Small Sample GEE Estimation of regression parameters for Longitudinal Data

Sudhir Paul* Xuemao Zhang†

Abstract
Longitudinal or clustered response data arise in many applications such as biostatistics, epidemiology, and environmental studies. The repeated responses cannot, in general, be assumed to be independent. Generalized estimating equations (GEE) is a widely used method to estimate marginal regression parameters for correlated responses. The advantage of the GEE is that the estimates of the regression parameters are asymptotically unbiased even if the correlation structure is misspecified, although their small sample properties are not known. In this paper, two bias adjusted GEE estimators of the regression parameters in longitudinal data are investigated when the number of subjects is small. One is based on bias correction, and the other is based on a bias reduction. Simulations show that both methods do well in reducing bias and have, in general, higher efficiency than the GEE estimates. Analysis of data involving a small number of subjects also show improvement in bias, MSE, and standard errors of the estimates by the two bias adjusted methods over the GEE estimates. However, the bias corrected estimate is preferable over the bias reduced estimate as the former is computationally simpler.

Key Words: Bias correction, Bias reduction, Generalized estimating equations, Longitudinal data, Marginal model

1. Introduction

Longitudinal data arise in many epidemiological and bio-statistical practices in which a number of repeated count/binary responses are observed on a number of individuals (subjects). Longitudinal studies are characterized by repeated measures over a period of time from each individual. Usually the subjects are assumed to be independent while the repeated measurements taken on each subject are correlated. The complication of longitudinal data analysis is partly due to the lack of a rich class of models such as the multivariate gaussian for the joint distribution of the correlated responses (Liang and Zeger, 1986 and Zeger and Liang, 1986). Liang and Zeger (1986) introduced the generalized estimating equations (GEE) approach for analyzing longitudinal data in which a working correlation matrix for the responses of each individual is used. The GEE approach requires specification of only the first two moments of a subject's responses rather than the full specification of the joint distribution. The advantage of the GEE is that the estimates of the regression parameters are asymptotically unbiased even if the correlation structure is misspecified. However, the small sample properties of the GEE estimates are not known (Sharples and Breslow, 1992).

The purpose of this paper is to obtain estimates of the regression parameters which will have better bias and efficiency properties in comparison to the GEE estimates when the number of subjects is small. The maximum likelihood (ML) estimators are, in general, consistent. However, they are not in general unbiased. Cox and Snell (1968) provided general results for the first-order correction of bias of the ML estimators of parameters under any distribution. Cordeiro and Klein (1994) further simplified the results of Cox and Snell (1968). Firth (1993) showed that the order $1/n$ bias of the ML estimator can be removed by introducing an appropriate bias term into the likelihood score function. The

*University of Windsor, Windsor, ON N9B 3P4 Canada
†Worcester Polytechnic Institute, Worcester, MA 01609 US

bias correction method of Cox and Snell (1968) is corrective in the sense that it adds (or substracts) a correction to the ordinary ML estimators and that of Firth (1993) is preventive in the sense that it solves a modified likelihood score equation. In this paper, we obtain bias corrected estimates (Cox and Snell, 1968) and bias reduced estimates (Firth, 1993) of the regression parameters for longitudinal data by treating the generalized estimating functions as if they were likelihood score functions.

The bias corrected and the bias-reduced estimates are derived in Section 2. Applications to longitudinal binary and Poisson data are shown in Section 3. A simulation study is conducted in Section 4. Three examples are given in Section 5 and a discussion follows in Section 6.

2. Estimates of the Regression Parameters Based on Bias-correction and Bias-reduction for Longitudinal Data

Let $\mathbf{y}_n = (y_{n1}, \ldots, y_{nd})^T$ be the vector of responses with a $d \times p$ design matrix $\mathbf{X}_n = (\mathbf{x}_{n1}, \ldots, \mathbf{x}_{nd})^T$ for the nth subject, $n = 1, \ldots, N$. Assume that the N subjects are independent while the repeated measurements y_{nj} taken on each subject are correlated, $j = 1, \ldots, d$. Define $\boldsymbol{\mu}_n = \mathrm{E}(\mathbf{y}_n|\mathbf{X}_n) = (\mu_{n1}, \ldots, \mu_{nd})^T$ to be the expectation of $\mathbf{y}_n$ conditional on $\mathbf{X}_n$ and suppose $\boldsymbol{\mu}_n = F(\mathbf{X}_n\boldsymbol{\beta})$, where $\boldsymbol{\beta}$ is a $p \times 1$ vector of regression parameters of interest and F^{-1} is the link function. Assume that the variance of y_{nj} is given by $\phi v(\mu_{nj})$, where v is the variance function and ϕ is the overdispersion parameter. Note that, for binary data, F is the standard normal cumulative distribution for probit link and a standard logistic cumulative distribution for logit link. For Poisson data with log link, F is the exponential function.

As is well known, the GEE method of Liang and Zeger (1986) for repeated measures uses a common working correlation matrix for the longitudinal responses of each subject. Using standard notations, let $R(\boldsymbol{\alpha})$ be a working correlation matrix completely specified by the parameter vector $\boldsymbol{\alpha}$. Then, $\phi\boldsymbol{W}_n = \phi\mathbf{A}_n^{1/2}R(\boldsymbol{\alpha})\mathbf{A}_n^{1/2}$ is the corresponding working covariance matrix, where $\mathbf{A}_n(\boldsymbol{\beta}) = \mathrm{diag}\{v(\mu_{nj})\}$, $j = 1, \ldots, d$, $n = 1, \ldots, N$. For given consistent estimates of ϕ and $\boldsymbol{\alpha}$, the GEE estimate of $\boldsymbol{\beta}$, denoted by $\hat{\boldsymbol{\beta}}$, is obtained by solving the generalized estimating equations

$$\sum_{n=1}^{N}(\mathbf{y}_n - \boldsymbol{\mu}_n)^T\mathbf{W}_n^{-1}\mathbf{D}_n = 0, \tag{1}$$

where $\mathbf{D}_n = \frac{\partial\boldsymbol{\mu}_n}{\partial\boldsymbol{\beta}} = \boldsymbol{\Delta}_n\mathbf{X}_n$, $\boldsymbol{\Delta}_n = \mathrm{diag}(f(\mathbf{x}_{n1}^T\boldsymbol{\beta}), \ldots, f(\mathbf{x}_{nd}^T\boldsymbol{\beta}))$ with $f = F'$, $n = 1, \cdots, N$.

