The Canadian Journal of Statistics
Vol. 29, No. 2, 2001, Pages ???-???
La revue canadienne de statistique

On the application of extended quasi-likelihood to the clustered data case

Daniel B. HALL

Key words and phrases: Egee; estimating equations; generalized linear model; longitudinal, repeated measures.
AMS 1991 subject classifications: Primary 62J12; secondary 62H99.

ABSTRACT

The author describes the relationship between the extended generalized estimating equations (EGEEs) of Hall & Severini (1998) and various similar methods. He proposes a true extended quasi-likelihood approach for the clustered data case and explores restricted maximum likelihood-like versions of the EGEE and extended quasi-likelihood estimating equations. He also presents simulation results comparing the various estimators in terms of mean squared error of estimation based on three moderate sample size, discrete data situations.

RÉSUMÉ

L'auteur décrit la relation entre les équations d'estimation généralisées étendues (EEGE) de Hall & Severini (1998) et plusieurs méthodes similaires. Il propose une véritable approche de quasi-vraisemblance étendue adaptée au cas de données regroupées et étudie des versions de type maximum de vraisemblance restreint des EEGE et des équations d'estimation de quasi-vraisemblance. Il présente de plus des résultats de simulation comparant les différents estimateurs en terme d'erreur quadratique moyenne pour des échantillons de données discrètes de taille moyenne.

1. INTRODUCTION

Recently, Hall & Severini (1998) proposed estimating equations for clustered data that are motivated as an application of extended quasi-likelihood to the problem of fitting a marginal generalized linear model (GLM) to data that are correlated within each of K independent units, or clusters. This work fits into the generalized estimating equation (GEE) literature that dates to the landmark paper of Liang & Zeger (1996), who proposed an approach to fitting marginal GLMs to longitudinal data that combines an estimating equation for regression (first moment) parameters with method of moment (MOM) estimators for association (second moment) parameters. As originally proposed, GEEs provide consistent, but not necessarily fully efficient, estimators of regression parameters regardless of whether or not the assumed "working" covariance model is correct.

Prentice & Zhao (1991) extended Liang and Zeger's work in two important ways. First, they replaced the MOM approach to estimating second moment parameters with an "ad hoc" estimating equation for these quantities. This generalized Liang and Zeger's approach since their MOM estimators can be recovered as special cases of the ad hoc estimating equations. The generalization of MOM to an estimating equation procedure has two important benefits: increased efficiency, particularly with respect to second moment parameters; and increased flexibility to model the covariances among within-cluster responses. Following the usage of Liang *et al.* (1992), we refer to this approach as GEE1, whose defining characteristic is that "it operates as if [regression and association parameters] are orthogonal to one another even when they are not" (Liang *et al.* 1992, p. 10), ensuring consistency of the regression parameter estimator even when the covariance model is misspecified. As pointed out by Godambe (1992), the use of the term "orthogonal" here does not coincide with the usual meaning of parameter orthogonality in parametric and semiparametric models (Cox & Reid 1987, Godambe 1991). However, what is meant here is simple enough. In GEE1, first and second moment elementary estimating functions are stacked into vector form, and then linearly combined using block diagonal weights, i.e., weights that ignore both the functional and statistical dependence between the first and second moment elementary estimating functions.

The second major contribution of Prentice & Zhao (1991) was to present a unified approach to regression and association parameter estimation that typically leads to greater efficiency. In their "GEE2" estimating equations, the imposed "orthogonality" of the GEE1 approach is dropped, and the functional and statistical dependence between the first and second moment elementary estimating functions is incorporated into their weight matrices. A working structure is assumed for third and fourth-order moments which need not be correct to obtain consistent estimators of regression and association parameters. However, a price is paid in that the consistency of the regression parameter estimator is lost when the covariance model is incorrect.

Like Prentice and Zhao in their GEE2 work, Hall & Severini (1998) attempted to improve on GEE1 by taking a unified approach regression and association parameter to estimation. While Liang and Zeger's original work has a close connection to quasi-likelihood (Wedderburn 1972), and Prentice and Zhao motivated their GEE2 work through the ideas of pseudo maximum likelihood (Gouriéroux *et al.* 1984), Hall and Severini's estimating equations were justified as an application of extended quasi-likelihood (Nelder & Pregibon 1987). In reference to this motivating idea, they call their method extended GEE, or EGEE for short. Through simulations and asymptotic relative efficiency comparisons, they demonstrated that EGEE has efficiency properties comparable to GEE2 and often substantially better than GEE1, depending on what particular second moment estimating function is being used in GEE1. EGEE was also shown to have the consistency properties of GEE1: namely that EGEE regression parameter estimators are consistent under a misspecified covariance structure.

In this paper, we briefly review the aforementioned methods and explore their relationships. In particular, EGEE is seen to be a special case of GEE1. It also amounts to estimation of the correlation parameter by maximizing a Gaussian likelihood function, an approach that Crowder (1992) advocated for clustered data GLMs. In addition, we develop the method of extended quasi-likelihood (EQL) for the clustered data case, and we propose a restricted maximum likelihoood (REML)-like version of EQL that also applies to EGEE. We compare these methods and two GEE2 approaches via a small simulation study.

2. A BRIEF REVIEW OF GEE-BASED METHODS

Let y_{ij} be a scalar response and $\mathbf{x}_{ij}$ a $p \times 1$ vector of covariates for clusters $i = 1, \ldots, K$, and measurements $j = 1, \ldots, n_i$ on cluster i. The response vector for the i-th cluster is $\mathbf{y}_i = (y_{i1}, \ldots, y_{in_i})'$ with mean $\boldsymbol{\mu}_i = (\mu_{i1}, \ldots, \mu_{in_i})'$ and variance-covariance matrix $\text{var}(\mathbf{y}_i) = \mathbf{V}_{i11}$. We assume $\boldsymbol{\mu}_i$ depends upon covariates through the unknown $p-$dimensional regression parameter $\boldsymbol{\beta}$ via a known link function $g(\boldsymbol{\mu}_i) = \mathbf{x}_i\boldsymbol{\beta} \equiv \boldsymbol{\eta}_i$. The marginal variance is related to the marginal mean through a known variance function, v, via $\text{var}(y_{ij}) = \phi v(\mu_{ij})$, where ϕ is an unknown dispersion parameter. We model the marginal covariance within $\mathbf{y}_i$ with an $s-$dimensional parameter $\boldsymbol{\alpha}$. This covariance model combined with the variance-mean relationship implies that $\text{var}(\mathbf{y}_i)$ depends on $\boldsymbol{\beta}$, $\boldsymbol{\alpha}$ and ϕ, so we write $\mathbf{V}_{i11}\{\boldsymbol{\mu}_i(\boldsymbol{\beta}), \tilde{\boldsymbol{\alpha}}\}$, where $\tilde{\boldsymbol{\alpha}} = (\boldsymbol{\alpha}', \phi)'$. Here we have combined the second moment parameters $\boldsymbol{\alpha}$ and ϕ for notational convenience and because the distinction is unnecessary in the development of the estimating equations presented below.

2.1 GEE1.

Based on the marginal GLM described above, a quasi-score function for $\boldsymbol{\beta}$ based on all of the data is

$$\sum_{i=1}^{K} \mathbf{D}_{i11}' \mathbf{V}_{i11}^{-1}\{\boldsymbol{\mu}_i(\boldsymbol{\beta}), \tilde{\boldsymbol{\alpha}}\}(\mathbf{y}_i - \boldsymbol{\mu}_i), \tag{1}$$

where $\mathbf{D}_{i11} = \partial\boldsymbol{\mu}_i/\partial\boldsymbol{\beta}$. Since (1) involves the typically unknown parameter $\tilde{\boldsymbol{\alpha}}$, it is not a suitable estimating function for $\boldsymbol{\beta}$. Therefore, Liang & Zeger (1986) proposed their GEE for $\boldsymbol{\beta}$ as equation (1) with $\tilde{\boldsymbol{\alpha}}$ replaced by an estimator, $\hat{\tilde{\boldsymbol{\alpha}}}(\boldsymbol{\beta})$, that is $\sqrt{K}$-consistent given $\boldsymbol{\beta}$.

In Liang and Zeger's original paper, the marginal covariance model was specified through the marginal correlation matrix $\mathbf{R}(\boldsymbol{\alpha})$, making use of the decomposition

$$\mathbf{V}_{i11}(\boldsymbol{\mu}_i, \tilde{\boldsymbol{\alpha}}) = \phi \mathbf{A}_i^{1/2}(\boldsymbol{\mu}_i)\mathbf{R}(\boldsymbol{\alpha})\mathbf{A}_i^{1/2}(\boldsymbol{\mu}_i), \tag{2}$$

where $\mathbf{A}_i(\boldsymbol{\mu}_i) = \text{diag}\{v(\mu_{i1}), \ldots, v(\mu_{in_i})\}$. For the estimation of the nuisance parameters ϕ and $\boldsymbol{\alpha}$, they considered several possible choices for $\mathbf{R}(\boldsymbol{\alpha})$, and proposed method of moment estimators for these parameters on a case-by-case basis.

Prentice & Zhao (1991) generalized Liang and Zeger's MOM approach by suggesting *ad hoc* estimating equations for $\tilde{\boldsymbol{\alpha}}$ similar in form to (1). The term GEE1 refers to either of these approaches. Let $\mathbf{s}_i = \text{vech}\{(\mathbf{y}_i - \boldsymbol{\mu}_i)(\mathbf{y}_i - \boldsymbol{\mu}_i)'\}$ and $\boldsymbol{\sigma}_i = E(\mathbf{s}_i) = \text{vech}(\mathbf{V}_{i11})$, where vech is the vector-half function that stacks the columns of its argument including only those elements on or below the diagonal (cf., e.g., Searle 1982). Then the first and second moment GEE1 estimating equations are given by

$$\begin{aligned} \sum_{i=1}^{K} \mathbf{D}_{i11}' \mathbf{V}_{i11}^{-1}(\mathbf{y}_i - \boldsymbol{\mu}_i) &= \mathbf{0}, &(3)\\ \sum_{i=1}^{K} \mathbf{D}_{i22}' \mathbf{V}_{i22}^{-1}(\mathbf{s}_i - \boldsymbol{\sigma}_i) &= \mathbf{0}, &(4) \end{aligned}$$

where $\mathbf{D}_{i22} = \partial\boldsymbol{\sigma}_i/\partial\tilde{\boldsymbol{\alpha}}'$ and $\mathbf{V}_{i22}$ is an assumed form for $\text{cov}(\mathbf{s}_k)$, or more generally, a weight matrix. If we let

$$\mathbf{D}_i = \begin{pmatrix} \partial\boldsymbol{\mu}_i/\partial\boldsymbol{\beta}' & \mathbf{0} \\ \mathbf{0} & \partial\boldsymbol{\sigma}_i/\partial\tilde{\boldsymbol{\alpha}}' \end{pmatrix} = \begin{pmatrix} \mathbf{D}_{i11} & \mathbf{0} \\ \mathbf{0} & \mathbf{D}_{i22} \end{pmatrix},$$

$$\mathbf{V}_i = \begin{pmatrix} \mathbf{V}_{i11} & \mathbf{0} \\ \mathbf{0} & \mathbf{V}_{i22} \end{pmatrix}, \quad \mathbf{f}_i = \begin{pmatrix} \mathbf{y}_i - \boldsymbol{\mu}_i \\ \mathbf{s}_i - \boldsymbol{\sigma}_i \end{pmatrix},$$

then (3) and (4) can be combined into a single estimating equation, GEE1:

$$\sum_{i=1}^{K} \mathbf{D}_i' \mathbf{V}_i^{-1} \mathbf{f}_i = \mathbf{0}. \tag{5}$$

Let $\boldsymbol{\gamma}$ denote the combined first and second moment parameter vector $\boldsymbol{\gamma} = (\boldsymbol{\beta}', \tilde{\boldsymbol{\alpha}}')'$, and let $\hat{\gamma}_{\mathrm{G1}} = (\hat{\boldsymbol{\beta}}_{\mathrm{G1}}', \hat{\tilde{\boldsymbol{\alpha}}}_{\mathrm{G1}}')'$ be a solution to (5). Standard estimating function asymptotics can be used to show that $\sqrt{K}(\hat{\boldsymbol{\gamma}}_{\mathrm{G1}} - \boldsymbol{\gamma})$ has an asymptotic normal distribution with mean zero and covariance matrix

$$\lim_{K\to\infty} K\Sigma^{-1} \left\{ \sum_{i=1}^{K} \mathbf{D}_i' \mathbf{V}_i^{-1} \mathrm{cov}\begin{pmatrix} \mathbf{y}_i \\ \mathbf{s}_i \end{pmatrix} \mathbf{V}_i^{-1} \mathbf{D}_i \right\} \Sigma^{-1\prime}, \tag{6}$$

where

$$\Sigma = \sum_{i=1}^{K} \mathbf{D}_i' \mathbf{V}_i^{-1} E\left(\frac{\partial \mathbf{f}_i}{\partial \boldsymbol{\gamma}'}\right) = \sum_{i=1}^{K} \begin{pmatrix} \mathbf{D}_{i11}' \mathbf{V}_{i11}^{-1} \mathbf{D}_{i11} & \mathbf{0} \\ \mathbf{D}_{i22}' \mathbf{V}_{i22}^{-1} \frac{\partial \boldsymbol{\sigma}_i}{\partial \boldsymbol{\beta}'} & \mathbf{D}_{i22}' \mathbf{V}_{i22}^{-1} \mathbf{D}_{i22} \end{pmatrix}.$$

Note that this expression corrects the formula given by Prentice & Zhao (1991). A robust covariance estimator may be obtained by evaluating all quantities in (6) at $\hat{\boldsymbol{\gamma}}_{\mathrm{G1}}$ and substituting $\mathrm{cov}\binom{\mathbf{y}_i}{\mathbf{s}_i} = \mathbf{f}_i \mathbf{f}_i'$.

