



**HAL**  
open science

# Bias and sampling covariances of estimates of variance components due to maternal effects

K Meyer

► **To cite this version:**

K Meyer. Bias and sampling covariances of estimates of variance components due to maternal effects. Genetics Selection Evolution, 1992, 24 (6), pp.487-509. hal-00893972

**HAL Id: hal-00893972**

**<https://hal.science/hal-00893972>**

Submitted on 11 May 2020

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

## Bias and sampling covariances of estimates of variance components due to maternal effects

K Meyer \*

*Edinburgh University, Institute for Cell, Animal and Population Biology,  
West Mains Road, Edinburgh EH9 3JT, Scotland, UK ;  
University of New England, Animal Genetics and Breeding Unit,  
Armidale, NSW 2351, Australia*

(Received 13 September 1991; accepted 26 August 1992)

**Summary** – The sampling behaviour of Restricted Maximum Likelihood estimates of (co)variance components due to additive genetic and environmental maternal effects is examined for balanced data with different family structures. It is shown that sampling correlations between estimates are high and that sizeable data sets are required to allow reasonably accurate estimates to be obtained, even for designs specifically formulated for the estimation of maternal effects. Bias and resulting mean square error when fitting the wrong model of analysis are investigated, showing that an environmental dam-offspring covariance, which is often ignored in the analysis of growth data for beef cattle, has to be quite large before its effect is statistically significant. The efficacy of embryo transfer in reducing sampling correlations direct and maternal genetic (co)variance components is illustrated.

**maternal effect / variance component / sampling covariance**

**Résumé** – Biais et covariances d'échantillonnage des estimées de composantes de variance dues à des effets maternels. *Les propriétés d'échantillonnage des estimées du maximum de vraisemblance restreint des variances-covariances dues à des effets maternels génétiques additifs et de milieu sont examinées sur des données d'un dispositif équilibré et avec différentes structures familiales. On montre que les corrélations d'échantillonnage entre les estimées sont élevées et qu'un volume de données important est requis pour obtenir des estimées raisonnablement précises, même avec des dispositifs établis spécifiquement pour estimer des effets maternels. L'étude du biais et de l'erreur quadratique moyenne résultant de l'ajustement d'un modèle incorrect montre qu'une covariance mère-fille due au*

*milieu, souvent ignorée dans l'analyse des données de croissance des bovins à viande, doit être très grande pour que son effet soit statistiquement significatif. L'efficacité du transfert d'embryon pour réduire les corrélations d'échantillonnage entre les variances-covariances génétiques directes et maternelles est illustrée.*

effet maternel / composante de variance / covariance d'échantillonnage

## INTRODUCTION

The importance of maternal effects, both genetic and environmental, for the early growth and development of mammals has long been recognised. For post-natal growth, these represent mainly the dam's milk production and mothering ability, though effects of the uterine environment and extra-chromosomal inheritance may contribute. Detailed biometrical models have been suggested. Willham (1963) distinguished between the animal's and its mother's, *ie* direct and maternal, additive genetic, dominance and environmental effects affecting the individual's phenotype. Allowing for direct-maternal covariances between each of the 3 effects, this gave a total of 9 causal (co)variance components contributing to the resemblance between relatives. Willham (1972) described an extension to include grand-maternal effects and recombination loss.

Estimation of maternal effects and the pertaining genetic parameters is inherently problematic. Unless embryo transfer or crossfostering has taken place, direct and maternal effects are generally confounded. Moreover, the expression of maternal effects is sex-limited, occurs late in life of the female and lags by one generation (Willham, 1980). Methods to estimate (co)variances due to maternal effects have been reviewed by Foulley and Lefort (1978). Early work relied on estimating covariances between relatives separately, equating these to their expectations and solving the resulting system of linear equations. However, this ignored the fact that the same animal might have contributed to different types of covariances and that different observational components might have different sampling variances, *ie* combined information in a non-optimal way. In addition, sampling variances of estimates could not be derived (Foulley and Lefort, 1978).

Thompson (1976) presented a maximum likelihood (ML) procedure which overcomes these problems and showed how it could be applied to designs found in the literature. He considered the ML method most useful when data were balanced due to computational requirements in the unbalanced case. Over the last decade, ML estimation, in particular Restricted Maximum Likelihood (REML) as first described by Patterson and Thompson (1971), has found increasing use in the estimation of (co)variance components and genetic parameters. Especially for animal breeding applications this almost invariably involves unbalanced data. Recently, analyses under the so-called animal model, fitting a random effect for the additive genetic value of each animal, have become a standard procedure. To a large extent, this was facilitated by the availability of a derivative-free REML algorithm (Graser *et al*, 1987) which made analysis involving thousands of animals feasible.

Maternal effects, both genetic and environmental, can be accommodated in animal model analyses by fitting appropriate random effects for each animal or each dam with progeny in the data. Conceptually, this simplifies the estimation of genetic parameters for maternal effects. Rather than having to determine the types of covariances between relatives arising from the data and their expectations, to estimate each of them and to equate them to their expectations, we can estimate maternal (co)variance components in the same way as additive genetic (co)variances with the animal model, namely as variances due to random effects in the model of analysis (or covariances between them). The derivative-free REML algorithm extends readily to this type of analyses (Meyer, 1989).

As emphasised by Foulley and Lefort (1978), estimates of genetic parameters are likely to be imprecise. Thompson (1976) suggested that in the presence of maternal effects, sampling variances of estimates of the direct heritability would be increased 3–5-fold over those which would be obtained if only direct additive genetic effects existed. Special experimental designs to estimate (co)variances due to maternal effects have been described, for instance, by Eisen (1967) and Bondari *et al* (1978). Thompson (1976) applied his ML procedure to these designs and showed that for Bondari *et al*'s (1978) data, estimates of maternal components had not only large standard errors but also high sampling correlations.

In the estimation of maternal effects for data from livestock improvement schemes, non-additive genetic effects and a direct-maternal environmental covariance have largely been ignored. In part, this has been due to the fact that often the types of covariances between relatives available in the data do not have sufficiently different expectations to allow all components of Willham's (1963) model to be estimated. Even for Bondari *et al*'s (1978) experiment, providing 11 types of relationships between animals, Thompson (1976) emphasised that only 7 parameters, 6 (co)variances and a linear function of the direct and maternal dominance variance and the maternal environmental variance, could be estimated. In field data, the contrasts between relatives available are likely to be fewer, thus limiting the scope to separate the various maternal components.

In the analysis of pre-weaning growth traits in beef cattle, components estimated have generally been restricted to the direct additive genetic variance ( $\sigma_A^2$ ), the maternal additive genetic variance ( $\sigma_M^2$ ), the direct-maternal additive genetic covariance ( $\sigma_{AM}$ ), the maternal environmental variance ( $\sigma_C^2$ ) and the residual error variance ( $\sigma_E^2$ ) or a subset thereof; see Meyer (1992) for a recent summary. Using data from an experimental herd which supplied various "unusual" relationships, Cantet *et al* (1988) attempted to estimate all components. There has been concern about a negative direct-maternal environmental covariance ( $\sigma_{EC}$ ) in this case (Koch, 1972) which, if ignored, is likely to bias estimates of the other components and corresponding genetic parameters, in particular the direct-maternal genetic correlation ( $r_{AM}$ ). Summarising literature results in- and excluding information from the dam-offspring covariance, the only observational component affected by  $\sigma_{EC}$ , Baker (1980) reported mean values of  $r_{AM}$  of  $-0.42$  and  $0.0$  for birth weight,  $-0.45$  and  $-0.05$  for daily gain from birth to weaning and  $-0.72$  and  $-0.07$  for weaning weight, respectively.

While the modern methods of analysis together with the availability of high speed computers and the appropriate software make it easier to estimate genetic

parameters due to maternal effects, they might make it all too easy to ignore the inherent problems of this kind of analyses and to ensure that all parameters fitted can be estimated accurately. Unexpected or inconsistent estimates have been attributed to high sampling correlations between parameters or bias due to some component not taken into account without any quantification of their magnitude (eg Meyer, 1992). The objective of this paper was to examine REML estimates of genetic parameters due to maternal effects, investigating both sampling (co)variances and potential bias due to fitting the wrong model of analysis.

