A COMPARISON OF VARIANCE COMPONENT ESTIMATORS

by

R. R. Corbeil and S. R. Searle
Cornell University

BU-563-M June, 1975

Abstract

Explicit solutions are given for the maximum likelihood (ML) and restricted maximum likelihood (REML) equations under normality for four common variance components models with balanced (equal subclass numbers) data. Solutions of the REML equations are identical to analysis of variance (AOV) estimators. The ratio of mean squared errors of REML and ML solutions are also given.

Unbalanced (unequal subclass numbers) data are used in a series of numerical trials to compare ML and REML procedures with 3 other estimation methods using a 2-way crossed classification mixed model with no interaction and 0 or 1 observation per cell. Results are similar to those reported by Hocking and Kutner [1975] for the BIB design. Collectively, these studies and those of Klotz, Milton, and Zacks [1969] point, with few exceptions, to the greater efficiency of ML estimators under a range of experimental settings.

1. Introduction

Maximum likelihood (ML) estimators of variance components are solutions of equations that maximize the likelihood over the positive space of the variance components parameters. Direct solutions of the maximizing equations, ignoring this

---

Paper No. BU-563-M in the Biometrics Unit Mimeo Series, Department of Plant Breeding and Biometry, Cornell University, Ithaca, New York 14853.

positivity requirement, are not necessarily the maximum likelihood estimators. Taking this positivity requirement into account is not always straightforward (e.g., Herbach [1959] and Thompson [1961, 1962]), particularly when considering such properties as bias, sampling variance and mean square error. In order to study these properties of different estimators, this paper uses the phrase "ML estimators" to mean "solutions to the ML equations" (and similarly for REML), thus ignoring the positivity requirement. We acknowledge that solutions are not true ML estimators, but since they are when a set of estimates are all positive we see value in comparing "solutions" with other estimators; and in doing so, calling them ML estimators is clearly a convenience.

Miller [1973] points out that closed form ML estimators of variance components do exist in some cases. We have obtained these and REML estimators (Corbeil and Searle [1974]), their biases and sampling variances and mean square errors for 4 oft-encountered balanced (equal subclass numbers) data models, and compared them using the ratio of their mean squared errors. This is of particular interest because in all 4 models the REML estimators are identical to the long-standing analysis of variance (AOV) estimators.

For further comparison, with unbalanced data, where REML and AOV estimators are not the same, and where closed form ML and REML estimators cannot be derived, numerical studies were made using a 2-way crossed classification, mixed model. The results are presented in section 4.

2. The Models

The general mixed model can be represented as

$$\underset{\sim}{y} = \underset{\sim}{X}\underset{\sim}{\mu} + \sum_{i=1}^{c+1} \underset{\sim}{U}_i \underset{\sim}{b}_i \;, \tag{1}$$

where $\underset{\sim}{b}_{c+1} = \underset{\sim}{e}$, a vector of error terms, and $\underset{\sim}{U}_{c+1} = \underset{\sim}{I}$, an identity matrix. c is the number of variance components excluding the error variance, $\underset{\sim}{y}$ is an N-vector of observations, $\underset{\sim}{X}$ is an $N \times k$ matrix of known constants, $\underset{\sim}{\mu}$ is a k-vector of fixed effects, $\underset{\sim}{U}_i$ is an $N \times m_i$ design matrix of zeros and ones, and, with the usual assumptions, the $\underset{\sim}{b}_i$'s are mutually independent random vectors each having a multivariate normal distribution with mean $\underset{\sim}{0}$ and variance $\sigma_i^2\underset{\sim}{I}$ for $i = 1, 2, \cdots, c + 1$. It is also assumed that $\underset{\sim}{X}$ has full rank k; $N \geq k + c + 1$; and $[\underset{\sim}{X}|\underset{\sim}{U}_i]$ has rank $> k$. Thus $\underset{\sim}{y}$ is distributed as a multivariate normal $(\underset{\sim}{X}\underset{\sim}{\mu}, \sigma^2\underset{\sim}{H})$ where σ^2, σ_e^2, and σ_{c+1}^2 are used interchangeably, and $\underset{\sim}{H} = \sum_{i=1}^{c+1} \gamma_i \underset{\sim}{U}_i \underset{\sim}{U}_i'$ for $\gamma_i = \sigma_i^2/\sigma^2$.