The characteristic feature of GEE1 is that both $\mathbf{D}_i$ and $\mathbf{V}_i$ are block-diagonal. This ensures consistency of $\hat{\boldsymbol{\beta}}_{\mathrm{G1}}$ in the presence of a misspecified model for $\mathrm{var}(\mathbf{y}_i)$ provided that $\hat{\tilde{\boldsymbol{\alpha}}}_{\mathrm{G1}}$ tends stochastically to some limit as $K \to \infty$ (Crowder 1995).

2.2 GEE2.

Suppose that $\mathbf{y}_i$ arises according to a quadratic exponential model of the form

$$P(\mathbf{y}_i; \boldsymbol{\mu}_i, \boldsymbol{\sigma}_i) = \Delta_i^{-1} \exp\{\mathbf{y}_i' \boldsymbol{\theta}_i + \mathbf{w}_i' \boldsymbol{\lambda}_i + c_i(\mathbf{y}_i)\}, \quad i = 1, \ldots, K,$$

where $\mathbf{w}_i = \mathrm{vech}(\mathbf{y}_i \mathbf{y}_i')$, $c_i(\cdot)$ is a "shape" function, Δ_i is a normalization constant, and $\boldsymbol{\theta}_i$ and $\boldsymbol{\lambda}_i$ are canonical parameters which may be expressed as functions of the mean and covariance parameters, $\boldsymbol{\mu}_i$ and $\boldsymbol{\sigma}_i$. In this case, Prentice & Zhao (1991) showed that the score equation for $\boldsymbol{\gamma}$ is

$$\sum_{i=1}^{K} \mathbf{D}_i' \mathbf{V}_i^{-1} \mathbf{f}_i = \mathbf{0}, \tag{7}$$

where now

$$\mathbf{D}_i = \frac{\partial \binom{\boldsymbol{\mu}_i}{\boldsymbol{\sigma}_i}}{\partial \boldsymbol{\gamma}'} = \begin{pmatrix} \partial \boldsymbol{\mu}_i / \partial \boldsymbol{\beta}' & \mathbf{0} \\ \partial \boldsymbol{\sigma}_i / \partial \boldsymbol{\beta}' & \partial \boldsymbol{\sigma}_i / \partial \tilde{\boldsymbol{\alpha}}' \end{pmatrix} = \begin{pmatrix} \mathbf{D}_{i11} & \mathbf{D}_{i12} \\ \mathbf{D}_{i21} & \mathbf{D}_{i22} \end{pmatrix},$$

$$\mathbf{V}_i = \mathrm{cov}\begin{pmatrix} \mathbf{y}_i \\ \mathbf{s}_i \end{pmatrix} = \begin{pmatrix} \mathrm{var}(\mathbf{y}_i) & \mathrm{cov}(\mathbf{y}_i, \mathbf{s}_i) \\ \mathrm{cov}(\mathbf{s}_i, \mathbf{y}_i) & \mathrm{var}(\mathbf{s}_i) \end{pmatrix} = \begin{pmatrix} \mathbf{V}_{i11} & \mathbf{V}_{i12} \\ \mathbf{V}_{i21} & \mathbf{V}_{i22} \end{pmatrix}.$$

The corresponding information matrix is $K^{-1} \sum_{i=1}^{K} \mathbf{D}_i' \mathbf{V}_i^{-1} \mathbf{D}_i$. For $\hat{\boldsymbol{\gamma}}_{\mathrm{G2}}$ solving (7), the asymptotic covariance matrix of $\sqrt{K}(\hat{\gamma}_{\mathrm{G2}} - \boldsymbol{\gamma})$ can be consistently estimated by

$$\left(\sum_{i=1}^{K} \hat{\mathbf{D}}_i' \hat{\mathbf{V}}_i^{-1} \hat{\mathbf{D}}_i \right)^{-1} \left(\sum_{i=1}^{K} \hat{\mathbf{D}}_i' \hat{\mathbf{V}}_i^{-1} \mathbf{f}_i \mathbf{f}_i' \hat{\mathbf{V}}_i^{-1} \hat{\mathbf{D}}_i \right) \left(\sum_{i=1}^{K} \hat{\mathbf{D}}_i' \hat{\mathbf{V}}_i^{-1} \hat{\mathbf{D}}_i \right)^{-1}$$

Note that this formula differs from the corresponding expression for GEE1.

When the estimating equation (7) corresponds to the first derivative of a misspecified likelihood in the quadratic exponential family, solving (7) to obtain an estimate of $\boldsymbol{\gamma}$ is known as pseudo maximum likelihood (PML). Of course, if the likelihood is correctly specified then the procedure is maximum likelihood. More generally, the GEE2 procedure proceeds by letting $\mathbf{V}_{i21}$ and $\mathbf{V}_{i22}$ be "working" matrices which may or may not correspond to a legitimate choice of the shape function $c_i(\mathbf{y}_i)$ in the quadratic exponential family. When third and fourth moments are known and used to form $\mathbf{V}_{i21}$ and $\mathbf{V}_{i22}$, estimating equation (7) is optimal (Godambe & Thompson 1989). Alternatively, we may plug in $\sqrt{K}-$consistent estimators of any third and fourth moment parameters that may appear in our working matrices $\mathbf{V}_{i21}$ and $\mathbf{V}_{i22}$, thus obtaining a procedure analogous to Liang & Zeger's (1986) original idea. In the PML case, the consistency of $\hat{\boldsymbol{\gamma}}_{\mathrm{G2}}$ was proven by Gouriéroux *et al.* (1984). In the more general GEE2 case, consistency of $\hat{\boldsymbol{\gamma}}_{\mathrm{G2}}$ follows from the theory laid out in Crowder (1986).

Note that a special case of GEE2 occurs when the multivariate Gaussian likelihood is used as a criterion function to be maximized with respect to both $\boldsymbol{\beta}$ and $\tilde{\boldsymbol{\alpha}}$. This approach has been examined in the context of clustered binomial data by Crowder (1985) and Paul & Islam (1998).

2.3 EGEE.

The GEE1 procedure disussed in Section 2.1 may be seen as an application of quasi-likelihood to longitudinal data. Equation (1) is the quasi-score equation which results upon differentiation of quasi-likelihood functions $Q_i(\boldsymbol{\mu}_i; \mathbf{y}_i)$, $i = 1, \ldots, K$, defined below, and summation over i. The partial derivatives of these Q_i's have score-like properties when taken with respect to $\boldsymbol{\beta}$ but not with respect to $\tilde{\boldsymbol{\alpha}}$ (McCullagh & Nelder 1989, p. 325). Hall & Severini (1998) proposed estimating equations for $\boldsymbol{\gamma}$ formed from the derivatives of an extended quasi-likelihood function $Q_i^+(\boldsymbol{\mu}_i, \tilde{\boldsymbol{\alpha}}; \mathbf{y}_i)$; Q_i^+ is an extended quasi-likelihood function (Nelder & Pregibon 1987) in the sense that its partial derivatives with respect to both $\boldsymbol{\beta}$ and $\tilde{\boldsymbol{\alpha}}$ have properties similar to an efficient score vector.

Specifically, the extended quasi-likelihood function is assumed to be of the form $Q_i^+(\boldsymbol{\mu}_i, \tilde{\boldsymbol{\alpha}}; \mathbf{y}_i) = Q_i(\boldsymbol{\mu}_i; \mathbf{y}_i) + f_{i1}(\tilde{\boldsymbol{\alpha}}) + f_{i2}(\mathbf{y}_i)$, which ensures that $\partial Q_i^+/\partial \boldsymbol{\beta} = \partial Q_i/\partial \boldsymbol{\beta}$. The requirement $E\left(\partial Q_i^+/\partial \tilde{\boldsymbol{\alpha}}\right) = \mathbf{0}$ leads to

$$\frac{\partial f_{i1}(\tilde{\boldsymbol{\alpha}})}{\partial \tilde{\alpha}_u} \approx \frac{1}{2}\,\mathrm{tr}\left(\mathbf{V}_{i11}\frac{\partial \mathbf{V}_{i11}^{-1}}{\partial \tilde{\alpha}_u}\right), \quad u = 1, \ldots, s+1, \tag{8}$$

which equals $\frac{1}{2}\,\mathrm{tr}\left(\mathbf{R}\partial\mathbf{R}^{-1}/\partial\tilde{\alpha}_u\right)$ under decomposition (2). A true extended quasi-likelihood approach would now proceed by taking partial derivatives of Q_i^+ with respect to the components of $\boldsymbol{\beta}$ and $\tilde{\boldsymbol{\alpha}}$ to obtain estimating functions for these parameters. For $\boldsymbol{\beta}$, this leads to the estimating equation given by (3). For the estimation of $\boldsymbol{\alpha}$ though, this approach has two drawbacks: (i) $\partial Q_i/\partial\boldsymbol{\alpha}$, the first term in $\partial Q_i^+/\partial\boldsymbol{\alpha}$, involves a line integral that is, in general, not easy to evaluate; (ii) the resulting estimating function is biased and in general yields an inconsistent parameter estimator.

To expand on point (i), the quasi-likelihood function Q_i, when it exists, is given by

$$Q_i(\boldsymbol{\mu}_i; \mathbf{y}_i) \quad = \quad (\mathbf{y}_i - \boldsymbol{\mu}_i)' \int_{\mathbf{t}(s)=\mathbf{y}_i}^{\mathbf{t}(s)=\boldsymbol{\mu}_i} \mathbf{V}_{i11}^{-1}\{\mathbf{t}(s)\} d\mathbf{t}(s)$$

$$= \quad -(\mathbf{y}_i - \boldsymbol{\mu}_i)' \left[\int_0^1 s \mathbf{V}_{i11}^{-1}\{\mathbf{t}(s)\} ds \right] (\mathbf{y}_i - \boldsymbol{\mu}_i). \tag{9}$$

The second equality in (9) follows if we choose the path of integration to be the straight-line path $\mathbf{t}(s) = \mathbf{y}_i + (\boldsymbol{\mu}_i - \mathbf{y}_i)s$ (McCullagh & Nelder 1989, p. 335). Under decomposition (2), we require the matrix

$$\int_0^1 s \mathbf{A}_i^{-1/2}\{\mathbf{t}(s)\} \frac{\partial\{\phi^{-1}\mathbf{R}^{-1}(\boldsymbol{\alpha})\}}{\partial \tilde{\alpha}_u} \mathbf{A}_i^{-1/2}\{\mathbf{t}(s)\} ds, \quad u = 1, \ldots, s+1 \tag{10}$$

whose (j,k)-th element is some function of $\boldsymbol{\alpha}$ (depending upon the chosen working correlation structure) times

$$\int_0^1 \frac{s ds}{\sqrt{v\{y_{ij} + (\mu_{ij} - y_{ij})s\}}\sqrt{v\{y_{ik} + (\mu_{ik} - y_{ik})s\}}} \,. \tag{11}$$

Aside from the normal-theory case $v(\mu) = 1$, both analytical expressions for these integrals and numerical integration methods are complicated by the endpoint singularites in (11).

