have little meaning in this setting. We also present the mean over all the simulations of the estimated degrees of freedom from each adjusted t-test. It can be seen that the expected negative bias in the unadjusted estimate of variance is present, but is not large for the slope parameter β_1. The adjustment does succeed in removing most of the bias. The high (10%) unadjusted test sizes can be attributed to the infinite degrees of freedom implicit in the chi-squared approximation. The adjusted test sizes are perfectly acceptable given an approximate standard deviation of 0.2% for the observed sizes. It is interesting that the mean of the estimated effective degrees of freedom is notably smaller for the intercept than for the slope, the degree of freedom for the latter approach the residual degrees of freedom among subjects. If the data were balanced, these tests would have exact t-distributions with 23 degrees of freedom.

4.4 *Ante-Dependence Analysis for Repeated Measurements*

The preceding simulation studies have been based on linear covariance structures. Our final study uses a nonlinear structure. We would like an example in which we expect the estimation of the covariance structure to have a nonnegligible impact on the precision of the fixed effects estimates. The ante-dependence (AD) structure for repeated measurements (Gabriel, 1962; Kenward, 1987) is a potential candidate for this in that the number of parameters is of the same order as the number of times of measurement. For a definition of the AD structure, see Gabriel (1962); it can be derived in several ways, for example using conditional independence arguments. In the simulations, a first-

Table 4

Random-coefficient regression model: percentage relative bias in the variance estimates, mean effective degrees of freedom, and actual size of nominal 5% Wald tests from the simulation study

Parameter	ρ	Percentage bias in variance estimates		Mean effective d.f.	Observed test size	
		Asymptotic	Adjusted		Asymptotic	Adjusted
β_0	0.5	−13	−1	13.7	9.1	5.0
	1.0	−12	−3	15.6	9.3	5.8
β_1	0.5	−2	1	22.6	6.4	4.7
	1.0	−5	−3	22.8	6.9	5.3

order AD structure is used. This is equivalent to assuming that the inverse of the variance–covariance matrix is symmetric, positive definite, and tridiagonal, implying a Markov structure among the repeated measurements. No other constraints, such as stationarity, are imposed and the matrix is therefore defined by $2T-1$ parameters given T times of measurement. The chosen matrix has the following variance–correlation form:

$$
\begin{array}{cccccc}
1.00 & & & & & \\
0.54 & 2.81 & & & & \\
0.33 & 0.61 & 4.80 & & & \\
0.20 & 0.37 & 0.61 & 6.35 & & \\
0.12 & 0.22 & 0.35 & 0.58 & 6.79 & \\
0.06 & 0.11 & 0.18 & 0.29 & 0.50 & 6.70
\end{array}
$$

For the mean structure, we take the simple situation in which there are 2 groups of 10 units, each unit being measured notionally on the same $T = 6$ occasions. A number of values have been assumed to be missing through dropout (a monotone data structure) with the numbers of observations per unit as follows:

Group	Number of observations per unit									
1	6	6	6	6	5	4	4	4	3	3
2	6	6	6	6	4	4	3	3	3	2

We fit a model including time and group main effects and time-by-group interaction. As well as estimating the effects in the model, we test the overall time-by-group interaction on five degrees of freedom. The time effects have been parameterized in terms of orthogonal polynomials. The simulation run consisted of 20,000 samples. An EM algorithm was used to fit the model and there were no convergence failures. The adjustments use the expected information matrix.

The results from the simulation study are summarized in Table 5. The bias among the unadjusted variance estimates fluctuates considerably among the different contrasts, but typically lies between −5 and −10%. The adjustment effectively removes this small bias. As with the bias, the mean effective degrees of freedom vary among the different polynomial contrasts, reflecting the differing relationships of the contrasts with the covariance structure as modified by the dropout pattern. Note that the degrees of freedom nowhere exceed 18, the maximum residual degrees of freedom at any one time of measurement. The unadjusted t-tests have an observed size that is too large, given an approximate standard deviation of 0.2% for these figures. Again, this is corrected effectively using the adjustment.

The unadjusted F-test for the treatment-by-time interaction on 5 degrees of freedom produced an observed size of 19.6%, while the adjusted test has an observed size of 5.8% with estimated effective degrees of freedom of 14.1, in the middle of the range of the degrees of freedom for the component t-tests. The scale factor had a mean value of 0.8, with a very small variance over the simulations (<0.001). The comparatively small size of the scale factor reflects the number of parameters in the covariance structure. Note that, if the Wald statistic were not scaled, the estimated effective degrees of freedom would have had a mean of 6.6, a much smaller figure than that obtained above and bearing little relation to the degrees of freedom of the component t-tests.