ML estimators $\hat{\sigma}_i^2$ are obtained by solving

$$\hat{\underset{\sim}{\mu}} = (\underset{\sim}{X}'\underset{\sim}{H}^{-1}\underset{\sim}{X})^{-1}\underset{\sim}{X}'\underset{\sim}{H}^{-1}\underset{\sim}{y} , \tag{2}$$

$$\hat{\sigma}^2 = \frac{1}{N}\left[\underset{\sim}{y}'\underset{\sim}{H}^{-1}\underset{\sim}{y} - \hat{\underset{\sim}{\mu}}'(\underset{\sim}{X}'\underset{\sim}{H}^{-1}\underset{\sim}{y})\right] , \tag{3}$$

and

$$\text{trace}\left[\underset{\sim}{U}_i'\underset{\sim}{H}^{-1}\underset{\sim}{U}_i\right] = (\underset{\sim}{y} - \underset{\sim}{X}\hat{\underset{\sim}{\mu}})'\underset{\sim}{H}^{-1}\underset{\sim}{U}_i\underset{\sim}{U}_i'(\underset{\sim}{y} - \underset{\sim}{X}\hat{\underset{\sim}{\mu}})/\hat{\sigma}^2, \tag{4}$$

equations which are given for example as (11), (12), and (9) respectively, in Hemmerle and Hartley [1973]. REML estimators $\tilde{\sigma}_i^2$ are obtained by solving equations (17) and (16) of Corbeil and Searle [1974]:

$$\tilde{\sigma}^2 = \underset{\sim}{y}'\underset{\sim}{T}'(\underset{\sim}{T}\underset{\sim}{H}\underset{\sim}{T}')^{-1}\underset{\sim}{T}\underset{\sim}{y}/(N - k) \tag{5}$$

and

$$\text{trace}\left[\underset{\sim}{U}_i'\underset{\sim}{T}'(\underset{\sim}{T}\underset{\sim}{H}\underset{\sim}{T}')^{-1}\underset{\sim}{T}\underset{\sim}{U}_i\right] = \underset{\sim}{y}'\underset{\sim}{T}'(\underset{\sim}{T}\underset{\sim}{H}\underset{\sim}{T}')^{-1}\underset{\sim}{T}\underset{\sim}{U}_i\underset{\sim}{U}_i'\underset{\sim}{T}'(\underset{\sim}{T}\underset{\sim}{H}\underset{\sim}{T})^{-1}\underset{\sim}{T}\underset{\sim}{y}/\tilde{\sigma}^2, \tag{6}$$

where

$$\underset{\sim}{T} = \sum_{t=1}^{k}{}^{+}\left[\underset{\sim}{I}_{\lambda_t} - n_t^{-1}\underset{\sim}{J}_{\lambda_t} \;\vdots\; -n_t^{-1}\underset{\sim}{1}_{\lambda_t}\right] \tag{7}$$

is a matrix of order $(N - k) \times N$ and where Σ^+ represents a direct sum of matrices. n_t is the number of observations on the t^{th} fixed effect, for $t = 1, \cdots, k$. $N = \sum_{t=1}^{k} n_t$, $\lambda_t = n_t - 1$, $\underset{\sim}{J}_{\lambda_t}$ is a square matrix of order λ_t with every element unity and $\underset{\sim}{1}_{\lambda_t}$ is a vector of λ_t unities.

3. The 4 balanced data models and their results

ML and REML estimators for the balanced data models were obtained by explicitly solving equations (2) - (4) and (5), (6) respectively. Equations (4) and (6) involved especially tedious algebra. Results are shown in Tables 1 and 2.

A brief description of the 4 models is as follows. In all cases the error terms are represented by e with appropriate subscripts, having zero mean and variance covariance matrix $\sigma^2\underset{\sim}{I}$; all random elements are uncorrelated with each other and with error terms, have zero mean and uniform variance σ^2 with appropriate subscript: in short, the well-known conditions of traditional variance components models [e.g., Searle (1971, Chapters 9-11)].

[1] The 1-way classification, random model

$$y_{ij} = \mu + \alpha_i + e_{ij}, \tag{8}$$

for $i = 1, \cdots, a$, and $j = 1, \cdots, n$, with α's random, and σ^2_α and σ^2 to be estimated. In terms of the general model (1), $k = 1$, $c = 1$, $m_1 = a$, and $m_2 = an$. The sums of squares are, using the familiar dot notation for totals, e.g., $y_{i.} = \sum_{j=1}^{n} y_{ij}$,

$$SSA = \frac{1}{n} \sum_{i=1}^{a} y_{i.}^2 - \frac{1}{an} y_{..}^2 \ , \tag{8a}$$

$$SSE = \sum_{i=1}^{a} \sum_{j=1}^{n} y_{ij}^2 - \frac{1}{n} \sum_{i=1}^{a} y_{i.}^2 \ . \tag{8b}$$

As seen in Table 1, ML and REML estimators are very similar in this model, especially for large a, although on the MSE criterion the ML estimator of σ_α^2 is always more efficient.