To avoid these complications and to produce unbiased estimating functions for $\tilde{\boldsymbol{\alpha}}$, Hall and Severini approximated the integrand in (10) with a first-order Taylor series expansion in $\mathbf{y}_i$ taken at $\boldsymbol{\mu}_i$. Note that this is the same approximation used to obtain the relationship (8). The resulting estimating equations are given by $\sum_{i=1}^K U_i(\boldsymbol{\gamma}) = \mathbf{0}$, where $U_i(\boldsymbol{\gamma}) = (U_i(\boldsymbol{\gamma};\boldsymbol{\beta})', U_i(\boldsymbol{\gamma};\tilde{\alpha}_1), \ldots, U_i(\boldsymbol{\gamma};\tilde{\alpha}_{s+1}))'$, and

$$\begin{aligned} U_i(\boldsymbol{\gamma};\boldsymbol{\beta}) &= \mathbf{D}_{i11}'\mathbf{V}_{i11}^{-1}(\mathbf{y}_i - \boldsymbol{\mu}_i), \\ U_i(\boldsymbol{\gamma};\tilde{\alpha}_u) &= -(\mathbf{y}_i - \boldsymbol{\mu}_i)' \frac{\partial \mathbf{V}_{i11}^{-1}}{\partial \tilde{\alpha}_u}(\mathbf{y}_i - \boldsymbol{\mu}_i) + \mathrm{tr}\left(\mathbf{V}_{i11} \frac{\partial \mathbf{V}_{i11}^{-1}}{\partial \tilde{\alpha}_u} \right), \end{aligned} \tag{12}$$

$u = 1, \ldots, s+1$. Note that it will often be desirable to make use of the fact that $\mathbf{V}_{i11}\partial\mathbf{V}_{i11}^{-1}/\partial\tilde{\alpha}_u = -\left(\partial\mathbf{V}_{i11}/\partial\tilde{\alpha}_u\right)\mathbf{V}_{i11}^{-1}$ when working with $U_i(\boldsymbol{\gamma};\tilde{\boldsymbol{\alpha}}_u)$ to avoid differentiating $\mathbf{V}_{i11}^{-1}$.

3. RELATIONSHIPS AMONG EGEE, GEE1, AND GEE2

As described in Section 2.3, in the EGEE approach an estimating function for $\boldsymbol{\beta}$ is obtained as the first derivative with respect to $\boldsymbol{\beta}$ of the extended quasi-likelihood function $\sum_i Q_i^+$ where

$$Q_i^+ = -(\mathbf{y}_i - \boldsymbol{\mu}_i)' \left[\int_0^s s \mathbf{V}_{i11}^{-1}\{\mathbf{t}(s)\} ds \right] (\mathbf{y}_i - \boldsymbol{\mu}_i) - \frac{1}{2} \log |\mathbf{V}_{i11}(\boldsymbol{\mu}_i, \tilde{\boldsymbol{\alpha}})|.$$

The estimating equation for $\tilde{\boldsymbol{\alpha}}$ is obtained similarly, but where the first term in Q_i^+ is replaced by $-(\mathbf{y}_i - \boldsymbol{\mu}_i)'\mathbf{V}_{i11}^{-1}(\boldsymbol{\mu}, \tilde{\boldsymbol{\alpha}})(\mathbf{y}_i - \boldsymbol{\mu}_i)/2$. With this substitution, it is apparent that $\hat{\tilde{\boldsymbol{\alpha}}}$ is chosen to maximize a multivariate normal log-likelihood. For this reason, Hall and Severini's method is, to be precise, neither strictly extended QL nor strictly pseudo maximum likelihood. Rather it is something of a hybrid between the two, combining the quasi-score equation for $\boldsymbol{\beta}$ with Gaussian estimation for $\tilde{\boldsymbol{\alpha}}$.

Of course, the strictly pseudo maximum likelihood approach is GEE2, which implies that the EGEE and GEE2 estimating functions for $\tilde{\boldsymbol{\alpha}}$ should be equal when GEE2 is implemented with a Gaussian working structure for third and fourth order moments. This is indeed the case, although it is not obvious from expressions (7) and (12). Under the Gaussian working structure, GEE2 becomes

$$\sum_{i=1}^{K} \begin{pmatrix} \partial\boldsymbol{\mu}_i'/\partial\boldsymbol{\beta} & \partial\boldsymbol{\sigma}_i'/\partial\boldsymbol{\beta} \\ \mathbf{0} & \partial\boldsymbol{\sigma}_i'/\partial\tilde{\boldsymbol{\alpha}} \end{pmatrix} \begin{pmatrix} \mathbf{V}_{i11}^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{V}_{i22}^{-1} \end{pmatrix} \begin{pmatrix} \mathbf{y}_i - \boldsymbol{\mu}_i \\ \mathbf{s}_i - \boldsymbol{\sigma}_i \end{pmatrix} = \mathbf{0}, \tag{13}$$

where $\mathbf{V}_{i22} = \text{var}(\mathbf{s}_i) = \mathbf{H}(\mathbf{I} + \mathbf{K}_n)(\mathbf{V}_{i11} \otimes \mathbf{V}_{i11})\mathbf{H}'$ by Theorem 4.4 of Magnus & Neudecker (1979). In this expression, $\mathbf{I}$ is the identity matrix, $\mathbf{K}_n$ is the $n^2 \times n^2$ commutation matrix which, for arbitrary $n \times n$ matrix $\mathbf{A}$, transforms vec$\mathbf{A}$ into vec$\mathbf{A}'$, and $\mathbf{H}$ is a (non-unique) matrix with the property vech$\mathbf{A} = \mathbf{H}$vec$\mathbf{A}$ for $\mathbf{A}$ symmetric (cf. Searle 1982, Section 12.9). Therefore, the GEE2 estimating function for $\tilde{\alpha}_u$ in this case is

$$\frac{\partial\boldsymbol{\sigma}_i'}{\partial\tilde{\alpha}_u} \left\{\mathbf{H}(\mathbf{I} + \mathbf{K}_n)(\mathbf{V}_{i11} \otimes \mathbf{V}_{i11})\mathbf{H}'\right\}^{-1} (\mathbf{s}_i - \boldsymbol{\sigma}_i).$$

Using properties of Kronecker products and the vec and vech operators (cf., e.g., Magnus & Neudecker 1979 and Searle 1982), $-U_i(\boldsymbol{\gamma};\tilde{\alpha}_u)/2$ can be written as

$$\begin{aligned} &\frac{1}{2}\left(\text{vec}\frac{\partial\mathbf{V}_{i11}^{-1}}{\partial\tilde{\alpha}_u}\right)' [\text{vec}\{(\mathbf{y}_i - \boldsymbol{\mu}_i)(\mathbf{y}_i - \boldsymbol{\mu}_i)'\} - \text{vec}\mathbf{V}_{i11}] \\ &= \frac{1}{2}\left(\text{vech}\frac{\partial\mathbf{V}_{i11}}{\partial\tilde{\alpha}_u}\right)' \mathbf{G}'(\mathbf{V}_{i11} \otimes \mathbf{V}_{i11})^{-1}\mathbf{G} \\ &\qquad \times [\text{vech}\{(\mathbf{y}_i - \boldsymbol{\mu}_i)(\mathbf{y}_i - \boldsymbol{\mu}_i)'\} - \text{vech}\mathbf{V}_{i11}], \end{aligned} \tag{14}$$

where $\mathbf{G}$ is the unique matrix such that for symmetric matrix $\mathbf{A}$, vec$\mathbf{A} = \mathbf{G}$vech$\mathbf{A}$. Since $\mathbf{G}'(\mathbf{V}_{i11} \otimes \mathbf{V}_{i11})^{-1}\mathbf{G}/2 = \mathbf{V}_{i22}^{-1}$, it follows that the EGEE estimating equations for $\tilde{\boldsymbol{\alpha}}$ are equivalent to the corresponding GEE2 equations when a Gaussian working structure is used. The two approaches differ in that EGEE treats $\boldsymbol{\beta}$ and $\tilde{\boldsymbol{\alpha}}$ orthogonally. In fact, if we replace $\partial\boldsymbol{\sigma}_i'/\partial\boldsymbol{\beta}$ with $\mathbf{0}$ in (13), we obtain EGEE in a different, but equivalent, form from that given in Section 2.3.

A simple consequence of this result is that EGEE and GEE1 are equivalent under a Gaussian form for $\mathbf{V}_{i22}$. That is, EGEE is a special case of GEE1. This fact was unrecognized in Hall & Severini (1998), although it should have been revealed in the efficiency comparisons made by those authors. That is, I) ratios of asymptotic variances in Table 2 of their paper should have been one under the "GAUSS1" specification for "Model higher correlations" and II) all mean square error ratios in Table 5 of their paper should have been one as well. Fact I was not observed because the incorrect formula given in Prentice & Zhao (1991) (corrected by expression (6) of the current work) was used to compute the asymptotic variance-covariance matrix of $\hat{\boldsymbol{\gamma}}$ in Hall & Severini (1998). Fact II was not observed because those authors applied a bias correction in the estimation of ϕ. That is, $\hat{\phi} = (N - p)^{-1}\sum_{i=1}^{K}(\mathbf{y}_i - \hat{\boldsymbol{\mu}}_i)'\mathbf{V}_{i11}(\hat{\boldsymbol{\mu}}_i, \hat{\boldsymbol{\alpha}})(\mathbf{y}_i - \hat{\boldsymbol{\mu}}_i)$, where $N = \sum_i n_i$, was used rather than $(N-p)\hat{\phi}/N$ in the EGEE computations, but not in the GEE1 computations.

For computational and theoretical purposes, a convenient form for EGEE is suggested by the left-hand side of (14). EGEE can be written as a single estimating equation of the form

$$\sum_{i=1}^{K} U_i(\boldsymbol{\gamma}) = \sum_{i=1}^{K} \begin{pmatrix} \mathbf{D}_{i11}' & \mathbf{0} \\ \mathbf{0} & \mathbf{F}_i' \end{pmatrix} \begin{pmatrix} \mathbf{V}_{i11}^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{I} \end{pmatrix} \begin{pmatrix} \mathbf{y}_i - \boldsymbol{\mu}_i \\ \mathbf{s}_i^* - \boldsymbol{\sigma}_i^* \end{pmatrix} = \mathbf{0}, \tag{15}$$

where $\mathbf{s}_i^* = \text{vec}\{(\mathbf{y}_i - \boldsymbol{\mu}_i)(\mathbf{y}_i - \boldsymbol{\mu}_i)'\}$, $\boldsymbol{\sigma}_i^* = \text{vec}\mathbf{V}_{i11}$, and $\mathbf{F}_i = \partial\{\text{vec}\mathbf{V}_{i11}^{-1}\}/\partial\tilde{\boldsymbol{\alpha}}'$. From this expression it follows easily that a consistent estimate of the asymptotic variance of the EGEE estimator $\hat{\boldsymbol{\gamma}}_{\text{EG}}$, the solution to (15), is given by

$$\{U_\cdot^*(\hat{\boldsymbol{\gamma}}_{\text{EG}})\}^{-1}\left\{\sum_i U_i(\hat{\boldsymbol{\gamma}}_{\text{EG}})U_i(\hat{\boldsymbol{\gamma}}_{\text{EG}})'\right\}\{U_\cdot^*(\hat{\boldsymbol{\gamma}}_{\text{EG}})\}^{-1},$$

where

$$U_\cdot^*(\boldsymbol{\gamma}) = \sum_{i=1}^K \begin{pmatrix} \mathbf{D}_{i11}' & \mathbf{0} \\ \mathbf{0} & \mathbf{F}_i' \end{pmatrix} \begin{pmatrix} \mathbf{V}_{i11}^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{I} \end{pmatrix} \begin{pmatrix} \mathbf{D}_{i11} & \mathbf{0} \\ \partial\boldsymbol{\sigma}_i^*/\partial\boldsymbol{\beta}' & \partial\boldsymbol{\sigma}_i^*/\partial\tilde{\boldsymbol{\alpha}}' \end{pmatrix}.$$

