Generalized Linear Mixed Models: A Pseudo-Likelihood Approach

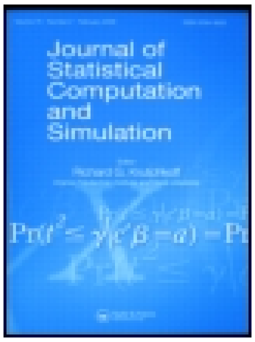
Article *in* Journal of Statistical Computation and Simulation · December 1993

DOI: 10.1080/00949659308811554

2 authors:

Russ Wolfinger
SAS Institute

Michael O'Connell
TIBCO



Taylor & Francis
Taylor & Francis Group

Journal of Statistical Computation and Simulation

ISSN: 0094-9655 (Print) 1563-5163 (Online) Journal homepage: http://www.tandfonline.com/loi/gscs20

Generalized linear mixed models a pseudo-likelihood approach

Russ Wolfinger & Michael O'connell

J. Statist. Comput. Simul., Vol. 48, pp. 233–243
Reprints available directly from the publisher
Photocopying permitted by license only

GENERALIZED LINEAR MIXED MODELS: A PSEUDO-LIKELIHOOD APPROACH

RUSS WOLFINGER

SAS Institute, Inc., SAS Campus Drive, Cary, North Carolina 27513-2414, U.S.A.

MICHAEL O'CONNELL

Becton Dickinson Research Center, Research Triangle Park, NC 27709-201, U.S.A.

(*Received July 1, 1993; in final form September 5, 1993*)

A useful extension of the generalized linear model involves the addition of random effects and/or correlated errors. A pseudo-likelihood estimation procedure is developed to fit this class of mixed models based on an approximate marginal model for the mean response. The procedure is implemented via iterated fitting of a weighted Gaussian linear mixed model to a modified dependent variable. The approach allows for flexible specification of covariance structures for both the random effects and the correlated errors. An estimate of an additional dispersion parameter for underlying exponential family distributions is optionally automatic. The method allows for subject-specific and population-averaged inference, and the Salamander data example from McCullagh and Nelder (1989) is used to illustrate both.

KEY WORDS: Covariance structure; Newton-Raphson; Restricted maximum likelihood.

1. INTRODUCTION

Generalized linear models enjoy much popularity since their introduction in Nelder and Wedderburn (1972), primarily because of their ability to handle discrete data via an extension of the familiar Gaussian regression model to models based on underlying exponential family distributions.

For the generalized linear mixed model (GLMM) consider a data vector $\mathbf{y}$ of length n satisfying

$$\mathbf{y} = \boldsymbol{\mu} + \mathbf{e}$$

and a differentiable monotonic link function $g(.)$ such that

$$g(\boldsymbol{\mu}) = \mathbf{X}\boldsymbol{\alpha} + \mathbf{Z}\boldsymbol{\beta}$$

where $\boldsymbol{\alpha}$ is a vector of unknown fixed effects with known model matrix $\mathbf{X}$ of rank p, and $\boldsymbol{\beta}$ is a vector of unknown random effects with known model matrix $\mathbf{Z}$. Assume $E(\boldsymbol{\beta}) = \mathbf{0}$ and $\text{cov}(\boldsymbol{\beta}) = \mathbf{D}$, where $\mathbf{D}$ is unknown.

Also, $\mathbf{e}$ is a vector of unobserved errors with $E(\mathbf{e}|\boldsymbol{\mu}) = \mathbf{0}$ and

$$\text{cov}(\mathbf{e}|\boldsymbol{\mu}) = \mathbf{R}_{\mu}^{1/2}\mathbf{R}\mathbf{R}_{\mu}^{1/2}$$

Here $\mathbf{R}_{\mu}$ is a diagonal matrix containing evaluations at $\boldsymbol{\mu}$ of a known variance function for the generalized linear model under consideration and $\mathbf{R}$ is unknown.

There has been considerable recent interest in the above extension of generalized linear models to include the random effects, $\boldsymbol{\beta}$. Several authors focus on longitudinal data (Stiratelli, Laird and Ware, 1984; Liang and Zeger, 1986; Zeger and Liang, 1986; Zeger, Liang and Albert 1988). Primary interest here is in the estimation of fixed effects; however, Liang and Zeger (1986) and Zeger, Liang and Albert (1988) discuss the interpretion of their estimates in both a subject-specific (SS) and population-averaged (PA) sense. An SS approach focuses on the estimates of $\boldsymbol{\beta}$ for individuals and their relation to the population parameters $\boldsymbol{\alpha}$, whereas under a PA approach interest is primarily in $\boldsymbol{\alpha}$ and variability arising from the random effects is treated essentially as a nuisance parameter. An example of an SS methodology is best linear unbiased prediction (BLUP) and Stein-type shrinkage estimation (Robinson, 1991), and an example of a PA approach to count data is given by Thall and Vail (1990).