The left hand side of equation (1), which can be written as

$$U(\boldsymbol{\beta}; \boldsymbol{\alpha}, \phi) = \sum_{n=1}^{N}(\mathbf{y}_n - \boldsymbol{\mu}_n)^T\mathbf{W}_n^{-1}\frac{\partial\boldsymbol{\mu}_n}{\partial\boldsymbol{\beta}}, \tag{2}$$

is the generalized estimating function for $\boldsymbol{\beta}$ given $\boldsymbol{\alpha}$ and ϕ. Let $U(\boldsymbol{\beta}; \boldsymbol{\alpha}, \phi) = (U_1, U_2, \ldots, U_p)$. For obtaining bias-corrected (Cox and Snell, 1968) and bias-reduced (Firth, 1993) GEE estimates, we treat U_i as if it were a likelihood score function for β_i, $i = 1, \ldots, p$.

Now, define $\kappa_{ij} = \mathrm{E}(\partial U_i/\partial\beta_j)$ for $i, j = 1, \ldots, p$. Further, define $\kappa_{ijl} = \mathrm{E}(\partial^2 U_i/\partial\beta_j\partial\beta_l)$ and $\kappa_{ij}^{(l)} = \partial\kappa_{ij}/\partial\beta_l$ for $i, j, l = 1, \ldots, p$. Derivation of the quantities κ_{ij}, $\kappa_{ij}^{(l)}$, and κ_{ijl} are given in the Appendix A. Then, the Fisher information matrix analogue of order p for $\boldsymbol{\beta}$ is $\mathbf{I} = \{-\kappa_{ij}\}$. Now, let $\mathbf{I}^{-1} = \{\kappa^{ij}\}$ be the inverse of $\mathbf{I}$. Then, following Cordeiro and

Klein (1994) the bias of $\hat{\beta}_s$ can be expressed as

$$b_s(\boldsymbol{\beta}) = \sum_{i=1}^{p} \kappa^{si} \sum_{j,l=1}^{p} \Big[\kappa_{ij}^{(l)} - \frac{1}{2}\kappa_{ijl}\Big]\kappa^{jl}, \quad s = 1, \ldots, p \tag{3}$$

and the bias corrected estimate $\tilde{\beta}_s$ of β_s is given by $\tilde{\beta}_s = \hat{\beta}_s - b(\hat{\beta}_s)$. The estimates $\tilde{\beta}_s$ will also be referred to as GEEBc estimates.

Following the "preventive" method of Firth (1993), by introducing a bias term into the score function $U(\boldsymbol{\beta}; \boldsymbol{\alpha}, \phi)$, the modified score function is

$$U^*(\boldsymbol{\beta}; \boldsymbol{\alpha}, \phi) = U(\boldsymbol{\beta}; \boldsymbol{\alpha}, \phi) - \mathbf{I}\mathbf{b}(\boldsymbol{\beta}),$$

where $b(\boldsymbol{\beta}) = (b_1(\boldsymbol{\beta}), \ldots, b_p(\boldsymbol{\beta}))^{\mathrm{T}}$.

The bias reduced GEE estimate, denoted by $\boldsymbol{\beta}^*$, of $\boldsymbol{\beta}$ using the method of Firth (1993) is obtained by solving the modified score equation

$$U^*(\boldsymbol{\beta}; \boldsymbol{\alpha}, \phi) = 0. \tag{4}$$

An iterative method of solving the above equation is described in Appendix B. The estimates $\boldsymbol{\beta}^*$ will also be referred to as BcGEE estimates.

3. Calculation of Bias in Equation (3) for Binary and Count data

In this section, we discuss the application of the formula given in equation (3) for the calculation of bias of the estimates of the regression parameters for binary and count data.

3.1 Binary data

For the vector of binary responses $\mathbf{y}_i$, the variance function is given by $v(\boldsymbol{\mu}) = \boldsymbol{\mu}(1 - \boldsymbol{\mu})$. We consider the probit and logit link functions. For the probit link, $F = \Phi$ is the cumulative distribution function of the standard normal distribution. Thus, $f = F' = \phi$ is the density function of the standard normal distribution. For the logit link, $F(x) = \frac{\exp(x)}{1+\exp(x)}$ is the standard logistic cumulative distribution function and $f(x) = F'(x) = \frac{\exp(x)}{(1+\exp(x))^2}$.

Let I_d and I_p be a d-dimensional and p-dimensional identity matrix, respectively, and let K_{dp} be a $dp \times dp$ commutation matrix. Then, the quantities required for the calculation of the bias $b_s(\boldsymbol{\beta})$ are given by

$$\mathbf{I} = \{-\kappa_{ij}\} = \sum_{n=1}^{N} (\boldsymbol{\Delta}_n \mathbf{X}_n)^T \mathbf{W}_n^{-1} \boldsymbol{\Delta}_n \mathbf{X}_n, \tag{5}$$

$$\begin{aligned}\Big(\{\kappa_{ij}^{(1)}\}, \{\kappa_{ij}^{(2)}\}, \cdots, \{\kappa_{ij}^{(p)}\}\Big)^T = -\sum_{n=1}^{N} (\mathbf{X}_n^T \otimes \mathbf{X}_n^T) \big[(\boldsymbol{\Delta}_n \mathbf{W}_n^{-1}) \otimes I_d \\ + I_d \otimes (\boldsymbol{\Delta}_n \mathbf{W}_n^{-1})\big] \frac{\partial \boldsymbol{\Delta}_n}{\partial \boldsymbol{\mu}_n}\end{aligned} \tag{6}$$

