

# Bayesian inference about dispersion parameters of univariate mixed models with maternal effects: theoretical considerations

RJC Cantet \*, RL Fernando, D Gianola \*\*

*University of Illinois, Department of Animal Sciences, Urbana, IL 61801, USA*

(Received 14 January 1991; accepted 5 January 1992)

**Summary** – Mixed linear models for maternal effects include fixed and random elements, and dispersion parameters (variances and covariances). In this paper a Bayesian model for inferences about such parameters is presented. The model includes a normal likelihood for the data, a “flat” prior for the fixed effects and a multivariate normal prior for the direct and maternal breeding values. The prior distribution for the genetic variance-covariance components is in the inverted Wishart form and the environmental components follow inverted  $\chi^2$  prior distributions. The kernel of the joint posterior density of the dispersion parameters is derived in closed form. Additional numerical and analytical methods of interest that are suggested to complete a Bayesian analysis include Monte-Carlo Integration, maximum entropy fit, asymptotic approximations, and the Tierney-Kadane approach to marginalization.

**maternal effect / Bayesian method / dispersion parameter**

**Résumé** – Inférence bayésienne des paramètres de dispersion de modèles mixtes univariates avec effets maternels : considérations théoriques. Les modèles linéaires mixtes avec effets maternels comprennent des éléments fixés et aléatoires, et des paramètres de dispersion (variances et covariances). Dans cet article est présenté un modèle bayésien pour l'estimation de ces paramètres. Le modèle comprend une vraisemblance normale pour les données, un a priori uniforme pour les effets fixés et un a priori multivariate normal pour les valeurs génétiques directes et maternelles. La distribution a priori des composantes de variance-covariance est une distribution de Wishart inverse et les composantes de milieu

---

\* Correspondence and reprints; present address: Facultad de Agronomía, Universidad de Buenos Aires, Departamento de Zootecnia, 1417 Buenos Aires, Argentina.

\*\* Present address: University of Wisconsin, Department of Meat and Animal Science, Madison, WI 53706, USA.

*suivent des distributions a priori de  $\chi^2$  inverse. Le noyau de la densité conjointe a posteriori des paramètres de dispersion est explicite. En outre, des méthodes numériques et analytiques sont proposées pour compléter l'analyse bayésienne: intégration par des méthodes de Monte-Carlo, ajustement par le maximum d'entropie, approximations asymptotiques et la méthode de marginalisation de Tierney-Kadane.*

**effet maternel / méthode bayésienne / paramètre de dispersion**

## INTRODUCTION

Mixed linear models for the study of quantitative traits include, in addition to fixed and random effects, the necessary dispersion parameters. Suppose one is interested in making inferences about variance and covariance components. Except in trivial cases, it is impossible to derive the exact sampling distribution of estimators of these parameters (Searle, 1979) so, at best, one has to resort to asymptotic results. Theory (Cramer, 1986) indicates that the joint distribution of maximum likelihood estimators of several parameters is asymptotically normal, and therefore so are their marginal distributions. However, this may not provide an adequate description of the distribution of estimators with finite sample sizes. On the other hand, the Bayesian approach is capable of producing exact joint and marginal posterior distributions for any sample size (Zellner, 1971; Box and Tiao, 1973), which give a full description of the state of uncertainty posterior to data.

In recent years, Bayesian methods have been developed for variance component estimation in animal breeding (Gianola and Fernando, 1986; Foulley *et al*, 1987; Macedo and Gianola, 1987; Carriquiry, 1989; Gianola *et al* 1990a, b). All these studies found analytically intractable joint posterior distributions of (co)variance components, as Broemeling (1985) has also observed. Further marginalization with respect to dispersion parameters seems difficult or impossible by analytical means. However, there are at least 3 other options for the study of marginal posterior distributions: 1), approximations; 2), integration by numerical means; and 3), numerical integration for computing moments followed by a fit of the density using these numerically obtained expectations. Recent advances in computing have encouraged the use of numerical methods in Bayesian inference. For example, after the pioneering work of Kloek and Van Dijk (1978), Monte Carlo integration (Hammersley and Handscomb, 1964; Rubinstein, 1981) has been employed in econometric models (Bauwens, 1984; Zellner *et al*, 1988), seemingly unrelated regressions (Richard and Steel, 1988) and binary responses (Zellner and Rossi, 1984).

Maternal effects are an important source of genetic and environmental variation in mammalian species (Falconer, 1981). Biometrical aspects of the associated theory were first developed by Dickerson (1947), and quantitative genetic models were proposed by Kempthorne (1955), Willham (1963, 1972) and Falconer (1965). Evolutionary biologists have also become interested in maternal effects (Cheverud, 1984; Riska *et al*, 1985; Kirkpatrick and Lande, 1989; Lande and Price, 1989). There is extensive animal breeding literature dealing with biological aspects and with estimation of maternal effects (*eg*, Foulley and Lefort, 1978; Willham, 1980;

Henderson, 1984, 1988). Although there are maternal sources of variation within and among breeds, we are concerned here only with the former sources.

The purpose of this expository paper is to present a Bayesian model for inference about variance and covariance components in a mixed linear model describing a trait affected by maternal effects. The formulation is general in the sense that it can be applied to the case where maternal effects are absent. The joint posterior distribution of the dispersion parameters is derived. Numerical methods for integration of dispersion parameters regarded as "nuisances" in specific settings are reviewed. Among these, Monte Carlo integration by "importance sampling" (Hammersley and Handscomb, 1964; Rubinstein, 1981) is discussed. Also, fitting a "maximum entropy" posterior distribution (Jaynes, 1957, 1979) using moments obtained by numerical means (Mead and Papanicolaou, 1984; Zellner and Highfield, 1988) is considered. Suggestions on some approximations to marginal posterior distributions of the (co)variance components are given. Asymptotic approximations using the Laplace method for integrals (Tierney and Kadane, 1986) are also described as a means for obtaining approximate posterior moments and marginal densities. Extension of the methods studied here to deal with multiple traits is possible but the algebra is more involved.

## THE BAYESIAN MODEL

### *Model and prior assumptions about location parameters*

The maternal animal model (Henderson, 1988) considered is:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}_o\mathbf{a}_o + \mathbf{Z}_m\mathbf{a}_m + \mathbf{E}_m\mathbf{e}_m + \mathbf{e}_o \quad [1]$$

where  $\mathbf{y}$  is an  $n \times 1$  vector of records and  $\mathbf{X}$ ,  $\mathbf{Z}_o$ ,  $\mathbf{Z}_m$  and  $\mathbf{E}_m$  are known, fixed,  $n \times p$ ,  $n \times a$ ,  $n \times a$  and  $n \times d$  matrices, respectively; without loss of generality, the matrix  $\mathbf{X}$  is assumed to have full-column rank. The vectors  $\boldsymbol{\beta}$ ,  $\mathbf{a}_o$ ,  $\mathbf{a}_m$  and  $\mathbf{e}_m$  are unknown fixed effects, additive direct breeding values, additive maternal breeding values and maternal environmental deviations, respectively. The  $n \times 1$  vector  $\mathbf{e}_o$  contains environmental deviations as well as any discrepancy between the "structure" of the model ( $\mathbf{X}\boldsymbol{\beta} + \mathbf{Z}_o\mathbf{a}_o + \mathbf{Z}_m\mathbf{a}_m + \mathbf{E}_m\mathbf{e}_m$ ) and the data  $\mathbf{y}$ . As in Gianola *et al.* (1990b), the vectors  $\boldsymbol{\beta}$ ,  $\mathbf{a}_o$ ,  $\mathbf{a}_m$  and  $\mathbf{e}_m$  are formally viewed as location parameters of the conditional distribution  $\mathbf{y}|\boldsymbol{\beta}, \mathbf{a}_o, \mathbf{a}_m, \mathbf{e}_m$ , but a distinction is made between  $\boldsymbol{\beta}$  and the other 3 vectors depending on the state of uncertainty prior to observing data. It is assumed *a priori* that  $\boldsymbol{\beta}$  follows a uniform distribution, so as to reflect vague prior knowledge on this vector. Polygenic inheritance is often assumed for  $\mathbf{a} = [\mathbf{a}'_o, \mathbf{a}'_m]'$  (Falconer, 1981; Bulmer, 1985) so it is reasonable to postulate *a priori* that  $\mathbf{a}$  follows the multivariate normal distribution:

$$\mathbf{a}|\mathbf{G}, \mathbf{A} \sim N_a(\mathbf{0}, \mathbf{G} \otimes \mathbf{A}) \quad [2]$$

where  $\mathbf{G}$  is a  $2 \times 2$  matrix with diagonal elements  $\sigma_{A_o}^2$  and  $\sigma_{A_m}^2$ , the variance components for additive direct and maternal genetic effects, respectively, and off-diagonal elements  $\sigma_{A_o A_m}$ , the covariance between additive direct and maternal

effects. The  $a \times a$  positive-definite matrix  $\mathbf{A}$  has elements equal to Wright's coefficients of additive relationship or twice Melecot's coefficients of co-ancestry (Willham, 1963). Maternal environmental deviations, presumably caused by the joint action of many factors having relatively small effects are also assumed to be normally, independently distributed (Quaas and Pollak, 1980; Henderson, 1988) as:

$$\mathbf{e}_m \sim N_d(\mathbf{0}, \mathbf{I}_d \sigma_{E_m}^2) \quad [3]$$

where  $\sigma_{E_m}^2$  is the maternal environmental variance. It is assumed that *a priori*  $\boldsymbol{\beta}$ ,  $\mathbf{a}$  and  $\mathbf{e}_m$  are mutually independent. For the vector  $\mathbf{y}$ , it will be assumed that:

$$\mathbf{y} | \boldsymbol{\beta}, \mathbf{a}_o, \mathbf{a}_m, \mathbf{e}_m, \sigma_{E_o}^2 \sim N_n(\mathbf{X}\boldsymbol{\beta} + \mathbf{Z}_o\mathbf{a}_o + \mathbf{Z}_m\mathbf{a}_m + \mathbf{E}_m\mathbf{e}_m, \mathbf{I}_n\sigma_{E_o}^2) \quad [4]$$

where  $\sigma_{E_o}^2$  is the variance of the direct environmental effects. It should be noted that [1–4] complete the specification of the classical mixed linear model (Henderson, 1984), but in the latter, distributions [2] and [3] have a frequentist interpretation. A simplifying assumption made in this model, for analytical reasons, is that the direct and maternal environmental effects are uncorrelated.

#### **Prior assumptions about variance parameters**

Variance and covariance components, the main focus of this study, appear in the distributions of  $\mathbf{a}$ ,  $\mathbf{e}_m$  and  $\mathbf{e}_o$ . Often these components are unknown. In the Bayesian approach, a joint prior distribution must be specified for these, so as to reflect uncertainty prior to observing  $\mathbf{y}$ . "Flat" prior distributions, although leading to inferences that are equivalent to those obtained from likelihood in certain settings (Harville, 1974, 1977) can cause problems in others (Lindley and Smith, 1972; Thompson, 1980; Gianola *et al*, 1990b). In this study, informative priors of the type of proper conjugate distributions (Raiffa and Schlaiffer, 1961) are used. A prior distribution is said to be conjugate if the posterior distribution is also in the same family. For example, a normal prior combined with a normal likelihood produces a normal posterior (Zellner, 1971; Box and Tiao, 1973). However, as shown later for the variance-covariance structure under consideration, the posterior distribution of the dispersion parameters is not of the same type as their joint prior distribution. This was also found by Macedo and Gianola (1987) and by Gianola *et al* (1990b) who studied a mixed linear model with several variance components employing normal-gamma conjugate prior distributions.

An inverted-Wishart distribution (Zellner, 1971; Anderson, 1984; Foulley *et al*, 1987) will be used for  $\mathbf{G}$ , with density:

$$p(\mathbf{G} | n_g, \mathbf{G}_h) \propto |\mathbf{G}|^{-\frac{1}{2}(n_g+3)} \exp \left[ -\frac{1}{2} \text{tr}(\mathbf{G}^{-1}\mathbf{G}^*) \right] \quad [5]$$

where  $\mathbf{G}^* = n_g\mathbf{G}_h$ . The  $2 \times 2$  matrix  $\mathbf{G}_h$  of "hyperparameters", interpretable as prior values of the dispersion parameters, has diagonal elements  $s_{A_o}^2$  and  $s_{A_m}^2$ , and off-diagonal elements  $S_{A_oA_m}$ . The integer  $n_g$  is analogous to degrees of freedom and reflects the "degree of belief" on  $\mathbf{G}_h$  (Chen, 1979). Choosing hyperparameter

values may be difficult in many applications. Gianola *et al* (1990b) suggested fitting the distribution to past estimates of the (co)variance components by *eg* a method of moments fit. For traits such as birth and weaning weight in cattle there is a considerable number of estimates of the necessary (co)variance components in the literature (Cantet *et al*, 1988). Clearly, the value of  $\mathbf{G}_h$  influences posterior inferences unless the prior distribution is overwhelmed by the likelihood function (Box and Tiao, 1973).

Similarly, as in Hoeschele *et al* (1987) the inverted  $\chi^2$  distribution (a particular case of the inverted Wishart distribution) is suggested for the environmental variance components, and the densities are:

$$p(\sigma_{Em}^2 | n_m, s_{Em}^2) \propto (\sigma_{Em}^2)^{-\frac{1}{2}(n_m+2)} \exp \left[ -\frac{n_m s_{Em}^2}{2\sigma_{Em}^2} \right] \quad [6]$$

and

$$p(\sigma_{Eo}^2 | n_o, s_{Eo}^2) \propto (\sigma_{Eo}^2)^{-\frac{1}{2}(n_o+2)} \exp \left[ -\frac{n_o s_{Eo}^2}{2\sigma_{Eo}^2} \right] \quad [7]$$

The prior variances  $s_{Em}^2$  and  $s_{Eo}^2$  are the scalar counterparts of  $\mathbf{G}_n$ , and  $n_o$  and  $n_m$  are the corresponding degrees of belief. The marginal distribution of any diagonal element of a Wishart random matrix is  $\chi^2$  (Anderson, 1984). Likewise, the marginal distribution of the diagonal of an inverted-Wishart random matrix is inverted  $\chi^2$  (Zellner, 1971). Note that the 2 variances in [6] and [7] cannot be arranged in matrix form similar to the additive (co)variance components in  $\mathbf{G}$  to obtain an inverted Wishart density, unless  $n_o = n_m$ . Setting  $n_g, n_o$  and  $n_m$  to zero makes the prior distributions for all (co)variance components "uninformative", in the sense of Zellner (1971).