## MATERIAL AND METHODS

### Theory

Consider a mixed liner model,

$$\mathbf{y} = \mathbf{X}\mathbf{b} + \mathbf{Z}\mathbf{u} + \mathbf{e} \quad [1]$$

where  $\mathbf{y}$ ,  $\mathbf{b}$ ,  $\mathbf{u}$  and  $\mathbf{e}$  denote the vector of observations, fixed effects, random effects and residual errors, respectively, and  $\mathbf{X}$  and  $\mathbf{Z}$  are the incidence matrices pertaining to  $\mathbf{b}$  and  $\mathbf{u}$ . Let  $\mathbf{V}$  denote the variance matrix of  $\mathbf{y}$ . The REML log likelihood ( $\mathcal{L}$ ) is then

$$\log \mathcal{L} = \text{constant} - 1/2[\log|\mathbf{V}| + \log|\mathbf{X}^*\mathbf{V}^{-1}\mathbf{X}^*| + \mathbf{y}'\mathbf{P}\mathbf{y}] \quad [2]$$

with  $\mathbf{X}^*$  a full-rank submatrix of  $\mathbf{X}$  and

$$\mathbf{P} = \mathbf{V}^{-1} - \mathbf{V}^{-1}\mathbf{X}(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1} \quad [3]$$

For the majority of REML algorithms employed in the analysis of animal breeding data, [2] and its derivatives have been re-expressed in terms arising in the mixed model equations pertaining to [1]. An alternative, based on the principle of constructing independent sums of squares (SS) and crossproducts (CP) of the data as for analyses of (co)variances, has been described by Thompson (1976, 1977). As a simple example, he considered data with a balanced hierarchical full-sib structure and records available on both parents and offspring, showing that the SS within dams, between dams within sires and between sires, as utilised in an analysis of variance (for data on offspring only), could be extended to include information on parents. This was accomplished by augmenting the later 2 by rows and columns for dams and sires, yielding a  $2 \times 2$  and a  $3 \times 3$  matrix, respectively, with the additional elements representing offspring parent CP, and SS/CP among parents; see Thompson (1977) for a detailed description.

More generally, let the data be represented by  $p$  independent matrices of SS/CP  $\mathbf{S}_k$ , each with associated degrees of freedom  $d_k$  ( $k = 1, \dots, p$ ). The corresponding matrices of mean squares and products are then  $\mathbf{M}_k = \mathbf{S}_k/d_k$  with expected values  $\mathbf{V}_k$ , and [2] can be rewritten as (Thompson, 1976):

$$\log \mathcal{L} = \text{constant} - 1/2 \sum_{k=1}^p d_k [\log|\mathbf{V}_k| + \text{tr}(\mathbf{M}_k \mathbf{V}_k^{-1})] \quad [4]$$

In the estimation of (co)variance components,  $\mathbf{V}$  and the matrices  $\mathbf{V}_k$  are usually linear functions of the parameters to be estimated,  $\boldsymbol{\theta} = \{\theta_i\}$  with  $i = 1, \dots, t$ , *ie*

$$\mathbf{V}_k = \sum_{i=1}^t \mathbf{F}_{ki} \theta_i$$

REML estimates of  $\boldsymbol{\theta}$  can then be determined as iterative solutions to

$$\mathbf{B}\hat{\boldsymbol{\theta}} = \mathbf{q} \tag{5}$$

(Thompson, 1976) with  $\mathbf{B} = \{b_{ij}\}$  and  $\mathbf{q} = \{q_i\}$  for  $i, j = 1, \dots, t$ , and

$$b_{ij} = \sum_{k=1}^p d_k \text{tr}(\mathbf{V}_k^{-1} \mathbf{F}_{ki} \mathbf{V}_k^{-1} \mathbf{F}_{kj}) \tag{6}$$

$$q_i = \sum_{k=1}^p d_k \text{tr}(\mathbf{V}_k^{-1} \mathbf{F}_{ki} \mathbf{V}_k^{-1} \mathbf{M}_k) \tag{7}$$

This is an algorithm utilising second derivatives of  $\log \mathcal{L}$ . At convergence, an estimate of the large sample covariance matrix of  $\hat{\boldsymbol{\theta}}$  is given by  $-2\mathbf{B}^{-1}$ . As emphasised by Thompson (1976),  $\mathbf{B}$  is singular if a linear combination of the matrices  $\mathbf{F}_{ki}$  is zero for all  $k$ , which implies that not all parameters can be estimated.

This methodology can be employed readily to examine the properties for REML estimates for various models. Consider data consisting of records for  $f$  independent families. Hence  $\mathbf{V}_k$ ,  $\mathbf{M}_k$  and the  $\mathbf{F}_{ki}$  can be evaluated for one family at a time. If the data are “balanced”, *ie* all families are of size  $n$  and have the same structure, these calculations, involving matrices of size  $n \times n$ , are required only once, *ie*  $p = 1$ . Fitting an overall mean as the only fixed effect, the associated degrees of freedom of  $\mathbf{S}_1$  are then  $f - 1$ .

Let a record  $y_j$  for animal  $j$  with dam  $j'$  be 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  $e_j$ , *ie*:

$$y_j = \mu + a_j + m_{j'} + c_{j'} + e_j \tag{8}$$

with  $\mu$  denoting the overall mean. Assume

$$V(a_j) = \sigma_A^2$$

$$V(m_j) = \sigma_M^2$$

$$V(c_j) = \sigma_C^2$$

$$V(e_j) = \sigma_E^2$$

$$\text{Cov}(a_j, m_j) = \sigma_{AM}$$

$$\text{Cov}(e_j, c_{j'}) = \sigma_{EC}$$

with all remaining covariances equal to zero. Letting, in turn, maternal effects  $m_j$  and  $c_j$  be present or absent and covariances  $\sigma_{AM}$  and  $\sigma_{EC}$  be zero or not, yields a total of 9 models of analysis as summarised in table I.

**Table I.** Random effects fitted and (co)variance components estimated in the 9 models of analysis.

Model No	Fit <sup>a</sup>			Estimate <sup>b</sup>					
	<i>a</i>	<i>m</i>	<i>c</i>	$\sigma_A^2$	$\sigma_M^2$	$\sigma_{AM}$	$\sigma_C^2$	$\sigma_{EC}$	$\sigma_E^2$
1	x			x					x
2	x		x	x			x		x
3	x	x		x	x				x
4	x	x		x	x	x			x
5	x	x	x	x	x		x		x
6	x	x	x	x	x	x	x		x
7	x		x	x			x	x	x
8	x	x	x	x	x		x	x	x
9	x	x	x	x	x	x	x	x	x

<sup>a</sup>*a*: Direct additive genetic effect; *m*: maternal additive genetic effect; *c*: maternal environmental effect; <sup>b</sup> $\sigma_A^2$ : direct additive genetic variance;  $\sigma_M^2$ : maternal additive genetic variance;  $\sigma_{AM}$ : direct maternal additive genetic covariance;  $\sigma_C^2$ : maternal environmental variance;  $\sigma_{EC}$ : maternal environmental dam–daughter covariance;  $\sigma_E^2$ : error variance.

Clearly,  $M_k$  in [4] above represents the contribution of the data to  $\log \mathcal{L}$ , *ie* relates to the “true” model describing the data. Conversely,  $V_k$  is determined by the “assumed” model of analysis, *ie* the effect of fitting an inappropriate model can be examined deriving  $V_k$  under the wrong model. Furthermore, the information contributed by individual records can be assessed by “omitting” these records from the analysis which operationally is simply achieved by setting the corresponding rows and columns in  $V_k$  and  $M_k$  to zero.

**Analyses**

In total, 6 family structures were considered. The first, denoted by FS1, was a simple hierarchical full-sib design with records for both parents and offspring for *f* sires mated to *d* dams each with *m* offspring per dam, *ie* *f* families of size  $n = 1 + d(1 + m)$ . As shown in table II, this yielded only 5 types of covariances between relatives, *ie* not all 9 models of analysis could be fitted. Linking pairs of such families by assuming the sire of family 1 to be a full sib (FS<sub>2</sub>F) or paternal half sib (FS<sub>2</sub>H) to one of the dams mated to sire 2 then added up to 3 further relationships (see table II). With *s* = 2 sires per family, this gave a family size of  $n = 2(1 + d(1 + m))$ .

The fourth design examined was design I of Bondari *et al* (1978) . As depicted in figure 1, this was created by mating 2 unrelated grand-dams to the same grand-sire and recording 1 male and one female offspring for each dam. Paternal half-sibs of opposite sex were then chosen among these 4 animals and each of these

**Table II.** Covariances between relatives arising in the 6 designs considered and their composition<sup>a</sup>.

	Causal (co) variance <sup>b</sup>					Design						
	$\sigma_A^2$	$\sigma_M^2$	$\sigma_{AM}$	$\sigma_C^2$	$\sigma_{EC}$	$\sigma_E^2$	FS1	FS2H	FS2F	B1	B1P	E1
Individual	1	1	1	1	0	1	x	x	x	x	x	x
Full sibs	1/2	1	1	1	0	0	x	x	x	x	x	x
Paternal half sibs	1/4	0	0	0	0	0	x	x	x	x	x	x
Sire-offspring	1/2	0	1/4	0	0	0	x	x	x	x	x	x
Dam-offspring	1/2	1/2	5/4	0	1	0	x	x	x	x	x	x
Paternal uncle <sup>c</sup>	1/4	0	1/4	0	0	0			x	x	x	x
Paternal half uncle <sup>c</sup>	1/8	0	0	0	0	0	x			x	x	
Maternal uncle <sup>c</sup>	1/8	0	1/4	0	0	0			x	x	x	x
Maternal half uncle <sup>c</sup>	1/8	0	1/4	0	0	0	x			x	x	
Single first cousins <sup>d</sup>	1/8	0	0	0	0	0						x
	1/8	0	1/4	0	0	0			x			x
Double first cousins	1/4	1/2	1/2	0	0	0						x
Half first cousins <sup>e</sup>	1/16	0	1/8	0	0	0	x			x	x	
PHS + SFC <sup>f</sup>	3/8	1/2	1/2	0	0	0						x
Three-quarter sibs	5/16	1/4	1/4	0	0	0						x
SFC + HFC <sup>g</sup>	3/16	1/4	1/4	0	0	0						x
Pat grandsire-Off	1/4	0	0	0	0	0					x	
Mat grandsire-Off	1/4	0	1/8	0	0	0					x	
Pat granddam-Off	1/4	0	0	0	0	0					x	
Mat granddam-Off	1/4	1/4	5/8	0	0	0					x	