and

$$\begin{aligned}&(\{\kappa_{ij1}\}, \{\kappa_{ij2}\}, \cdots, \{\kappa_{ijp}\})^T \\ &= -\sum_{n=1}^{N} \bigg\{ (\mathbf{X}_n^T \otimes \mathbf{X}_n^T) \Big[(\boldsymbol{\Delta}_n \mathbf{W}_n^{-1}) \otimes I_d + I_d \otimes (\boldsymbol{\Delta}_n \mathbf{W}_n^{-1})\Big] \frac{\partial \boldsymbol{\Delta}_n}{\partial \boldsymbol{\mu}_n} \\ &+ \Big[(\frac{\partial \boldsymbol{\Delta}_n}{\partial \boldsymbol{\mu}_n} \boldsymbol{\Delta}_n \mathbf{X}_n)^T \otimes I_p\Big] (I_d \otimes K_{dp})(\mathrm{vec}(\mathbf{X}_n^T) \otimes I_d) \cdot \mathbf{W}_n^{-1} \bigg\} \boldsymbol{\Delta}_n \mathbf{X}_n,\end{aligned} \tag{7}$$

where $\mathbf{\Delta}_n = \text{diag}\{\phi(\mathbf{x}_{n1}^T\boldsymbol{\beta}), \ldots, \phi(\mathbf{x}_{nd}^T\boldsymbol{\beta})\}$ and $\frac{\partial \mathbf{\Delta}_n}{\partial \boldsymbol{\mu}_n}$ is a $d^2 \times d$ dimensional sparse matrix with non-zero quantities $\phi^{'}(\Phi^{-1}(\mu_{nj}))(\Phi^{-1})^{'}(\mu_{nj})$ in the $[(j-1)d+j, j]$ term for the probit link, and $\mathbf{\Delta}_n = \text{diag}\{\frac{\exp(\mathbf{x}_{n1}^T\boldsymbol{\beta})}{(1+\exp(\mathbf{x}_{n1}^T\boldsymbol{\beta}))^2}, \ldots, \frac{\exp(\mathbf{x}_{nd}^T\boldsymbol{\beta})}{(1+\exp(\mathbf{x}_{nd}^T\boldsymbol{\beta}))^2}\}$ and $\frac{\partial \mathbf{\Delta}_n}{\partial \boldsymbol{\mu}_n}$ is a $d^2 \times d$ dimensional sparse matrix with non-zero quantities $(1-2\mu_{nj})$ in the $[(j-1)d+j, j]$ term for the logit link, $j = 1, \ldots, d$, $n = 1, \ldots, N$.

3.2 Count data

For the vector of Poisson responses $\mathbf{y}_n$, $n = 1, \ldots, N$, the variance function is given by $v(\boldsymbol{\mu}) = \boldsymbol{\mu}$. Then, the expressions of the quantities required for the bias $b_s(\boldsymbol{\beta})$ in equation (3) for count data are the same as (5), (6), and (7) with the exceptions of $\mathbf{\Delta}_n$ and $\frac{\partial \mathbf{\Delta}_n}{\partial \boldsymbol{\mu}_n}$ which are different. If we use a log link function, we have $f(x) = F^{'}(x) = F(x) = \exp(x)$, $\mathbf{\Delta}_n = \text{diag}\{\exp(\mathbf{x}_{n1}^T\boldsymbol{\beta}), \ldots, \exp(\mathbf{x}_{nd}^T\boldsymbol{\beta})\}$, and $\frac{\partial \mathbf{\Delta}_n}{\partial \boldsymbol{\mu}_n}$ is a $d^2 \times d$ dimensional sparse matrix with 1 in the $[(j-1)d+j, j]$ term, $j = 1, \ldots, d$, $n = 1, \ldots, N$.

4. Simulations

4.1 Binary data

We first study properties of the regression parameters in a marginal model for correlated binary data with $p = 3$, x_{i1j} and x_{i2j}, $j = 1, \ldots, d$, generated as uniform random variables. That is, the design matrix for subject i is of the form

$$\mathbf{X}_i = \begin{pmatrix} 1 & \cdots & 1 \\ x_{i11} & \cdots & x_{i1d} \\ x_{i21} & \cdots & x_{i2d} \end{pmatrix}^T .$$

A factor that could affect the magnitude of the bias of the GEE estimator is whether the covariate is a between or a within cluster(subject) variable. To see this, we generated a within cluster(subject) covariate $\mathbf{x}_{i1}$ with x_{i1j} uniform random variables in the interval $[-2, 2]$ and a cluster(subject) level covariate $\mathbf{x}_{i2}$ with x_{i2j} identical uniform random variables in the interval $[-2, 2]$, $j = 1, \ldots, d$, $i = 1, \ldots, N$.

We consider $\boldsymbol{\beta} = (\beta_0, \beta_1, \beta_2)$ with $\beta_0 = 0.0$, $\beta_1 = 0.5$, $\beta_2 = 1.0$, and a logit link function. Given a correlation matrix $\Omega(\gamma)$, where γ is the correlation parameter, the correlated binary responses $\mathbf{y}_i = (y_{i1}, \ldots, y_{id})^T$ are generated using the method by Qaqish (2003) such that $\mathbf{y}_i$ has the marginal means $\mu_{ij} = P(y_{ij} = 1|\mathbf{X}_i) = exp(\beta_0 + \beta_1 x_{i1j} + \beta_2 x_{i2j})/\left(1 + exp(\beta_0 + \beta_1 x_{i1j} + \beta_2 x_{i2j})\right)$ and correlation matrix $\Omega(\gamma)$, $j = 1, \ldots, d$.