## POSTERIOR DENSITIES

### Joint posterior density of all parameters

The posterior density of all parameters (Zellner, 1971; Box and Tiao, 1973) is proportional to the product of the densities corresponding to the distributions in [2], [3] and [4] times [5], [6] and [7]. One obtains:

$$\begin{aligned} & p(\boldsymbol{\beta}, \mathbf{a}_o, \mathbf{a}_m, \mathbf{e}_m, \mathbf{G}, \sigma_{Em}^2, \sigma_{Eo}^2 | \mathbf{y}, \mathbf{G}_h, s_{Eo}^2, s_{Em}^2, n_g, n_o, n_m) \propto \\ & \exp \left[ -\frac{1}{2\sigma_{Eo}^2} (\mathbf{y} - \mathbf{X}\boldsymbol{\beta} - \mathbf{Z}_o\mathbf{a}_o - \mathbf{Z}_m\mathbf{a}_m - \mathbf{E}_m\mathbf{e}_m)' (\mathbf{y} - \mathbf{X}\boldsymbol{\beta} - \mathbf{Z}_o\mathbf{a}_o - \mathbf{Z}_m\mathbf{a}_m - \mathbf{E}_m\mathbf{e}_m) \right] \times \\ & \exp \left[ -\frac{1}{2} \mathbf{a}' (\mathbf{G}^{-1} \otimes \mathbf{A}^{-1}) \mathbf{a} \right] |\mathbf{G}|^{-\frac{1}{2}(n_g+\mathbf{a}+3)} \exp \left[ -\frac{1}{2} \text{tr}(\mathbf{G}^{-1}\mathbf{G}^*) \right] \times \\ & (\sigma_{Eo}^2)^{-\frac{1}{2}(n_o+n+2)} \exp \left[ -\frac{n_o s_{Eo}^2}{2\sigma_{Eo}^2} \right] (\sigma_{Em}^2)^{-\frac{1}{2}(n_m+d+2)} \exp \left[ -\frac{n_m s_{Em}^2 + \mathbf{e}'_m \mathbf{e}_m}{2\sigma_{Em}^2} \right] \quad [8] \end{aligned}$$

with:

$$-\infty < \beta_i < \infty, i = 1, \dots, p; \quad -\infty < a_j < \infty, j = 1, \dots, 2a; \quad -\infty < e_{mk} < \infty, k = 1, \dots, d;$$

$$\mathbf{G} \text{ positive definite}; \quad \sigma_{E_m}^2 > 0; \quad \sigma_{E_o}^2 > 0$$

To facilitate marginalization of [8], and as in Gianola *et al* (1990a), let  $\mathbf{W} = [\mathbf{X}|\mathbf{Z}_o|\mathbf{Z}_m|\mathbf{E}_m]$ ,  $\boldsymbol{\theta}' = [\boldsymbol{\beta}'|\mathbf{a}'|\mathbf{e}'_m]$  and define  $\hat{\boldsymbol{\theta}}$  such that

$$(\mathbf{W}'\mathbf{W} + \boldsymbol{\Sigma})\hat{\boldsymbol{\theta}} = \mathbf{W}'\mathbf{y}$$

where the  $p + 2a + d$  square matrix  $\boldsymbol{\Sigma}$  is given by:

$$\boldsymbol{\Sigma} = \begin{bmatrix} \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{G}^{-1} \otimes \mathbf{A}^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{I}_d(\sigma_{E_m}^2)^{-1} \end{bmatrix} \sigma_{E_o}^2$$

Using this, one can write:

$$\begin{aligned} & (\mathbf{y} - \mathbf{X}\boldsymbol{\beta} - \mathbf{Z}_o\mathbf{a}_o - \mathbf{Z}_m\mathbf{a}_m - \mathbf{E}_m\mathbf{e}_m)'(\mathbf{y} - \mathbf{X}\boldsymbol{\beta} - \mathbf{Z}_o\mathbf{a}_o - \mathbf{Z}_m\mathbf{a}_m - \mathbf{E}_m\mathbf{e}_m) + \left[ \mathbf{a}'(\mathbf{G}^{-1} \otimes \mathbf{A}^{-1})\mathbf{a} + \frac{\mathbf{e}'_m\mathbf{e}_m}{\sigma_{E_m}^2} \right] \sigma_{E_o}^2 \\ &= (\mathbf{y} - \mathbf{W}\boldsymbol{\theta})'(\mathbf{y} - \mathbf{W}\boldsymbol{\theta}) + \boldsymbol{\theta}'\boldsymbol{\Sigma}\boldsymbol{\theta} \\ &= \mathbf{y}'\mathbf{y} - 2\boldsymbol{\theta}'\mathbf{W}'\mathbf{y} + \boldsymbol{\theta}'(\mathbf{W}'\mathbf{W} + \boldsymbol{\Sigma})\boldsymbol{\theta} \\ &= \mathbf{y}'\mathbf{y} - 2\boldsymbol{\theta}'(\mathbf{W}'\mathbf{W} + \boldsymbol{\Sigma})\hat{\boldsymbol{\theta}} + \boldsymbol{\theta}'(\mathbf{W}'\mathbf{W} + \boldsymbol{\Sigma})\boldsymbol{\theta} \\ &= \mathbf{y}'\mathbf{y} - \hat{\boldsymbol{\theta}}'(\mathbf{W}'\mathbf{W} + \boldsymbol{\Sigma})\hat{\boldsymbol{\theta}} + (\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})'(\mathbf{W}'\mathbf{W} + \boldsymbol{\Sigma})(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}) \end{aligned} \tag{9}$$

Gianola *et al* (1990a) noted that

$$\mathbf{y}'\mathbf{y} - \hat{\boldsymbol{\theta}}'(\mathbf{W}'\mathbf{W} + \boldsymbol{\Sigma})\hat{\boldsymbol{\theta}} = \mathbf{y}'\mathbf{y} - \hat{\boldsymbol{\theta}}'\mathbf{W}'\mathbf{y}$$

in [9] can be interpreted as a “mixed model residual sum of squares”. Using [9] in [8] the joint posterior density becomes:

$$\begin{aligned} & p(\boldsymbol{\beta}, \mathbf{a}_o, \mathbf{a}_m, \mathbf{e}_m, \mathbf{G}, \sigma_{E_m}^2, \sigma_{E_o}^2 | \mathbf{y}, \mathbf{G}_h, s_{E_o}^2, s_{E_m}^2, n_g, n_o, n_m) \propto \\ & \exp \left[ -\frac{1}{2\sigma_{E_o}^2}(\mathbf{y}'\mathbf{y} - \hat{\boldsymbol{\theta}}'\mathbf{W}'\mathbf{y}) \right] \exp \left[ -\frac{1}{2\sigma_{E_o}^2}(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})'(\mathbf{W}'\mathbf{W} + \boldsymbol{\Sigma})(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}) \right] \times \\ & \quad |\mathbf{G}|^{-\frac{1}{2}(n_g+a+3)} \exp \left[ -\frac{1}{2} \text{tr}(\mathbf{G}^{-1}\mathbf{G}^*) \right] \times \\ & \quad (\sigma_{E_o}^2)^{-\frac{1}{2}(n_o+n+2)} \exp \left[ -\frac{n_o s_{E_o}^2}{2\sigma_{E_o}^2} \right] (\sigma_{E_m}^2)^{-\frac{1}{2}(n_m+d+2)} \exp \left[ -\frac{n_m s_{E_m}^2}{2\sigma_{E_m}^2} \right] \end{aligned} \tag{10}$$

**Posterior density of the (co)variance components**

To obtain the marginal posterior distribution of  $\mathbf{G}, \sigma_{Em}^2$  and  $\sigma_{Eo}^2, \boldsymbol{\theta}$  must be integrated out of [10]. This can be accomplished noting that the second exponential term in [10] is the kernel of the  $(p + 2a + d)$ -variate normal distribution

$$\boldsymbol{\theta} \sim N_{p+2a+d}[\widehat{\boldsymbol{\theta}}, (\mathbf{W}'\mathbf{W} + \boldsymbol{\Sigma})^{-1}\sigma_{Eo}^2]$$

and the variance-covariance matrix is non-singular because  $\mathbf{X}$  has full-column rank. The remaining terms in [10] do not depend on  $\boldsymbol{\theta}$ . Therefore, with  $R_{\boldsymbol{\theta}}$  being the range of  $\boldsymbol{\theta}$ , using properties of the normal distribution we have:

$$\int_{R_{\boldsymbol{\theta}}} \exp \left[ -\frac{1}{2\sigma_{Eo}^2}(\boldsymbol{\theta} - \widehat{\boldsymbol{\theta}})'(\mathbf{W}'\mathbf{W} + \boldsymbol{\Sigma})(\boldsymbol{\theta} - \widehat{\boldsymbol{\theta}}) \right] d\boldsymbol{\theta} = (2\pi)^{\frac{p+2a+d}{2}} |\mathbf{W}'\mathbf{W} + \boldsymbol{\Sigma}|^{-\frac{1}{2}} (\sigma_{Eo}^2)^{\frac{1}{2}(p+2a+d)}$$

The marginal posterior distribution of all (co)variance components then is:

$$\begin{aligned} p(\mathbf{G}, \sigma_{Em}^2, \sigma_{Eo}^2 | \mathbf{y}, \mathbf{G}_h, s_{Eo}^2, s_{Em}^2, n_g, n_o, n_m) &\propto \\ \exp \left[ -\frac{1}{2\sigma_{Eo}^2} (\mathbf{y}'\mathbf{y} - \widehat{\boldsymbol{\theta}}'\mathbf{W}'\mathbf{y} + n_o s_{Eo}^2) \right] &|\mathbf{W}'\mathbf{W} + \boldsymbol{\Sigma}|^{-\frac{1}{2}} \times \\ |\mathbf{G}|^{-\frac{1}{2}(n_g+a+3)} \exp \left[ -\frac{1}{2} tr(\mathbf{G}^{-1}\mathbf{G}^*) \right] \exp \left[ -\frac{n_m s_{Em}^2}{2\sigma_{Em}^2} \right] &\times \\ (\sigma_{Eo}^2)^{-\frac{1}{2}(n_o+n-p-2a-d+2)} (\sigma_{Em}^2)^{-\frac{1}{2}(n_m+d+2)} & \end{aligned} \tag{11}$$

The structure of [11] makes it difficult or impossible to obtain by analytical means the marginal posterior distribution of  $\mathbf{G}, \sigma_{Eo}^2$  or  $\sigma_{Em}^2$ . Therefore, in order to make marginal posterior inferences about the elements of  $\mathbf{G}$  or the environmental variances, approximations or numerical integration must be used. The latter may give accurate estimates of posterior moments, but in multiparameter situations computations can be prohibitive.

There are 2 basic approaches to numerical integration in Bayesian analysis. The first one is based on classical methods such as quadrature (Naylor and Smith, 1982, 1988; Wright, 1986). Increased power of computers has made Monte Carlo numerical integration (MCI), the second approach, feasible in posterior inferences in econometric models (Kloek and Van Dijk, 1978; Bauwens, 1984; Bauwens and Richard, 1985; Zellner *et al*, 1988) and in other models (Zellner and Rossi, 1984; Geweke, 1988; Richard and Steel, 1988). In MCI the error is inversely proportional to  $N^{1/2}$ , where  $N$  is the number of points where the integrand is evaluated (Hammersley and Handscomb, 1964; Rubinstein, 1981). Even though this "convergence" of the error to zero is not rapid, neither the dimensionality of the integration region nor the degree of smoothness of the function evaluated enter into the determination of the error (Haber, 1970). This suggests that as the number of dimensions of integration increases the advantage of MCI over classical methods should also increase. A brief description of MCI in the context of maternal effects models is discussed next.

## POSTERIOR MOMENTS VIA MONTE CARLO INTEGRATION

Consider finding moments of parameters having the joint posterior distribution with density [11]. Let  $\Gamma' = [\sigma_{A_0}^2, \sigma_{A_m}^2, \sigma_{A_0 A_m}, \sigma_{E_m}^2, \sigma_{E_0}^2]$ , and let  $g(\Gamma)$  be either a scalar, vector or matrix function of  $\Gamma$  of which we would like to compute its posterior expectation. Also, let [11] be represented as  $p(\Gamma | \mathbf{y}, \mathbf{H})$ , where  $\mathbf{H}$  stands for hyperparameters. Then:

$$\mathbf{E}[g(\Gamma) | \mathbf{y}, \mathbf{H}] = \int g(\Gamma) p(\Gamma | \mathbf{y}, \mathbf{H}) d\Gamma \quad [12]$$

assuming the integrals in [12] exist.

Different techniques can be used with MCI to achieve reasonable accuracy. An appealing one for computing posterior moments (Kloek and Van Dijk, 1978; Bauwens, 1984; Zellner and Rossi, 1984; Richard and Steel, 1988) is called "importance sampling" (Hammersley and Handscomb, 1964; Rubinstein, 1981). Let  $I(\Gamma)$  be a known probability density function defined on the space of  $\Gamma$ ;  $I(\Gamma)$  is called the importance sampling function. Following Kloek and Van Dijk (1978) let  $M(\Gamma)$  be:

$$M(\Gamma) = \frac{g(\Gamma) p(\Gamma | \mathbf{y}, \mathbf{H})}{I(\Gamma)} \quad [13]$$

with [13] defined in the region where  $I(\Gamma) > 0$ . Then [12] is expressible as:

$$\mathbf{E}[g(\Gamma) | \mathbf{y}, \mathbf{H}] = \int M(\Gamma) I(\Gamma) d\Gamma = E[M(\Gamma)] \quad [14]$$

where the expectation is taken with respect to the importance density  $I(\Gamma)$ .

Using a standard Monte Carlo procedure (Hammersley and Handscomb, 1964; Rubinstein, 1981), values of  $\Gamma$  are drawn at random from the distribution with density  $I(\Gamma)$ . Then the function  $M(\Gamma)$  is evaluated for each drawn value of  $\Gamma$ ,  $\Gamma_j$  ( $j = 1, \dots, m$ ) say. For sufficiently large  $m$ :

$$\mathbf{E}[M(\Gamma)] \approx \frac{\sum_{j=1}^m M(\Gamma_j)}{m} \quad [15]$$

The critical point is the choice of the density function  $I(\Gamma)$ . The closer  $I(\Gamma)$  is to  $p(\Gamma | \mathbf{y}, \mathbf{H})$ , the smaller is the variance of  $M(\Gamma)$ , and the number of drawings needed to obtain a certain accuracy (Hammerley and Handscomb, 1964; Rubinstein, 1981).

Another important requirement is that random drawings of  $\Gamma$  should be relatively simple to obtain from  $I(\Gamma)$  (Kloek and Van Dijk, 1978; Bauwens, 1984). For location parameters, the multivariate normal, multivariate and matrix-variate  $t$  and poly- $t$  distributions have been used as importance functions (Kloek and Van Dijk, 1978; Bauwens, 1984; Bauwens and Richard, 1985; Richard and Steel, 1988; Zellner *et al*, 1988). Bauwens (1984) developed an algorithm for obtaining random samples



from the inverted Wishart distribution. There are several problems yet to be solved and the procedure is still experimental (Richard and Steel, 1988). However, results obtained so far make MCI by importance sampling promising (Bauwens, 1984; Zellner and Rossi, 1984; Richard and Steel, 1988; Zellner *et al*, 1988).