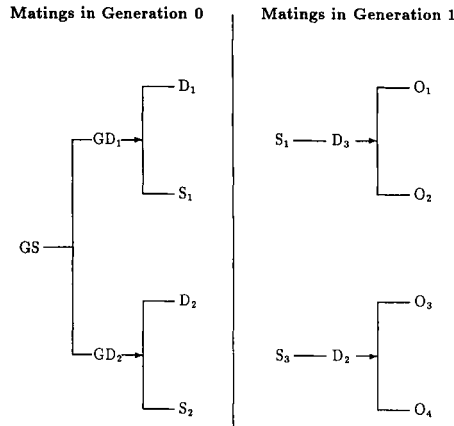
<sup>a</sup> Modified from Thompson (1976); <sup>b</sup> table I for notation; <sup>c</sup> or aunt with nephew or niece; <sup>d</sup> 1st line: sires full sibs; 2nd line: parents of opposite sex full-sibs; <sup>e</sup> parents of opposite sex half sibs; <sup>f</sup> paternal half sibs and single first cousins; <sup>g</sup> single first cousins and half first cousins. Off: offspring.

2 mated to a random, unrelated animal. From each of these matings, 2 offspring were recorded. For Bondari's design I (B1), records on grand-parents and random mates were assumed unknown, yielding a family size of  $n = 8$  and 10 types of relationships between animals. Assuming, for this study, the former to be known then increased the family size for design B1P to  $n = 13$  and added grand-parent offspring covariances to the observational components available.

The last design chosen was Eisen's design 1 (E1). For this, each family consisted of  $s$  sires which were full-sibs and each sire was mated to  $d_1$  dams from an unrelated full-sib family and to  $d_2$  dams from an unrelated half-sib family. Each dam had  $m$  offspring which yielded a family size of  $n = s(1 + d_1 + d_2)(1 + n)$ . As shown in table II, this produced a total of 13 different types of relationships between animals. Figure 2 illustrates the mating structure for this design.

For each design and set of genetic parameters considered, the matrix of mean squares and products,  $M_k$  was constructed assuming the population (co)variance components to be known and "estimates" under various models of analysis were obtained using the Method of Scoring (MSC) algorithm outlined above (see [5]).



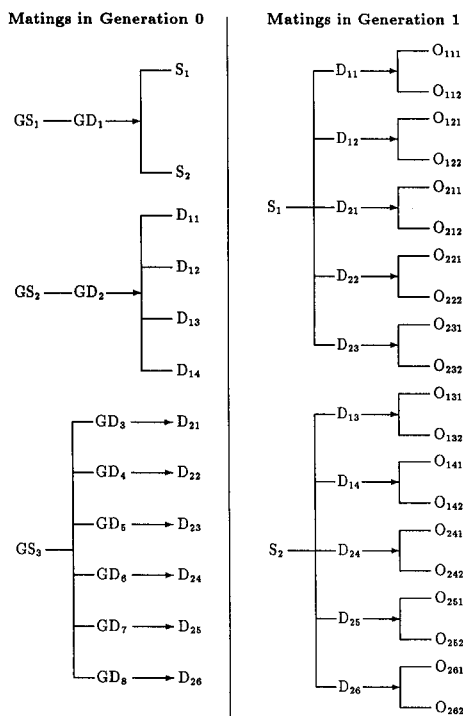


**Fig 1.** Mating plan for Bondari's design I (B1) which includes the 8 records for dams D<sub>1</sub> and D<sub>2</sub>, sires S<sub>1</sub> and S<sub>2</sub> and offspring O<sub>k</sub> (k = 1, . . . , 4); for the augmented design (B1P), records on grand-sire (GS), grand-dams (GD<sub>i</sub>) and the random mates (S<sub>3</sub> and D<sub>3</sub>) are added.

Results obtained in this way are equivalent to those obtained as means over many replicates. Large sample values of sampling errors and sampling correlations between parameter estimates were then obtained from the inverse of the information matrix,  $\mathbf{F} = -2\mathbf{B}^{-1}$ . This is commonly referred to as the formation matrix (Edwards, 1966).

Simulation was carried out by sampling matrices  $\mathbf{M}_k^*$  from an appropriate Wishart distribution with covariance matrix  $\mathbf{M}_k$  and  $f - 1$  degrees of freedom and obtaining estimates of (co)variance components and their sampling variances and correlations using the MSC algorithm. However, this did not guarantee estimates to be within parameter space. Hence, if estimates out of bounds occurred, estimation was repeated using a derivative-free (DF) algorithm, calculating  $\log \mathcal{L}$  as given in [4] and locating its maximum using the Simplex procedure due to Nelder and Mead (1965). This allowed estimates to be restrained to the parameter space simply by assigning a very large, negative value to  $\log \mathcal{L}$  for non-permissible vectors of parameters (Meyer, 1989).

Large sample 95% confidence intervals were calculated as estimate  $\pm 1.96 \times$  the lower bound sampling error obtained from the information matrix. Corresponding likelihood based confidence limits (Cox and Hinkley, 1974) were determined, as described by Meyer and Hill (1992), as the points on the profile likelihood curve for each parameter for which the log profile likelihood differed from the maximum by  $-1.92$ , *ie* the points for which the likelihood ratio test criterion would be equal to the  $\chi^2$  value pertaining to one degree of freedom and an error probability of 5% ( $\chi_{1,5\%}^2 = 3.84$ ).



**Fig 2.** Mating plan for Eisen's design 1 (E1) with  $s = 2$  sires ( $S_i$ ) per family, each mated to  $d_1 = 2$  full-sib ( $D_{1j}$ ) and  $d_2 = 3$  half-sib ( $D_{2j}$ ) dams with  $m = 2$  offspring ( $O_{ijk}$ ) per mating (no records on grand-sires ( $GS_k$ ) and grand-dams ( $GD_m$ )).

## RESULTS AND DISCUSSION

### Sampling covariances

Sampling errors ( $SE$ ) of (co)variance component estimates based on 2 000 records from analyses under Model 6, *ie* when both genetic and environmental maternal effects are present and there is a direct-maternal genetic covariance, are summarised in table III for data sets of 3 designs, and 2 sets of population (co)variances. For comparison, values which would be obtained for equal heritability and phenotypic variance in the absence of maternal effects (Model 1) are given.

The most striking feature of table III is the magnitude of sampling errors even for a quite large data set and for designs like B1 and E1 which have been especially formulated for the estimation of maternal effects components. In all cases,  $SE(\hat{\sigma}_A^2)$  under Model 6 is about twice that under Model 1. FS2F and E1 yield considerably more accurate estimates than B1 under Model 1, with virtually no difference between the former 2 for parameter set I. Estimates from design E1 with the most contrasts between relatives available have an average variance about a quarter of those from FS2F and a third of those from B1 for parameter set I,

*ie* a high direct heritability and low negative direct-maternal correlation, and are comparatively even less variable for parameter set II, *ie* a low direct and medium maternal heritability and a moderate to high positive genetic correlation.

**Table III.** Sampling errors for estimates of (co)variance components<sup>a</sup> based on 2 000 records for 3 designs and 2 sets of genetic parameters<sup>b</sup> (I:  $h^2 = 0.40$ ,  $m^2 = 0.20$ ,  $c_{AM} = -0.05$  and  $c^2 = 0.15$ ; II:  $h^2 = 0.10$ ,  $m^2 = 0.30$ ,  $c_{AM} = 0.10$  and  $c^2 = 0.05$ ) for a model without maternal effects (Model 1) and fitting both genetic and environmental maternal effects (Model 6).

Design <sup>c</sup>	Model 1		Model 6				
	$\sigma_A^2$	$\sigma_E^2$	$\sigma_A^2$	$\sigma_M$	$\sigma_{AM}$	$\sigma_C^2$	$S_E^2$
Set I <sup>d</sup>	40.00	60.00	40.00	20.00	-5.00	15.00	30.00
FS2F	4.79	3.85	9.56	22.90	11.23	14.43	5.49
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
Set II	10.00	90.00	10.00	30.00	10.00	5.00	45.00
FS2F	3.56	4.29	6.90	23.78	10.78	14.91	4.37
B1	4.33	4.96	10.82	18.47	9.95	11.74	5.84
E1	3.36	4.10	6.69	9.43	5.01	5.95	4.10

<sup>a</sup>See table I; <sup>b</sup> $h^2$ : direct heritability;  $m^2$ : maternal heritability;  $c_{AM}$ : direct-maternal genetic covariance as proportion of phenotypic variance;  $c^2$ : maternal environmental variance as proportion of phenotypic variance; <sup>c</sup>FS2F: hierarchical full-sib design with  $s = 2$ ,  $d = 5$  and  $m = 2$ , *ie*  $n = 32$ , with sire 1 and dam 21 full-sibs; B1: Bondari's design 1 ( $n = 8$ ); and E1: Eisen's design 1 with  $s = 2$ ,  $d_1 = 2$ ,  $d_2 = 3$  and  $m = 2$ , *ie*  $n = 32$ ; <sup>d</sup>population value of (co)variance component.

