Extended Generalized Estimating Equations for Clustered Data

Daniel B. HALL and Thomas A. SEVERINI

Typically, analysis of data consisting of multiple observations on a cluster is complicated by within-cluster correlation. Estimating equations for generalized linear modeling of clustered data have recently received much attention. This article proposes an extension to the generalized estimating equation method proposed by Liang and Zeger, which treats within-cluster correlations as nuisance parameters. Using ideas from extended quasi-likelihood, estimating equations for regression and association parameters are provided simultaneously. The resulting estimators are proven to be asymptotically normal and consistent under certain conditions. The consistency of regression estimators allows incorrect modeling of the correlation among repeated responses. The method is illustrated with an analysis of data from a developmental toxicity study.

KEY WORDS: Correlation; Extended quasi-likelihood; Generalized linear model; Longitudinal data; Marginal model; Quasi-likelihood.

1. INTRODUCTION

The analysis of clustered data is complicated by the correlation that typically exists among observations within the same cluster. The introduction of generalized estimating equations (GEEs) for the analysis of such data has stimulated a great deal of interest during the last decade. The GEE approach was originally proposed for the situation in which it is reasonable to assume that the marginal mean response conforms to a generalized linear model (GLM) (Nelder and Wedderburn 1972) where the coefficients of the linear model are of primary interest. In this setting, the within-cluster correlation is treated as a nuisance. GEEs arise naturally as an application of quasi-likelihood (QL) estimation (Wedderburn 1974) of the regression parameters when correlation parameters are consistently estimated separately.

As noted by Liang, Zeger, and Qaqish (1992), the original GEE approach (hereafter referred to as GEE1, following the usage of Liang et al. 1992) can yield estimators of the correlation parameters that have low asymptotic efficiency relative to maximum likelihood. This should not be surprising, because GEE1 treats correlations as a nuisance and estimates them using consistent, but ad hoc estimators. An alternative approach (GEE2) has been suggested by Prentice and Zhao (1991) and Zhao and Prentice (1990). These authors suggested estimating equations for the simultaneous estimation of regression and association parameters. These estimating equations arise from a pseudo–maximum likelihood approach to the estimation of the mean and covariance parameters when the responses are assumed to follow a quadratic exponential model (Gourieroux, Monfort, and Trognon 1984). Although GEE2 estimators of the association parameters are considerably more efficient than those obtained using GEE1 (Liang et al. 1992), this improvement comes at some cost. The consistency of GEE2 regression parameter estimators requires correct specification of the covariance structure. In addition, an assumption must be made concerning the structure of the covariance matrix of the vector of observed responses and covariances. That is, a "working" model for third- and fourth-order moments is necessary. Although the correctness of this working model does not affect the consistency of GEE2 regression and association parameter estimators, it does affect their efficiency.

This article proposes an alternative set of estimating equations for the simultaneous estimation of regression and association parameters. These estimating equations are motivated using the ideas of extended QL (McCullagh and Nelder 1989; Nelder and Pregibon 1987) rather than pseudo–maximum likelihood. In this way the approach proposed here is a natural extension of the original GEE methodology. The method makes only first- and second-moment assumptions and does not require a correct covariance specification for the consistency of its regression parameter estimators.

2. GENERALIZED ESTIMATING EQUATIONS FOR REGRESSION PARAMETERS

To establish notation, suppose that a scalar response y_{it} and a p-dimensional vector of covariates $\mathbf{x}_{it}$ are observed for clusters $i=1,\dots,K$ and cluster subunits $t=1,\dots,T_i$. For the ith cluster, let $\mathbf{y}_i = (y_{i1},\dots,y_{iT_i})^T$ be the response vector with mean $\boldsymbol{\mu}_i = (\mu_{i1},\dots,\mu_{iT_i})^T$ and let $\mathbf{x}_i = (x_{i1},\dots,x_{iT_i})^T$ be the $T_i \times p$ matrix of covariates. Now suppose that $g(\mu_{it}) = \mathbf{x}_{it}^T\boldsymbol{\beta}$ and $\text{var}(\mathbf{y}_i) = \phi\mathbf{V}_i(\boldsymbol{\mu}_i,\boldsymbol{\alpha})$, where $\mathbf{V}_i$ is a $T_i \times T_i$ covariance matrix with the structure

$$\mathbf{V}_i(\boldsymbol{\mu}_i,\boldsymbol{\alpha}) = \mathbf{A}_i^{1/2}(\boldsymbol{\mu}_i)\mathbf{R}(\boldsymbol{\alpha})\mathbf{A}_i^{1/2}(\boldsymbol{\mu}_i). \qquad (1)$$

Here $\mathbf{A}_i(\boldsymbol{\mu}_i) = \text{diag}(v_i(\mu_{i1}),\dots,v_i(\mu_{1T_i}))$ is a $T_i \times T_i$ diagonal matrix of variance functions, $\mathbf{R}(\boldsymbol{\alpha})$ is the parametric form chosen for the working correlation matrix of $\mathbf{y}_i$, and $\boldsymbol{\alpha}$ is an $s \times 1$ vector of unknown parameters. This formulation leads to estimating equations analogous to the joint

Daniel B. Hall is Assistant Professor, Department of Statistics, University of Georgia, Athens, GA 30602 (E-mail: dhall@stat.uga.edu). Thomas A. Severini is Associate Professor, Department of Statistics, Northwestern University, Evanston, IL 60208 (E-mail: severini@nwu.edu). This research was supported in part by National Institute of Mental Health grant MH15168-18.

Journal of the American Statistical Association
December 1998, Vol. 93, No. 444, Theory and Methods

quasi-score equations under independence,

$$\frac{1}{\phi}\sum_{i=1}^{K}\mathbf{D}_i^T\mathbf{V}_i^{-1}(\mathbf{y}_i-\boldsymbol{\mu}_i)=\mathbf{0}, \tag{2}$$

where $\mathbf{D}_i=\partial\boldsymbol{\mu}_i/\partial\boldsymbol{\beta}$. Notice that the function $U_i(\boldsymbol{\beta},\boldsymbol{\alpha},\phi)\equiv\phi^{-1}\mathbf{D}_i^T\mathbf{V}_i^{-1}(\mathbf{y}_i-\boldsymbol{\mu}_i)$ in (2) depends on unknown parameters $\boldsymbol{\beta}$, $\boldsymbol{\alpha}$, and ϕ. Therefore, Liang and Zeger (1986) proposed estimating equations for $\boldsymbol{\beta}$ based on (2) where ϕ is replaced by estimator $\tilde{\phi}(\boldsymbol{\beta})$, which is $K^{1/2}$ consistent given $\boldsymbol{\beta}$, and $\boldsymbol{\alpha}$ is replaced by an estimator $\tilde{\alpha}(\boldsymbol{\beta},\phi)$, which is $K^{1/2}$ consistent given $\boldsymbol{\beta}$ and ϕ. These substitutions led to Liang and Zeger's original GEEs,

$$\sum_{i=1}^{K}U_i[\boldsymbol{\beta},\tilde{\alpha}\{\boldsymbol{\beta},\tilde{\phi}(\boldsymbol{\beta})\},\tilde{\phi}(\boldsymbol{\beta})]=\mathbf{0}. \tag{3}$$

These authors indicated that under mild regularity conditions, $K^{1/2}(\tilde{\boldsymbol{\beta}}-\boldsymbol{\beta})$ is asymptotically $\mathrm{N}(\mathbf{0},\mathbf{V}_1)$, where $\tilde{\boldsymbol{\beta}}$ is the solution of (3) and $\mathbf{V}_1$ equals

$$\lim_{K\to\infty}K\left(\sum_{i=1}^{K}\mathbf{D}_i^T\mathbf{V}_i^{-1}\mathbf{D}_i\right)^{-1}\times\left\{\sum_{i=1}^{K}\mathbf{D}_i^T\mathbf{V}_i^{-1}\operatorname{var}(\mathbf{y}_i)\mathbf{V}_i^{-1}\mathbf{D}_i\right\}\left(\sum_{i=1}^{K}\mathbf{D}_i^T\mathbf{V}_i^{-1}\mathbf{D}_i\right)^{-1}.$$

A consistent estimator $\tilde{\mathbf{V}}_1$ of this covariance matrix can be obtained by replacing $\operatorname{var}(\mathbf{y}_i)$ with $(\mathbf{y}_i-\boldsymbol{\mu}_i)(\mathbf{y}_i-\boldsymbol{\mu}_i)^T$ and $\boldsymbol{\beta}$, $\boldsymbol{\alpha}$, and ϕ with their estimates in the expression for $\mathbf{V}_1$.

