Institute of Mathematical Statistics

LECTURE NOTES — MONOGRAPH SERIES

ON ESTIMATING FUNCTION APPROACH IN THE GENERALIZED LINEAR MIXED MODEL

Brajendra C. Sutradhar
Memorial University of Newfoundland

V. P. Godambe
University of Waterloo

ABSTRACT

Waclawiw and Liang (1993) develop an estimating function-based approach to component estimation in the generalized linear mixed model with univariate random effects and a vector of fixed effects. In their approach they utilize the standard optimal estimating functions to estimate the fixed effects and a so-called Stein-type form of estimating functions to estimate both the random effects and their variance. In this paper, we provide a semiparametric solution to the estimation problem dealt by Waclawiw and Liang. The solution is obtained under two set-up by utilizing the standard theory of optimal estimating functions (Godambe and Thompson, 1989). Under the first set-up, the solution is obtained in three steps. In the first step, the estimating functions for the regression parameters, and the random effects are developed by treating the random effects as fixed effects. In the second step, we obtain the prediction of the random effects by taking their true nature of randomness into account. These predicted random effects are then used in the estimating equations for the regression parameters, of Step 1, to obtain their improved estimates. In the third step, the estimating function for the variance of the random effects is developed based on the true nature of the random effects. Under the second set-up, the estimating functions for the regression parameters, random effects and their variance are developed by utilizing the true nature of the random effects directly. Results of a small simulation study based on the performance of the proposed estimating function-based approaches are reported.

Key Words: Random effects; variance component of the random effects; semi-parametric solutions; standard estimating function approach; unconditional and conditional mixed methods; corrected conditional mixed method.

1 INTRODUCTION

The generalized linear model (McCullah and Nelder (1989)) neatly synthesizes likelihood-based approaches to regression analysis for a variety of outcome measures. The underlying distribution of the outcome variables is assumed to be of the exponential family form, and a link function transformation of the expectation is modelled as a linear function of observed covariates. Several recent extensions of this useful theory involve models with random terms in the linear expectation. Such generalized linear mixed models are useful for accommodating the overdispersion often observed among outcomes that nominally have binomial (Williams (1982)) or Poisson (Breslow (1984)) distributions; and for modelling the dependence among outcome variables inherent in longitudinal or repeated measures designs (cf. Laird and Ware (1982), Stiratelli, Laird and Ware (1984), Zeger, Liang and Albert (1988), Zeger and Karim (1991)).

Consider a set of repeated observations consisting of a response y_{ij} as the jth $(j = 1, \dots, n_i)$ repeated observation on individual $i(i = 1, \dots, k)$ and a $p \times 1$ vector x_{ij} of covariates associated with that response. Let β denote a $p \times 1$ vector of unknown fixed effect parameters associated with covariate x_{ij}. Further, let γ_i be a univariate random effect such that for a given γ_i, n_i observations due to the ith individual are independent.

Under the assumption that the conditional density of y_i, the $n_i \times 1$ vector of responses for individual i, given γ_i is of the exponential form

$$f(y_i|\gamma_i) = \exp\left\{\sum_{j=1}^{n_i} \eta_{ij} y_{ij} - \sum_{j=1}^{n_i} \phi(\eta_{ij})\right\}, \tag{1.1}$$

with $\eta_{ij} = x_{ij}^T\beta + \gamma_i$, and $\gamma_i \overset{\text{iid}}{\sim} N(0, \sigma^2)$, recently Waclawiw and Liang (1993) have used a three-step iterative procedure to estimate all three unknowns β, γ_i $(i = 1, \dots, k)$, and σ^2. The three steps are:

1. Assuming an initial value for σ^2, the fixed effects β are computed using the generalized estimating equation of the form

$$U(\beta, \sigma^2) = \sum_{i=1}^{k} \left(\frac{\partial \mu_i(\beta, \sigma^2)}{\partial \beta}\right)^T V_i^{-1}(\beta, \sigma^2)(y_i - \mu_i(\beta, \sigma^2)) = 0, \tag{1.2}$$

   where $\mu_i(\beta, \sigma^2)$ and $V_i(\beta, \sigma^2)$ are, respectively, marginal mean and variance-covariance matrix of the response vector y_i for the ith individual.

2. Assuming that σ^2 and β are fixed, the equation $\tilde{g}_i = 0$ is solved for γ_i, where

$$\tilde{g}_i(y_i, \gamma_i, \beta) = g_i(a_i y_i + b_i, \gamma_i, \beta) \tag{1.3}$$

is a class of estimating functions for γ_i, $g_i(y_i, \gamma_i, \beta)$ being the score equation given by

$$\begin{aligned} g_i(y_i, \gamma_i, \beta) &= \partial \log f(y_i|\gamma_i)/\partial \gamma_i \\ &= \sum_{j=1}^{n_i} (y_{ij} - \phi'(\eta_{ij})) = 0, \end{aligned} \tag{1.4}$$

where $\phi'(\eta_{ij})$ is the first derivative of $\phi(\cdot)$ in (1.1) with respect to η_{ij}, which is, in fact, the conditional mean of y_{ij} given γ_i. Note that in (1.3), they have obtained a_{ij} and b_i by following certain optimal criteria due to Godambe (1960) [see also Ferreira (1982)]. Let a^*_{ij} and b^*_i be such solutions. An optimal function $\tilde{g}^*_i$ is obtained by substituting a^*_{ij} and b^*_i in $\tilde{g}_i$ given in (1.3). Next a Stein-type estimation for γ_i, γ^*_i is achieved by solving $\tilde{g}^*_i = 0$.

3. The variance component of the random effects σ^2 is estimated by using the relationship

$$E(\gamma_i^{*^2}) \simeq E(\gamma_i^2) - E(\gamma^*_i - \gamma_i)^2, \tag{1.5}$$

where $(\gamma^*_i - \gamma_i)$ is obtained by first expanding the optimal estimating function to the first order approximation and then solving the resulting optimal estimating equation. Note that in this step, one actually obtains a recursive form for the estimation of σ^2, which must be updated with changes in β and γ_i. Let σ^{*^2} be the estimator of σ^2.

The above three steps of the iterative procedure describe a complete cycle or one full iteration. The cycles of iteration continue until convergence (if exists) is achieved. However, convergence is to be investigated and further work is needed to guarantee convergence. Neither in Waclawiw and Liang (1993) nor in the present paper, this is attempted.