Table IV gives means and empirical deviations of estimates of (co)variance components and their sampling errors under Model 6 for 1 000 replicates for a data set of size 2 000 for parameter set I. While MSC estimates agree closely with the population values, corresponding mean DF estimates are, by definition, biased due to the restriction on the parameter space imposed. This is particularly noticeable for designs FS2F and B1 with 355 and 258 replicates for which estimates needed to be constrained. Overall, however, corresponding estimates of the asymptotic lower bound errors appear little affected: means over all replicate and considering replicates within the parameter space only (MSC\*) show only small differences, except for FS2F, and agree with the population values given in table III. Moreover, standard deviations over replicates for these (not shown) are small and virtually the same for MSC and MSC\*, ranging from 0.22 ( $SE(\hat{\sigma}_E^2)$  for B1) to 1.19 ( $SE(\hat{\sigma}_M^2)$  for FS2F). In turn, empirical standard deviations of MSC estimates agree well with their expected values, being on average slightly higher. Those of the DF estimates, however, are in parts substantially lower, demonstrating clearly that constraining estimates alters their distribution, *ie* that large sample theory does not hold at the bounds of the parameter space.

Table V presents both large sample (LS) and profile likelihood (PL) derived confidence intervals corresponding to parameter estimates in table IV, determined

**Table IV.** Mean estimates of (co)variance components<sup>a</sup> and large sample sampling errors, E(SD), together with empirical sampling errors (SD) over 1 000 replicates for data sets of 3 designs<sup>b</sup> of size 2 000.

		$\sigma_A^2$	$\sigma_M^2$	$\sigma_{AM}$	$\sigma_C^2$	$\sigma_E^2$
FS2F						
Mean	MSC <sup>c</sup>	39.78	19.95	-4.87	15.09	30.17
	MSC* <sup>d</sup>	40.44	22.52	-6.17	13.75	29.72
	DF <sup>e</sup>	40.45	21.15	-5.72	14.44	29.79
E(SD)	MSC	9.46	22.59	11.08	14.25	5.44
	MSC*	9.54	22.94	11.23	14.44	5.48
SD	MSC	9.67	23.49	11.39	14.80	5.60
	MSC*	9.05	12.19	7.22	7.69	5.38
	DF	8.92	15.90	7.99	10.13	5.26
B1						
Mean	MSC	40.21	20.25	-5.20	14.82	29.94
	MSC*	40.69	21.97	-6.48	14.17	29.63
	DF	40.81	21.01	-5.94	14.52	29.61
E(SD)	-MSC	11.37	18.79	10.56	12.17	6.03
	MSC*	11.39	18.81	10.59	12.18	6.03
SD	MSC	11.50	18.69	10.75	12.28	6.08
	MSC*	10.84	11.64	8.05	8.06	5.76
	DF	10.82	14.36	8.65	9.73	5.73
E1						
Mean	MSC	40.50	20.55	-5.27	14.75	29.75
	MSC*	40.60	20.78	-5.45	14.68	29.67
	DF	40.50	20.58	-5.32	14.74	29.72
E(SD)	MSC	8.82	8.84	6.16	5.78	4.98
	MSC*	8.84	8.86	6.18	5.78	4.99
SD	MSC	9.35	9.27	6.53	5.94	5.14
	MSC*	9.30	8.69	6.34	5.60	5.09
	DF	9.30	9.07	6.44	5.83	5.10

<sup>a</sup>see table I for notation; population values are 40, 20, -5, 15 and 30, respectively; <sup>b</sup>see table III; <sup>c</sup>method of scoring estimates, all replicates; <sup>d</sup>method of scoring estimates, ignoring estimates out of bounds; values given pertain to 645, 742 and 976 replicates within bounds for FS2F, B1 and E1, respectively; <sup>e</sup>derivative-free estimates, constrained to parameter space, all replicates.

for the population (co)variances. As noted for other examples by Meyer and Hill (1992), unless bounds of the parameter space are exceeded, predicted lengths of the interval from the 2 methods agree consistently better than values for the position of the confidence bounds. Lower PL limits for  $\hat{\sigma}_M^2$  and  $\hat{\sigma}_C^2$  for designs FS2F and B1 could not be determined (as the log profile likelihood curve to the left of the estimates was so flat that it did not deviate from the maximum by -1.92), and were thus set to zero, the bound of the parameter space. While differences between PL and LS intervals are small for all designs for larger data sets (not shown), considerable deviations occur for the 2 000 record case, particularly for  $\hat{\sigma}_{AM}$  and the upper limits for  $\hat{\sigma}_M^2$  and  $\hat{\sigma}_C^2$  for FS2F and B1.

**Table V.** Large sample and likelihood derived 95% confidence intervals of (co)variance components<sup>a</sup> (calculated for population values) for data sets of 3 designs<sup>b</sup> (2 000 records); together with number of replicates (out of 1 000) for which estimates fall into the respective interval.

		$\sigma_A^2$	$\sigma_M^2$	$\sigma_{AM}$	$\sigma_C^2$	$\sigma_E^2$
FS2F						
Low <sup>c</sup>	LS <sup>d</sup>	22.98	-24.89	-27.00	-13.27	19.25
	PL	21.27	0.00	-25.25	0.00	18.66
UPP <sup>e</sup>	LS	58.73	64.89	17.00	43.27	40.75
	PL	60.51	52.96	10.18	33.38	40.25
Len <sup>f</sup>	LS	37.46	89.78	44.43	56.55	21.51
	PL	37.53	52.96	35.43	33.38	21.59
Sym <sup>g</sup>	PL	0.830	0.607	1.334	0.816	1.107
No repl <sup>h</sup>		962	990	988	990	965
		963	1 000	993	1 000	960
		948	944	938	947	948
B1						
Low	LS	17.66	-16.99	-25.78	-8.96	18.16
	PL	18.09	0.00	-26.12	0.00	18.06
Upp	LS	62.34	56.00	15.78	38.96	41.84
	PL	62.90	52.38	11.98	34.30	41.80
Len	LS	44.69	73.97	41.56	47.92	23.69
	PL	44.81	52.38	38.10	34.30	23.74
Sym	PL	0.957	0.618	1.244	0.777	1.012
No repl		962	993	981	990	961
		958	1 000	984	1 000	962
		950	958	947	947	950
E1						
Low	LS	22.71	2.59	-17.12	3.63	20.23
	PL	24.31	3.79	-18.16	3.65	19.53
Up	LS	57.29	37.41	7.12	26.37	39.77
	PL	59.20	38.62	6.20	26.49	39.27
Len	LS	34.59	34.81	24.24	22.74	19.53
	PL	34.89	34.83	24.36	22.84	19.74
Sym	PL	0.817	0.871	1.176	0.988	1.129
No repl		934	939	937	943	943
		949	942	941	942	945
		938	938	936	939	943

<sup>a</sup>see table I for notation; population values are 40, 20, -5, 15 and 30, respectively; <sup>b</sup>see table III; <sup>c</sup>lower confidence limit; <sup>d</sup>LS: large sample value, PL: profile likelihood derived value; <sup>e</sup>upper confidence limit; <sup>f</sup>length of confidence interval; <sup>g</sup>symmetry of confidence interval = length of lower tail as proportion of length the upper tail; <sup>h</sup>No of replicates with estimates within predicted confidence limit; 1<sup>st</sup> line: estimates constrained within parameter space; PL interval, 2<sup>nd</sup> line: estimates constrained within parameter space, LS interval; 3<sup>rd</sup> line: unconstrained estimates, LS interval.

Corresponding empirical and expected sampling correlations between DF estimates of (co)variance components are contrasted in table VI. Mean expected values over replicates were in all cases equal (to the second or third decimal place) to those derived from the information matrix of the population parameters. While empirical values for larger data sets (not shown) again agree well with their theoretical counterparts, those based on 2 000 records deviate, again reflecting the effect of constraining estimates to the parameter space on their sampling distribution. Deviations are in places considerable for parameter set II, for which estimates from 773, 726 and 553 replicates for FS2F, B1 and E1, respectively, needed to be constrained to the parameter space.

Overall, however, some of the sampling correlations (expected values) show remarkably little variation between designs or differences between parameter sets considered.  $\hat{\sigma}_A^2$  and  $\hat{\sigma}_E^2$  are consistently highly negatively correlated, with values of  $\approx -0.8$  to  $-0.9$ , while  $\hat{\sigma}_A^2$  and the maternal effects components,  $\hat{\sigma}_M^2$  and  $\hat{\sigma}_C^2$ , show comparatively little (though more variable) association, correlations ranging from 0 to  $\approx 0.4$  and 0 to  $\approx -0.3$ , respectively. Similarly, correlations between  $\hat{\sigma}_M^2$  and  $\hat{\sigma}_E^2$  are low and negative and between  $\hat{\sigma}_C^2$  and  $\hat{\sigma}_E^2$  are low and positive or negative and close to zero.