[2] The 2-way nested classification, random model

$$y_{ij\ell} = \mu + \alpha_i + \beta_{ij} + e_{ij\ell}, \tag{9}$$

for i = 1, $\cdots$, a, j = 1, $\cdots$, b, and ℓ = 1, $\cdots$, n, with α's and β's random and σ_α^2, σ_β^2, and σ^2 to be estimated. In the general model, k = 1, c = 2, m_1 = a, m_2 = ab, and m_3 = abn. The sums of squares are

$$SSA = \frac{1}{bn} \sum_{i=1}^{a} y_{i..}^2 - \frac{1}{abn} y_{...}^2 , \tag{9a}$$

$$SSB(A) = \frac{1}{n} \sum_{i=1}^{a} \sum_{j=1}^{b} y_{ij.}^2 - \frac{1}{bn} \sum_{i=1}^{a} y_{i..}^2 , \tag{9b}$$

$$SSE = \sum_{i=1}^{a} \sum_{j=1}^{b} \sum_{\ell=1}^{n} y_{ij\ell}^2 - \frac{1}{n} \sum_{i=1}^{a} \sum_{j=1}^{b} y_{ij.}^2 . \tag{9c}$$

[3] The 2-way crossed classification, mixed model, no interaction

$$y_{ij\ell} = \mu + \alpha_i + \beta_j + e_{ij\ell} \tag{10}$$

for i = 1, $\cdots$, a, j = 1, $\cdots$, b, and ℓ = 1, $\cdots$, n, with the α's fixed and the β's random, and σ_β^2 and σ_e^2 to be estimated. Here k = a, c = 1, m_1 = b, and m_2 = abn. The sums of squares are

$$SSB = \frac{1}{an} \sum_{j=1}^{b} y_{.j.}^2 - \frac{1}{abn} y_{...}^2 , \tag{10a}$$

$$SSE = \sum_{i=1}^{a} \sum_{j=1}^{b} \sum_{\ell=1}^{n} y_{ij\ell}^2 - \frac{1}{bn} \sum_{i=1}^{a} y_{i\cdot\cdot}^2 - \frac{1}{an} \sum_{j=1}^{b} y_{\cdot j\cdot}^2 + \frac{1}{abn} y_{\cdot\cdot\cdot}^2 . \tag{10b}$$

In this model the ML estimators of both variance components are biased (see Table 2).

Due primarily to this bias, the ratio $MSE(\hat{\sigma}^2)/MSE(\tilde{\sigma}^2) > 1$ (Table 1) under any one of 3 conditions:

$$\begin{aligned} &\text{(i) when } a \geq 6 \text{ and } b > 3 \\ \text{or } &\text{(ii) when } a \geq 7 \text{ and } b > 2 \\ \text{or } &\text{(iii) when } a \geq 8 \text{ and } b \geq 2 . \end{aligned} \tag{10c}$$

Because $[MSE(\hat{\sigma}_\beta^2)/MSE(\tilde{\sigma}_\beta^2)] < 1$, $[MSE(\hat{\sigma}^2) + MSE(\hat{\sigma}_\beta^2)] < [MSE(\tilde{\sigma}^2) + MSE(\tilde{\sigma}_\beta^2)]$ only when the ratio σ_β^2/σ^2 exceeds a constant which is in the neighborhood of 1.

[4] <u>The 2-way crossed classification, mixed model, with interaction</u>

$$y_{ij\ell} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + e_{ijk}, \tag{11}$$

for $i = 1, \cdots, a$, $j = 1, \cdots, b$, and $\ell = 1, \cdots, n$, with the α's fixed, and the β's and $(\alpha\beta)$'s random, and σ_β^2, $\sigma_{\alpha\beta}^2$, and σ^2 to be estimated. Here $k = a$, $c = 2$, $m_1 = b$, $m_2 = ab$, and $m_3 = abn$. The sums of squares are

$$SSB = \frac{1}{an} \sum_{j=1}^{b} y_{\cdot j\cdot}^2 - \frac{1}{abn} y_{\cdot\cdot\cdot}^2 , \tag{11a}$$