Consider calculating the mean of  $\mathbf{G}$ ,  $\sigma_{Eo}^2$  and  $\sigma_{Em}^2$  with joint posterior density as given in [11]. From [13] and [14]:

$$E[\mathbf{\Gamma} | \mathbf{y}, \mathbf{H}] = \int \left[ \frac{\mathbf{\Gamma} p(\mathbf{\Gamma} | \mathbf{y}, \mathbf{H})}{I(\mathbf{\Gamma})} \right] I(\mathbf{\Gamma}) d\mathbf{\Gamma} \tag{16}$$

Let now:

$$I(\mathbf{\Gamma}) = I_1(\mathbf{\Gamma}) I_2(\mathbf{\Gamma}) I_3(\mathbf{\Gamma}) \tag{17}$$

where:

- $I_1(\mathbf{\Gamma})$  = prior density of  $\mathbf{G}$  ([5] times  $k_1$ , the integration constant),
- $I_2(\mathbf{\Gamma})$  = prior density of  $\sigma_{Em}^2$  ([6] times  $k_2$ , the integration constant),
- $I_3(\mathbf{\Gamma})$  = prior density of  $\sigma_{Eo}^2$  [7] times  $k_3$ , the integration constant).

Then:

$$M(\mathbf{\Gamma}) = \left( \frac{k_o}{k_1 k_2 k_3} \right) \mathbf{\Gamma} (\sigma_{Eo}^2)^{-\frac{1}{2}(n-p-2a-d)} (\sigma_{Em}^2)^{-\frac{d}{2}} \times \exp \left[ -\frac{1}{2\sigma_{Eo}^2} (\mathbf{y}'\mathbf{y} - \hat{\boldsymbol{\theta}}' \mathbf{W}'\mathbf{y}) \right] |\mathbf{W}'\mathbf{W} + \boldsymbol{\Sigma}|^{-\frac{1}{2}} |\mathbf{G}|^{-\frac{n}{2}} \tag{18}$$

where  $k_o$  is the constant of integration of [11]. Evaluating  $E(\mathbf{\Gamma} | \mathbf{y}, \mathbf{H})$  then entails the following steps:

a) draw at random the elements of  $\mathbf{\Gamma}$  from distributions with densities  $I_1(\mathbf{\Gamma})$  (inverted Wishart),  $I_2(\mathbf{\Gamma})$  (inverted  $\chi^2$ ) and  $I_3(\mathbf{\Gamma})$  (inverted  $\chi^2$ ). This can be done using, for example, the algorithm of Bauwens (1984).

b) Evaluate  $k_o = \left[ \int [11] d\mathbf{\Gamma} \right]^{-1}$ . Now,

$$\int [11] d\mathbf{\Gamma} = \int \frac{[11]}{I(\mathbf{\Gamma})} I(\mathbf{\Gamma}) d\mathbf{\Gamma} = E(M_o) = k_o^{-1}$$

Note that  $M_o$  is [18] without  $\mathbf{\Gamma}$ . Then  $k_o$  can be evaluated by MCI by computing the average of  $M_o$ , and taking its reciprocal.

c) Once  $M_o$  is evaluated, then compute  $M(\mathbf{\Gamma}) = \mathbf{\Gamma} M_o$ . In order to perform steps (b) and (c), the mixed model equations and the determinant of  $\mathbf{W}'\mathbf{W} + \boldsymbol{\Sigma}$  need to be solved and evaluated, repeatedly, for each drawing. The mixed model equations can be solved iteratively and diagonalization or sparse matrix factorization (Misztal, 1990) can be employed to advantage in the calculation of the determinant.

This procedure can be used to calculate any function of  $\mathbf{\Gamma}$ . For example, the posterior variance-covariance matrix is:

$$\text{var} [\mathbf{\Gamma} | \mathbf{y}, \mathbf{H}] = E[\mathbf{\Gamma}\mathbf{\Gamma}' | \mathbf{y}, \mathbf{H}] - E[\mathbf{\Gamma} | \mathbf{y}, \mathbf{H}] E[\mathbf{\Gamma}' | \mathbf{y}, \mathbf{H}] \tag{19}$$

so the additional calculation required would be evaluating  $M'(\mathbf{\Gamma}) = \mathbf{\Gamma}\mathbf{\Gamma}' M_o$ .

## MAXIMUM ENTROPY FIT OF MARGINAL POSTERIOR DENSITIES

A full Bayesian analysis requires finding the marginal posterior distribution of each of the (co)variance components. Probability statements and highest posterior density intervals are obtained from these distributions (Zellner, 1971; Box and Tiao, 1973). Marginal posterior densities can be obtained using the Monte Carlo method (Kloek and Van Dijk, 1978) but it is computationally expensive. An alternative is to compute by MCI some moments (for instance, the first 4) of each parameter, and then fit a function that approximates the necessary marginal distribution. A method that gives a reasonable fit, "Maximum entropy" (ME), has been used by Mead and Papanicolaou (1984) and Zellner and Highfield (1988). Choosing the ME distribution means assuming the "least" possible (Jaynes, 1979), *ie*, using information one has but not using what one does not have. An ME fit based on the first 4 moments implies constructing a distribution that does not use information beyond that conveyed by these moments. Jaynes (1957) set the basis for what is known as the "ME formalism" and found a role for this to play in Bayesian statistics.

The entropy ( $W$ ) of a continuous distribution with density  $p(x)$  is defined (Shannon, 1948; Jaynes, 1957, 1979) to be:

$$W = - \int \log [p(x)]p(x)dx \quad [20]$$

The ME distribution is obtained from the density that maximizes [20] subject to the conditions:

$$\int x^i p(x)dx = \mu_i \quad i = 0, 1, \dots, 4 \quad [21]$$

where  $\mu_0 = 1$  (by definition of a proper density function) and  $\mu_i (i = 1, \dots, 4)$  are the first 4 moments of the distribution of  $x$ . Zellner and Highfield (1988) expressed the function to be maximized as the Lagrangian:

$$L[x, p(x), l] = \int p(x) \log [p(x)]dx + \sum_{i=0}^4 l_i \left( \int x^i p(x)dx - \mu_i \right) \quad [22]$$

where the  $l_i (i = 0, \dots, 4)$  are Lagrange multipliers and  $l = [l_0, l_1, l_2, l_3, l_4]'$ . Note that [22] involves integrals whose integrands depend on the unknown function  $p(x)$ , and on functions of it ( $\log p(x)$ ). Rewrite [22] as:

$$L[x, p(x), l] = \int [p(x) \log [p(x)] + \sum_{i=0}^4 l_i x^i p(x)]dx - [l_0 + l_1 \mu_1 + l_2 \mu_2 + l_3 \mu_3 + l_4 \mu_4] \quad [23]$$

On defining  $F[x, p(x)] = p(x) \log p(x)$ ,  $S_i[x, p(x)] = x^i p(x)$ , and

$$H[x, p(x)] = F[x, p(x)] + \sum_{i=0}^4 l_i S_i[x, p(x)]$$

formula [23] is expressible as:

$$\begin{aligned}
 L[x, p(x), l] &= \int \{F[x, p(x)] + \sum_{i=0}^4 l_i S_i[x, p(x)]\} dx - constant \\
 &= \int H[x, p(x)] dx - constant \tag{24}
 \end{aligned}$$

Using Euler's equation (Hildebrand, 1972) the condition for a stationary point is:

$$\frac{d}{dx} \left[ \frac{\partial H}{\partial p'(x)} \right] - \frac{\partial H}{\partial p(x)} = 0 \tag{25}$$

Because  $H$  does not depend on  $p'(x)$ , [25] holds only if  $\partial H/\partial p(x) = 0$ , ie, if:

$$\begin{aligned}
 \frac{\partial H}{\partial p(x)} &= \frac{\partial F}{\partial p(x)} + \sum_{i=0}^4 l_i \left[ \frac{\partial S_i}{\partial p(x)} \right] \\
 &= \frac{\partial [p(x) \log [p(x)]]}{\partial p(x)} + \sum_{i=0}^4 l_i \left[ \frac{\partial [x^i p(x)]}{\partial p(x)} \right] = 0
 \end{aligned}$$

Hence, the condition for a stationary point is:

$$\log [p(x)] + 1 + l_0 + l_1 x + l_2 x^2 + l_3 x^3 + l_4 x^4 = 0 \tag{26}$$

plus the 5 constraints given in [21]. From [26], the density of the ME distribution of  $x$  has the form:

$$p(x | l) = \exp[-(1 + l_0 + l_1 x + l_2 x^2 + l_3 x^3 + l_4 x^4)] \tag{27}$$

To specify the ME distribution completely  $l$  must be found. Zellner and Highfield (1988) suggested a numerical solution based on Newton's method. Using [27] the side conditions [21] can be written as:

$$G_i(l) = \mu_i \quad i = 0, \dots, 4$$

where:

$$G_i(l) = \int x^i - \exp[-(1 + l_0 + l_1 x + l_2 x^2 + l_3 x^3 + l_4 x^4)] dx \tag{28}$$

Expanding  $G_i(l)$  with a Taylor series about  $l_0$ , a trial value for  $l$ , and retaining the linear terms leads to:

$$G_i(l) \sim G_i(l_0) + \sum_{j=0}^4 \left[ \frac{\partial G_i(l)}{\partial l_j} \right]_{l=l_0} (l_j - l_j^0) \tag{29}$$

Now

$$\begin{aligned} \frac{\partial G_i(\mathbf{l})}{\partial l_j} &= \int x^i \frac{\partial \exp [-(1 + l_0 + l_1 x + l_2 x^2 + l_3 x^3 + l_4 x^4)]}{\partial l_j} dx \\ &= - \int x^{i+j} \exp [-(1 + l_0 + l_1 x + l_2 x^2 + l_3 x^3 + l_4 x^4)] dx \quad i, j = 0, 1, \dots, 4 \end{aligned}$$

These derivatives are simply moments (with negative sign) of the maximum entropy distribution.

Putting

$$\left[ \frac{\partial G_i(\mathbf{l})}{\partial l_j} \right] = -G_{i+j}(\mathbf{l})$$

in [29] and setting this equal to [28] one obtains the linear system in  $l$ :

$$\sum_{j=0}^4 [G_{i+j}(\mathbf{l})]_{l_i=l_0} (l_j - l_j^0) = G_i(\mathbf{l}_0) - \mu_i \quad i = 0, 1, \dots, 4$$

This system can be solved for  $l_j (j = 0, 1, \dots, 4)$  to obtain a new set of trial values and, thus, an iteration is established. Defining

$$\delta_j^{[t]} = l_j^{[t]} - l_j^{[t-1]}$$

and observing that  $0 \leq i + j \leq 8$ , the above system can be written in matrix notation as:

$$\begin{bmatrix} G_0 & G_1 & G_2 & G_3 & G_4 \\ G_1 & G_2 & G_3 & G_4 & G_5 \\ G_2 & G_3 & G_4 & G_5 & G_6 \\ G_3 & G_4 & G_5 & G_6 & G_7 \\ G_4 & G_5 & G_6 & G_7 & G_8 \end{bmatrix}_{l=l^{[t-1]}} \begin{bmatrix} \delta_0 \\ \delta_1 \\ \delta_2 \\ \delta_3 \\ \delta_4 \end{bmatrix}^{[t]} = \begin{bmatrix} G_0 \\ G_1 \\ G_2 \\ G_3 \\ G_4 \end{bmatrix}_{l=l^{[t-1]}} - \begin{bmatrix} \mu_0 \\ \mu_1 \\ \mu_2 \\ \mu_3 \\ \mu_4 \end{bmatrix} \quad [30]$$

This system is solved for  $\delta^{[t]}$  to obtain  $l^{[t]} = l^{[t-1]} + \delta^{[t]}$ , the vector of new trial values. Iteration continues until  $\delta$  becomes appropriately small. Zellner and Highfield (1988) showed that coefficient matrix in [30] is positive definite, so solutions are unique. In summary, the method includes 3 types of computations. First, the moments  $\mu_1 - \mu_4$  must be computed by some method such as MCI; this is done only once. Second, the  $G_i$  values ( $i = 0, 1, \dots, 8$ ) are computed at every round of iteration carrying out unidimensional integrations, as indicated in [28]. Third, the  $5 \times 5$  system [30] is solved. At convergence, the ME density [27] is employed to approximate marginal inferences about the appropriate element of  $\Gamma$ .

### SOME ANALYTICAL APPROXIMATIONS TO MARGINAL POSTERIOR DENSITIES

Because numerical integration can be computationally expensive and the accuracy of MCI in this type of problem is still unknown, we consider several approximations to marginal posterior distributions.

The mode of the posterior density [11] can be found by maximizing this jointly with respect to  $\mathbf{G}$ ,  $\sigma_{E_m}^2$  and  $\sigma_{E_o}^2$ . Foulley *et al* (1987), Gianola *et al* (1990b) and Macedo and Gianola (1987) showed how this could be done with a simple algorithm based on first derivatives. Additional algorithms can be constructed using second derivatives, and the necessary expression are given in the *Appendix*. The solutions can be viewed as weighted averages of REML “estimators” of dispersion parameters and of the hyperparameters  $\mathbf{G}_h$ ,  $s_{E_o}^2$  and  $s_{E_m}^2$ . Let the modal values so obtained be  $\hat{\mathbf{G}}$ ,  $\hat{\sigma}_{E_o}^2$  and  $\hat{\sigma}_{E_m}^2$ , or  $\hat{\Gamma}$ , in compact.

Consider approximations to the marginal density of  $\mathbf{G}$  because this matrix contains the parameters of primary interests. One can write:

$$p(\mathbf{G} | y, \mathbf{H}) = \int_{R_{\sigma_{E_o}^2}} \int_{R_{\sigma_{E_m}^2}} p(\mathbf{G} | \sigma_{E_m}^2, \sigma_{E_o}^2, y, \mathbf{H}) p(\sigma_{E_m}^2, \sigma_{E_o}^2 | y, \mathbf{H}) d\sigma_{E_m}^2 d\sigma_{E_o}^2 \quad [31]$$

where  $p(\sigma_{E_m}^2, \sigma_{E_o}^2 | y, \mathbf{H})$  is the posterior density of  $\sigma_{E_m}^2, \sigma_{E_o}^2$  obtained after integrating  $\mathbf{G}$  out of [11]. It seems impossible to carry out this integration analytically. Following ideas in Gianola and Fernando (1986), we propose as first approximation:

$$p(\mathbf{G} | y, \mathbf{H}) \approx p(\mathbf{G} | \sigma_{E_m}^2 = \hat{\sigma}_{E_m}^2, \sigma_{E_o}^2 = \hat{\sigma}_{E_o}^2, y, \mathbf{H}) \quad [32]$$

It would be better to use the modal values of  $p(\sigma_{E_m}^2, \sigma_{E_o}^2 | y, \mathbf{H})$  rather than  $\hat{\sigma}_{E_m}^2$  and  $\hat{\sigma}_{E_o}^2$ , but finding this distribution does not seem feasible. Using [32] in [11] one obtains:

$$p_1(\mathbf{G} | y, \mathbf{H}) \propto |\mathbf{G}|^{-\frac{1}{2}(n_a+a+3)} \exp \left[ \frac{1}{2} \left[ \text{tr}(\mathbf{G}^{-1}\mathbf{G}^*) + \frac{(y'y - \hat{\boldsymbol{\theta}}' \mathbf{W}'y)}{\hat{\sigma}_{E_o}^2} \right] \right] |\mathbf{W}'\mathbf{W} + \hat{\boldsymbol{\Sigma}}|^{-\frac{1}{2}} \quad [33]$$

It should be noted that now  $\hat{\boldsymbol{\theta}} = f(\mathbf{G}, \hat{\sigma}_{E_m}^2, \hat{\sigma}_{E_o}^2)$  and  $\hat{\boldsymbol{\Sigma}} = h(\mathbf{G}, \hat{\sigma}_{E_m}^2, \hat{\sigma}_{E_o}^2)$ . Then, the MCI method can be used to compute moments of [33]. The additional degree of marginalization with respect to [11] achieved in this approximation may be small, but savings in computing accrue because drawing values of  $\sigma_{E_m}^2$  and  $\sigma_{E_o}^2$  from  $I_2(\Gamma)$  and  $I_3(\Gamma)$ , respectively, is no longer necessary.