A restricted maximum likelihood (REML)-like version of EGEE can be used to adjust the estimator of $\tilde{\boldsymbol{\alpha}}$ for lost degrees of freedom in estimating $\boldsymbol{\beta}$. For this REML-like procedure, we obtain parameter estimates, $\hat{\boldsymbol{\gamma}}_{\text{REG}}$, say, as the solution of

$$\sum_{i=1}^K \begin{pmatrix} \partial\boldsymbol{\mu}_i'/\partial\boldsymbol{\beta} & \mathbf{0} \\ \mathbf{0} & \mathbf{F}_i' \end{pmatrix} \begin{pmatrix} \mathbf{V}_{i11}^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{I} \end{pmatrix} \begin{pmatrix} \mathbf{y}_i - \boldsymbol{\mu}_i \\ \mathbf{s}_i^* - \boldsymbol{\rho}_i \end{pmatrix} = \mathbf{0},$$

where $\boldsymbol{\rho}_i = \text{vec}(\mathbf{V}_i\mathbf{P}_i\mathbf{V}_i)$ and

$$\mathbf{P}_i = \mathbf{V}_i^{-1} - \mathbf{V}_i^{-1}\mathbf{D}_{i11}\left(\sum_{i=1}^K \mathbf{D}_{i11}'\mathbf{V}_{i11}^{-1}\mathbf{D}_{i11}\right)^{-1}\mathbf{D}_{i11}'\mathbf{V}_{i11}^{-1}.$$

Notice that $\boldsymbol{\rho}_i = \boldsymbol{\sigma}_i^* - \boldsymbol{\zeta}_i$, where $\boldsymbol{\zeta}_i = \text{vec}\{\mathbf{D}_{i11}(\sum_i \mathbf{D}_{i11}'\mathbf{V}_{i11}^{-1}\mathbf{D}_{i11})^{-1}\mathbf{D}_{i11}'\}$. Therefore, the REML estimating function for $\tilde{\boldsymbol{\alpha}}$ involves a correction term which corrects for bias in the EGEE profile estimating function for $\tilde{\boldsymbol{\alpha}}$. A similar REML-like bias correction for second moment parameters has been described by Breslow & Clayton (1993, equation 13) in a generalized linear mixed model context, and also by Wolfinger & Lin (1997) and Hall & Bailey (2000) in nonlinear mixed model contexts. We will consider a similar bias correction for the extended quasi-likelihood (EQL) estimating functions proposed in Section 4.

Again, a sandwich-type estimator of $\text{var}(\hat{\boldsymbol{\gamma}}_{\text{REG}})$ can be used, with "bread" equal to the inverse of the matrix

$$\sum_{i=1}^K \begin{pmatrix} \mathbf{B}_{i11} & \mathbf{0} \\ \mathbf{B}_{i21} & \mathbf{B}_{i22} \end{pmatrix},$$

where $\mathbf{B}_{i11} = \mathbf{D}_{i11}'\mathbf{V}_{i11}^{-1}\mathbf{D}_{i11}$, $\mathbf{B}_{i21}$ is the $p \times (s+1)$ matrix with (j,k)-th element

$$\left(\frac{\partial\text{vec}\mathbf{V}_{i11}^{-1}}{\partial\tilde{\alpha}_j}\right)'\frac{\partial\boldsymbol{\rho}_i}{\beta_k} - \left(\frac{\partial^2\text{vec}\mathbf{V}_{i11}^{-1}}{\partial\tilde{\alpha}_j\partial\beta_k}\right)'\boldsymbol{\zeta}_i,$$

$\mathbf{B}_{i22}$ is the $(s+1)\times(s+1)$ matrix with (j,k)-th element

$$\left(\frac{\partial\text{vec}\mathbf{V}_{i11}^{-1}}{\partial\tilde{\alpha}_j}\right)'\frac{\partial\boldsymbol{\rho}_i}{\tilde{\alpha}_k} - \left(\frac{\partial^2\text{vec}\mathbf{V}_{i11}^{-1}}{\partial\tilde{\alpha}_j\partial\tilde{\alpha}_k}\right)'\boldsymbol{\zeta}_i,$$

and $\mathbf{B}_{i11}, \mathbf{B}_{i21}, \mathbf{B}_{i22}$ are all evaluated at $\hat{\boldsymbol{\gamma}}_{\text{REG}}$. The "meat" is given by

$$\sum_{i=1}^K \begin{pmatrix} \mathbf{D}_{i11}'\mathbf{V}_{i11}^{-1}(\mathbf{y}_i - \boldsymbol{\mu}_i)(\mathbf{y}_i - \boldsymbol{\mu}_i)'\mathbf{V}_{i11}^{-1}\mathbf{D}_{i11} & \mathbf{D}_{i11}'\mathbf{V}_{i11}^{-1}(\mathbf{y}_i - \boldsymbol{\mu}_i)(\mathbf{s}_i^* - \boldsymbol{\rho}_i^*)'\mathbf{F}_i \\ \mathbf{F}_i'(\mathbf{s}_i^* - \boldsymbol{\rho}_i^*)(\mathbf{y}_i - \boldsymbol{\mu}_i)'\mathbf{V}_{i11}^{-1}\mathbf{D}_{i11} & \mathbf{F}_i'\{(\mathbf{s}_i^* - \boldsymbol{\rho}_i^*)(\mathbf{s}_i^* - \boldsymbol{\rho}_i^*)' - \boldsymbol{\zeta}_i\boldsymbol{\zeta}_i'\}\mathbf{F}_i \end{pmatrix},$$

which is again evaluated at $\hat{\gamma}_{\text{REG}}$.

4. EXTENDED QUASI-LIKELIHOOD FOR CLUSTERED DATA

As described in Section 2.3, a true extended quasi-likelihood (EQL) approach to the estimation of $\tilde{\boldsymbol{\alpha}}$ would solve the estimating equation $\sum_{i=1}^{K} \partial Q_i^{+}(\boldsymbol{\mu}_i, \tilde{\boldsymbol{\alpha}}; \mathbf{y}_i)/\partial\tilde{\boldsymbol{\alpha}} = \mathbf{0}$ for $\tilde{\boldsymbol{\alpha}}$. This estimating equation takes the form

$$\sum_{i=1}^{K}(\mathbf{E}_i'\mathbf{s}_i^* - \mathbf{F}_i'\boldsymbol{\sigma}_i^*) = \mathbf{0}, \tag{16}$$

where $\mathbf{s}_i^*$, $\boldsymbol{\sigma}_i^*$, $\mathbf{F}_i$ are as defined in Section 3, $\mathbf{E}_i$ is the $n^2 \times (s+1)$ matrix with u-th column equal to $2\text{vec}(\mathbf{W}_{iu})$, and

$$\mathbf{W}_{iu} = \frac{\partial}{\partial\tilde{\alpha}_u}\left[\int_0^1 s\mathbf{V}_i^{-1}\{\mathbf{t}(s)\}ds\right].$$

Note that, as in the univariate case where extended quasi-likelihood has been proposed for estimation of a scalar overdispersion parameter (Nelder & Pregibon 1987), the EQL estimating function in (16) is biased, leading to a generally inconsistent estimator of $\tilde{\boldsymbol{\alpha}}$ (Davidian & Carroll 1988). Furthermore, correcting the bias in (16) would require higher than second moment information to approximate $E\{\partial Q_i(\boldsymbol{\mu}_i; \mathbf{y}_i)/\partial\tilde{\boldsymbol{\alpha}}\}$, undermining the appeal of the EQL approach. Despite this theoretical drawback, Nelder & Lee (1992, 1997) have described examples in which the EQL overdispersion parameter estimator out-performs alternative estimators in terms of both asymptotic, and finite-sample efficiency. A natural question, therefore, is whether EQL can be extended to the clustered data case to yield improved estimators of $\tilde{\boldsymbol{\alpha}}$.

Let $\hat{\boldsymbol{\gamma}}_{\text{EQ}}$ be the EQL estimator obtained by solving (3) and (16). Under regularity conditions as in Takagi & Inagaki (1993), the asymptotic variance-covariance matrix of $\sqrt{K}(\hat{\gamma}_{\text{EQ}} - \boldsymbol{\gamma})$ is given by $\mathbf{J}^{-1}\mathcal{I}\left(\mathbf{J}^{-1}\right)'$, where

$$\mathbf{J} = \lim_{K\to\infty}\frac{1}{K}\sum_{i=1}^{K}\frac{\partial}{\partial\boldsymbol{\gamma}'}\begin{pmatrix}\mathbf{D}_{i11}'\mathbf{V}_{i11}^{-1}(\mathbf{y}_i - \boldsymbol{\mu}_i)\\ \mathbf{E}_i'\mathbf{s}_i^* - \mathbf{F}_i'\boldsymbol{\sigma}_i^*\end{pmatrix}$$

and

$$\mathcal{I} = \lim_{K\to\infty}\frac{1}{K}\sum_{i=1}^{K}\text{var}\begin{pmatrix}\mathbf{D}_{i11}'\mathbf{V}_{i11}^{-1}(\mathbf{y}_i - \boldsymbol{\mu}_i)\\ \mathbf{E}_i'\mathbf{s}_i^* - \mathbf{F}_i'\boldsymbol{\sigma}_i^*\end{pmatrix}.$$

From this result, it follows that the usual sandwich variance-covariance estimator obtained by evaluating $\mathbf{J}$ and $\mathcal{I}$ at $\boldsymbol{\gamma} = \hat{\boldsymbol{\gamma}}_{\text{EQ}}$ is consistent for $\text{var}(\hat{\boldsymbol{\beta}}_{\text{EQ}})$ and $\text{cov}(\hat{\boldsymbol{\beta}}_{\text{EQ}}, \hat{\tilde{\boldsymbol{\alpha}}}_{\text{EQ}})$. In general, it is necessary to consistently estimate the vector $E(\mathbf{E}_i'\mathbf{s}_i^*)$ to form a consistent estimator of $\text{var}(\hat{\tilde{\boldsymbol{\alpha}}}_{\text{EQ}})$. This task is complicated both by the general inconsistency of $\hat{\tilde{\boldsymbol{\alpha}}}$ and the mathematical inconvenience of the forms taken by $\mathbf{E}_i$ under specific variance function specifications (see below). However, based on the simulation results of Section 5 and the theory available from the univariate case (Davidian & Carroll 1988, Smyth & Verbyla 1999), we expect the asymptotic bias in $\hat{\tilde{\boldsymbol{\alpha}}}_{\text{EQ}}$ to be small in many problems of practical interest. In such cases, the usual sandwich variance estimator should provide a reasonable approximation to $\text{var}(\hat{\tilde{\boldsymbol{\alpha}}}_{\text{EQ}})$.

The specific form of the EQL estimating equation (16) depends upon the variance function v because $\mathbf{W}_{iu}$ has (j,k)-th element

$$\frac{\partial(R^{jk}/\phi)}{\tilde{\alpha}_u}\int_0^1 \frac{sds}{\sqrt{v(y_{ij}+(\mu_{ij}-y_{ij})s)}\sqrt{v(y_{ik}+(\mu_{ik}-y_{ik})s)}},$$

where R^{jk} is the (j,k)-th element of $\mathbf{R}^{-1}(\boldsymbol{\alpha})$. For the constant variance function $v(\mu)=1$, we recover EGEE. However, for other common variance function, the EQL estimating equations and EGEE differ. Next, we present the specific form of the integral in $\mathbf{W}_{iu}$ for the most commonly used variance functions, $v(\mu)=\mu^2$ (gamma distribution), $v(\mu)=\mu$ (Poisson distribution), and $v(\mu)=\mu(1-\mu)$ (binomial).