just as in (10a),

$$SSAB = \frac{1}{n} \sum_{i=1}^{a} \sum_{j=1}^{b} y_{ij\cdot}^2 - \frac{1}{bn} \sum_{i=1}^{a} y_{i\cdot\cdot}^2 - \frac{1}{an} \sum_{j=1}^{b} y_{\cdot j\cdot}^2 + \frac{1}{abn} y_{\cdot\cdot\cdot}^2 , \tag{11b}$$

and

$$SSE = \sum_{i=1}^{a} \sum_{j=1}^{b} \sum_{\ell=1}^{n} y_{ij\ell}^2 - \frac{1}{n} \sum_{i=1}^{a} \sum_{j=1}^{b} y_{ij.}^2 \ , \qquad (11c)$$

as in (10b).

Again, two of the ML estimators are biased as seen in Table 2, leading to the ratio $[MSE(\hat{\sigma}_{\alpha\beta}^2)/MSE(\tilde{\sigma}_{\alpha\beta}^2)] > 1$ in Table 1 for conditions (10c); but $[MSE(\hat{\sigma}_{\beta}^2)/MSE(\tilde{\sigma}_{\beta}^2)] < 1$. Using some generalized MSE criteria such as $\sum_{i=1} MSE(\hat{\sigma}_i^2)$ it follows that ML is more efficient than REML when $\sum_{i=1}^{3} MSE(\hat{\sigma}_i^2) < \sum_{i=1}^{3} MSE(\tilde{\sigma}_i^2)$ and this is achieved only when $\sigma_{\beta}^2/\sigma_{\alpha\beta}^2$ exceeds some constant.

For large values of b the biases in $\hat{\sigma}_{\alpha\beta}^2$ and $\hat{\sigma}_{\beta}^2$ are small and also, variances of the ML and REML estimators then approach equality.

For balanced data, inversion of $\underset{\sim}{H}$ or $\underset{\sim}{THT}'$ is trivial, whereas it is not for unbalanced data; indeed in the then necessarily iterative procedures that are used it can be an advantage of REML estimators that the matrices to be inverted have order k less than those for ML estimators, Σm_i and $\Sigma m_i + k$ respectively (see Hemmerle and Hartley [1973] and Corbeil and Searle [1974]).

4. Numerical Studies on Unbalanced Data

Data in variance components settings (e.g., genetics) are frequently unbalanced, thus prompting questions about the relative efficiency of different methods of estimation from unbalanced data of different degrees of unbalancedness. Because with balanced data the 2-way crossed classification, mixed model with no interaction [Eqs. (10)] is in some sense a marginal case insofar as the relative efficiency of ML or REML estimators is concerned, this model was selected for a numerical study with unbalanced data of either 0 or 1 observation per cell. There is, for such data, no closed form solution to equations (2) - (4) and (5), (6) and so the study is based on iterative solutions using, respectively, procedures of Hemmerle and Hartley [1973] and Corbeil and Searle [1974].

In order to make comparisons wider than just between ML and REML, we chose 5 methods of estimation: (i) the fitting constants method (Henderson's method 3), (ii) an iterative method after Thompson [1969], (iii) Henderson's method 2, (iv) ML, and (v) REML. The reader is referred to Searle [1971] for details of methods (i) - (iii). Further, we confined ourselves to the case of 6 levels of the fixed effects factor and 10 levels of the random effects factor, choosing these numbers of levels in order to satisfy conditions (10c). Our study is based on 20 basic data sets of $6 \times 10 = 60$ values. From designs having 10%, 30% or 60% of cells empty we selected, in each case, 3 designs at random; and for each empty cell in such designs the corresponding datum of each of the 20 basic data sets was dropped. Mean square errors (MSE) and variances of the resulting estimators $\mathring{\sigma}^2_\beta$ and $\mathring{\sigma}^2$ were then calculated for each of the 3 designs and averaged over those 3 designs; i.e., if $\mathring{\sigma}^2_{\beta,pq}$ is the estimate of σ^2_β in the p^{th} data set for design q, we calculated

$$\begin{aligned} \mathrm{MSE}(\mathring{\sigma}^2_\beta) &= \frac{1}{3} \sum_{q=1}^{3} \Big[\sum_{p=1}^{20} (\mathring{\sigma}^2_{\beta,pq} - \sigma^2_\beta)^2 \Big] / 19 \\ &= \sum_{q=1}^{3} \sum_{p=1}^{20} (\mathring{\sigma}^2_{\beta,pq} - \sigma^2_\beta)^2 / 57. \end{aligned} \tag{12}$$

Thus $\mathrm{MSE}(\mathring{\sigma}^2_\beta)$ represents an estimate of the mean square error of σ^2_β over 3 different samplings (insofar as to which cells may be empty) of 20 data sets. The same was done for $\mathrm{MSE}(\mathring{\sigma}^2_e)$.