More recently, Sutradhar and Qu (1997) have shown for a Poisson mixed model that even if γ^*_i in Step 2 is computed based on large n_i, the estimator σ^{*^2} of σ^2 obtained from the third step, does not converge to σ^2. These authors proposed a likelihood approximation (valid for small σ^2) to estimate all three parameters β, γ_i and σ^2 of this special model. The estimation is carried out in two steps. In the first step, they utilize a small σ^2 based approximate likelihood function to estimate the fixed effect parameters and σ^2. In the second step, they estimate the random effects γ_i by its posterior mean $E(\gamma_i|y_i)$, which is, in fact, the minimum mean square error prediction of γ_i. It was shown by Sutradhar and Qu (1997) through a simulation study that their likelihood estimation approach performs much better than Waclawiw and Liang's three steps estimation approach in estimating all three parameters β, γ_i and σ^2. The computation of the likelihood function is, however, not easy in general.

This paper, unlike Waclawiw and Liang (1993), and Sutradhar and Qu (1997), provides a semi-parametric solution to the estimation problem dealt by these authors. That is, we do not make any distributional assumption for the random effects. The solution is obtained under two set-up by utilizing the standard theory of optimal estimating functions [cf. Godambe and Thompson (1989), Godambe and Kale (1991)]. Under the first set-up, the optimal estimating functions for the regression parameters and the random effects are developed by treating random effects γ_i fixed. The random nature of γ_i is, however, taken into account when estimating equations are solved for the parameters. Next the estimating function for σ^2 is developed based on the true nature of the random effects. Under the second set-up, estimating functions for the regression parameters, random effects and their variance are developed by treating γ_i as random effects as they should be.

The performance of the estimators obtained under these two set-up are also compared through a simulation experiment, for Poisson mixed models.

2 OPTIMAL ESTIMATION WHEN RANDOM EFFECTS ARE TREATED INITIALLY AS FIXED EFFECTS

Assume that given γ_i $(i = 1, \ldots, k)$, the response y_{ij} has the mean $\phi'(\eta_{ij})$ and variance $\phi''(\eta_{ij})$ with $\eta_{ij} = x_{ij}^T\beta + \gamma_i$, where $\phi'(\eta_{ij})$ is the first derivative of $\phi(\cdot)$ with respect to η_{ij}, as in (1.4), and $\phi''(\eta_{ij})$ is the second derivative, $\phi(\cdot)$ being a known functional form. Further assume that γ_i's are independently and identically distributed with zero mean and variance σ^2, but the specific form of the distribution of γ_i is not known.

Now by holding γ_i $(i = 1, \ldots, k)$ fixed, we construct the optimal estimating functions for β and γ_i, following Godambe and Thompson (1989). These functions are, respectively, given by

$$g_1 = \sum_{i=1}^{k} \sum_{j=1}^{n_i} w_{1ij} h_{1ij}, \tag{2.1}$$

and

$$g_{2i} = \sum_{j=1}^{n_i} w_{2ij} h_{1ij}, \tag{2.2}$$

where h_{1ij} are the elementary functions defined as

$$h_{1ij} = y_{ij} - E_2(y_{ij}|\gamma_i) \tag{2.3}$$

$$= y_{ij} - \phi'(\eta_{ij}),$$

and w_{1ij} and w_{2ij} are given by

$$w_{1ij} = \frac{E_2(\partial h_{1ij}/\partial\beta|\beta,\gamma_i)}{E_2(h_{1ij}^2|\beta,\gamma_i)} = \frac{-\phi''(\eta_{ij})x_{ij}}{\phi''(\eta_{ij})} = -x_{ij}, \tag{2.4}$$

and

$$w_{2ij} = \frac{E_2(\partial h_{1ij}/\partial\gamma_i|\beta,\gamma_i)}{E_2(h_{1ij}^2|\beta,\gamma_i)} = \frac{-\phi''(\eta_{ij})}{\phi''(\eta_{ij})} = -1, \tag{2.5}$$

respectively. Note that in the equations (2.3) - (2.5), E_2 denotes the conditional expectation of y_{ij} for given γ_i. Further note that $E(g_1) = E(g_{2i}) = 0$ for all $i = 1, \dots, k$, because $E(h_{1ij}|\gamma_i) = 0$. That is, g_1 and g_{2i} are unbiased estimating functions. Now joint estimation of the parameters β and γ_i can be achieved by solving estimating equations $g_1 = 0$ and $g_{2i} = 0$ for the observed data (x_{ij}, y_{ij}) $(i = 1, \dots, k;\ j = 1, \dots, n_i)$.

The solutions of $g_1 = 0$ and $g_{2i} = 0$, denoted by $\hat\beta$ and $\hat\gamma_i$ respectively, may be obtained by the customary Newton-Raphson method. To begin, we assume that $\gamma_i = 0$ for $i = 1, \dots, k$. Given the value $\hat\beta(u)$ at the uth iteration, $\hat\beta(u+1)$ is obtained as

$$\hat\beta(u+1) = \hat\beta(u) - [(\partial g_1/\partial\beta)^T]_u^{-1}[g_1]_u, \tag{2.6}$$

where $[\ ]_u$ denotes that the expression within the brackets is evaluated at $\hat\beta(u)$. Next we use this estimate $\hat\beta$ in g_{2i} and solve $g_{2i}(\hat\beta) = 0$ for γ_i. Given the value $\hat\gamma_i(u)$ at the uth iteration, $\hat\gamma_i(u+1)$ is obtained as

$$\hat\gamma_i(u+1) = \hat\gamma_i(u) - [\partial g_{2i}(\hat\beta)/\partial\gamma_i]_u^{-1}[g_{2i}(\hat\beta)]_u, \tag{2.7}$$

where $[\ ]_u$ denotes that the expression within the brackets is evaluated at $\hat\gamma_i(u)$.