Liang and Zeger (1986) suggested estimating the unknown correlation parameters, $\boldsymbol{\alpha}$ and ϕ, using method-of-moments estimators based on the Pearson residuals, $r_{it}=(y_{it}-\hat{\mu}_{it})/v_i(\hat{\mu}_{it})^{1/2}$. The exact form of these residual-based estimators depends on the assumed form of $\mathbf{R}(\boldsymbol{\alpha})$. Examples for several correlation structures have been considered by Liang and Zeger (1986). Alternatively, Prentice (1988) and Prentice and Zhao (1991) suggested ad hoc estimating equations for the association parameters. This approach generalizes Liang and Zeger's method-of-moments techniques, because the residual-based parameter estimators considered by Liang and Zeger (1986) can be obtained as the solution of appropriately chosen estimating equations. One advantage of using such second-moment estimating equations is that it allows for modeling in the association parameters. For example, correlations can be assumed to be related to covariates through a generalized linear model; such an approach is considered in Section 5. Another advantage is that greater efficiency can be achieved in the estimation of second-moment parameters over the method-of-moments approach (Mancl 1992).

It is important to note that the method-of-moments approach of Liang and Zeger (1986) and the ad hoc estimating equation approach of Prentice and Zhao (1991, sec. 2) are both GEE1 techniques. In these methods, first- and second-moment parameters are treated as though they are orthogonal to one another even when they are not. In contrast, both GEE2 and the extended GEE technique proposed here approach the estimation of $\boldsymbol{\beta}$ and $\boldsymbol{\alpha}$ simultaneously.

3. EXTENDED GENERALIZED ESTIMATING EQUATIONS

As discussed in Section 2, GEE1 may be seen as an application of QL to clustered data. Equation (2) is a quasi-score equation for $\boldsymbol{\beta}$ when $\boldsymbol{\alpha}$ and ϕ are known. When significant scientific interest is focused on the association among responses, this approach is inadequate, because "GEE1 can be extremely inefficient for the estimation of α" (Liang et al. 1992, p. 16). As detailed by Liang et al. (1992), when $\operatorname{var}(\mathbf{y}_i)$, $i=1,\ldots,K$, is correctly modeled, GEE2 provides estimators of $\boldsymbol{\beta}$ and $\boldsymbol{\alpha}$ that outperform GEE1-based estimators in terms of asymptotic efficiency relative to maximum likelihood. However, GEE2 has two major drawbacks: third- and fourth-moment assumptions on the responses are required, and correct specification of the mean and covariance models is necessary to ensure consistent estimation of $\boldsymbol{\beta}$ and $\boldsymbol{\alpha}$ (Prentice and Zhao 1991). This section introduces an alternative set of estimating equations that, like GEE2, efficiently estimate $\boldsymbol{\alpha}$ but, unlike GEE2, do not require correct covariance specification for the consistent estimation of $\boldsymbol{\beta}$.

3.1 Form of Extended Generalized Estimating Equations

The estimating equations are presented here using index notation (e.g., McCullagh 1987). Let $\gamma=(\boldsymbol{\beta}^T,\boldsymbol{\alpha}^T)^T$, where now the overdispersion parameter ϕ is included in $\boldsymbol{\alpha}$. For the ith individual, let

$$U_i(\gamma;\beta^b)=(y^j-\mu^j)V_{ja}\mu^{a,b},\qquad b=1,\ldots,p$$

and

$$U_i(\gamma;\alpha^u)=-(y^j-\mu^j)(y^k-\mu^k)V_{jk}^u+V^{jk}V_{jk}^u,\qquad u=1,\ldots,s \tag{4}$$

be the extended generalized estimating functions for the components of β and α. Here the superscripts on β, α, y, and μ indicate vector components; V^{jk} and V_{jk} are the (j,k)th components of $\mathbf{V}_i(\boldsymbol{\mu}_i,\boldsymbol{\alpha})=\operatorname{var}(\mathbf{y}_i)$ and $\mathbf{V}_i^{-1}(\boldsymbol{\mu}_i,\boldsymbol{\alpha})$; $V_{jk}^u=\partial V_{jk}/\partial\alpha^u$; and $\mu^{a,b}=\partial\mu^a/\partial\beta^b$. Then

$$\boldsymbol{\Psi}_K(\gamma)\equiv\frac{1}{K}\sum_{i=1}^{K}\mathbf{U}_i(\gamma)=\mathbf{0} \tag{5}$$

defines an extended generalized estimating equation (EGEE) for γ, where

$$\mathbf{U}_i(\gamma)=\begin{bmatrix}\mathbf{U}_i(\gamma;\boldsymbol{\beta})\\ \mathbf{U}_i(\gamma;\boldsymbol{\alpha})\end{bmatrix}.$$

A motivation for (5) may be found in the ideas of extended QL. For independent observations, $\mathbf{y}=(y_1,\ldots,y_K)^T$, with mean $\boldsymbol{\mu}(\boldsymbol{\beta})$ and covariance matrix $\phi\mathbf{V}(\boldsymbol{\mu})$, McCullagh and Nelder (1989) and Nelder and Pregibon (1987) have discussed an extended QL function Q^+ that has likelihood-like properties with respect to both $\boldsymbol{\beta}$ and the overdispersion parameter ϕ. Nelder and Lee (1992) have provided Monte Carlo evidence that favors an extended QL approach toward estimating dispersion parameters in the univariate case. In the clustered data setting, $\mathbf{y} = (\mathbf{y}_1^T, \ldots, \mathbf{y}_K^T)^T$, where $\mathbf{y}_i$ is a vector, $\mathbf{y}_i = (y_{i1}, \ldots, y_{iT_i})^T$, and $\text{var}(\mathbf{y}) = \oplus_{i=1}^K \mathbf{V}_i(\boldsymbol{\mu}_i, \boldsymbol{\alpha})$. An extended QL for $\mathbf{y}$ having likelihood-like properties with respect to $\boldsymbol{\beta}$ and $\boldsymbol{\alpha}$ may be obtained by defining such extended QL functions for each vector of observations $\mathbf{y}_i$ and summing over all i.

Regardless of whether var(y) is diagonal, an equation of the form

$$\mathbf{U}(\boldsymbol{\beta}) \equiv \frac{\partial \boldsymbol{\mu}^T}{\partial \boldsymbol{\beta}} \text{var}^{-1}(\mathbf{y})(\mathbf{y} - \boldsymbol{\mu}) = \mathbf{0}$$

may be used as an estimating equation for $\boldsymbol{\beta}$. However, for $\mathbf{U}(\boldsymbol{\beta})$ to correspond to a QL function, it is necessary and sufficient that the matrix $\partial\mathbf{U}(\boldsymbol{\beta})/\partial\boldsymbol{\beta}$ by symmetric. Because in general var(y) need not be diagonal, the QL function may not exist. Given that a QL function does exist, however, it is given by

$$Q(\boldsymbol{\mu}, \mathbf{y}) = -(\mathbf{y} - \boldsymbol{\mu})^T \left(\int_0^1 s[\mathbf{V}\{\mathbf{t}(s)\}]^{-1} \, ds \right) (\mathbf{y} - \boldsymbol{\mu}), \quad (6)$$

where $\mathbf{t}$ is the straight-line path $\mathbf{t}(s) = \mathbf{y} + (\boldsymbol{\mu} - \mathbf{y})s$, for $0 \le s \le 1$ (McCullagh and Nelder 1989, p. 335).

To obtain the extended QL function $Q^+(\boldsymbol{\mu}, \boldsymbol{\alpha}, \mathbf{y})$, let

$$Q_i^+(\boldsymbol{\mu}_i, \boldsymbol{\alpha}, \mathbf{y}_i) = Q_i(\boldsymbol{\mu}_i, \mathbf{y}_i) + f(\boldsymbol{\alpha}, \mathbf{y}_i), \quad (7)$$

where $f(\boldsymbol{\alpha}, \mathbf{y}_i) = f_1(\boldsymbol{\alpha}) + f_2(\mathbf{y}_i), Q_i(\boldsymbol{\mu}_i, \mathbf{y}_i)$ is as in (6), and $\mathbf{V}_i(\boldsymbol{\mu}_i) = \text{var}(\mathbf{y}_i)$. Dropping the individual index i and switching to index notation, (7) becomes

$$Q^+(\boldsymbol{\mu}, \boldsymbol{\alpha}, \mathbf{y}) = -\int_0^1 sV_{jk}\{\mathbf{t}(s)\}(y^j - \mu^j)(y^k - \mu^k)\, ds + f_1(\boldsymbol{\alpha}) + f_2(\mathbf{y}).$$

For Q^+ to be a QL for $\boldsymbol{\alpha}$ in addition to $\boldsymbol{\beta}$, it is necessary that

$$E\left(\frac{\partial Q^+}{\partial \alpha^u}\right) = 0, \quad \text{where} \quad u = 1, \ldots, s. \quad (8)$$

Assuming that it is possible to differentiate inside the integral,

$$\frac{\partial Q^+}{\partial \alpha^u} = -\int_0^1 sV_{jk}^u\{\mathbf{t}(s)\}(y^j - \mu^j)(y^k - \mu^k)\, ds + \frac{\partial f_1(\boldsymbol{\alpha})}{\partial \alpha^u}.$$

By a Taylor series expansion,

$$E\left(\frac{\partial Q^+}{\partial \alpha^u}\right) = -\int_0^1 s\{V_{jk}^u(\boldsymbol{\mu})\Sigma^{jk} + V_{jk}^l(\boldsymbol{\mu})\Sigma^{jkl} + \cdots\}\, ds + \frac{\partial f_1(\boldsymbol{\alpha})}{\partial \alpha^u}, \quad (9)$$

where $\Sigma^{jk} = E\{(y^j - \mu^j)(y^k - \mu^k)\}$, $\Sigma^{jkl} = E\{(y^j - \mu^j)(y^k - \mu^k)(y^l - \mu^l)\}$, and so on, and $V_{jk}^l = \partial V_{jk}/\partial \mu^l$.