The study as so far described was repeated for 4 values of σ^2_β, namely $\frac{1}{4}$, 1, 4, and 9. The error component, σ^2, was taken as $\sigma^2 = 1$ on all occasions. The 20 basic data sets for each of the 4 values of σ_β^2 differed, naturally, depending on σ_β^2. However, the same seed for the random number generation was used for each value of σ_β^2, so that if for $\sigma_\beta^2 = \frac{1}{4}$ and for 2 pseudo random numbers r_1 and r_2 the random part of a datum was $r_1 + \frac{1}{2}r_2$, then for $\sigma_\beta^2 = 1$ the corresponding part is $r_1 + r_2$. This certainly introduces correlation between the data sets for different values of σ_β^2, but if the criteria we are interested in, e.g., MSE($\mathring{\sigma}_\beta^2$), depend in any way on σ_β^2 then possibly this manner of generating data may exhibit that dependence more readily than generating data for one value of σ_β^2 without reference to those data generated for some other value.

Results are shown in Tables 3 and 4. Table 3 shows values of [MSE($\mathring{\sigma}_\beta^2$) + MSE($\mathring{\sigma}^2$)] in the manner of Hocking and Kutner [1975]. Using this expression as a measure of efficiency, scrutiny of Table 3 shows that one procedure, ML estimation, stands out even with such preliminary data as are used here. It is 8% to 21% more efficient than the next best procedure except when $\sigma_\beta^2 = \frac{1}{4}$ with 60% of the cells having no datum. In developing this table we started with values for each of the 3 designs separately and were able to observe that averaging over the 3 designs in each case had no effect on the ranking of the ML method among the other methods. However, it is emphasized that the table entries are not independent of one another. Not only is there correlation between methods, because they were used on the same data, but there is also correlation between results for the 10%, 30%, and 60% of cells empty cases since these represent different subsets of the same basic data. In further support of ML estimation we observed that with another set of 20 separate and independent trials at $\sigma_\beta^2/\sigma^2 = 9$ with 15-25% of the cells empty, the ML procedure was 13% more efficient than the next best one. Bias plays an important role in estimating MSE in these relatively small designs and in these few trials the unbiased estimators (fitting constants and REML estimators) cannot be expected to produce zero deviation from the assigned parameter value.

Expressions for sampling variances are given in Searle [1971, Ch. 11] for the fitting constants estimators and, under large sample theory, are given in Hartley and Rao [1967] and Corbeil and Searle [1974] for the ML estimators and REML estimators respectively. Results of calculating these expressions are shown in Table 4, in the form var($\mathring{\sigma}^2_\beta$) + var($\mathring{\sigma}^2$) and labeled 'theory', alongside results of calculating simple sample variances of the estimates, labeled 'sample'.

Keeping in mind that the 'theory' sampling variances are exact for the fitting constants procedure, several comments about Tables 3 and 4 are in order:

(i) For the fitting constants method, the estimated MSE in Table 3 appears to be inflated by the frequent overestimates of σ^2_β that we observed in our samples and also by a larger than expected estimated variance (Table 4). Note that such samples tend to favor the ML procedures. Recall that in section 3 with balanced data ML characteristically underestimated as a result of its larger than degrees-of-freedom divisor.

(ii) Considering that there are only 10 levels of the random factor and from 24 to 54 observations in each 6×10 design of our data, the sampling variances of the ML estimates calculated from 20 trials are in reasonably good agreement with the sampling variances calculated from large sample theory (Table 4).

(iii) It can be seen in Table 4 that the 'theory' variances calculated for the REML estimates are always equal to or less than the corresponding variances for the fitting constants estimates. However, the theory variances for both ML and REML may well be too low here as a result of using large sample theory on what are really quite small samples.

5. Conclusions

The analytic results presented in section 3 for balanced data favor the ML procedure by the MSE criterion and this appears to carry over to unbalanced data designs at least to the extent presented in section 4, and also in Hocking and Kutner [1975]. It is less clear, however, how the five procedures would compare in larger design settings. Certainly with a large number of random effects, bias would be expected to play a significantly lesser role than in our results here, but the effects of severe imbalance in the data is not predictable. Besides the consideration of efficient estimation, other practical concerns help determine the feasibility of a method of estimation. The fitting constants method and especially Henderson's method 2 are appealing for their simplicity and relative ease of use, particularly in the model dealt with here. Although more cumbersome, the ML and REML methods are uniquely defined for an arbitrary number of components, but again they have stricter distributional requirements than methods derived from the calculations of least squares fitting of data.