Notice that γ_i's are estimated so far by treating them as fixed effects. But, in the present mixed model, they are random by nature. We now propose an adhoc estimator of the random effect γ_i, say $\hat{\hat\gamma}_i$, obtained as

$$\hat{\hat\gamma}_i = \hat E(\gamma_i|\hat\gamma_i) = E(\gamma_i) + \hat E(\gamma_i\hat\gamma_i)\{\hat v(\hat\gamma_i)\}^{-1}(\hat\gamma_i - \hat E(\hat\gamma_i)), \tag{2.8}$$

which is the posterior mean of γ_i given the data through $\hat\gamma_i$, provided γ_i and $\hat\gamma_i$ have jointly bivariate normal distribution. Now, as

$$E(\hat\gamma_i - \gamma_i)^2 = E\hat\gamma_i^2 - 2E\hat\gamma_i\gamma_i + \sigma^2, \tag{2.9}$$

we may estimate $E(\hat\gamma_i\gamma_i)$ by

$$\hat E(\hat\gamma_i\gamma_i) = \frac{1}{2}\left[\sum_{i=1}^{k}\hat\gamma_i^2/k + \hat\sigma^2 - \hat E(\hat\gamma_i - \gamma_i)^2\right], \tag{2.10}$$

where $E(\hat{\gamma}_i - \gamma_i)^2$ may be obtained easily by expanding $g_{2i}(\hat{\gamma}_i)$ about γ_i and noting that $g_{2i}(\hat{\gamma}_i) = 0$. Next by using $\hat{E}(\hat{\gamma}_i) = \sum_{i=1}^{k} \hat{\gamma}_i/k = \bar{\gamma}$ and $\hat{v}(\hat{\gamma}_i) = \sum_{i=1}^{k} [\hat{\gamma}_i - \bar{\gamma}]^2/k$ it follows from (2.8) and (2.10) that

$$\hat{\hat{\gamma}}_i = \frac{k}{2}\left[\sum_{i=1}^{k} \hat{\gamma}_i^2/k + \hat{\sigma}^2 - \hat{E}(\hat{\gamma}_i - \gamma_i)^2\right](\gamma_i - \bar{\gamma})/\sum_{i=1}^{k}(\hat{\gamma}_i - \bar{\gamma})^2, \tag{2.11}$$

where $\hat{\sigma}^2$ is a suitable estimate of σ^2. These values of $\hat{\hat{\gamma}}_i$ may, in turn, be used in (2.6) to improve the estimate of the regression parameters.

Now to obtain the estimate of σ^2, we solve the estimating equation $g_3 = 0$, where

$$g_3 = \sum_{i=1}^{k} w_{3i} h_{2i}, \tag{2.12}$$

is the optimal estimating function for σ^2, with $h_{2i} = \gamma_i^2 - \sigma^2$ as the elementary function, and w_{3i} as the respective weight given by $w_{3i} = E(\partial h_{3i}/\partial\sigma^2)/E(h_{3i}^2) = -1/k_4$, where $k_4 = E(\gamma_i^4) - \sigma^2$. Note that to solve $g_3 = 0$ for σ^2, it is not necessary to know k_4, i.e., $E(\gamma_i^4)$. The solution of $g_3 = 0$ for σ^2 yields $\sigma^2 = \Sigma\gamma_i^2/k$. We, thus, obtain

$$\hat{\sigma}^2 = \sum_{i=1}^{k} \hat{\hat{\gamma}}_i^2/k, \tag{2.13}$$

where $\hat{\hat{\gamma}}_i$ is the prediction of the true random effects, given by (2.11).

Notice that $\hat{\sigma}^2$ in (2.13), $\hat{\hat{\gamma}}_i$ in (2.11) have to be computed iteratively. As mentioned above, we then go back to (2.6) to improve the estimate of β by using $\gamma_i = \hat{\hat{\gamma}}_i$ instead of $\hat{\gamma}_i$. The cycles of iteration continues until convergence is achieved for β and σ^2. Let $\tilde{\beta}$, $\tilde{\gamma}_i$ and $\tilde{\sigma}^2$ be the final estimates.

3 OPTIMAL ESTIMATION WHEN RANDOM EFFECTS ARE TREATED TRULY AS RANDOM EFFECTS

In this approach, optimal estimating functions for the regression parameters are developed under the fact that γ_i's are independently distributed with zero mean and unknown variance σ^2. For the time being, suppose that we can compute the unconditional mean and variance-covariance matrix of the response vector y_i. That is

$$\mu_i(\beta, \sigma^2) = E(y_i) = E_1 E_2(y_i|\gamma_i) \tag{3.1}$$

and

$$V_i(\beta, \sigma^2) = E(y_i - \mu_i(\beta, \sigma))(y_i - \mu_i(\beta, \sigma))^T \tag{3.2}$$

are computable, where E_2 in (3.1) denotes the conditional expectation of y_i for fixed γ_i as in (2.4) and E_1 denotes the expectation over γ_i when they are random. We now consider an elementary estimating function for β as

$$h_{1i}^* = y_i - \mu_i(\beta, \sigma^2) \tag{3.3}$$

and construct an optimal estimating function as

$$g_1^* = \sum_{i=1}^{k} w_{1i}^* h_{1i}^* \tag{3.4}$$

where

$$w_{1i}^* = E(\partial h_{1i}^{*T} / \partial \beta)[E h_{1i}^* h_{1i}^{*T}]^{-1}.$$

For given σ^2, the estimating equation

$$g_1^* = -\sum_{i=1}^{k} \left(\frac{\partial \mu_i(\beta, \sigma^2)}{\partial \beta} \right)^T V_i^{-1}(\beta, \sigma^2)(y_i - \mu_i(\beta, \sigma^2)) = 0 \tag{3.5}$$

is solved for β.

Note that the estimating equation $g_1^* = 0$ in (3.5) is the same as the estimating equation for β considered by Waclawiw and Liang (1993). Further note that in the manner similar to that for the estimation of β, we could construct an optimal estimating function for σ^2, but this will require calculations of higher moments for the response vector, which may not be easy. Consequently, we choose to estimate σ^2 by using the predicted random effects, where the prediction of the random effects is made by exploiting their true randomness nature.

More specifically, the joint computational steps for β, γ_i and σ^2 are as follows. First, for an initial σ^2 value, β estimate at the $(u+1)$th iteration is obtained as

$$\beta^*(u+1) = \beta^*(u) - [(\partial g_1^*(\sigma^2)/\partial \beta)^T]_u^{-1} [g_1^*(\sigma^2)]_u \tag{3.6}$$

where $g_1^*(\sigma^2)$ is the same estimating function as g_1^* in (3.4) except that now σ^2 has a specified value, and $[\ ]_u$ denotes that the expression within the brackets is evaluated at $\beta^*(u)$. Next we use this estimate β^* in (3.8) below to obtain an optimal prediction γ_i^* for γ_i under the assumption that γ_i's are truly random. The estimating function for γ_i is constructed as follows.

