SIMPLE ESTIMATIONS OF THE VARIANCE COMPONENTS AND THE FIXED AND RANDOM EFFECTS IN MIXED, THREE-STAGE, HIERARCHAL MODELS

C. Philip Cox

Recommended Citation

Cox, C. Philip (1993). "SIMPLE ESTIMATIONS OF THE VARIANCE COMPONENTS AND THE FIXED AND RANDOM EFFECTS IN MIXED, THREE-STAGE, HIERARCHAL MODELS," *Conference on Applied Statistics in Agriculture*. https://doi.org/10.4148/2475-7772.1375

SIMPLE ESTIMATIONS OF THE VARIANCE COMPONENTS
AND THE FIXED AND RANDOM EFFECTS
IN MIXED, THREE-STAGE, HIERARCHAL MODELS

C. Philip Cox
Iowa State University, Department of Statistics
312 Snedecor Hall, Ames, IA 50011-1210

Abstract

FRR (Fixed, Random, Random) hierarchal models in which the first-stage elements are fixed and the second and third-stage elements are random, are used in analyses of comparative experiments and, extensively, in animal breeding contexts where, in the latter, estimates of the second-stage elements and of combinations of them with first-stage elements, are of practical interest. The two procedures, i) empirical BLUP (Best Linear Unbiased Prediction) and ii) a Bayesian approach, used when the ratio of the within-second-stages and the within-third-stages variances is unknown are 'computationally intensive'. When the ratio of the second- to the third-stage variances is large, an alternative and computationally simpler procedure considered here is applicable. This approach provides estimates, including those of the variance components, which jointly maximize likelihood. Another simple method is proposed for further investigation. In all the procedures, estimates of the second-stage elements are obtained by centering or shrinkage translations from the observed means. It is shown that the validity of these adjustments is critically dependent on the distributional assumption made for the second-stage elements. The adjustments will not be centering unless the distribution is Gaussian, in particular, or 'centri-modal' in general.

Key words: Bayesian inference; best linear unbiased prediction; distributional assumptions; fixed-random-random mixed hierarchal models; variance components.

1. Introduction

In the FRR (Fixed, Random, Random) models to be considered, the elements in the first stage are fixed quantities while the elements in the second and third stages are independently (and, here, normally) distributed variates. Data based on such models are commonly analyzed to make comparisons between the fixed quantities when observations on them are subject to two additive sources of variability. Hulting and Harville (1991) have described applications of FRR models in analyses of data from comparative experiments. FRR models are also widely used in animal breeding contexts in which estimations of the fixed elements, of the random second stage elements and of the variance components are of interest. Such applications, together with basic contributions by himself and others, are reviewed and exposited in Henderson (1984). Relatedly, using data on birth weights of lambs and on milk yields of dairy cows, Harville and Carriquiry (1992) have described and assessed two procedures - the empirical BLUP (best linear unbiased prediction) and a Bayesian

Applied Statistics in Agriculture

procedure – available for analyses of FRR data when estimations of linear combinations of fixed and random effects are of interest. Both procedures are described as "<u>computationally intensive</u>", a qualification related to the influence of the two variances or, in effect, of their ratio. For investigations with only one first-stage element Cox (1992) proposed a simple alternative procedure for obtaining estimates – of the overall first stage mean, of the random, quasi-parametric second-stage elements, and of the two variance components – which jointly maximize likelihood. In balanced cases the estimates are explicit and in other cases estimates are obtained without intensive computation. Additionally, a new and very simple, though not maximum likelihood, estimation procedure was noted for further investigation. Extensions of these procedures to FRR experiments with more than one first-stage element are examined here.

2. The model and the estimation equations

Suffixes i, j and k are used for the first, second and third stage identifications, respectively, of a specimen observation y_{ijk}. It is supposed that there are N_1 first-stage quantities, N_{2i} second-stage elements in the ith first stage and N_{3ij} observations in i,jth second stage. The observation is then modeled as:

$$y_{ijk} = \mu_i + (M_{ij} - \mu_i) + (y_{ijk} - M_{ij}) \qquad (1$$

$$k = 1,\ldots,N_{3ij} \; ; \; j = 1,2,\ldots,N_{2i} \; ; \; i = 1,2,\ldots,N_1,$$

where

μ_i is the fixed population mean of the ith first-stage, and

M_{ij} is the jth second-stage mean in the ith first-stage.