6. References

Corbeil, R. R. and S. R. Searle [1974]. Restricted maximum likelihood (REML) estimation of variance components in the mixed model. BU-538-M in the Biometrics Unit Mimeo Series, Cornell University. Technometrics (in press).

Hartley, H. O. and J. N. K. Rao [1967]. Maximum likelihood estimation for the mixed analysis of variance model. Biometrika 54:93-108.

Hemmerle, W. J. and H. O. Hartley [1973]. Computing maximum likelihood estimates for the mixed A.O.V. model using the W-transformation. Technometrics 15:819-832.

Herbach, L. H. [1959]. Properties of type II analysis of variance tests. Ann. Math. Stat., 30:939-959.

Hocking, R. R. and M. H. Kutner [1975]. Some analytical and numerical comparisons of estimators for the mixed A.O.V. model. Biometrics 31:19-27.

Klotz, J. H., R. C. Milton, and S. Zacks [1969]. Mean square efficiency of estimators of variance components. J.A.S.A. 64:1383-1402.

Miller, J. J. [1973]. Asymptotic properties and computation of maximum likelihood estimates in the mixed model of the analysis of variance. Technical Report No. 12, Stanford University.

Searle, S. R. [1971]. Linear Models. Wiley and Sons, New York.

Thompson, R. [1969]. Iterative estimation of variance components for non-orthogonal data. Biometrics 25:767-773.

Thompson, W. A., Jr. [1961]. Negative estimates of variance components: an introduction. Bulletin, International Institute of Statistics 34:1-4.

Thompson, W. A., Jr. [1962]. The problem of negative estimates of variance components. Ann. Math. Stat., 33:273-289.

Table 1: ML and REML estimators in 4 balanced data models

ML estimators	REML estimators (Identical to AOV estimators in these 4 models	$\frac{\text{MSE(ML)}}{\text{MSE(REML)}}$
1. The 1-way, random model [Eqs. (8)]*		
$\hat{\sigma}^2 = \frac{\text{SSE}}{a(n-1)}$	$\tilde{\sigma}^2 = \hat{\sigma}^2$	1
$\hat{\sigma}^2_\alpha = \frac{1}{n}\left[\frac{\text{SSA}}{a} - \hat{\sigma}^2\right]$	$\tilde{\sigma}^2_\alpha = \frac{1}{n}\left[\frac{\text{SSA}}{a-1} - \tilde{\sigma}^2\right]$	< 1
2. The 2-way nested, random model [Eqs. (9)]		
$\hat{\sigma}^2 = \frac{\text{SSE}}{ab(n-1)}$	$\tilde{\sigma}^2 = \hat{\sigma}^2$	1
$\hat{\sigma}^2_\beta = \frac{1}{n}\left[\frac{\text{SSB(A)}}{a(b-1)} - \hat{\sigma}^2\right]$	$\tilde{\sigma}^2_\beta = \hat{\sigma}^2_\beta$	1
$\hat{\sigma}^2_\alpha = \frac{1}{bn}\left[\frac{\text{SSA}}{a} - \frac{\text{SSB(A)}}{a(b-1)}\right]$	$\tilde{\sigma}^2_\alpha = \frac{1}{bn}\left[\frac{\text{SSA}}{a-1} - \frac{\text{SSB(A)}}{a(b-1)}\right]$	< 1
3. The 2-way crossed, mixed model, no interaction [Eqs. (10)]		
$\hat{\sigma}^2 = \frac{\text{SSE}}{b(an-1)}$	$\tilde{\sigma}^2 = \frac{\text{SSE}}{abn-a-b+1}$	> 1 see text
$\hat{\sigma}^2_\beta = \frac{1}{an}\left[\frac{\text{SSB}}{b} - \hat{\sigma}^2\right]$	$\tilde{\sigma}^2_\beta = \frac{1}{an}\left[\frac{\text{SSB}}{b-1} - \tilde{\sigma}^2\right]$	< 1
4. The 2-way crossed, mixed model, with interaction [Eqs. (11)]		
$\hat{\sigma}^2 = \frac{\text{SSE}}{abn(n-1)}$	$\tilde{\sigma}^2 = \hat{\sigma}^2$	1
$\hat{\sigma}^2_{\alpha\beta} = \frac{1}{n}\left[\frac{\text{SSAB}}{(a-1)b} - \hat{\sigma}^2\right]$	$\tilde{\sigma}^2_{\alpha\beta} = \frac{1}{n}\left[\frac{\text{SSAB}}{(a-1)(b-1)} - \tilde{\sigma}^2\right]$	> 1 see text
$\hat{\sigma}^2_\beta = \frac{1}{an}\left[\frac{\text{SSB}}{b} - \frac{\text{SSAB}}{(a-1)b}\right]$	$\tilde{\sigma}^2_\beta = \frac{1}{an}\left[\frac{\text{SSB}}{b-1} - \frac{\text{SSAB}}{(a-1)(b-1)}\right]$	< 1