To begin, γ_i's are treated as fixed as in the previous section. Next the prior information that γ_i's are independently and identically distributed with zero mean and variance σ^2, is used to take the randomness nature of γ_i into account. The optimal estimating function for γ_i is then written as

$$g_{2i}^* = g_{2i} + g_{3i}, \tag{3.7}$$

where g_{2i} is as in (2.2) and g_{3i} is given by $g_{3i} = w_{3i}h_{3i}$, with $h_{3i} = \gamma_i - E(\gamma_i)$ as the elementary function and $w_{3i} = E(\partial h_{3i}/\partial\gamma_i)/E(h_{3i}^2) = 1/\sigma^2$, as its weight (Godambe, 1994; Naik-Nimbalkar and Rajarshi, 1995). It then follows that for $\beta = \beta^*$ obtained from (3.6) and for known σ^2, the predicted value of γ_i at the $(u+1)$th iteration is obtained as

$$\gamma_i^*(u+1) = \gamma_i^*(u) - [\partial g_{2i}^*(\beta^*, \sigma^*)/\partial\gamma_i]_u^{-1}[g_{2i}^*(\beta^*, \sigma^2)]_u, \tag{3.8}$$

where $[\ ]_u$ denotes that the expression within the brackets is evaluated at $\gamma_i^*(u)$, the uth iteration value of γ_i.

Notice that it has been assumed in (3.8) that σ^2 is known. When σ^2 is unknown, it may be estimated, as in the previous section, by using the optimal estimating function g_3 given in (2.12). The corresponding estimating equation $g_3 = 0$ yields

$$\sigma^{*^2} = \sum_{i=1}^{k} \gamma_i^{*^2}/k, \tag{3.9}$$

where γ_i^* is obtained from (3.8) for $\beta = \beta^*$ and for a given value of σ^2. Now σ^{*^2} is put back into (3.6) and (3.8) to obtain the improved estimates of β and γ_i. The improved estimate of γ_i is then used in (3.9) to obtain an improved estimate of σ^2. The cycle of iteration continues until convergence is achieved for β and σ^2. Let $\tilde{\tilde{\beta}}$, $\tilde{\tilde{\gamma}}_i$ and $\tilde{\tilde{\sigma}}^2$ be the final estimates.

Now turning back to the issue of computations for the mean and covariance matrix of y_i, it is well known that in general, exact expressions for the marginal means and variances may not be easily computable. But for $\gamma_i \overset{\text{iid}}{\sim} N(0, \sigma^2)$, expressions for the marginal means and variances simplify or may be easily approximated for the standard link functions [see Zeger et al (1988)]. For example, for the log link as in the Poisson case,

$$\begin{aligned} E(y_{ij}) &= E_1 E_2(y_{ij}|\gamma_i) \\ &= E_1\{\exp(x_{ij}^T\beta + \gamma_i)\} \\ &= \exp(x_{ij}^T\beta + \frac{1}{2}\sigma^2), \end{aligned} \tag{3.10}$$

and

$$\begin{aligned}\mathrm{var}(y_{ij}) &= \mathrm{var}\{E_2(y_{ij}|\gamma_i)\} + E_1\{\mathrm{var}(y_{ij}|\gamma_i)\} \\ &= \exp(x_{ij}^T\beta + \frac{1}{2}\sigma^2)\{1 + \exp(x_{ij}^T\beta + \frac{1}{2}\sigma^2)(\exp(\sigma^2) - 1)\}.\end{aligned} \tag{3.11}$$

If it is assumed that for given γ_i, Poisson responses y_{ij} and $y_{ij'}$ are independent, for $j \neq j'$, $j, j' = 1, \ldots, n_i$, then unconditional covariance between y_{ij} and $y_{ij'}$ is given by

$$\begin{aligned}\mathrm{cov}(y_{ij}, y_{ij'}) &= \{\exp(x_{ij}^T\beta + \frac{1}{2}\sigma^2)\}\{\exp(x_{ij'}^T\beta + \frac{1}{2}\sigma^2)\} \\ &\quad \times\{\exp(\sigma^2) - 1\}.\end{aligned} \tag{3.12}$$

The mean vector $\mu_i(\beta, \sigma^2)$ and variance-covariance matrix $V_i(\beta, \sigma^2)$ are then easily computed.

Note that for the cases when $\gamma_i \stackrel{\text{iid}}{\sim} N(0, \sigma^2)$ and σ^2 is assumed to be small, one may develop an approximate likelihood function for β and σ^2 and compute the likelihood estimate for these parameters [see Sutradhar and Qu (1997)]. The random effects γ_i ($i = 1, \ldots, k$) may be estimated by using the posterior likelihood of γ_i given the data.

For the general case when γ_i's are independently and identically distributed with zero mean and variance σ^2, one may still obtain the approximate marginal means and variances, provided σ^2 is small (cf. Sutradhar and Rao (1996)]. Rewrite the conditional density (1.1) of y_{ij} given γ_i as

$$f(y_{ij}|\theta_{ij}^*) = \exp\{a(\theta_{ij}^*)y_{ij} - g(\theta_{ij}^*) + c(y_{ij})\}, \tag{3.13}$$

where θ_{ij}^*'s, with $\theta_{ij}^* = x_{ij}^T\beta + \gamma_i$, are independent random variates with

$$E(\theta_{ij}^*) = \theta_{ij} = x_{ij}^T\beta, \text{ and } \mathrm{var}(\theta_{ij}^*) = \sigma^2.$$

Now by expanding $f(y_{ij}|\theta_{ij}^*)$ in (3.13) about θ_{ij} and taking expectation over the distribution of θ_{ij}^*, one first obtains the density function of y_{ij}, which may then be exploited to compute the marginal means and variances. After some algebra it follows that

$$\begin{aligned}E(y_{ij}) &\simeq m_1 + \frac{\sigma^2}{2}[(a')^2\{m_3 + m_1m_2\} + a''m_2 \\ &\quad -m_1\{g'' - m_1a''\},\end{aligned} \tag{3.14}$$

with $m_1 = g'/a'$, $m_2 = [g'' - a''m_1]/(a')^2$, $m_3 = -(a')^{-3}[3a'a''m_2 + a'''m_1 - g''']$, where, for example, m_1, a' and g' are used for the functions $m_1(\theta_{ij})$, $a'(\theta_{ij})$ and $g'(\theta_{ij})$ respectively, by suppressing their dependence on θ_{ij}. By similar calculations, one obtains

$$\mathrm{var}(y_{ij}) = M_2(\theta_{ij}, \sigma^2) - \{E(y_{ij})\}^2, \tag{3.15}$$

where

$$\begin{aligned} M_2(\theta_{ij}, \sigma^2) &= (m_2 + m_1^2) + \frac{\sigma^2}{2}[(a')^2\{m_4 + 2m_1m_3 + m_1^2m_2\} \\ &\quad + a''\{m_3 + 3m_1m_2 + m_1^3\} - g''(m_2 + m_1^2)], \end{aligned}$$

with

$$\begin{aligned} m_4 &= -(a')^{-4}[6(a')^2\{a''m_3m_3 + (a''m_1 - g'')m_2\} \\ &\quad + 3\{(a'')^2 + (a''m_1 - g'')^2\} \\ &\quad + 4a'a'''m_2 + a^{IV}m_1 - g^{IV}]. \end{aligned}$$

