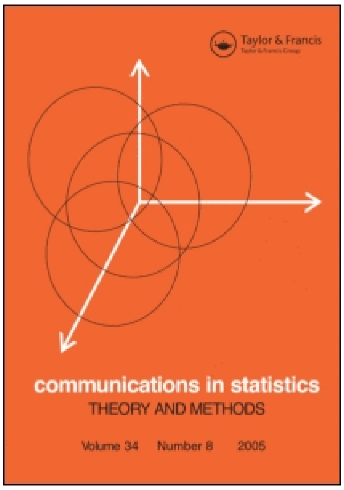

Communications in Statistics - Theory and Methods
Publication details, including instructions for authors and subscription information:
http://www.informaworld.com/smpp/title~content=t713597238

On gee-based regression estimators under first moment misspegification

Daniel B. Hall [a]
[a] Department of Statistics, University of Georgia, Athens, Georgia

To cite this Article Hall, Daniel B.(1999) 'On gee-based regression estimators under first moment misspegification', Communications in Statistics - Theory and Methods, 28: 5, 1021 — 1042
To link to this Article: DOI: 10.1080/03610929908832341
URL: http://dx.doi.org/10.1080/03610929908832341

COMMUN. STATIST.—THEORY METH., 28(5), 1021–1042 (1999)

ON GEE-BASED REGRESSION ESTIMATORS UNDER FIRST MOMENT MISSPECIFICATION

Daniel B. Hall

Department of Statistics, University of Georgia
Athens, Georgia 30602

Key Words: *clustered data, estimand, longitudinal data, quasi-likelihood.*

ABSTRACT

Many of the classical estimation methods of statistics lead to estimators that solve an equation. In common special cases, estimating equation-based estimators have appeal because they correspond to the maximum or minimum of an objective function. In such cases, an intuitively reasonable criterion for estimation, such as minimizing the Euclidean distance between the observation vector and the fitted value, motivates the procedure. In general, however, solutions of estimating equations need not minimize an objective function. Therefore, when the assumed model for the data is inaccurate, it is unclear what aspect of the data is being described by an estimating equation-based estimator. Since the landmark article of Liang and Zeger (1986), there has been considerable interest in using estimating equations for longitudinal and other clustered data. In this paper we examine the form of the regression parameter estimand under model misspecification when estimating equations related to Liang and Zeger's generalized estimating equations (GEEs) are used for model fitting. Closed form expressions are presented for these estimands in simple examples. These results indicate that

the estimands for GEE1 (Liang and Zeger, 1986) and extended GEE (Hall and Severini, 1998) are intuitively reasonable, but estimands based on GEE2 (Prentice and Zhao, 1991) in some cases are considerably more difficult to justify than their GEE1 counterparts.

1. INTRODUCTION

In clustered data correlation typically exists among within-cluster responses. This correlation complicates the analysis of such data. An approach to this problem which has received much attention recently is to analyze the data using a generalized linear marginal model which has parameters whose estimates are obtained as the solutions of equations. This approach was introduced by Liang and Zeger (1986) who proposed "generalized estimating equations" (GEEs) for the estimation of parameters related to the marginal mean response when the within-cluster correlation may be viewed as a nuisance. Subsequent authors have introduced alternative estimating equations for the situation in which correlation parameters are of interest. Prentice and Zhao (1991) proposed equations for the simultaneous estimation of regression and association parameters which have the same form as the score equations when the responses follow a quadratic exponential density (Gourieroux *et al.*, 1984). Following Liang *et al.* (1992), the Liang and Zeger (1986) and Prentice and Zhao (1991) approaches will hereafter be referred to as GEE1 and GEE2, respectively. Hall and Severini (1998) proposed "extended generalized estimating equations" (EGEEs) for regression and association parameters which are based on the ideas of extended quasi-likelihood.

A weakness of estimating equation-based parameter estimation is that, in general, solutions of estimating equations do not minimize an objective function. That is, the solution of an estimating equation does not necessarily minimize some measure of inaccuracy such as the mean squared deviations or the median absolute deviations of the responses from the regression surface. This means that when the assumed model for the data is inaccurate, it is unclear what aspect of the data is being described by an estimating equation-based estimator. That is, under model misspecification the *estimand* is unknown. Given that, to some extent, most models are unrealistic representations of the data they describe, this issue calls into question the appropriateness of GEE-based methods. In this paper GEE-based estimation of regression parameters is examined under model misspecification.

2. GEE-BASED METHODS

In the discussion that follows it will be assumed that the data to be analyzed are clustered due to a longitudinal study design. That is, multiple observations are made through time on each of several individuals. However, the estimating equation methods under consideration in this paper apply more generally to other types of clustered data.

To establish notation, let y_{it} be the scalar response and $\mathbf{x}_{it}$ be the $P\times 1$ vector of covariates for individuals $i = 1, \ldots, K$, and times $t = 1, \ldots, T_i$. The response vector for the i^{th} subject is $\mathbf{y}_i = (y_{i1}, \ldots, y_{iT_i})^T$ with mean $\boldsymbol{\mu}_i = (\mu_{i1}, \ldots, \mu_{iT_i})^T$, and the $T_i \times P$ matrix of covariates is $\mathbf{x}_i = (\mathbf{x}_{i1}, \ldots, \mathbf{x}_{iT_i})^{\mathrm{T}}$. Let $\boldsymbol{\beta}$ be a $P\times 1$ vector of parameters describing the relationship between the mean and the covariates, and let $\boldsymbol{\alpha}$ be an $s\times 1$ vector of parameters describing the associations among repeated observations. In each of the approaches it is assumed that the marginal mean is related to covariates through a known link function, g, via $g(\mu_{it}) = \mathbf{x}_{it}^T\boldsymbol{\beta}$, and the marginal variance is related to the marginal mean through a known variance function, v, via $\text{var}(y_{it}) = \phi v(\mu_{it})$, where ϕ is an unknown dispersion parameter.