In the above general set-up for a GLMM, variance modeling in $\boldsymbol{\beta}$ and $\mathbf{D}$ corresponds to an SS approach, and modeling in $\mathbf{R}$ corresponds to a PA approach. Our framework thus accounts for both methods, and an attractive feature is the flexibility in specifying covariance structures for $\mathbf{D}$ and/or $\mathbf{R}$. Jennrich and Schluchter (1986) discuss a variety of possible covariance structures, including unstructured, simple random effects, AR(1), and compound symmetry. For example, one can fit a generalized linear model with correlated random coefficients or autocorrelated errors, or even both.

Fitting a linear mixed model in a likelihood setting usually consists of specifying a Gaussian distribution for the random effects, and then estimating the unknown parameters using maximum likelihood (ML) or restricted/residual/marginal maximum likelihood (REML). For non-linear responses or cases other than an identity link function, this typically involves numerical integration. Numerical integration has been handled directly using quadrature (Anderson and Aitkin, 1985) and Gibbs sampling (Zeger and Karim, 1991). Stiratelli, Laird and Ware (1984), using the technique of Laird and Louis (1982), present a Gaussian approximation to the posterior distribution of the fixed and random effects that circumvents the need for numerical integration. Lindstrom and Bates (1990) apply this approximation to nonlinear mixed models. Schall (1991) uses the same approximation along with Harville's (1977, p. 328) modification of Henderson's (1984, Ch. 12) method of successive approximations to derive an algorithm for estimation in generalized linear models with random effects. Knuiman and Laird (1990) review many of the methods from a Bayesian perspective. Nearly all of the above authors focus on the case where either $\mathbf{D}$ or $\mathbf{R}$ is diagonal.

In this article we give pseudo-likelihood (PL) and restricted pseudo-likelihood (REPL) procedures for the estimation of $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$, $\mathbf{D}$, and $\mathbf{R}$. To quote Carroll and

Ruppert (1988, p. 71):

> Pseudo-likelihood estimates of θ are based on pretending that the regression parameter β is known and equal to the current estimate $\hat{\beta}_*$, and then estimating θ by maximum likelihood assuming normality . . .

Implementation of the PL procedure involves estimating β and θ in turn and iterating until convergence. For the GLMM, θ represents the unknown parameters in $\mathbf{D}$ and $\mathbf{R}$ and β represents $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$. In our approach, $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ are estimated from the mixed-model equations, and $\mathbf{D}$ and $\mathbf{R}$ are estimated using either ML or REML, thus yielding either PL or REPL, respectively.

Since the submission of this article, several others have appeared detailing similar approaches. Breslow and Clayton (1993) present two estimation procedures for GLMMs. They refer to these as penalized quasi-likelihood (PQL) and marginal quasi-likelihood (MQL). Although not mentioned as such, these methods in fact correspond to the SS and PA models of Zeger et al. (1988), respectively, and implementation of both PQL and MQL procedures may be achieved with our algorithm. In fact, a more general version of MQL is also possible, and this is discussed in the context of the Salamander example. Breslow and Clayton motivate their estimation procedures from a quasi-likelihood viewpoint, using approximations based on Laplace's method, slowly varying weights, and Pearson residuals. This is in contrast to our PL/REPL motivation based on a Gaussian approximation and Taylor's theorem.

Engel and Keen (1992) estimate the GLMM using a combination of quasi-likelihood and REML. The consider only the SS variance-components case, and motivate their algorithm by Taylor's approximations to various moments. An additional dispersion parameter is also available, as in our method. Their technique appears to be identical to PQL.

Waclawiw and Liang (1993) predict the random effects of a GLMM by iteratively solving a set of Stein-type estimating equations. This SS approach is similar to ours in its iterative nature, although they replace our mixed-model and ML/REML equations with optimal estimating equations for the fixed effects, random effects, and variance parameters.

Davidian and Giltinan (1993) discuss two classes of estimation methods for nonlinear mixed effects models which they call pooled two stage (PTS) and linearized mixed effects (LME) methods. PTS combines estimates obtained separately from each individual and the LME approach is a generalization of the linearization method of Vonesh and Carter (1992). Our approach is more similar to LME except that the LME linearization with respect to the random effects is about 0, in contrast to Lindstrom and Bates' (1990) and our approach which both linearize about a current estimate. The simpler LME linearization allows estimation of heteroscedastic intraindividual errors with a GLS-style implementation.

The approaches of Waclawiw and Liang (1993) and Davidian and Giltinan (1993) could potentially be more robust than ours to model misspecification, although this requires investigation. However these and the other developments discussed above appear to be more complicated than the PL/REPL motivation given in the next section. Furthermore, our technique provides a unified framework for both SS and PA inference, and includes PQL and MQL as special cases. Also, more general models are available by using non-trivial covariance structures for both **D** and **R**. Finally, the PL/REPL based algorithms can be readily implemented using recently developed mixed linear models software (SAS Institute Inc., 1992).

