BIOMETRICS 49, 677–689
September 1993

REML Estimation with Exact Covariance in the Logistic Mixed Model

Melinda L. Drum

School of Public Health, University of Illinois at Chicago, 2121 West Taylor, Chicago, Illinois 60612, U.S.A.

and

Peter McCullagh

Department of Statistics, University of Chicago, 5734 University Avenue, Chicago, Illinois 60637, U.S.A.

SUMMARY

Residual maximum likelihood (REML) estimation is adapted to certain logistic mixed models for which representation of the unconditional mean as a linear function of the fixed effects is possible. Only the first two moments of the unconditional distribution need be evaluated, and except for the form of the covariance, the maximization algorithm carries over directly from linear models. The exact unconditional covariance is computed from the logistic–normal mixture for input into the algorithm. Residual log-likelihood plots provide a means of inference for the dispersion components. The Taylor series approximation to this covariance, besides exhibiting insufficient accuracy, fails to be positive definite for large values of the dispersion components. As a consequence, REML log-likelihood plots based on this approximate covariance attribute misleadingly high precision to the dispersion component estimates. The method is presented in the context of a salamander mating experiment in which random effects corresponding to male and female animals occur in a crossed design. Analyses of these data by several methods are compared.

1. Introduction

Logistic and probit regression models may be extended to accommodate correlated observations by adding random effects to the linear predictor. Estimation in these mixed models is a challenging problem that has generated a number of innovative approaches for dealing with the marginal likelihood, which is a complicated multidimensional integral that must be evaluated numerically. This paper describes adaptation of residual maximum likelihood (REML) estimation to certain logistic mixed models for which representation of the unconditional mean as a linear function of the fixed effects is possible. Only the first two moments of the unconditional distribution need be evaluated, and except for the form of the covariance, the maximization algorithm carries over directly from linear models. The method accommodates crossed, as well as nested, designs for the random effects.

REML estimation using a Taylor series approximation for the unconditional covariance has been described previously (Drum, Technical Report No. 285, Department of Statistics, The University of Chicago, 1990). The accuracy of the approximation, however, is questionable. Moreover, this approximate covariance fails to be positive definite for large values of the dispersion components. (An upper bound on individual components, σ_r^2, beyond which positive definiteness fails, is $\min\{1/[\pi_i(1-\pi_i)]\}$ where π is the unconditional mean of the observations.) As a consequence, REML log-likelihood plots used to evaluate and compare estimates, when based on this approximate covariance, are considerably steeper than the corresponding plots based on the exact covariance and give a misleading indication of the precision of the dispersion component estimates. Therefore, the

Key words: Components of dispersion; Correlated binary data; Logistic mixed model; Quasi-likelihood; Random effects; REML estimation.

exact unconditional covariance is calculated from the logistic–normal mixture defined by the mixed model and used as input to the algorithm.

The method is presented in the context of an experiment on the mating habits of Appalachian salamanders, which is described below.

2. Residual Maximum Likelihood Estimation in the Logistic Mixed Model

2.1 *The Salamander Mating Experiment*

In the summer and fall of 1986, S. Arnold and P. Verrell of the Department of Ecology and Evolution at the University of Chicago conducted experiments to investigate whether inhibitions against cross-breeding exist in two populations of Appalachian salamanders, Rough Butt and White Side, whose natural habitats do not overlap. Ten males and ten females from each population (for a total of 40 animals) were paired in a crossed design. Each animal was sequestered overnight on six different occasions with an animal of the opposite sex, alternately from the same and from the opposite population. Thus four types of crosses were observed: RR, RW, WR, and WW with the letters indicating female and male populations, respectively. The occurrence of mating was recorded as the binary response. Three replications of the experiment were conducted, one in the summer of 1986 and two in the fall of the same year. The same animals were used in the summer and first fall experiments, but since this is atypical, the data are treated as if fresh animals had been used each time. The proposed method could be adapted to accommodate correlation between observations on the same animals in different seasons, but we have not done this here. The data are given in Tables 14.4 through 14.6 in McCullagh and Nelder (1989).

In the analysis, the cross effect is chosen as most relevant to the purpose of the experiment, with particular interest in the contrast between the mixed crosses, RW and WR. The effects of the individual salamanders are regarded as random, with female and male effects modeled separately.

2.2 *The Model*

Conditioning on the random effects, observations are assumed to be independent and to follow a logistic model,

$$\operatorname{logit} \mathrm{E}(Y \mid b) = X\alpha + Z\beta, \tag{1}$$

where X and Z are design matrices for fixed and random effects respectively, α is the parameter vector of fixed effects, and b is the vector of unobservable random effects. For a design with c random factors, the components of the random term partition as $Z = [Z_1, \ldots, Z_c]$ and $b^{\mathrm{T}} = (b_1^{\mathrm{T}}, \ldots, b_c^{\mathrm{T}})$, where Z_r and b_r are the $n \times q_r$ incidence matrix and the $q_r \times 1$ vector of random effects of the rth random factor with q_r levels. The random effects are assumed to have zero mean and covariance $D = \operatorname{diag}(\sigma_1^2 I_{q_1}, \ldots, \sigma_c^2 I_{q_c})$. For the Taylor series approximation to the unconditional covariance, these first two moments of b suffice, but to obtain the exact covariance the distribution of b must be specified. We shall assume $b \sim N(0, D)$. For the salamander design, $\operatorname{logit} \mathrm{E}(Y \mid b) = X\alpha + Z_{\mathrm{f}} b_{\mathrm{f}} + Z_{\mathrm{m}} b_{\mathrm{m}}$, $D = \operatorname{diag}(\sigma_{\mathrm{f}}^2 I_{20}, \sigma_{\mathrm{m}}^2 I_{20})$, and $\alpha = (\alpha_{\mathrm{RR}}, \alpha_{\mathrm{RW}}, \alpha_{\mathrm{WR}}, \alpha_{\mathrm{WW}})$, where "f" and "m" denote female and male, respectively.

When X is the incidence matrix for a single factor or, equivalently, for the interaction of all fixed factors in the model, the unconditional mean, which is $\mathrm{E}[\mathrm{E}(Y \mid b)] = \mathrm{E}[F(X\alpha + Zb)]$, where F is the logistic cumulative distribution function (cdf), can be written $\mathrm{E}(Y) = X\pi$. This linearity permits application of methods for linear mixed models, adapted to allow functional dependence of the variance on the mean. In order to make this adaptation, it is essential for the marginal model to be linear.

2.3 *Estimation of Fixed Effects*

With linearity on the identity scale, the quasi-likelihood estimating equations for dependent observations (McCullagh and Nelder, 1989, Chap. 9) take the same form as normal theory score equations, and the quasi-likelihood estimates of fixed effects take the same form as Gauss–Markov estimates, with the exception that the appropriate covariance is functionally dependent on the mean.