2.1 *GEE1*

In the original GEE approach it is assumed that $\text{cov}(\mathbf{y}_i) = \phi\mathbf{V}_{i11}(\boldsymbol{\mu}_i, \boldsymbol{\alpha})$ where $\mathbf{V}_{i11}$ is a $T_i \times T_i$ covariance matrix which has the structure

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

Here $\mathbf{A}_i(\boldsymbol{\mu}_i) = \text{diag}(v(\mu_{i1}), \ldots, v(\mu_{1T_i}))$ is a $T_i \times T_i$ diagonal matrix of variance functions, and $\mathbf{R}(\boldsymbol{\alpha})$ is a "working" specification for $\text{corr}(\mathbf{y}_i)$. A quasi-likelihood approach based on these assumptions leads to the quasi-score equation

$$\sum_{i=1}^{K} \phi^{-1}\mathbf{D}_{i11}^T\mathbf{V}_{i11}^{-1}(\mathbf{y}_i - \boldsymbol{\mu}_i) = \mathbf{0}, \tag{2}$$

where $\mathbf{D}_{i11} = \partial\boldsymbol{\mu}_i/\partial\boldsymbol{\beta}$. Since (2) involves the unknown parameters $\boldsymbol{\beta}$, $\boldsymbol{\alpha}$, and ϕ, it is not an estimating equation for $\boldsymbol{\beta}$. Therefore, Liang and Zeger (1986) proposed their GEEs for $\boldsymbol{\beta}$ based on (2) where ϕ is replaced by an estimator, $\hat{\phi}(\boldsymbol{\beta})$, which is $\sqrt{K}$-consistent given $\boldsymbol{\beta}$, and $\boldsymbol{\alpha}$ is replaced by an estimator, $\hat{\boldsymbol{\alpha}}(\boldsymbol{\beta}, \phi)$, which is $\sqrt{K}$-consistent given $\boldsymbol{\beta}$ and ϕ. These substitutions yield GEE1:

$$\sum_{i=1}^{K} \hat{\phi}^{-1}(\boldsymbol{\beta})\mathbf{D}_{i11}^T\mathbf{V}_{i11}^{-1}[\boldsymbol{\mu}_i, \hat{\boldsymbol{\alpha}}\{\boldsymbol{\beta}, \hat{\phi}(\boldsymbol{\beta})\}](\mathbf{y}_i - \boldsymbol{\mu}_i) = \mathbf{0}. \tag{3}$$

For the estimation of the nuisance parameters ϕ and $\boldsymbol{\alpha}$, Liang and Zeger proposed method of moment estimators based on residuals. Prentice and Zhao (1991) generalized the method of moment approach by suggesting *ad hoc* estimating equations for $\boldsymbol{\alpha}$ similar in form to (3). The term GEE1 refers to either of these approaches, the critical feature being that GEE1 methods treat "$\boldsymbol{\beta}$ and $\boldsymbol{\alpha}$ as orthogonal to one another even when they are not" (Liang *et al.*, 1992, p.10). Under certain regularity conditions $\hat{\boldsymbol{\beta}}_{\text{GEE1}}$, the solution to (2), is consistent and asymptotically multivariate normal (Liang and Zeger, 1986) regardless of whether or not $\text{cov}(\mathbf{y}_i)$, $i = 1, \ldots, K$, have been modelled correctly.

2.2 *GEE2*

In the GEE2 and EGEE methods the decomposition (1) is dropped and the over-dispersion parameter ϕ is included in $\boldsymbol{\alpha}$. Hence $\boldsymbol{\alpha}$ describes

$\text{var}(\mathbf{y}_i)$ rather than $\text{corr}(\mathbf{y}_i)$. The GEE2 approach is based on the observation that when $\mathbf{y}_i$, $i = 1, \ldots, K$, are assumed to follow a quadratic exponential model, the score equation for $(\boldsymbol{\beta}^T, \boldsymbol{\alpha}^T)^T$ has the following form:

$$\sum_{i=1}^{K} \mathbf{D}_i^T \mathbf{V}_i^{-1} \begin{pmatrix} \mathbf{y}_i - \boldsymbol{\mu}_i \\ \mathbf{s}_i - \boldsymbol{\sigma}_i \end{pmatrix} = \mathbf{0}, \tag{4}$$

where $\boldsymbol{\sigma}_i = \boldsymbol{\sigma}_i(\boldsymbol{\beta}, \boldsymbol{\alpha}) = (\sigma_{i11}, \sigma_{i12}, \ldots, \sigma_{iT_iT_i})^T$ is the covariance matrix $\mathbf{V}_{i11}$ in vector form, and $\mathbf{s}_i = \mathbf{s}_i(\boldsymbol{\beta}) = (s_{i11}, s_{i12}, \ldots, s_{iT_iT_i})$, where $s_{ijk} = (y_{ij} - \mu_{ij})(y_{ik} - \mu_{ik})$. In addition, in (4)

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

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

One approach to estimation of $\boldsymbol{\beta}$ and $\boldsymbol{\alpha}$ would be to specify the form of the quadratic exponential model (which implies forms for the third and fourth moments appearing in $\mathbf{V}_i$) and solve (4). Depending on whether or not the true probability density function is assumed to belong to this model this approach corresponds to either maximum likelihood or pseudo maximum likelihood estimation (Gourieroux *et al.*, 1984). Since this procedure is typically expensive computationally, Prentice and Zhao (1991) suggested obtaining estimates by solving (4) assuming a "working" form for $\mathbf{V}_i$, $i = 1, \ldots, K$, where third and fourth moments are modelled in terms of $\boldsymbol{\mu}_i$ and $\boldsymbol{\sigma}_i$. This GEE2 approach yields consistent and asymptotically normal estimators provided that the mean and covariance models, $\mu_i(\boldsymbol{\beta})$, $\boldsymbol{\sigma}_i(\boldsymbol{\beta}, \boldsymbol{\alpha})$, $i = 1, \ldots, K$, have been correctly specified.

2.3 *EGEE*

The GEE1 procedure disussed in section 2.1 may be seen as an application of quasi-likelihood to longitudinal data. Equation (2) is the quasi-score equation which results upon differentiation of quasi-likelihood functions

$Q_i(\boldsymbol{\mu}_i, \mathbf{y}_i)$, $i = 1, \ldots, K$, and summation over i. These Q_i, $i = 1, \ldots, K$, have likelihood-like properties with respect to $\boldsymbol{\beta}$ but not with respect to $\boldsymbol{\alpha}$ and ϕ. Hall and Severini (1998) proposed estimating equations for $\boldsymbol{\gamma} = (\boldsymbol{\beta}^T, \boldsymbol{\alpha}^T)^T$ formed from the derivatives of an extended quasi-likelihood function for γ, $Q_i^+(\boldsymbol{\mu}_i, \boldsymbol{\alpha}, \mathbf{y}_i)$. Q_i^+ is an extended quasi-likelihood function (Nelder and Pregibon, 1987) in the sense that its partial derivatives with respect to both $\boldsymbol{\beta}$ and $\boldsymbol{\alpha}$ have properties similar to an efficient score vector.