In what follows, Section 2 describes the estimation procedure and its associated approximations, Section 3 presents the PL and REPL-based algorithms for the GLMM, Section 4 discusses the Salamander example, and Section 5 contains some final remarks.

2. ESTIMATION

We now motivate the PL and REPL methods of fitting the GLMM by using three approximations, two analytic and one probabilistic.

For the first analytic approximation, let $\hat{\boldsymbol{\alpha}}$ and $\hat{\boldsymbol{\beta}}$ be known estimates of $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$, and define

$$\hat{\boldsymbol{\mu}} = g^{-1}(\mathbf{X}\hat{\boldsymbol{\alpha}} + \mathbf{Z}\hat{\boldsymbol{\beta}})$$

which is a vector consisting of evaluations of g^{-1} at each component of $\mathbf{X}\hat{\boldsymbol{\alpha}} + \mathbf{Z}\boldsymbol{\beta}$. Now let

$$\tilde{\mathbf{e}} = \mathbf{y} - \hat{\boldsymbol{\mu}} - (g^{-1})'(\mathbf{X}\hat{\boldsymbol{\alpha}} + \mathbf{Z}\hat{\boldsymbol{\beta}})(\mathbf{X}\boldsymbol{\alpha} - \mathbf{X}\hat{\boldsymbol{\alpha}} + \mathbf{Z}\boldsymbol{\beta} - \mathbf{Z}\hat{\boldsymbol{\beta}})$$

where $(g^{-1})'(\mathbf{X}\hat{\boldsymbol{\alpha}} + \mathbf{Z}\hat{\boldsymbol{\beta}})$ is a diagonal matrix with elements consisting of evaluations of the first derivative of g^{-1}. Note that $\tilde{\mathbf{e}}$ is a Taylor series approximation to $\mathbf{e} = \mathbf{y} - \boldsymbol{\mu}$, expanding about $\hat{\boldsymbol{\alpha}}$ and $\hat{\boldsymbol{\beta}}$.

Next, for the probabilistic approximation, we follow Laird and Louis (1982) and Lindstrom and Bates (1990), and approximate the conditional distribution of $\tilde{\mathbf{e}}$ given $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ with a Gaussian distribution having the same first two moments as $\mathbf{e}|\boldsymbol{\alpha},\boldsymbol{\beta}$ which we assume corresponds to $\mathbf{e}|\boldsymbol{\mu}$. In particular, we assume that $\tilde{\mathbf{e}}|\boldsymbol{\alpha},\boldsymbol{\beta}$ is Gaussian with mean $\mathbf{0}$ and variance $\mathbf{R}_{\mu}^{1/2}\mathbf{R}\mathbf{R}_{\mu}^{1/2}$.

The second and final analytic approximation is substituting $\hat{\boldsymbol{\mu}}$ for $\boldsymbol{\mu}$ in the variance matrix. Then, since

$$(g^{-1})'(\mathbf{X}_i\hat{\boldsymbol{\alpha}} + \mathbf{Z}_i\hat{\boldsymbol{\beta}}) = \frac{1}{g'(\hat{\mu}_i)}$$

for each component i, we can write

$$g'(\hat{\boldsymbol{\mu}})(\mathbf{y} - \hat{\boldsymbol{\mu}})|\boldsymbol{\alpha}, \boldsymbol{\beta} \sim N[\mathbf{X}\boldsymbol{\alpha} - \mathbf{X}\hat{\boldsymbol{\alpha}} + \mathbf{Z}\boldsymbol{\beta} - \mathbf{Z}\hat{\boldsymbol{\beta}}, g'(\hat{\boldsymbol{\mu}})\mathbf{R}_{\hat{\mu}}^{1/2}\mathbf{R}\mathbf{R}_{\hat{\mu}}^{1/2}g'(\hat{\boldsymbol{\mu}})]$$

where $g'(\hat{\boldsymbol{\mu}})$ is a diagonal matrix with elements constructed as above. If we define

$$\boldsymbol{\nu} = g(\hat{\boldsymbol{\mu}}) + g'(\hat{\boldsymbol{\mu}})(\mathbf{y} - \hat{\boldsymbol{\mu}})$$

Table 1 Exponential family canonical links and weights

Distribution	$g(\mu)$	*Weight*
Binomial	$\log[\mu/(1-\mu)]$	$\mu(1-\mu)$
Poisson	$\log(\mu)$	μ
Gamma	μ^{-1}	μ^2
Gaussian	μ	1

then we can equivalently specify

$$\boldsymbol{\nu}|\boldsymbol{\alpha},\boldsymbol{\beta} \sim N[\mathbf{X}\boldsymbol{\alpha} + \mathbf{Z}\boldsymbol{\beta}, g'(\hat{\boldsymbol{\mu}})\mathbf{R}_{\hat{\mu}}^{1/2}\mathbf{R}\mathbf{R}_{\hat{\mu}}^{1/2}g'(\hat{\boldsymbol{\mu}})]$$