For exponential dispersion models with dispersion parameter ϕ,

$$\kappa_{r+1} = \kappa_r' \kappa_2, \quad \text{for} \quad r \ge 2, \quad (10)$$

where κ_r is the rth cumulant, $\kappa_2 = \phi V(\mu_i)$, and the prime denotes differentiation with respect to μ_i (Jorgensen 1987; Morris 1982). Property (10) implies that $\mu_3 = O(\phi^2), \mu_4 = O(\phi^2), \mu_5 = O(\phi^3), \mu_6 = O(\phi^3), \ldots,$ where μ_r is the rth central moment. Applying Holder's inequality, it is possible to bound the Σ terms in (9) such that $\Sigma^{jkl} = O(\phi^{3/2})$, $\Sigma^{jklm} = O(\phi^2)$, and so on. Ignoring such terms gives

$$E\left(\frac{\partial Q^+}{\partial \alpha^u}\right) \approx -\frac{1}{2} V^{jk} V_{jk}^u + \frac{\partial f_1(\boldsymbol{\alpha})}{\partial \alpha^u}. \quad (11)$$

Hence f_1 should be chosen to satisfy $\partial f_1(\boldsymbol{\alpha})/\partial \alpha^u \approx V^{jk}V_{jk}^u/2$. Therefore, an extended QL function for $\boldsymbol{\beta}$ and also $\boldsymbol{\alpha}$ for the entire response vector $\mathbf{y}$ is

$$Q^+(\boldsymbol{\mu}, \boldsymbol{\alpha}, \mathbf{y}) = -\sum_{i=1}^K (\mathbf{y}_i - \boldsymbol{\mu}_i)^T \times \left(\int_0^1 s[\mathbf{V}_i\{\mathbf{t}(s)\}]^{-1}\, ds \right) (\mathbf{y}_i - \boldsymbol{\mu}_i) + \frac{K}{2} \log[\det\{\mathbf{V}^{-1}(\boldsymbol{\mu}, \boldsymbol{\alpha})\}],$$

where the notation $\det(\mathbf{M})$ indicates the determinant of $\mathbf{M}$. This extended QL function may not exist for all choices of the covariance matrix $\mathbf{V}$. Regardless, the first derivatives of Q^+ given by (4) provide valid estimating functions for $\boldsymbol{\beta}$ and $\boldsymbol{\alpha}$.

3.2 Properties of Extended Generalized Estimating Equations Estimators

Let $\hat{\gamma} = (\hat{\boldsymbol{\beta}}^T, \hat{\boldsymbol{\alpha}}^T)^T$ be the solution of (5) and let γ_0 denote the true γ. The following theorems establish the consistency and asymptotic multivariate normality of $\hat{\gamma}$. Both theorems assume the conditions given in the Appendix and are proven under the simplifying assumption that $T_i = T$, $i = 1, \ldots, K$. The more general case requires additional conditions to those appearing in the Appendix.

Theorem 1. Suppose that the $N \times p$ matrix of derivatives, $\partial\boldsymbol{\mu}/\partial\boldsymbol{\beta}$ has rank p when $\gamma = \gamma_0$; then as $K \to \infty$, $\hat{\gamma} \xrightarrow{p} \gamma_0$.

Theorem 2. The asymptotic distribution of $K^{1/2}(\hat{\gamma} - \gamma_0)$ is multivariate normal with mean 0 and covariance matrix $\mathbf{V}_2$ given by

$$\mathbf{V}_2 = [E\{\mathbf{D}\mathbf{U}_i(\gamma_0)\}]^{-1} \lim_{K \to \infty} K \left[\sum_{i=1}^K \text{var}\{\mathbf{U}_i(\gamma_0)\} \right] \times [E\{\mathbf{D}\mathbf{U}_i(\gamma_0)\}]^{-1^T}, \quad (12)$$

where $\mathbf{D}f$ denotes the Jacobian of f.

Proofs of these theorems are outlined in the Appendix. The proof of Theorem 1 requires the assumption that the equation $E_0\{\mathbf{U}_1(\gamma)\} = \mathbf{0}$ has a unique solution. Whether this assumption holds for $\boldsymbol{\beta}$ does not depend on the correct specification of the covariance matrix $\mathbf{V}$. Therefore, like GEE1 but unlike GEE2, the consistency of $\hat{\boldsymbol{\beta}}$ does not depend on whether $\text{var}(\mathbf{y}_i)$, $i = 1, \ldots, K$, has been correctly modeled. Of course, the consistency of $\hat{\alpha}$ requires correct covariance specification. An alternative proof of consistency is available if $\hat{\gamma}$ is defined as a one-step estimator (Lehmann 1983, sec. 6.3) based on a $K^{1/2}$-consistent estimator $\tilde{\gamma}$. In this case the proof requires weaker conditions than those given in the Appendix.

Although property (10) plays a role in the motivation of the estimating equations presented in Section 3.1, it should be noted that Theorems 1 and 2 are proved without assuming such a condition. Although terms of order $O(\phi^{3/2})$ are neglected in the determination of function f_1, the role of small dispersion asymptotics (Jorgensen 1987) is simply to suggest the form of an estimating equation for $\boldsymbol{\alpha}$; they are not used to establish the consistency and asymptotic normality of $\hat{\gamma}$. The same approach is used in the derivation of the extended QL function (McCullagh and Nelder 1989).

Expression (12) may be rewritten as

$$\mathbf{V}_2 = [E\{\mathbf{DU}_i(\gamma_0)\}]^{-1} \lim_{K\to\infty} K \left\{\sum_{i=1}^{K} \mathbf{M}_i(\gamma_0)\right\} \times [E\{\mathbf{DU}_i(\gamma_0)\}]^{-1^T}, \quad (13)$$

where $\mathbf{M}_i$ is the symmetric matrix $E\{\mathbf{U}_i(\gamma_0)\cdot \mathbf{U}_i(\gamma_0)^T\}$ given by

$$\mathbf{M}_i(\gamma_0) = \begin{bmatrix} \mathbf{M}_{i1} & \mathbf{M}_{i2} \\ \cdot & \mathbf{M}_{i3} \end{bmatrix}.$$

Here $\mathbf{M}_{i1}$ is the $p\times p$ matrix with (b,d)th element $\Sigma^{jk} V_{jl} V_{km} \mu^{l,b} \mu^{m,d}$, $\mathbf{M}_{i2}$ is the $p\times s$ matrix with (b,u)th element $-\Sigma^{jkl} V_{kl}^{u} V_{jm} \mu^{m,b}$, and $\mathbf{M}_{i3}$ is the $s\times s$ matrix with (u,v)th element such that

$$\Sigma^{jklm} V_{jk}^{u} V_{lm}^{v} - \Sigma^{jk} V_{jk}^{u} V^{lm} V_{lm}^{v} - \Sigma^{jk} V_{jk}^{v} V^{lm} V_{lm}^{u} + V^{jk} V_{jk}^{u} V^{lm} V_{lm}^{v}.$$

A covariance estimator $\hat{\mathbf{V}}_2$ can be obtained by replacing γ_0 with $\hat{\gamma}$ in $\mathbf{V}_2$. This covariance estimator is robust with respect to the specification var($\mathbf{y}$) when expected central moments in $\mathbf{M}_{i1}$, $\mathbf{M}_{i2}$, and $\mathbf{M}_{i3}$ are replaced with observed central moments. Given that no specifications have been made for third- and fourth-central moments, this "robust" estimator of var($\hat{\gamma}$) is the only one available. That is, without specifications for Σ^{jkl} and Σ^{jklm}, a model-based estimator of var($\hat{\gamma}$) is unobtainable.