Differences between designs and the amount of information available to separate not only direct and maternal but also maternal genetic and environmental components as apparent in table III, however, are clearly exhibited in the correlations among  $\hat{\sigma}_M^2$ ,  $\hat{\sigma}_{AM}^2$  and  $\hat{\sigma}_C^2$ . While the correlation between  $\hat{\sigma}_M^2$  and  $\hat{\sigma}_{AM}^2$  is as high as  $-0.9$  for FS2F, it is reduced in magnitude to  $-0.8$  for B1 and  $-0.7$  to  $-0.6$  for E1. Correspondingly, a high positive correlation between  $\hat{\sigma}_{AM}^2$  and  $\hat{\sigma}_C^2$  for FS2F ( $\approx 0.9$ ) and B1 ( $\approx 0.7$ ) is reduced substantially for E1 (to  $\approx 0.4$ ).

## FAMILY STRUCTURE

The relationship between family structure and sampling (co)variances for a given number of records is further investigated in table VII, for analyses under Models 1 and 4. The total genetic variance is defined as  $\sigma_G^2 = \sigma_A^2 + 1/2\sigma_M^2 + 3/2\sigma_{AM}^2$  (Willham, 1963), *ie* is the same as  $\sigma_A^2$  for Model 1. As in table III, differences between designs are small for Model 1, but increase with the number of parameters estimated. In particular, including the direct-maternal genetic covariance has a pronounced effect.

For FS1 with only one offspring per dam (column 6), there are no full-sibs in the data. Though the remaining 4 observational covariances between relatives still allow all 4 components under Model 4 to be estimated, this causes an almost complete sampling correlation between  $\hat{\sigma}_M^2$  and both  $\hat{\sigma}_E^2$  and  $\hat{\sigma}_{AM}^2$ , and correspondingly high sampling variances. Conversely, with the same number of offspring but only one dam per sire (column 3) there are no paternal half-sibs. However, as the expectation of the pertaining covariance involves only  $\sigma_A^2$ , estimates of the maternal components are thus much less affected, though  $SE(\hat{\sigma}_A^2)$  and the sampling correlation between  $\hat{\sigma}_A^2$  and  $\hat{\sigma}_E^2$  are largest amongst those for the FS1 designs.

As noted above for Model 1, Bondari's design 1 gives less accurate estimates than most full-sib family structures (unless  $d = 1$  or  $m = 1$ ) for the simpler models of analysis, even at equal family size. While all other cases considered in table VII

Table VI. Expected and empirical sampling correlations between estimates of (co)variance components based on 2 000 records under Model 6<sup>a</sup>.

Design <sup>b</sup>	Correlation between <sup>c</sup>													
	$\hat{\sigma}_A^2$	$\hat{\sigma}_M^2$	$\hat{\sigma}_{A,AM}$	$\hat{\sigma}_A^2$	$\hat{\sigma}_M^2$	$\hat{\sigma}_{M,AM}$	$\hat{\sigma}_A^2$	$\hat{\sigma}_M^2$	$\hat{\sigma}_{M,C}$	$\hat{\sigma}_M^2$	$\hat{\sigma}_C^2$	$\hat{\sigma}_{AM}$	$\hat{\sigma}_C^2$	
	$\tau_{A,M}$	$\tau_{M,M}$	$\tau_{A,AM}$	$\tau_{A,C}$	$\tau_{M,C}$	$\tau_{A,E}$	$\tau_{M,AM}$	$\tau_{M,C}$	$\tau_{M,E}$	$\tau_{M,C}$	$\tau_{M,E}$	$\tau_{AM,C}$	$\tau_{AM,E}$	$\tau_{C,E}$
	Set I <sup>b</sup> : $h^2 = 0.40, m^2 = 0.20, c_{AM} = -0.05, c^2 = 0.15$													
FS2F	E <sup>e</sup>	0.36	-0.63	-0.31	-0.88	-0.91	-0.97	-0.33	0.86	0.23	0.56	0.23	0.56	0.23
	S	0.22	-0.60	-0.16	-0.86	-0.85	-0.94	-0.22	-0.74	0.08	0.55	0.08	0.55	0.08
B1	E	0.31	-0.64	-0.32	-0.93	-0.83	-0.94	-0.31	0.70	0.27	0.61	0.27	0.61	0.27
	S	0.20	-0.60	-0.22	-0.92	-0.76	-0.89	-0.19	0.56	0.17	0.56	0.17	0.56	0.17
E1	E	0.24	-0.71	-0.09	-0.87	-0.65	-0.81	-0.25	0.34	-0.01	0.65	-0.01	0.65	-0.01
	S	0.25	-0.72	-0.08	-0.88	-0.65	-0.81	-0.24	0.35	-0.03	0.64	-0.03	0.64	-0.03
	Set II <sup>b</sup> : $h^2 = 0.10, m^2 = 0.30, c_{AM} = -0.10, c^2 = 0.05$													
FS2F	E	0.34	-0.56	-0.34	-0.80	-0.93	-0.98	-0.28	0.90	0.21	0.46	0.21	0.46	0.21
	S	-0.52	-0.30	0.44	-0.85	-0.29	-0.88	0.35	0.06	-0.38	0.32	-0.38	0.32	-0.38
B1	E	0.26	-0.62	-0.32	-0.92	-0.82	-0.94	-0.26	0.74	0.27	0.59	0.27	0.59	0.27
	S	-0.49	-0.11	0.28	-0.90	-0.45	-0.82	0.39	0.10	-0.27	0.14	-0.27	0.14	-0.27
E1	E	0.02	-0.56	-0.03	-0.80	-0.57	-0.85	-0.07	0.39	-0.08	0.49	-0.08	0.49	-0.08
	S	-0.29	-0.20	0.20	-0.76	-0.55	-0.77	0.17	0.25	-0.29	0.20	-0.29	0.20	-0.29

<sup>a</sup>1 000 replicates, estimates constrained to parameter space; <sup>b</sup>see table III; <sup>c</sup>see table I; <sup>e</sup>E: expected value from formation matrix; S: empirical value.

involve data on only 2 generations, B1P includes records on grand-parents, *ie* 3 generations in total. Though the coefficients of  $\sigma_A^2$  and  $\sigma_{AM}$  in the expectation of the grand-parent offspring covariances are comparatively small (see table II), this clearly reduces the sampling errors of all components estimated and the magnitude of sampling correlations between  $\hat{\sigma}_A^2$  and both  $\sigma_M^2$  and  $\hat{\sigma}_E^2$ . As for FS1, sampling errors for E1 are markedly increased when one or several of the covariances between relatives are missing ( $s = 1$  or  $d_1 = 1$  or  $d_2 = 1$ ), the more the more parameters are estimated. Sampling correlations follow a similar pattern as for FS1. Based on 8 000 records, design 12 provides the most accurate estimates among the 12 data structures examined. Some discussion on the optimal choice of  $s, d_1$  and  $d_2$  for Eisen's (1967) designs is given by Thompson (1976).

### BIAS AND MEAN SQUARE ERROR

So far, only analyses under the "true" model describing the data have been considered. In some instances, analyses are carried out, however, fitting the wrong model. A particular example as discussed above is the analysis of growth traits in beef cattle where an environmental correlation between a dam and her daughter, though assumed to exist, is generally ignored. Figure 3 shows the effect of such environmental covariance ( $b_{EC} = \sigma_{EC}/\sigma_C^2$ ) on the estimates of (co)variance components and the direct-maternal genetic correlation ( $r_{AM}$ ) under Model 6 when the true model describing the data is Model 9, for parameter set I and 3 designs.

While  $\hat{\sigma}_A^2$  and  $\hat{\sigma}_E^2$  are generally little affected, even for large (absolute) values of  $b_{EC}$ , all the maternal components are substantially biased. The pattern of biases differs between designs, reflecting clearly the differences in covariances between relatives available and the information contributed by each of them. For design FS1 (not shown), estimates of  $\sigma_A^2, \sigma_{AM}$  and  $\sigma_E^2$  were unbiased while  $\hat{\sigma}_M^2$  and  $\hat{\sigma}_C^2$  were biased by  $-2\sigma_{EC}$  and  $+2\sigma_{EC}$ , respectively (unless estimates exceeded the bounds of the parameter space and were constrained).

Figure 4 shows the corresponding differences in  $\log \mathcal{L}$  from analyses under models 6 and 9. For the parameter set examined, the magnitude of  $b_{EC}$  needs to exceed 0.3 for design E1 before a likelihood ratio test would be expected to identify a significantly better fit of Model 9 than of Model 6 (at an error probability of 5%; the dashed line in figure 2 marks the significance level). While estimates from EFS2H and EFS2F (not shown) differ little, the higher coefficients in the observational covariances due to the across family relationships for FS2F clearly increase the scope to identify a non-zero  $\sigma_{EC}$ .