Now by letting $\boldsymbol{\alpha}$ be an unknown parameter and assuming $\boldsymbol{\beta} \sim N(\mathbf{0}, \mathbf{D})$, this takes the form of a weighted linear mixed model with diagonal weight matrix

$$\hat{\mathbf{W}} = \mathbf{R}_{\hat{\mu}}^{-1}[g'(\hat{\boldsymbol{\mu}})]^{-2}$$

For canonical link functions $\hat{\mathbf{W}} = \mathbf{R}_{\hat{\mu}}$. The form of the diagonal elements of $\mathbf{W}$ are given in Table 1 for common canonical link functions.

The new response $\boldsymbol{\nu}$ is a Taylor series approximation to the linked response $g(\mathbf{y})$. It is analogous to the modified dependent variable used in the iteratively-reweighted-least-squares algorithm of Nelder and Wedderburn (1972). The Gaussian log likelihood corresponding to the linear mixed model for $\boldsymbol{\nu}$ is the following:

$$l(\boldsymbol{\alpha},\phi,\mathbf{D}^*,\mathbf{R}^*;\ \boldsymbol{\nu}) = -\frac{1}{2}\log|\phi\mathbf{V}| - \frac{1}{2}\phi^{-1}(\boldsymbol{\nu} - \mathbf{X}\boldsymbol{\alpha})^T\mathbf{V}^{-1}(\boldsymbol{\nu} - \mathbf{X}\boldsymbol{\alpha}) - \frac{n}{2}\log 2\pi$$

where

$$\mathbf{V} = \mathbf{W}^{-1/2}\mathbf{R}^*\mathbf{W}^{-1/2} + \mathbf{Z}\mathbf{D}^*\mathbf{Z}^T$$

Here we have introduced an additional scale/dispersion parameter, ϕ, and $\mathbf{D}^*$ and $\mathbf{R}^*$ are reparameterized versions of $\mathbf{D}$ and $\mathbf{R}$ in terms of ratios with ϕ. The inclusion of ϕ provides a quasi-likelihood-style extension of the generalized linear model, and it can be omitted if desired.

The log likelihood can be maximized analytically for $\boldsymbol{\alpha}$ and ϕ, yielding the profile/concentrated log likelihood

$$l(\mathbf{D}^*,\mathbf{R}^*;\ \boldsymbol{\nu}) = -\frac{1}{2}\log|\mathbf{V}| - \frac{n}{2}\log \mathbf{r}^T\mathbf{V}^{-1}\mathbf{r} - \frac{n}{2}[1 + \log(2\pi/n)] \tag{1}$$

where $\mathbf{r} = \boldsymbol{\nu} - \mathbf{X}(\mathbf{X}^T\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}^T\mathbf{V}^{-1}\boldsymbol{\nu}$. The analogous restricted/residual/marginal log likelihood is given by

$$l_R(\mathbf{D}^*,\mathbf{R}^*;\, \boldsymbol{\nu}) = -\frac{1}{2}\log|\mathbf{V}| - \frac{n-p}{2}\log \mathbf{r}^T\mathbf{V}^{-1}\mathbf{r} - \frac{1}{2}\log|\mathbf{X}^T\mathbf{V}^{-1}\mathbf{X}| - \frac{n-p}{2}\{1 + \log[2\pi/(n-p)]\} \tag{2}$$

Numerical methods are generally required to maximize l and l_R over the parameters in $\mathbf{D}^*$ and $\mathbf{R}^*$. Upon obtaining $\hat{\mathbf{D}}^*$ and $\hat{\mathbf{R}}^*$, estimates for $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$, and ϕ are computed as

$$\begin{aligned} \hat{\boldsymbol{\alpha}} &= (\mathbf{X}^T\hat{\mathbf{V}}^{-1}\mathbf{X})^{-1}\mathbf{X}^T\hat{\mathbf{V}}^{-1}\boldsymbol{\nu} \\ \hat{\boldsymbol{\beta}} &= \hat{\mathbf{D}}^*\mathbf{Z}^T\hat{\mathbf{V}}^{-1}\hat{\mathbf{r}} \\ \hat{\phi} &= \hat{\mathbf{r}}^T\hat{\mathbf{V}}^{-1}\hat{\mathbf{r}}/n^* \end{aligned} \tag{3}$$

where n^* equals n for ML and $n - p$ for REML. With $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$ and ϕ set to these estimates, (1) or (2) is maximized again to obtain updated estimates of $\mathbf{D}$ and $\mathbf{R}$. Iteration between (1) and (3) yields the PL estimation procedure and between (2) and (3) the REPL procedure.