The expected Jacobian $E\{\mathbf{DU}_i(\hat{\gamma})\}$, necessary to compute $\hat{\mathbf{V}}_2$, can be estimated by

$$E\{\mathbf{DU}_i(\hat{\gamma})\} = \begin{bmatrix} \mathbf{A}_{i1} & \mathbf{0} \\ \mathbf{A}_{i2} & \mathbf{A}_{i3} \end{bmatrix},$$

where $\mathbf{A}_{i1}$ is the $p\times p$ matrix with (b,d)th element $-\mu^{j,d} V_{jk} \mu^{k,b}$, $\mathbf{A}_{i2}$ is the $s\times p$ matrix with (u,b)th element $V^{jk,b} V_{jk}^{u}$, and $\mathbf{A}_{i3}$ is the $s\times s$ matrix with (u,v)th element $V^{jk,v} V_{jk}^{u}$. Here the superscripts u and v indicate partial derivatives with respect to the elements of $\boldsymbol{\alpha}$ and the superscript b indicates the partial derivative with respect to β^b. Note that in the expressions for $\mathbf{M}_{i1}$, $\mathbf{M}_{i2}$, and $\mathbf{M}_{i3}$ and $\mathbf{A}_{i1}$, $\mathbf{A}_{i2}$, and $\mathbf{A}_{i3}$, the individual subscript i has been dropped for notational convenience. It should be understood, however, that these expressions are individual specific.

3.3 Computation

A Fisher scoring algorithm may be used to solve the EGEEs given by (5). Given an initial value $\gamma^{(0)}$, the updated parameter estimate, $\gamma^{(1)}$, is given by

$$\gamma^{(1)} = \gamma^{(0)} - [E\{\mathbf{DU}_i(\gamma^{(0)})\}]^{-1} \left\{\sum_{i=1}^{K} \mathbf{U}_i(\gamma^{(0)})\right\},$$

where the expectation is with respect to $\gamma^{(0)}$. This procedure is iterated until the correction term becomes sufficiently small.

4. EFFICIENCY CONSIDERATIONS

In this section we address the relative efficiencies of parameter estimators based on EGEE, GEE1, and GEE2. We consider results for both GEE1 with method-of-moment estimators and for GEE1 with ad hoc estimating equation–based estimators of association parameters. We refer to the latter approach as GEE1* to differentiate the two methods. In Section 4.1 we present ratios of asymptotic variances (AVARs) for a particular generalized linear model. In Section 4.2 we give ratios of mean squared errors (MSEs) from a small-sample simulation study. These results demonstrate advantages of the EGEE method in some of the situations considered.

4.1 Asymptotic Relative Efficiency

Liang and Zeger (1986) considered the asymptotic relative efficiency (ARE) of the GEE1 estimator $\tilde{\boldsymbol{\beta}}$ under the model $g(\mu_{it}) = \beta_1 + \beta_2 t/10$, where $\beta_1 = \beta_2 = 1$ and $T_i = 10$, for all i. Here we return to this model, assuming further a log link and identity variance function,

$$\log(\mu_{it}) = \beta_1 + \beta_2 t/10, \qquad \text{var}(y_{it}) = \phi\mu_{it},$$
$$\text{where} \quad i = 1,\ldots,K \quad \text{and} \quad t = 1,\ldots,10 \quad (14)$$

as in Poisson regression. Here we present ratios of the asymptotic variances of GEE1*, GEE2, and EGEE estimators of parameters β_1, β_2, α, and ϕ from this model, under several choices of the true and assumed structure for third- and fourth-order moments. We consider four of the correlation structures discussed by Liang and Zeger (1986), and we assume that in each case, the working correlation structure has been specified correctly. For all of these results, α was taken to be equal to .5.

For these methods, a true structure for third- and fourth-order moments must be assumed to obtain asymptotic variances. That is, $E[(y_{ij}-\mu_{ij})(y_{ik}-\mu_{ik})(y_{il}-\mu_{il})]$ and $E[(y_{ij}-\mu_{ij})(y_{ik}-\mu_{ik})(y_{il}-\mu_{il})(y_{im}-\mu_{im})]$ must be specified for all i, j, k, l, and m. Three cases are considered here, corresponding to higher-moment structures suggested by Prentice and Zhao (1991) as possible working variance models for GEE2. In the first case (INDEP), the structure for third- and fourth-order moments is assumed to be consistent with independent elements in the response

Table 1. Ratios of AVAR(GEE2) to AVAR(EGEE) Under Model (14)

True higher correlations	Model higher correlations	Parameter	1-Dependence	Exchangeable	AR(1)	Independence
INDEP	INDEP	β_1	.85	.93	.83	.92
		β_2	.82	.89	.79	.90
		α	.24	.16	.15	
		ϕ	.26	1.69	3.40	11.25
INDEP	GAUSS1	β_1	3.29	1.01	1.07	.92
		β_2	3.81	1.02	1.09	.90
		α	.75	.16	.40	
		ϕ	1.81	1.71	3.73	11.25
GAUSS1	INDEP	β_1	1.03	1.02	1.17	.92
		β_2	1.04	1.04	1.21	.90
		α	63.02	1.34	2.45	
		ϕ	1.52	1.76	1.47	1.22
GAUSS1	GAUSS1	β_1	.88	.95	.87	.92
		β_2	.85	.91	.84	.90
		α	1.03	1.34	1.09	
		ϕ	1.27	1.75	1.36	1.22
GAUSS2	INDEP	β_1	1.15	1.26	1.24	1.09
		β_2	1.11	1.47	1.22	1.04
		α	12.46	1.42	4.28	
		ϕ	3.76	1.74	.37	1.36
GAUSS2	GAUSS1	β_1	1.23	1.07	1.01	1.09
		β_2	1.17	1.11	.95	1.04
		α	.95	1.42	1.19	
		ϕ	1.19	1.75	1.51	1.36

vectors $\mathbf{y}_i$, $i = 1, \ldots, K$. In the second case (GAUSS1), third- and fourth-order moments are consistent with $\mathbf{y}_i$, $i = 1, \ldots, K$, each distributed as a multivariate normal random vector. The third case (GAUSS2) is a generalization of GAUSS1 incorporating third- and fourth-order correlations that are assumed to be common across clusters (see Prentice and Zhao 1991 for more detail on these structures). In GAUSS2, third- and fourth-order correlations are given by

$$\gamma_{jkl} = \frac{E[(y_{ij} - \mu_{ij})(y_{ik} - \mu_{ik})(y_{il} - \mu_{il})]}{\sqrt{\sigma_{jj}\sigma_{kk}\sigma_{ll}}} \quad (15)$$

and

$$\delta_{jklm} = \frac{E[((y_{ij} - \mu_{ij})(y_{ik} - \mu_{ik})(y_{il} - \mu_{il})(y_{im} - \mu_{im})]}{\sqrt{\sigma_{jj}\sigma_{kk}\sigma_{ll}\sigma_{mm}}} - \frac{(\sigma_{jk}\sigma_{lm} + \sigma_{jl}\sigma_{km} + \sigma_{jm}\sigma_{kl})}{\sqrt{\sigma_{jj}\sigma_{kk}\sigma_{ll}\sigma_{mm}}}. \quad (16)$$

Rather than arbitrarily choosing values for γ_{jkl} and δ_{jklm}, we obtained the particular values of these parameters used here to specify the true model by replacing expected values with sample means in (15) and (16). We computed sample means on a large ($K = 1{,}000$) dataset randomly generated from (14) with an exchangeable correlation matrix with $\alpha = .5$.

Table 1 gives ratios of the asymptotic variances of GEE2 estimators to the asymptotic variances of EGEE estimators for parameters β_1, β_2, α, and ϕ. We used the INDEP, GAUSS1, and GAUSS2 structures as true higher-moment models, and the INDEP and GAUSS1 models as working structures for GEE2 and GEE1*. As mentioned earlier, the true structures for third- and fourth-order moments considered here correspond to the working structures suggested for GEE2 by Prentice and Zhao (1991). Although these structures may be effective as working higher-moment models when estimating first- and second-moment parameters, there is no reason to expect them to be accurate repre-

Table 2. Ratios of AVAR(GEE1*) to AVAR(EGEE) Under Model (14)

True higher correlations	Model higher correlations	Parameter	1-Dependence	Exchangeable	AR(1)	Independence
INDEP	INDEP	α	.24	.16	.15	
		ϕ	.22	1.07	2.61	10.14
INDEP	GAUSS1	α	.76	.16	.40	
		ϕ	1.41	1.07	2.93	10.14
GAUSS1	INDEP	α	63.02	1.34	2.45	
		ϕ	1.33	1.48	1.26	1.10
GAUSS1	GAUSS1	α	1.03	1.34	1.09	
		ϕ	1.08	1.48	1.16	1.10
GAUSS2	INDEP	α	12.46	1.42	4.28	
		ϕ	4.82	2.00	.65	1.77
GAUSS2	GAUSS1	α	.95	1.42	1.19	
		ϕ	2.08	2.00	1.88	1.77

Table 3. Ratios of MSE_{GEE1}/MSE_{EGEE} Under Model (17)

True **R**	Fitted **R**	Parameter					
		β_1	β_2	β_3	β_4	α_1	ϕ
Independence	Independence	1.000	1.000	1.000	1.000		1.000
	Unspecified	.998	.998	.997	.996		.986
	Exchangeable	1.000	1.000	1.000	1.000		1.000
	AR(1)	1.000	1.000	1.000	1.000		1.000
Unspecified	Independence	1.000	1.000	1.000	1.000		1.000
	Unspecified	1.021	1.018	1.012	1.011	1.353	1.001
	Exchangeable	1.000	1.000	1.000	1.000		.974
	AR(1)	1.001	1.001	1.000	1.000		.986
Exchangeable	Independence	1.000	1.000	1.000	1.000		1.000
	Unspecified	1.023	1.020	1.016	1.015		.993
	Exchangeable	1.000	1.000	1.000	1.000	1.080	.981
	AR(1)	1.002	1.001	1.000	1.000		.975
AR(1)	Independence	1.000	1.000	1.000	1.000		1.000
	Unspecified	1.014	1.012	1.010	1.008		1.019
	Exchangeable	1.000	1.000	1.000	1.000		.986
	AR(1)	1.001	1.001	1.001	1.000	1.206	1.001

sentations of the true structure of data likely to be encountered in practice. On the contrary, the GEE-based methods under consideration in this article are designed for clustered data for which generalized linear models are appropriate; that is, dependent, nonnormal data. This is an important issue to keep in mind while considering the results presented here. In particular, it implies that the situations in which the working model for GEE2 has been incorrectly specified are most relevant.