Simulations were conducted with an exchangeable correlation structure $\Omega(\gamma)$ with values of $\gamma = 0(0.1)$, …, 0.8 and with the AR(1) correlation structure $\Omega(\gamma)$ with values of $\gamma = -0.8(0.1), \ldots, 0.8$. The largest correlation strength 0.8 or -0.8 is chosen to avoid data generation difficulties. The numbers of subjects taken are $N = 15, 20, 30$, and 50, each subject having $d = 4$ observations. For each N, we simulated 5000 samples.

We calculated bias of the GEE, GEEBc, and BcGEE estimates $\hat{\boldsymbol{\beta}}$, $\tilde{\boldsymbol{\beta}}$, and $\boldsymbol{\beta}^*$ and efficiency of the estimates $\tilde{\boldsymbol{\beta}}$ and $\boldsymbol{\beta}^*$. Bias, for example, of $\hat{\beta}_1$ was calculated as bias($\hat{\beta}_1$) $=\sum_{i=1}^{5000}(\hat{\beta}_{1i} - \beta_1)/5000$ and efficiency, for example, of $\tilde{\beta}_1$ was calculated as RE($\tilde{\beta}_1$) $=$ MSE($\hat{\beta}_1$)/MSE($\tilde{\beta}_1$), where, for example, MSE($\hat{\beta}_1$) $= \sum_{i=1}^{5000}(\hat{\beta}_{1i} - \beta_1)^2/5000$.

We first discuss the bias results. Bias properties of the estimates of β_1 and β_2 are very similar for all three methods and for both the correlation structures. So, we present bias results for the estimates of β_1 and β_2 for the exchangeable correlation structure only. The results for bias of the estimates of β_1 are summarized in Figure 1, and those for β_2

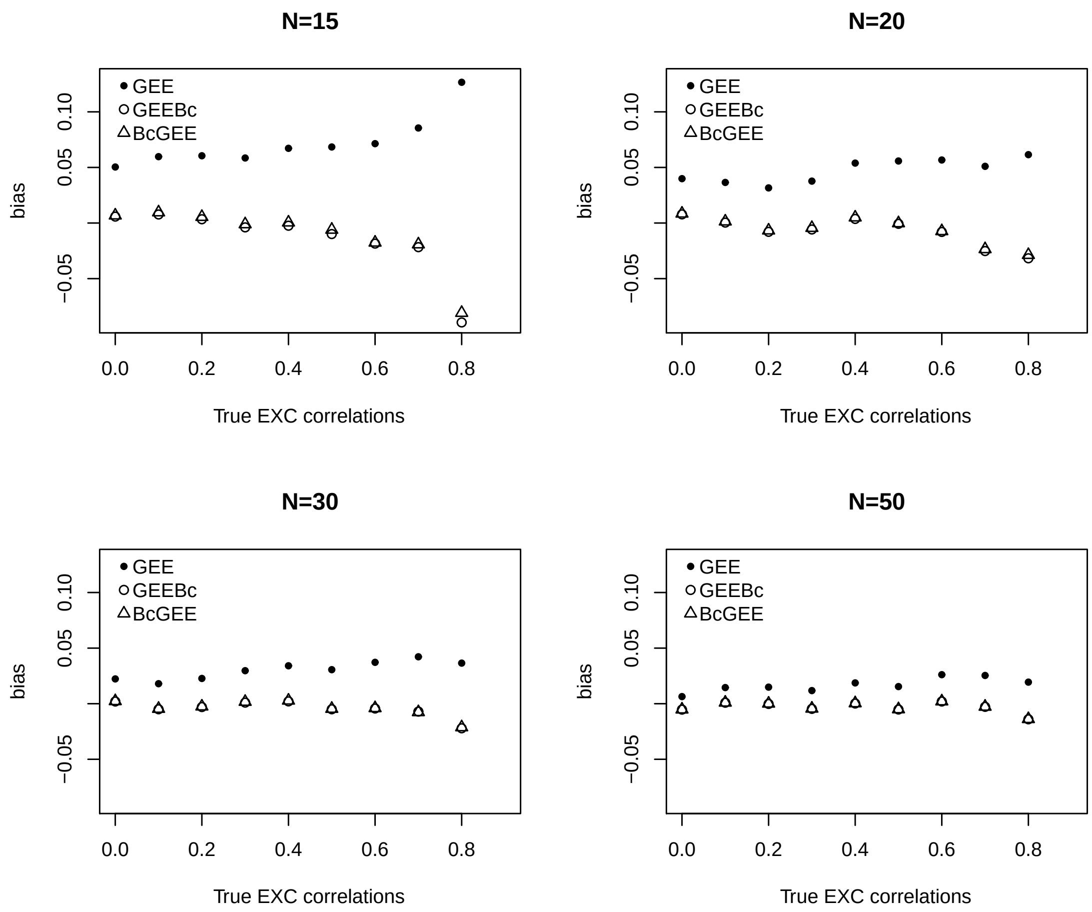

Figure 1: Biases of GEE, GEEBc, and BcGEE estimates of β_1 with true exchangeable correlations.

are summarized in Figure 2. By comparing the two figures, we see that the bias of the GEE estimates of the between cluster regression parameter β_2 is much larger than that of the GEE estimates of the within cluster regression parameter β_1, whereas the estimates by the two bias corrected methods seem to remain little affected. Furthermore, from the figures we see that GEE estimates have the largest biases for all values of the correlation considered. However, the biases of the estimates of GEEBc and BcGEE seem to be similar. The difference between the biases of the GEE estimates of β_1 and β_2 and those of GEEBc and BcGEE estimates diminish as the number of subjects increases.

We now compare efficiency of the GEEBc and BcGEE estimates $\tilde{\boldsymbol{\beta}}$ and $\boldsymbol{\beta}^*$ in relation to the GEE estimate $\hat{\boldsymbol{\beta}}$. From the simulations, again, the MSE of the GEE estimates of the between cluster regression parameter is much larger than that of the within cluster regression parameter. The comparative efficiency results of the GEEBc and BcGEE estimates are similar for both the parameters β_1 and β_2, so we present results for the estimates of β_1 only. These are summarized in Figure 3 for the exchangeable correlation structure.