Table 5

Ante-dependence analysis: percentage relative bias in the variance estimates, mean effective degrees of freedom, and actual size of nominal 5% Wald tests from the simulation study

	% Relative bias in variance estimates		Mean effective	Observed test size (nominal 5%)	
Parameter	Asymptotic	Adjusted	d.f.	Asymptotic	Adjusted
constant	−5	−1	14.8	7.7	5.2
lin	−8	−1	10.3	9.0	5.4
qua	−5	−1	16.1	7.5	5.1
cub	−6	0	14.0	7.9	5.1
4th	−9	−1	13.7	8.3	5.1
5th	−5	1	17.5	7.4	5.2
group	−4	0	14.9	7.5	5.0
g×lin	−9	−1	10.7	8.9	5.4
g×qua	−5	0	16.0	7.7	5.1
g×cub	−6	0	14.5	7.8	5.0
g×4th	−10	−2	14.0	8.5	5.3
g×5th	−5	1	17.7	7.3	5.1

5. Examples

5.1 *A Four-Period Crossover Trial*

Jones and Kenward (1989, p. 232) describe a crossover trial with the aim of comparing the effects of three active treatments (A, B, and C) and a placebo on subjects with intermittent claudication. A four-period design was used in which a different ordering of the four treatments was used for each subject. The design and data from one response measurement, the left ventricular ejection time (in milliseconds), are given in Table 5.23 of Jones and Kenward (1989). The covariance structure of these data is far from uniform and previous analyses (e.g., Jones and Kenward, 1989, Section 7.6) have explored the use of other patterned and unstructured covariance matrices for these data. As an illustration of the small sample bias that arises from variability in the estimated covariance parameters, we show the results from fitting an unstructured covariance matrix. A linear model is used with period and direct treatment effects and the small sample adjustments have been calculated using the expected information matrix. The treatment effects, presented as differences from placebo, were estimated as follows, together with asymptotic and adjusted standard errors and corresponding estimated degrees of freedom:

		SE		
Effect	Estimate	Asymptotic	Adjusted	d.f.
A–P	−52.0	12.0	15.8	24.7
B–P	−48.7	12.2	16.1	26.3
C–P	−66.1	12.1	15.9	25.4

As a comparison with the simulation results, the percentage relative bias in the variance estimates for the three treatment effects is −42%. A summary of two Wald tests are presented below: the unadjusted and adjusted statistics for testing overall treatment differences adjusted for period effects and for periods adjusted for treatments.

	Periods	Treatments
Numerator d.f.	3	3
Scale factor	0.83	0.99
Effective denominator d.f.	9.8	24.7
Scaled Wald statistic	4.82	6.43
Asymptotic χ^2_3 probability	0.0006	0.0002
F probability, scaled statistic	0.025	0.002

The probabilities from the asymptotic tests are far too small and imply rather different conclusions about the statistical significance of the treatment and period effects. Note, in addition, the difference in the scale factor and denominator degrees of freedom for the two tests. These quantities reflect to a great extent the interplay between the covariance structure and the effects being tested. The scale factor for the treatment test is close to one, implying that the covariance structure of these estimated effects is, up to a constant multiplier, almost fixed. This is a consequence of the approximate balance of the treatments with respect to periods. In contrast, the scale factor for the period test is well below one and the denominator degrees of freedom are much smaller than for the treatment test. Here the levels of the factor coincide exactly with the variables defining the covariance structure and so the test is approximately of the form of a Hotelling's T^2. In the absence of treatment effects, this would be an exact Hotelling's T^2 and the scale factor and degrees of freedom would then be equal to 0.85 and 11, close to the observed values.