Further, as in this general case we do not specify the joint distribution of $y_{i1}, \ldots, y_{ij}, \ldots, y_{in_i}$, one may use the 'working' covariance matrix

$$\Sigma_i(\beta, \sigma^2) = D_i^{\frac{1}{2}} R(\alpha) D_i^{\frac{1}{2}}, \tag{3.16}$$

in place of the true covariance matrix $V_i(\beta, \sigma^2)$ without sacrificing the consistency of $\tilde{\tilde{\beta}}$ through the generalized estimating equation approach [Liang and Zeger (1986)]. In (3.16), $R(\alpha)$ is referred to as a 'working' correlation matrix of y_i, and $D_i = \text{diag}[\text{var}(y_{i1}) \ldots, \text{var}(y_{ij}) \ldots, \text{var}(y_{in_i})]$.

4 SIMULATION STUDY

To examine the performance of the proposed approaches, we executed a small simulation study under the Poisson mixed model, with

$$\log\{E(y_{ij}|\gamma_i)\} = \sum_{\ell=1}^{p} \beta_\ell x_{ij\ell} + \gamma_i, \tag{4.1}$$

for $j = 1, \ldots, n_i$; $i = 1, \ldots, k$. The parameters to be controlled in the simulation study are as follows: (a) k, the number of independent clusters or individuals; (b) n_i, the number of observations under each cluster or individual; (c) $\beta_1, \ldots, \beta_p$, the regression effects of the p covariates; (d) γ_i ($i = 1, \ldots, k$), the random effects; and (e) σ^2, the variance of the random effects.

We take the number of clusters as $k = 100$, and consider two sets of values of n_i and p, namely, $n_i = 4$, $p = 4$; and $n_i = 10$, $p = 2$, for all $i = 1, \ldots, k$. For the first set of values of n_i and p, we take

$$\beta_1 = 2.5, \quad \beta_2 = -1.0, \quad \beta_3 = 1.0 \quad \text{and} \quad \beta_4 = 0.5;$$

and

$$x_{ij1} = 1, \quad \text{for } j = 1, \ldots, n_i; \quad x_{ij2} = \begin{cases} 1, & \text{for} \quad j = 1, \ldots, \frac{n_i}{2} \\ 0, & \text{for} \quad j = \frac{n_i}{2} + 1, \ldots, n_i \end{cases}$$

$$x_{ij3} = j - \frac{n_i + 1}{2}; \quad j = 1, \ldots, n_i; \quad \text{and } x_{ij4} = x_{ij2} x_{ij3}.$$

For the second set of values of n_i and p, we take the first two values of β's, i.e., $\beta_1 = 2.5$ and $\beta_2 = -10$; and first two covariates x_{ij1} and x_{ij2}. For $k = 100$, the γ_i's were independently generated from a normal distribution with mean 0 and variance σ^2. Five values of $\sigma^2 = 0.1$, 0.3, 0.50, 0.75 and 1.00 were considered. The responses $(y_{i1}, \ldots, y_{in_i})$ for each cluster i were generated as realizations of Poisson model (1.1) with mean and variance equal to $\exp\left\{\sum_{\ell=1}^{p} \beta_\ell x_{ij\ell} + \gamma_i\right\}$. The simulated data (y_{ij}), $j = 1, \ldots, n_i$; $i = 1, \ldots k$, and the covariates (x_{iju}), $u = 1, \ldots, p$; $j = 1, \ldots, n_i$; $i = 1, \ldots k$ were used to compute the estimates of the fixed effect parameters β, variance component σ^2 of the random effects, and the random effects γ_i $(i = 1, \ldots, k)$, based on both approaches discussed in Sections 2 and 3. The simulation was repeated 2,000 times in order to obtain the mean value and standard errors of the parameter estimates.

For simplicity, we refer to the estimation method discussed in Section 2 as corrected conditional mixed method (CCMM) and the estimation method discussed in Section 3 as unconditional mixed method (UMM). We now spell out the formulas for the estimation of the parameters of the Poisson mixed model by CCMM and UMM.

In CCMM, to begin with, the random effects γ_i are treated as fixed effects and estimating equations for β and γ_i are developed by conditioning on these fixed effects. Folowing (2.6) and (2.7), these estimating equations are

$$\hat{\beta}(u+1) = \hat{\beta}(u) + \left[\sum_{i=1}^{k}\sum_{j=1}^{n_i} \exp(x_{ij}^T\beta + \gamma_i) x_{ij} x_{ij}^T\right]_u^{-1} \left[\sum_{i=1}^{k}\sum_{j=1}^{n_i} \left\{(y_{ij} - \exp(x_{ij}^T\beta + \gamma_i)\right\} x_{ij}\right]_u, \tag{4.2}$$

and

$$\hat{\gamma}_i(u+1) = \hat{\gamma}_i(u) + \left[\sum_{j=1}^{n_i} \exp(x_{ij}^T\beta + \gamma_i)\right]_u^{-1} \left[\sum_{j=1}^{n_i} \{y_{ij} - \exp(x_{ij}^T\beta + \gamma_i)\}\right]_u, \tag{4.3}$$

where $[\ ]_u$ in (4.2) and (4.3) denotes that the expression within the brackets are evaluated at uth iterated value $\hat{\beta}(u)$ and $\hat{\gamma}_i(u)$ respectively. Note that

$\hat{\gamma}_i$ in (4.3) are computed by treating γ_i $(i = 1, \ldots, k)$ as fixed, although in reality, under the present model, they are random. We now make a correction to take this random nature of γ_i into account and estimate them by (2.11) by noting that for the Poisson model

$$\hat{E}(\hat{\gamma}_i - \gamma_i)^2 = \left[\sum_{j=1}^{n_i}\{y_{ij} - \exp(x_{ij}^T\beta + \hat{\gamma}_i)\}\right]^{-1}. \tag{4.4}$$

This result in (4.4) is obtained by expanding

$$g_{2i}(\hat{\gamma}_i) = \sum_{j=1}^{n_i}\{y_{ij} - \exp(x_{ij}^T\beta + \hat{\gamma}_i)\}$$

about γ_i and noting that $g_{21}(\hat{\gamma}_i) = 0$. Now by using (4.4) in (2.11), and exploiting (4.2), (4.3), (2.11) and (2.13) iteratively, we obtain the solutions for β, γ_i and σ^2. They are referred to as $\tilde{\beta}$, $\tilde{\gamma}_i$ and $\tilde{\sigma}^2$ respectively.