The effect of an over- or underparameterized model of analysis on estimates of (co)variances, their lower bound sampling errors and the resulting mean square error (MSE), defined as bias squared plus prediction error variance, is further illustrated in table VIII. Clearly, estimating a (co)variance when it is not present increases the sampling errors of all components unnecessarily. Similarly, when the bias introduced by ignoring a component is small, MSEs under the wrong model may be considerably smaller than under the correct model. As the deviations in  $\log \mathcal{L}$  from the value under Model 9 show, none of the analyses would be expected to identify a  $\sigma_{EC}$  different from zero.



**Table VII.** Large sample sampling errors and sampling correlations for estimates of (co)variance components based on 8 000 records, for 12 data structures and 2 models of analysis.

Design <sup>a</sup>	FS1	FS1	FS1	FS1	FS1	FS1	FS1	FS1	FS1	B1	B1P	E1	E1	E1
$\tau^b$	1 000	800	667	616	500	381	258	1 000	616	250	308	2/2	308	2/2
d	1	3	1	2	5	10	10			2+3	2	5	5	2
m	6	2	10	5	2	1	2				32	26	26	26
n	8	10	12	13	16	21	31	8	13	8	9	10	11	12
	1	2	3	4	5	6	7	8	9	8	9	10	11	12
<i>Sampling errors</i>														
$\sigma^2_A$	M1 <sup>d</sup>	2.45	2.35	2.68	2.43	2.38	2.53	2.46	2.74	2.34	2.41	2.36	2.29	2.29
	M4	5.15	4.31	5.93	4.87	4.39	5.09	4.75	5.07	3.93	4.16	4.72	4.12	4.12
$\sigma^2_M$	M4	2.49	2.97	2.68	2.40	2.93	19.21	2.92	3.16	2.80	2.51	2.27	2.39	2.39
$\sigma_{AM}$	M4	3.37	2.91	3.84	3.10	2.86	8.81	2.92	3.53	2.63	2.69	3.01	2.60	2.60
$\sigma^2_E$	M1	1.85	1.94	1.87	1.82	1.93	2.17	1.95	2.36	1.92	1.89	1.73	1.78	1.78
	M4	2.89	2.95	3.19	2.79	2.95	13.30	3.09	2.92	2.72	2.54	2.68	2.51	2.51
$\sigma^2_G$	M4	4.55	4.15	4.95	4.35	4.14	7.32	4.24	4.68	3.92	3.99	4.26	3.92	3.92
<i>Sampling correlations</i>														
$\tau_{A,E}$	M1	-0.70	-0.71	-0.71	-0.69	-0.71	-0.63	-0.72	-0.80	-0.70	-0.70	-0.66	-0.66	-0.66
	M4	-0.90	-0.80	-0.93	-0.89	-0.81	-0.61	-0.83	-0.87	-0.77	-0.81	-0.89	-0.81	-0.81
$\tau_{A,M}$	M4	0.29	0.28	0.30	0.32	0.28	0.48	0.30	0.04	0.23	0.24	0.35	0.25	0.25
$\tau_{A,AM}$	M4	-0.77	-0.75	-0.77	-0.76	-0.75	-0.65	-0.77	-0.61	-0.70	-0.71	-0.77	-0.74	-0.74
$\tau_{M,E}$	M4	-0.40	-0.61	-0.36	-0.44	-0.61	-0.98	-0.60	-0.22	-0.58	-0.49	-0.46	-0.50	-0.50
$\tau_{M,AM}$	M4	-0.66	-0.69	-0.66	-0.68	-0.70	-0.97	-0.70	-0.70	-0.67	-0.64	-0.67	-0.61	-0.61
$\tau_{E,AM}$	M4	0.74	0.75	0.74	0.74	0.76	0.98	0.78	0.59	0.70	0.66	0.73	0.70	0.70

<sup>a</sup>FS1: hierarchical full-sib family structure, unrelated families; B1: Bondari's design 1; B1P: Bondari's design 1 augmented by records on grandparents and random mates; E1: Eisen's design 1 with 2 sires per family; <sup>b</sup>No of families; <sup>d</sup>: No of dams per sire for FS1; no of full-sib + no of half-sib dams for E1, *m*: No of progeny per dam; *n*: family size; <sup>c</sup>see table I;  $\sigma^2_G$ : total genetic variance; <sup>d</sup>M1: Model 1, M4: Model 4; <sup>e</sup>correlation between (co)variances with the subscripts as given for *r*: see table VI.

## EMBRYO TRANSFER

With a dam affecting the phenotype of her offspring both through half her direct additive genetic value and her maternal genotype as well as her maternal environmental effect, high sampling correlations among the genetic and maternal (co)variance components are invariable, even with the best experimental design. Fortunately, modern reproductive technology allows some of these correlations to be reduced. As a simple illustration, consider the hierarchical full-sib design (FS1) with one sire per family. Assume now that the sire has been mated to only one out of the  $d$  dams with  $md$  full-sib offspring resulting from this mating. Further, assume that each dam raises  $m$  of these offspring (design FS1ET). This gives rise to 3 different dam-offspring covariances, namely:

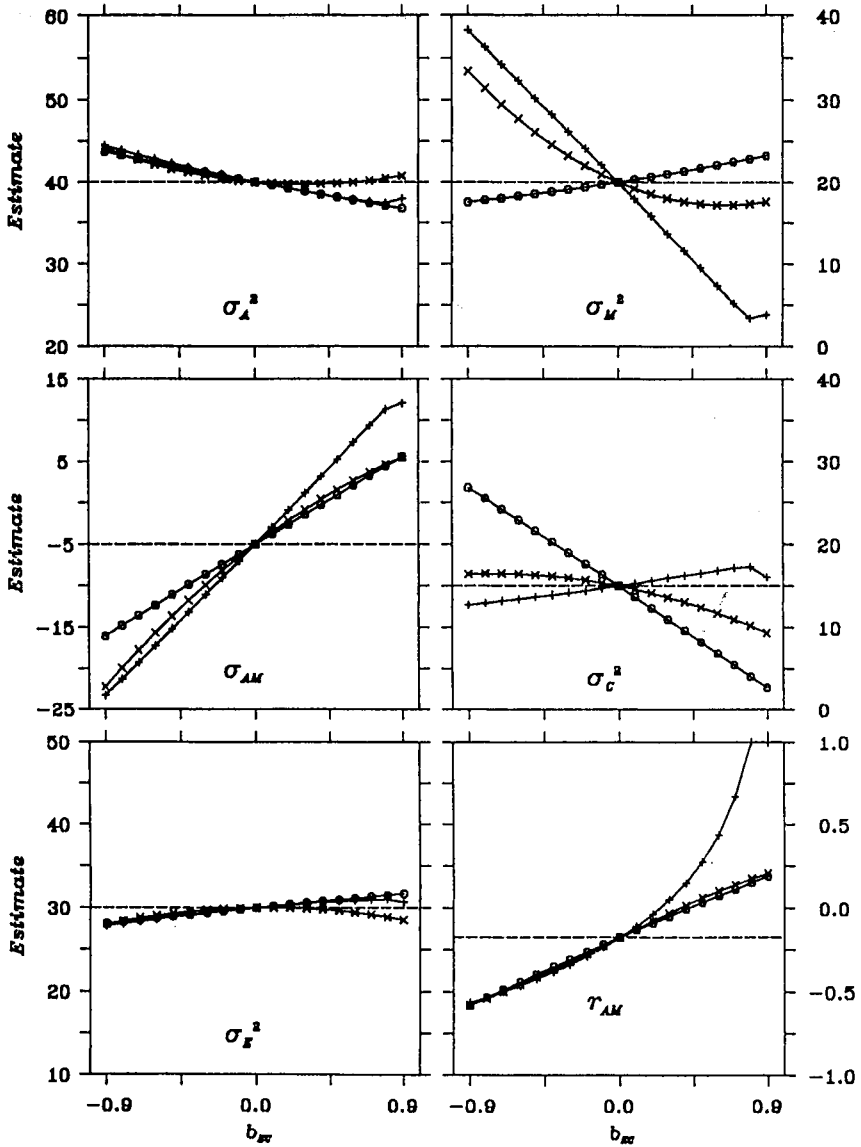
- the "usual" covariance between a dam and her offspring raised by her, with expectation  $\sigma_A^2/2 + 5\sigma_{AM}/4 + \sigma_M^2/2 + \sigma_{EC}$ ;
- the covariance between a dam and her offspring raised by another, recipient dam, with expectation  $\sigma_A^2/2 + \sigma_{AM}/4$ , *ie* the same as the sire-offspring covariance; and
- the covariance between a recipient dam and the offspring (of another dam) which she raised, with expectation  $\sigma_{AM} + \sigma_M^2/2 + \sigma_{EC}$ .