5.2 *Nodule Isolation Effectiveness Trial*

This example is from an experiment testing the intrinsic nitrogen-fixing ability of rhizobia bacteria from sludge-treated soils. Nitrogen fixation was measured by dry-weight yields of white clover plants that had been inoculated with the bacteria. The soil for the trial was removed from a sludge incorporation field and used in the experiment at three rates: low, medium, and high. Lime was used as a second treatment factor: absent and present. The trial soil was placed into pots in a greenhouse, with 12 pots per treatment, and sown with either white clover or rye grass. After approximately 2 months growth, the plant matter was removed from 6 pots and 50 white clover nodules (isolates) were extracted for each treatment. Each isolate was used to inoculate six white clover seeds, which were then placed in pairs into plant agar test tubes. Hence, for each treatment, there were 50 isolates and for each isolate there were 3 tubes. A further 150 tubes were prepared in the same way, as a control, using a commercial white clover inoculant, TA1. The test tubes were arranged in a glass house in 3 positional groups or blocks, each of 350 tubes, 50 from each of the 7 treatments. Note that the groups do not represent genuine replicates, as the same isolates contribute to each group.

We base the analysis on the dry matter yields from each tube, of which 115 are missing, producing an unbalanced structure. One feature of the measurements is the large component of variability among isolates from the sludge, which is not present among the isolates of the commercial inoculant. It is therefore necessary to incorporate a variance component in the analysis that applies to the treatment isolates only. For the analysis, a conventional mixed model is used with the seven treatments as fixed effects. In additional to the residual component, three other variance components are included: one for position, one for treatments within position, and one for the noncontrol isolates. These are estimated as follows:

Component	Estimate	SE
Residual	12.11	6.73
Position	0.29	0.40
Position/treatment	0.48	0.31
Noncontrol isolates	6.36	0.95

In this example the estimate of precision of the fixed effects should not be much influenced by the small sample adjustment because the design is not far from balance and the variance components play a very small role in the estimation of the fixed effects. However, the presence of random effects associated with some treatments but not others does have implications for the degrees of freedom, and it is interesting to see how these are estimated for the treatment means. These, together with their estimated asymptotic and unadjusted standard errors and degrees of freedom, are presented in Table 6. As expected, the sludge treatment means are more variable than the control mean, reflecting the presence of the additional variance component. Their associated degrees of freedom are correspondingly greater, reflecting the additional information in the extra variance component. The standard errors of treatment differences and their associated degrees of freedom (not presented) show the same pattern, the particular combination (whether sludge treatment versus control or sludge treatment versus sludge treatment) determining the precision and degrees of freedom. The small differences in precision and degrees of freedom among the sludge treatments reflect the small departures from balance in the design due to missing values. The difference in the unadjusted and adjusted standard errors are, in this example, negligible.

Table 6
Nodule isolation effectiveness trial: estimated treatment means, standard errors, and effective degrees of freedom

Treatment	Estimate	SE		d.f.
		Asymptotic	Adjusted	
(1) Sludge absent, lime absent	18.12	0.70	0.70	18.4
(2) Sludge low, lime absent	14.72	0.69	0.69	18.2
(3) Sludge high, lime absent	16.38	0.69	0.69	18.0
(4) Sludge absent, lime present	15.37	0.69	0.69	17.3
(5) Sludge low, lime present	15.02	0.72	0.72	21.1
(6) Sludge high, lime present	15.21	0.71	0.71	20.0
(7) Control	17.03	0.58	0.58	9.1

5.3 *Effect of Nitrogen on a Soil Element*

The third example is an illustration of the use of the adjustment with two nonlinear covariance structures, one with many and one with few parameters. The data are from a series of field experiments exploring the effect of applied nitrogen on the concentration of certain elements in the soil. Here we consider the data on one particular element from a single repeated measurements experiment. Four replicate plots received each of four added levels of nitrogen. Nine equally spaced sampling occasions were used, and many values were missing. Only 2 of the 16 plots had a complete set of 9 measurements and, in total, 32 of the 144 (22%) observations were missing. The missing values were scattered quite evenly over treatment groups and times, and the resulting data set is very unbalanced.

There is little average plot-to-plot variation with these data, but the local correlation over time is quite high. A stationary first-order autoregressive (AR(1)) covariance structure (two parameters) provides a good fit. Preliminary analysis shows no appreciable interaction between nitrogen level and time of measurement, and so an additive fixed effects model has been used. Overall Wald tests of effects associated with time and nitrogen level are presented. As a comparison, the same tests have been made using the first-order ante-dependence (AD(1)) structure (17 parameters). This generalizes the AR(1) structure and so necessarily provides as good a fit but, in this instance, is grossly over-parameterized. The details of the tests are as follows:

Effect	AD(1)		AR(1)	
	N	Time	N	Time
Numerator degrees of freedom	3	8	3	8
Scale factor	1.00	0.79	1.00	0.99
Denominator degrees of freedom	32.9	20.4	34.0	81.4
Scaled Wald statistic	1.46	0.79	2.14	1.25

In this case, none of the effects are statistically significant, whether the asymptotic or adjusted tests are used. Again, however, the influence of the covariance structure and design can be seen in the scale factors and degrees of freedom. Note for the time effect the large increase in denominator degrees of freedom associated with the more parsimonious covariance parameterization and the corresponding difference in scale factor. The treatment main effect statistics involve averages over time and depend less on the covariance parameters, hence their smaller dependence on the choice of structure and associated unit scale factors.

