SAMPLING PROPERTIES OF ESTIMATES OF (CO)VARIANCE COMPONENTS DUE TO MATERNAL EFFECTS

K. MEYER

Institute for Cell, Animal and Population Biology,
West Mains Road, Edinburgh EH9 3JT, Scotland

and

Animal Genetics and Breeding Unit[1], University of New England,
Armidale, NSW 2351

SUMMARY

Theoretical (large sample) and empirical sampling variances and correlations are examined for animal model analyses fitting genetic and permanent environmental maternal effects, considering different family structures and models of analysis. It is shown that substantial amounts of data are required to obtain accurate estimates, even for experimental data specifically designed to estimate maternal components. The impact of embryo transfer on sampling correlations is illustrated and biases due to a non-zero direct-maternal environmental covariances are examined.

INTRODUCTION

The estimation of maternal effects and the pertaining genetic parameters is inherently problematic as direct and maternal effects are generally confounded. Thompson (1976) suggested that in the presence of maternal effects, sampling variances of direct heritability estimates would be increased by three to five times over those obtained if only direct effects existed. Recently, Restricted Maximum Likelihood (REML) analyses under an animal model including maternal genetic or permanent environmental effects in addition to animals' direct genetic effects have found increasing use. This paper examines sampling properties of estimates of (co)variance components from such analyses.

MATERIAL AND METHODS

Thompson (1976, 1977) described the calculation of the REML likelihood and estimation of variance components using a Method of Scoring (MSC) algorithm for data which can be represented by independent matrices of sums of squares (SS) and crossproducts (CP). This methodology is easily modified for a derivative-free (DF) algorithm and can be employed readily to examine the properties for REML estimates; see Meyer (1992) for details.

Fitting an overall mean as the only fixed effect, consider data consisting of records for f independent families, and let all families be of size n and have the same structure. Assume a record y_j for animal j with dam j' is determined by the animal's (direct) additive genetic value a_j, its dam's maternal genetic effect $m_{j'}$, its dam's maternal environmental effect $c_{j'}$ and a residual error e_j, i.e.

$$y_j = \mu + a_j + m_{j'} + c_{j'} + e_j \qquad (1)$$

with μ the overall mean, $V(a_j) = \sigma^2_A$, $V(m_j) = \sigma^2_M$, $V(c_j) = \sigma^2_C$, $V(e_j) = \sigma^2_E$, $Cov(a_j, m_j) = \sigma_{AM}$, $Cov(e_j, c_{j'}) = \sigma_{EC}$ and all remaining covariances equal to zero. Letting selected effects or covariances be equal to zero, then yields 5 models (non-zero population values of (co)variances given in brackets): Model 1 (M1) fits a_j only ($\sigma^2_A = 40$, $\sigma^2_E = 60$), Model 2 (M2) includes a_j and $m_{j'}$ assuming $\sigma_{AM} = 0$ ($\sigma^2_A = 40$, $\sigma^2_M = 20$, $\sigma^2_E = 60$), Model 3 (M3) is as M2 but allows for σ_{AM} ($\sigma^2_A = 40$, $\sigma^2_M = 20$, $\sigma_{AM} = -5$, $\sigma^2_E = 45$), Model 4 (M4) includes all effects for $\sigma_{EC} = 0$ ($\sigma^2_A = 40$, $\sigma^2_M = 20$, $\sigma_{AM} = -5$, $\sigma^2_C = 15$, $\sigma^2_E = 30$), and Model 5 (M5) was as M4 allowing for $\sigma_{EC} \neq 0$ ($\sigma^2_A = 40$, $\sigma^2_M = 20$, $\sigma_{AM} = -5$, $\sigma^2_c = 15$, $\sigma_{EC} = -4.5, 7.5$, $\sigma^2_E = 30$).