In the second approximation, we write the expression in the exponent of [33] as:

$$\begin{aligned} \text{tr}(\mathbf{G}^{-1}\mathbf{G}^*) + \frac{(y'y - \hat{\boldsymbol{\theta}}' \mathbf{W}'y)}{\hat{\sigma}_{E_o}^2} &= \text{tr}(\mathbf{G}^{-1}\mathbf{G}^*) + \text{tr}(I_2) \left[ \frac{y'y - \hat{\boldsymbol{\theta}}' \mathbf{W}'y}{2\hat{\sigma}_{E_o}^2} \right] \\ &= \text{tr} \left[ \mathbf{G}^{-1}\mathbf{G}^* + \mathbf{G} \frac{(y'y - \hat{\boldsymbol{\theta}}' \mathbf{W}'y)}{2\hat{\sigma}_{E_o}^2} \right] \end{aligned}$$

In the preceding, replace

$$\mathbf{G}^* + \mathbf{G} \frac{(y'y - \hat{\boldsymbol{\theta}}' \mathbf{W}'y)}{2\hat{\sigma}_{Eo}^2} = n_g \mathbf{G}_h + \mathbf{G} \left( \frac{y'y - \hat{\boldsymbol{\theta}}' \mathbf{W}'y}{2\hat{\sigma}_{Eo}^2} \right)$$

by

$$n_g + a \left[ \left( \frac{n_g}{n_g + a} \right) \hat{\mathbf{G}} + \hat{\mathbf{G}} \left( \frac{y'y - \hat{\boldsymbol{\theta}}' \mathbf{W}'y}{2\hat{\sigma}_{Eo}^2(n_g + a)} \right) \right] = (n_g + a) \mathbf{G}_*$$

where

$$\mathbf{G}_* = \left[ \left( \frac{n_g}{n_g + a} \right) \hat{\mathbf{G}} + \hat{\mathbf{G}} \left( \frac{y'y - \hat{\boldsymbol{\theta}}' \mathbf{W}'y}{2\hat{\sigma}_{Eo}^2(n_g + a)} \right) \right]$$

Defining

$$\mathbf{G}_*^* = (n_g + a) \mathbf{G}_*$$

and using the preceding developments in [33] we write, after neglecting  $|\mathbf{W}'\mathbf{W} + \hat{\boldsymbol{\Sigma}}|^{-1/2}$

$$p_2(\mathbf{G} | y, \mathbf{H}) \propto |\mathbf{G}|^{-\frac{1}{2}(n_g+a+3)} \exp \left[ -\frac{1}{2} \text{tr}(\mathbf{G}^{-1} \mathbf{G}_*^*) \right] \tag{34}$$

This density is in the inverted Wishart form, with parameters  $n'_g = n_g + a$  and  $\mathbf{G}_*^*$ , provided  $\mathbf{G}_*^*$  is positive definite. If not, one can “bend” this matrix following the ideas of Hayes and Hill (1981). The computational advantage of [34] over [33] is that  $y'y - \hat{\boldsymbol{\theta}}' \mathbf{W}'y$  would be evaluated only once at  $\hat{\mathbf{G}}, \hat{\sigma}_{Em}^2, \hat{\sigma}_{Eo}^2$ . Further, the inverted Wishart form of [34] yields an analytical solution for the (approximate) marginal posterior densities of  $\sigma_{Ao}^2$  and  $\sigma_{Am}^2$ , so approximate probability statements about elements of  $\mathbf{G}$  can be made with relative ease.

A third approximation would be writing [34] as

$$p_3(\mathbf{G} | y, \mathbf{H}) \propto |\mathbf{G}|^{-\frac{1}{2}(n_g+a+3)} \exp \left[ -\frac{1}{2} \text{tr}[\mathbf{G}^{-1} \hat{\mathbf{G}}(n_g + a)] \right] \tag{35}$$

so we would have an inverted Wishart distribution with hyperparameters  $n''_g = n_g + a$  and  $\hat{\mathbf{G}}$ . If  $\hat{\mathbf{G}}$  is obtained with an algorithm that guarantees positive semi-definiteness such as EM (Dempster *et al*, 1977), this would circumvent the potential problem posed by  $\mathbf{G}_*^*$  in [34].

The fourth approximation involves the matrix of second derivatives ( $\mathbf{C}$ , say) of the logarithm of [11] with respect to the unique elements of  $\mathbf{G}, \sigma_{Em}^2$  and  $\sigma_{Eo}^2$  and then evaluating  $\mathbf{C}$  at  $\hat{\mathbf{G}}, \hat{\sigma}_{Em}^2$  and  $\hat{\sigma}_{Eo}^2$ . The second derivatives are in the *Appendix*. Invoking the asymptotic normality property of posterior distributions (Zellner, 1971), one would approximately have :

$$\Gamma | y, \mathbf{H} \sim N[\hat{\Gamma}, (-\mathbf{C})^{-1}] \tag{36}$$

where it is assumed that the matrix  $-\mathbf{C} = f(\hat{\mathbf{G}}, \hat{\sigma}_{Em}^2, \hat{\sigma}_{Eo}^2)$  has full rank. The approximate marginal distributions of  $\sigma_{Ao}^2, \sigma_{Am}^2, \sigma_{AoAm}, \sigma_{Em}^2$  and  $\sigma_{Eo}^2$  follow directly from [36]: all are univariate normal.

## THE TIERNEY-KADANE APPROXIMATIONS

The approximation in [36] produces reasonable results when the posterior distribution is unimodal, which holds for large enough samples. Tierney and Kadane (1986) described another approximation (based on Laplace's method for integrals), and this is reviewed in the following section.

### Single parameter situation

Let  $g(\Gamma) = g$  be a positive function of the scalar parameter  $\Gamma$ . Then

$$E[g(\Gamma)|y] = \int g(\Gamma) p(\Gamma | y) d\Gamma \quad [37]$$

with

$$p(\Gamma | y) = l \cdot \pi \cdot c \quad [38]$$

where  $l$  is the likelihood function,  $\pi$  is the prior density and  $c$  is the integration constant

$$c = \left[ \int l\pi d\Gamma \right]^{-1} \quad [39]$$

With  $n$  being sample size, let

$$L = \frac{[\log(\pi) + \log(l)]}{n}$$

so that

$$\exp(\log(\pi) + \log(l)) = \pi l = \exp[nL] \quad [40]$$

Employing this in [39] and [38]:

$$c = \left[ \int \exp(nL) d\Gamma \right]^{-1} \quad [41]$$

and

$$p(\Gamma | y) = \frac{\exp[nL]}{\int \exp[nL] d\Gamma} \quad [42]$$

Then

$$\begin{aligned} g p(\Gamma | y) &= \frac{g \exp[nL]}{\int \exp[nL] d\Gamma} \\ &= \frac{\exp[\log(g) + nL]}{\int \exp[nL] d\Gamma} \end{aligned}$$

$$\begin{aligned}
&= \frac{\exp \left[ n \left[ \frac{\log(g)}{n} + L \right] \right]}{\int \exp [nL] d\Gamma} \\
&= \frac{\exp [nL^*]}{\int \exp [nL] d\Gamma}
\end{aligned} \tag{43}$$

where

$$L^* = \frac{\log(g)}{n} + L = \frac{\log(g) + \log(\pi) + \log(l)}{n} \tag{44}$$

Using [44] in [37]:

$$E[g | y] = \frac{\int \exp [nL^*] d\Gamma}{\int \exp [nL] d\Gamma} \tag{45}$$

The method of Tierney and Kadane (1986) continues as follows. Let  $\Gamma_m$  be the posterior mode (which is also the maximum of  $L$ ),  $L'(\Gamma)$  and  $L''(\Gamma)$  be the first and second derivatives of  $L$  with respect to  $\Gamma$  and let  $\sigma^2 = -1/L''(\Gamma_m)$ . Using a Taylor series expansion for  $nL(\Gamma)$  about  $\Gamma_m$  we have:

$$nL(\Gamma) = nL(\Gamma_m) + nL'(\Gamma_m)(\Gamma - \Gamma_m) - \left( \frac{n}{2\sigma^2} \right) (\Gamma - \Gamma_m)^2 + \dots$$

Noting that

$$L'(\Gamma_m) = \left[ \frac{\partial L(\Gamma)}{\partial \Gamma} \right]_{\Gamma=\Gamma_m} = 0$$

and retaining terms up to second-order, the expansion becomes:

$$nL(\Gamma) \approx nL(\Gamma_m) - \left( \frac{n(\Gamma - \Gamma_m)^2}{2\sigma^2} \right)$$

Using this, the denominator in [45] can be approximated as:

$$\begin{aligned}
\int \exp [nL(\Gamma)] d\Gamma &\approx \int \exp \left[ nL(\Gamma_m) - \left( \frac{n(\Gamma - \Gamma_m)^2}{2\sigma^2} \right) \right] d\Gamma \\
&= (2\pi)^{\frac{1}{2}} \sigma n^{-\frac{1}{2}} \exp [nL(\Gamma_m)]
\end{aligned} \tag{46}$$

In the same way, if  $\Gamma_m^*$  is the maximum of  $L^*$  and  $\sigma^{*2} = -1/L^{*''}(\Gamma_m^*)$

$$\int \exp [nL^*(\Gamma)] d\Gamma = (2\pi)^{-\frac{1}{2}} \sigma^* n^{-\frac{1}{2}} \exp [nL^*(\Gamma_m^*)] \tag{47}$$



Taking the ratio between [47] and [46] as required in [45] then, approximately, we have:

$$\widehat{E}[g(\Gamma) | y] \approx \left(\frac{\sigma^*}{\sigma}\right) \exp [n[L^*(\Gamma_m^*) - L(\Gamma_m)]] \quad [48]$$

An interesting aspect of this approximation is that only first and second order derivatives are needed, and this is less tedious than other approximations suggested by *eg*, Mosteller and Wallace (1964) and Lindley (1980), requiring evaluation of third derivatives. The posterior variance can also be approximated by finding the posterior mean of  $g^2$ . The only modification needed is to define  $L^*$  as

$$L^* = \frac{\log(g^2) + \log(\pi) + \log(l)}{n}$$

Then

$$\widehat{\text{Var}}[g(\Gamma) | y] \approx \widehat{E}[g^2(\Gamma) | y] - \widehat{E}^2[g(\Gamma) | y] \quad [49]$$

**The multiparameter case**

When  $\Gamma$  is a vector, as in this paper, [48] generalizes to:

$$\widehat{E}[g(\Gamma) | y] \approx |\mathbf{H}^* \mathbf{H}^{-1}|^{\frac{1}{2}} \exp [n[L^*(\Gamma_m^*) - L(\Gamma_m)]] \quad [50]$$

where  $\Gamma_m^*$  and  $\Gamma_m$  maximize  $L^*$  and  $L$ , respectively, and  $\mathbf{H}^*$  and  $\mathbf{H}$  are minus the inverse matrices of second derivatives of  $L^*$  and  $L$  with respect to  $\Gamma$ , evaluated at  $\Gamma_m^*$  and  $\Gamma_m$ , respectively.

**Marginal posterior densities**

The method can also be used to approximate marginal posterior densities of individual parameters of  $\Gamma$ . Partition  $\Gamma'$  as  $[\Gamma_1, \Gamma_2']$ . If the order of  $\Gamma$  is  $p$ , say, then  $\Gamma_2$  is of order  $p - 1$  (4 in our case). The marginal posterior density of  $\Gamma_1$  is:

$$p(\Gamma_1 | y) = \frac{\int \exp [\log(l)] \pi(\Gamma_1, \Gamma_2) d\Gamma_2}{\int \exp [\log(l)] \pi(\Gamma) d\Gamma} \quad [51]$$

where  $\pi(\Gamma_1, \Gamma_2)$  is the joint posterior density of  $\Gamma$ . From preceding developments, the denominator in [51] is expressible as:

$$\begin{aligned} & \int \exp [\log(l)] \pi(\Gamma) d\Gamma = \int \exp [nL(\Gamma)] d\Gamma \\ & \approx \int \exp [nL(\Gamma_m) - \frac{n}{2}(\Gamma - \Gamma_m)'[-L''(\Gamma_m)](\Gamma - \Gamma_m)] d\Gamma \end{aligned} \quad [52]$$

where  $\Gamma_m$  is the mode of the posterior distribution of  $\Gamma$ , and  $L''(\Gamma_m)$  is the matrix of second derivatives of  $L$  with respect to  $\pi$ . Then:

$$\begin{aligned} \int \exp[\log(l)]\pi(\Gamma) d\Gamma &\approx (2\pi)^{\frac{p}{2}} | -nL''(\Gamma_m) |^{-\frac{1}{2}} \exp[nL(\Gamma_m)] \\ &= (2\pi)^{\frac{p}{2}} n^{-\frac{p}{2}} | -L''(\Gamma_m) |^{-\frac{1}{2}} \exp[nL(\Gamma_m)] \end{aligned} \quad [53]$$

However

$$\exp[nL(\Gamma_m)] = \pi(\Gamma_m) \exp[\log[l(\Gamma_m)]]$$

where  $l(\Gamma_m)$  is the log-likelihood evaluated at  $\Gamma_m$ . Hence, [53] becomes:

$$\int \exp[\log(l)]\pi(\Gamma) d\Gamma \approx (2\pi)^{\frac{p}{2}} n^{-\frac{p}{2}} | -L''(\Gamma_m) |^{-\frac{1}{2}} \pi(\Gamma_m) \exp[\log[l(\Gamma_m)]] \quad [54]$$

Consider now the numerator of [51], and write it as:

$$\int \exp[\log(l)]\pi(\Gamma_1, \Gamma_2) d\Gamma_2 = \int \exp[nB(\Gamma_1, \Gamma_2)] d\Gamma_2 \quad [55]$$

where

$$B(\Gamma_1, \Gamma_2) = \frac{\log[\pi(\Gamma_1, \Gamma_2)] + \log[l(\Gamma_1, \Gamma_2)]}{n}$$

is a function where  $\Gamma_1$  is fixed. Define  $\Gamma_{2m}(\Gamma_1)$  to be the  $(p-1) \times 1$  vector that maximizes this function. This maximizer can be found employing the derivatives in the appendix. Then, similar to [53], we can write.

$$\begin{aligned} \int \exp[nB(\Gamma_1, \Gamma_2)] d\Gamma_2 &\approx (2\pi)^{\frac{p-1}{2}} n^{-\frac{p-1}{2}} | \\ &\quad -B''(\Gamma_{2m}(\Gamma_1)) |^{-\frac{1}{2}} \exp[nB(\Gamma_1, \Gamma_{2m}(\Gamma_1))] \\ &= (2\pi)^{\frac{p-1}{2}} n^{-\frac{p-1}{2}} | -B''(\Gamma_{2m}(\Gamma_1)) |^{-\frac{1}{2}} \pi(\Gamma_1, \Gamma_{2m}(\Gamma_1)) \\ &\quad \exp[\log[l(\Gamma_1, \Gamma_{2m}(\Gamma_1))] \end{aligned} \quad [56]$$

where  $B''(\Gamma_1, \Gamma_{2m}(\Gamma_1))$  is the  $(p-1) \times (p-1)$  matrix of second derivatives of  $B$  with respect to  $\Gamma_2$ .