With a certain degree of balance, estimation of fixed effects is independent of the variance. That is, the least squares estimator $P_I Y$ is the same as the Gauss–Markov estimator $P_W Y$, where P_I and P_W are the projection matrices $X(X^{\mathrm{T}}X)^{-1}X^{\mathrm{T}}$ and $X(X^{\mathrm{T}}WX)^{-1}X^{\mathrm{T}}W$, and $W^{-1} = V = \operatorname{cov}(Y)$. A sufficient balance criterion is that for each x in the column space $\mathscr{X}$ of X, $Vx \in \mathscr{X}$ (Kruskal, 1968). In matrix terminology, this is equivalent to $VX = XB$ for some $p \times p$ matrix B. Note that the estimate $P_I Y$ is simply the vector of averages, or observed proportions, for the levels of the fixed effect. The salamander data, as modeled, meet this balance criterion so the estimates of fixed effects are simply the sample proportions of successful matings for each cross type.

2.4 *Estimation of Dispersion Components*

In the first analysis of these data, McCullagh and Nelder (1989, Chap. 14) employed the moment method of equating selected quadratic forms to their expectation, using the Taylor series approximation to the unconditional covariance.

REML estimates of dispersion components, θ, are obtained by maximizing the log-likelihood for θ based on residuals, or error contrasts,

$$l_R(\theta; y) = -\tfrac{1}{2}\log \det(V) - \tfrac{1}{2}\log \det(X^T W X) - \tfrac{1}{2} R^T W(I - P_{W'})R, \tag{2}$$

where, as above, $V = \text{cov}(Y)$, $W = V^{-1}$, $P_{W'}$ is the weighted projection matrix $X(X^T WX)^{-1}X^T W$, and R is the vector of residuals from an arbitrary projection P onto the column space of X. We shall use ordinary least squares residuals.

This log-likelihood may be obtained as the singular normal likelihood of the residuals R. If $Y \sim N(X\alpha, V)$, then

$$R = Y - X\hat{\alpha} \sim N(0, (I - P)V(I - P)^T). \tag{3}$$

Since the variance $(I - P)V(I - P)^T$ of R is not of full rank, the extra term $-\frac{1}{2}\log\det(X^T V^{-1} X)$ enters the log-likelihood. The residual log-likelihood differs by this term from the profile log-likelihood that yields the maximum likelihood estimate of θ. The quantity $W(I - P_{W'})$ in the last term of (2) is the annihilating generalized inverse of $(I - P)V(I - P)^T$. Residual maximum likelihood estimation adjusts for degrees of freedom lost in estimating α. For the ordinary linear model, $s^2 = \text{RSS}/(n - p)$ is an REML estimate.

Adaptation of this method is possible because the quasi-likelihood estimate of the mean is a projection onto the column space of X, so that the residuals are error contrasts with the mean–variance structure given in (3) that leads to the log-likelihood (2). Although the residual log-likelihood is a normal theory likelihood, the REML estimating equations can be derived without the assumption of normality.

For linear models, the residual log-likelihood is free of fixed effect parameters so that a single LS/REML/WLS cycle completes the estimation process. For binary data, however, because fixed effect parameters enter the residual log-likelihood through functional dependence of the variance on the mean, iteration between the REML and WLS steps is required. For the remainder of this paper, only the balanced case in which the least squares and Gauss–Markov estimates are equal so that this level of iteration is avoided, will be considered. The unbalanced case requires the following adjustments. In the score and information given by (5) or (10), P_I is replaced by $P_{W'}$, and the last term of the score takes form $R^T W(I - P_{W'})(\partial V/\partial\theta_r)(1 - P_{W'})WR$, which reduces to $R^T W(\partial V/\partial\theta_r)WR$ when weighted least squares residuals are used. Additionally, estimation of dispersion components is conditional on current estimates of fixed effects, which are regarded as constant within the REML step.

The maximization algorithm used in the REML step is Newton–Raphson iteration with Fisher scoring.

3. REML Estimation with the Approximate Covariance

The first-order Taylor series approximation to the unconditional covariance takes form

$$V_B(\pi) + \sum_{r=1}^{c} \sigma_r^2 V_B(\pi)(Z_r Z_r^T - I)V_B(\pi) = V_B(\pi) + \sum_{r=1}^{c} \sigma_r^2 V_r(\pi), \tag{4}$$

where $V_B(\pi)$ is the Bernoulli variance (McCullagh and Nelder, 1989, pp. 446–447).

This approximate covariance serves as an advantageous starting point, to some extent due to ease of computation, but primarily because of its form, which is linear in the dispersion components. This form facilitates the adaptations of methods from linear mixed models that were described previously. In the moment method of McCullagh and Nelder (1989), which consists of equating selected quadratic forms to their expectations, this form of the variance is crucial. Because the variance is linear in the dispersion components, so too are expectations of the quadratic forms. Thus, once the necessary expectations have been derived algebraically, estimates of dispersion components, as in linear models, can be obtained as solutions to a system of linear equations.

This linear form also makes possible a straightforward adaptation of the REML algorithm. The derivatives of the variance with respect to the dispersion components are the matrices $V_r(\pi)$, which are functionally independent of the σ_r^2. Thus, derivatives in the score and Fisher information remain constant from one Newton–Raphson iteration to the next. Looking at the score and information using the Taylor series approximation for the variance,

$$
\begin{aligned}
s_r &= -\tfrac{1}{2}\mathrm{tr}[W(I - P_I)V_r(\pi)] + \tfrac{1}{2}R^{\mathrm{T}}WV_r(\pi)WR, \\
I_{rs} &= -\tfrac{1}{2}\mathrm{tr}[W(I - P_I)V_r(\pi)W(I - P_I)V_s(\pi)],
\end{aligned} \tag{5}
$$

we see that the computational effort at each step consists of recalculating and inverting the $n \times n$ covariance matrix and multiplying this inverse by constant matrices or vectors. Recalculation of the variance, $V = V_{\mathrm{B}}(\pi) + \sum_{r=1}^{c} \sigma_r^2 V_r(\pi)$, is simple scalar multiplication and addition, since only the σ_r^2 change from iteration to iteration. The important computational difference between the REML algorithm for linear models and for the logistic mixed model with Taylor covariance lies in the form of the matrix coefficients $V_r(\pi)$ of the σ_r^2: $V_r(\pi) = V_{\mathrm{B}}(\pi)(Z_rZ_r^{\mathrm{T}} - I)V_{\mathrm{B}}(\pi)$ versus $V_r = (Z_rZ_r^{\mathrm{T}})$. Because the $V_r(\pi)$ are not semidefinite and hence do not factor, the dimension-reducing W-transform of Hemmerle and Hartley (1973) is not applicable. An inferential difference arises from the functional dependence of the covariance on the mean in the logistic mixed model. This characteristic results in estimates of dispersion components that are conditional on estimated values of the fixed effects.

Although the Taylor series approximation provides a variance with the same additive form as the variance for linear models with random effects, it has the serious flaw of failing to be positive definite for values of the dispersion components that are sufficiently large. There is also some question as to its accuracy. It turns out, in fact, that for these data, the difference between the exact covariance and this approximation is considerable for values of σ_{f}^2 and σ_{m}^2 in the range of the estimates. The next step, then, is to incorporate the exact covariance into the REML algorithm.