Specifically, the extended quasi-likelihood function is assumed to be of the form

$$Q_i^+(\boldsymbol{\mu}_i, \boldsymbol{\alpha}, \mathbf{y}_i) = Q_i(\boldsymbol{\mu}_i, \mathbf{y}_i) + f_1(\boldsymbol{\alpha}) + f_2(\mathbf{y}_i),$$

which ensures that Q_i^+ retain the log-likelihood-like properties of Q_i with respect to $\boldsymbol{\beta}$ since $\partial Q_i^+/\partial \boldsymbol{\beta} = \partial Q_i/\partial \boldsymbol{\beta}$. In the following, we drop the individual index i for simplicity of notation. The requirement

$$\mathrm{E}\left(\frac{\partial Q^+}{\partial \boldsymbol{\alpha}}\right) = \mathbf{0}$$

leads to

$$\frac{\partial f_1(\boldsymbol{\alpha})}{\partial \alpha^u} \approx \frac{1}{2} V^{jk} V^u_{jk} = \frac{1}{2} R_{jk} R^{jk,u}, \quad u = 1, \ldots, s.$$

Here we use index notation (see for example, Barndorff-Nielsen and Cox, 1989, ch.5) where subscripts and superscripts indicate vector or matrix components and a summation is implied over any index which is repeated in a subscript and a superscript. In our notation V^{jk} is the $(j,k)^{\text{th}}$ element of $\mathbf{V}_{i11}$, V_{jk} is the $(j,k)^{\text{th}}$ element of $\mathbf{V}_{i11}^{-1}$, R_{jk} is the $(j,k)^{\text{th}}$ element of $\mathbf{R}(\boldsymbol{\alpha})$, and R^{jk} is the $(j,k)^{\text{th}}$ element of $\mathbf{R}^{-1}(\boldsymbol{\alpha})$. The superscript u represents a partial derivative with respect to the u^{th} element of $\boldsymbol{\alpha}$, α^u. The partial derivatives of the resulting extended quasi-likelihood function with respect to the components of $\boldsymbol{\beta}$ and $\boldsymbol{\alpha}$ give the following estimating functions:

$$\begin{aligned} U(\gamma; \beta^b) &= (y^j - \mu^j) V_{ja} (\partial \mu^a / \partial \beta^b), \quad b = 1, \ldots, p, \\ U(\gamma; \alpha^u) &= -(y^j - \mu^j)(y^k - \mu^k) V^u_{jk} + R_{jk} R^{jk,u}, \quad u = 1, \ldots, s. \end{aligned} \tag{5}$$

Stacking these estimating functions and summing over independent clusters yields the EGEE for γ:

$$\sum_{i=1}^{K} U_i(\gamma) = \mathbf{0}, \tag{6}$$

where

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

It can easily be seen by shifting to matrix notation that the estimating equation for $\boldsymbol{\beta}$ corresponding to (6), $\sum_{i=1}^{K} U_i(\gamma; \boldsymbol{\beta}) = \mathbf{0}$, is exactly the same as (2). Therefore, as functions of an estimator for $\boldsymbol{\alpha}$, the GEE1 and EGEE estimators of $\boldsymbol{\beta}$ are the same. For this reason we will emphasize comparisons between GEE1 and GEE2 in what follows, with the understanding that the results hold for comparisons of EGEE versus GEE2 as well.

2.4 *Relationship with Optimal Estimating Equations*

See Desmond (1997) for a thorough treatment of the issues addressed in this subsection.

Assuming that $\boldsymbol{\alpha}$ and ϕ are known, the quasi-score equation (2) is the optimal linear estimating equation for $\boldsymbol{\beta}$ in the sense of Godambe (1960). However, the situation Liang and Zeger focused on when proposing their original GEEs is one in which the intracluster correlation is a nuisance and the true structure for $\text{cov}(\mathbf{y}_i)$ is unknown. Therefore, these authors proposed the suboptimal estimating equations given in (3), trading optimality for robustness. Here we take "robustness" to mean that inference for $\boldsymbol{\beta}$ "is insensitive to whether or not other parts of the probability mechanism are misspecified" (Liang and Rathouz, 1997, p.113). In addition, the *ad hoc* estimating equation of Prentice and Zhao (1991) that can be used to supplement (3) in the GEE1 approach is optimal for $\boldsymbol{\alpha}$ for fixed $\boldsymbol{\beta}$. However, for the combined estimation of $\boldsymbol{\alpha}$ and ϕ, GEE1 is not optimal.

Among the class of unbiased estimating functions that are linear combinations of $\mathbf{y}_i$ and $\mathbf{s}_i$, it is known that equation (4) is optimal if $\mathbf{V}_i = \text{cov}(\mathbf{y}_i, \mathbf{s}_i)$ is a function of $\boldsymbol{\beta}$ and $\boldsymbol{\alpha}$ only (Liang *et al*, 1992). Typically, however, $\mathbf{V}_i$ will involve third and fourth moment parameters. If such higher order parameters are known, then (4) is again optimal. Since knowledge of these third and fourth moments is typically unavailable, Prentice and Zhao (1991) proposed using "working" models for $\text{cov}(\mathbf{y}_i, \mathbf{s}_i)$ and $\text{var}(\mathbf{s}_i)$ in their estimating equations, again trading optimality for robustness.

Hall and Severini (1998) use a different approach than Prentice and Zhao to avoid the assumption of models for third and fourth order moments. The alternatives offered by these pairs of authors parallels the situation in the literature on estimation of overdispersion parameters in generalized linear models. In overdispersed GLMs the optimal estimating function of Godambe and Thompson (1989) (which they call the extended quasi-score function in a different use of the terminology of Nelder and Pregibon, 1987) requires knowledge of skewness and kurtosis. To avoid such higher moment specifications, two prominent approaches are extended quasi-likelihood (Nelder and Pregibon, 1987) and pseudo-likelihood (Carroll and Ruppert, 1982; Davidian and Carroll, 1988).

The connection between EGEE and extended quasi-likelihood (EQL) has already been established. As described in Prentice and Zhao (1991), the GEE2 approach is closely connected with the pseudo maximum likelihood (PML) approach of Gourieroux et al. (1984). PML differs from, but is closely related to, Carroll and Ruppert's pseudo likelihood (PL). In PML, estimators for first and second moment parameters are obtained by maximizing a density in the quadratic exponential family when the true density does not necessarily belong to that family. Gourieroux *et al* showed that this procedure yields consistent estimators, and that the quadratic exponential family was unique in implying such consistency. In PL, one fixes $\boldsymbol{\beta}$ at a preliminary estimate and estimates second order parameters by maximizing the normal theory likelihood. Clearly, the two methods are closely related. In each case consistent estimators are obtained by maximizing a possibly incorrect likelihood in the exponential family. Currently, there seems to be some evidence in favor of each of the methods, EQL and PL, in the overdispersed GLM setting. Further investigation of the relative merits of GEE2 and EGEE in the clustered data setting seems warranted.