Gamma Variance Function, $v(\mu)=\mu^2$: In this case, $\mathbf{W}_{iu}$ has (j,k)-th element equal to $\partial(R^{jk}/\phi)/\partial\tilde{\alpha}_u$ times

$$\int_0^1 \frac{sds}{\{y_{ij}+(\mu_{ij}-y_{ij})s\}\{y_{ik}+(\mu_{ik}-y_{ik})s\}}.$$

The latter integral is equal to $1/(2y_{ij}y_{ik})$ if $y_{ij}=\mu_{ij}$ and $y_{ik}=\mu_{ik}$; furthermore, it reduces to

$$\frac{(\mu_{ik}-y_{ik})+y_{ik}\log(y_{ik}/\mu_{ik})}{y_{ij}(\mu_{ik}-y_{ik})^2}$$

if $y_{ij}=\mu_{ij}$ and $y_{ik}=\mu_{ik}$. Finally, it is equal to

$$\frac{y_{ij}(\mu_{ik}-y_{ik})\log(y_{ij}/\mu_{ij})-y_{ik}(\mu_{ij}-y_{ij})\log(y_{ik}/\mu_{ik})}{(\mu_{ij}-y_{ij})(\mu_{ik}-y_{ik})\{y_{ik}(\mu_{ij}-y_{ij})-y_{ij}(\mu_{ik}-y_{ik})\}}$$

when $y_{ij}\neq\mu_{ij}$ and $y_{ik}\neq\mu_{ik}$.

Poisson Variance Function, $v(\mu)=\mu$: In this case, we need

$$\int_0^1 \frac{sds}{\sqrt{y_{ij}+(\mu_{ij}-y_{ij})s}\sqrt{y_{ik}+(\mu_{ik}-y_{ik})s}}. \tag{17}$$

When $y_{ij}=\mu_{ij}>0$ and $y_{ik}=\mu_{ik}>0$, (17) equals $(y_{ij}y_{ik})^{-1/2}/2$, and when $y_{ij}=\mu_{ij}>0$ and $\mu_{ik}\neq y_{ik}\geq 0$, (17) equals

$$\frac{\mu_{ik}^{3/2}+2y_{ik}^{3/2}-3y_{ik}\mu_{ik}^{1/2}}{3y_{ij}^{1/2}(\mu_{ik}-y_{ik})^2/2}.$$

Otherwise, (17) is given by

$$\frac{(\mu_{ij}\mu_{ik})^{1/2}-(y_{ij}y_{ik})^{1/2}}{s_{jk}}-\frac{y_{ij}(\mu_{ik}-y_{ik})+y_{ik}(\mu_{ij}-y_{ij})}{2|s_{jk}|^{3/2}}A,$$

where $s_{jk}=(\mu_{ij}-y_{ij})(\mu_{ik}-y_{ik})$ and

$$A=2\log\left\{\frac{|(s_{jk}\mu_{ij})^{1/2}+(\mu_{ij}-y_{ij})\mu_{ik}^{1/2}|}{|(s_{jk}y_{ij})^{1/2}+(\mu_{ij}-y_{ij})y_{ik}^{1/2}|}\right\},$$

if $y_{ij}, y_{ik} \geq 0$ and $s_{jk} > 0$, and

$$A = \arcsin\left(\frac{y_{ij}\mu_{ik} + y_{ik}\mu_{ij} - 2\mu_{ij}\mu_{ik}}{y_{ik}\mu_{ij} - y_{ij}\mu_{ik}}\right) - \arcsin\left(\frac{y_{ij}\mu_{ik} + y_{ik}\mu_{ij} - 2y_{ij}y_{ik}}{y_{ij}\mu_{ik} - y_{ik}\mu_{ij}}\right),$$

if $y_{ij}, y_{ik} \geq 0$, $\mu_{ij} > y_{ij}$ and $\mu_{ik} < y_{ik}$.

Binomial Variance Function, $v(\mu) = \mu(1-\mu)$: In this case, we require

$$\int_0^1 \prod_{\ell \in \{j,k\}} \{y_{i\ell} + (\mu_{i\ell} - y_{i\ell})s\}^{-1/2}\{1 - y_{i\ell} - (\mu_{i\ell} - y_{i\ell})s\}^{-1/2} s ds. \qquad (18)$$

Expression (18) is an example of an elliptic integral of the third kind, and it can be evaluated using formulas given in standard tables of elliptic integrals (e.g., Abramowitz & Stegun 1964). However, with such tables, it is necessary to use several formulas, corresponding to a variety of subcases of (18) depending upon the relationships between y_{ij}, y_{ik}, μ_{ij} and μ_{ik}. A more succinct representation for the integral is given by Carlson (1988), who presents a table of formulas for elliptic integrals of the third kind in terms of "standard" elliptic integrals, $R_F(x, y, z)$ and $R_J(x, y, z, \rho)$, defined in that paper. Computer code (in C) for evaluating these standard integrals is available in Press *et al.* (1992). In Carlson's notation, the integral is given by

$$\begin{aligned}(18) \quad = \quad & 2\left\{\frac{-d_{12}d_{13}d_{14}}{3(\mu_{ij} - y_{ij})} R_J(U_{12}^2, U_{13}^2, U_{14}^2, W_1^2) + R_F(P_1^2, Q_1^2, Q_1^2)\right. \\ & \left. -y_{ij} R_F(U_{12}^2, U_{13}^2, U_{14}^2)\right\} \Big/ (\mu_{ij} - y_{ij}), \qquad (19)\end{aligned}$$

for all cases (although it is assumed that $\mu_{ij}, \mu_{ik} \notin \{0, 1\}$). The notation in (19) is defined in the appendix.

A REML-like version of the EQL estimating function (16) can be obtained by correcting the bias arising in (16) when evaluated at $\hat{\boldsymbol{\beta}}$. Using some small dispersion asymptotics as in Hall & Severini (1998, p. 1367), it can be shown that for exponential dispersion models with dispersion parameter ϕ, the bias in the i-th cluster's contribution to the profile version of (16) is equal to $-\mathbf{F}_i'\boldsymbol{\zeta}_i + \mathrm{O}(\phi^{3/2})$. This result agrees with the corresponding EGEE result to the same order of approximation. Therefore, we suggest replacing (16) by

$$\sum_{i=1}^{K}(\mathbf{E}_i'\mathbf{s}_i^* - \mathbf{F}_i'\boldsymbol{\rho}_i) = \mathbf{0}, \qquad (20)$$

for a REML-like version of EQL. Equation (20) generalizes the approximate REML estimating function suggested by both Lee & Nelder (1998) and Smyth & Verbyla (1999) for estimation of dispersion parameters in "double" generalized linear models.

As we will see in Section 5, both EQL and REML-EQL do provide substantial gains in terms of finite-sample mean squared error for second moment (dispersion and correlation) parameters over available alternative methods of estimation for some problems. This comes despite the absence of a general consistency result for the EQL estimator of $\tilde{\alpha}$. Some indication of when EQL can be expected to perform well is provided by Smyth & Verbyla (1999). These authors examine the accuracy of the extended quasi-likelihood function as an approximation to the true log-likelihood in the univariate problem in which the GLM dispersion parameter ϕ_i is allowed to vary across observations. They suggest that the approximation should be satisfactory when $\tau_i \leq 1/3$ for all i, where $\tau_i = \phi_i v(y_i)/(y_i - \text{boundary})^2$, and "boundary" represents the closest boundary of support for y_i (0 in the Poisson and gamma distributions; 0 or m_i, whichever is closer, in the case that y_i is a binomial random variable based on m_i trials).

Smyth and Verbyla give both a theoretical and a heuristic justification for this rule of thumb in their paper, and they establish bounds on the relative error of the saddle-point approximation underlying the extended quasi-likelihood function. This relative error is less than 2.8% when $\phi_i < 1/3$ for the gamma distribution, and when $y_i > 3$ for the Poisson distribution. For the binomial, the upper bound on the relative error of the saddle point approximation is given as 4% when $y_i \geq 3$ and $m_i - y_i \geq 3$. Such a rule of thumb is consistent with the arguments of Nelder & Lee (1992) and Davidian & Carroll (1988) suggesting that the asymptotic bias when using EQL goes to 0 for certain power of the mean variance functions (including the Poisson variance function) for μ large.

5. FINITE SAMPLE PROPERTIES OF EQL FOR SIMULATED DATA

Here we report the results of a simulation study conducted to investigate the performance of the newly proposed methods, EQL, REML-EQL, and REML-EGEE, relative to the alternatives reviewed in this paper. The purpose is not to present an exhaustive investigation of the relative performance of these methods over a range of models, sample sizes and degrees of correlation, but rather to give some preliminary basis of comparison based on three clustered data problems chosen to be typical of the class of problems to which these methods might be applied.

5.1 Correlated Poisson Data.

In the first simulation, we randomly generated 500 data sets in the structure of the epilepsy data reported by Thall & Vail (1990). These data consist of seizure counts in each of four consecutive two-week periods in addition to an eight-week baseline seizure count, a treatment group indicator, the subject's age, and other covariates measured on each of $K = 59$ individuals. Correlated Poisson data in this configuration (i.e., with $K = 59$, $n_i = 4$, $i = 1, \ldots, 59$, and with the original covariate values) were generated according to the algorithm described by Sim (1993) to have marginal mean given by

$$\mu_{ij} = \exp(0.21 + 0.96\text{BASE}_{ij} - 1.11\text{TRT}_{ij} + 0.41\text{BASE}_{ij}\text{TRT}_{ij}), \qquad (21)$$

where $1 \leq i \leq 59$, $1 \leq j \leq 4$, $\text{BASE} = \log(.25 \times \text{baseline seizure count})$ and TRT is an indicator for inclusion in the active treatment group. The value of $\boldsymbol{\beta}$ in (21) was taken from an EGEE fit with exchangeable correlation structure to the original data. The marginal covariance structure was taken to be as in (2), with dispersion parameter $\phi = 1$, and variance function $v(\mu) = \mu$, and four specifications of the correlation matrix $\mathbf{R}(\boldsymbol{\alpha})$. The correlation structures used were independence, equicorrelation (exchangeable intracluster responses), AR(1), and an unspecified or completely general structure (cf. Liang & Zeger 1986 for a description of these structures). In the non-independence structures, a moderate level of correlation was chosen, with values of $\boldsymbol{\alpha}$ equal to .5, .5, and $(.5, .45, .4, .55, .45, .48)'$ for the exchangeable, AR(1), and unspecified structures, respectively.

In Tables 1 and 2, we present mean square errors for the parameters of model (21) based on fitting this model to the same data using EGEE, REML-EGEE (REGEE), EQL, REML-EQL (REQL), GEE2 with a Gaussian working structure for third and fourth order moments (GEE2-G), and GEE2 with an independence working structure for third and fourth order moments (GEE2-I). Because it is not possible to generate correlated Poisson deviates with Sim's method when the correlation structure is MA(1) (1-dependence), we did not include this structure among our choices for the true correlation matrix. Otherwise, data were generated and fit using all combinations of the structures independence (Indep.), unspecified (Unspec.), exchangeable (Exch.), AR(1), and MA(1). That is, we examined cases in which the correlation structure was correctly specified (Table 1), as well as cases in which it was misspecified (Table 2). Of course, only in the correct specification case is it meaningful to examine the MSE for $\boldsymbol{\alpha}$.

Table 1: Poisson simulation results under correct specification of covariance structure. MSEs for EGEE, REML-EGEE (REGEE), EQL, REML-EQL (REQL), GEE2 with Gaussian working higher moment structure (GEE2-G), and GEE2 with independence working higher moment structure (GEE2-I). These results are based on true values of $\boldsymbol{\beta}$ from an EGEE fit to the original data. Smallest values are underlined.