Similarly, we now need to distinguish between 4 types of covariances between full-sibs:

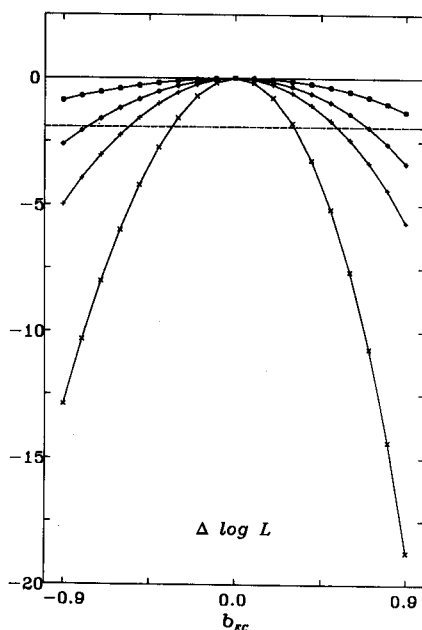
- the "usual" covariance between full-sibs raised by their genetic dam, with expectation  $\sigma_A^2/2 + \sigma_{AM} + \sigma_M^2 + \sigma_C^2$ ;
- the covariance between full-sibs raised by the same recipient dam (not their genetic dam), with expectation  $\sigma_A^2/2 + \sigma_M^2 + \sigma_C^2$ ;
- the covariance between full-sibs raised by different dams, with one of them being their genetic dam, with expectation  $\sigma_A^2/2 + \sigma_{AM}/2$ ; and
- the covariance between full-sibs raised by different recipient dams, none of which is their genetic dam, with expectation  $\sigma_A^2/2$ .

Table IX compares the expected sampling errors for FS1 and FS1ET for 3 family structures and Table X contains the corresponding sampling correlations. Results from analyses under Models 3, 4, 5 (not shown) and 6 were contrasted. For Model 3, with low correlations between  $\hat{\sigma}_M^2$  and the other components, FS1ET yields slightly less accurate estimates than FS1. However, as soon as a direct-maternal genetic covariance is fitted, FS1ET gives considerably smaller sampling errors than FS1 as it reduces the high sampling correlations between  $\hat{\sigma}_A^2$  and  $\hat{\sigma}_M^2$  (Model 4, 5 and 6),  $\sigma_M^2$  and  $\hat{\sigma}_E^2$  (Model 4 and 5),  $\hat{\sigma}_C^2$  and  $\hat{\sigma}_E^2$  (Model 5),  $\hat{\sigma}_A^2$  and  $\hat{\sigma}_{AM}$  (Model 4 and 6),  $\hat{\sigma}_M^2$  and  $\hat{\sigma}_{AM}$  (Model 4 and slightly for Model 6), or  $\hat{\sigma}_A^2$  and  $\hat{\sigma}_C^2$  (Model 5 and 6). Clearly, however, FS1ET does not allow genetic and environmental maternal effects to be separated any better than FS1, and sampling correlations between  $\hat{\sigma}_M^2$  and  $\hat{\sigma}_C^2$  (Model 5 and 6) are still large and negative.

Other designs involving genetically more diverse "litter mates" and related parents or recipients will provide more types of covariances between relatives and thus allow even better separation of genetic and environmental, and direct and maternal effects. While the expectation of all observational components in table II which involve  $\sigma_M^2$  also include  $\sigma_A^2$  and  $\sigma_{AM}$ , the covariance between 2 unrelated animals, for instance, raised by different recipient dams (unrelated to



**Fig 3.** Estimates of (co)variance components ( $\sigma_A^2$ : direct additive genetic variance,  $\sigma_M^2$ : maternal additive genetic variance,  $\sigma_{AM}$ : direct-maternal additive genetic covariance,  $\sigma_E^2$ : error variance) and the direct-maternal genetic correlation ( $r_{AM}$ ) based on 8 000 records and from analyses under Model 6, ignoring a direct-maternal environmental covariance ( $b_{EC} = \sigma_{EC}/\sigma_C^2$ ). (“o”: Hierarchical full-sib design with  $f = 250$ ,  $s = 2$ ,  $d = 5$  and  $m = 2$ , sire 1 and dam 1 mated to sire 2 are paternal half sibs; “x”: Eisen’s design 1 with  $f = 250$ ,  $s = 2$ ,  $d_1 = 2$ ,  $d_2 = 3$  and  $m = 2$ ; “+”: Bondari’s design 1).



**Fig 4.** Difference in log likelihood from analyses under Model 6 and Model 9 for different values of a direct-maternal environmental covariance ( $b_{EC}$ ). ("o": Hierarchical full-sib design with  $f = 250$ ,  $s = 2$ ,  $d = 5$  and  $m = 2$ , sire 1 and dam 1 mated to sire 2 are paternal half sibs; " $\wedge$ ": Hierarchical full-sib design with  $f = 250$ ,  $s = 2$ ,  $d = 5$  and  $m = 2$ , sire 1 and dam 1 mated to sire 2 are full sibs; " $\times$ ": Eisen's design 1 with  $f = 250$ ,  $s = 2$ ,  $d_1 = 2$ ,  $d_2 = 3$  and  $m = 2$ ; "+": Bondari's design 1).

them) which are full-sibs or maternal half-sibs, is solely due to maternal genetic effects (expectation  $\sigma_M^2/4$ ).

## CONCLUSIONS

It has been shown that estimates of (co)variance components are subject to large sampling variances and high sampling correlations, even for a "reduced" model ignoring dominance effects and family structures providing numerous types of covariances between relatives which have been specifically designed for the estimation of maternal effects. For small data sets and models of analysis fitting both genetic and maternal environmental effects or a direct-maternal covariance, this frequently induces the need to constrain estimates to the parameter space. Consequently, large sample theory predictions of sampling errors and correlations estimates do not agree with empirical results. Further research is required to evaluate the implications of such large sampling (co)variances on the accuracy of selection indexes including both direct and maternal effects, *ie* the expected loss in selection response because inaccurately estimated parameters have been used deriving index weights.

**Table VIII.** Estimates of (co)variance components<sup>a</sup>, lower bound sampling errors and resulting mean square error from analyses under Models 6, 7, 8 and 9 when the true model describing the data is Model 9 or 6; Eisen's design 1: 250 families with  $s = 2$ ,  $d_1 = 2$ ,  $d_2 = 3$  and  $m = 2$  (8 000 records).

		$b_{EC} = 0^b$			$b_{EC} = -0.30$			$b_{EC} = 0.50$		
		<i>Est</i> <sup>c</sup>	SE <sup>d</sup>	MSE <sup>e</sup>	<i>Est</i>	SE	MSE	<i>Est</i>	SE	MSE
$\sigma_A^2$	6	40.00	4.40	19.39	40.79	4.50	20.84	39.85	4.26	18.18
	7	40.71	3.21	10.84	40.96	3.23	11.34	40.36	3.17	10.20
	8	36.71	3.47	22.85	37.73	3.49	22.89	36.67	3.42	22.74
$\sigma_M^2$	9	40.00	4.41	19.47	40.00	4.47	19.99	40.00	4.29	18.44
	6	20.00	4.43	19.64	23.22	4.53	30.86	17.35	4.27	25.32
	8	16.11	3.43	26.88	16.07	3.40	27.04	16.20	3.45	26.37
$\sigma_{AM}$	9	20.00	4.66	21.68	20.00	4.73	22.34	20.00	4.51	20.31
	6	-5.00	3.09	9.52	-9.92	3.23	34.58	1.57	2.85	51.28
	9	-5.00	3.96	15.71	-5.00	4.09	16.76	-5.00	3.71	13.80
$\sigma_C^2$	6	15.00	2.89	8.38	15.94	2.93	9.44	12.40	2.82	14.72
	7	28.16	1.71	176.38	28.21	1.72	177.48	28.06	1.67	173.44
	8	15.50	2.94	8.89	15.50	2.94	8.88	15.49	2.91	8.71
	9	15.00	2.96	8.79	15.00	2.96	8.78	15.00	2.93	8.60
$\sigma_{EC}$	7	0.38	1.90	3.74	-4.43	1.92	3.70	8.28	1.82	3.92
	8	-2.00	1.92	7.68	-6.53	1.92	7.77	5.56	1.88	7.31
	9	0.00	2.47	6.10	-4.50	2.52	6.33	7.50	2.35	5.55
$\sigma_E^2$	6	30.00	2.49	6.18	29.72	2.53	6.49	29.63	2.41	5.94
	7	30.18	2.01	4.08	29.96	2.01	4.06	30.53	2.00	4.29
	8	31.60	2.10	6.98	31.61	2.11	7.03	31.59	2.08	6.87
$\Delta \log \mathcal{L}^f$	9	30.00	2.49	6.18	30.00	2.52	6.33	30.00	2.43	5.90
	6	0			-1.57			-5.17		
	7	-14.44			-14.50			-14.63		
	8	-0.86			-0.80			-0.98		

<sup>a</sup>See table I for notation; <sup>b</sup> $b_{EC} = \sigma_{EC}/\sigma_C^2$ ; <sup>c</sup>estimate; <sup>d</sup>sampling error; <sup>e</sup>mean square error; <sup>f</sup>difference in log likelihood to analysis under Model 9.

The efficiency of search procedures used in derivative-free REML algorithms is highly dependent on the correlation structure of the parameters to be estimated, being most effective if these are uncorrelated. The fact that expected sampling correlations between some components for a given model of analysis varied little between designs (see table VI) suggested that a reparameterisation to linear functions of the (co)variance components might improve the convergence rate of such algorithms. Inspection of eigenvalues and eigenvectors of the formation matrices, however, failed to identify any general guidelines.