3. MODEL-MISSPECIFICATION

The *only* requirement that is typically made of estimating functions is that they have zero mean at the true parameter point (see, for example, McCullagh, 1991). Consider the quasi-score function $\mathbf{D}_{i11}^T\mathbf{V}_{i11}^{-1}\{\mathbf{y}_i - \boldsymbol{\mu}_i(\boldsymbol{\beta})\}$ for a single cluster (the i^{th}) assuming, for the moment, that the association parameter $\boldsymbol{\alpha}$ is known. Under misspecification of the model for $\mathrm{E}(\mathbf{y}_i)$, the quasi-score function is clearly not an estimating function for the true parameter value, call it $\boldsymbol{\beta}_0$. For example, suppose that the assumed model is $\boldsymbol{\mu}_i(\boldsymbol{\beta}) = (\beta_i, \ldots, \beta_i)^T$, but under the true model $\mathrm{E}(\mathbf{y}_i) = \boldsymbol{\mu}_{0i}(\boldsymbol{\beta}_0) = (\beta_{0i1}, \ldots, \beta_{0iT_i})^T$. In this case, the quasi-score function does not have zero mean at the true parameter point because it cannot be evaluated there; $\boldsymbol{\beta}$ and $\boldsymbol{\beta}_0$ do not have the same dimension. Certainly,

$$\mathrm{E}_0[\mathbf{D}_{i11}^T\mathbf{V}_{i11}^{-1}\{\mathbf{y}_i - \boldsymbol{\mu}_i(\boldsymbol{\beta})\}] = \mathbf{0}$$

does not imply that $\boldsymbol{\beta} = \boldsymbol{\beta}_0$.

By the same argument, equation (2), the GEE1 equation when $\boldsymbol{\alpha}$ and ϕ are known, is not an estimating equation for $\boldsymbol{\beta}_0$, and the GEE2 equation (4) is not an estimating equation for $(\boldsymbol{\beta}_0^T, \boldsymbol{\alpha}_0^T)^T$ under first moment model misspecification. However, both (2) and (4) are estimating functions with respect to some quantities, and it is of interest to examine what those quantities are. The quantity for which (2) is an estimating function satisfies

$$\mathrm{E}_0\left[\sum_{i=1}^{K}\mathbf{D}_{i11}^T\mathbf{V}_{i11}^{-1}\{\mathbf{y}_i - \boldsymbol{\mu}_i(\boldsymbol{\beta})\}\right] = \mathbf{0},$$

or

$$\begin{aligned}\sum_{i=1}^{K}&\mathbf{D}_{i11}^T\{\boldsymbol{\mu}_i(\boldsymbol{\beta})\}\mathbf{V}_{i11}^{-1}\{\boldsymbol{\mu}_i(\boldsymbol{\beta}),\boldsymbol{\alpha}\}\boldsymbol{\mu}_i(\boldsymbol{\beta})\\ &=\sum_{i=1}^{K}\mathbf{D}_{i11}^T\{\boldsymbol{\mu}_i(\boldsymbol{\beta})\}\mathbf{V}_{i11}^{-1}\{\boldsymbol{\mu}_i(\boldsymbol{\beta}),\boldsymbol{\alpha}\}\boldsymbol{\mu}_{0i}.\end{aligned} \tag{7}$$

Here E_0 indicates that the expectation is taken with respect to the true distribution of $\mathbf{y}_i$. We call the solution of (7) the estimand for GEE1 and denote it $\boldsymbol{\beta}^*$. The estimand of GEE2 is the solution of the equation obtained by applying E_0 to the left-hand side of (4), which yields

$$\sum_{i=1}^{K}\mathbf{D}_i^T\mathbf{V}_i^{-1}\begin{pmatrix}\boldsymbol{\mu}_i(\boldsymbol{\beta})\\ \boldsymbol{\sigma}_i(\boldsymbol{\beta})\end{pmatrix} = \sum_{i=1}^{K}\mathbf{D}_i^T\mathbf{V}_i^{-1}\begin{pmatrix}\boldsymbol{\mu}_{0i}\\ \boldsymbol{\sigma}_{0i}\end{pmatrix}. \tag{8}$$

The GEE2 estimand solving (8) we denote as $(\boldsymbol{\beta}^{**T}, \boldsymbol{\alpha}^{**T})^T$.

In general, it is not possible to solve either (7) or (8) in closed form. When, according to the assumed model, $\boldsymbol{\mu}_i(\boldsymbol{\beta}) = \mathbf{x}_i\boldsymbol{\beta}$, that is, the link function is linear, or when $g(\mu_{it})$ is equal to a single component of the parameter vector (e.g., as in a means model for the one-way layout), the model mean $\boldsymbol{\mu}_i(\boldsymbol{\beta})$ has the form $\mathbf{x}_i\boldsymbol{\lambda}$, where $\boldsymbol{\lambda} = g^{-1}(\boldsymbol{\beta})$. In this case, the form of $\boldsymbol{\lambda}^* = g^{-1}(\boldsymbol{\beta}^*)$ implied by equation (7) is easy to obtain, and for some simple models, (8) can also be solved to yield a closed form expression for $\boldsymbol{\lambda}^{**} = g^{-1}(\boldsymbol{\beta}^{**})$.

In sections 4 and 5 two simple examples are considered where equations (7) and (8) yield solutions. Results depend upon the model chosen for $\mathrm{var}(\mathbf{y}_i)$. In the examples discussed below, covariance models corresponding to (1) with exchangeable, 1-dependence, and AR(1) correlation structures are considered (see Liang and Zeger, 1986, for a description of these correlation structures). In addition, three variance functions are utilized:

$$v_1(\mu) = \mu, \quad v_2(\mu) = 1, \quad v_3(\mu) = \mu^2.$$

In general, $\boldsymbol{\lambda}^*$ depends on $\boldsymbol{\alpha}$, which is assumed known.

For GEE2, the form of $\boldsymbol{\lambda}^{**}$ depends not only on the model for $\mathrm{cov}(\mathbf{y}_i)$, but also on the working models for the third and fourth order moments appearing in $\mathbf{V}_{i21}$ and $\mathbf{V}_{i22}$. The independence and Gaussian working models for higher order moments are used here. These models will be denoted GEE2(I) and GEE2(G), respectively, in what follows. A detailed description of these models may be found in Prentice and Zhao (1991). For GEE2, the first moment estimand, $\boldsymbol{\lambda}^{**}$, typically depends upon $\boldsymbol{\alpha}^{**}$. For most of the results that follow, we do not assume that the correlation model has been correctly specified. In such cases, $\boldsymbol{\alpha}^{**}$ is not necessarily equal to the true correlation parameter $\boldsymbol{\alpha}_0$. As a result, most of the expressions for $\boldsymbol{\lambda}^{**}$ involve both $\boldsymbol{\alpha}^{**}$ and $\boldsymbol{\alpha}_0$.

4. SUBJECT-SPECIFIC MEANS ASSUMED CONSTANT OVER TIME

Consider the case in which K individuals are measured at each of three time points; that is $T_i = 3$, $i = 1, \ldots, K$. Suppose that $\mathbf{y} = (y_{11}, y_{12}, y_{13}, y_{21}, \ldots, y_{K3})$ has true mean $\boldsymbol{\mu}_0 = (\mu_{011}, \mu_{012}, \mu_{013}, \mu_{021}, \ldots, \mu_{0K3})$ which is not of the form $\boldsymbol{\mu}(\boldsymbol{\beta}) = g^{-1}(\mathbf{X}\boldsymbol{\beta})$ specified by the GLM. Suppose that $\mathbf{X}$ is a design matrix grouping responses according to individual. That is, $\mathbf{X}$ is the $3K \times K$ matrix $\mathbf{I}_K \otimes \mathbf{1}_3$ where $\mathbf{1}_a$ is a $a \times 1$ vector of ones, $\mathbf{I}_a$ is the $a \times a$ identity matrix, and $\otimes$ denotes the Kronecker product. Here, β_p, the parameter corresponding to the p^{th} column of $\mathbf{X}$, has the interpretation that the mean response for individual p equals $\lambda_p = g^{-1}(\beta_p)$, $p = 1, \ldots, P$, $P = K$.