In UMM, regression effects β are estimated by using (3.6). For the Poisson model, equation (3.6) reduces to

$$\beta^*(u+1) = \beta^*(u) - g_{1\beta}^{*^{-1}} g_1^*, \tag{4.5}$$

where

$$g_1^* = \sum_{i=1}^{k}\left[\sum_{j=1}^{n_i} y_{ij}x_{ij} - \frac{\alpha + \sum_{j=1}^{n_i} y_{ij}}{\lambda + \sum_{j=1}^{n_i}\exp(x_{ij}^T\beta)}\left\{\sum_{j=1}^{n_i}\exp(x_{ij}^T\beta)x_{ij}\right\}\right],$$

and

$$g_{1\beta}^* = \sum_{i=1}^{k}\left[\left(\alpha + \sum_{j=1}^{n_i} y_{ij}\right)\left\{\frac{\sum_{j=1}^{n_i}\exp(x_{ij}^T\beta)x_{ij}\sum_{j=1}^{n_i}\exp(x_{ij}^T\beta)x_{ij}^T}{\left(\lambda + \sum_{j=1}^{n_i}\exp x_{ij}^T\beta\right)^2}\right.\right.$$

$$\left.\left. - \frac{\sum_{j=1}^{n_i}\exp(x_{ij}^T\beta)x_{ij}x_{ij}^T}{(\lambda + \sum_{j=1}^{n_i}\exp(x_{ij}^T\beta))}\right\}\right],$$

with $\alpha = 1/\{\exp(\sigma^2) - 1\}$, and $\lambda = 1/\{\exp(\sigma^2)(\exp(\sigma^2) - 1)\}$. Next, by (3.8), the iterative equations for λ_i $(i = 1, \ldots, k)$ for the Poisson model, reduces to

$$\gamma_i^*(u+1) = \gamma_i^*(u) + \left[\sum_{j=1}^{n_i}\exp(x_{ij}^T\beta + \gamma_i) + 1/\sigma^2\right]_u^{-1}$$

$$\times\left[\sum_{j=1}^{n_i}\{y_{ij} - \exp(x_{ij}^T\beta + \gamma_i)\} - \gamma_i/\sigma^2\right]_u, \tag{4.6}$$

where $[\ ]_u$ denotes that the expression within the brackets are evaluated at uth iterative value $\gamma_i^*(u)$. Further, by (3.9), the estimating equation for σ^2 is given by

$$\sigma^{*^2} = \sum_{i=1}^{k} \gamma_i^{*^2}/k, \tag{4.7}$$

which is similar to the estimating equation (2.13) for σ^2 in CCMM.

The above three estimating equations (4.5), (4.6) and (4.7) are solved iteratively as follows. For a given value of σ^2, we first solve (4.5) for β. This estimate of β is then used in (4.6) to obtain the estimate of γ_i. In the third step, these values of γ_i $(i = 1, \dots, k)$ are used in (4.7) to obtain an estimate of σ^2. We then put back this estimate of σ^2 in (4.5) and (4.6) to improve the estimates of β and γ_i. This cycle of iteration continues until convergence is achieved for β and σ^2. The final estimates are denoted by $\tilde{\tilde{\beta}}$, $\tilde{\tilde{\gamma}}_i$ $(i = 1, \dots, k)$ and $\tilde{\tilde{\sigma}}^2$.

In the simulation study, we also include the results based on the conditional mixed method (CMM). In this method, unlike the CCMM, β and γ_i are estimated by treating the random effects γ_i as fixed effects, although in reality they are not so. The appropriate estimating equations are (2.6) and (2.7) for β and γ_i respectively. Next, the variance of the random components is estimated by treating γ_i as random effects, but using the estimate of the fixed γ_i for the unobservable random γ_i. The appropriate equation for the variance component is

$$\hat{\sigma}^2 = \sum_{i=1}^{k} \hat{\gamma}_i^2/k, \tag{4.8}$$

instead of (2.13). In (4.8), $\hat{\gamma}_i$ is obtained from (2.7) instead of (2.11).

4.1 Estimate of β

Table 1 reports the simulated values of the percent relative bias (RB%) of the regression estimators computed by: (1) the conditional mixed method (CMM), (2) the corrected conditional mixed method (CCMM), and (3) the unconditional mixed method (UMM). The percent relative bias (RB%) of the estimator $\tilde{\beta}_1$, for example, is given by $100 \times RB(\tilde{\beta}_1)$, where

$$RB(\tilde{\beta}_1) = |\bar{E}(\tilde{\beta}_1) - \beta_1|/\bar{\sigma}(\tilde{\beta}_1),$$

with $\bar{E}(\tilde{\beta}_1)$ and $\bar{\sigma}(\tilde{\beta}_1)$ as the simulated mean and standard errors of the estimator $\tilde{\beta}_1$. It is clear from the table that in estimating all the regression parameters β_1, β_2, β_3 and β_4 for $p = 4$, β_1 and β_2 for p=2, the UMM leads to very large reduction in RB relative to the conditional methods CMM and CCMM. Between the two conditional methods, the corrected method

(CCMM) yields slightly better estimates for the regression parameters, as compared to the uncorrected conditional method (CMM). For large values of σ^2, both conditional and unconditional methods may have convergence problems. For example, for $n_i = 4$, $p = 4$, CMM and CCMM do not converge when $\sigma^2 = .75$ and 1.00. The convergence problems are shown by putting '*' in the tables against the parameter values for which convergence are not achieved. Similarly, for $n_i = 10$, $p = 2$, the UMM fails to yield the estimates of β_1 and β_2 when $\sigma^2 = 1.00$.

Table 1. Comparison of percent relative bias (RB%) of the regression estimates for selected values of σ^2; $k = 100$; $n_i = 4$ $(i = 1, \ldots, k)$, $p = 4$; $n_i = 10$ $(i = 1, \ldots, k)$, $p = 2$; true values of the regression parameters: $\beta_1 = 2.5$, $\beta_2 = -1.0$, $\beta_3 = 1.0$ and $\beta_4 = 0.5$; 2,000 simulations.