Examination of bias, sampling variances and resulting mean square errors when fitting the wrong model of analysis showed that, in some instances, ignoring some component(s) can lead to considerably smaller MSE without biasing the (co)variances estimated substantially or reducing the likelihood significantly over that under the true model. In particular, investigating the effect of ignoring an environmental, direct-maternal covariance for a parameter set which might be

**Table IX.** Approximate lower bound sampling errors for estimates of (co)variance components<sup>a</sup> from data with a hierarchical full-sib design with (FS1ET) and without (FS1) embryo transfer, for 3 models of analyses.

	<i>Design FS1</i>			<i>Design FS1ET</i>		
f <sup>b</sup>	500	320	258	500	320	258
d <sup>c</sup>	5	4	10	5	4	10
d <sup>d</sup>	2	5	2	2	5	2
n <sup>e</sup>	16	25	31	16	25	31
<i>Model 3</i>						
$\sigma_A^2$	2.89	3.29	3.02	2.94	3.56	3.36
$\sigma_M^2$	2.05	1.70	2.04	1.59	1.37	1.53
$\sigma_E^2$	1.84	1.87	1.84	2.20	2.15	2.41
$\sigma_G^2$	2.44	2.61	2.49	2.53	2.75	2.70
<i>Model 4</i>						
$\sigma_A^2$	4.39	5.08	4.75	2.97	3.58	3.37
$\sigma_M^2$	2.93	2.39	2.92	1.66	1.37	1.60
$\sigma_{AM}$	2.86	3.08	2.92	1.47	1.45	1.46
$\sigma_E^2$	2.95	2.88	3.09	2.24	2.23	2.51
$\sigma_G^2$	4.14	4.38	4.24	3.31	3.52	3.50
<i>Model 6</i>						
$\sigma_A^2$	5.52	6.74	5.29	3.25	3.89	3.60
$\sigma_M^2$	20.88	25.32	19.66	6.09	5.40	6.88
$\sigma_{AM}$	9.75	11.80	8.86	2.73	2.01	3.18
$\sigma_G^2$	12.93	15.62	12.08	6.25	5.40	7.09
$\sigma_E^2$	3.15	3.55	3.06	2.59	2.37	3.21
$\sigma_G^2$	7.70	8.56	7.40	4.19	4.02	4.58

<sup>a</sup>See table I;  $\sigma_G^2$ : total genetic variance; <sup>b</sup>No of families; <sup>c</sup>No of dams per family; <sup>d</sup>No of offspring (raised) per dam; <sup>e</sup>Family size.

appropriate for a growth trait in beef cattle, suggested that for a data set of size 8 000, the covariance should amount to at least 30% of the permanent environmental variance due to the dam before a likelihood ratio test would be expected to distinguish it from zero (at 5% error probability).

Results presented here reinforce earlier warnings about the inaccuracy of estimates of maternal effects and the pertaining variance components (Thompson, 1976; Foulley and Lefort, 1978). Clearly, use of an estimation procedure with "built-in" optimality characteristics like REML will not alleviate the need for large data sets supplying numerous types of covariances between relatives when attempting to estimate these components. Use of modern reproductive techniques such as embryo transfer may provide data where direct and maternal effects are less confounded. Most cases examined here considered data from 2 generations only, and includ-

**Table X.** Expected sampling correlations between estimates of (co)variance components<sup>a</sup> from data with a hierarchical full-sib design with (FS1ET) and without (FS1) embryo transfer.

f <sup>b</sup>	<i>Design FS1</i>			<i>Design FS1ET</i>		
	500	320	258	500	320	258
d <sup>c</sup>	5	4	10	5	4	10
d <sup>d</sup>	2	5	2	2	5	2
n <sup>e</sup>	16	25	31	16	25	31
<i>Model 3</i>						
$r_{A,M}$	-0.43	-0.30	-0.44	-0.12	-0.13	-0.10
$r_{A,E}$	-0.58	-0.80	-0.61	-0.72	-0.85	-0.78
$r_{M,E}$	-0.20	-0.06	-0.16	-0.31	-0.03	-0.27
<i>Model 4</i>						
$r_{A,M}$	0.28	0.36	0.30	-0.06	-0.11	-0.08
$r_{A,AM}$	-0.75	-0.77	-0.77	-0.21	-0.21	-0.09
$r_{A,E}$	-0.81	-0.90	-0.83	-0.65	-0.73	-0.72
$r_{M,AM}$	-0.70	-0.70	-0.70	-0.35	-0.18	-0.37
$r_{M,E}$	-0.61	-0.46	-0.60	-0.21	-0.04	-0.14
$r_{AM,E}$	0.76	0.75	0.78	-0.17	-0.28	-0.27
<i>Model 6</i>						
$r_{A,M}$	0.58	0.57	0.40	0.29	0.11	0.26
$r_{A,AM}$	-0.72	-0.71	-0.59	-0.38	-0.25	-0.30
$r_{A,C}$	-0.55	-0.54	-0.37	-0.34	-0.18	-0.30
$r_{A,E}$	-0.90	-0.96	-0.90	-0.32	-0.66	-0.32
$r_{M,AM}$	-0.97	-0.97	-0.96	-0.85	-0.68	-0.89
$r_{M,C}$	-0.99	-0.99	-0.99	-0.96	-0.95	-0.97
$r_{M,E}$	-0.53	-0.55	-0.37	0.56	-0.36	0.66
$r_{AM,C}$	0.95	0.95	0.94	0.84	0.68	0.88
$r_{AM,E}$	0.65	0.69	0.53	-0.63	-0.46	-0.73
$r_{C,E}$	0.47	0.52	0.30	-0.59	-0.34	-0.67

<sup>a</sup>See table VI; <sup>b</sup>No of families; <sup>c</sup>No of dams per family; <sup>d</sup>No of offspring (raised) per dam; <sup>e</sup>family size.

ing several generations would provide further contrast which might help to reduce the biologically induced high sampling correlations. Implications for the scope of fitting more detailed models, accounting, for instance, for dominance effects, recombination loss or variance due to new mutation, and of estimating the appropriate (co)variance components are somewhat discouraging.

## ACKNOWLEDGMENTS

Financial support for this study was provided under the MRC (Australia) grant UNE15 and by the Agricultural and Food Research Council (UK). I am grateful to WG Hill for comments on the manuscript.

## REFERENCES

- Baker RL (1980) The role of maternal effects on the efficiency of selection in beef cattle: a review. *Proc NZ Soc Anim Prod* 40, 285-303
- Bondari K, Willham RL, Freeman AE (1978) Estimates of direct and maternal genetic correlations for pupa weight and family size of *Tribolium*. *J Anim Sci* 47, 358-365
- Cantet RJC, Kress DD, Anderson DC, Doornbos DE, Burfening PJ, Blackwell RL (1988) Direct and maternal phenotypic effects on preweaning growth of beef cattle. *J Anim Sci* 66, 648-660
- Cox DR, Hinkley DV (1974) *Theoretical Statistics*. Chapman and Hall, London
- Edwards AWF (1966) *Likelihood*. Cambridge University Press
- Eisen EJ (1967) Mating designs for estimating direct and maternal genetic variances and direct-maternal covariances. *Can J Genet Cytol* 9, 13-22
- Foulley JL, Lefort G (1978) Méthodes d'estimation des effets directs et maternels en sélection animale. *Ann Génét Sél Anim* 10, 475-496
- Graser HU, Smith SP, Tier B (1987) A derivative-free approach for estimating variance components in animal models by Restricted Maximum Likelihood. *J Anim Sci* 64, 1362-1370
- Koch RM (1972) The role of maternal effects in animal breeding. VI. Maternal effects in beef cattle. *J Anim Sci* 35, 1316-1323
- Meyer K (1989) Restricted Maximum Likelihood to estimate variance components for animal models with several random effects using a derivative-free algorithm. *Genet Sel Evol* 21, 317-340
- Meyer K (1992) Variance components due to direct and maternal effects for growth traits of Australian beef cattle. *Livest Prod Sci* 31, 179-204
- Meyer K, Hill WG (1992) Approximation of sampling variances and confidence intervals for maximum likelihood estimates of variance components. *J Anim Breed Genet* 109, 264-280
- Nelder JA, Mead R (1965) A simplex method for function minimization. *Computer J* 7, 147-151
- Patterson HD, Thompson R (1971) Recovery of inter-block information when block sizes are unequal. *Biometrika* 58, 545-554
- Thompson R (1976) The estimation of maternal genetic variances. *Biometrics* 32, 903-917
- Thompson R (1977) Estimation of quantitative genetic parameters. In: *Proc Int Conf Quantitative Genetics*, (Pollak E, Kempthorne O, Bailey TB, eds) Iowa State Univ Press, Ames, 639-657
- Willham RL (1963) The covariance between relatives for characters composed of components contributed by related individuals. *Biometrics* 19, 18-27
- Willham RL (1972) The role of maternal effects in animal breeding: III. Biometrical aspects of maternal effects in animals. *J Anim Sci* 35, 1288-1293
- Willham RL (1980) Problems in estimating maternal effects. *Livest Prod Sci* 7, 405-418