		Parameter					
True $\mathbf{R}$	Method	β_1	β_2	β_3	β_4	$\boldsymbol{\alpha}'$	ϕ
Indep.	EGEE	$\underline{.0114}$	$\underline{.0019}$	$\underline{.0252}$	$\underline{.0040}$		.0090
	REGEE	$\underline{.0114}$	$\underline{.0019}$	$\underline{.0252}$	$\underline{.0040}$		.0091
	EQL	$\underline{.0114}$	$\underline{.0019}$	$\underline{.0252}$	$\underline{.0040}$		.0112
	REQL	$\underline{.0114}$	$\underline{.0019}$	$\underline{.0252}$	$\underline{.0040}$		.0134
	GEE2-G	.0122	.0021	.0265	.0041		$\underline{.0089}$
	GEE2-I	.0122	.0021	.0265	.0041		$\underline{.0089}$
Unspec.	EGEE	.0238	.0040	.0551	.0082	a	.0152
	REGEE	.0238	.0040	.0550	.0082	b	.0137
	EQL	$\underline{.0237}$	$\underline{.0039}$	$\underline{.0549}$	$\underline{.0081}$	c	$\underline{.0135}$
	REQL	$\underline{.0237}$	$\underline{.0039}$	$\underline{.0549}$	$\underline{.0081}$	d	.0174
	GEE2-G	.0265	.0044	.0611	.0089	e	.0152
	GEE2-I	.0745	.0097	.1138	.0163	f	2.1362
Exch.	EGEE	$\underline{.0267}$	$\underline{.0045}$	$\underline{.0559}$	$\underline{.0086}$	.0070	.0162
	REGEE	$\underline{.0267}$	$\underline{.0045}$	$\underline{.0559}$	$\underline{.0086}$	.0054	.0146
	EQL	$\underline{.0267}$	$\underline{.0045}$	$\underline{.0559}$	$\underline{.0086}$	.0068	$\underline{.0135}$
	REQL	$\underline{.0267}$	$\underline{.0045}$	$\underline{.0559}$	$\underline{.0086}$	$\underline{.0052}$	.0172
	GEE2-G	.0296	.0049	.0631	.0095	.0069	.0162
	GEE2-I	.1175	.0154	.1474	.0251	.0076	5.7815
AR(1)	EGEE	.0246	$\underline{.0041}$	$\underline{.0465}$	$\underline{.0076}$	.0045	$\underline{.0130}$
	REGEE	.0246	$\underline{.0041}$	$\underline{.0465}$	$\underline{.0076}$	.0040	$\underline{.0130}$
	EQL	$\underline{.0245}$	$\underline{.0041}$	$\underline{.0465}$	$\underline{.0076}$	.0043	.0137
	REQL	$\underline{.0245}$	$\underline{.0041}$	$\underline{.0465}$	$\underline{.0076}$	$\underline{.0038}$	.0182
	GEE2-G	.0269	.0045	.0538	.0085	.0044	.0131
	GEE2-I	.0508	.0064	.1001	.0171	.0072	4.2305

$a = (.0110, .0165, .0159, .0102, .0143, .0123)$; $b = (.0094, .0141, .0137, .0087, .0124, .0108)$;
$c = (.0102, .0147, .0144, .0094, .0128, .0116)$; $d = (\underline{.0088}, \underline{.0126}, \underline{.0126}, \underline{.0080}, \underline{.0113}, \underline{.0104})$;
$e = (.0110, .0163, .0156, .0101, .0141, .0122)$; $f = (.0140, .0183, .0171, .0142, .0157, .0152)$.

Table 2: Poisson simulation results under incorrectly specified covariance structure. MSEs for EGEE, REML-EGEE (REGEE), EQL, REML-EQL (REQL), GEE2 with Gaussian working higher moment structure (GEE2-G), and GEE2 with independence working higher moment structure (GEE2-I). These results are based on true values of $\boldsymbol{\beta}$ from an EGEE fit to the original data. Smallest values are underlined.

			Parameter				
True **R**	Fitted **R**	Method	β_1	β_2	β_3	β_4	ϕ
Indep.	Unspec.	EGEE	.0111	.0019	.0233	.0037	.0096
		REGEE	.0111	.0019	.0233	.0037	.0093
		EQL	.0110	.0019	.0233	.0037	.0107
		REQL	.0110	.0019	.0232	.0037	.0143
		GEE2-G	.0121	.0021	.0259	.0040	.0095
		GEE2-I	.0124	.0021	.0261	.0040	.0096
	Exch.	EGEE	.0113	.0018	.0250	.0038	.0111
		REGEE	.0113	.0018	.0250	.0038	.0090
		EQL	.0113	.0018	.0250	.0038	.0208
		REQL	.0113	.0018	.0250	.0038	.0131
		GEE2-G	.0119	.0019	.0272	.0041	.0108
		GEE2-I	.0119	.0019	.0274	.0042	.0108
	AR(1)	EGEE	.0103	.0017	.0229	.0035	.0265
		REGEE	.0103	.0017	.0229	.0035	.0093
		EQL	.0103	.0017	.0229	.0035	.0466
		REQL	.0103	.0017	.0229	.0035	.0132
		GEE2-G	.0109	.0018	.0254	.0038	.0260
		GEE2-I	.0109	.0018	.0256	.0038	.0260
	MA(1)	EGEE	.0104	.0017	.0222	.0035	.0127
		REGEE	.0103	.0017	.0222	.0035	.0088
		EQL	.0103	.0017	.0222	.0035	.0089
		REQL	.0103	.0017	.0222	.0035	.0129
		GEE2-G	.0106	.0018	.0240	.0037	.0128
		GEE2-I	.0107	.0018	.0241	.0037	.0128
Unspec.	Indep.	EGEE	.0274	.0046	.0568	.0087	.0422
		REGEE	.0274	.0046	.0568	.0087	.0158
		EQL	.0274	.0046	.0568	.0087	.0654
		REQL	.0274	.0046	.0568	.0087	.0159
		GEE2-G	.0288	.0049	.0619	.0093	.0412
		GEE2-I	.0288	.0049	.0619	.0093	.0412
	Exch.	EGEE	.0290	.0048	.0532	.0078	.0137
		REGEE	.0290	.0048	.0532	.0078	.0140
		EQL	.0290	.0048	.0532	.0078	.0223
		REQL	.0290	.0048	.0532	.0078	.0184
		GEE2-G	.0315	.0052	.0597	.0086	.0136
		GEE2-I	.0985	.0169	.1100	.0164	.7149
	AR(1)	EGEE	.0287	.0049	.0565	.0084	.0563
		REGEE	.0287	.0049	.0566	.0085	.0133
		EQL	.0287	.0049	.0566	.0084	.0844
		REQL	.0287	.0049	.0567	.0085	.0153
		GEE2-G	.0304	.0052	.0619	.0091	.0552
		GEE2-I	.1672	.0256	.1424	.0258	4.8107

Table 2 (continued)

True **R**	Fitted **R**	Method	Parameter β_1	β_2	β_3	β_4	ϕ
Exch.	MA(1)	EGEE	.0271	.0046	.0562	.0090	.0173
		REGEE	.0271	.0046	.0562	.0090	.0151
		EQL	.0271	.0046	.0562	.0090	.0284
		REQL	.0271	.0046	.0562	.0090	.0125
		GEE2-G	.0284	.0048	.0618	.0097	.0168
		GEE2-I	.0363	.0061	.0799	.0119	.0283
Exch.	Indep.	EGEE	.0265	.0047	.0603	.0097	.0171
		REGEE	.0265	.0047	.0603	.0097	.0152
		EQL	.0265	.0047	.0603	.0097	.0152
		REQL	.0265	.0047	.0603	.0097	.0183
		GEE2-G	.0277	.0049	.0627	.0100	.0172
		GEE2-I	.0277	.0049	.0627	.0100	.0172
	Unspec.	EGEE	.0321	.0054	.0621	.0096	.0200
		REGEE	.0321	.0054	.0621	.0096	.0152
		EQL	.0319	.0054	.0618	.0096	.0353
		REQL	.0319	.0054	.0618	.0096	.0182
		GEE2-G	.0329	.0055	.0688	.0104	.0200
		GEE2-I	.0987	.0171	.1299	.0200	.7304
	AR(1)	EGEE	.0316	.0052	.0591	.0091	.0193
		REGEE	.0317	.0052	.0592	.0091	.0160
		EQL	.0316	.0052	.0591	.0091	.0193
		REQL	.0317	.0052	.0592	.0091	.0174
		GEE2-G	.0333	.0055	.0665	.0099	.0196
		GEE2-I	.1530	.0200	.1668	.0308	8.0767
	MA(1)	EGEE	.0285	.0047	.0623	.0094	.0280
		REGEE	.0285	.0047	.0623	.0094	.0154
		EQL	.0285	.0047	.0623	.0094	.0167
		REQL	.0285	.0047	.0623	.0094	.0117
		GEE2-G	.0312	.0050	.0687	.0102	.0284
		GEE2-I	.0387	.0061	.0898	.0127	.0223
AR(1)	Indep.	EGEE	.0238	.0041	.0528	.0079	.0148
		REGEE	.0238	.0041	.0528	.0079	.0114
		EQL	.0238	.0041	.0528	.0079	.0114
		REQL	.0238	.0041	.0528	.0079	.0136
		GEE2-G	.0257	.0044	.0569	.0084	.0150
		GEE2-I	.0257	.0044	.0569	.0084	.0150
	Unspec.	EGEE	.0225	.0040	.0483	.0076	.0125
		REGEE	.0225	.0040	.0482	.0075	.0120
		EQL	.0223	.0039	.0482	.0076	.0119
		REQL	.0223	.0039	.0482	.0076	.0155
		GEE2-G	.0244	.0042	.0559	.0086	.0124
		GEE2-I	.0505	.0063	.1045	.0170	3.8770

Table 2 (continued)

True **R**	Fitted **R**	Method	Parameter				
			β_1	β_2	β_3	β_4	ϕ
AR(1)	Exch.	EGEE	.0232	.0038	.0498	.0078	.0511
		REGEE	.0232	.0038	.0498	.0078	.0139
		EQL	.0232	.0038	.0498	.0078	.0318
		REQL	.0232	.0038	.0498	.0078	.0180
		GEE2-G	.0261	.0043	.0603	.0091	.0510
		GEE2-I	.1122	.0124	.1949	.0324	4.8007
	MA(1)	EGEE	.0227	.0039	.0426	.0067	.0219
		REGEE	.0227	.0039	.0426	.0067	.0120
		EQL	.0227	.0039	.0426	.0067	.0398
		REQL	.0227	.0039	.0425	.0067	.0101
		GEE2-G	.0241	.0041	.0494	.0075	.0216
		GEE2-I	.0288	.0049	.0642	.0092	.0324

Based on the results summarized in Tables 1 and 2, we make several observations. When the working correlation structure was taken to be independence, GEE2-I and GEE2-G were equivalent. Otherwise, GEE2-I produced MSEs that were similar to, or much larger than, those produced by GEE2-G. In several cases, the GEE2-I results were very poor. For these reasons, the independence working structure for third and fourth order moments in GEE2 is not recommended and we give no further consideration to this method in the simulation studies of this paper.

GEE2-G consistently yielded larger MSEs for $\boldsymbol{\beta}$ than the methods that treat first and second moment parameters orthogonally (EGEE, EQL and their REML-like modifications). While this deficiency was often counter-balanced by more accurate estimation of ϕ than EGEE and EQL, these gains in the estimation of the dispersion parameter were slight in comparison to losses with respect to the parameter typically of primary interest, $\boldsymbol{\beta}$. In addition, REML-EGEE MSEs for all first and second moment parameters were always similar to, or substantially smaller than, those based on GEE2-G.

EGEE, REML-EGEE, EQL, and REML-EQL were practically equivalent for the estimation of $\boldsymbol{\beta}$. With respect to this parameter, these methods were always as good or substantially better than the GEE2 methods.

When the true correlation structure was correctly modelled, both REML-EGEE and REML-EQL were superior to their non-REML counterparts with respect to the estimation of $\boldsymbol{\alpha}$ in all cases. Surprisingly, the REML modification to EGEE did not always improve the estimation of ϕ, and, for EQL, in Table 1 we see that the REML modification always increased the MSEs for ϕ. In the incorrect working correlation matrix cases reported in Table 2, the REML modification to EGEE typically improved the estimator of ϕ, often by a substantial amount. This was also unexpected, given that the REML modification is derived under the assumption of a correctly specified covariance structure. The REML modification to EQL produced mixed results with respect to ϕ in Table 2.