Cluster				Parameter			
Size (n_i)	p	σ^2	Method	β_1	β_2	β_3	β_4
4	4	0.10	CMM	2432	353.7	433.3	285.3
			CCMM	2380	353.7	422.2	282.4
			UMM	54.3	1.7	0	2.7
		0.30	CMM	2368	443.8	650.0	321.9
			CCMM	2305	428.6	650.0	318.8
			UMM	100	1.8	5.0	8.3
		0.50	CMM	1848	540.5	837.5	375.0
			CCMM	1819	538.1	837.5	362.1
			UMM	148.5	3.7	0	3.0
		0.75	CMM	*	*	*	*
			CCMM	*	*	*	*
			UMM	200.0	1.9	0	1.0
		1.00	CMM	*	*	*	*
			CCMM	*	*	*	*
			UMM	259.4	4.2	5.9	4.4
10	2	0.10	CMM	5463	705.3		
			CCMM	4950	636.4		
			UMM	123.1	0		
		0.30	CMM	6513	668.4		
			CCMM	6275	688.9		
			UMM	235.7	0		
		0.50	CMM	4360	615.8		
			CCMM	605.8	221.7		
			UMM	320.0	4.5		
		0.75	CCM	5467	623.5		
			CCMM	1615	536.8		
			UMM	364.7	0		
		1.00	CMM	2422	564.7		
			CCMM	2300	564.7		
			UMM	*	*		

4.2 Prediction of the Random Effects

The results of Table 2 show the performance of the predictors of γ_i ($i = 1, \ldots, k$) by all three methods CMM, CCMM and UMM. Here we examine the empirical distribution form of the estimates of γ_i ($i = 1, \ldots, k$), given that γ_i's are generated from the normal distribution with mean 0 and variance σ^2. This is done by studying the empirical mean, median, skewness and kurtosis of the predictors of γ_i. Table 2, in its last column, also reports the simulated total mean square error (TMSE) of the random effect predictors based on the CMM, CCMM, and UMM. Let $\hat{\gamma}_{is}$ be the CMM estimator of the random effect γ_i in the sth ($s = 1, \ldots, 2000$) simulation. Then the TMSE of the CMM predictors is defined by $\sum_{i=1}^{k} \left\{ \sum_{s=1}^{2000} (\hat{\gamma}_{is} - \gamma_i)^2/2000 \right\}$, where $k = 100$ is the number of independent clusters. Similarly, the TMSE of the CCMM and UMM predictors are defined by $\sum_{i=1}^{k} \left\{ \sum_{s=1}^{2000} (\tilde{\gamma}_{is} - \gamma_i)^2/2000 \right\}$ and $\sum_{i=1}^{k} \left\{ \sum_{s=1}^{2000} (\tilde{\tilde{\gamma}}_{is} - \gamma_i)^2/2000 \right\}$ respectively. It is interesting to note from the table that the corrected conditional method (CCMM) performs extremely well in the prediction of the random effects as compared to the unconditional method UMM. Between the two conditional mixed methods, as expected, the corrected conditional mixed method performs much better as compared to the uncorrected conditional mixed method. This leads to the fact that correction for randomness of γ_i ($i = 1, \ldots, k$) is quite important as the true γ_i's are random. When the corrected approach is compared with the uncorrected approach for checking for the normality of the predictors of the random effects, both approaches appear to yield the normal predictors. But, the mean value of the predictors by CMM is quite away from the mean value zero for γ_i ($i = 1, \ldots, k$), whereas CCMM yields zero mean value similar to that of the distribution of γ_i ($i = 1, \ldots, k$). Next,

Table 2. Comparison of Mean, Median, Skewness, Kurtosis and total mean square errors (TMSE) of the random effect predictions for selected values of σ^2; $k = 100$; $n_i = 4$ $(i = 1, \ldots, k)$, $p = 4$; $n_i = 10$ $(i = 1, \ldots, k)$, $p = 2$; true values of the regression parameters: $\beta_1 = 2.5$, $\beta_2 = -1.0$, $\beta_3 = 1.0$ and $\beta_4 = 0.5$; 2,000 simulations.

Cluster Size	p	σ^2	Method	Mean	Median	Skewness	Kurtosis	TMSE
4	4	0.10	CMM	0.473	0.445	0.320	2.655	23.327
			CCMM	0.000	-0.028	0.332	2.660	0.178
			UMM	0.000	0.004	0.184	2.284	0.813
		0.30	CMM	0.337	0.288	0.321	2.654	12.565
			CCMM	0.000	-0.049	0.324	2.654	0.211
			UMM	0.000	0.007	0.153	2.264	1.050
		0.50	CMM	0.153	0.091	0.324	2.651	3.252
			CCMM	0.000	-0.064	0.325	2.651	0.244
			UMM	0.002	0.014	0.142	2.252	1.293
		0.75	CMM	*	*	*	*	*
			CCMM	*	*	*	*	*
			UMM	0.006	0.029	0.135	2.246	1.639
		1.00	CMM	*	*	*	*	*
			CCMM	*	*	*	*	*
			UMM	0.014	0.038	0.135	2.256	2.046
10	2	0.10	CMM	0.536	0.507	0.342	2.654	30.817
			CCMM	0.000	0.014	-0.089	2.833	1.106
			UMM	0.000	-0.002	0.238	2.324	1.160
		0.30	CMM	0.460	0.408	0.362	2.658	23.793
			CCMM	0.000	-0.056	0.429	2.661	1.315
			UUM	0.000	0.005	0.183	2.264	1.496
		0.50	CMM	0.380	0.314	0.363	2.651	17.318
			CCMM	0.000	-0.069	0.386	2.658	1.462
			UMM	0.001	0.014	0.164	2.252	1.771
		0.75	CMM	0.274	0.194	0.366	2.649	10.455
			CCMM	0.000	-0.081	0.372	2.650	1.669
			UMM	0.005	0.015	0.159	2.256	2.170
		1.00	CCM	0.168	0.074	0.371	2.642	5.548
			CCMM	0.000	-0.094	0.373	2.642	1.845
			UMM	*	*	*	*	*

when the unconditional approach UMM is compared with the corrected conditional mixed method (CCMM) for normality, the UMM appears to produce much better mean, median and skewness values for the predictors of the random effects than those yielded by the CCMM approach. The CCMM, however, appears to yield much better kurtosis value (close to 3) as compared to the unconditional mixed method (UMM).