In the first four lines of Table 1 the working higher moment structure has been correctly specified; hence GEE2 clearly outperforms EGEE with respect to $\boldsymbol{\beta}$ and α. Results for ϕ are mixed. In the other situation in which GEE2 is used with a correct higher-moment structure (GAUSS1), GEE2 outperforms EGEE with respect to $\boldsymbol{\beta}$ but has substantially higher asymptotic variances for α and ϕ. In the misspecified situations, EGEE generally has somewhat smaller asymptotic variances for both regression and association parameters. The exception to this is when the true structure is INDEP, and GEE2 is used with a GAUSS1 working structure. In this case, GEE2 provides estimators of α with much smaller asymptotic variances. A possible explanation for this result is that the INDEP and GAUSS1 structures are very similar, and the minor misspecification involved in choosing GAUSS1 when the true higher-moment structure is INDEP somehow is not relevant to the estimation of $\boldsymbol{\alpha}$.

Table 2 presents ratios of the asymptotic variances of GEE1* estimators to the asymptotic variances of EGEE estimators for α and ϕ from (14). Such ratios are 1 for $\boldsymbol{\beta}$ and have been omitted. Ratios in Table 2 are identical to those in Table 1 for α. For ϕ, when the true structure of third- and fourth-order moments is INDEP or GAUSS1, the ratios are less than those in Table 1. For the GAUSS2 structure, the asymptotic variance ratios for ϕ are higher than those in Table 1, indicating a greater advantage to using EGEE. These results indicate that GEE1* and EGEE are equally efficient for estimating regression parameters. GEE1* can be more efficient than EGEE for estimating association parameters if weight matrices (approximating the third- and fourth-moment structures) are chosen well. However, if these weight matrices do not approximate the true higher moment structure, then substantial gains can be made

Table 4. Ratios of MSE_{GEE1^*}/MSE_{EGEE} Under Model (17) GEE1* Fit With an Independence Working Structure for Higher-Order Moments

True **R**	Fitted **R**	Parameter					
		β_1	β_2	β_3	β_4	α_1	ϕ
Independence	Independence	1.000	1.000	1.000	1.000		1.000
	Unspecified	.999	.999	.999	.999		.990
	Exchangeable	1.000	1.000	1.000	1.000		1.000
	AR(1)	1.000	1.000	1.000	1.000		1.000
Unspecified	Independence	1.000	1.000	1.000	1.000		1.000
	Unspecified	1.025	1.022	1.014	1.013	1.360	1.024
	Exchangeable	1.000	1.000	1.000	1.000		1.015
	AR(1)	1.012	1.011	1.012	1.012		1.369
Exchangeable	Independence	1.000	1.000	1.000	1.000		1.048
	Unspecified	1.028	1.025	1.021	1.019		1.031
	Exchangeable	1.000	1.000	1.000	1.000	1.061	1.024
	AR(1)	1.030	1.029	1.033	1.030		1.641
AR(1)	Independence	1.000	1.000	1.000	1.000		1.045
	Unspecified	1.017	1.014	1.014	1.012		1.047
	Exchangeable	1.000	1.000	1.000	1.000		1.030
	AR(1)	1.001	1.001	1.002	1.001	1.572	1.068

Table 5. Ratios of MSE_{GEE1^*}/MSE_{EGEE} Under Model (17) GEE1 Fit With a Gaussian Working Structure for Higher-Order Moments*

True **R**	Fitted **R**	Parameter β_1	β_2	β_3	β_4	α_1	ϕ
Independence	Independence	1.000	1.000	1.000	1.000		1.000
	Unspecified	.999	.999	.999	.999		.990
	Exchangeable	1.000	1.000	1.000	1.000		1.000
	AR(1)	1.000	1.000	1.000	1.000		1.000
Unspecified	Independence	1.000	1.000	1.000	1.000		1.042
	Unspecified	1.000	1.000	1.000	1.000	1.028	1.010
	Exchangeable	1.000	1.000	1.000	1.000		1.015
	AR(1)	.999	1.000	1.000	1.000		1.021
Exchangeable	Independence	1.000	1.000	1.000	1.000		1.048
	Unspecified	1.000	1.000	1.000	1.000		1.024
	Exchangeable	1.000	1.000	1.000	1.000	1.061	1.024
	AR(1)	.999	.999	.999	.999		1.030
AR(1)	Independence	1.000	1.000	1.000	1.000		1.045
	Unspecified	.999	.999	.999	.999		1.020
	Exchangeable	1.000	1.000	1.000	1.000		.957
	AR(1)	1.000	1.000	1.000	1.000	1.045	1.027

in the efficiency of association parameter estimators by using EGEE. Further ARE computations with higher levels of intracluster correlation (not reported here) indicate that such gains increase with the amount of within-cluster dependence.

The results of this section suggest that when there is strong prior knowledge regarding the third and fourth moments of the responses, either GEE1* or GEE2 is an excellent choice. In such cases, when substantial interest is centered on the association parameters, we concur with Liang et al. (1992) in the recommendation of GEE2. But when covariance parameters are of secondary interest, GEE1* is preferrable. When higher-moment information is not available, however, EGEE appears to be the best choice, combining efficient inference for second-moment parameters without risking the consistency of regression parameter estimators.

4.2 Simulation Results

To compare finite-sample MSEs, datasets were generated in the structure of the data reported by Leppik et al. (1985). These data consist of seizure counts measured repeatedly over four 2-week periods in addition to an 8-week baseline seizure count, a treatment group indicator, the subject's age, and other covariates measured on each of $K = 59$ individuals. The simulation data were generated with this structure (i.e., $K = 59$ and $T_i = 4$), according to the model

$$\log(\mu_{it}) = \beta_1 + \text{BASE}\beta_2 + \text{TRT}\beta_3 + \text{BASE} \times \text{TRT}\beta_4, \quad (17)$$

where BASE $= \log(.25 \times$ baseline seizure count) and TRT is a binary indicator for inclusion in the treatment group. Data were generated with this mean structure, identity variance function, and several correlation structures. These correlation structures included the exchangeable (equicorrelation), AR(1), unspecified, and independence correlation structures discussed by Liang and Zeger (1986). The values of $\boldsymbol{\alpha}$ for these structures were taken to be .5, .5, and $(.5, .45, .40, .55, .45, .48)^T$ for the first three of these structures. These data were generated according to the algorithm of Sim (1993) for Poisson deviates with given marginal means and covariance matrix.

Table 6. Ratios of MSE_{GEE2}/MSE_{EGEE} Under Model (17) GEE2 Fit With a Gaussian Working Structure for Higher-Order Moments

True **R**	Fitted **R**	Parameter β_1	β_2	β_3	β_4	α_1	ϕ
Independence	Independence	1.109	1.103	1.156	1.132		.993
	Unspecified	1.093	1.088	1.123	1.106		.982
	Exchangeable	1.105	1.100	1.145	1.123		.994
	AR(1)	1.108	1.102	1.150	1.127		.993
Unspecified	Independence	1.054	1.053	1.090	1.067		1.040
	Unspecified	1.100	1.100	1.173	1.138	1.006	1.015
	Exchangeable	1.113	1.111	1.187	1.147		1.023
	AR(1)	1.098	1.098	1.149	1.116		1.026
Exchangeable	Independence	1.054	1.050	1.091	1.070		1.042
	Unspecified	1.121	1.113	1.156	1.131		1.026
	Exchangeable	1.124	1.116	1.165	1.143	1.035	1.025
	AR(1)	1.083	1.078	1.132	1.104		1.031
AR(1)	Independence	1.065	1.067	1.100	1.086		1.042
	Unspecified	1.124	1.130	1.186	1.162		1.018
	Exchangeable	1.152	1.159	1.210	1.183		1.032
	AR(1)	1.143	1.149	1.210	1.184	1.024	1.026

Table 7. Summary of Responses to 2,4,5-T Among CD-1 Mice*

Dose (mg/kg)	Responses	Implants	Litters
0	59	802	73
30	124	955	87
45	338	1120	98
60	390	806	76
75	372	482	44
90	242	254	25

* Replicates 7–9, 11–15.