6. Discussion

The appearance of REML procedures in major statistical packages has meant that techniques such as weighted and generalized least squares and the recovery of between-block information are being applied in an increasingly wide range of settings, some involving very small sets of data (see, e.g., Chi, 1991; Brown and Kempton, 1994). It is our belief that the asymptotic nature of these techniques as conventionally described, implemented, and used is not always fully appreciated. For example, if between-block information is recovered in an incomplete block experiment, the asymptotic estimate of standard error will always be smaller than the conventional estimate from the fixed block analysis. The naive user of these methods will get the impression that the former

analysis must be superior. Two questions follow. First, to what extent is the asymptotic approximation of precision, and associated methods of inference, inadequate in any given situation? Second, in what situations is it not worth using generalized least squares as opposed to a simpler unweighted procedure? We have not addressed the second question directly in this paper, but the results presented in Section 2 that bear directly on the first can be used to provide a more meaningful approach to answering it. We propose that the adjusted estimates of precision, together with their associated Wald statistics, be used to compare methods of estimation when samples cannot be regarded as large. The size of the adjustment itself provides a measure of when a sample might be regarded as small and can be used in this way in data analysis as well.

We have seen in the simulation study that there are practically relevant situations in which a small sample adjustment is a necessity, and we have seen that the proposed measures of precision and Wald-type pivots perform acceptably well in a range of settings. Simulation studies of simpler situations not presented here, such as incomplete block designs and other crossover trials, produced the same conclusions.

The calculation of the scale factor for the Wald statistics allows us to include as special cases those settings where the statistics have exact F distributions. Further, it leads to consistency among the degrees of freedom associated with Wald statistics for nested linear combinations of fixed effects. The overall procedure can therefore be applied in an automatic way to construct tests and confidence intervals for fixed effects without first having to separate out special cases, such as analysis of variance F-ratios. All such special forms will be reproduced exactly by the procedure. The degrees of freedom and scale factor also provide some insight into the structure of the data and model.

In the simulations, it was seen that the estimated scale factor is very stable indeed, exhibiting very little variation among runs. The estimated denominator degrees of freedom of the F approximation were more variable than the scale factor, but not excessively so and were well within acceptable limits.

The observed performance of the proposed procedure should not be taken to imply, however, that it will always be so successful, and we conjecture that nonlinear covariance structures are more likely to provide examples where the procedure fails to behave acceptably well. In our experience, the parameterization of a nonlinear structure can have a great influence on the behavior. The Taylor series expansions on which the approximations are based may also prove less acceptable when variance or covariance components are constrained to be positive or positive definite and when there is a nonnegligible probability of estimates falling on a boundary.

Acknowledgements

We are grateful to Dr Brian Cullis for helpful comments and Ms Kellie Munn for permission to use the nodule isolation effectiveness example (both are with New South Wales Agriculture, Australia). We also thank Dr Colin Webster of IACR-Rothamsted for permission to use the data in Section 5.3. We appreciate a referee's comments, which led to a considerable number of improvements.

Résumé

Le maximum de vraisemblance restreint (REML) est maintenant bien établi comme une méthode d'estimation des paramètres du modèle linéaire général Gaussien avec une structure de matrice de covariance, en particulier pour les modèles mixtes. En général la précision des estimateurs et l'inférence pour les effets fixes sont fondés sur leurs distributions asymptotiques dont on sait qu'elles ne conviennent pas dans les cas de petits échantillons. Dans cet article nous présentons une statistique de Wald, ayant aussi une approximation de sa distribution d'échantillonnage par un F, dont on montre qu'elle est performante pour une plage de petits échantillons. Cette statistique utilise un estimateur ajusté de la matrice de covariance ce qui réduit le biais d'échantillonnage. Cette approche présente l'avantage qu'on retrouve à la fois les statistiques et les distributions F dans les situations où on a les distributions exactes, c'est-à-dire pour les statistiques de type T^2 de Hotteling et pour les rapports de Fisher de l'analyse de variance. La performance des statistiques modifiées est mise en évidence avec des études de simulation de quatre différentes analyses avec le REML et les méthodes sont illustrées avec trois exemples.