* See also Klotz, Milton, and Zacks [1969].

Table 2: Bias of ML estimators and variances of ML and REML estimators in 4 balanced data models

ML estimators

$b(\hat{\theta}) \equiv$ bias in $\hat{\theta} = E(\hat{\theta}) - \theta$

1. The 1-way, random model [Eqs. (8)]

$\hat{\sigma}^2$ is unbiased $\qquad v(\hat{\sigma}^2) = \dfrac{2\sigma^4}{a(n-1)}$

$$b(\hat{\sigma}^2_\alpha) = \frac{-(n\sigma^2_\alpha + \sigma^2)}{an} \qquad v(\hat{\sigma}^2_\alpha) = \frac{2}{n^2}\left[\frac{(a-1)(n\sigma^2_\alpha + \sigma^2)^2}{a^2} + \frac{\sigma^4}{a(n-1)}\right]$$

2. The 2-way nested, random model [Eqs. (9)]

$\hat{\sigma}^2$ is unbiased $\qquad v(\hat{\sigma}^2) = \dfrac{2\sigma^4}{ab(n-1)}$

$\hat{\sigma}^2_\beta$ is unbiased $\qquad v(\hat{\sigma}^2_\beta) = \dfrac{2}{an^2}\left[\dfrac{(n\sigma^2_\beta + \sigma^2)^2}{b-1} + \dfrac{\sigma^4}{b(n-1)}\right]$

$$b(\hat{\sigma}^2_\alpha) = \frac{-(bn\sigma^2_\alpha + n\sigma^2_\beta + \sigma^2)}{abn} \qquad v(\hat{\sigma}^2_\alpha) = \frac{2(a-1)(bn\sigma^2_\alpha + n\sigma^2_\beta + \sigma^2)^2}{a^2b^2n^2} + \frac{2(n\sigma^2_\beta + \sigma^2)^2}{ab^2(b-1)n^2}$$

3. The 2-way crossed, mixed model, no interaction [Eqs. (10)]

$$b(\hat{\sigma}^2) = \frac{-(a-1)\sigma^2}{b(an-1)} \qquad v(\hat{\sigma}^2) = \frac{2(abn-a-b+1)\sigma^4}{b^2(an-1)^2}$$

$$b(\hat{\sigma}^2_\beta) = \frac{-n(an-1)\sigma^2_\beta - (n-1)\sigma^2}{bn(an-1)} \qquad v(\hat{\sigma}^2_\beta) = \frac{2(b-1)(an\sigma^2_\beta + \sigma^2)^2}{a^2b^2n^2} + \frac{v(\hat{\sigma}^2)}{a^2n^2}$$

4. The 2-way crossed, mixed model, with interaction [Eqs. (11)]

$\hat{\sigma}^2$ is unbiased $\qquad v(\hat{\sigma}^2) = \dfrac{2\sigma^4}{ab(n-1)}$

$$b(\hat{\sigma}^2_{\alpha\beta}) = \frac{n\sigma^2_{\alpha\beta} + \sigma^2}{bn} \qquad v(\hat{\sigma}^2_{\alpha\beta}) = \frac{2(b-1)(n\sigma^2_{\alpha\beta} + \sigma^2)^2}{(a-1)b^2n^2} + \frac{2\sigma^4}{abn^2(n-1)}$$

$$b(\hat{\sigma}^2_\beta) = \frac{\sigma^2_\beta}{b} \qquad v(\hat{\sigma}^2_\beta) = \frac{2(b-1)}{a^2b^2n^2}\left[(an\sigma^2_\beta + n\sigma^2_{\alpha\beta} + \sigma^2)^2 + \frac{(n\sigma^2_{\alpha\beta} + \sigma^2)^2}{a-1}\right]$$