The model is completed with the distributional assumptions:

$(M_{ij} - \mu_i) = \delta_{ij} \sim NI(0,\sigma_2^2); \; (y_{ijk} - M_{ij}) = \epsilon_{ijk} \sim NI(0,\sigma_3^2);$

$COV(\delta_{ij},\delta_{ij'}) = 0, \; j \neq j', \; COV(\delta_{ij},\epsilon_{ijk}) = 0.$

Notationally also,

$$N_2 = \sum_{i=1}^{N} N_{2i} \text{ is the total number of second-stage elements.}$$

$$N_{3i} = \sum_{j=1}^{N_{2i}} N_{3ij} \text{ is the total number of third-stage observations in the}$$

ith first stage.

$$N_3 = \sum_{i=1}^{N_1} N_{3i} \text{ is the total number of third-stage observations.}$$

Now, since the observations are obtained conditionally on the particular M_{ij} realizations, the likelihood function can be expressed as:

$$\prod_{i=1}^{N_1} \prod_{j=1}^{N_{2i}} (2\pi\sigma_2^2)^{-1/2} \exp(-1/2\sigma_2^2)(M_{ij} - \mu_i)^2 \text{ x}$$

$$(2\pi\sigma_3^2)^{-N_{3ij}/2} \exp\left\{(-1/2\sigma_3^2) \sum_{k=1}^{N_{3ij}} (y_{ijk} - M_{ij})^2\right\}$$

so that, except for a constant, the log-likelihood is:

$$\ell = (-1/2\sigma_2^2) \sum_i \sum_j (M_{ij} - \mu_i)^2 - (1/2\sigma_3^2) \sum_i \sum_j \sum_k (y_{ijk} - M_{ij})^2 \qquad (2$$

$$-(N_2/2)\ln\sigma_2^2 - (N_3/2)\ln\sigma_3^2$$

The estimation equations $\delta\ell/\delta\mu_i = \delta\ell/\delta M_{ij} = \delta\ell/\delta\sigma_2^2 = \delta\ell/\delta\sigma_3^2 = 0$, then give the estimates $\hat{\mu}_i$, m_{ij}, s_2^2 and s_3^2, of μ_i, M_{ij}, σ_2^2 and σ_3^2 respectively, as the solutions of the following equations:

$$\hat{\mu}_i = \sum_j m_{ij}/N_{2i}, \qquad\qquad i = 1,\ldots,N_1 \qquad (3$$

$$(m_{ij} - \hat{\mu}_i)/s_2^2 = N_{3ij}(\bar{y}_{ij.} - m_{ij})/s_3^2 \qquad (4$$

$$j = 1,\ldots,N_{2i} \; ; \; i = 1,\ldots,N_1$$

$$s_2^2 = N_2^{-1} \sum_i \sum_j (m_{ij} - \hat{\mu}_i)^2 \qquad (5$$

$$s_3^2 = N_3^{-1} \sum_i \sum_j \sum_k (y_{ijk} - m_{ij})^2 \qquad (6$$

where, in (4, $\bar{y}_{ij.}$ is the mean of the N_{3ij} observations in the i,jth third-stage. Next, with:

$$a = s_3^2/s_2^2 \text{ and } w_{ij} = N_{3ij}/(a + N_{3ij}) \qquad (7$$

the equations (4 show that:

$$N_{3ij}(\bar{y}_{ij.} - m_{ij}) = a(m_{ij} - \hat{\mu}_i) \qquad (8$$

and

$$m_{ij} = w_{ij}\bar{y}_{ij.} + (1 - w_{ij})\hat{\mu}_i \tag{9}$$

so that the estimates m_{ij}, of the M_{ij} quantities, are weighted means of the $\bar{y}_{ij.}$ and $\hat{\mu}_i$ quantities.

It also follows from (3 and (9 that

$$\hat{\mu}_i = \sum_j w_{ij}\bar{y}_{ij.} / \sum_j w_{ij} \tag{10}$$

so that, if $a = s_3^2/s_2^2$ is found, the w_{ij}, the $\hat{\mu}_i$ and, from (9, the m_{ij} can all be calculated.

Next, because $(y_{ijk} - m_{ij}) = (y_{ijk} - \bar{y}_{ij.}) + (\bar{y}_{ij.} - m_{ij})$, it follows from (6 that

$$N_3 s_3^2 = \sum_i \sum_j \sum_k \{(y_{ijk} - \bar{y}_{ij.})^2 + (\bar{y}_{ij.} - m_{ij})^2\}$$

$$= W + \sum_i \sum_j N_{3ij}(\bar{y}_{ij.} - m_{ij})^2$$

$$= W + a^2 \sum_i \sum_j (m_{ij} - \hat{\mu}_i)^2/N_{3ij} \tag{11}$$

using (8, where W is the within-second-stages sum of squares in the regular ANOVA.