Taking the ratio between [56] and [54] the posterior density of  $\Gamma_1$  in [51] is approximately:

$$\hat{p}(\Gamma_1 | y) \approx \frac{\pi(\Gamma_1, \Gamma_{2m}(\Gamma_1)) \exp[\log[l(\Gamma_1, \Gamma_{2m}(\Gamma_1))] | -L''(\Gamma_m)]^{\frac{1}{2}}}{\pi(\Gamma_m) \exp[\log[l(\Gamma_m)]] | -B''(\Gamma_{2m}(\Gamma_1)) |^{\frac{1}{2}}} \left(\frac{n}{2\pi}\right)^{\frac{1}{2}} \quad [57]$$

The moments of the posterior distribution of  $\Gamma_1$ , must be found numerically employing the methods discussed in earlier sections.

### Remarks

It has been shown that the method of Tierney and Kadane (1986) has less error than the usual normal approximation centered at the posterior mode with the order of approximation being  $O(n^{-2})$ . However, it also requires that the functions to be expanded be either unimodal or dominated by a single mode, so sample size must be sufficiently large for this to hold.

The requirement that  $g(\Gamma)$  be a positive function is restrictive. Tierney and Kadane (1986) pointed out that for the approximation to be accurate for a function  $g$  taking both positive and negative values, the posterior distribution of  $g$  must be concentrated almost entirely on one side of the origin. However, Tierney *et al* (1988) extended the method to apply to expectations and variances of non-positive functions. To obtain a second-order approximation to  $\mathbf{E}[g(\Gamma)]$ , they used the method of Tierney and Kadane (1986) to approximate the moment generating function  $\mathbf{E}\{\exp[sg(\Gamma)]\}$ , whose integrand is positive, and then the result was differentiated.

Another difficulty arises in the approximation [49] to the posterior variance of  $g(\Gamma)$ . Unless computations are made with sufficient precision, [49] can have a large error or turn up negative. Similar problems can arise in the computations of posterior covariance, *ie*

$$\widehat{\text{Cov}}[g(\Gamma), h(\Gamma) | y] \approx \widehat{E}[g(\Gamma) h(\Gamma) | y] - \widehat{E}[g(\Gamma) | y] \widehat{E}[h(\Gamma) | y] \quad [58]$$

as a covariance matrix computed from [58] may not be positive semi-definite.

### CONCLUSION

This paper presents theory and techniques for carrying out a Bayesian analysis of dispersion parameters in a univariate model for maternal effects. However, implementation of the methods suggested here poses difficulties to quantitative geneticists interested in analysis of large data sets. The development of feasible computing techniques is a challenge to researchers in the area of application of numerical methods to animal breeding.

Research is underway to identify more promising algorithms to approximate marginal moments of posterior distributions, a non-trivial problem as new techniques are developed and there is little indication on the choice to make for estimating (co)variance components under "non-exchangeability" of model [1]. Recently Gelfand and Smith (1990) and Gelfand *et al* (1990) described the Gibbs sampler, a potential competitor of the methods presented here.

### REFERENCES

- Anderson TW (1984) *An Introduction to Multivariate Statistical Analysis*. J Wiley and Sons, New York, NY
- Bauwens L (1984) *Bayesian Full Information Analysis of Simultaneous Equations Models Using Integration by Monte Carlo*. Springer-Verlag, Berlin

- Bauwens L, Richard JF (1985) A 1-1 poly-t random variable generator with application to Monte Carlo integration. *J Econometrics* 29, 19-46
- Box GEP, Tiao GC (1973) *Bayesian Inference in Statistical Analysis*. Addison-Wesley Publishing Co, Reading, MA
- Broemeling LD (1985) *Bayesian Analysis of Linear Models*. Marcel Dekker, New York, NY
- Bulmer MG (1985) *The Mathematical Theory of Quantitative Genetics*. Clarendon Press, Oxford, UK, 2nd edn
- Cantet RJC, Kress DD, Anderson DC, Doornbos DE, Burfening PJ, Blackwell RL (1988) Direct and maternal variances and covariances and maternal phenotypic effects on preweaning growth of beef cattle. *J Anim Sci* 66, 648-660
- Carriquiry AL (1989) Bayesian prediction and its application to the genetic evaluation of livestock. Ph D thesis, Iowa State University, Ames, IA
- Chen CF (1979) Bayesian inference for a normal dispersion matrix and its application to stochastic multiple regression analysis. *J R Stat Soc Ser B* 41, 235-248
- Cheverud JM (1984) Evolution by kin selection: a quantitative genetic model illustrated by maternal performance in mice. *Evolution* 38, 766-777
- Cramer JS (1986) *Econometric Applications of Maximum Likelihood Methods*. Cambridge University Press, Cambridge, UK
- Dempster AP, Laird NM, Rubin DB (1977) Maximum likelihood from incomplete data via the EM algorithm. *J R Stat Soc Ser B* 39, 1-38
- Dickerson GE (1947) Composition of hog carcasses as influenced by heritable differences in rate and economy of gain. *Iowa Agric Exp Stat, Ames, IA, Res Bull* 354, 489-524
- Falconer DS (1965) Maternal effects and selection response. In: *Genetics Today. Proc XIth Int Congr Genetics*. The Hague, The Netherlands, 763-774
- Falconer DS (1981) *Introduction to Quantitative Genetics*. Longman, London, UK
- Foulley JL, Im S, Gianola D, Hoschele I (1987) Empirical Bayes estimation of parameters for *n* polygenic binary traits. *Genet Sel Evol* 19, 197-224
- Foulley JL, Lefort G (1978) Methodes d'estimation des effets directs et maternels en sélection animale. *Ann Genet Sel Anim* 10, 475-496
- Gelfand AE, Smith AFM (1990) Sampling-based approaches to calculating marginal densities. *J Am Stat Assoc* 85, 398-409
- Gelfand AE, Hills SE, Racine-Poon A, Smith AFM (1990) Illustration of Bayesian inference in normal data models using Gibbs sampling. *J Am Stat Assoc* 85, 972-985
- Geweke J (1988) Antithetic acceleration of Monte Carlo integration in Bayesian inference. *J Econometrics* 38, 73-89
- Gianola D, Fernando RL (1986) Bayesian methods in animal breeding. *J Anim Sci* 63, 217-244
- Gianola D, Im S, Fernando RL, Foulley JL (1990a) Mixed model methodology and the Box-Cox theory of transformations: a Bayesian approach. In: *Advances in Statistical Methods for the Genetic Improvement of Livestock* (Gianola D, Hammond K, eds) Springer-Verlag, Berlin, 15-40
- Gianola D, Im S, Macedo FW (1990b) A framework for prediction of breeding value. In: *Advances in Statistical Methods for the Genetic Improvement of Livestock* (Gianola D, Hammond K, eds) Springer-Verlag, Berlin, 210-238

- Haber S (1970) Numerical evaluation of multiple integrals. *SIAM Rev* 12, 481-526
- Hammersley JM, Handscomb DC (1964) *Monte Carlo Methods*. Methuen, London, UK
- Harville DA (1974) Bayesian inference for variance components using only error contrasts. *Biometrika* 61, 383-384
- Harville DA (1977) Maximum likelihood approaches to variance component estimation and to related problems. *J Am Stat Assoc* 72, 320-340
- Hayes JF, Hill WG (1981) Modification of estimates of parameters in the construction of genetic selection indices ("bending"). *Biometrics* 37, 483-493
- Henderson CR (1984) *Applications of Linear Models in Animal Breeding*. Univ of Guelph, Guelph, Canada
- Henderson CR (1988) Theoretical basis and computational methods for a number of different animal models. *J Dairy Sci* 71, 1-16
- Hildebrand FB (1972) *Advanced Calculus for Applications*. Addison-Wesley Publishing Co, Reading, MA, 2nd edn
- Hoeschele I, Gianola D, Foulley JL (1987) Estimation of variance components with quasi-continuous data using Bayesian methods. *J Anim Breed Genet* 104, 334-349
- Jaynes ET (1957) Information theory and statistical mechanics. *Physiol Rev* 106, 620-630
- Jaynes ET (1979) Where do we stand on maximum entropy. In: *The Maximum Entropy Formalism* (Levine RD, Tribus M, eds) The MIT Press, Cambridge, MA
- Kempthorne O (1955) The correlation between relatives in random mating populations. *Cold Spring Harbor Symp Quant Biol* 22, 60-78
- Kirkpatrick M, Lande R (1969) The evolution of maternal characters. *Evolution* 43, 585-503
- Kloek T, Van Dijk HK (1978) Bayesian estimates of equation system parameters: an application of integration by Monte Carlo. *Econometrica* 46, 1-19
- Lande R, Price T (1989) Genetic correlations and maternal effect coefficients obtained from offspring-parent regression. *Genetics* 122, 915-922
- Lindley DV (1980) Approximate Bayesian methods. In: *Bayesian Statistics* (Bernardo JM, Degroot MH, Lindley DV, Smith AFM, eds) Valencia University Press, Valencia, Spain
- Lindley DV, Smith AFM (1972) Bayes estimates for the linear model. *J R Stat Soc Ser B* 34, 1-18
- Macedo FW, Gianola D (1987) Bayesian analysis of univariate mixed models with informative priors. Eur Assoc Animal Prod XVIII Annu Meet, Lisbon, Portugal
- Mead LR, Papanicolaou N (1984) Maximum entropy in the problem of moments. *J Math Phys* 25, 2404-2417
- Misztal I (1990) Restricted maximum likelihood estimation of variance components in animal model using sparse matrix inversion and a supercomputer. *J Dairy Sci* 73, 163-172
- Mosteller F, Wallace D (1964) *Inference and Disputed Authorship: The Federalist*. Addison-Wesley Publishing Co, Reading, MA
- Naylor JC, Smith AFM (1982) Applications of a method for the efficient computation of posterior distributions. *Appl Stat* 31, 214-225
- Naylor JC, Smith AFM (1988) Econometric illustrations of novel numerical integration strategies for Bayesian inference. *J Econometrics* 38, 103-125

- Quaas RL, Pollak EJ (1980) Mixed model methodology for farm and ranch beef cattle testing program. *J Anim Sci* 51, 1277-1287
- Raiffa H, Schlaiffer R (1961) *Applied Statistical Decision Theory*. Harvard University Press, Boston, MA
- Richard JF, Steel MFJ (1988) Bayesian analysis of systems of seemingly unrelated regression equations under a recursive extended natural conjugate prior density. *J Econometrics* 38, 7-37
- Riska B, Rutledge JJ, Atchley WR (1985) Covariance between direct and maternal genetic effects in mice, with a model of persistent environmental influences. *Genet Res* 45, 287-297
- Rubinstein RY (1981) *Simulation and the Monte Carlo Method*. J Wiley & Sons, New York, NY
- Searle SR (1979) Notes on variance component estimation. A detailed account of maximum likelihood and kindred methodology. Paper BU-673-M, Biometrics Unit, Cornell University, Ithaca, NY
- Shannon CE (1948) A mathematical theory of communications. *Bell Syst Tech J* 27, 379-423; reprinted In: *The Mathematical Theory of Communications* (Shannon CE, Weaver W, eds) The University of Illinois Press, Urbana, IL
- Thompson R (1980) Maximum likelihood estimation of variance components. *Math Operationforsch Stat*, 11, 95-103
- Tierney L, Kadane JB (1986) Accurate approximations for posterior moments and marginal densities. *J Am Stat Assoc* 81, 82-86
- Tierney L, Kass RE, Kadane JB (1986) Fully exponential Laplace approximations of expectations and variances of non-positive functions. Tech Rep 418, Dept Stat, Carnegie Mellon University, Pittsburgh, PA, USA
- 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. *J Anim Sci* 35, 1288-1293
- Wright DE (1986) A note on the construction of highest posterior density intervals. *Appl Stat* 35, 49-53
- Zellner A (1971) *An Introduction to Bayesian Inference in Econometrics*. J Wiley & Sons, New York, NY
- Zellner A, Bauwens L, Van Dijk HK (1988) Bayesian specification analysis and estimation of simultaneous equations models using Monte Carlo methods. *J Econometrics* 38, 39-72
- Zellner A, Highfield RA (1988) Calculation of maximum entropy distributions and approximation of marginal posterior distributions. *J Econometrics* 37, 195-209
- Zellner A, Rossi P (1984) Bayesian analysis of dichotomous quantal response models. *J Econometrics* 23, 365-394

**APPENDIX**

**First and second derivatives of the log-posterior of all (co)variance components**

The log of [11] is:

$$\begin{aligned} \log [p(\Gamma | y)] = & -1/2\{[(y'y - \hat{\theta}'\mathbf{W}'y + n_o s_{E_o}^2)/\sigma_{E_o}^2] + \log |\mathbf{W}'\mathbf{W} + \Sigma| \\ & + (n_g + a + 3) \log |\mathbf{G}| + \text{tr}(\mathbf{G}^{-1}\mathbf{G}^*) + n_m s_{E_m}^2/\sigma_{E_m}^2 \quad \text{[A.1]} \\ & + (n_o + n - p - 2a - d + 2) \log [\sigma_{E_o}^2] + (n_m + d + 2) \log [\sigma_{E_m}^2]\} + \text{constant} \end{aligned}$$

Let  $\mathbf{C} = (\mathbf{W}'\mathbf{W} + \Sigma)^{-1}$  be partitioned as

$$\mathbf{C} = \begin{bmatrix} \mathbf{C}^{\beta\beta} & \mathbf{C}^{\beta a} & \mathbf{C}^{\beta e} \\ \mathbf{C}^{a\beta} & \mathbf{C}^{aa} & \mathbf{C}^{ae} \\ \mathbf{C}^{e\beta} & \mathbf{C}^{ea} & \mathbf{C}^{ee} \end{bmatrix} \quad \text{[A.2]}$$

Let  $\mathbf{M}' = [0 | \mathbf{I}_{2a} | 0]$  be a  $2a \times (p + 2a + d)$  matrix such that  $\mathbf{M}'\hat{\theta} = \hat{\mathbf{a}}$ . In the same way,  $\mathbf{N}' = [0 | 0 | \mathbf{I}_d]$  be a  $d \times (p + 2a + d)$  matrix such that  $\mathbf{N}'\hat{\theta} = \hat{\mathbf{e}}_m$ . The  $\mathbf{0}$  represents a matrix of appropriate order with all elements equal to zero.