From the figure, we see that the efficiencies of $\tilde{\beta}_1$ and β_1^* are very similar. In general, both estimates are more efficient than the GEE estimates $\hat{\beta}_1$ for a small number of subjects ($N = 15$, $N = 20$). As the number of subjects increases ($N = 50$), relative efficiencies of

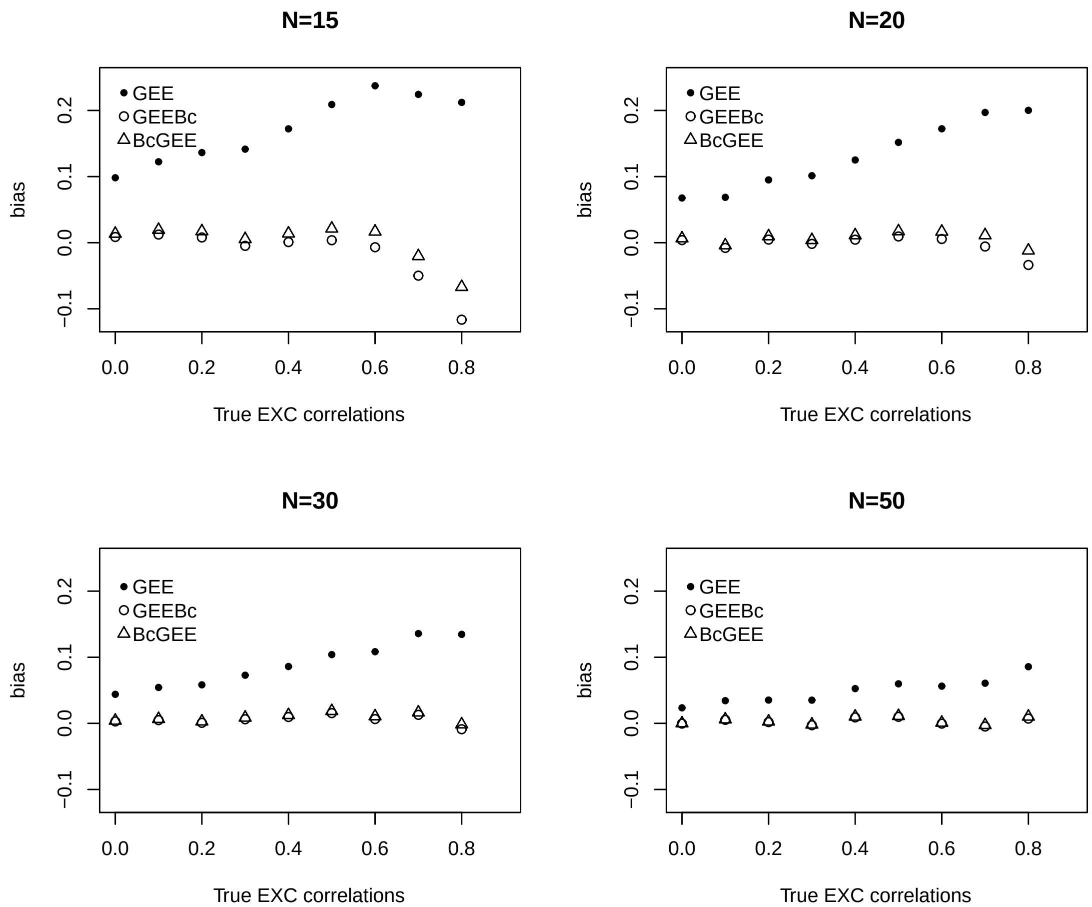

Figure 2: Biases of GEE, GEEBc, and BcGEE estimates of β_2 with true exchangeable correlations.