3. THE ALGORITHMS

Algorithms for the PL and REPL estimation procedures are as follows:

1. Obtain an initial estimate of $\boldsymbol{\mu}$, $\hat{\boldsymbol{\mu}}$, using either the original data or an adjustment that permits the application of the link function. For example, a binary datum y can be replaced with $(y + 0.5)/2$.
2. Compute

$$\boldsymbol{\nu} = g(\hat{\boldsymbol{\mu}}) + (\mathbf{y} - \hat{\boldsymbol{\mu}})g'(\hat{\boldsymbol{\mu}})$$

3. Using either a ML or REML estimation procedure as per (1) or (2), fit a weighted linear mixed model with response variable $\boldsymbol{\nu}$, fixed effects model matrix $\mathbf{X}$, random effects model matrix $\mathbf{Z}$, and diagonal weight matrix

$$\hat{\mathbf{W}} = \mathbf{R}_{\hat{\mu}}^{-1}[g'(\hat{\boldsymbol{\mu}})]^{-2}$$

   This fitting procedure yields the estimates $\hat{\mathbf{D}}^*$ and $\hat{\mathbf{R}}^*$ which are then converted to $\hat{\mathbf{D}}$ and $\hat{\mathbf{R}}$ by using $\hat{\phi}$ from (3).
4. Compare the old estimates of $\hat{\mathbf{D}}$ and $\hat{\mathbf{R}}$ with the new ones. If the difference is sufficiently small, then stop; otherwise, go to the next step.

5. Solve the mixed-model equations for $\hat{\boldsymbol{\alpha}}$ and $\boldsymbol{\beta}$:

$$\mathbf{H}\begin{bmatrix}\hat{\boldsymbol{\alpha}}\\ \hat{\boldsymbol{\beta}}\end{bmatrix} = \begin{bmatrix}\mathbf{X}^T\hat{\mathbf{W}}^{1/2}\hat{\mathbf{R}}^{-1}\hat{\mathbf{W}}^{1/2}\boldsymbol{\nu}\\ \mathbf{Z}^T\hat{\mathbf{W}}^{1/2}\hat{\mathbf{R}}^{-1}\hat{\mathbf{W}}^{1/2}\boldsymbol{\nu}\end{bmatrix} \tag{4}$$

where

$$\mathbf{H} = \begin{bmatrix}\mathbf{X}^T\hat{\mathbf{W}}^{1/2}\hat{\mathbf{R}}^{-1}\hat{\mathbf{W}}^{1/2}\mathbf{X} & \mathbf{X}^T\hat{\mathbf{W}}^{1/2}\hat{\mathbf{R}}^{-1}\hat{\mathbf{W}}^{1/2}\mathbf{Z}\\ \mathbf{Z}^T\hat{\mathbf{W}}^{1/2}\hat{\mathbf{R}}^{-1}\hat{\mathbf{W}}^{1/2}\mathbf{X} & \mathbf{Z}^T\hat{\mathbf{W}}^{1/2}\hat{\mathbf{R}}^{-1}\hat{\mathbf{W}}^{1/2}\mathbf{Z} + \hat{\mathbf{D}}^{-1}\end{bmatrix}$$

These estimates are the same as those given by (3).

6. Compute the new estimate of $\boldsymbol{\mu}$ by substituting $\hat{\boldsymbol{\alpha}}$ and $\hat{\boldsymbol{\beta}}$ in the expression

$$\hat{\boldsymbol{\mu}} = g^{-1}(\mathbf{X}\hat{\boldsymbol{\alpha}} + \mathbf{Z}\hat{\boldsymbol{\beta}})$$

and go to step 2.

In Step 3, Newton-Raphson (NR) is preferred to Expectation-Maximization (EM) for reasons discussed in Lindstrom and Bates (1988). First, the number of NR iterations required for convergence is often orders of magnitude less than that required for EM. Also, NR enables the use of an orthogonality convergence criterion, as well as automatically providing the Hessian matrix for approximate standard errors. Neither of these benefits are available without modification to the EM method used by Stiratelli, Laird, and Ware (1984) and Schall (1991).

The elements of $\hat{\mathbf{D}}$ and $\hat{\mathbf{R}}$ are used in Step 5 instead of $\hat{\boldsymbol{\alpha}}$ and $\hat{\boldsymbol{\beta}}$ because $\hat{\mathbf{D}}$ and $\hat{\mathbf{R}}$ are inherently less precise. It is tacitly assumed that the number of unknown parameters in $\mathbf{D}$ and $\mathbf{R}$ is small enough to permit an NR approach. This is not extremely restrictive in practice because often both $\mathbf{D}$ and $\mathbf{R}$ are block-diagonal. Efficiency can be gained by using the estimates from Step 5 as starting values for the subsequent Step 3. The final convergence of the algorithms is subject to the usual caveats involved in numerical optimization (McCullagh and Nelder, 1989, p. 117).