In the correctly specified covariance structure case, EQL and REML-EQL did yield more accurate estimates of $\boldsymbol{\alpha}$ in comparison to EGEE and REML-EGEE. However, these gains were modest, particularly in the comparison of the REML-like approaches, and they are offset by cases of poorer performance with respect to ϕ. In the incorrectly specified covariance structure case of Table 2, a comparison of the EQL-based approaches versus the EGEE-based approaches to the estimation of ϕ is somewhat inconclusive.

Although these results do not demonstrate a clear advantage to using EQL over EGEE, it is important to keep in mind the nature of the simulated data. These results were based on data generated from model (21), which involves a regression parameter vector $\boldsymbol{\beta} = (.21, .96, -1.11, .41)'$ which corresponds to an EGEE fit to the original data. Given the relatively large number of zeros and other small counts in those data, this example is not one to which EQL is particularly well-suited (25% of the seizure counts in the epilepsy data set are less than 3). Under these circumstances, it is surprising how well EQL and REML-EQL estimators performed.

5.2 Overdispersed and Correlated Binomial Data.

The second simulation study is based on data from a study of the effects of salinity and temperature on the hatch rate for fish eggs (Alderice & Forrester 1968). These data are presented and analyzed in Lindsey (1999a); see also Lindsey (1999b). In this study, 18 tanks, each divided into four cells, were assigned to 17 different combinations of temperature and salinity (in one case, two tanks received the same treatment). In each cell, the hatch proportion from a large number of English sole (*Parophrys vetulus*) eggs (mean = 622.9, standard deviation = 103.3) was recorded. These data exhibit substantial overdispersion relative to a model based on an assumption of independent binomials. This is likely due to dependence among the cells within a given tank, and may also result from heterogeneity in the egg-specific hatch outcomes both at the cell level and the tank level. Lindsey (1999b) makes use of a beta-binomial model to analyze these data. Although his model accounts for overdispersion in the cell-specific proportions and correlation among the egg-specific hatch outcomes within a given cell, correlation among the cells within the same tank is ignored. A marginal GEE-type model for these data provides greater flexibility to separately model within tank correlation and overdispersion.

For simulations, we generated data based on the structure of the fish egg data set where the probability π_{ij} associated with the j-th cell in the i-th tank was assumed to follow the model

$$\text{logit}(\pi_{ij}) = -278 + 110x_{1i} - 37x_{1i}^2 + 234x_{2i} - 88x_{2i}^2 + 18x_{1i}x_{2i},$$

where $x_{1i} = \text{salinity}^{0.2}$ and $x_{2i} = \text{temperature}^{0.2}$. This is the response surface model considered by Lindsey (1999b) with regression coefficients taken from an EGEE fit to the original data assuming an exchangeable correlation structure. For each of several covariance structure specifications, 500 data sets were generated, each consisting of $K = 18$ four-variate observation vectors $\mathbf{y}_i$ with elements $y_{ij} \in \{0, 1, \ldots, m_{ij}\}$ with mean $\mu_{ij} = m_{ij}\pi_{ij}$. To speed up the data generation, we used values of m_{ij} equal to one sixth the number of eggs (rounded to the nearest integer) in the (i, j)-th cell in the original data set. Marginally, $\mathbf{y}_i$ was generated to have variance-covariance matrix $\mathbf{V}_{i11}(\boldsymbol{\mu}_i, \boldsymbol{\alpha}, \phi)$ as in (2) with binomial variance function and correlation structures that again included the independence, unspecified, exchangeable, and AR(1) structures. This time, however, no misspecified covariance models were examined because of the computationally intensive and time-consuming nature of the simulations. Values of $\boldsymbol{\alpha}$ used to generate the data were the same as reported in Section 5.1.

Although we are not aware of any published algorithms for generating correlated, overdispersed binomials, given a method for generating correlated binary deviates, an obvious technique is as follows. Let $X_1 = Z_{11} + \cdots + Z_{1m_1}$, $X_2 = Z_{21} + \cdots + Z_{2m_2}$, where the Z_{ij}'s are Bernoulli random variables with $E(Z_{1i}) = \pi_1$, $i = 1, \ldots, m_1$, $E(Z_{2j}) = \pi_2$, $j = 1, \ldots, m_2$, $\mathrm{corr}(Z_{1i}, Z_{1i'}) = \rho_{11}$, $i \neq i'$, $\mathrm{corr}(Z_{2j}, Z_{2j'}) = \rho_{22}$, $j \neq j'$, and $\mathrm{corr}(Z_{1i}, Z_{2j}) = \rho_{12}$ $\forall i, j$. Then $\mathrm{var}(X_1) = m_1\pi_1(1-\pi_1)\{1+(m_1-1)\rho_{11}\}$, $\mathrm{var}(X_2) = m_2\pi_2(1-\pi_2)\{1+(m_2-1)\rho_{22}\}$, and

$$\mathrm{corr}(X_1, X_2) = \frac{\sqrt{m_1 m_2}\rho_{12}}{\sqrt{\{1+(m_1-1)\rho_{11}\}\{1+(m_2-1)\rho_{22}\}}}.$$

Therefore, to generate integer-valued deviates X_1, X_2 in $[0, m_i]$, $i = 1, 2$, respectively, with means $m_i\pi_i$, variances $\phi_i m_i \pi_i(1-\pi_i)$, $i = 1, 2$ and correlation δ_{12}, we simply generate a $m_1 + m_2$ dimensional vector $\mathbf{Z} = (Z_{11}, \ldots, Z_{1m_1}, Z_{21}, \ldots, Z_{2m_2})'$ of binary deviates with mean $E(\mathbf{Z}) = (\pi_1 \mathbf{1}_{1,m_1}, \pi_2 \mathbf{1}_{1,m_2})'$ and correlation matrix

$$\begin{pmatrix} \rho_{11}\mathbf{1}_{m_1,m_1} + (1-\rho_{11})\mathbf{I}_{m_1} & \rho_{12}\mathbf{1}_{m_1,m_2} \\ \rho_{12}\mathbf{1}_{m_2,m_1} & \rho_{22}\mathbf{1}_{m_2,m_2} + (1-\rho_{22})\mathbf{I}_{m_2} \end{pmatrix}$$

and then compute $X_i = \sum_j Z_{ij}$, $i = 1, 2$. Here $\mathbf{1}_{a,b}$ is the $a \times b$ matrix of ones, $\mathbf{I}_a$ is the $a-$dimensional identity matrix, $\rho_{ii} = (\phi_i - 1)/(m_i - 1)$, $i = 1, 2$ and $\rho_{12} = \delta_{12}\sqrt{\phi_1\phi_2}/\sqrt{m_1 m_2}$, Extension to generate greater than two such deviates follows in an obvious way. We utilized the algorithm of Emrich & Piedmonte (1991) to generate the correlated binary deviates.

Table 3: Binomial simulation results. MSEs for EGEE, REML-EGEE (REGEE), EQL, REML-EQL (REQL), and GEE2 with Gaussian working higher moment structure (GEE2-G). Smallest values are underlined.

True $\mathbf{R}$	Method	Parameter β_1	β_2	β_3	β_4	β_5	β_6	$\boldsymbol{\alpha}'$	ϕ
Indep.	EGEE	$\underline{351.6}$	$\underline{102.1}$	$\underline{6.244}$	$\underline{331.5}$	$\underline{36.46}$	$\underline{16.49}$		.5133
	REGEE	$\underline{351.6}$	$\underline{102.1}$	$\underline{6.244}$	$\underline{331.5}$	$\underline{36.46}$	$\underline{16.49}$		.5065
	EQL	$\underline{351.6}$	$\underline{102.1}$	$\underline{6.244}$	$\underline{331.5}$	$\underline{36.46}$	$\underline{16.49}$		$\underline{.4722}$
	REQL	$\underline{351.6}$	$\underline{102.1}$	$\underline{6.244}$	$\underline{331.5}$	$\underline{36.46}$	$\underline{16.49}$		.4844
	GEE2-G	369.4	107.0	6.628	359.3	40.30	19.20		.5048
Unspec.	EGEE	637.2	173.9	10.09	629.3	66.04	29.35	a	.6970
	REGEE	627.8	172.5	10.03	620.6	64.97	29.27	b	.6723
	EQL	634.8	173.3	10.01	627.0	65.57	29.23	c	$\underline{.6391}$
	REQL	$\underline{626.8}$	$\underline{172.1}$	$\underline{9.956}$	$\underline{620.1}$	$\underline{64.78}$	$\underline{29.13}$	d	.6592
	GEE2-G	669.5	179.6	10.56	676.2	72.89	33.68	e	.6951
Exch.	EGEE	626.9	171.6	9.725	616.9	68.09	28.20	.0379	.8222
	REGEE	624.7	$\underline{170.7}$	9.728	615.5	67.97	$\underline{28.06}$	.0211	.7029
	EQL	626.8	171.5	$\underline{9.724}$	616.8	68.09	28.20	.0364	.7701
	REQL	$\underline{624.5}$	$\underline{170.7}$	9.728	$\underline{615.4}$	$\underline{67.96}$	28.07	$\underline{.0209}$	$\underline{.6786}$
	GEE2-G	660.0	179.3	10.20	682.7	77.46	31.66	.0371	.8385
AR(1)	EGEE	710.0	191.1	11.59	676.0	74.45	27.77	.0312	.9640
	REGEE	$\underline{708.8}$	$\underline{190.8}$	$\underline{11.49}$	673.6	74.10	$\underline{27.60}$	.0159	.7732
	EQL	709.8	191.0	11.59	675.7	74.41	27.77	.0296	.8965
	REQL	$\underline{708.8}$	$\underline{190.8}$	11.50	$\underline{673.5}$	$\underline{74.06}$	27.62	$\underline{.0152}$	$\underline{.7628}$
	GEE2-G	766.3	198.6	12.45	759.2	85.36	32.96	.0301	.9521

a = (.0835, .0774, .0816, .0871, .0794, .0974); b = (.0583, .0551, .0626, .0590, .0569, .0690);
c = (.0812, .0737, .0792, .0836, .0758, .0935); d = ($\underline{.0574}$, $\underline{.0532}$, $\underline{.0616}$, $\underline{.0572}$, $\underline{.0552}$, $\underline{.0671}$);
e = (.0822, .0762, .0807, .0863, .0787, .0964).

Results from these simulations appear in Table 3. For all four correlation structures, MSEs for the regression parameter $\boldsymbol{\beta}$ were similar for EGEE, REML-EGEE, EQL, and REML-EQL, but substantially higher for GEE2-G. With respect to α and ϕ, there is a fairly consistent pattern that EQL provides a modest improvement in MSE over EGEE and GEE2-G, and the REML versions of EGEE and EQL provide substantial reductions in MSE that are large enough to improve the quality of the corresponding estimators of $\boldsymbol{\beta}$.

5.3 Correlated Binary Data.

The third simulation study is based on longitudinal data from a clinical trial that compared two treatments for a respiratory illness. These data, presented and analyzed using GEE1 in Stokes *et al.* (1995), include a dichotomous response, $Y = 1$ (respiratory status "good") or $Y = 0$ (respiratory status "poor"), measured at four occasions for each of $K = 111$ subjects recruited at two different centers. In our simulation study, we generated 500 data sets based on these data to conform to the logistic regression model

$$\text{logit}(\pi_{ij}) = 0.35 - 1.14\text{TRT}_{ij} + 1.87\text{BASE}_{ij} - 0.55\text{CENTER}_{ij},$$

where $1 \le i \le 111$, $1 \le j \le 4$, $\text{TRT}_{ij}, \text{BASE}_{ij}, \text{CENTER}_{ij}$ are indicator variables for placebo treatment, "good" respiratory status at baseline, and center 1, respectively. The true regression coefficients, $\boldsymbol{\beta} = (.35, -1.14, 1.87, -.55)'$, were taken from an EGEE fit to the original data set. Again, the independence, unspecified, exchangeable, and AR(1) marginal correlation structures were used to generate and model the data. In this case, correlated binary data were generated using the algorithm of Park *et al.* (1996) based on the same true values of $\boldsymbol{\alpha}$ as were used in Sections 5.1 and 5.2. Results appear in Table 4.

TABLE 4: Binary data simulation results. MSEs for EGEE, REML-EGEE (REGEE), EQL, REML-EQL (REQL), and GEE2 with Gaussian working higher moment structure (GEE2-G). Smallest values are underlined.