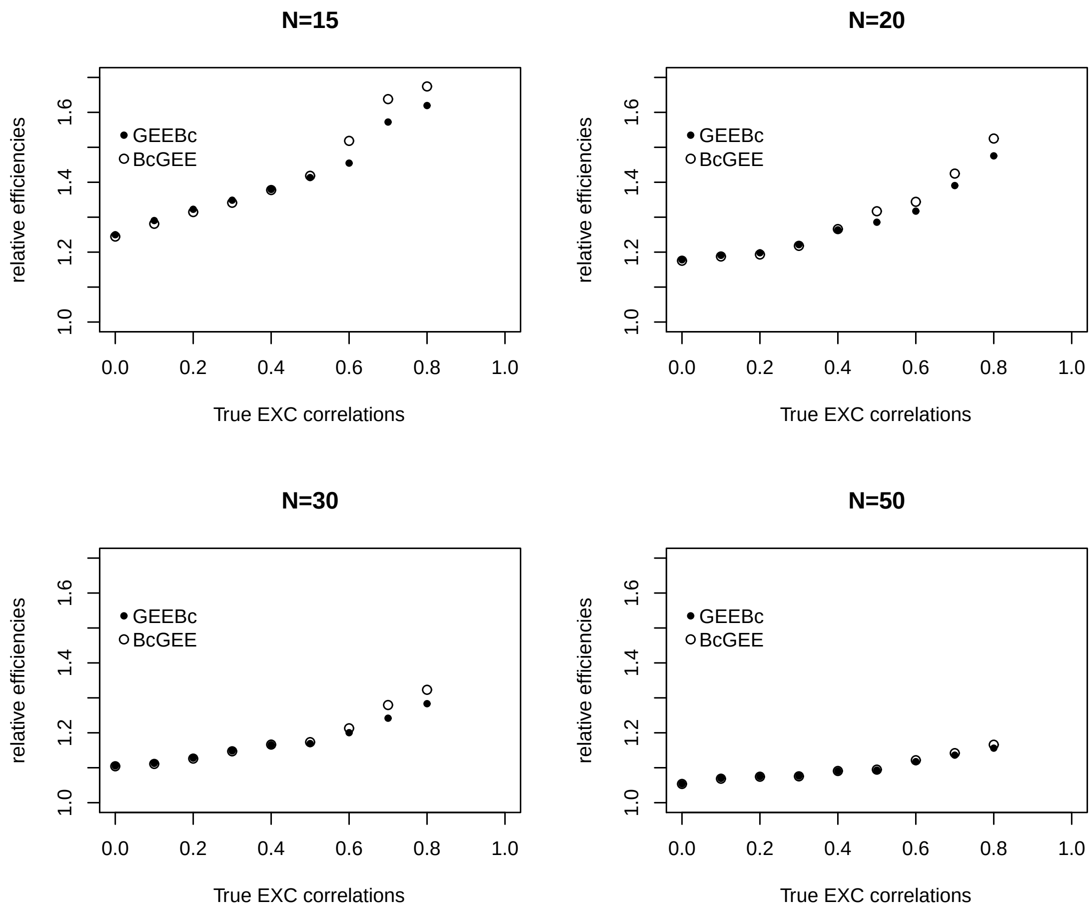

Figure 3: Relative efficiency of GEEBc and BcGEE estimates of β_1 with true exchangeable correlations

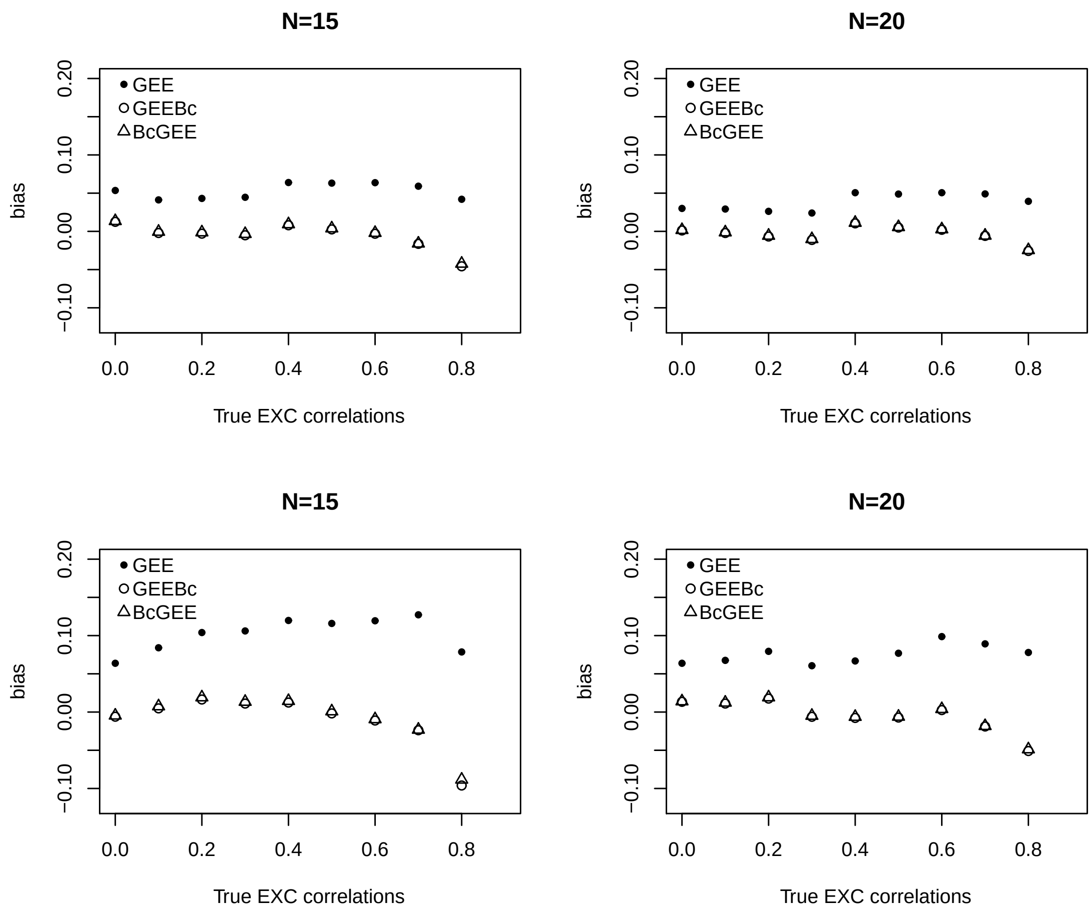

Figure 4: Biases of estimates of β_1 (top panel) and β_2 (bottom panel) by GEE, GEEBc, and BcGEE estimation

$\tilde{\beta}_1$ and β_1^* become closer to 1. This indicates the benefit of the bias correction procedure for small clusters situations. Results for the AR(1) correlation structure, which are not presented in the paper, are similar to those for the exchangeable correlation structure, except in some cases in which $\tilde{\boldsymbol{\beta}}$ is slightly more efficient.

Another concern is if the covariate type affects the bias of the estimates of the regression parameters. To see this, we extended the simulation study. We generated continuous covariate $\mathbf{x}_{i1}$ with x_{i1j} uniform random variables in the interval $[-2, 2]$ and discrete covariate $\mathbf{x}_{i2}$ with x_{i2j} binary random variables with success probabilities varying between 0.1 and 0.9, $j = 1, \ldots, d$, $i = 1, \ldots, N$. Bias results of the simulation are shown in Figure 4. The top panel shows bias of $\hat{\beta}_1$, $\tilde{\beta}_1$, and β_1^* and the bottom panel shows bias of $\hat{\beta}_2$, $\tilde{\beta}_2$, and β_2^*. It can be seen that the bias of the GEE estimates of the regression parameter for the binary covariates is a little larger than that of the regression parameter for the continuous covariates. Whereas, these estimates by the two bias corrected methods seem to be little affected.