4. REML Estimation with the Exact Covariance

The changes in the REML algorithm due to use of the exact covariance lie entirely with computation of the covariance and its derivatives. The exact covariance can be written as

$$
V_{\mathrm{B}}(\pi) + V_c(\pi, \sigma^2), \tag{6}
$$

where $V_c(\pi, \sigma^2)$ is zero on the main diagonal and contains covariances off-diagonal. If each pair of observations has in common a level of at most one random factor, then

$$
V(\pi, \sigma^2) = V_{\mathrm{B}}(\pi) + \sum_{r=1}^{c} V_r(\pi, \sigma^2), \tag{7}
$$

where $V_r(\pi, \sigma^2)$ has as elements covariances of distinct observations at the same level of random factor, r. Notice that each V_r is a function of the entire vector of dispersion components, even though the relevant pairs of observations are correlated only through the rth random effect. For the salamander example,

$$
V = V_{\mathrm{B}}(\pi) + V_{\mathrm{f}}(\pi, \sigma_{\mathrm{f}}^2, \sigma_{\mathrm{m}}^2) + V_{\mathrm{m}}(\pi, \sigma_{\mathrm{f}}^2, \sigma_{\mathrm{m}}^2). \tag{8}
$$

In general, if the design is such that distinct observations might share the same level of more than one random factor, then

$$
V_c = \sum_{r=1}^{c} V_r(\pi, \sigma^2) + \sum_{r=1}^{c} \sum_{s=r}^{c} V_{r,s}(\pi, \sigma^2) + \cdots + V_{1,\ldots,c}(\pi, \sigma^2), \tag{9}
$$

where the subscripts indicate that the elements are covariances of pairs of distinct observations with levels of exactly those random factors in common. In the score and information,

$$
\begin{aligned}
s_r &= -\frac{1}{2}\,\mathrm{tr}\left[W(I - P_I)\frac{\partial V(\pi, \sigma^2)}{\partial \sigma_r^2}\right] + \frac{1}{2}\,R^{\mathrm{T}}W\frac{\partial V(\pi, \sigma^2)}{\partial \sigma_r^2}\,WR, \\
I_{rs} &= -\frac{1}{2}\,\mathrm{tr}\left[W(I - P_I)\frac{\partial V(\pi, \sigma^2)}{\partial \sigma_r^2}\,W(I - P_I)\frac{\partial V(\pi, \sigma_r^2)}{\partial \sigma_s^2}\right],
\end{aligned} \tag{10}
$$

the derivatives are no longer constant with respect to the dispersion components and must be recalculated at each iteration. Each unique element of V and $\partial V/\partial \sigma_r^2$ also must be calculated individually. There are, of course, a limited number of distinct elements of each term in V_c. These correspond to the distinct ordered pairs of levels of the fixed factor that can occur at the same levels of the relevant random factors. In the salamander model, V_{f} and V_{m} each contain six distinct elements. Both contain elements corresponding to pairs of observations with the same cross type; in addition, V_{f} contains elements corresponding to RR/RW and WR/WW pairs, and V_{m} to RR/WR and RW/WW pairs of observations.

Since, in the proposed model, $\hat{\pi}$ and hence $V_B(\hat{\pi})$ are functionally independent of σ^2, the term $V_B(\hat{\pi})$ of $V(\hat{\pi}, \sigma^2)$ is constant over iterations, and

$$\frac{\partial V(\hat{\pi}, \sigma^2)}{\partial \sigma_s^2} = \sum_{r=1}^{c} \frac{\partial V_r(\hat{\pi}, \sigma^2)}{\partial \sigma_s^2} + \cdots + \frac{\partial V_{1,\ldots,c}(\hat{\pi}, \sigma^2)}{\partial \sigma_s^2}, \tag{11}$$

or, for the salamander model,

$$\frac{\partial V(\hat{\pi}, \sigma^2)}{\partial \sigma_f^2} = \frac{\partial V_f(\hat{\pi}, \sigma^2)}{\partial \sigma_f^2} + \frac{\partial V_m(\hat{\pi}, \sigma^2)}{\partial \sigma_f^2}, \quad \frac{\partial V(\hat{\pi}, \sigma^2)}{\partial \sigma_m^2} = \frac{\partial V_f(\hat{\pi}, \sigma^2)}{\partial \sigma_m^2} + \frac{\partial V_m(\hat{\pi}, \sigma^2)}{\partial \sigma_m^2}. \tag{12}$$

Details concerning calculation of the covariance and its derivatives are given in the Appendix. The Newton–Raphson algorithm is programmed in S with calls to FORTRAN to compute the elements of the exact covariance and its derivatives. Matrix operations are carried out explicitly, which involves inversion and multiplication of $n \times n$ matrices. Known dimension-reducing techniques for linear models, such as the Hemmerle–Hartley W transform, again do not apply since they are based on an additive form, $\sum \sigma_r^2 V_r$, with known, positive definite V_r, for the covariance.

Adaptation of the algorithm to accommodate correlation between random effects (in the salamander case, between observations on the same animals in different seasons) would require estimation of covariance as well as variance components at the REML step for input into calculation of the exact covariance.

5. Analysis of the Salamander Data

The focus here is on comparison of estimators of dispersion components and fixed effects from methods that have been applied to the salamander mating data, and thus that accommodate crossed designs for random effects. These methods include a Bayesian analysis implemented by Gibbs sampling (Karim and Zeger, 1992) and penalized quasi-likelihood analyses by Schall (1991) and Breslow and Clayton (1993). Inferences about the salamander experiments based on estimates produced by these methods are qualitatively the same as those of McCullagh and Nelder (1989) in the original analysis. The emphasis here will be on differences among estimates and on how these differences relate to characteristics of the methods. Primary interest will be in dispersion components.

5.1 *Methods of Estimation*

Table 1 briefly summarizes the methods under consideration. The basic model is the logistic mixed model with random effects added to the linear predictor. Use of the exact covariance requires the additional assumption of a distribution for the random effects, typically Gaussian.

A distinguishing feature of the methods presented in the previous sections is the transformation from the logistic mixed model to a marginal model that is essentially linear. Methods applied to this model will be referred to as marginal methods in the ensuing discussion. Fixed effects estimated in this model are probabilities, on the same scale as the data. The other methods produce estimates on the logistic scale—that is, of regression parameters in the linear predictor of the logistic model. Given estimates of dispersion components, one can easily transform from one to the other.

Although derivations and specific algorithms differ for the Schall and Breslow/Clayton methods, the procedures are equivalent and can be simply described as application of best linear unbiased estimation (BLUE) of fixed and random effects and REML estimation of dispersion components in the "linear" model defined by the expression for the working variate in the usual generalized linear model fitting algorithm. This working variate includes the random effect term from the linear predictor.

The marginal methods are based on a quasi-likelihood approach that requires only first and second moments as, evidently, is penalized quasi-likelihood. The Bayesian approach of Karim and Zeger, on the other hand, is akin to a full-likelihood approach in that the complete distribution of the data enters into the estimation procedure. The computational intractability of explicit calculations involving the likelihood, or more precisely the posterior distribution, is avoided through implementation via Gibbs sampling Monte Carlo, although the method is still computationally intensive with respect to both programming and computing time.