We generated 3,000 datasets of this description and fit them to model (17) using GEE1, GEE1*, GEE2, and EGEE. All four of the true correlation structures were used as working correlation structures. GEE1* was fit using both the INDEP and GAUSS1 working structures for third- and fourth-order moments. For GEE2, convergence problems when fitting (17) with the INDEP structure resulted in unreliable results, so we report only the GAUSS1 results.

Tables 3–6 present MSE ratios for the methods where the EGEE MSE always appears in the denominator. (Ratios greater than 1 indicate superior performance for EGEE). Note that in these tables, results for $\alpha_2, \ldots, \alpha_6$ have been omitted in the unspecified case to save space. Tables 3, 4, and 5 show that GEE1, GEE1*, and EGEE are essentially identical for the estimation of $\boldsymbol{\beta}$. But with respect to the association parameters, EGEE gives smaller MSEs, particularly in Tables 3 and 4 for $\boldsymbol{\alpha}$ and particularly in Table 4 for ϕ.

Finally, in Table 6, GEE2 is fit with a GAUSS1 working structure. This procedure leads to MSEs for α that are similar to EGEE. Comparing the MSE ratios for α in Tables 3 and 6, we see that GEE2 has resulted in an increase in efficiency over GEE1 with respect to α. This result is consistent with results of other authors (e.g., Liang et al. 1992). However, at least in the Poisson regression example considered here, this increase in efficiency for α comes at the cost of decreased accuracy with respect to $\boldsymbol{\beta}$. Such is not the case for EGEE.

5. EXAMPLE

In developmental toxicity studies in laboratory animals, intralitter correlation typically is present. For a binary response measured on littermates, this correlation often induces extrabinomial variation. Historically, the beta-binomial distribution has been used with a logistic dose–response model to account for such overdispersion. Recently, however, results have been established indicating that an incorrect assumed correlation structure in such a model can lead to severely biased parameter estimates and underestimated parameter variances (Kupper, Portier, Hogan, and Yamamoto 1986; Williams 1988). To avoid these problems, several authors have considered using GEEs. Lefkopoulou, Moore, and Ryan (1989) and Liang and Hanfelt (1994) considered GEE1 for teratology data, whereas Zhu, Krewski, and Ross (1994) used GEE1* and Bowman, Chen, and George (1995) used GEE2 in this setting. In this section we apply the EGEE method to data considered by Bowman et al. and Zhu et al. As discussed by Bowman et al. (1995), flexible modeling of the intralitter correlation is important in these studies to obtain adequate fit in the first-moment model and to eliminate bias and increase precision in the estimates of regression parameters. Furthermore, although the first-moment relationships are of primary interest in developmental toxicity studies, "a dose–response function of the intralitter correlation can, in addition to the mean, provide further opportunity to evaluate developmental effects of a toxic substance" (Bowman et al. 1995, p. 1523). Such considerations motivate the use of estimating equations for the association parameters so that dose dependence in the intralitter correlations can be accounted for and to ensure efficient inference on the intralitter correlations themselves.

In a study of 2,4,5-trichlorophenoxyacetic acid (2,4,5-T) conducted at the National Center for Toxicological Research (Holson et al. 1992), developmental effects including reduced fetal weight, cleft palate, and embryolethality were observed among the fetuses of five strains of pregnant mice exposed to 2,4,5-T. Following Bowman et al. (1995), we analyze only the data from the CD-1 strain, replicates 7–9 and 11–15. Each of these mice was randomly assigned to receive a dosage of 0, 30, 45, 60, 75, or 90 mg/kg of the toxin during pregnancy. Before the end of gestation, the mice were sacrificed and examined to identify responses to the toxin among the offspring. In this analysis a response is defined as death, resorption, or cleft palate malformation. A summary of these data appears in Table 7.

A logit link function was used to relate μ_{it} (the probability of response for the tth fetus in the ith litter) to x_i, the dose for the ith litter as follows:

$$\log\left(\frac{\mu_{it}}{1-\mu_{it}}\right) = \beta_1 + \beta_2 x_i, \quad \text{where} \quad i = 1, \ldots, K \quad \text{and} \quad t = 1, \ldots, T_i.$$

Two correlation models were considered. For all i, let $\rho_i = \operatorname{corr}(y_{it}, y_{it'})$ when $t \neq t'$. Model 1 is the constant correlation model $\rho_i = \alpha$, $i = 1, \ldots, K$, and model 2 is the dose-dependent logistic correlation model $\operatorname{logit}((\rho_i + 1)/2) = \alpha_1 + \alpha_2 x_i$, $i = 1, \ldots, K$. Both models were fitted using GEE2 with the Gaussian working correlation model described in Section 4.1, and EGEE. Parameter estimates and asymptotic standard errors appear in Table 8.

The estimates of β_2 and α_2 in Table 8 are positive and large compared to their standard errors. Thus the probability of response and the intralitter correlation increase with dose. The intralitter correlations for model 2 are −.005, .285, .416, .531, .630, and .711 for the 0, 30, 45, 60, 75, and

Table 8. Parameter Estimates and Standard Errors (in Parentheses) for 2,4,5-T Data

	GEE2		EGEE	
	Model 1	Model 2	Model 1	Model 2
β_1	−3.3863(.204)	−2.6105(.149)	−3.5065(.198)	−2.8108(.143)
β_2	.0580(.004)	.0455(.004)	.0600(.004)	.0467(.003)
α_1	.2595(.028)	−.0109(.032)	.2402(.024)	−.0064(.038)
α_2		.0199(.002)		.0220(.002)
ϕ	1.1366(.051)	.9323(.027)	1.1944(.040)	1.0480(.026)

90 mg/kg dose groups, using GEE2. These correlations are $-.003$, .316, .456, .576, .676, and .756 for model 2 fitted with EGEE.

6. DISCUSSION

We have proposed a procedure for estimating regression parameters ($\boldsymbol{\beta}$) and association parameters ($\boldsymbol{\alpha}$) in GLMs for clustered data. The proposed methodology is based on the ideas of extended QL and is a natural extension of the work of Liang and Zeger (1986) and others to the situation when precise estimates of both the regression and association parameters are desired. This approach may be seen as an alternative to that of Prentice and Zhao (1991), who proposed estimating equations based on a quadratic exponential model that require a "working" specification for the third and fourth moments of the responses. Resulting estimators of $\boldsymbol{\beta}$ and $\boldsymbol{\alpha}$ have been found to have good asymptotic efficiencies in some cases (Liang et al. 1992) and can be expected to have good efficiency when these higher-order moments have been well specified.

We have shown that under certain regularity conditions, EGEE-based estimators are asymptotically normal and consistent. The consistency of the EGEE-based estimator of $\boldsymbol{\beta}$ does not require that the covariance of the response vector be modeled correctly. The procedure proposed here also has the advantage that estimates can be obtained without making third- and fourth-moment assumptions about the responses. Comparisons of asymptotic variances and simulation-based MSEs have demonstrated advantages of the EGEE method over alternative estimating equation-based procedures for some situations of interest.

An alternative procedure for the analysis of GLMs for longitudinal data that has not been discussed here is the full-likelihood procedure introduced by Fitzmaurice and Laird (1993). These authors proposed an approach based on a log-linear representation of the joint distribution of a correlated binary response vector. This representation is a more general version of Zhao and Prentice's (1990) quadratic exponential density in which the canonical parameters are partitioned into first-order parameters and into second- and higher-order parameters. Separate models for the mean and the second- and higher-canonical association parameters are specified, and parameter estimation proceeds via a two-stage iterative procedure that alternates between iterative proportional fitting and Fisher scoring.

Clearly, this full-likelihood approach offers some advantages over EGEE in certain situations in which the assumed model is correct. For example, such a procedure yields likelihood ratio tests and parametric variance formulas. In addition, the canonical association parameters in the Fitzmaurice and Laird approach have interpretations as log-conditional odds ratios and hence are unconstrained, unlike the marginal correlations used in EGEE. On the other hand, in contrast to the marginal association parameters of EGEE, the log-conditional odds ratios do not have convenient interpretations independent of cluster size. This is of particular concern in the case considered here in which both mean and association parameters are of scientific interest. A related drawback is that the distribution considered by Fitzmaurice and Laird is not "reproducible" in the sense described by Liang et al. (1992). Therefore, the full-likelihood procedure is inappropriate for data in which clusters are of various sizes. Furthermore, EGEE avoids the doubly iterative nature of the maximum likelihood estimation procedure, and it shares robustness in the estimation of $\boldsymbol{\beta}$ to misspecification of second- and higher-moment models. Adaptation of EGEE to a parameterization of intracluster dependence in terms of marginal odds ratios appears to be straightforward.