Note that all the bias and efficiency results presented here are based on the logit link function to be consistent with the logit link used in the analysis of the data in Section 5. However, simulations using the probit link function were also conducted. The results are not given here as the comparative performance of the estimates by all three methods were similar.

4.2 Poisson data

We now study the performance of GEE, GEEBc, and BcGEE estimates $\hat{\beta}_1$, $\tilde{\beta}_1$, and β_1^* for longitudinal Poisson data. For the generation of longitudinal Poisson data, we consider N subjects, each with $d = 4$ repeated responses such that $\mu_{ij} = \exp(\beta_0 + \beta_1 x_{ij})$ and $\mathrm{var}(y_{ij}) = \mu_{ij}$, where $\beta_0 = 0$, $\beta_1 = 0.5$, x_{ij} is generated from a uniform distribution on $[j-1, j]$, $j = 1, \ldots, 4$, $i = 1, \ldots, N$. Exchangeable and AR(1) Poisson data y_{ij} are generated using the method of Yahav and Shmueli (2012).

For large N, bias and efficiency properties of GEE, GEEBc, and BcGEE estimates $\hat{\beta}_1$, $\tilde{\beta}_1$ and β_1^* are very similar. Also, for the data generated from the AR(1) correlation structure, these properties of the three estimates do not differ much, even for a small N. As N increases, in general, biases of all three estimates become closer to 0. Also, efficiencies of $\tilde{\beta}_1$ and β_1^* in relation to $\hat{\beta}_1$ become closer to 1. That is, all three estimates become equally efficient. These results are not presented here. For a small N and when data are generated from the exchangeable correlation structure, the GEE estimates $\hat{\beta}_1$ show some bias, whereas the bias of the other two becomes closer to 0 (see Figure 5). In this case, the efficiencies of $\tilde{\beta}_1$ and β_1^* in relation to $\hat{\beta}_1$ are almost identical and increase as the true exchangeable correlation increases. It then shows that both these estimates are more efficient than the GEE estimates.

5. Examples

Example 1: We consider data of a clinical trial on cerebrovascular deficiency with a crossover design. The data set is from Diggle, Liang, and Zeger (1994). The purpose of this crossover trial was to compare an active drug (A) and a placebo (B). A total of 67 patients were enrolled into the clinical trial of which 34 patients received the active drug (A) followed by the placebo (B), and another 33 patients were treated in the reverse order. The response variable is defined as 0 for an abnormal and 1 for a normal electrocardiogram reading. Conceptually, the 2×2 crossover trial can be viewed as a longitudinal study with 2 observations for each patient. The two major covariates, period (x_{i1}) and treatment (x_{i2}), are both time-dependent. They are coded as

$$x_{i1} = \begin{cases} 1, & \text{period 2} \\ 0, & \text{period 1,} \end{cases}$$

and

$$x_{i2} = \begin{cases} 1, & \text{active drug (A)} \\ 0, & \text{placebo (B),} \end{cases}$$

respectively. The data is summarized in Table 1. An analysis of a full regression model by

Table 1: Data from a crossover trial on cerebrovascular deficiency (Diggle, Liang, and Zeger,1994).

	Responses					Period	
Group	(1,1)	(0,1)	(1,0)	(0,0)	Total	1	2
AB	22	0	6	6	34	28	22
BA	18	4	2	9	33	20	22

Diggle, Liang, and Zeger (1994) shows little support for a treatment-by-period interaction. Therefore, we consider the logit regression model

$$\text{logit}\,\Pr(Y_{ij} = 1) = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2}. \tag{8}$$