We can think of the K subjects as belonging to K treatment groups with distinct means. In applications we are more likely to encounter the situation in which we have distinct means in g treatment groups with K_i subjects in the i^{th} group ($K = K_1 + \cdots + K_g$). The estimands in the g treatment example are more difficult to obtain, but some examples suggest that the results for the K group example presented here are suggestive of what occurs in the g group case.

4.1 *Case 1: Exchangeable Correlation Structure*

For the exchangeable model, (7) can be solved to yield

$$\lambda_p^* = \frac{1}{3}(\mu_{0p1} + \mu_{0p2} + \mu_{0p3}), \quad p = 1, \ldots, P, \tag{9}$$

for all three choices of the variance function. Clearly the arithmetic average of the true means at each time point is a reasonable form for the GEE1 regression parameter estimand.

The corresponding GEE2(I) and GEE2(G) estimators depend upon the variance function v. Results for GEE2(I) will be considered first. Using variance function v_1, equation (8) yields two solutions: one corresponding to $\lambda_p^{**} = \lambda_p^*$ given by (9) and the other given by

$$\lambda_p^{**} = \frac{\alpha_{01}\sqrt{\mu_{0p1}\mu_{0p2}} + \alpha_{02}\sqrt{\mu_{0p1}\mu_{0p3}} + \alpha_{03}\sqrt{\mu_{0p2}\mu_{0p3}}}{3\alpha^{**}}, \quad p = 1, \ldots, P.$$

Here α^{**} is the model-based estimator which (with $\boldsymbol{\beta}^{**}$) solves (8), and α_{01}, α_{02}, α_{03} are the true correlations; that is

$$\text{corr}_0(\mathbf{y}_i) = \begin{pmatrix} 1 & \alpha_{01} & \alpha_{02} \\ \alpha_{01} & 1 & \alpha_{03} \\ \alpha_{02} & \alpha_{03} & 1 \end{pmatrix}.$$

For v_2, $\boldsymbol{\lambda}^{**} = \boldsymbol{\lambda}^*$. For v_3, (8) leads to solutions of the form

$$\lambda_{pk}^{**} = \pm\sqrt{\frac{s_{1k}\alpha_{01}\mu_{0p1}\mu_{0p2} + s_{2k}\alpha_{02}\mu_{0p1}\mu_{0p3} + s_{3k}\alpha_{03}\mu_{0p2}\mu_{0p3}}{3\alpha^{**}}},$$

$p = 1, \ldots, P$, $k = 1, \ldots, 4$, where $\mathbf{s}_1 = (1, -1, -1, 1)^T$, $\mathbf{s}_2 = (1, 1, -1, -1)^T$, and $\mathbf{s}_3 = (1, -1, 1, -1)^T$.

Using GEE2(G) and the identity variance function, v_1, leads to

$$\lambda_p^{**} = \frac{c(2\alpha^{**2}+1) - \mu_{0p\cdot}(\alpha^{**2}+2\alpha^{**})}{3\alpha^{**}(2\alpha^{**}+1)(\alpha^{**}-1)}, \quad p = 1, \ldots, P,$$

where $c = \alpha_{01}\sqrt{\mu_{0p1}\mu_{0p2}} + \alpha_{02}\sqrt{\mu_{0p1}\mu_{0p3}} + \alpha_{03}\sqrt{\mu_{0p2}\mu_{0p3}}$ and $\mu_{0p\cdot} = \sum_{j=1}^{3}\mu_{0pj}$. Variance function v_2 again leads to the solution given by (9). When v_3 is used equation (8) yields four solutions corresponding to

$$\lambda_{pk}^{**} = \frac{2(s_{1k}m_{p1} + s_{2k}m_{p2} + s_{3k}m_{p3})(\alpha^{**2}+\alpha^{**}+1) - 3d\alpha^{**}(\alpha^{**}+1)}{\mu_{0p\cdot}\alpha^{**}(2\alpha^{**}+1)(\alpha^{**}-1)},$$

$p = 1, \ldots, P$, $k = 1, \ldots, 4$, where $m_{p1} = \alpha_{01}\mu_{0p1}\mu_{0p2}$, $m_{p2} = \alpha_{02}\mu_{0p1}\mu_{0p3}$, $m_{p3} = \alpha_{03}\mu_{0p2}\mu_{0p3}$, and $d = \sum_{j=1}^{3}\mu_{0pj}^2$.

The results from this example under the exchangeable working correlation structure are summarized in table 1. While for the identity variance function, v_2, the GEE2 regression parameter estimands correspond to an intuitively appealing form agreeing with GEE1 and EGEE, in general GEE2 can lead to quite complicated and difficult to interpret estimands when the regression model is misspecified. This disadvantage to the use of GEE2 will be seen in several of the other examples considered below.

4.2 *Case 2: 1-Dependence Correlation Structure*

For the 1-dependence correlation model, (7) leads to

$$\lambda_p^{*} = \frac{1}{4\alpha^{*}-3}\{\mu_{0p1}(\alpha^{*}-1) + \mu_{0p2}(2\alpha^{*}-1) + \mu_{0p3}(\alpha^{*}-1)\}, \quad p = 1, \ldots, P, \tag{10}$$

for all choices of the variance function. This form is a weighted average of the true means. For $\alpha^* = 0$ the weights are equal as in the exchangeable case. For large α^*, more weight is put upon the second observation than the first or third. This differential weighting reflects the fact that under the 1-dependence correlation structure, observation two is correlated with both of the other observations and observations one and three are each correlated