APPENDIX: PROOFS OF THEOREMS 1 AND 2

Included here are a statement of the assumed regularity conditions, some comments regarding these conditions, and proofs of the two theorems.

Conditions.

1. Let Γ be the set of all possible γ. Assume that Γ is a compact subset of $\Re^q$.

2. Let $\mathbf{U}\cdot(\gamma) = \sum_{i=1}^K \mathbf{U}_i(\gamma)$ and let $\boldsymbol{\Psi}(\gamma) = E_0\{\mathbf{U}_1(\gamma)\}$, where E_0 is the expectation with respect to the true γ, γ_0. Assume that $\boldsymbol{\Psi}(\gamma) = \mathbf{0}$ has a unique solution, γ_0^*, in the interior of Γ.

3. For every $\varepsilon > 0$, there exists a $\delta > 0$ such that $\|\boldsymbol{\Psi}(\tilde{\gamma})\| \leq \delta$ implies that $\|\tilde{\gamma} - \gamma_0^*\| \leq \varepsilon$ for each $\tilde{\gamma} \in \Gamma$.

4. For any open set $S \subset \Gamma$, let $\mathbf{a}$ and $\mathbf{b}$ be two points in S such that the line segment $L(\mathbf{a},\mathbf{b}) \subset S$. For each $K > 0$, let Ψ_{K_i} be the ith component of $\boldsymbol{\Psi}_K$, $i = 1,\ldots,q$, and let $\boldsymbol{\nabla}\Psi_{K_i}(\tilde{\gamma}_i)$ be the gradient of Ψ_{K_i} at the point $\tilde{\gamma}_i \in L(\mathbf{a},\mathbf{b})$. Let

$$W_K = \left\{ \sum_{i=1}^{q} \sup_{\tilde{\gamma}_i \in L(\mathbf{a},\mathbf{b})} \|\boldsymbol{\nabla}\Psi_{K_i}(\tilde{\gamma}_i)\|^2 \right\}^{1/2}.$$

Assume that there exists a K_0 such that for all $K > K_0$, $E_0(W_K) < \infty$.

5. There exists an open set $S \subset \Gamma$ containing γ_0 such that for any two points, $\mathbf{a}$ and $\mathbf{b}$ in S such that the line segment $L(\mathbf{a},\mathbf{b}) \subset S$ and $\gamma_0 \in L(\mathbf{a},\mathbf{b})$, $\partial\Psi_{K_i}(\tilde{\gamma})/\partial\tilde{\gamma}$, and $\partial^2\Psi_{K_i}(\tilde{\gamma})/\partial\tilde{\gamma}^2$ exist almost surely for each $i = 1,\ldots,q$, and all $\tilde{\gamma} \in L(\mathbf{a},\mathbf{b})$.

6. $E_0(\sup_K \|\boldsymbol{\Psi}_K(\gamma_0^*) - \boldsymbol{\Psi}(\gamma_0^*)\|) < \infty$.

7. There exists an open set $S \subset \Gamma$ containing γ_0 such that for any two points $\mathbf{a}$ and $\mathbf{b}$ in S, such that the line segment $L(\mathbf{a},\mathbf{b}) \subset S$ and $\gamma_0 \in L(\mathbf{a},\mathbf{b})$, the result

$$(\tilde{\gamma} - \gamma_0)^T[\boldsymbol{\nabla}^2 U^k \cdot \{t\tilde{\gamma} + (1-t)\gamma_0\}](\tilde{\gamma} - \gamma_0) \qquad \text{(A.1)}$$

is a continuous function of t, for all $\tilde{\gamma} \in L(\mathbf{a},\mathbf{b})$ and each $i = 1,\ldots,q$. Here $\boldsymbol{\nabla}^2 f(\mathbf{a})$ denotes the matrix of second derivatives of $f(\mathbf{a})$—$\boldsymbol{\nabla}^2 f(\mathbf{a})$ has (i,j)th element $\partial^2 f(\mathbf{a})/\partial\mathbf{a}^i\partial\mathbf{a}^j$—and superscripts in (A.1) indicate vector components.

8. There exists an open set $S \subset \Gamma$ containing γ_0 such that for any two points $\mathbf{a}$ and $\mathbf{b}$ in S, such that $L(\mathbf{a},\mathbf{b}) \subset S$ and $\gamma_0 \in L(\mathbf{a},\mathbf{b})$,

$$\frac{1}{K}\left|\frac{\partial^2 U^k \cdot (\tilde{\gamma})}{\partial\tilde{\gamma}^i\partial\tilde{\gamma}^j}\right| \leq g_{ijk}(\mathbf{y}),$$

where $E\{g_{ijk}(\mathbf{y})\} < \infty$ for each $i,j,k \in \{1,\ldots,q\}$ and all $\tilde{\gamma} \in L(\mathbf{a},\mathbf{b})$.

9. It is assumed that $E\{\mathbf{DU}\cdot(\gamma_0)\}$, is nonsingular.

10. The matrices $\mathbf{x}_i$, $i = 1,\ldots,K$, are assumed to be iid random matrices.

Comments about Conditions

Conditions 1–6 are assumed to establish the consistency of $\hat{\gamma}$.

Condition 2 holds for β when the link function is linear, because when $\partial\boldsymbol{\mu}/\partial\beta$ has rank p, $\boldsymbol{\mu}_0 = \boldsymbol{\mu}$ is the unique solution of $E_0\{\mathbf{U}_1(\gamma;\beta)\} = \mathbf{0}$. Here $\boldsymbol{\mu}_0$ is the mean, $g^{-1}(\mathbf{X}\beta_0)$, at the true regression parameter, β_0. In the nonlinear link situation $\boldsymbol{\mu}_0 = \boldsymbol{\mu}$ can be expected to be the unique solution when a good linear approximation to the link function exists for values of β of interest. (See, e.g., Seber and Wild 1989 for a discussion of this issue in the nonlinear regression context.) It is not required that the assumed covariance matrix be correct.

However, when the covariance matrix has been correctly specified, a unique solution for $\boldsymbol{\alpha}$ and ϕ is ensured if the equation

$$\operatorname{tr}\left\{R(\boldsymbol{\alpha})\frac{\partial R^{-1}(\boldsymbol{\alpha})}{\partial\boldsymbol{\alpha}}\right\} - \frac{T\operatorname{tr}\left\{R(\boldsymbol{\alpha}_0)\frac{\partial R^{-1}(\boldsymbol{\alpha})}{\partial\boldsymbol{\alpha}}\right\}}{\operatorname{tr}\{R(\boldsymbol{\alpha}_0)R^{-1}(\boldsymbol{\alpha})\}} = 0 \qquad (A.2)$$

has the unique solution $\boldsymbol{\alpha} = \boldsymbol{\alpha}_0$. This equation results from the simultaneous solution of $E_0\{\mathbf{U}\cdot(\gamma;\boldsymbol{\alpha})\} = \mathbf{0}$ and $E_0\{\mathbf{U}\cdot(\gamma;\phi)\} = \mathbf{0}$. The fact that (A.2) implies that $\boldsymbol{\alpha} = \boldsymbol{\alpha}_0$ can be verified for simple cases such as the AR(1) and exchangeable correlation structures. However, it has not yet been established whether or not the condition holds in general.

Conditions 5 and 7–10 are used to prove asymptotic normality. Condition 10 allows an iid proof and is assumed for simplicity. The other conditions are standard regularity conditions of the sort often used to establish asymptotic normality. (The interested reader is referred to LeCam 1970 for a discussion of these conditions.)

Proof of Theorem 1

By the weak law of large numbers, as $K \to \infty$, $\boldsymbol{\Psi}_K(\gamma) \xrightarrow{p} \boldsymbol{\Psi}(\gamma)$ for each $\gamma \in \Gamma$. Condition 3 implies that for every $\varepsilon > 0$ and each $K > 0$, there exists a $\delta > 0$ such that

$$\Pr(\|\hat{\gamma} - \gamma_0^*\| > \varepsilon) \le \Pr\{\sup_{\gamma} \|\Psi_K(\gamma) - \Psi(\gamma)\| > \delta\}$$

which goes to 0 as $K \to \infty$ provided that $\boldsymbol{\Psi}_K$ converges uniformly to $\boldsymbol{\Psi}$. Uniform convergence follows from the relative compactness of $\{\boldsymbol{\Psi}_K\}$ (Billingsley 1968), which requires that (a) $\{\boldsymbol{\Psi}_K(\tilde{\gamma})\}$ be tight for some $\tilde{\gamma} \in \Gamma$, and (b) for each $\varepsilon > 0$, $\eta > 0$, there exists a δ such that $0 < \delta < 1$ and an integer K_0 such that

$$\Pr\{\sup_{\|\mathbf{a}-\mathbf{b}\|<\delta} \|\boldsymbol{\Psi}_K(\mathbf{a}) - \boldsymbol{\Psi}_K(\mathbf{b})\| \ge \varepsilon\} \le \eta, \qquad K \ge K_0.$$

Tightness of $\{\boldsymbol{\Psi}_K\}$ follows from condition 6, and (b) follows from condition 4 using the mean-value theorem.