As will be seen shortly, use of an approximate rather than an exact covariance does have a considerable effect on the estimates. Use of the exact posterior covariance is implicit in the Bayes/Gibbs method. Marginal REML may be implemented using either the exact unconditional covariance or a Taylor series approximation to it. The PQL methods utilize an approximate covariance, and it is not apparent how they might be extended to incorporate an exact covariance.

Table 1
Methods of estimation for logistic mixed models with crossed random effects

Method of Estimation of Dispersion Components	Model	Assumptions, Constraints	Variance
REML	Logistic–Normal mixed model → Marginal model with identity link	Normally-distributed random effects; unconditional mean linear in fixed effects	Exact, from logistic–normal mixture
Moment method (equating quadratic forms to expectations): McCullagh and Nelder	Logistic mixed model → Marginal model with identity link	Unconditional mean linear in fixed effects	Taylor series approximation to expectation of conditional covariance
1-iteration REML-T	Logistic mixed model → Marginal model with identity link	Unconditional mean linear in fixed effects	Taylor series approximation to expectation of conditional covariance
REML-T	Logistic mixed model → Marginal model with identity link	Unconditional mean linear in fixed effects	Taylor series approximation to expectation of conditional covariance
Bayes/Gibbs: Karim and Zeger	Logistic–normal mixed model	Normally distributed random effects	Exact posterior
Penalized quasi-likelihood (REML applied to working variate in GLM algorithm): Schall, Breslow and Clayton	Logistic mixed model		Approximate covariance of working variate

The approximate covariance that is used arises directly from the particular approximation of the quasi-likelihood that produces the penalized quasi-likelihood and the corresponding estimating equations.

The marginal REML and Bayes/Gibbs methods provide means of evaluating dispersion component estimates through profile, or residual, log-likelihood plots and posterior density plots, respectively.

5.2 *Results of the Analysis*

Mating probability estimates. In this particular example, due to the balance in the design, all four marginal methods produce the same least squares estimates of mating probabilities, which are simply sample proportions, given in Table 2. For unbalanced designs, different dispersion component estimates would lead to differences in the estimates of fixed effects. Estimates reported by Schall (1991) and Karim and Zeger (1992) are given in Table 3. These probabilities are transformations of regression parameters that were estimated directly. The Bayes/Gibbs estimates are in fairly close agreement with the quasi-likelihood estimates in Table 2. Schall's estimates differ more markedly, although as yet no measure of variability has been given with which to judge the differences.

The estimates cannot be expected to coincide exactly because the probabilities pertain to a marginal model and the regression parameters to a conditional model, and transforming from one to the other involves some degree of approximation. Strictly speaking, we want $\hat{\pi} = \mathrm{E}[F(\hat{\alpha} + \hat{\sigma}_{\mathrm{f}}^2 e_{\mathrm{f}} + \hat{\sigma}_{\mathrm{m}}^2 e_{\mathrm{m}})]$. When transforming from the logistic to the mean scale, Zeger, Liang, and Albert (1988) make an adjustment to account for the attenuating effect of the random term in the linear predictor by computing $\hat{\pi}$ as $F(a_l(D)\hat{\alpha})$, where $a_l(D) = (c^2(\sigma_{\mathrm{f}}^2 + \sigma_{\mathrm{m}}^2) + 1)^{-1/2}$ and $c = 16\sqrt{3}/(15\pi)$. It is likely that the Karim and Zeger estimates were calculated in this way. Duplication of the Schall analysis confirms that the estimates given above were calculated as $\hat{\pi} = F(\hat{\alpha})$, which is not an estimate of the correct parameter. The adjustment for attenuation, however, seems to overcompensate; the adjusted estimates differ from the least squares estimates in the opposite direction. The influence of the approximation for the covariance used in penalized quasi-likelihood may also contribute to the difference. Breslow

Table 2
Mating probability estimates

Experiment	$\hat{\pi}_{RR}$	$\hat{\pi}_{RW}$	$\hat{\pi}_{WR}$	$\hat{\pi}_{WW}$
Summer '86	.7333	.6667	.2333	.7000
Fall 1, '86	.6000	.4667	.2333	.6667
Fall 2, '86	.6667	.5333	.1667	.6333
Pooled estimate	.6667	.5556	.2111	.6667

Table 3
Mating probability estimates from alternative models

	$\hat{\pi}_{RR}$		$\hat{\pi}_{RW}$		$\hat{\pi}_{WR}$		$\hat{\pi}_{WW}$	
Experiment	K&Z	Schall	K&Z	Schall	K&Z	Schall	K&Z	Schall
Summer '86	.74	.7619	.67	.6865	.22	.1959	.72	.7340
Fall 1, '86	.59	.6101	.47	.4629	.21	.1830	.67	.6955
Fall 2, '86	.67	.6985	.54	.5402	.16	.1339	.65	.6584
Pooled estimate	.67		.56		.20		.68	

Estimates from Karim and Zeger (1992) and Schall (1991).

and Clayton reported regression parameters for the pooled analysis and for an alternative model with a seasonal effect, neither of which corresponds to a model analyzed by Schall. Their estimates on the logistic scale should be the same, however.

Dispersion component estimates. Estimates of the dispersion components by various methods for each of the three experiments and for the pooled data are given in Table 4. Figures 1 and 2 contain residual log-likelihood plots for the pooled data based on the exact and approximate covariances. These plots are representative of plots for the individual experiments, with the expected difference in precision and some particular features that are noted in the discussion below.

In the summer and first fall experiments, when estimates based on the exact covariance are compared, variability appears to be greater between female salamanders than between males, with an increase in variability between animals for both sexes in the fall. Conversely, in the second fall experiment, variability between males is greater than between females. Although the fact that the same salamanders were used in the summer and first fall experiments was ignored in the analysis, it is interesting to note the same pattern of relative variability in contrast to the opposite pattern in the second fall experiment with a fresh set of animals. Overall, the pooled estimate indicates similar variances for the two sexes.

The log-likelihood plots do not rule out negative values for the male dispersion component in the summer experiment or the female dispersion component in the second fall experiment. The random effects model cannot give rise to negative correlation, but it would be feasible in the physical context of these experiments. Either sex could be less inclined to mate following a successful mating on the previous occasion and conversely, more likely to mate after a period of celibacy. Or, it might be that after a number of successful matings (fewer than the number of occasions observed), a salamander's capacity is either satisfied or exhausted.

Table 4
Dispersion component estimates

	Summer		Fall 1		Fall 2		Pooled	
Method of estimation	σ_f^2	σ_m^2	σ_f^2	σ_m^2	σ_f^2	σ_m^2	σ_f^2	σ_m^2
REML: Exact covariance	1.68	.34	2.46	1.44	.69	2.40	1.67	1.50
Moment estimate:								
McCullagh and Nelder	1.37	.70	.98	.60	.40	1.34	.91	.88
1-iteration REML: Taylor covariance	1.30	.63	.92	.54	.33	1.28	.80	.75
REML: Taylor covariance	1.09	.14	.89	.55	.29	1.09	.71	.65
Bayes/Gibbs: Karim & Zeger[a]	2.35	.14	2.99	1.42	.33	2.89	1.50	1.36
PQL: Schall								
Breslow & Clayton	1.41	.09	1.26	.62	.26	1.50	.72	.63

[a] Median of posterior distribution.