As in McCullagh and Nelder (1989, pp. 40–43), each loop through the algorithms consists of a Fisher scoring step to optimize a likelihood. In the case of canonical link functions, as those listed in Table 1, this is equivalent to a full NR approach because the exact and expected Hessians coincide.

$\mathbf{H}^{-1}$ from Step 5 is an approximate covariance matrix for $\hat{\boldsymbol{\alpha}}$ and $\hat{\boldsymbol{\beta}}$. Suitable modifications can be made to the coefficient matrix if $\mathbf{D}$ is singular (Henderson, 1984, p. 48). The fixed-effects estimate, $\hat{\boldsymbol{\alpha}}$, can be interpreted in a SS sense if $\mathbf{D}$ is nontrivial, and in a PA sense if $\mathbf{R}$ is nontrivial. The random-effects estimate, $\hat{\boldsymbol{\beta}}$, is available in SS analyses and has interpretations in both empirical Bayes and BLUP contexts.

4. EXAMPLE

We now apply the PL and REPL procedures to the Salamander mating data analyzed in McCullagh and Nelder (1989, Ch. 14.5), Schall (1991), and Breslow and

Clayton (1993). The experiment involves two populations of salamanders, Rough Butt and Whiteside. The crossed mating design involves 20 males and females from each population and is described in Table 14.3 of McCullagh and Nelder (1989). The result is 120 binary observations indicating whether or not mating occurs. The logit of the mating probability is modeled as a sum of a fixed effect for crossing and random effects for the male and female. For example,

$$\log \frac{\pi_{RRij}}{1 - \pi_{RRij}} = \alpha_{RR} + \beta_{Fi} + \beta_{Mj}$$

where μ_{RRij} is the mating probability of the ith Rough Butt female and the jth Rough Butt male, α_{RR} is the average logit of Rough Butt-Rough Butt mating, and β_{Fi} and β_{Mj} are random effects from the female and male individuals in the pair.

The experiment is repeated three times, and we analyze the data separately for each repetition for comparison purposes. Though not done here, it is quite an easy matter to pool the data and find joint estimates (see Breslow and Clayton, 1993). The fact that the first two experiments involve the same animals can then be accounted for by appropriate variance models.

Calculations in Steps 3 and 4 of the PL and REPL methods were performed using the recently developed SAS/STAT MIXED procedure (1992) on a Hewlett-Packard series 700 workstation, and other calculations were performed in SAS data sets. All problems converged in less than 6 iterations of algorithms, and each iteration took less than 5 seconds real time. The internal NR step within each iteration took 1-4 iterations.

We perform both SS and PA analyses. The SS results are reported in Table 2, excluding the random effects estimates. The estimation procedure of Schall (1991) and the PQL method of Breslow and Clayton (1993) produce estimates equivalent to those from our algorithms with ϕ constrained to equal one. For all three experiments, the unconstrained estimate of ϕ is around 0.65. This is accompanied by somewhat larger estimates for the dispersion components than those obtained from Schall's approach. Our algorithms converge in much fewer iterations than Schall's EM algorithm because of the more computationally intensive NR optimization performed in Step 3.

The PA results obtained with $\mathbf{D} = \mathbf{Z} = \mathbf{0}$ and different forms for $\mathbf{R}$ are given in Table 3. The method-of-moments estimates of McCullagh and Nelder (1989) are included for comparison. The REPL estimates of the dispersion components are always larger than those from PL. Both PL and REPL estimates are smaller than those of McCullagh and Nelder for Experiments 1 and 3 and both are larger for Experiment 2. Because of the crossed nature of the female and male variance components, it was necessary to specify an $\mathbf{R}$ matrix that was not block-diagonal. For larger problems this kind of specification may be prohibitive computationally.

An interesting generalization of the MQL method of Breslow and Clayton (1993) is illustrated in Table 3 by the rows labeled REPL* and REPL*-CS, both representing models that omit the male random effect. The REPL* row corresponds to the MQL approach, in which the weight matrix does not multiply the component of $\mathbf{R}$ corresponding to the random effects, that is

$$\mathbf{R} = \sigma_F^2 \mathbf{Z}_F \mathbf{Z}_F^T + \phi \mathbf{W}^{-1}$$

Table 2 Salamander mating data: summary of subject-specific parameter estimates for three experiments