TABLE 1 – GEE-BASED ESTIMANDS UNDER MODEL MISSPECIFICATION

Exchangeable Working Correlation Structure

Method	v	λ_p^* (GEE1) or λ_p^{**} (GEE2)
GEE1/EGEE	v_{const}	$\frac{1}{3}(\mu_{0p1} + \mu_{0p2} + \mu_{0p3}) = \mu_{0p\cdot}/3$
GEE1/EGEE	v_{id}	same as above
GEE1/EGEE	v_{sq}	same as above
GEE2(I)	v_{const}	$\frac{1}{3}(\mu_{0p1} + \mu_{0p2} + \mu_{0p3})$
GEE2(I)	v_{id}	$\frac{1}{3}(\mu_{0p1} + \mu_{0p2} + \mu_{0p3})$ and $\frac{\alpha_{01}\sqrt{\mu_{0p1}\mu_{0p2}} + \alpha_{02}\sqrt{\mu_{0p1}\mu_{0p3}} + \alpha_{03}\sqrt{\mu_{0p2}\mu_{0p3}}}{3\alpha^{**}}$
GEE2(I)	v_{sq}	$\pm\sqrt{\frac{s_{1k}\alpha_{01}\mu_{0p1}\mu_{0p2} + s_{2k}\alpha_{02}\mu_{0p1}\mu_{0p3} + s_{3k}\alpha_{03}\mu_{0p2}\mu_{0p3}}{3\alpha^{**}}}$
GEE2(G)	v_{const}	$\frac{1}{3}(\mu_{0p1} + \mu_{0p2} + \mu_{0p3})$
GEE2(G)	v_{id}	$\frac{(\alpha_{01}\sqrt{\mu_{0p1}\mu_{0p2}} + \alpha_{02}\sqrt{\mu_{0p1}\mu_{0p3}} + \alpha_{03}\sqrt{\mu_{0p2}\mu_{0p3}})(2\alpha^{**2} + 1) - \mu_{0p\cdot}(\alpha^{**2} + 2\alpha^{**})}{3\alpha^{**}(2\alpha^{**} + 1)(\alpha^{**} - 1)}$
GEE2(G)	v_{sq}	$\frac{2(s_{1k}\alpha_{01}\mu_{0p1}\mu_{0p2} + s_{2k}\alpha_{02}\mu_{0p1}\mu_{0p3} + s_{3k}\alpha_{03}\mu_{0p2}\mu_{0p3})(\alpha^{**2} + \alpha^{**} + 1) - 3\alpha^{**}(\alpha^{**} + 1)\sum_{j=1}^{3}\mu_{0pj}^2}{\mu_{0p\cdot}\alpha^{**}(2\alpha^{**} + 1)(\alpha^{**} - 1)}$

with only one other observation. Hence, in some sense, when α^* is large observation two contains the most information about the overall mean.

When using GEE2(I), (8) yields

$$\lambda_p^{**} = \frac{\alpha_{01}\sqrt{\mu_{0p1}\mu_{0p2}} + \alpha_{03}\sqrt{\mu_{0p2}\mu_{0p3}}}{2\alpha^{**}}, \quad p = 1, \ldots, P,$$

when v_1 is the variance function. When v_2 is used, the estimand for GEE2(I) again coincides with the estimand for GEE1; that is, $\boldsymbol{\lambda}^{**}$ is as in (10). When v_3 is used, $\boldsymbol{\lambda}^{**}$ is given by

$$\lambda_{pk}^{**} = \sqrt{\frac{s_{1k}(\alpha_{01}\mu_{0p1}\mu_{0p2} - s_{2k}\alpha_{03}\mu_{0p2}\mu_{0p3})}{2\alpha^{**}}}, \quad p = 1, \ldots, P, \quad k = 1, \ldots, 4.$$

Under GEE2(G), v_1 leads to λ_p^{**} equal to

$$\{2\alpha^{**2}(\alpha_{01}\sqrt{\mu_{0p1}\mu_{0p2}} + \alpha_{03}\sqrt{\mu_{0p2}\mu_{0p3}}) + (\alpha_{01}\sqrt{\mu_{0p1}\mu_{0p2}} + \alpha_{03}\sqrt{\mu_{0p2}\mu_{0p3}}) - \alpha^{**}(2\alpha_{02}\sqrt{\mu_{0p1}\mu_{0p3}} + \mu_{0p\cdot} + \mu_{0p2})\}/\{2\alpha^{**}(2\alpha^{**2} - 1)\}$$

as the solution to (8). The constant variance function v_2 again yields $\boldsymbol{\lambda}^{**}$ as in (10). With v_3, GEE2(G) leads to the p^{th} component of $\boldsymbol{\lambda}^{**}$ corresponding to one of the four solutions:

$$\Big\{\alpha^{**}(-4\alpha^{**2} - 2\alpha^{**} + 3)(\mu_{0p2}^2 - 2s_{1k}m_{p2}) - 2\alpha^{**}(\alpha^{**} - 1)d - (2\alpha^{**4} + 4\alpha^{**3} - \alpha^{**2} + 2\alpha^{**} - 3)(s_{3k}m_{p1} - s_{2k}m_{p3}) - 4s_{1k}\alpha^{**5}m_{p2} - 2\alpha^{**5}(\mu_{0p1}^2 + \mu_{0p3}^2)\Big\}\Big/\alpha^{**}(2\alpha^{**2} - 1)\{\alpha^{**}(\mu_{0p\cdot} + \mu_{0p2}) - d\},$$

$k = 1, \ldots, 4$. The results from this section are summarized in table 2. In this case as in the previous one, the GEE2 estimands under non-identity variance functions appear to be difficult to justify.

4.3 *Case 3: AR(1) Correlation Structure*

When the correlation structure is assumed to follow an AR(1) model (7) has solution which, again, is a weighted average:

$$\lambda_p^* = \frac{\mu_{0p1} + \mu_{0p2}(1 - \alpha^*) + \mu_{0p3}}{3 - \alpha^*}, \quad p = 1, \ldots, P. \tag{11}$$

TABLE 2 – GEE-BASED ESTIMANDS UNDER MODEL MISSPECIFICATION
1-dependence Working Correlation Structure

Method	v	λ_p^*
GEE1/EGEE	v_{const}	$\frac{1}{4\alpha^* - 3}\{\mu_{0p1}(\alpha^* - 1) + \mu_{0p2}(2\alpha^* - 1) + \mu_{0p3}(\alpha^* - 1)\}$
GEE1/EGEE	v_{id}	same as above
GEE1/EGEE	v_{sq}	same as above
GEE2(I)	v_{const}	$\frac{1}{4\alpha^{**} - 3}\{\mu_{0p1}(\alpha^{**} - 1) + \mu_{0p2}(2\alpha^{**} - 1) + \mu_{0p3}(\alpha^{**} - 1)\}$
GEE2(I)	v_{id}	$\frac{\alpha_{01}\sqrt{\mu_{0p1}\mu_{0p2}} + \alpha_{03}\sqrt{\mu_{0p2}\mu_{0p3}}}{2\alpha^{**}}$
GEE2(I)	v_{sq}	$\sqrt{\frac{s_{1k}(\alpha_{01}\mu_{0p1}\mu_{0p2} - s_{2k}\alpha_{03}\mu_{0p2}\mu_{0p3})}{2\alpha^{**}}}$
GEE2(G)	v_{const}	$\frac{1}{4\alpha^{**} - 3}\{\mu_{0p1}(\alpha^{**} - 1) + \mu_{0p2}(2\alpha^{**} - 1) + \mu_{0p3}(\alpha^{**} - 1)\}$
GEE2(G)	v_{id}	$\frac{(2\alpha^{**2} + 1)(\alpha_{01}\sqrt{\mu_{0p1}\mu_{0p2}} + \alpha_{03}\sqrt{\mu_{0p2}\mu_{0p3}}) - \alpha^{**}(2\alpha_{02}\sqrt{\mu_{0p1}\mu_{0p3}} + \mu_{0p\cdot} + \mu_{0p2})}{2\alpha^{**}(2\alpha^{**2} - 1)}$
GEE2(G)	v_{sq}	$\Big\{\alpha^{**}(-4\alpha^{**2} - 2\alpha^{**} + 3)(\mu_{0p2}^2 - 2s_{1k}\alpha_{02}\mu_{0p1}\mu_{0p3}) - 2\alpha^{**}(\alpha^{**} - 1)\sum_{j=1}^{3}\mu_{0pj}^2 - 2\alpha^{**5}(\mu_{0p1}^2 + \mu_{0p3}^2)$ $-(2\alpha^{**4} + 4\alpha^{**3} - \alpha^{**2} + 2\alpha^{**} - 3)(s_{3k}\alpha_{01}\mu_{0p1}\mu_{0p2} - s_{2k}\alpha_{03}\mu_{0p2}\mu_{0p3}) - 4s_{1k}\alpha^{**5}\alpha_{02}\mu_{0i1}\mu_{0i3}\Big\}$ $\Big/\alpha^{**}(2\alpha^{**2} - 1)\left\{\alpha^{**}(\mu_{0p\cdot} + \mu_{0p2}) - \sum_{j=1}^{3}\mu_{0pj}^2\right\}$