		Parameter				
True **R**	Method	β_1	β_2	β_3	β_4	$\boldsymbol{\alpha}'$
Unspec.	EGEE	.1218	.1346	.1493	<u>.1347</u>	*a*
	REGEE	.1217	.1345	.1492	<u>.1347</u>	*b*
	EQL	<u>.1213</u>	<u>.1342</u>	<u>.1486</u>	.1353	*c*
	REQL	<u>.1213</u>	<u>.1342</u>	<u>.1486</u>	.1352	*d*
	GEE2-G	.1701	.1876	.1601	.2233	*e*
Exch.	EGEE	<u>.1235</u>	<u>.1435</u>	.1564	<u>.1210</u>	.0064
	REGEE	<u>.1235</u>	<u>.1435</u>	.1564	<u>.1210</u>	.0061
	EQL	<u>.1235</u>	<u>.1435</u>	.1564	<u>.1210</u>	.0052
	REQL	<u>.1235</u>	<u>.1435</u>	.1564	<u>.1210</u>	<u>.0049</u>
	GEE2-G	.1751	.1745	<u>.1506</u>	.1986	.0054
AR(1)	EGEE	<u>.0976</u>	<u>.1072</u>	.1134	<u>.1025</u>	.0038
	REGEE	<u>.0976</u>	<u>.1072</u>	.1134	<u>.1025</u>	.0036
	EQL	.0977	<u>.1072</u>	<u>.1133</u>	.1026	.0036
	REQL	<u>.0976</u>	<u>.1072</u>	<u>.1133</u>	.1026	<u>.0033</u>
	GEE2-G	.1566	.1576	.1300	.1649	.0036

a = (.0125, .0132, .0121, .0114, .0131, .0118); b = (.0120, .0126, .0115, .0111, .0124, .0113); c = (.0104, .0103, .0096, .0092, .0110, .0101); d = (<u>.0099</u>, <u>.0098</u>, <u>.0091</u>, <u>.0089</u>, <u>.0104</u>, <u>.0096</u>); e = (.0120, .0120, .0110, .0108, .0123, .0111).

Results in Table 4 are generally quite similar to those in Table 3 based on binomial data. GEE2-G performs appreciably more poorly than the other methods with respect to $\boldsymbol{\beta}$. In the binary response context, overdispersion is not possible, so ϕ is necessarily 1 and need not be estimated. For $\boldsymbol{\alpha}$, though, EQL provides slightly better estimates than EGEE and both REML-like methods provide some reduction in MSE as well.

6. DISCUSSION

The simulation results suggest that within the guidelines of Smyth & Verbyla (1999) described in Section 4, EQL is capable of producing more accurate estimates of second moment parameters than EGEE and GEE2. Furthermore, even for data that violate Smyth and Verbyla's rule of thumb (e.g., the epileptic seizure data, binary data of the sort simulated in section 5.3), EQL can perform quite well with respect to estimation of the correlation parameter. These gains come without compromising either the consistency or efficiency of estimators of $\boldsymbol{\beta}$. However, the magnitude of the gains offered by EQL may not be sufficient to justify the increased computational complexity of the method and to compensate for the lack of a general consistency result for second moment parameter estimators. EGEE appears to perform well with respect to estimation of both first and second moment parameters across a range of situations. In addition, when the second moment structure is known, the REML-like adjustment to EGEE proposed in Section 3 often results in MSEs than are nearly as small or smaller than those obtained with EQL and GEE2. Therefore, we recommend EGEE for general use and EQL only for particular situations in which the observed data are not close to their boundary of support and second moment inference is of substantial interest.

APPENDIX

Here we present the notation for expression (19). Carlson's standard elliptic integrals of the first and third kind are defined as

$$\begin{aligned} R_F(x,y,z) &= \frac{1}{2}\int_0^\infty \{(t+x)(t+y)(t+z)\}^{-1/2}dt, \quad \text{and} \\ R_J(x,y,z,\rho) &= \frac{3}{2}\int_0^\infty \{(t+x)(t+y)(t+z)\}^{-1/2}(t+\rho)^{-1}dt, \end{aligned}$$

respectively. Other quantities in expression (19) are defined as follows:

$$\begin{aligned} U_{12} &= \{\mu_{ij}(1-\mu_{ij})y_{ik}(1-y_{ik})\}^{1/2} + \{y_{ij}(1-y_{ij})\mu_{ik}(1-\mu_{ik})\}^{1/2}, \\ U_{13} &= \{\mu_{ij}\mu_{ik}(1-y_{ij})(1-y_{ik})\}^{1/2} + \{y_{ij}y_{ik}(1-\mu_{ij})(1-\mu_{ik})\}^{1/2}, \\ U_{14} &= \{\mu_{ij}(1-\mu_{ik})(1-y_{ij})y_{ik}\}^{1/2} + \{y_{ij}(1-y_{ik})(1-\mu_{ij})\mu_{ik}\}^{1/2}, \end{aligned}$$

$$\begin{aligned} d_{12} &= y_{ij}(y_{ij}-\mu_{ij}) - (1-y_{ij})(\mu_{ij}-y_{ij}), \\ d_{13} &= y_{ij}(\mu_{ik}-y_{ik}) - y_{ik}(\mu_{ij}-y_{ij}), \\ d_{14} &= y_{ij}(y_{ik}-\mu_{ik}) - (1-y_{ik})(\mu_{ij}-y_{ij}), \end{aligned}$$

$$W_1^2 = U_{12}^2 - \frac{(y_{ij}-\mu_{ij})d_{13}d_{14}}{\mu_{ij}-y_{ij}}\,, \quad Q_1^2 = \frac{W_1^2}{\mu_{ij}y_{ij}}\,,$$

$$P_1^2 = Q_1^2 + \frac{(y_{ij}-\mu_{ij})(\mu_{ik}-y_{ik})(y_{ik}-\mu_{ik})}{\mu_{ij}-y_{ij}}\,.$$

ACKNOWLEDGEMENTS

The author wishes to thank Ishwar Basawa, Tom Severini, and Anand Vidyashankar for discussions which led to improvements in the paper.

REFERENCES

M. Abramowitz & I. A. Stegun (1964). *Handbook of Mathematical Functions*, National Bureau of Standards, Applied Mathematics Series, Number 55. U.S. Government Printing Office, Washington, DC.

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

B. C. Carlson (1988). A table of elliptic integrals of the third kind. *Mathematics of Computation*, **51**, 267–280.

D. R. Cox & N. M. Reid (1987). Parameter orthogonality and approximate conditional inference (with discussion). *Journal of the Royal Statistical Society Series B*, **49**, 1–39.

M. Crowder (1985). Gaussian estimation for correlated binomial data. *Journal of the Royal Statistical Society Series A*, **47**, 229–237.

M. Crowder (1986). On consistency and inconsistency of estimating equations. *Econometric Theory*, **2**, 305–330.

M. Crowder (1992). Comments on "Multivariate regression analyses for categorical data." *Journal of the Royal Statistical Society Series B*, **54**, 26–27.

M. Crowder (1995). On the use of a working correlation matrix in using generalised linear models for repeated measures. *Biometrika*, **82**, 407–10.

M. Davidian & R. J. Carroll (1988). A note on extended quasi-likelihood. *Journal of the Royal Statistical Society Series B*, **50**, 74–82.

L. J. Emrich & M. R. Piedmonte (1991). A method of generating high-dimensional multivariate binary deviates. *The American Statistician*, **45**, 302–304.

V. P. Godambe (1991). Orthogonality of estimating functions and nuisance parameters. *Biometrika*, **78**, 143–151.

V. P. Godambe (1992). Comments on "Multivariate regression analyses for categorical data." *Journal of the Royal Statistical Society Series B*, **54**, 33.

V. P. Godambe & M.E. Thompson (1989). An extension of quasi-likelihood estimation. *Journal of Statistical Planning and Inference*, **22**, 137–152.

C. Gouriéroux, A. Monfort & A. Trognon (1984). Pseudomaximum likelihood methods: theory. *Econometrica*, **52**, 681–700.

D. B. Hall & R. L. Bailey (2000). *Estimation and Prediction of Forest Growth Variables Based on Multilevel Nonlinear Mixed Models.* Department of Statistics Technical Report #2000-5, University of Georgia, Atlanta, GA.

D. B. Hall & T. A. Severini (1998). Extended generalized estimating equations for clustered data. *Journal of the American Statistical Association*, **93**, 1365–1375.

Y. Lee & J. A. Nelder (1998). Generalized linear models for the analysis of quality-improvement experiments. *The Canadian Journal of Statistics*, **26**, 95–105.

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

K. Y. Liang, S. L. Zeger & B. Qaqish (1992). Multivariate regression analyses for categorical data (with discussion). *Journal of the Royal Statistical Society Series B*, **54**, 3–40.

J. K. Lindsey (1999a). *Models for Repeated Measurements*, Second Edition. Oxford University Press, Oxford.

J. K. Lindsey (1999b). Response surfaces for overdispersion in the study of the conditions for fish eggs hatching. *Biometrics*, **55**, 149–155.

J. R. Magnus & H. Neudecker (1979). The commutation matrix: some properties and applications. *The Annals of Statistics*, **7**, 381–394.

P. J. McCullagh (1991). Quasi-likelihood and estimating functions. In *Statistical Theory and Modelling*, D. V. Hinkley, N. M. Reid Ẽ. J. Snell (Eds) pp. 265–28. Chapman and Hall, London.

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

J. A. Nelder & Y. Lee (1992). Likelihood, quasi-likelihood and pseudolikelihood: some comparisons. *Journal of the Royal Statistical Society Series B*, **54**, 273–284.

J. A. Nelder & Y. Lee (1997). Extended quasi-likelihood and estimating equations. In *Selected Proceedings of the Symposium on Estimating Functions*, I. V. Basawa, V. P. Godambe & and R. L. Taylor (Eds) pp. 139–148. Institute of Mathematical Statistics, Hayward, CA.

J. A. Nelder & D. Pregibon (1987). An extended quasi-likelihood function. *Biometrika*, **74**, 221–231.

C. G. Park, T. Park & D. W. Shin (1996). A simple method for generating correlated binary variates. *The American Statistician*, **50**, 306–310.

S. R. Paul & A. S. Islam (1998). Joint estimation of the mean and dispersion parameters in the analysis of proportions: a comparison of efficiency and bias. *The Canadian Journal of Statistics*, **26**, 83–94.

R. L. Prentice & L. P. Zhao (1991). Estimating equations for parameters in means and covariances of multivariate discrete and continuous responses. *Biometrics*, **47**, 825–839.

W. H. Press, S .A. Teukolsky, W. T. Vetterling & B. P. Flannery (1992). *Numerical Recipes in C*, Second Edition. Cambridge University Press, Cambridge.

S. R. Searle (1982). *Matrix Algebra Useful for Statistics*. Wiley, New York.

C. H. Sim (1993). Generation of Poisson and gamma random vectors with given marginals and covariance matrix. *Journal of Statistical Computation and Simulation*, **47**, 1–10.

M. E. Stokes, C. S. Davis & G. G. Koch (1995). *Categorical Data Analysis Using the SAS System*. SAS Institute Inc, Cary, NC.

G. K. Smyth & A. P. Verbyla (1999). Adjusted likelihood methods for modelling dispersion in generalized linear models. *Environmetrics*, **10**, 695–709.

Y. Takagi & N. Inagaki (1993). Estimating function with asymptotic bias and its estimator. *Annals of the Institute of Statistical Mathematics*, **45**, 499–510.

P. F. Thall, P. F. & S .C. Vail (1990). Some covariance models for longitudinal count data with overdispersion. *Biometrics*, **46**, 657–671.

R. W. M. Wedderburn (1974). Quasi-likelihood functions, generalized linear models, and the Gauss-Newton method. *Biometrika*, **61**, 439–47.

R. D. Wolfinger & X. Lin (1997). Two Taylor-series approximation methods for nonlinear mixed models. *Computational Statistics and Data Analysis*, **25**, 465–490.

Received 17 April 2000
Accepted 18 August 2000

Daniel B. HALL: `dhall@stat.uga.edu`
Department of Statistics
University of Georgia
Athens, GA 30602-1952, USA