At this point it has been established that $\hat{\gamma} \xrightarrow{p} \gamma_0^*$. By condition 2, $\boldsymbol{\Psi}(\gamma) = \mathbf{0}$ has a unique solution. Therefore, $\gamma_0^* = \gamma_0$. Hence $\hat{\gamma}$ is consistent.

Proof of Theorem 2

A Taylor series expansion of $\mathbf{U}\cdot(\hat{\gamma})$ around γ_0 is given by

$$\mathbf{U}\cdot(\hat{\gamma}) = \mathbf{U}\cdot(\gamma_0) + \mathbf{DU}\cdot(\gamma_0)(\hat{\gamma} - \gamma_0) + \mathbf{P}(\hat{\gamma} - \gamma_0),$$

where $\mathbf{P}$ is a $q \times q$ matrix with kth row given by

$$\int_0^1 (1-t)(\hat{\gamma} - \gamma_0)^T[\boldsymbol{\nabla}^2 U^k \cdot \{t\hat{\gamma} + (1-t)\gamma_0\}]\,dt, \qquad (A.3)$$

and $\boldsymbol{\nabla}^2 f(\mathbf{a})$ denotes the matrix of second derivatives of $f(\mathbf{a})$; $\boldsymbol{\nabla}^2 f(\mathbf{a})$ has (i,j)th element $\partial^2 f(\mathbf{a})/\partial\mathbf{a}^i\partial\mathbf{a}^j$. Note that superscripts in (A.3) refer to vector elements rather than powers. Because $\mathbf{U}\cdot(\hat{\gamma}) = \mathbf{0}$,

$$-\mathbf{U}\cdot(\gamma_0) = \mathbf{DU}\cdot(\gamma_0)(\hat{\gamma} - \gamma_0) + \mathbf{P}(\hat{\gamma} - \gamma_0). \qquad (A.4)$$

Adding and subtracting $E\{\mathbf{DU}\cdot(\gamma_0)\}(\hat{\gamma} - \gamma_0)$ to the right side of (A.4) gives

$$\begin{aligned} K^{-1/2}\mathbf{U}\cdot(\gamma_0) &= \left(\frac{1}{K}[\mathbf{DU}\cdot(\gamma_0) - E\{\mathbf{DU}\cdot(\gamma_0)\}]\right. \\ &\qquad \left. + \frac{1}{K}E\{\mathbf{DU}\cdot(\gamma_0)\} + \frac{1}{K}\mathbf{P}\right) \\ &\quad \times K^{1/2}(\gamma_0 - \hat{\gamma}). \end{aligned}$$

Note that $K^{-1}\mathbf{P} = o_p(1)$ and $K^{-1}[\mathbf{DU}\cdot(\gamma_0) - E\{\mathbf{DU}\cdot(\gamma_0)\}] = o_p(1)$. It follows that

$$-K^{-1/2}\mathbf{U}\cdot(\gamma_0) = K^{1/2}(\hat{\gamma}_0 - \gamma_0)\left[\frac{1}{K}E\{\mathbf{DU}\cdot(\gamma_0)\} + o_p(1)\right],$$

Thus for large values of K, $K^{1/2}(\hat{\gamma} - \gamma_0)$ can be approximated by

$$-K^{1/2}\left[\sum_{i=1}^{K} E\{\mathbf{DU}_i(\gamma_0)\}\right]^{-1}\mathbf{U}\cdot(\gamma_0),$$

which is asymptotically multivariate normal with mean 0 and covariance matrix $\mathbf{V}_2$.

[Received December 1995. Revised April 1998.]

REFERENCES

Billingsley, P. (1968), *Convergence of Probability Measures*, New York: Wiley.

Bowman, D., Chen, J. J., and George, E. O. (1995), "Estimating Variance Functions in Developmental Toxicity Studies," *Biometrics*, 51, 1523–1528.

Fitzmaurice, G. M., and Laird, N. M. (1993), "A Likelihood-Based Method for Analysing Longitudinal Binary Responses," *Biometrika*, 80, 141–151.

Gourieroux, C., Monfort, A., and Trognon, A. (1984), "Pseudomaximum Likelihood Methods: Theory," *Econometrika*, 52, 681–700.

Holson, J. F., Gaines, T. B., Nelson, C. J., LaBorde, J. B., Gaylor, D. W., Sheehan, D. M., and Young, J. F. (1992), "Developmental Toxicity of 2,4,5-Trichlorophenoxyacetic Acid (2,4,5-T): I. Multireplicated Dose-Response Studies in Four Inbred Strains and One Outbred Stock of Mice," *Fundamental and Applied Toxicology*, 19, 286–297.

Jorgensen, B. (1987), "Exponential Dispersion Models" (with discussion), *Journal of the Royal Statistical Society*, Ser. B, 49, 127–162.

Kupper, L. L., Portier, C., Hogan, M. D., and Yamamoto, F. (1986), "The Impact of Litter Effects on Dose Response Modeling in Teratology," *Biometrics*, 42, 85–98.

LeCam, L. (1970), "On the Assumptions Used To Prove Asymptotic Normality of Maximum Likelihood Estimates," *The Annals of Mathematical Statistics*, 41, 802–828.

Lefkopoulou, M., Moore, D., and Ryan, L. (1989), "The Analysis of Multiple Correlated Binary Outcomes: Application to Rodent Teratology Experiments," *Journal of the American Statistical Association*, 84, 810–815.

Lehman, E. L. (1983), *Theory of Point Estimation*, New York: Wiley.

Leppik, I. E., Dreifuss, F. E., Bowman-Cloyd, T., Santilli, N., Jacobs, M., Crosby, C., Cloyd, J., Stockman, J., Graves, N., Sutula, T., Welty, T., Vickery, J., Brundage, R., Gumnit, R., and Gutierres, A. (1985), "A Double-Blind Crossover Evaluation of Progabide in Partial Seizures," *Neurology*, 35 (Suppl.), 285.

Liang, K. Y., and Hanfelt, J. (1994), "On the Use of the Quasi-Likelihood Method in Teratological Experiments," *Biometrics*, 50, 872–880.

Liang, K. Y., and Zeger, S. L. (1986), "Longitudinal Data Analysis Using Generalized Linear Models," *Biometrika*, 73, 13–22.

Liang, K. Y., Zeger, S. L., and Qaqish, B. (1992), "Multivariate Regression Analyses for Categorical Data" (with discussion), *Journal of the Royal Statistical Society*, Ser. B, 45, 3–40.

Mancl, L. (1992), "Regression Analysis of Correlated Discrete and Continuous Data: Evaluation of an Estimating Equation Approach," unpublished Ph.D. dissertation, University of Washington, Dept. of Biostatistics.

McCullagh, P. (1987), *Tensor Methods in Statistics*, London: Chapman and Hall.

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

Morris, C. N. (1982), "Natural Exponential Families With Quadratic Variance Functions," *The Annals of Statistics*, 10, 65–80.

Nelder, J. A., and Lee, Y. (1992), "Likelihood, Quasi-Likelihood, and Pseudolikelihood: Some Comparisons," *Journal of the Royal Statistical Society*, Ser. B, 54, 273–284.

Nelder, J. A., and Pregibon, D. (1987), "An Extended Quasi-Likelihood Function," *Biometrika*, 74, 221–232.

Nelder, J. A., and Wedderburn, R. W. M. (1972), "Generalized Linear Models," *Journal of the Royal Statistical Society*, Ser. A, 135, 370–384.

Prentice, R. L. (1988), "Correlated Binary Regression With Covariates Specific to Each Binary Observation," *Biometrics*, 44, 1033–1048.

Prentice, R. L., and Zhao, L. P. (1991), "Estimating Equations for Parameters in Means and Covariances of Multivariate Discrete and Continuous Responses," *Biometrics*, 47, 825–839.

Seber, G. A. F., and Wild, C. J. (1989), *Nonlinear Regression*, New York: Wiley.

Sim, C. H. (1993), "Generation of Poisson and Gamma Random Vectors With Given Marginals and Covariance Matrix," *Journal of Statistical Computing and Simulation*, 47, 1–10.

Wedderburn, R. W. M. (1974), "Quasi-Likelihood Functions, Generalized Linear Models, and the Gauss–Newton Method," *Biometrika*, 61, 439–47.

Williams, D. A. (1988), "Estimation Bias Using the Beta-Binomial Distribution in Teratology," *Biometrics*, 44, 305–309.

Zhao, L. P., and Prentice, R. L. (1990), "Correlated Binary Regression Using a Quadratic Exponential Model," *Biometrika*, 77, 642–648.

Zhu, Y., Krewski, D., and Ross, W. H. (1994), "Dose–Response Models for Correlated Multinomial Data From Developmental Toxicity Studies," *Applied Statistics*, 43, 583–598.