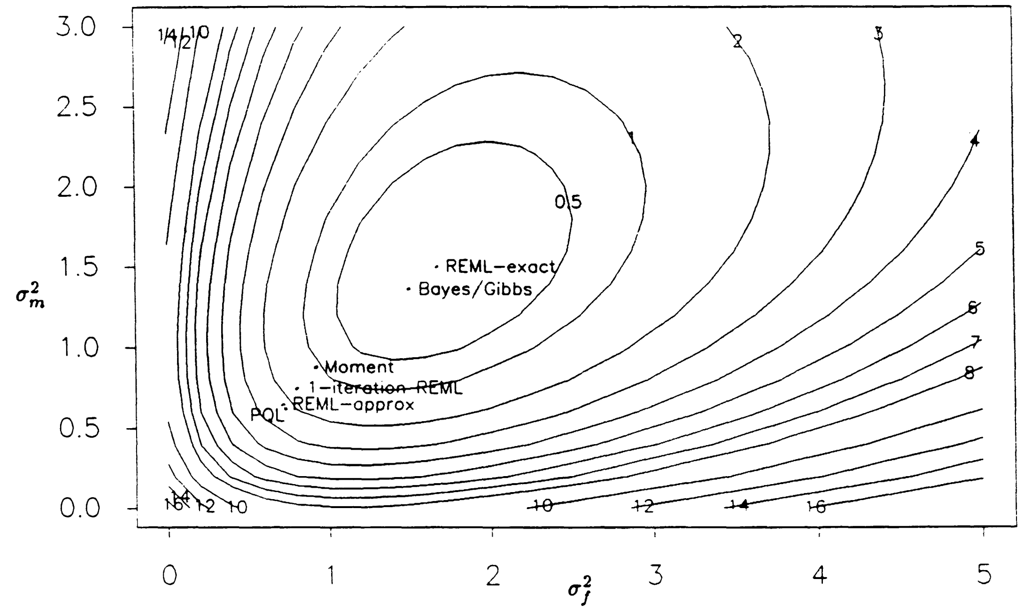

Figure 1. Residual log-likelihood plot for the pooled salamander data: Exact covariance.

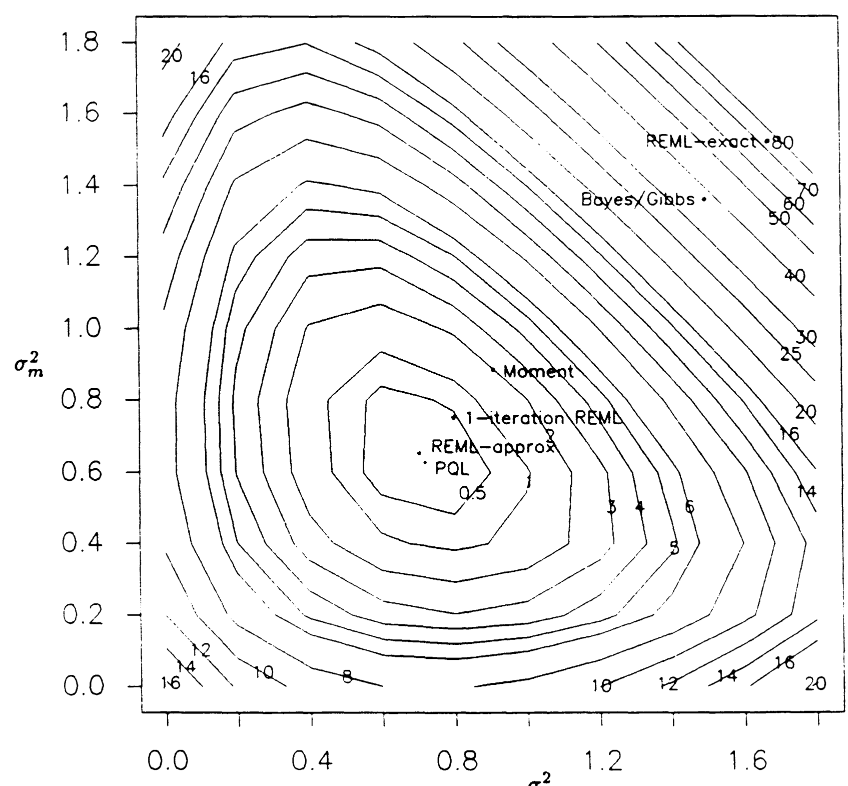

Figure 2. Residual log-likelihood plot for the pooled salamander data: Taylor series approximation.

The patterns commented on above are generally consistent for all methods, although less pronounced for estimates based on an approximate covariance. The estimates themselves, however, differ noticeably. For the fall experiments and the pooled data, estimates from methods employing the Taylor series approximation for the covariance and the PQL methods cluster together and are smaller in at least one component than the REML-exact and Bayes/Gibbs estimates, both of which are based on an exact covariance. For the summer experiment, the clustering is not apparent, but the relative size of the estimates holds for the female component. In both instances where an estimate based on the approximate covariance is larger than one based on the exact covariance, the REML estimates of the component are not significantly different from zero, according to the log-likelihood plot. In all cases, when the two REML estimates are compared, the estimate based on the approximate covariance is smaller in both components. Estimating equations are biased if the exact covariance is not used, so substantial differences are to be expected.

The residual log-likelihood plots provide a framework for evaluating the differences among the various estimates. Plots based on the exact covariance will be used for reasons that will be discussed subsequently. Even though estimates from different methods differ by as much as a factor of 2 or more in both components and by considerably more when single components are compared, for individual experiments the differences are small on the log-likelihood scale—that is, with respect to the precision of the REML-exact estimate. With the greater precision of the pooled analysis, these differences are more substantial. The log-likelihood differences between the REML estimate based on the approximate covariance and the two estimates based on the exact covariance do in fact exceed 2. Although this is a comparison between the largest and smallest of a number of estimates, it is also a comparison of inherent interest and is suggestive of a real difference due to the approximation, especially since this discrepancy is consistent. The approximation appears to lead to underestimation of the dispersion components. The estimates from Karim and Zeger (1992) are medians of a skewed posterior distribution, and might be expected to be larger than a (restricted) maximum likelihood estimate. No such pattern is apparent for these data, however.