References

Brown, H. K. and Kempton, R. A. (1994). The application of REML in clinical trials. *Statistics in Medicine* **13,** 1601–1618.

Chi, M. C. (1991). Recovery of inter-block information in cross-over trials. *Statistics in Medicine* **10,** 1115–1122.

Cox, D. R. and Hinkley, D. V. (1974). *Theoretical Statistics.* London: Chapman and Hall.

Dempster, A. P., Laird, N. M., and Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm (with discussion). *Journal of the Royal Statistical Society, Series B* **39,** 1–38.

Gabriel, K. R. (1962). Ante dependence analysis of an ordered set of variables. *Annals of Mathematical Statistics* **33,** 201–212.

Giesbrecht, F. G. and Burns, J. C. (1985). Two-stage analysis based on a mixed model: Large sample asymptotic theory and small-sample simulation results. *Biometrics* **41,** 477–486.

Gilmour, A. R., Thompson, R., and Cullis, B. R. (1995). An efficient algorithm for REML estimation in linear mixed models. *Biometrics* **51,** 1440–1450.

Harville, D. A. (1977). Maximum likelihood approaches to variance component estimation and to related problems. *Journal of the American Statistical Association* **72,** 320–340.

Harville, D. A. (1985). Decomposition of prediction error. *Journal of the American Statistical Association* **80,** 132–138.

Harville, D. A. and Carriquiry, A. L. (1992). Classical and Bayesian predication as applied to an unbalanced mixed linear model. *Biometrics* **48,** 987–1003.

Harville, D. A. and Jeske, D. R. (1992). Mean squared error of estimation or prediction under a general linear model. *Journal of the American Statistical Association* **87,** 724–731.

Hulting, F. L. and Harville, D. A. (1991). Some Bayesian and non-Bayesian procedures for the analysis of comparative experiments and for small area estimation: Computational aspects, frequentist properties, and relationships. *Journal of the American Statistical Association* **86,** 557–568.

Jeske, D. R. and Harville, D. A. (1988). Prediction-interval procedures and (fixed-effects) confidence-interval procedures for mixed linear models. *Communications in Statistics—Theory and Methods* **17,** 1053–1087.

John, J. A. and Eccleston, J. A. (1986). Row-column α-designs. *Biometrika* **73,** 301–306.

Jones, B. and Kenward, M. G. (1989). *The Design and Analysis of Cross-Over Trials.* London: Chapman and Hall.

Kackar, A. N. and Harville, D. A. (1984). Approximations for standard errors of estimators of fixed and random effects in mixed linear models. *Journal of the American Statistical Association* **79,** 853–862.

Kenward, M. G. (1987). A method for comparing profiles of repeated measurements. *Applied Statistics* **36,** 296–308.

Krzanowski, W. J. (1988). *Principles of Multivariate Analysis.* Oxford, U.K.: Oxford University Press.

Laird, N. M. and Ware, J. H. (1982). Random-effects models for longitudinal data. *Biometrics* **38,** 963–974.

McLean, R. A and Sanders, W. L. (1988). Approximating the degrees of freedom for SE's in mixed linear models. *Proceedings of the Statistical Computing Section of the American Statistical Association,* 50–59. New Orleans: American Statistical Association.

Nguyen, N.-K. and Williams, E. R. (1993). An algorithm for constructing optimal resolvable row-column designs. *Australian Journal of Statistics* **35,** 363–370.

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

Robinson, G. K. (1991). That BLUP is a good-thing: The estimation of random effects. *Statistical Science* **6,** 15–51.

Satterthwaite, F. F. (1941). Synthesis of variance. *Psychometrika* **6,** 309–316.

Thompson, R. (1980). Maximum likelihood estimation of variance components. *Mathematische Operationsforschung und Statistik, Series Statistics* **11,** 545–561.

Welham, S. J. and Thompson, R. (1997). A likelihood ratio test for fixed model terms using residual maximum likelihood. *Journal of the Royal Statistical Society, Series B,* in press.

Received November 1994; *revised September* 1996; *accepted February* 1997.