[1]a joint unit with NSW Agriculture and Fisheries; financial support for this work was provided by the AFRC (U.K.) and by MRC grant UNE35

Table 1 : Lower bound sampling errors for estimates of (co)variance components[a]

	Design[a]	Model 1 σ^2_A	Model 1 σ^2_E	Model 4 σ^2_A	Model 4 σ^2_M	Model 4 σ_{AM}	Model 4 σ^2_C	Model 4 σ^2_E
2000	FS2F	4.79	3.85	9.56	22.90	11.23	14.43	5.49
rec.s	B1	5.49	4.73	11.40	18.87	10.60	12.22	6.04
	E1	4.84	3.78	8.82	8.88	6.18	5.80	4.98
10000	FS2F	2.16	1.69	4.26	10.21	5.00	6.43	2.45
rec.s	B1	2.45	2.11	5.09	8.43	4.73	5.46	2.70
	E1	2.16	1.69	3.93	3.96	2.76	2.59	2.22

[a]see text for notation

Two mating structures designed specifically to facilitate the estimation of maternal effects are examined. In Bondari's (1973) design I (B1), parental-half sibs of opposite sex are mated to unrelated animals and 2 offspring are recorded per mating, yielding a family size of n=8 and 10 different types of covariances between relatives (c.f. Thompson 1976). In Eisen's (1967) design I (E1), each family consists of $s = 2$ sires which are full-sibs, each mated to $d_1 = 2$ dams from an unrelated full-sib and $d_2 = 3$ from an unrelated half-sib family, with $m = 2$ offspring per dam, i.e. the family size is $n = 32$ and there are 13 types of covariances betwen relatives. These designs are contrasted with a balanced hierarchical full-sib design (FS2F) with $d = 5$ dams per sire and $m = 2$ offspring per dam, linking two sire families ($s = 2$) by assuming sire 1 and 1 mated to sire 2 are full sibs. The effects of embryo transfer (ET) are illustrated contrasting a simple full-sib family structure (FS1 : $s = 1$, d dams/sire, m offspring per dam) with a design where all md progeny per sire have the same genetic dam but are transferred so that each of the other $d - 1$ dams per sire raises m of her progeny (FS1ET).

Using population values of variance components to construct the matrices of SS/CP for each family, expected values of estimates and their sampling (co)variance matrices can be derived, while sampling them from appropriate Wishart distributions allows empirical results to be obtained.

RESULTS AND DISCUSSION

Sampling errors (SE) for analyses under M1 and M4 are given in Table 1 for the 3 designs and two sizes of data sets. Even for a large data set and designs specially formulated for the estimation of maternal effects, SE amount to 10% or more of the estimates (except for σ^2_E). While differences in accuracy of estimation are small under M1, E1 with the most types of covariances between relatives provides the best estimates under M4.

Table 2 contrasts mean estimates, expected SE and their empirical counterparts for M4 for estimates obtained using a MSC and a DF algorithm. Clearly, constraining estimates to the parameter space or excluding replicates with estimates out of bounds biases estimates and causes empirical SE to be substantially less than expected, in particular for design B1. This demonstrates that large sample theory does not hold at the bounds of the parameter space. Corresponding sampling correlations, as expected from population values and between DF estimates over replicates, are given in Table 3. For both designs, there are strong negative correlations between the maternal components (except σ_{AM} and σ^2_C) and between σ^2_A and σ^2_E. For B1 and E1, deviations between expected and empirical values are considerably smaller than for sampling variances, while some of the corresponding empirical values for FS2F (not shown) deviate substantially from the large sample values, especially for the small data set (N=2000) and correlations involving σ_{AM}.

Bias and mean square error (MSE = SE^2 + bias^2) from ignoring an environmental direct-maternal covariance are shown in Table 4 for design E1. For $\sigma_{EC} = 0$, analysis under M5 rather than M4, i.e. estimating an extra component unnecessarily, increases SE of the maternal components, in particular σ_{AM}. Ignoring a non-zero σ_{EC} has comparatively little effect on estimates and thus MSEs for σ^2_A and σ^2_E, while σ_{AM} is biased substantially. For this example, σ_{EC} needs to exceed 30% of σ^2_C before M5 gives a significant improvement in likelihood over M4 which demonstrates again how inaccurate estimates under this kind of models can be.