Further,

$$(m_{ij} - \hat{\mu}_i) = w_{ij}(\bar{y}_{ij.} - \hat{\mu}_i)$$

so that (11 gives

$$N_3 s_3^2 = W + a^2 \sum_i \sum_j w_{ij}^2(\bar{y}_{ij.} - \hat{\mu}_i)^2/N_{3ij} \tag{12}$$

and (5 gives

$$N_2 s_2^2 = \sum_i \sum_j w_{ij}^2(\bar{y}_{ij.} - \hat{\mu}_i)^2 \tag{13}$$

Division of (12 by (13 then gives

$$(N_3 a/N_2)\sum_i \sum_j w_{ij}^2(\bar{y}_{ij.} - \hat{\mu}_i)^2$$

$$= W + a^2 \sum_i \sum_j w_{ij}^2(\bar{y}_{ij.} - \hat{\mu}_i)^2/N_{3ij}$$

which can be equivalently written as the quasi-quadratic equation in $\underline{a}$:

$$a^2 \sum_i \sum_j w_{ij}^2 (\bar{y}_{ij.} - \hat{\mu}_i)^2 / N_{3ij} - a(N_3/N_2) \sum_i \sum_j w_{ij}^2 (\bar{y}_{ij.} - \hat{\mu}_i)^2 + W = 0 \quad (14$$

The special case when $N_1 = 1$

If there is only one first stage, that is $N_1 = 1$, (14 reduces to

$$a^2 \sum_j w_{1j}^2 (\bar{y}_{1j.} - \hat{\mu}_1)^2 / N_{31j} - a(N_3/N_2) \sum_j w_{1j}^2 (\bar{y}_{1j.} - \hat{\mu}_1)^2 + W = 0 \quad (15$$

with $w_{1j} = N_{31j}/(a + N_{31j})$.

Apart from notational differences, equation (15 is exactly that previously obtained and discussed in Cox (1992) for this case.

3. Solution possibilities

For the $N_1 = 1$ case it has been shown, loc. cit., that explicit solutions for $\underline{a}$, and hence the remaining estimates are obtainable:

i) in the equal replication case when N_{3ij} is constant for all j,

ii) when there are only two second-stage elements, that is $N_{21} = 2$, whether or not $N_{311} = N_{312}$,

and that iterative solutions can be obtained in some, though not all, more general cases.

For three-stage investigations with $N_1 \geq 1$ first-stage elements, the equal replication situation analogous to i) above is that N_{3ij} is constant for all i and j. Writing $N_{3ij} = n$, it follows from (7 that $\hat{w}_{ij} = n/(a+n)$ = w, a constant, and hence from (10 $\hat{\mu}_i$ can be directly calculated as

$$\hat{\mu}_i = \sum_j \bar{y}_{ij.}/N_{2i} = \bar{y}_{i..} \quad (16$$

Other simplifications are that $N_3/N_2 = n$ and that $\sum \sum n(\bar{y}_{ij.} - \bar{y}_{i..})^2$ = B_2 obtainable from the ANOVA as the sum of squares between the second-stages within the first-stages. Equation (14 can then be reduced to:

$$a^2(B_2 + W) - an(nB_2 - 2W) + n^2W = 0 \quad (17$$

which is explicitly soluble and corresponds to the equation, obtained for $N_1 = 1$ and equal replication in Cox (1992).

The analogue of case (ii) above is that there are two second stage elements within each first stage, that is $N_{2i} = 2$. It appears, however, that despite some simplifications, an explicitly soluble quadratic equation for $\underline{a}$ is not obtained unless again, all the weights, that is the numbers N_{3ij}, are also equal. Attempts to solve (13 iteratively are therefore indicated for both (ii) and (iii) above.

Applied Statistics in Agriculture

In some cases – when differences between the second-stage means are small relative to the third-stage variability – the quadratic equation will not have real roots. One implication is then that at least some of the second-stage parameters, the μ_i, are superfluous in that the data can be 'explained' in terms of a degenerate model, one with fewer parameters, as discussed for the $N_1 = 1$ case in Cox (1992).

4. Examples

The data in Table 1 are a small part adapted (ignoring age of dam) from those used in Harville and Carriquiry (1992). The two first-stage quantities are distinct population lines, numbers 4 and 5, within each of which there were 2 rams, these being the second-stage elements of interest. The third-stage elements are the birth-weights of lambs, sired by the rams out of, in every case, a different dam.