The estimated covariance of fixed effects and standard errors of contrasts. Figure 3 gives (a multiple of) the estimated covariance of $\hat{\Pi}$ for the pooled data, written as a sum in which the first term is the variance under independence and the second consists of the variability from the random effects. The increase in the variance of the parameter estimates when random effects are incorporated in the analysis ranges from approximately 40% to over 70%. Table 5 gives the estimate of the contrast of primary interest, $\hat{\pi}_{RW} - \hat{\pi}_{WR}$, by the methods under consideration together with the standard error

REML—Exact covariance

$$90\ \mathrm{cov}_{est}(\hat{\Pi}) = \begin{bmatrix} .2222 & & & \\ & .2469 & & \\ & & .1665 & \\ & & & .2222 \end{bmatrix} + \begin{bmatrix} .1416 & .1203 & .0727 & \\ .1203 & .1653 & & .1077 \\ .0727 & & .0923 & .0805 \\ & .1077 & .0805 & .1416 \end{bmatrix}$$

REML—Taylor covariance

$$90\ \mathrm{cov}_{est}(\hat{\Pi}) = \begin{bmatrix} .2222 & & & \\ & .2469 & & \\ & & .1665 & \\ & & & .2222 \end{bmatrix} + \begin{bmatrix} .1343 & .1169 & .0722 & \\ .1169 & .1658 & & .1070 \\ .0722 & & .0754 & .0788 \\ & .1070 & .0788 & .1343 \end{bmatrix}$$

Figure 3. The estimated covariance of mating probability estimates for the pooled data. Estimated probabilities are for the four cross types, RR, RW, WR, and WW. The first term is the Bernoulli variance for a single observation; the second consists of the variability from the random effects.

Table 5
The contrast between mixed cross types, $\hat{\pi}_{RW} - \hat{\pi}_{WR}$

	Summer		Fall 1		Fall 2		Pooled	
Method of estimation	Est.	s.e.	Est.	s.e.	Est.	s.e.	Est.	s.e.
REML: Exact covariance	.433	.141	.233	.156	.367	.145	.345	.0863
Moment estimate	.433	.157	.233	.155	.367	.143	.345	.0904
1-iteration REML	.433	.154	.233	.153	.367	.150	.345	.0875
REML: Taylor covariance	.433	.142	.233	.152	.367	.150	.345	.0853
Bayes/Gibbs	.43[a]	[b]	.26[a]	[b]	.37[a]	[b]	.36[a]	[b]
Schall	.491		.280		.406			

[a] Median of posterior distribution.
[b] 5th and 95th percentiles of posterior distributions: (.18, .64), (−.02, .49), (.10, .58), and (.21, .49).

produced by each method. The Bayes/Gibbs method produces 5th and 95th percentiles of the posterior distribution of the contrast rather than standard errors.

The different methods of estimation produce essentially the same estimates and standard errors for $\hat{\pi}_{RW} - \hat{\pi}_{WR}$, except for Schall's, as a direct consequence of the difference in the parameter estimates noted earlier. The largest difference between the Bayes/Gibbs and marginal estimates of the contrast is less than 12%. In all cases, the inference is qualitatively the same. The estimated standard errors with σ^2 set to zero are .116, .119, .114, and .068 for the three experiments and for the pooled data, respectively. In one case, the first fall experiment, this smaller standard error changes the direction of the inference.

To get some idea of the sensitivity of the above standard error to the values of the dispersion components, the standard error was computed for all three experiments at the estimates in the table and at the four "extreme" points, north, south, east, and west of the REML-exact estimate that differed from it by roughly 2 in log-likelihood units. The variation in the standard error was small. The maximum difference in this standard error among any of the above values was less than the differences for any of these values and the estimate that assumes independence.

Correlations. Correlations between pairs of observations corresponding to the REML-exact parameter estimates range from approximately .04 when the dispersion component for the effect in common is smallest (summer, $\sigma_f^2 = .34$) to .20–.30 for the larger estimates. The relationship is not monotone since the correlations are functions of both male and female dispersion components as well as the fitted probabilities.

5.3 *The Exact Covariance Versus the Taylor Series Approximation*

The distinct elements of the exact covariance and of the Taylor series approximation were calculated for the summer data at values of the dispersion components corresponding to the estimates arrived at by the various methods in Table 1. The Taylor series approximation was from 30% larger to more than twice as large as the exact value—mostly, 50%–60% larger. The discrepancy was greatest when the sum of the variance components was large. Reversing this relationship, for a given covariance among observations, the Taylor series approximation would correspond to a smaller dispersion component. The exact covariance elements that were calculated were monotone increasing in the dispersion component for the effect in common. This relationship is obvious for the Taylor series approximation.

In every case, the log-likelihood using the exact covariance is considerably flatter than the log-likelihood based on the Taylor series approximation, $V_B + \sum_{r=1}^{c} V_B(Z_r Z_r^T - I)V_B$, which ceases to be positive definite if σ^2 is sufficiently large. An upper bound on the individual dispersion components, beyond which the approximation is not positive definite, is $\min\{1/[\hat{\pi}(1 - \hat{\pi})]\}$. Thus, the log-likelihood plots based on this approximation are unreliable for large values of σ^2, giving a misleading indication of the precision of the approximate REML estimates and magnifying the difference between estimates from various methods.

For the summer data, with the exact covariance, estimates by all methods considered fall within .3 log-likelihood units of the maximum, even though the estimates themselves appear to differ markedly. When the approximation is used, the other two REML estimates and the moment estimate differ from the REML-Taylor estimate by approximately 2 log-likelihood units. The results are similar for the fall experiments. Although the discrepancies are somewhat greater, all of the estimates lie within 1.5 units of the maximum of the plot based on the exact covariance. When the approximation is used, the REML-exact and Bayes/Gibbs estimates are beyond the boundaries of the plot, differing from the REML-Taylor estimate by more than 7 for the first fall experiment and by more than 3 for the second. For the pooled analysis, as described above, the log-likelihood based on the exact covariance gives a strong indication of a difference between the REML-exact (and Bayes/Gibbs) and the REML-Taylor estimates. The log-likelihood plot based on the approximate covariance shows the same differences, but to an extreme degree—more than 30 log-likelihood units when comparing the REML-Taylor to the Bayes/Gibbs estimate and more than 70 when comparing it to the REML-exact estimate.

The variability in the REML estimates based on the exact covariance as indicated by the log-likelihood is of approximately the same order as the variability in the Bayes/Gibbs estimates of Karim and Zeger as indicated by their 5th and 95th posterior percentiles.

6. Summary

Correlation due to repeated observations on the same animals needs to be taken into account in some way. In the analysis of the salamander data, from 30% to over 40% of the variability in the estimates of the mating probabilities for the four cross types is attributable to variability between salamanders. The standard errors of the main contrast of interest differ less among the various methods examined than between estimates from any of these methods and those based on the assumption of independence.

In spite of their similarity for this specific application, the methods are not without important differences. Use of the exact covariance is preferable to use of the Taylor series approximation. The latter not only tends to produce estimates of dispersion components that are too small, but also gives an indication of their precision that misleadingly excludes estimates based on the exact covariance. The failure of this approximate covariance to be positive definite for large values of the dispersion components is a serious drawback to its use. This consideration would lead to the Bayes/Gibbs method or to marginal REML with exact covariance. Both of these methods also yield plots that are useful for inference about dispersion components. The Bayes/Gibbs method has a broader range of applicability with respect to the designs that it will accommodate, but for designs with a simple fixed effects structure such as the salamander example, comparable results are obtained from REML at a fraction of the computational cost.