4.3 Estimate of σ^2

Table 3 reports the simulated mean values and standard errors of the estimates of σ^2. It is clear from the table that the unconditional mixed method and the corrected conditional mixed method compete each other in estimating the variance σ^2 of the random effects. The uncorrected conditional mixed method performs worse, as expected, as compared to its counterpart CCMM. This is because, unlike the CCMM, the CMM treats the random effects as fixed effects and use them to estimate the variance component of the random effects. Between the CCMM and UMM, σ^2 estimates of the UMM always have the smaller bias but larger standard errors than the estimates of the corrected conditional mixed method (CCMM).

Table 3. Comparison of simulated mean values and standard error (SE) of the estimates of variance components of random effects for selected values of σ^2; $k = 100$; $n_i = 4$ $(i = 1, \ldots, k)$, $p = 4$; $n_i = 10$ $(i = 1, \ldots, k)$, $p = 2$; true values of the regression parameters: $\beta_1 = 2.5$, $\beta_2 = -1.0$, $\beta_3 = 1.0$ and $\beta_4 = 0.5$; 2,000 simulations.

Cluster				σ^2				
Size	p	Method		0.10	0.30	0.50	0.75	1.00
4	4	CMM	Mean	0.329	0.427	0.545	*	*
			SE	0.006	0.007	0.007	*	*
		CCMM	Mean	0.104	0.312	0.520	*	*
			SE	0.003	0.005	0.007	*	*
		UMM	Mean	0.095	0.304	0.511	0.772	1.032
			SE	0.006	0.011	0.015	0.022	0.026
10	2	CMM	Mean	0.401	0.530	0.667	0.853	1.059
			SE	0.004	0.010	0.013	0.021	0.027
		CCMM	Mean	0.112	0.306	0.512	0.774	1.029
			SE	0.007	0.011	0.018	0.026	0.030
		UMM	Mean	0.090	0.298	0.506	0.763	*
			SE	0.008	0.014	0.020	0.026	*

5 SUMMARY AND DISCUSSION

Our limited simulation study has shown for the Poisson mixed model that the unconditional mixed method (UMM) is superior to the corrected conditional mixed method (CCMM) in estimating the fixed effects parameters. The CCMM, on the other hand, performs better than the UMM, in predicting the random effects, as the total mean square errors (TMSE) yielded by the CCMM were always found to be smaller than those produced by the UMM. In estimating the variance of the random effects, both UMM and CCMM were found to be almost the same. The CMM always performs poorly in estimating any parameters β or γ_i ($i = 1, \ldots, k$) or σ^2. This uncorrected CMM, therefore, should not be used in estimating the parameters of the mixed model.

Note that it has been assumed in the simulation study that $\gamma_i \overset{\text{iid}}{\sim} N(0, \sigma^2)$. But, in general, γ_i ($i = 1, \ldots, k$) may also follow non-normal distributions. In such cases, the performance of the adhoc estimates $E(\gamma_i|\tilde{\gamma}_i)$ or $E(\gamma_i|\tilde{\tilde{\gamma}}_i)$, computed by pretending that $\gamma_i \overset{\text{iid}}{\sim} N(0, \sigma^2)$, may not be satisfactory. This shows the necessity for further investigation for the construction of a robust corrected predictor for γ_i ($i = 1, \ldots, k$), irrespective of the distribution of γ_i. This problem, however, does not arise in the unconditional mixed method (UMM). But again, it does not mean that the UMM is problem free. This is because, for large σ^2, it may be extremely difficult to compute the unconditional marginal mean and variance of the response variable, which are required for the construction of the estimating equations for the regression parameters.

Further note that when the elementary functions, such as h_{1ij} in (2.1) and h_{2i} in (2.12) are conditionally correlated, Durairajan (1992) has given a closed form of optimal estimating function under certain conditions and utilizing this, Bai and Durairajan (1996) obtained optimal estimating function for means and variances of one-type and two-type branching processes. The result of Durairajan (1992) may be used in the generalized linear mixed model also. However, we have not considered this approach in this paper.

Acknowledgements

The research was partially supported by grants from the Natural Sciences and Engineering Research Council of Canada. The authors would like to thank a referee for constructive comments.

References

Bai, K. and Durairajan, T. M. (1996). Estimating functions for branching processes. *J. Statist. Plann. Inference*, 53, 21-23.

Breslow, N. E. (1984). Extra-Poisson variation in log-linear models. *Applied Statistics*, 33, 38-44.

Durairajan, T. M. (1992). Optimal estimating function for non-orthogonal model. *J. Statist. Plann. Inference*, 33, 381-384.

Ferreira, P. E. (1982). Estimating equations in the presence of prior knowledge. *Biometrika*, 69, 667-669.

Godambe, V. P. (1960). An optimum property of regular maximum likelihood estimation. *Annals of Mathematical Statistics*, 31, 1208-1211.

Godambe, V. P. (1994). Linear Bayas and optimal estimation. Tech. Report STAT-94-11, University of Waterloo.

Godambe, V. P. and Kale, B. K. (1991). Estimating functions: An overview. *Estimating Functions*, (V. P. Godambe, ed.), Oxford University Press, New York, 47-63.

Godambe, V. P. and Thompson, M. E. (1989). An extension of quasi-likelihood estimation (with discussion). *Journal of Statistical Planning and Inference*, 22, 137-72.

Laird, N. M. and Ware, J. H. (1982). Random effects models for longitudinal data. *Biometrics*, 38, 963-974.

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

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

Naik-Nimbalkar, U. V. and Rajarshi, M. D. (1995). Filtering and smoothing via estimating functions. *Journal of the American Statistical Association*, 90, 301-306.

Stiratelli, R., Laird, N. M. and Ware, J. H. (1984). Random effects models for serial observations with binary responses. *Biometrics*, 40, 961-971.

Sutradhar, B. C. and Qu, Z. (1997). On approximate likelihood inference in Poisson mixed model. *Canadian Journal of Statistics*, to appear.

Sutradhar, B. C. and Rao, R. P. (1996). On joint estimation of regression and overdispersion parameters in generalized linear models for longitudinal data. *Journal of Multivariate Analysis*, Vol. 56, No. 1, 90-119.

Waclawiw, M. A. and Liang, K.-Y. (1993). Prediction of random effects; a Gibbs sampling approach. *Journal of the American Statistical Association*, 88, 171-178.

Williams, D. A. (1982). Extra-binomial variation in logistic linear models. *Applied Statistics*, 31, 144-148.

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