Since this is a case of equal replication, equation (17 for $\underline{a}$ is directly applicable and, of its solutions, $\underline{a}$ = 0.176 and 2.83, the lower one maximizes the likelihood as noted in Cox (1992). Then, since the weights w_{ij} are equal, the estimates $\hat{\mu}_i$ are, from (16, the simple averages of the means $\bar{y}_{ij\,.}$; the estimates m_{ij} can then be calculated from (9 and the variance estimates from (5 and (6.

For the above data the estimates of μ_1, μ_{11}, μ_{12}; μ_2, μ_{21}, μ_{22}; σ_2^2, σ_3^2 are 10.075, 9.64, 10.51; 11.425, 13.10, 9.75; 1.497, 0.2395, respectively. In a preliminary examination of the iteration procedure in an unbalanced case, the observation $y_{113} = 10.6$ was restored to the data in Table 1. Then, beginning with the trial value $a_0 = 0.18$, two (simple) iterations gave $a_1 = 0.186$ and $a_2 = 0.1868$ and one iteration from $a_0 = 0.19$ gave $a_1 = 0.188$ as indications that the iterations were converging in this case.

The above examples are given to outline the simple computational procedures. In practice, the estimates of the second-stage quantities, μ_{ij}, as the mean weights of progeny, provide estimates of breeding value (Falconer, 1989; Harville and Carriquiry, 1992).

5. A potentially alternative procedure

An alternative procedure initiated in Cox (1992) can be adapted to this, the $N_1 \geq 1$ situation, and in particular very simply for the equal replication case with $N_{3ij} = n$ say. For, because $(\bar{y}_{ij\,.} - \mu_i)/$

$\sqrt{\sigma_2^2 + (\sigma_3^2/n)}$ and $(M_{ij} - \mu_i)/\sqrt{\sigma_2^2}$ are standard Gaussian z-variates, probability plot lines for $\bar{y}_{ij\,.}$ and M_{ij} should be concurrent with ordinate

μ_i and slopes $\sqrt{\sigma_2^2 + (\sigma_3^2/n)}$ and σ_2 respectively. Hence, using $(\cdot)$ to distinguish an order statistic, if $\bar{y}_{i(j)\,.}$ is the ordinate at $z_{(i)}$ to a

should–be–straight line through the $z_{(i)}$, $\bar{y}_{i(j)}$. pairs, and $w = n/(a+n)$, an alternative estimate of M_{ij} can be constructed as

$$\tilde{m}_{ij} = \sqrt{w}\, \tilde{y}_{i(j)}. + (1 - \sqrt{w})\hat{\mu}_i \qquad (18$$

This procedure is clearly very simple, it has not yet, however, been examined fully.

6. Assessments

Most of the assessments made for the $N_1 = 1$ case in Cox (1992) apply, *mutatis mutandis* in the more general case and for completeness they will only be briefly recapitulated here.

i) Because, from (9

$$m_{ij} = w_{ij}\, \bar{y}_{ij}. + (1 - w_{ij})\, \hat{\mu}_i$$

with $0 < w_{ij} < 1$, the centering ('shrinkage') property, that $|m_{ij} - \hat{\mu}_i| < |\bar{y}_{ij}. - \hat{\mu}_i|$ obtains and the amount of centering:

$$|(\bar{y}_{ij}. - \hat{\mu}_i) - (m_{ij} - \hat{\mu}_i)| = (1 - w_{ij})|\bar{y}_{ij}. - \hat{\mu}_i|$$
$$= |\bar{y}_{ij}. - \hat{\mu}_i|(a/a + n_{ij})$$

decreases as n_{ij} increases and as a decreases. Since real solutions of the quadratic equation for a only occur if a is small it is somewhat paradoxical that when there is strong support for centering only small shifts are indicated.

ii) Since the estimates of the M_{ij} depend on the variance ratio a they will be non–linear.

iii) Considerable further research is required to evaluate the procedure, to obtain interval estimations, to make comparisons with centering procedures based on equation (9 and (18 using variance components calculated from ANOVA mean squares and to examine possibilities of using REML methods which, it is conjectured, may relax the conditions required here to give real roots and iterative convergence.