To simplify the derivation, we will decompose [A.1] into components, take derivatives with respect to an element of  $\mathbf{G}$  ( $g_{ij}$  say),  $\sigma_{E_m}^2$  or  $\sigma_{E_o}^2$ , and collect results to obtain the desired expressions.

**Derivatives of  $(y'y - \hat{\theta}'\mathbf{W}'y)$**

The term  $y'y$  does not depend on  $\Gamma$ . The other term is

$\hat{\theta}'\mathbf{W}'y = y'\mathbf{W}\mathbf{C}\mathbf{W}'y$ , so that:

$$\begin{aligned} \partial \hat{\theta}'\mathbf{W}'y / \partial g_{ij} &= y'\mathbf{W}[\partial \mathbf{C} / \partial g_{ij}] \mathbf{W}'y \\ &= y'\mathbf{W}[-\mathbf{C}(\partial(\mathbf{W}'\mathbf{W} + \Sigma) / \partial g_{ij})\mathbf{C}] \mathbf{W}'y \\ &= -y'\mathbf{W}\mathbf{C}(\partial \Sigma / \partial g_{ij})\mathbf{C}\mathbf{W}'y \\ &= -y'\mathbf{W}\mathbf{C}\mathbf{D}[0, -\mathbf{G}^{-1}\mathbf{E}_{ij}\mathbf{G}^{-1} \otimes \mathbf{A}^{-1}, 0]\mathbf{C}\mathbf{W}'y\sigma_{E_o}^2 \end{aligned}$$

where  $\mathbf{E}_{ij}$  is a  $2 \times 2$  matrix with all elements equal to zero, with the exception of a one in position  $i, j$ . Note that if  $\mathbf{e}_i$  ( $\mathbf{e}_j$ ) is a  $2 \times 1$  vector with a 1 in the  $i$ -th ( $j$ -th) position  $\mathbf{E}_{ij} = \mathbf{e}_i\mathbf{e}_j'$ . The notation  $\mathbf{D}[\mathbf{M}_1, \dots, \mathbf{M}_s]$  stands for a block diagonal matrix with the  $s$  blocks being equal to  $\mathbf{M}_i, (i = 1, \dots, s)$ . Since  $\hat{\theta} = \mathbf{C}\mathbf{W}'y$  and  $\hat{\theta}' = [\hat{\beta}' | \hat{\mathbf{a}} | \hat{\mathbf{e}}_m]$ , we can write the above expression as:

$$\begin{aligned} &= \hat{\theta}'\mathbf{D}[0, \mathbf{G}^{-1}\mathbf{E}_{ij}\mathbf{G}^{-1} \otimes \mathbf{A}^{-1}, 0]\hat{\theta}\sigma_{E_o}^2 \\ &= \hat{\mathbf{a}}'(\mathbf{G}^{-1}\mathbf{E}_{ij}\mathbf{G}^{-1} \otimes \mathbf{A}^{-1})\hat{\mathbf{a}}\sigma_{E_o}^2 \quad \text{[A.3]} \end{aligned}$$

In a similar way

$$\begin{aligned}\partial\hat{\boldsymbol{\theta}}'\mathbf{W}'y/\partial\sigma_{E_m}^2 &= -\hat{\boldsymbol{\theta}}'\mathbf{D}[0, 0, -\mathbf{I}_d(\sigma_{E_m}^2)^{-2}]\hat{\boldsymbol{\theta}}\sigma_{E_o}^2 \\ &= \hat{\mathbf{e}}'_m\hat{\mathbf{e}}_m(\sigma_{E_m}^2)^{-2}\sigma_{E_o}^2\end{aligned}\quad [\text{A.4}]$$

and

$$\begin{aligned}\partial\hat{\boldsymbol{\theta}}'\mathbf{W}'y/\partial\sigma_{E_o}^2 &= -\hat{\boldsymbol{\theta}}'\mathbf{D}[0, \mathbf{G}^{-1} \otimes \mathbf{A}^{-1}, \mathbf{I}_d(\sigma_{E_m}^2)^{-1}]\hat{\boldsymbol{\theta}} \\ &= -[\hat{\mathbf{a}}'(\mathbf{G}^{-1} \otimes \mathbf{A}^{-1})\hat{\mathbf{a}} + \hat{\mathbf{e}}'_m\hat{\mathbf{e}}_m(\sigma_{E_m}^2)^{-1}]\end{aligned}\quad [\text{A.5}]$$

Second derivatives are obtained from [A.3] to [A.5].

$$\begin{aligned}\partial^2(\hat{\boldsymbol{\theta}}'\mathbf{W}'y)/(\partial g_{ij})^2 &= \{2\hat{\mathbf{a}}'(\mathbf{G}^{-1}\mathbf{E}_{ij}\mathbf{G}^{-1} \otimes \mathbf{A}^{-1})(\partial\hat{\mathbf{a}}/\partial g_{ij}) \\ &\quad + \hat{\mathbf{a}}'[\partial(\mathbf{G}^{-1}\mathbf{E}_{ij}\mathbf{G}^{-1} \otimes \mathbf{A}^{-1})/\partial g_{ij}]\hat{\mathbf{a}}\}\sigma_{E_o}^2 \\ &= 2\{\hat{\mathbf{a}}'(\mathbf{G}^{-1}\mathbf{E}_{ij}\mathbf{G}^{-1} \otimes \mathbf{A}^{-1})[\mathbf{M}'(\partial\mathbf{C}/\partial g_{ij})\mathbf{W}'y] - \hat{\mathbf{a}}'(\mathbf{G}^{-1}\mathbf{E}_{ij}\mathbf{G}^{-1} \otimes \mathbf{A}^{-1})\hat{\mathbf{a}}\}\sigma_{E_o}^2 \\ &= 2\hat{\mathbf{a}}'[(\mathbf{G}^{-1}\mathbf{E}_{ij}\mathbf{G}^{-1} \otimes \mathbf{A}^{-1})\mathbf{M}'\mathbf{C}\mathbf{D}[0, (\mathbf{G}^{-1}\mathbf{E}_{ij}\mathbf{G}^{-1} \otimes \mathbf{A}^{-1}), \mathbf{0}] \\ &\quad \mathbf{C}\mathbf{W}'y\sigma_{E_o}^2](\mathbf{G}^{-1}\mathbf{E}_{ij}\mathbf{G}^{-1}\mathbf{E}_{ij}\mathbf{G}^{-1} \otimes \mathbf{A}^{-1})\hat{\mathbf{a}}\}\sigma_{E_o}^2 \\ &= 2\hat{\mathbf{a}}'[(\mathbf{G}^{-1}\mathbf{E}_{ij}\mathbf{G}^{-1} \otimes \mathbf{A}^{-1})\mathbf{C}^{aa}(\mathbf{G}^{-1}\mathbf{E}_{ij}\mathbf{G}^{-1} \otimes \mathbf{A}^{-1}) \\ &\quad \sigma_{E_o}^2 - (\mathbf{G}^{-1}\mathbf{E}_{ij}\mathbf{G}^{-1}\mathbf{E}_{ij}\mathbf{G}^{-1} \otimes \mathbf{A}^{-1})]\hat{\mathbf{a}}\sigma_{E_o}^2\end{aligned}\quad [\text{A.6}]$$

Also

$$\begin{aligned}\partial^2(\hat{\boldsymbol{\theta}}'\mathbf{W}'y)/(\partial\sigma_{E_m}^2)^2 &= \{2\hat{\mathbf{e}}'_m(\partial\hat{\mathbf{e}}_m/\partial\sigma_{E_m}^2)(\sigma_{E_m}^2)^{-2} - 2\hat{\mathbf{e}}'_m\hat{\mathbf{e}}_m\}^{-3}\sigma_{E_o}^2 \\ &= 2\{\hat{\mathbf{e}}'_m\mathbf{N}'(-\mathbf{C}\mathbf{D}[0, \mathbf{0}, -\mathbf{I}_d(\sigma_{E_m}^2)^{-2}]\sigma_{E_o}^2)\mathbf{C}\mathbf{W}'y(\sigma_{E_m}^2)^{-2} - \hat{\mathbf{e}}'_m\hat{\mathbf{e}}_m(\sigma_{E_m}^2)^{-3}\}\sigma_{E_o}^2 \\ &= 2\{\hat{\mathbf{e}}'_m\mathbf{C}^{ee}\hat{\mathbf{e}}_m(\sigma_{E_m}^2)^{-4}\sigma_{E_o}^2 - \hat{\mathbf{e}}'_m\hat{\mathbf{e}}_m(\sigma_{E_m}^2)^{-3}\}\sigma_{E_o}^2 \\ &= 2(\sigma_{E_m}^2)^{-3}\sigma_{E_o}^2\hat{\mathbf{e}}'_m(\mathbf{C}^{ee}(\sigma_{E_m}^2)^{-1}\sigma_{E_o}^2 - \mathbf{I}_d)\hat{\mathbf{e}}_m\end{aligned}\quad [\text{A.7}]$$

For the error component we have

$$\begin{aligned}\partial^2(\hat{\boldsymbol{\theta}}'\mathbf{W}'y)/(\partial\sigma_{E_o}^2)^2 &= -2\hat{\boldsymbol{\theta}}'\mathbf{D}[0, \mathbf{G}^{-1} \otimes \mathbf{A}^{-1}, \mathbf{I}_d(\sigma_{E_m}^2)^{-1}](\partial\hat{\boldsymbol{\theta}}/\partial\sigma_{E_o}^2) \\ &= -2\hat{\boldsymbol{\theta}}'\mathbf{D}[0, \mathbf{G}^{-1} \otimes \mathbf{A}^{-1}, \mathbf{I}_d(\sigma_{E_m}^2)^{-1}](-\mathbf{C}\mathbf{D}[0, \mathbf{G}^{-1} \otimes \mathbf{A}^{-1}, \mathbf{I}_d(\sigma_{E_m}^2)^{-1}])\hat{\boldsymbol{\theta}} \\ &= 2[\boldsymbol{\theta}'|\hat{\mathbf{a}}'(\mathbf{G}^{-1} \otimes \mathbf{A}^{-1})|\hat{\mathbf{e}}'_m(\sigma_{E_m}^2)^{-1}]\mathbf{C}[0|(\mathbf{G}^{-1} \otimes \mathbf{A}^{-1})\hat{\mathbf{a}}|\hat{\mathbf{e}}_m(\sigma_{E_m}^2)^{-1}] \\ &= 2[\hat{\mathbf{a}}'(\mathbf{G}^{-1} \otimes \mathbf{A}^{-1})\mathbf{C}^{aa}(\mathbf{G}^{-1} \otimes \mathbf{A}^{-1})\hat{\mathbf{a}} + 2\hat{\mathbf{a}} \\ &\quad + 2\hat{\mathbf{a}}'(\mathbf{G}^{-1} \otimes \mathbf{A}^{-1})\mathbf{C}^{ae}\hat{\mathbf{e}}_m(\sigma_{E_m}^2)^{-1} + \hat{\mathbf{e}}'_m\mathbf{C}^{ee}\hat{\mathbf{e}}_m(\sigma_{E_m}^2)^2]\end{aligned}\quad [\text{A.8}]$$



Additional second derivatives are:

$$\begin{aligned}
 \partial^2(\widehat{\boldsymbol{\theta}}' \mathbf{W}' y) / \partial g_{ij} \partial \sigma_{Em}^2 &= 2\widehat{\mathbf{a}}' (\mathbf{G}^{-1} \mathbf{E}_{ij} \mathbf{G}^{-1} \otimes \mathbf{A}^{-1}) (\partial \widehat{\mathbf{a}} / \partial \sigma_{Em}^2) \sigma_{Eo}^2 \\
 &= 2\widehat{\mathbf{a}}' (\mathbf{G}^{-1} \mathbf{E}_{ij} \mathbf{G}^{-1} \otimes \mathbf{A}^{-1}) [\mathbf{M}' \mathbf{C} \mathbf{D} [\mathbf{0}, \mathbf{0}, \mathbf{I}_d (\sigma_{Em}^2)^{-2} \sigma_{Eo}^2] \widehat{\boldsymbol{\theta}}] \sigma_{Eo}^2 \\
 &= 2\widehat{\mathbf{a}}' (\mathbf{G}^{-1} \mathbf{E}_{ij} \mathbf{G}^{-1} \otimes \mathbf{A}^{-1}) \mathbf{C}^{ae} \widehat{\mathbf{e}}_m (\sigma_{Eo}^2 / \sigma_{Em}^2)^2
 \end{aligned} \tag{A.9}$$

$$\begin{aligned}
 \partial^2(\widehat{\boldsymbol{\theta}}' \mathbf{W}' y) / \partial g_{ij} \partial \sigma_{Eo}^2 &= 2\widehat{\mathbf{a}}' (\mathbf{G}^{-1} \mathbf{E}_{ij} \mathbf{G}^{-1} \otimes \mathbf{A}^{-1}) (\partial \widehat{\mathbf{a}} / \partial \sigma_{Eo}^2) \sigma_{Eo}^2 + \widehat{\mathbf{a}}' (\mathbf{G}^{-1} \mathbf{E}_{ij} \mathbf{G}^{-1} \otimes \mathbf{A}^{-1}) \widehat{\mathbf{a}} \\
 &= -2\widehat{\mathbf{a}}' (\mathbf{G}^{-1} \mathbf{E}_{ij} \mathbf{G}^{-1} \otimes \mathbf{A}^{-1}) [\mathbf{C}^{aa} (\mathbf{G}^{-1} \otimes \mathbf{A}^{-1}) \widehat{\mathbf{a}} + \mathbf{C}^{ee} \widehat{\mathbf{e}}_m (\sigma_{Em}^2)^{-1}] \sigma_{Eo}^2 \\
 &\quad + \widehat{\mathbf{a}}' (\mathbf{G}^{-1} \mathbf{E}_{ij} \mathbf{G}^{-1} \otimes \mathbf{A}^{-1}) \widehat{\mathbf{a}} \\
 &= \widehat{\mathbf{a}}' \{ (\mathbf{G}^{-1} \mathbf{E}_{ij} \mathbf{G}^{-1} \otimes \mathbf{A}^{-1}) [\mathbf{I}_{2a} - 2\mathbf{C}^{aa} (\mathbf{G}^{-1} \otimes \mathbf{A}^{-1}) \sigma_{Eo}^2] \} \\
 &\quad \widehat{\mathbf{a}} - 2\widehat{\mathbf{a}}' (\mathbf{G}^{-1} \mathbf{E}_{ij} \mathbf{G}^{-1} \otimes \mathbf{A}^{-1}) \mathbf{C}^{ee} \widehat{\mathbf{e}}_m (\sigma_{Em}^2)^{-1} \sigma_{Eo}^2
 \end{aligned} \tag{A.10}$$

$$\begin{aligned}
 \partial^2(\widehat{\boldsymbol{\theta}}' \mathbf{W}' y) / \partial \sigma_{Em}^2 \partial \sigma_{Eo}^2 &= 2\widehat{\mathbf{e}}_m' (\partial \widehat{\mathbf{e}}_m / \partial \sigma_{Eo}^2) (\sigma_{Em}^2)^{-2} \sigma_{Eo}^2 + \widehat{\mathbf{e}}_m' \widehat{\mathbf{e}}_m (\sigma_{Em}^2)^{-2} \\
 &= 2\widehat{\mathbf{e}}_m' [-\mathbf{C}^{ea} (\mathbf{G}^{-1} \otimes \mathbf{A}^{-1}) \widehat{\mathbf{a}} - \mathbf{C}^{ee} \widehat{\mathbf{e}}_m (\sigma_{Em}^2)^{-2}] \sigma_{Eo}^2 + \widehat{\mathbf{e}}_m' \widehat{\mathbf{e}}_m (\sigma_{Em}^2)^{-2} \\
 &= \widehat{\mathbf{e}}_m' [\mathbf{I}_d - 2\mathbf{C}^{ee} \sigma_{Eo}^2] \widehat{\mathbf{e}}_m (\sigma_{Em}^2)^{-2} - 2\widehat{\mathbf{e}}_m' \mathbf{C}^{ea} (\mathbf{G}^{-1} \otimes \mathbf{A}^{-1}) \widehat{\mathbf{a}} \sigma_{Eo}^2
 \end{aligned} \tag{A.11}$$