Résumé

L'estimation du maximum de vraisemblance sur les résidus (REML) est adaptée à certains modèles logistiques mixtes pour lesquels il est possible de représenter la moyenne non conditionnelle comme une fonction linéaire des effets fixés. On doit évaluer seulement les deux premiers moments de la distribution non conditionnée, et, excepté pour la forme de la covariance, l'algorithme de maximisation travaille directement sur des modèles linéaires. La covariance non conditionnée est calculée sur un mélange logistique–normale pour l'entrée dans l'algorithme. Les tracés du maximum de vraisemblance sur les résidus donnent un moyen d'inférence sur les composantes de dispersion. L'approximation par série de Taylor de cette covariance, présentant de plus une précision insuffisante, n'est pas définie positive pour de grandes valeurs des composantes de dispersion. Comme conséquence, les tracés du log du maximum de vraisemblance restreint, fondés sur les covariance approchée, attribuent de façon erronée une précision élevée aux estimations des composantes de dispersion. On présente la méthode à l'aide d'une expérience sur le croisement de salamandres dans laquelle les effets aléatoires correspondant aux animaux mâles et femelles interviennent dans un protocole croisé. On compare les analyses de ces données obtenues par différentes méthodes.

References

Breslow, N. E. and Clayton, D. G. (1993). Approximate inference in generalized linear mixed models. *Journal of the American Statistical Association* **88,** 9–25.

Hemmerle, W. J. and Hartley, H. O. (1973). Computing maximum likelihood estimates for the mixed A.O.V. model using the *W* transformation. *Technometrics* **15,** 819–831.

Karim, M. R. and Zeger, S. L. (1992). Generalized linear models with random effects; salamander mating revisited. *Biometrics* **48,** 631–644.

Kruskal, W. (1968). When are Gauss–Markov and least squares estimators identical? A coordinate-free approach. *Annals of Mathematical Statistics* **39,** 70–75.

McCullagh, P. and Nelder, J. A. (1989). *Generalized Linear Models.* London: Chapman and Hall.

Monahan, J. F. and Stefanski, L. A. (1992). Normal scale mixture approximations to $F^*(z)$ and computation of the logistic–normal integral. In *Handbook of the Logistic Distribution*, N. Balakrishnan (ed.), 529–540. New York: Marcel Dekker.

Schall, R. (1991). Estimation in generalized linear models with random effects. *Biometrika* **78,** 719–727.

Thisted, R. A. (1988). *Elements of Statistical Computing.* London: Chapman and Hall.

Zeger, S. L., Liang, K.-Y., and Albert, P. S. (1988). Models for longitudinal data: A generalized estimating equation approach. *Biometrics* **44,** 1049–1060.

Received September 1991; *revised March* 1992; *accepted March* 1992.

Appendix

Calculation of the Exact Covariance and Its Derivatives

In the balanced case, the unconditional mean, and hence the unconditional variance, of individual observations are calculated by least squares and are functionally independent of the dispersion components. Covariances of distinct observations are obtained as moments of conditional means.

Because conditional independence is assumed,

$$\begin{aligned}\mathrm{cov}(Y_u, Y_v) &= \mathrm{cov}[\mathrm{E}(Y_u|b), \mathrm{E}(Y_v|b)] \\ &= \mathrm{E}[\mathrm{E}(Y_u|b)\mathrm{E}(Y_v|b)] - \mathrm{E}[\mathrm{E}(Y_u|b)]\mathrm{E}[\mathrm{E}(Y_v|b)] \\ &= \mathrm{E}[\mathrm{E}(Y_u|b)\mathrm{E}(Y_v|b)] - \mathrm{E}(Y_u)\mathrm{E}(Y_v). \end{aligned} \tag{13}$$

The second term is estimated by least squares, so only the first term need be obtained directly by integration. For the salamander model, the conditional expectation of the uth observation is $F(x_u\alpha + \sigma_f z_u^f \varepsilon + \sigma_m z_u^m \varepsilon)$, where F is the logistic distribution function, and $\sigma_f \varepsilon_f = b_f$ and $\sigma_m \varepsilon_m = b_m$ are the random effects for individual female and male salamanders, respectively. For notational convenience, let $x_u\alpha = \eta_u$, $z_u^f\varepsilon = \varepsilon_{f,u}$, and $z_u^m\varepsilon = \varepsilon_{m,u}$. Assuming that the random effects follow a Gaussian distribution, $\mathrm{E}(Y_uY_v)$ for observations u and v with a male in common is

$$\int F(\eta_u + \sigma_f\varepsilon_{f,u} + \sigma_m\varepsilon_{m,u})F(\eta_v + \sigma_f\varepsilon_{f,v} + \sigma_m\varepsilon_{m,v})\phi(\varepsilon_{f,u})\phi(\varepsilon_{f,v})\phi(\varepsilon_{m,u})\, d\varepsilon_{f,u}\, d\varepsilon_{f,v}\, d\varepsilon_{m,u}. \tag{14}$$

Note that this integral involves the quantities η_u, η_v, σ_f, and σ_m. Current estimates of the latter two are provided at each iteration of the REML algorithm, but values of η_u and η_v must be obtained in some other way. Since an estimate, $x_u\hat{\pi}$, of $\mathrm{E}(Y_u)$ is available,

$$\hat{\mathrm{E}}(Y_u) = x_u\hat{\pi} = \int F(\hat{\eta}_u + \hat{\sigma}_f\varepsilon_{f,u} + \hat{\sigma}_m\varepsilon_{m,u})\phi(\varepsilon_{f,u})\phi(\varepsilon_{m,u})\, d\varepsilon_{f,u}\, d\varepsilon_{m,u} \tag{15}$$

can be solved numerically for $\hat{\eta}_u$ at each iteration. This step requires that the derivatives of $\eta_u(\pi, \sigma^2)$ and $\eta_v(\pi, \sigma^2)$ enter into the derivative of $\mathrm{E}(Y_uY_v)$.

Monahan and Stefanski (1992) provide a means of evaluating integrals such as (13) and (14) that is an alternative to numerical integration. They express the logistic distribution function as a normal scale mixture,

$$F(t) = \frac{e^t}{1 + e^t} \approx \sum_{i=1}^{k} p_{k,i}\Phi(ts_{k,i}), \tag{16}$$

where $\{s_{k,i}, p_{k,i}\}_{i=1}^{k}$ are constants that are tabled for $k = 1, \ldots, 8$. This leads to the approximation

$$\int F(\eta_u + \sigma_f\varepsilon_{f,u} + \sigma_m\varepsilon_{m,u})\phi(\varepsilon_{f,u})\phi(\varepsilon_{m,u})\, d\varepsilon_{f,u}\, d\varepsilon_{m,u} \approx \sum_{j=1}^{k} p_{k,j}\Phi\left(\frac{\eta_u s_{k,j}}{\sqrt{1 + s_{k,j}^2(\sigma_f^2 + \sigma_m^2)}}\right). \tag{17}$$

For $k = 5$, the error in this approximation is less than 10^{-6}, so (16) is essentially exact.