Table 2 : Mean estimates of (co)variance components[a] under Model 4 with asymptotic lower bound sampling errors, E(SD), and empirical sampling errors (SD) over 1000 replicates for 2000 records

		Bondari's Design (B1)					Eisen's Design (E1)				
		σ^2_A	σ^2_M	σ_{AM}	σ^2_C	σ^2_E	σ^2_A	σ^2_M	σ_{AM}	σ^2_C	σ^2_E
Mean	MSC[b]	40.21	20.25	-5.20	14.82	29.94	40.50	20.55	-5.27	14.75	29.75
	MSC*	40.69	21.97	-6.48	14.17	29.63	40.60	20.78	-5.45	14.68	29.67
	DF	40.81	21.01	-5.94	14.52	29.61	40.50	20.58	-5.32	14.74	29.72
E(SD)	MSC	11.37	18.79	10.56	12.17	6.03	8.82	8.84	6.16	5.78	4.98
	MSC*	11.39	18.81	10.59	12.18	6.03	8.84	8.86	6.18	5.78	4.99
SD	MSC	11.50	18.69	10.75	12.28	6.08	9.35	9.27	6.53	5.94	5.14
	MSC*	10.84	11.64	8.05	8.06	5.76	9.30	8.69	6.34	5.60	5.09
	DF	10.82	14.36	8.65	9.73	5.73	9.30	9.07	6.44	5.83	5.10

[a]see text for notation

[b]MSC : Method of Scoring estimates, all replicates; MSC* : MSC ignoring replicates out of bounds; 742 (B1) and 976 (E1) replicates; DF : Derivative-free estimates, constrained to parameter space, all replicates

Table 3 : Expected (E) and empirical (S) sampling correlations ($r_{I,J}$[a]correlation between σ^2_I and $\sigma^{(2)}_J$) between (co)variance component estimates under Model 4[b]

		$r_{A,M}$	$r_{A,AM}$	$r_{A,C}$	$r_{A,E}$	$r_{M,AM}$	$r_{M,C}$	$r_{M,E}$	$r_{AM,C}$	$r_{AM,E}$	$r_{C,E}$
B1	E	0.31	-0.64	-0.32	-0.93	-0.83	-0.94	-0.31	0.70	0.61	0.27
	S[c]	0.20	-0.60	-0.22	-0.92	-0.76	-0.89	-0.19	0.56	0.56	0.17
		0.25	-0.62	-0.24	-0.92	-0.81	-0.93	-0.26	0.67	0.60	0.21
E1	E	0.24	-0.71	-0.09	-0.87	-0.65	-0.81	-0.25	0.34	0.65	-0.01
	S	0.25	-0.72	-0.08	-0.88	-0.65	-0.81	-0.24	0.35	0.64	-0.03
		0.25	-0.71	-0.11	-0.88	-0.64	-0.81	-0.27	0.35	0.66	0.01

[a])

[b]see text for notation

[c]1st line : 2000 records, 2nd line : 10,000 records

Tables 5 and 6 give SE of estimates and the corresponding sampling correlations for analyses under M2 and M3 with and without ET. For M2, ET reduces only $SE(\sigma^2_M)$ while it improves the accuracy of estimation markedly when σ_{AM} is fitted (M3), reducing all sampling correlations involving this component considerably. Analogous calculations for M4 show much bigger differences, ET reducing SEs of all components, by a factor of 3 or 4 for σ^2_M, σ_{AM} and σ^2_C, accompanied by markedly lower sampling correlations (Meyer 1992). This emphasizes again that the inaccuracy of parameter estimates from maternal effect models can largely be attributed to the biologically determined sampling correlations, i.e. the partial confounding of direct and maternal effects.

REFERENCES

BONDARI, K., WILLHAM, R.L. and FREEMAN, A.E. (1978) *J. Anim. Sci.* **47** : 358.