Here again, when $\alpha^* = 0$, $(\mu_{0p1}, \mu_{0p2}, \mu_{0p3})$ receive equal weights. When α^* is large and positive, μ_{0p2} is down-weighted. When α^* is large and negative, μ_{0p2} receives greater weight. As in the previous cases, all choices of the variance function v lead to the same GEE1 and EGEE estimand.

The GEE2(I) and GEE2(G) estimands depend on the choice for v. Using GEE2(I) with v_1 leads to

$$\lambda_p^{**} = \frac{\alpha_{01}\sqrt{\mu_{0p1}\mu_{0p2}} + 2\alpha_{02}\sqrt{\mu_{0p1}\mu_{0p3}} + \alpha_{03}\sqrt{\mu_{0p2}\mu_{0p3}}}{2\alpha^{**}(1+\alpha^{**2})}, \quad p = 1, \ldots, P.$$

When v_2 is used GEE2(I) yields $\boldsymbol{\lambda}^{**}$ as in (11). Using v_3 under GEE2(I) equation (8) yields four solutions for λ_p^{**}:

$$\lambda_{pk}^{**} = \frac{(\alpha^{**3} + \alpha^{**2} + \alpha^{**} + 1)(d + 2s_{1k}m_{p2}) + 4s_{1k}m_{p2}}{(1+\alpha^{**2})(\alpha^{**}\mu_{0p2} - \mu_{0p\cdot})} - \frac{(\alpha^{**5} + \alpha^{**4} - \alpha^{**} - 3)(s_{3k}m_{p1} - s_{2k}m_{p3})}{\alpha^{**}(1+\alpha^{**2})(\alpha^{**}\mu_{0p2} - \mu_{0p\cdot})},$$

$p = 1, \ldots, P$, $k = 1, \ldots, 4$.

Solutions of (8) obtained under GEE2(G) are as follows. When v_1 is the chosen variance function, λ_p^{**} equals

$$\frac{\alpha_{01}\sqrt{\mu_{0p1}\mu_{0p2}}(\alpha^{**2}+1) + \alpha_{03}\sqrt{\mu_{0p2}\mu_{0p3}}(\alpha^{**2}+1) - \alpha^{**}(\mu_{0p\cdot} + \mu_{0p2})}{2\alpha^{**}(\alpha^{**}-1)(\alpha^{**}+1)},$$

$p = 1, \ldots, P$. For v_2, equation (8) again yields the solution given by equation (11). Four possible estimands are obtained when variance function v_3 is utilized, given by λ_{pk}^{**} equal to

$$\frac{(-\alpha^{**4} - \alpha^{**3} + \alpha^{**2} + 3\alpha^{**})\mu_{0p2}^2 + 2\alpha^{**}d}{\alpha^{**}(1-\alpha^{**2})\{\mu_{0p1} + \mu_{0p2}(1-\alpha^{**}) + \mu_{0p3}\}} - \frac{(\alpha^{**3} - \alpha^{**2} - \alpha^{**} - 3)(s_{2k}m_{p1} + s_{1k}m_{p3})}{\alpha^{**}(1-\alpha^{**2})\{\mu_{0p1} + \mu_{0p2}(1-\alpha^{**}) + \mu_{0p3}\}},$$

for $p = 1, \ldots, P$, and $k = 1, \ldots, 4$. Results for this example are summarized in table 3.

TABLE 3 GEE-BASED ESTIMANDS UNDER MODEL MISSPECIFICATION

AR(1) Working Correlation Structure

Method	v	λ_p^*
GEE1/EGEE	v_{const}	$\frac{1}{3-\alpha^*}\{\mu_{0p1}+\mu_{0p2}(1-\alpha^*)+\mu_{0p3}\}$
GEE1/EGEE	v_{id}	same as above
GEE1/EGEE	v_{sq}	same as above
GEE2(I)	v_{const}	$\frac{1}{3-\alpha^{**}}\{\mu_{0p1}+\mu_{0p2}(1-\alpha^{**})+\mu_{0p3}\}$
GEE2(I)	v_{id}	$\frac{\alpha_{01}\sqrt{\mu_{0p1}\mu_{0p2}}+2\alpha_{02}\sqrt{\mu_{0p1}\mu_{0p3}}+\alpha_{03}\sqrt{\mu_{0p2}\mu_{0p3}}}{2\alpha^{**}(1+\alpha^{**2})}$
GEE2(I)	v_{sq}	$\frac{(\alpha^{**4}+\alpha^{**3}+\alpha^{**2}+\alpha^{**})(d+2s_{1k}m_{p2})-(\alpha^{**5}+\alpha^{**4}-\alpha^{**}-3)(s_{3k}m_{p1}-s_{2k}m_{p3})}{\alpha^{**}(1+\alpha^{**2})(\alpha^{**}\mu_{0p2}-\mu_{0p\cdot})}$ $+\frac{4s_{1k}m_{p2}}{(1+\alpha^{**2})(\alpha^{**}\mu_{0p2}-\mu_{0p\cdot})}$
GEE2(G)	v_{const}	$\frac{1}{3-\alpha^{**}}\{\mu_{0p1}+\mu_{0p2}(1-\alpha^{**})+\mu_{0p3}\}$
GEE2(G)	v_{id}	$\frac{\alpha_{01}\sqrt{\mu_{0p1}\mu_{0p2}}(\alpha^{**2}+1)+\alpha_{03}\sqrt{\mu_{0p2}\mu_{0p3}}(\alpha^{**2}+1)-\alpha^{**}(\mu_{0p\cdot}+\mu_{0p2})}{2\alpha^{**}(\alpha^{**}-1)(\alpha^{**}+1)}$
GEE2(G)	v_{sq}	$\frac{(-\alpha^{**4}-\alpha^{**3}+\alpha^{**2}+3\alpha^{**})\mu_{0p2}^2+2\alpha^{**}d-(\alpha^{**3}-\alpha^{**2}-\alpha^{**}-3)(s_{2k}m_{p1}+s_{1k}m_{p3})}{\alpha^{**}(1-\alpha^{**2})\{\mu_{0p1}+\mu_{0p2}(1-\alpha^{**})+\mu_{0p3}\}}$