Experiment	*Method*	*Parameter estimate*				*Variance estimate*		
		$\hat{\pi}_{RR}$	$\hat{\pi}_{RW}$	$\hat{\pi}_{WR}$	$\hat{\pi}_{WW}$	$\hat{\sigma}_F^2$	$\hat{\sigma}_M^2$	$\hat{\phi}$
1	PL-S	0.7563	0.6836	0.2028	0.7280	1.1511	0.0066	1.0000
	PL	0.7874	0.6973	0.1692	0.7556	1.6436	0.4704	0.6613
	REPL-S	0.7619	0.6865	0.1959	0.7340	1.4099	0.0896	1.0000
	REPL	0.7966	0.7017	0.1594	0.7641	2.0199	0.6318	0.6644
	REPL*-S	0.7609	0.6870	0.1956	0.7350	1.4679	0.0000	1.0000
	REPL*	0.7717	0.6952	0.1784	0.7521	1.8840	0.0000	0.7686
2	PL-S	0.6079	0.4638	0.1943	0.6897	0.9497	0.4404	1.0000
	PL	0.6187	0.4589	0.1375	0.7171	2.0770	1.2130	0.6358
	REPL-S	0.6101	0.4629	0.1830	0.6955	1.2584	0.6161	1.0000
	REPL	0.6203	0.4578	0.1272	0.7227	2.6079	1.5245	0.6397
3	PL-S	0.6917	0.5386	0.1401	0.6541	0.1517	1.1864	1.0000
	PL	0.7201	0.5462	0.1153	0.6716	0.5826	1.8587	0.6813
	REPL-S	0.6985	0.5402	0.1339	0.6584	0.2618	1.4988	1.0000
	REPL	0.7263	0.5482	0.1100	0.6757	0.7232	2.2396	0.6948

S, estimates from Schall (1991, Table 1).
*Results from fitting only the female effects.

Table 3 Salamander mating data: summary of population-averaged parameter estimates for three experiments

Experiment	*Method*	*Parameter estimate*				*Variance estimate*		
		$\hat{\pi}_{RR}$	$\hat{\pi}_{RW}$	$\hat{\pi}_{WR}$	$\hat{\pi}_{WW}$	$\hat{\sigma}_F^2$	$\hat{\sigma}_M^2$	$\hat{\phi}$
1	M&N	0.7333	0.6667	0.2333	0.7000	1.3704	0.6963	1.0000
	PL	0.7333	0.6667	0.2333	0.7000	1.0715	0.0926	0.7611
	REPL	0.7333	0.6667	0.2333	0.7000	1.2304	0.1386	0.7705
	REPL*	0.7333	0.6667	0.2333	0.7000	1.2919	0.0000	0.7923
	REPL*-CS	0.7333	0.6667	0.2333	0.7000	0.2611	0.0000	0.7914
	REPL*-AR(1)	0.7294	0.6712	0.2319	0.6892	0.2179	0.0000	1.0301
2	M&N	0.6000	0.4667	0.2333	0.6667	0.9787	0.5997	1.0000
	PL	0.6000	0.4667	0.2333	0.6667	1.0175	0.6108	0.6632
	REPL	0.6000	0.4667	0.2333	0.6667	1.1731	0.7100	0.6681
3	M&N	0.6667	0.5333	0.1667	0.6333	0.3954	1.3440	1.0000
	PL	0.6667	0.5333	0.1667	0.6333	0.2437	1.0939	0.7186
	REPL	0.6667	0.5333	0.1667	0.6333	0.3020	1.2597	0.7261

M&N, estimates from McCullagh and Nelder (1989, Table 14.10).
*Results from fitting only the female effects.
CS, compound-symmetry structure in **R**, estimate under $\hat{\sigma}_F^2$.
AR(1), first-order autoregressive structure in **R**, estimate under $\hat{\sigma}_F^2$.

where $\mathbf{Z}_F$ denotes the model matrix for the females; see (20) in Breslow and Clayton (1993). However, in the REPL*-CS row, the weight matrix does multiply the random effects component:

$$\mathbf{R} = \mathbf{W}^{-1/2}(\sigma_F^2 \mathbf{Z}_F \mathbf{Z}_F^T + \phi \mathbf{I})\mathbf{W}^{-1/2}$$

where **I** is the identity matrix. The REPL* estimates correspond to an equicovariance structure for **R** whereas the REPL*-CS estimates correspond to an equicorrelation structure for **R**. The resulting estimates of σ_F^2 are not directly comparable.

For the crossed dispersion-component design, the PA fixed-effects estimates agree with those of McCullagh and Nelder and, because of the balance in the design, do not depend on the valaues of the dispersion components (McCullagh and Nelder, 1989, Exercises 14.8 and 14.9). The same estimates are produced when the male dispersion component is deleted from the model, resulting in a block-diagonal **R** with blocks corresponding to females and each block structured as compound-symmetry. However, when each block has a first-order autoregressive structure, the fixed-effects estimates are influenced. For these data the influence is slight because the autoregressive dispersion parameter esitmate of 0.2179 corresponds to an **R** matrix that is close to that corresponding to the compound-symmetry estimate of 0.2611. Both the equicorrelation and AR(1) model estimates of the female variance component are small compared with the other model estimates because they are multiplied by the weight matrix and the others are not.