EISEN, E.J. (1967) *Can. J. Genet. Cytol.* **9** : 13–22.

MEYER, K. (1992). *Genet. Sel. Evol.* (submitted).

THOMPSON, R. (1976). *Biometrics* **32** : 903–917.

THOMPSON, R. (1977). In *Proc. Int. Conf. on Quantitative Genetics*, E. Pollak, O. Kempthorne and T.B. Bailey, ed.s, Iowa State University Press, Ames, pp. 639–657.

Table 4 : Estimates (Est.) of (co)variance components, lower bound sampling errors (SE) and mean square errors (MSE) from analyses under Models 4 and 5 for different environmental direct-maternal covariances; Eisen's design 1 (8000 records)[a]

		σ^2_A		σ^2_M		σ_{AM}		σ^2_C		σ_{EC}	σ^2_E	
σ_{EC}		M4	M5	M4	M5	M4	M5	M4	M5	M5	M4	M5
0	*SE*	4.4	4.4	4.4	4.7	3.1	4.0	2.9	3.0	2.5	2.5	2.5
	MSE	19.4	19.5	19.6	21.7	9.5	15.7	8.4	8.8	6.1	6.2	6.2
-4.5	Est.	40.8	40.0	23.2	20.0	-9.9	-5.0	15.9	15.0	-4.5	29.7	30.0
	SE	4.5	4.5	4.5	4.7	3.2	4.1	2.9	3.0	2.5	2.5	2.5
	MSE	20.8	20.0	30.9	22.3	34.6	16.8	9.4	8.8	6.3	6.5	6.3
7.5	Est.	39.9	40.0	17.4	20.0	1.6	-5.0	12.4	15.0	7.5	29.6	30.0
	SE	4.3	4.3	4.3	4.5	2.9	3.7	2.8	2.9	2.4	2.4	2.4
	MSE	18.2	18.4	25.3	20.3	51.3	13.8	14.7	8.6	5.6	5.9	5.9

[a]see text for notation

Table 5 : Sampling errors for estimates of (co)variance components and the total genetic variance ($\sigma^2_G=\sigma^2_A+\sigma^2_M/2+3\sigma_{AM}/2$) for data with a hierarchical full-sib design, without (1^{st} line) and with (2^{nd} line) embryo transfer[a]

				Model 2				Model 3				
f	d	m	n	σ^2_A	σ^2_M	σ^2_E	σ^2_G	σ^2_A	σ^2_M	σ^2_{AM}	σ^2_E	σ^2_G
500	5	2	10	2.89	2.05	1.84	2.44	4.39	2.93	2.86	2.95	4.14
				2.94	1.59	2.20	2.53	2.97	1.66	1.47	2.24	3.31
320	4	5	25	3.29	1.70	1.87	2.61	5.08	2.39	3.08	2.88	4.38
				3.56	1.37	2.15	2.75	3.58	1.37	1.45	2.23	3.52

[a]see text for notation

Table 6 : Sampling correlations between (co)variance estimates ($r_{I,J}$[a] for data with a hierarchical full-sib design, without (1^{st} line) and with (2^{nd} line) embryo transfer[b]

				Model 2			Model 3					
f	d	m	n	$r_{A,M}$	$r_{A,E}$	$r_{M,E}$	$r_{A,M}$	$r_{A,AM}$	$r_{A,E}$	$r_{M,AM}$	$r_{M,E}$	$r_{AM,E}$
500	5	2	10	-0.43	-0.58	-0.20	0.28	-0.75	-0.81	-0.70	-0.61	0.76
				-0.12	-0.72	-0.31	-0.06	-0.21	-0.65	-0.35	-0.21	-0.17
320	4	5	25	-0.30	-0.80	-0.06	0.36	-0.77	-0.90	-0.70	-0.46	0.75
				-0.13	-0.85	-0.03	-0.11	-0.21	-0.73	-0.18	0.04	-0.28

[a]correlation between σ^2_I and $\sigma^{(2)}_J$
[b]see text for notation