Table 2 (continued)

REML estimators (unbiased)

(Identical to AOV estimators in these 4 models)

1. The 1-way, random model [Eqs. (8)]

$$v(\tilde{\sigma}^2) = v(\hat{\sigma}^2)$$

$$v(\tilde{\sigma}_\alpha^2) = \frac{2}{n^2}\left[\frac{(n\sigma_\alpha^2 + \sigma^2)^2}{a-1} + \frac{\sigma^4}{a(n-1)}\right]$$

2. The 2-way nested, random model [Eqs. (9)]

$$v(\tilde{\sigma}^2) = v(\hat{\sigma}^2)$$

$$v(\tilde{\sigma}_\beta^2) = v(\hat{\sigma}_\beta^2)$$

$$v(\tilde{\sigma}_\alpha^2) = \frac{2(bn\sigma_\alpha^2 + n\sigma_\beta^2 + \sigma^2)^2}{(a-1)b^2n^2} + \frac{2(n\sigma_\beta^2 + \sigma^2)^2}{ab^2(b-1)n^2}$$

3. The 2-way crossed, mixed model, no interaction [Eqs. (10)]

$$v(\tilde{\sigma}^2) = \frac{2\sigma^4}{abn-a-b+1}$$

$$v(\tilde{\sigma}_\beta^2) = \frac{2(an\sigma_\beta^2 + \sigma^2)^2}{a^2n^2(b-1)} + \frac{v(\tilde{\sigma}^2)}{a^2n^2}$$

4. The 2-way crossed, mixed model, with interaction [Eqs. (11)]

$$v(\tilde{\sigma}^2) = v(\hat{\sigma}^2)$$

$$v(\tilde{\sigma}_{\alpha\beta}^2) = \frac{2(n\sigma_{\alpha\beta}^2 + \sigma^2)^2}{(a-1)(b-1)n^2} + \frac{2\sigma^4}{abn^2(n-1)}$$

$$v(\tilde{\sigma}_\beta^2) = \left(\frac{b}{b-1}\right)^2 v(\hat{\sigma}_\beta^2)$$

Table 3: Calculated values of $[MSE(\hat{\sigma}^2_\beta) + MSE(\hat{\sigma}^2)]$ based on equation (12)

Value of σ^2_β ($\sigma^2 = 1$)	Method of Estimation	Proportion of cells empty		
		10%	30%	60%
$\frac{1}{4}$	Fitting constants	.144	.197	.510
	Iterative	.133	.171	.416
	Henderson 2	.134	.189	.623
	ML	.110	.153	.464
	REML	.131	.176	.458
1	Fitting constants	.448	.525	.969
	Iterative	.448	.501	1.037
	Henderson 2	.452	.548	1.199
	ML	.353	.434	.890
	REML	.438	.546	.981
4	Fitting constants	4.626	4.572	5.816
	Iterative	4.637	4.581	6.208
	Henderson 2	4.512	4.689	6.047
	ML	3.595	3.683	4.965
	REML	4.609	4.644	6.173
9	Fitting constants	22.218	21.426	24.997
	Iterative	21.987	21.397	26.384
	Henderson 2	21.985	21.750	23.903
	ML	17.141	16.990	19.852
	REML	21.177	21.256	25.198

Table 4: Values of $var(\hat{\sigma}_\beta^2) + var(\hat{\sigma}^2)$ from sampling and from theory for 3 methods of estimation for which theoretical expressions are available

Value of σ_β^2 ($\sigma^2 = 1$)	Method of Estimation	Proportion of cells empty					
		10%		30%		60%	
		sample	theory	sample	theory	sample	theory
$\frac{1}{4}$	Fitting constants	.140	.096	.189	.135	.484	.407
	ML	.100	.085	.123	.114	.292	.245
	REML	.127	.096	.161	.135	.394	.364
1	Fitting constants	.438	.372	.508	.445	.927	.847
	ML	.346	.330	.416	.380	.730	.606
	REML	.429	.369	.523	.436	.913	.827
4	Fitting constants	4.533	4.045	4.487	4.399	5.697	5.884
	ML	3.547	3.559	3.653	3.692	4.779	4.261
	REML	4.538	3.961	4.521	4.136	5.924	5.107
9	Fitting constants	21.813	19.296	21.107	20.641	24.649	25.934
	ML	16.893	16.941	16.862	17.214	19.561	18.343
	REML	21.009	18.837	20.833	19.194	24.459	21.104