5. OCCASION-SPECIFIC MEANS, CONSTANT OVER SUBJECTS

Another simple case occurs when observations are grouped according to occasion. Suppose we have $T_i = 4$ observations per subject and design matrix $\mathbf{X} = \mathbf{1}_K \otimes \mathbf{I}_2$ grouping the first two observations on each subject together with mean $g^{-1}(\beta_1) = \lambda_1$ and the third and fourth observations on each subject together with common mean $g^{-1}(\beta^*) = \lambda_2$. In this case the expressions for $\boldsymbol{\lambda}^*$ implied by (7) do not depend on the choice of correlation structure or variance function. Using GEE1 or EGEE, λ_1^* is simply the average of the true means at occasions 1 and 2 and λ_2^* is simply the average of the true means at occasions 3 and 4. That is,

$$\begin{aligned} \lambda_1^* &= \frac{1}{2K}\sum_{i=1}^{K}(\mu_{0i1} + \mu_{0i2}) \\ \lambda_2^* &= \frac{1}{2K}\sum_{i=1}^{K}(\mu_{0i3} + \mu_{0i4}), \end{aligned} \tag{12}$$

regardless of the correlation structure and variance function.

For this example, the GEE2 estimands implied by (8) also do not depend on the choice of correlation structure, and they do not depend on the working third and fouth order moment structure. Using GEE2 with constant variance function v_2, λ_p^{**} is as in (12). For this result, we have not assumed that the correlation structure has been correctly specified. That is, using v_2, λ_p^{**} is a simple average as given by (12) under the assumption that the true correlation matrix has the following unstructured form

$$\text{corr}_0(\mathbf{y}_i) = \begin{pmatrix} 1 & \alpha_{01} & \alpha_{02} & \alpha_{03} \\ \alpha_{01} & 1 & \alpha_{04} & \alpha_{05} \\ \alpha_{02} & \alpha_{04} & 1 & \alpha_{06} \\ \alpha_{03} & \alpha_{05} & \alpha_{06} & 1 \end{pmatrix}.$$

Such an assumption in this example, however, leads to a complicated expression for equation (8) when either v_1 or v_3 is used. Hence, we were unable to obtain closed form expressions for the GEE2 estimands when either the identity or square variance functions were used without making the simplifying assumption that the true correlation structure corresponds to independent within-cluster responses. Clearly, it is unlikely that the true correlation matrix would be the identity matrix in most situations for which GEE-based methods would be used. However, the results under this simplifying assumption are of interest in the sense that if the GEE2 estimand is not an intuitively appealing combination of the true means in this case, then we can only expect it to be less attractive as nonzero correlations are introduced into the expression for the estimand.

Assuming the true correlation structure is the independence structure, v_1 leads to λ_p^{**} as in (12). When the square variance function v_3 is used, however, the GEE2 estimand λ_p^{**} is the solution to the following quadratic equation

$$4K\lambda_p^{**2} - \lambda_p^{**}\sum_{i=1}^{K}(\mu_{0i,2p-1} + \mu_{0i,2p}) - \sum_{i=1}^{K}(\mu_{0i,2p-1}^2 + \mu_{0i,2p}^2) = 0, \quad p = 1, 2.$$

6. DISCUSSION

From the evidence presented here it appears that when observations are grouped according to measurement occasion, GEE1 and EGEE-based estimators of group means estimate the average of the true means of the observations. When observations are grouped according to some time-invariant criterion, a weighted average of the true observation means is estimated. In general, the weights in this average depend upon the correlation among observations. In the exchangeable case these weights are equal regardless of $\boldsymbol{\alpha}$. In other cases small correlations result in nearly equal weights. When observations are substantially correlated weights are unequal. As an estimand, a weighted average of the true means appears to make some sense and the results presented here should be reassuring to users of GEE1 and EGEE.

In the examples considered here, the quantity that is estimated by solving a GEE2 equation is often not as appealing. The GEE2 estimand depends upon the choice of variance function, and particularly when GEE2 is used with a variance function other than $v(\mu) = 1$, the estimand can be difficult to justify.

ACKNOWLEDGEMENTS

This work was supported by NIMH grant MH15168-18. The author wishes to thank Tom Severini and Ishwar Basawa for discussions which led to improvements in the paper.

BIBLIOGRAPHY

Barndorff-Nielsen, O.E. and Cox, D.R. (1989) *Asymptotic Techniques for use in Statistics*, London: Chapman and Hall.

Carroll, R.J. and Ruppert, D. (1982) Robust estimation in heteroskedastic linear models. *Ann. Statist.*, **10**, 429–441.

Davidian, M. and Carroll, R. J. (1988) A note on extended quasi-likelihood. *J. R. Statist. Soc.* B, **50**, 74–82.

Desmond, A.F. (1997) Optimal estimating functions, quasi-likelihood and statistical modelling (with discussion). *J. Statist. Planng Inf.*, **60**, 77–121.

Godambe, V. P. (1960) An optimum property of regular maximum likelihood estimation. *Ann. Math. Statist.*, **31**, 1208–1212.

Godambe, V. P. and Thompson, M.E. (1989) An extension of quasi-likelihood (with discussion). *J. Statist. Planng Inf.*, **22**, 137–172.

Gourieroux, C., Monfort, A., and Trognon, A. (1984) Pseudomaximum likelihood methods: theory. *Econometrika*, **52**, 681–700.

Hall, D.B. and Severini, T.A. (1998) Extended generalized estimating equations for clustered data. *J. Am. Statist. Ass.*, (to appear).

Liang, K.Y. and Rathouz, P.J. (1997) Discussion on A.F. Desmond's paper: Optimal estimating functions, quasi-likelihood and statistical modelling. *J. Statist. Planng Inf.*, **60**, 113–116.

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

Liang, K. Y., Zeger, S. L., and Qaqish, B. (1992) Multivariate regression analyses for categorical data (with discussion). *J. R. Statist. Soc.* B, **45**, 3–40.

McCullagh, P. (1991) Quasi-likelihood and estimating functions. In: *Statistical Theory and Modelling* (ed. D.V. Hinkley, N. Reid and E.J. Snell), pp. 265–286. London: Chapman and Hall.

Nelder, J.A. and Pregibon, D. (1987) An extended quasi-likelihood function. *Biometrika*, **74**, 221–232.

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.

Received July, 1998.