The corresponding approximation for $\mathrm{E}(Y_uY_v)$ with a male in common is

$$\sum_{i=1}^{k}\sum_{i=1}^{k} p_{k,i}p_{k,j} \cdot \Phi_2\left\{\frac{s_{k,i}\eta_u}{\sqrt{1 + s_{k,i}^2(\sigma_f^2 + \sigma_m^2)}}, \frac{s_{k,j}\eta_v}{\sqrt{1 + s_{k,j}^2(\sigma_f^2 + \sigma_m^2)}}, \frac{s_{k,i}s_{k,j}\sigma_m^2}{\sqrt{1 + s_{k,i}^2(\sigma_f^2 + \sigma_m^2)}\sqrt{1 + s_{k,j}^2(\sigma_f^2 + \sigma_m^2)}}\right\}, \tag{18}$$

where Φ_2 is the bivariate standard normal distribution function. This finite sum was compared to numerical integration by Simpson's rule (Thisted, 1988, p. 271) for selected values that occurred in the analysis of the summer data, and the values were essentially the same.

Calculation of the derivatives from (17) proceeds term by term as follows. First,

$$\frac{\partial\Phi_2}{\partial\sigma_r^2} = \frac{\partial\Phi_2}{\partial x_1}\left[\frac{\partial x_1}{\partial\sigma_r^2} + \frac{\partial x_1}{\partial\eta_u}\frac{\partial\eta_u}{\partial\sigma_r^2}\right] + \frac{\partial\Phi_2}{\partial x_2}\left[\frac{\partial x_2}{\partial\sigma_r^2} + \frac{\partial x_2}{\partial\eta_v}\frac{\partial\eta_v}{\partial\sigma_r^2}\right] + \frac{\partial\Phi_2}{\partial\rho}\frac{\partial\rho}{\partial\sigma_r^2}, \tag{19}$$

where

$$x_1 = \frac{s_{k,i}\eta_u}{\sqrt{1 + s_{k,i}^2(\sigma_f^2 + \sigma_m^2)}}, \quad x_2 = \frac{s_{k,j}\eta_v}{\sqrt{1 + s_{k,j}^2(\sigma_f^2 + \sigma_m^2)}}, \quad \rho = \frac{s_{k,i}s_{k,j}\sigma_m^2}{\sqrt{1 + s_{k,i}^2(\sigma_f^2 + \sigma_m^2)}\sqrt{1 + s_{k,j}^2(\sigma_f^2 + \sigma_m^2)}}.$$

Calculation of $\partial x_1/\partial\sigma_r^2$, $\partial x_2/\partial\sigma_r^2$, and $\partial\rho/\partial\sigma_r^2$ as well as $\partial x_1/\partial\eta_u$ and $\partial x_2/\partial\eta_v$ is straightforward. The derivatives of the standard bivariate normal distribution function with respect to its arguments are

$$\frac{\partial\Phi_2}{\partial x_1} = \phi(x_1)\Phi\left(\frac{x_2 - \rho x_1}{\sqrt{1 - \rho^2}}\right), \quad \frac{\partial\Phi_2}{\partial x_2} = \phi(x_2)\Phi\left(\frac{x_1 - \rho x_2}{\sqrt{1 - \rho^2}}\right), \quad \frac{\partial\Phi_2}{\partial\rho} = \phi_2(x_1, x_2, \rho). \tag{20}$$

We now need $\partial\eta_u/\partial\sigma_r^2$ and $\partial\eta_v/\partial\sigma_r^2$. Because we have only an implicit expression for η_u as a function of θ (and π_u) where $\theta = (\sigma_1^2, \ldots, \sigma_c^2)$, its derivative is obtained indirectly by application of the inverse function theorem to the transformation f: $(\eta_u, \theta) \rightarrow (\pi_u, \theta)$. By this theorem, under regularity conditions involving continuous differentiability and nonzero Jacobian determinant that ensure existence of a local inverse, the Jacobian matrices of the transformation and its inverse are inverses. That is, $J_{f,(\pi_u,\theta)} = J^{-1}_{f^{-1},(\eta_u,\theta)}$, where

$$J_{f,(\pi_u,\theta)} = \begin{pmatrix} \partial\eta_u/\partial\pi_u & \partial\eta_u/\partial\theta \\ \partial\theta/\partial\pi_u & \partial\theta/\partial\theta \end{pmatrix} = \begin{pmatrix} \partial\eta_u/\partial\pi_u & \partial\eta_u/\partial\theta \\ 0 & I \end{pmatrix} \tag{21}$$

and

$$J_{f^{-1},(\eta_u,\theta)} = \begin{pmatrix} \partial\pi/\partial\eta & \partial\pi/\partial\theta \\ \partial\theta/\partial\eta & \partial\theta/\partial\theta \end{pmatrix} = \begin{pmatrix} \partial\pi_u/\partial\eta_u & \partial\pi_u/\partial\theta \\ 0 & I \end{pmatrix}. \tag{22}$$

Note that the element $\partial\eta_u/\partial\theta = (\partial\eta_u/\partial\sigma_1^2, \ldots, \partial\eta_u/\partial\sigma_c^2)$ of $J_{f,(\pi_u,\theta)}$ contains the derivative $\partial\eta_u/\partial\sigma_r^2$ that is needed for evaluation of (18) but that is not directly available. The elements of $J_{f^{-1},(\eta_u,\theta)}$, on the other hand, can be calculated readily from (16), as

$$\frac{\partial\pi_u}{\partial\eta_u} = \sum_{i=1}^{k} p_{k,i}\phi\left(\frac{\eta_u s_{k,i}}{\sqrt{1 + s_{k,i}^2(\sigma_\mathrm{f}^2 + \sigma_\mathrm{m}^2)}}\right)\left(\frac{s_{k,i}}{\sqrt{1 + s_{k,i}^2(\sigma_\mathrm{f}^2 + \sigma_\mathrm{m}^2)}}\right), \tag{23}$$

and

$$\frac{\partial\pi_u}{\partial\theta_r} = \frac{\partial\pi_u}{\partial\sigma_r^2} = -\sum_{i=1}^{k} p_{k,i}\eta_u\left(\frac{s_{k,i}}{\sqrt{1 + s_{k,i}^2(\sigma_\mathrm{f}^2 + \sigma_\mathrm{m}^2)}}\right)^3 \phi\left(\frac{\eta_u s_{k,i}}{\sqrt{1 + s_{k,i}^2(\sigma_\mathrm{f}^2 + \sigma_\mathrm{m}^2)}}\right). \tag{24}$$

Inverting the partitioned matrix $J_{f^{-1},(\eta_u,\theta_r)}$ and equating corresponding elements of the two matrices yields an explicit expression for $\partial\eta_u/\partial\theta_r$ in terms of these easily calculated derivatives:

$$\frac{\partial\eta_u}{\partial\theta_r} = -\frac{\partial\pi_u}{\partial\theta_r}\left(\frac{\partial\pi_u}{\partial\eta_u}\right)^{-1}. \tag{25}$$

Correlation between random effects would result in alternative expressions for the variances, $\sigma_\mathrm{f}^2 + \sigma_\mathrm{m}^2$, and covariance, σ_m^2, in (17) and (18) with consequent changes to succeeding calculations.