### Derivatives of $\log |\mathbf{W}' \mathbf{W} + \boldsymbol{\Sigma}|$

We use the result in Searle (1979):

$$\frac{\partial \log |\mathbf{Q}|}{\partial x} = \text{tr} \left[ \mathbf{Q}^{-1} \frac{\partial \mathbf{Q}}{\partial x} \right] \tag{A.12}$$

Using [A.12], the derivative of  $\log |\mathbf{W}' \mathbf{W} + \boldsymbol{\Sigma}|$  with respect to  $g_{ij}$  is

$$\begin{aligned}
 \partial \log |\mathbf{W}' \mathbf{W} + \boldsymbol{\Sigma}| / \partial g_{ij} &= \text{tr} \{ \mathbf{C} \mathbf{D} [\mathbf{0}, -\mathbf{G}^{-1} \mathbf{E}_{ij} \mathbf{G}^{-1} \otimes \mathbf{A}^{-1} \sigma_{Eo}^2, \mathbf{0}] \} \\
 &= -\text{tr} \{ \mathbf{C}^{aa} (\mathbf{G}^{-1} \mathbf{E}_{ij} \mathbf{C}^{-1} \otimes \mathbf{A}^{-1}) \} \sigma_{Eo}^2
 \end{aligned} \tag{A.13}$$

In a similar fashion

$$\begin{aligned}
 \partial \log |\mathbf{W}' \mathbf{W} + \boldsymbol{\Sigma}| / \partial \sigma_{Em}^2 &= \text{tr} \{ \mathbf{C} \mathbf{D} [\mathbf{0}, \mathbf{0}, -\mathbf{I}_d (\sigma_{Em}^2)^{-2} \sigma_{Eo}^2] \} \\
 &= -\text{tr} (\mathbf{C}^{ee}) (\sigma_{Em}^2)^{-2} \sigma_{Eo}^2
 \end{aligned} \tag{A.14}$$

and

$$\begin{aligned}
 \partial \log |\mathbf{W}' \mathbf{W} + \boldsymbol{\Sigma}| / \partial \sigma_{Eo}^2 &= \text{tr} \{ \mathbf{C} \mathbf{D} [\mathbf{0}, \mathbf{G}^{-1} \otimes \mathbf{A}^{-1}, \mathbf{I}_d (\sigma_{Em}^2)^{-1}] \} \\
 &= \text{tr} [\mathbf{C}^{aa} (\mathbf{G}^{-1} \otimes \mathbf{A}^{-1})] + \text{tr} [\mathbf{C}^{ee}] (\sigma_{Em}^2)^{-1}
 \end{aligned} \tag{A.15}$$

Taking derivatives of [A.13]-[A.15] again we obtain:

$$\begin{aligned}
\partial^2 \log |\mathbf{W}'\mathbf{W} + \Sigma| / (\partial g_{ij})^2 &= -\text{tr}\{(\partial \mathbf{C}^{aa} / \partial g_{ij})(\mathbf{G}^{-1} \mathbf{E}_{ij} \mathbf{G}^{-1} \otimes \mathbf{A}^{-1})\} \sigma_{E_o}^2 \\
&\quad + 2\text{tr}\{\mathbf{C}^{aa}(\mathbf{G}^{-1} \mathbf{E}_{ij} \mathbf{G}^{-1} \mathbf{E}_{ij} \mathbf{G}^{-1} \otimes \mathbf{A}^{-1})\} \sigma_{E_o}^2 \\
&= -\text{tr}\{[-\mathbf{M}'\mathbf{C}\mathbf{D}[0, -(\mathbf{G}^{-1} \mathbf{E}_{ij} \mathbf{G}^{-1} \otimes \mathbf{A}^{-1}) \sigma_{E_o}^2, \mathbf{0}]\mathbf{C}\mathbf{M}\}(\mathbf{G}^{-1} \mathbf{E}_{ij} \mathbf{G}^{-1} \otimes \mathbf{A}^{-1})\} \sigma_{E_o}^2 \\
&\quad + 2\text{tr}\{\mathbf{C}^{aa}(\mathbf{G}^{-1} \mathbf{E}_{ij} \mathbf{G}^{-1} \mathbf{E}_{ij} \mathbf{G}^{-1} \otimes \mathbf{A}^{-1})\} \sigma_{E_o}^2 \\
&= -\text{tr}\{\mathbf{C}^{aa}(\mathbf{G}^{-1} \mathbf{E}_{ij} \mathbf{G}^{-1} \otimes \mathbf{A}^{-1}) \mathbf{C}^{aa}(\mathbf{G}^{-1} \mathbf{E}_{ij} \mathbf{G}^{-1} \otimes \mathbf{A}^{-1})\} (\sigma_{E_o}^2)^2 \\
&\quad + 2\text{tr}\{\mathbf{C}^{aa}(\mathbf{G}^{-1} \mathbf{E}_{ij} \mathbf{G}^{-1} \mathbf{E}_{ij} \mathbf{G}^{-1} \otimes \mathbf{A}^{-1})\} \sigma_{E_o}^2 \\
&= \text{tr}\{\mathbf{C}^{aa}[2(\mathbf{G}^{-1} \mathbf{E}_{ij} \mathbf{G}^{-1} \mathbf{E}_{ij} \mathbf{G}^{-1} \otimes \mathbf{A}^{-1}) - (\mathbf{G}^{-1} \mathbf{E}_{ij} \mathbf{G}^{-1} \otimes \mathbf{A}^{-1}) \\
&\quad \mathbf{C}^{aa}(\mathbf{G}^{-1} \mathbf{E}_{ij} \mathbf{G}^{-1} \otimes \mathbf{A}^{-1}) \sigma_{E_o}^2]\} \sigma_{E_o}^2 \tag{A.16}
\end{aligned}$$

$$\begin{aligned}
\partial^2 \log |\mathbf{W}'\mathbf{W} + \Sigma| / \partial g_{ij} \partial g_{hk} &= -\text{tr}\{(\partial \mathbf{C}^{aa} / \partial g_{hk})(\mathbf{G}^{-1} \mathbf{E}_{ij} \mathbf{G}^{-1} \otimes \mathbf{A}^{-1})\} \sigma_{E_o}^2 \\
&\quad + \text{tr}\{\mathbf{C}^{aa}(\mathbf{G}^{-1} \mathbf{E}_{hk} \mathbf{G}^{-1} \mathbf{E}_{ij} \mathbf{G}^{-1} \otimes \mathbf{A}^{-1})\} \sigma_{E_o}^2 + \\
&\text{tr}\{\mathbf{C}^{aa}(\mathbf{G}^{-1} \mathbf{E}_{ij} \mathbf{G}^{-1} \mathbf{E}_{hk} \mathbf{G}^{-1} \otimes \mathbf{A}^{-1})\} \sigma_{E_o}^2 \\
&= -\text{tr}\{\mathbf{C}^{aa}(\mathbf{G}^{-1} \mathbf{E}_{hk} \mathbf{G}^{-1} \otimes \mathbf{A}^{-1}) \mathbf{C}^{aa}(\mathbf{G}^{-1} \mathbf{E}_{ij} \mathbf{G}^{-1} \otimes \mathbf{A}^{-1})\} (\sigma_{E_o}^2)^2 \\
&+ \text{tr}\{\mathbf{C}^{aa}(\mathbf{G}^{-1} \mathbf{E}_{hk} \mathbf{G}^{-1} \mathbf{E}_{ij} \mathbf{G}^{-1} \otimes \mathbf{A}^{-1})\} \sigma_{E_o}^2 \\
&\quad + \text{tr}\{\mathbf{C}^{aa}(\mathbf{G}^{-1} \mathbf{E}_{ij} \mathbf{G}^{-1} \mathbf{E}_{hk} \mathbf{G}^{-1} \otimes \mathbf{A}^{-1})\} \sigma_{E_o}^2
\end{aligned}$$

$$\begin{aligned}
\partial^2 \log |\mathbf{W}'\mathbf{W} + \Sigma| / (\partial \sigma_{E_m}^2)^2 &= -\text{tr}\{[-\mathbf{N}'\mathbf{C}\mathbf{D}[\mathbf{0}, \mathbf{0}, -\mathbf{I}_d(\sigma_{E_m}^2)^{-2}] \sigma_{E_o}^2 \mathbf{C}\mathbf{N}]\} (\sigma_{E_m}^2)^{-2} \sigma_{E_o}^2 \\
&\quad + 2\text{tr}\{\mathbf{C}^{ee}\} (\sigma_{E_m}^2)^{-3} \sigma_{E_o}^2 \\
&\quad - \text{tr}\{\mathbf{C}^{ee}(2\mathbf{I}_d - \mathbf{C}^{ee}(\sigma_{E_o}^2/\sigma_{E_m}^2))\} (\sigma_{E_m}^2)^{-3} \sigma_{E_o}^2 \tag{A.17}
\end{aligned}$$

$$\begin{aligned}
\partial^2 \log |\mathbf{W}'\mathbf{W} + \Sigma| / (\partial \sigma_{E_o}^2)^2 &= \partial \text{tr}\{\mathbf{C}^{aa}(\mathbf{G}^{-1} \otimes \mathbf{A}^{-1})\} / \partial \sigma_{E_o}^2 + \partial \text{tr}\{\mathbf{C}^{ee}\} (\sigma_{E_m}^2)^{-1} / \partial \sigma_{E_o}^2 \\
&= -\{\text{tr}\{\mathbf{C}^{aa}(\mathbf{G}^{-1} \otimes \mathbf{A}^{-1}) \mathbf{C}^{aa}(\mathbf{G}^{-1} \otimes \mathbf{A}^{-1})\} + 2\text{tr}\{\mathbf{C}^{ea}(\mathbf{G}^{-1} \otimes \mathbf{A}^{-1}) \mathbf{C}^{ae}\} (\sigma_{E_m}^2)^{-1} \\
&\quad + \text{tr}\{\mathbf{C}^{ee} \mathbf{C}^{ee}\} (\sigma_{E_m}^2)^{-2}\} \tag{A.18}
\end{aligned}$$

$$\begin{aligned}
\partial^2 \log |\mathbf{W}'\mathbf{W} + \Sigma| / \partial g_{ij} \partial \sigma_{E_m}^2 &= -\text{tr}\{(\partial \mathbf{C}^{aa} / \partial \sigma_{E_m}^2)(\mathbf{G}^{-1} \mathbf{E}_{ij} \mathbf{G}^{-1} \otimes \mathbf{A}^{-1})\} \sigma_{E_o}^2 \\
&= -\text{tr}\{\mathbf{C}^{ea}(\mathbf{G}^{-1} \mathbf{E}_{ij} \mathbf{G}^{-1} \otimes \mathbf{A}^{-1}) \mathbf{C}^{ae}\} (\sigma_{E_o}^2 / \sigma_{E_m}^2)^2 \tag{A.19}
\end{aligned}$$

$$\begin{aligned}
\partial^2 \log |\mathbf{W}'\mathbf{W} + \Sigma| / \partial g_{ij} \partial \sigma_{E_o}^2 &= -\{\text{tr}\{\mathbf{C}^{aa}(\mathbf{G}^{-1} \otimes \mathbf{A}^{-1}) \mathbf{C}^{aa}(\mathbf{G}^{-1} \mathbf{E}_{ij} \mathbf{G}^{-1} \otimes \mathbf{A}^{-1})\} \sigma_{E_o}^2 \\
&+ \text{tr}\{\mathbf{C}^{ea}(\mathbf{G}^{-1} \mathbf{E}_{ij} \mathbf{G}^{-1} \otimes \mathbf{A}^{-1}) \mathbf{C}^{ae}\} (\sigma_{E_o}^2 / \sigma_{E_m}^2) + \text{tr}\{\mathbf{C}^{aa}(\mathbf{G}^{-1} \mathbf{E}_{ij} \mathbf{G}^{-1} \otimes \mathbf{A}^{-1})\} \} \tag{A.20}
\end{aligned}$$

$$\begin{aligned}
\partial^2 \log |\mathbf{W}'\mathbf{W} + \Sigma| / \partial \sigma_{E_m}^2 \partial \sigma_{E_o}^2 &= -(\sigma_{E_m}^2)^{-2} \{\text{tr}\{\mathbf{C}^{ea}(\mathbf{G}^{-1} \otimes \mathbf{A}^{-1}) \mathbf{C}^{ae}\} \sigma_{E_o}^2 \\
&+ \text{tr}\{\mathbf{C}^{ee} \mathbf{C}^{ee}\} \sigma_{E_o}^2 (\sigma_{E_m}^2)^{-1} + \text{tr}\{\mathbf{C}^{ee}\} \} \tag{A.21}
\end{aligned}$$

### Other derivatives

We now consider the remaining derivatives and these are:

$$\partial[n_g + a + 3] \log |\mathbf{G}| + \text{tr}(\mathbf{G}^{-1} \mathbf{G}^*) / \partial g_{ij} = (n_g + a + 3) \mathbf{e}'_i \mathbf{G}^{-1} \mathbf{e}_j - \text{tr}(\mathbf{G}^{-1} \mathbf{G}^* \mathbf{G}^{-1} \mathbf{E}_{ij}) \quad [\text{A.22}]$$

with [A.12] used to obtain the second term on the right of [A.22].