5. DISCUSSION

The very successful application of generalized linear models and quasi-likelihood methods in the analysis of discrete data opens a wide range of possibilities for generalized linear mixed models. Breslow and Clayton (1993) present several interesting applications of the GLMM. These include examples of logistic regression incorporating overdispersion, Poisson regression with intrasubject correlation and overdispersion, and log-odds-ratio regression. As formulated here, the fitting procedure allows for very flexible specification of covariance structures for both the random effects and correlated errors. Our analysis of the Salamander data includes equicorrelation, equicovariance, and AR(1) covariance structures for **R**. Similar structuring of the covariance for the random effects in **D** is also straightforward. Further, since the approach is available via the recently developed SAS/STAT MIXED procedure, implementation is simple.

The motivation and derivation of PL and REPL procedures of this paper are very similar to the work of Lindstrom and Bates (1990). Their algorithm for nonlinear mixed models consists of iterating between a ML or REML fit of a linear mixed model to estimate **D** and **R**, and a nonlinear least-squares fit of a nonlinear fixed-effects model to estimate $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$. The first part of this procedure corresponds precisely to maximizing our equations (1) or (2), but the last part is replaced by solving the mixed-model equations (4). This solution can be viewed as the first Gauss-Newton step in a nonlinear optimization. Therefore, under suitable regularity conditions, it also appears that the PL or REPL approach can be applied to general nonlinear mixed models to produce Lindstrom-Bates estimates.

Acknowledgements

We thank Peter McCullagh for helpful comments.

References

Anderson, D. A. and Aitkin, M. (1985). Variance component models with binary response. *Journal of the Royal Statistical Society* B **47,** 203–10.

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

Carroll, R. J. and Ruppert, D. (1988). *Transformation and Weighting in Regression,* London: Chapman and Hall.

Davidian, M. and Giltinan, D. (1993). Some general estimation methods for nonlinear mixed effects models. *Journal of Biopharmaceutical Statistics* **3,** 23–55.

Engel, B. and Keen, A. (1992). A simple approach for the analysis of generalized linear mixed models. LWA-92-6, Agricultural Mathematics Group (GLW-DLO), Wageningen, The Netherlands.

Harville, D. A. (1977). Maximum likelihood approaches to variance component estimation and to related problems. *Journal of the American Statistical Association* **72,** 320–38.

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

Jennrich, R. I. and Schluchter, M. D. (1986). Unbalanced repeated-measures models with structured covariance matrices. *Biometrics* **42,** 805–20.

Knuiman, M. W. and Laird, N. M. (1990). Parameter estimation in variance component models for binary response data. *Advances in Statistical Methods for Genetic Improvement of Livestock.* D. Gianola and K. Hammond (Eds.). Springer-Verlag. 177–189.

Laird, N. M. and Louis, T. A. (1982). Approximate posterior distributions for incomplete data problems. *Journal of the Royal Statistical Society* B **44,** 190–200.

Liang, K. Y. and Zeger, S. L. (1986). Longitudinal data analysis using generalized linear models. *Biometrika* **73,** 13–22.

Lindstrom, M. J. and Bates, D. M. (1988). Newton-Raphson and EM algorithms for linear mixed-effects models for repeated-measures data. *Journal of the American Statistical Association* **83,** 1014–22.

Lindstrom, M. J. and Bates, D. M. (1990). Nonlinear mixed effects models for repeated measures data. *Biometrics* **46,** 673–87.

McCullagh, P. and Nelder, J. A. (1989). *Generalized Linear Models,* 2nd ed. London: Chapman and Hall.

Nelder, J. A and Wedderburn, R. W. M. (1972). Generalized linear models. *Journal of the Royal Statistical Society* A **135,** 370–84.

Robinson, G. K. (1991). That BLUP is a good thing: the estimation of random effects (with discussion). *Statistical Science* **6,** 15–51.

SAS Institute Inc. (1992). SAS Technical Report P-229. *SAS/STAT Software: Changes and Enhancements, Release 6.07.* Chapter 16: The MIXED Procedure. Cary, NC: SAS Institute Inc.

Schall, R. (1991). Estimation in generalized linear models with random effects. *Biometrika* **78,** 719–27.

Stiratelli, R., Laird, N. and Ware, J. H. (1984). Random-effects models for serial observations with binary response. *Biometrics* **40,** 961–71.

Thall, P. F. and Vail, S. C. (1990). Some covariance models for longitudinal count data with overdispersion. *Biometrics* **46,** 657–71.

Vonesh, E. F. and Carter, R. L. (1992). Mixed effects models for unbalanced repeated measures. *Biometrics* **48,** 1–18.

Waclawiw, M. A. and Liang, K. Y. (1993). Prediction of random effects in the generalized linear model. *Journal of American Statistical Association* **88,** 171–8.

Zeger, S. L. and Karim, M. R. (1991). Generalized linear models with random effects; a Gibbs sampling approach. *Journal of the American Statistical Association* **86,** 79–86.

Zeger, S. L. and Liang, K. Y. (1986). Longitudinal data analysis for discrete and continuous outcomes. *Biometrics* **42,** 121–30.

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