With respect to practical applications, two more general comments are noteworthy. First,

iv) The procedure here only gives real values for a when the between–second–stage sum of squares is large relative to the residual sum of squares, in effect, that is, when the ratio s_2^2/s_3^2 is large and a itself is small. By contrast the empirical BLUP and Bayesian approaches, compared in Harville and Carriquiry (1992) are appropriate when the reciprocal ratio s_3^2/s_2^2 is large in which case the procedure here is inapplicable. In such cases, however, the utility of centering may sometimes be an open question because large values of

s_3^2/s_2^2 can imply that $\sigma_2^2 = 0$ in the ANOVA F-test. For animal breeding applications, Henderson (1984) noted the requirement $s_2^2/s_3^2 \leq 1/3$ in order that $4s_2^2/(s_2^2 + s_3^2)$, as an estimate of heritability, should not exceed unity. If, however, $s_2^2/s_3^2 \leq 1/3$, the ANOVA F-statistic will be $\leq 1 + (N_3/3)$ which, for example if $N_1 = N_2 = 4$ and $N_3 = 3$, may fail to achieve 5% significance. This trouble will be less severe with large data sets for which, however, the validity of other assumptions may be dubious. Finally here, it is to be noted that the inference $\sigma_2^2 = 0$ would be inconsistent with either the presumption of a known, non-zero, ratio or the assumption of a prior distribution for the ratio.

v) An additional consideration arises if, in genetic contexts for example, the ratio s_3^2/s_2^2 is large because of selection in which case the assumption that the second-stage elements have Gaussian distributions may be invalid. Specific examinations of distributional assumptions do not seem to be common in literature for this area and it seems either tacitly taken for granted that the conveniently measured variates of current economical interest, such as animal weights, milk yields, milk compositions, do have Gaussian distributions or, as in the frequently cited text, Henderson (1984), where the issue is last addressed on page 1: '... a commonly used method is to assume as an approximation to the truth that the distribution is multivariate normal... See, for example, Cochran (1937).' The possibility that incorrect assumptions may _de facto_ introduce erroneous attributes to data has been noted in another context. In the present context the distributional assumption for the second-stage elements M_{ij} is critical because the adjustment from $\bar{y}_{ij}$. to m_{ij}, for example that given by (9, involves centering only if the M_{ij} - distribution is 'centri-modal'. This can be appreciated by noting that if $f(M_{ij}-\mu_i)$ is the probability density function for the M_{ij}, the y_{ijk} have a Gaussian distribution and ℓ is the logarithm of the likelihood function, the equation $\delta\ell/\delta m_{ij} = 0$ is:

$$\frac{f'(m_{ij}-\hat{\mu}_i)}{f(m_{ij}-\hat{\mu}_i)} + (N_{3ij}/\sigma_3^2)(\bar{y}_{ij}.-m_{ij}) = 0$$

so that

$$m_{ij} = \bar{y}_{ij}. + (\sigma_3^2/N_{3ij})f'(m_{ij}-\hat{\mu}_i)/f(m_{ij}-\hat{\mu}_i)$$

from which it can be seen that if the M_{ij} are uniformly distributed there is no adjustment and if their distribution is U-shaped de-centering is indicated. As an example of the latter, the arc-sine distribution

$$f(M_{ij}-\mu_i) = \{1 - (M_{ij}-\mu_i)^2\}^{-1/2}, \quad |M_{ij}-\mu_i| \leq 1, \text{ gives}$$

$$m_{ij} = \bar{y}_{ij}. + (\sigma_3^2/N_{3ij})(m_{ij}-\hat{\mu}_i)/\{1 - (m_{ij}-\hat{\mu}_i)^2\}^2$$

so that the adjustment has the sign of $(m_{ij}-\hat{\mu}_i)$.

Summary

Two alternatives to the Bayesian and BLUP (best linear unbiased prediction) procedures commonly used to estimate components of fixed-random-random (mixed) hierarchal models are proposed.

Acknowledgements

I wish to thank colleagues H. T. David and R. A. Groeneveld for examples of U-shaped distributions.

References

Cox, C. P. (1992). Some results on the estimation of normally distributed individual group effects in single classification experiments. Preprint No. 92-2 Statistical Laboratory, Iowa State University, Ames.

Falconer, D. S. (1989). Introduction to Quantitative Genetics. 3rd Edn. John Wiley, New York.

Harville, D. A. and Carriquiry, A. L. (1992). Classical and Bayesian Prediction as Applied to an Unbalanced Mixed Linear Model. Biometrics 48, 987.

Henderson, C. R. (1984). Applications of Linear Models in Animal Breeding. Guelph, Ontario: University of Guelph.

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.

Table 1

Birth weights (lb) of lambs

Line	Sire	Birth weight	Mean ($\bar{y}_{ij.}$)	N_{3ij}
4	1	9.2, 10.6	9.60	2
	2	10.2, 10.9	10.55	2
5	7	13.2, 13.3	13.25	2
	8	9.0, 10.2	9.6	2