Likewise

$$\begin{aligned} \partial\{(n_m s_{Em}^2 / \sigma_{Em}^2) + (n_m + d + 2) \log[\sigma_{Em}^2]\} / \partial \sigma_{Em}^2 = & -[n_m s_{Em}^2 / (\sigma_{Em}^2)^2] \\ & + [(n_m + d + 2) / \sigma_{Em}^2] \end{aligned} \quad [\text{A.23}]$$

$$\begin{aligned} \partial\{(n_o s_{Eo}^2 / \sigma_{Eo}^2) + (n_o + n - p - 2a - d + 2) \log[\sigma_{Eo}^2]\} / \partial \sigma_{Eo}^2 = & \\ -n_o s_{Eo}^2 / (\sigma_{Eo}^2)^2 + [(n_o + n - p - 2a - d + 2) / \sigma_{Eo}^2] & \end{aligned} \quad [\text{A.24}]$$

Second derivatives obtained from [A.22]-[A.24] are:

$$\begin{aligned} \partial^2[(n_g + a + 3) \log |\mathbf{G}| + \text{tr}(\mathbf{G}^{-1} \mathbf{G}^*)] / \partial g_{ij} \partial g_{kh} = & -(n_g + a + 3) \mathbf{e}'_k \mathbf{G}^{-1} \mathbf{E}_{ij} \mathbf{G}^{-1} \mathbf{e}_h \\ & + \mathbf{e}'_i \mathbf{G}^{-1} \mathbf{e}_h \mathbf{e}'_k \mathbf{G}^{-1} \mathbf{G}^* \mathbf{G}^{-1} \mathbf{e}_j + \mathbf{e}'_i \mathbf{G}^{-1} \mathbf{G}^* \mathbf{G}^{-1} \mathbf{e}_h \mathbf{e}'_k \mathbf{G}^{-1} \mathbf{e}_j \end{aligned} \quad [\text{A.25}]$$

$$\begin{aligned} \partial^2\{(n_m s_{Em}^2 / \sigma_{Em}^2) + (n_m + d + 2) \log[\sigma_{Em}^2]\} / (\partial \sigma_{Em}^2)^2 = & \\ [2n_m s_{Em}^2 / (\sigma_{Em}^2)^3] - [(n_m + d + 2) / (\sigma_{Em}^2)^2] & \end{aligned} \quad [\text{A.26}]$$

$$\begin{aligned} \partial^2\{(n_o s_{Eo}^2 / \sigma_{Eo}^2) + (n_o + n - p - 2a - d + 2) \log[\sigma_{Eo}^2]\} / (\partial \sigma_{Eo}^2)^2 = & \\ 2n_o s_{Eo}^2 / (\sigma_{Eo}^2)^3 - [(n_o + n - p - 2a - d + 2) / (\sigma_{Eo}^2)^2] & \end{aligned} \quad [\text{A.27}]$$

### First derivatives of the log-posterior

Using [A.3], [A.13] and [A.22] we have:

$$\begin{aligned} \partial \log[p(\Gamma | y)] / \partial g_{ij} = 1/2\{\hat{\mathbf{a}}'(\mathbf{G}^{-1} \mathbf{E}_{ij} \mathbf{G}^{-1} \otimes \mathbf{A}^{-1}) \hat{\mathbf{a}} + \text{tr}[\mathbf{C}^{aa}(\mathbf{G}^{-1} \mathbf{E}_{ij} \mathbf{G}^{-1} \otimes \mathbf{A}^{-1})]\} \sigma_{Eo}^2 \\ - (n_g + a + 3) \mathbf{e}'_i \mathbf{G}^{-1} \mathbf{e}_j + \text{tr}(\mathbf{G}^{-1} \mathbf{C}^* \mathbf{G}^{-1} \mathbf{E}_{ij}) \end{aligned} \quad [\text{A.28}]$$

Using [A.4], [A.14] and [A.23]:

$$\partial \log[p(\boldsymbol{\Gamma} | \mathbf{y})] / \partial \sigma_{E_m}^2 = 1/2 \{ (\sigma_{E_m}^2)^{-2} [\hat{\mathbf{e}}_m' \hat{\mathbf{e}}_m + \text{tr}(\mathbf{C}^{ee}) \sigma_{E_o}^2 + n_m s_{E_m}^2] - [(n_m + d + 2) / \sigma_{E_m}^2] \} \quad [\text{A.29}]$$

Using [A.5], [A.15] and [A.24]:

$$\begin{aligned} \partial \log[p(\boldsymbol{\Gamma} | \mathbf{y})] / \partial \sigma_{E_o}^2 = & 1/2 \{ [(y' \mathbf{y} - \hat{\boldsymbol{\theta}}' \mathbf{W}' \mathbf{y} + n_o s_{E_o}^2) / (\sigma_{E_o}^2)^2] \\ & - [\hat{\mathbf{a}}' (\mathbf{G}^{-1} \otimes \mathbf{A}^{-1}) \hat{\mathbf{a}} + \hat{\mathbf{e}}_m' \hat{\mathbf{e}}_m (\sigma_{E_m}^2)^{-1}] (\sigma_{E_o}^2)^{-1} \\ & - \text{tr}[\mathbf{C}^{aa} (\mathbf{G}^{-1} \otimes \mathbf{A}^{-1})] - \text{tr}[\mathbf{C}^{ee}] (\sigma_{E_m}^2)^{-1} \\ & - [(n_o + n - p - 2a - d + 2) / \sigma_{E_o}^2] \} \end{aligned} \quad [\text{A.30}]$$

### Second derivatives of the log-posterior

Using [A.6], [A.16] and [A.25]:

$$\begin{aligned} \partial^2 \log[p(\boldsymbol{\Gamma} | \mathbf{y})] / \partial g_{ij} \partial g_{kh} = & \hat{\mathbf{a}}' [\mathbf{G}^{-1} \mathbf{E}_{ij} \mathbf{G}^{-1} \otimes \mathbf{A}^{-1}] (\mathbf{C}^{aa} (\mathbf{G}^{-1} \mathbf{E}_{kh} \mathbf{G}^{-1} \\ & \otimes \mathbf{A}^{-1}) \sigma_{E_o}^2 - (\mathbf{G}^{-1} \mathbf{E}_{ij} \mathbf{G}^{-1} \mathbf{E}_{kh} \mathbf{G}^{-1} \otimes \mathbf{A}^{-1}) \hat{\mathbf{a}} \\ & + 1/2 \text{tr} \{ \mathbf{C}^{aa} [2(\mathbf{G}^{-1} \mathbf{E}_{ij} \mathbf{G}^{-1} \mathbf{E}_{kh} \mathbf{G}^{-1} \otimes \mathbf{A}^{-1}) - (\mathbf{G}^{-1} \mathbf{E}_{ij} \mathbf{G}^{-1} \\ & \otimes \mathbf{A}^{-1}) \mathbf{C}^{aa} (\mathbf{G}^{-1} \mathbf{E}_{kh} \mathbf{G}^{-1} \otimes \mathbf{A}^{-1}) \sigma_{E_o}^2] \} \sigma_{E_o}^2 \\ & + 1/2 (n_g + a + 3) \mathbf{e}'_k \mathbf{G}^{-1} \mathbf{E}_{ij} \mathbf{G}^{-1} \mathbf{e}_h - 1/2 [\mathbf{e}'_i \mathbf{G}^{-1} \mathbf{e}_h \mathbf{e}'_k \mathbf{G}^{-1} \mathbf{G}^* \mathbf{G}^{-1} \mathbf{e}_j \\ & + \mathbf{e}'_i \mathbf{G}^{-1} \mathbf{G}^* \mathbf{G}^{-1} \mathbf{e}_h \mathbf{e}'_k \mathbf{G}^{-1} \mathbf{e}_j] \end{aligned} \quad [\text{A.31}]$$

Using [A.17], [A.17] and [A.26]:

$$\begin{aligned} \partial^2 \log[p(\boldsymbol{\Gamma} | \mathbf{y})] / (\partial \sigma_{E_m}^2)^2 = & (\sigma_{E_m}^2)^{-3} \hat{\mathbf{e}}_m' (\mathbf{C}^{ee} (\sigma_{E_m}^2)^{-1} \sigma_{E_o}^2 - \mathbf{I}_d) \hat{\mathbf{e}}_m \\ & + 1/2 \text{tr} [\mathbf{C}^{ee} (2\mathbf{I}_d - \mathbf{C}^{ee} (\sigma_{E_o}^2 / \sigma_{E_m}^2))] (\sigma_{E_m}^2)^{-3} \sigma_{E_o}^2 - n_m s_{E_m}^2 / (\sigma_{E_m}^2)^3 \\ & + 1/2 (n_m + d + 2) / (\sigma_{E_m}^2)^2 \end{aligned} \quad [\text{A.32}]$$

Using [A.8], [A.18] and [A.27]:

$$\begin{aligned} \partial^2 \log[p(\boldsymbol{\Gamma} | \mathbf{y})] / (\partial \sigma_{E_o}^2)^2 = & -(y' \mathbf{y} - \hat{\boldsymbol{\theta}}' \mathbf{W}' \mathbf{y} + n_o s_{E_o}^2) / (\sigma_{E_o}^2)^3 \\ & + (1/2 \sigma_{E_o}^2) [\hat{\mathbf{a}}' (\mathbf{G}^{-1} \otimes \mathbf{A}^{-1}) \mathbf{C}^{aa} (\mathbf{G}^{-1} \otimes \mathbf{A}^{-1}) \hat{\mathbf{a}} \\ & + 2 \hat{\mathbf{a}}' (\mathbf{G}^{-1} \otimes \mathbf{A}^{-1}) \mathbf{C}^{ae} \hat{\mathbf{e}}_m (\sigma_{E_m}^2)^{-1} + \hat{\mathbf{e}}_m' \mathbf{C}^{ee} \hat{\mathbf{e}}_m (\sigma_{E_m}^2)^{-2}] \\ & + 1/2 \{ \text{tr} [\mathbf{C}^{aa} (\mathbf{G}^{-1} \otimes \mathbf{A}^{-1}) \mathbf{C}^{aa} (\mathbf{G}^{-1} \otimes \mathbf{A}^{-1})] \\ & + 2 \text{tr} [\mathbf{C}^{ea} (\mathbf{G}^{-1} \otimes \mathbf{A}^{-1}) \mathbf{C}^{ae}] (\sigma_{E_m}^2)^{-1} + \text{tr} [\mathbf{C}^{ee} \mathbf{C}^{ee}] (\sigma_{E_m}^2)^{-2} \} \\ & + (n_o + n - p - 2a - d + 2) / 2 (\sigma_{E_o}^2)^2 \end{aligned} \quad [\text{A.33}]$$

Using [A.9] and [A.19]:

$$\begin{aligned} \partial^2 \log[p(\boldsymbol{\Gamma} | \boldsymbol{y})] / \partial g_{ij} \partial \sigma_{E_m}^2 &= \{\hat{\mathbf{a}}'(\mathbf{G}^{-1} \mathbf{E}_{ij} \mathbf{G}^{-1} \otimes \mathbf{A}^{-1}) \mathbf{C}^{ae} \hat{\mathbf{e}}_m \\ &\quad + \text{tr}[\mathbf{C}^{ea}(\mathbf{G}^{-1} \otimes \mathbf{A}^{-1}) \mathbf{C}^{ae}]\} (\sigma_{E_m}^2)^{-2} \end{aligned} \quad [\text{A.34}]$$

Using [A.10] and [A.20]:

$$\begin{aligned} \partial^2 \log[p(\boldsymbol{\Gamma} | \boldsymbol{y})] / \partial g_{ij} \partial \sigma_{E_o}^2 &= 1/2 \{ \hat{\mathbf{a}}' \{ \mathbf{G}^{-1} \mathbf{E}_{ij} \mathbf{G}^{-1} \otimes \mathbf{A}^{-1} [\mathbf{I}_{2a} \\ &\quad - 2\mathbf{C}^{aa}(\mathbf{G}^{-1} \otimes \mathbf{A}^{-1}) \sigma_{E_o}^2] \} \hat{\mathbf{a}} (\sigma_{E_o}^2)^{-1} \\ &\quad - 2\hat{\mathbf{a}}'(\mathbf{G}^{-1} \mathbf{E}_{ij} \mathbf{G}^{-1} \otimes \mathbf{A}^{-1}) \mathbf{C}^{ae} \hat{\mathbf{e}}_m (\sigma_{E_m}^2)^{-1} \} \\ &\quad + 1/2 \{ \text{tr}[\mathbf{C}^{aa}(\mathbf{G}^{-1} \mathbf{E}_{ij} \mathbf{G}^{-1} \otimes \mathbf{A}^{-1}) \mathbf{C}^{aa}(\mathbf{G}^{-1} \mathbf{E}_{ij} \mathbf{G}^{-1} \otimes \mathbf{A}^{-1})] \sigma_{E_o}^2 \\ &\quad + \text{tr}[\mathbf{C}^{ea}(\mathbf{G}^{-1} \mathbf{E}^{ij} \mathbf{G}^{-1} \otimes \mathbf{A}^{-1}) \mathbf{C}^{ae}] (\sigma_{E_o}^2 / \sigma_{E_m}^2) \\ &\quad + \text{tr}(\mathbf{C}^{aa}(\mathbf{G}^{-1} \mathbf{E}_{ij} \mathbf{G}^{-1} \otimes \mathbf{A}^{-1})) \} \end{aligned} \quad [\text{A.35}]$$

Finally, using [A.11] and [A.21]:

$$\begin{aligned} \partial^2 \log[p(\boldsymbol{\Gamma} | \boldsymbol{y})] / \partial \sigma_{E_m}^2 \partial \sigma_{E_o}^2 &= 1/2 \hat{\mathbf{e}}_m' [\mathbf{I}_d - 2\mathbf{C}^{ee} \sigma_{E_o}^2] \hat{\mathbf{e}}_m (\sigma_{E_m}^2)^{-2} (\sigma_{E_o}^2)^{-1} \\ &\quad - \hat{\mathbf{e}}_m' \mathbf{C}^{ea} (\mathbf{G}^{-1} \otimes \mathbf{A}^{-1}) \hat{\mathbf{a}} \\ &\quad + 1/2 (\sigma_{E_m}^2)^{-2} \{ \text{tr}(\mathbf{C}^{ea}(\mathbf{G}^{-1} \otimes \mathbf{A}^{-1}) (\mathbf{C}^{ae}) \sigma_{E_o}^2 + \text{tr}(\mathbf{C}^{ee} \mathbf{C}^{ee}) \sigma_{E_o}^2 + \text{tr}(\mathbf{C}^{ee}) \} \end{aligned} \quad [\text{A.36}]$$