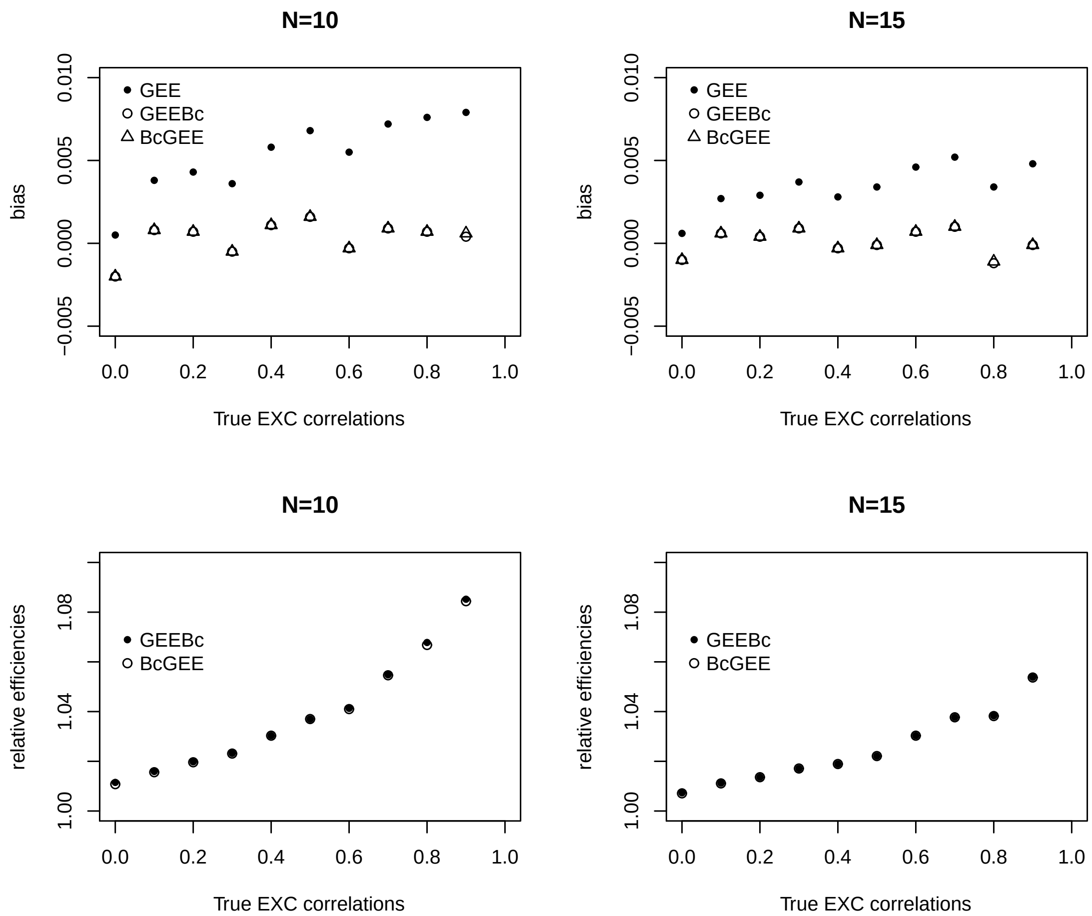

Figure 5: Biases and relative efficiencies of GEE, GEEBc, and BcGEE estimate of β_1 for exchangeable Poisson data

The GEE, GEEBc, and BcGEE estimates of the regression parameters β_0, β_1, and β_2 with their standard errors in parenthesis are given in Table 2.

The bias-correction of the robust covariance estimator was discussed by Kauermann and Carroll (2001), and Mancl and DeRouen (2001). When the number of subjects is not too small (≥ 10), the bias-correction method by Mancl and DeRouen generally performs better (Preisser *et al.*, 2007). So, the standard errors of the estimates are obtained by using the bias-corrected method by Mancl and DeRouen (2001).

Table 2: GEE, GEEBc, and BcGEE estimates of the regression parameters of Model (8).

	β_0	β_1	β_2
GEE	$0.6659(0.2939)$	$-0.2950(0.2382)$	$0.5689(0.2398)$
GEEBc	$0.6527(0.2924)$	$-0.2883(0.2367)$	$0.5557(0.2380)$
BcGEE	$0.6527(0.2924)$	$-0.2876(0.2367)$	$0.5556(0.2380)$

As in the simulation study, the GEEBc and BcGEE estimates are almost identical. There seems to be some difference between the GEE estimates and the two bias corrected GEE estimates, although the difference is minimal. This is not surprising as the sample size of 67 is quite large.

Example 2: To check what happens in small sample size situations, we investigated many simple random samples of the data in Example 1 of size 15 (7: 8) of which one sample is given in Table 3. For this sample, the GEE, GEEBc, and BcGEE estimates of the regression parameters β_0, β_1, and β_2 with their standard errors in parenthesis are given in Table 4.

Table 3: A subset of the 2×2 crossover trial data from Diggle, Liang, and Zeger (1994).

	Responses					Period	
Group	(1,1)	(0,1)	(1,0)	(0,0)	Total	1	2
AB	2	0	3	2	7	5	2
BA	2	2	1	3	8	3	4

Table 4: GEE, GEEBc, and BcGEE estimates of the regression parameters of Model (8) for the subset in Table 3.

	β_0	β_1	β_2
GEE	$-0.4006(0.6866)$	$-0.6655(0.7311)$	$1.1718(0.7316)$
GEEBc	$-0.3623(0.6776)$	$-0.5956(0.7158)$	$1.0554(0.7161)$
BcGEE	$-0.3656(0.6784)$	$-0.6022(0.7171)$	$1.0660(0.7173)$

As can be seen, again, there is not much difference between the GEEBc and BcGEE estimates. However, significant difference between the GEE estimates and those by the other two methods is observed. A similar conclusion was observed for most of the other samples investigated.

Example 3: We then wanted to see the average performance of the estimates of the regression parameters by the three methods in small samples for data similar to the clinical trials data in Example 1. For this, we conducted a small simulation study. In the simulation, we used the same logistic model as in Example 1 and generated data using the estimates of the regression parameters obtained from the full data of size 67 as the true parameter values. That is, the true values of the regression parameters taken were $\beta_0 = 0.6659$, $\beta_1 = -0.2950$, and $\beta_2 = 0.5689$.

Further, for generating correlated binary responses, we used the method of Qaqish (2003) with 0.3 as the true correlation parameter. This value of the correlation parameter is close to the GEE estimates of 0.327 of the correlation parameter from the full data. Moreover, the number of patients in each group is restricted to between 6 and 9 to balance the two treatment groups. We generated 5000 simple random samples of size 15, and for each sample the parameters were estimated by the three methods. The bias, MSE, and standard errors of the estimates of the parameters β_1 and β_2 by the three methods are given in Table 5.

Table 5: Bias and efficiency of the GEE, GEEBc, and BcGEE estimates of β_0, β_1, and β_2; Data generated using model (8); based on 5,000 replications.

	Bias			MSE			Average Standard Error		
	β_0	β_1	β_2	β_0	β_1	β_2	β_0	β_1	β_2
GEE	0.0525	-0.0313	0.0682	0.4960	0.6159	0.6052	0.7466	0.8046	0.8114
GEEBc	-0.0404	0.0147	-0.0242	0.3570	0.4432	0.4255	0.6984	0.7428	0.7432
BcGEE	-0.0278	0.0100	-0.0126	0.3751	0.4623	0.4466	0.7042	0.7509	0.7521

From this simulation study also we see that there is not much difference between the GEEBc and BcGEE estimates. However, in terms of all three measures bias, MSE, and standard errors of the estimates both the GEEBc and the BcGEE estimates substantially outperform the GEE estimates.

6. Discussion

In this paper, we obtain two bias corrected GEE estimates of the regression parameters in longitudinal data. One of these, the GEEBc estimate $\tilde{\boldsymbol{\beta}}$, is based on correcting the bias of the GEE estimates following Cox and Snell (1968). The other, the BcGEE estimate $\boldsymbol{\beta}^*$, is based on correcting the GEE following a method by Firth (1993). The bias correction method of Cox and Snell (1968) is corrective in the sense that it adjusts the ordinary ML estimators, and that of Firth (1993) is preventive in the sense that it solves a modified likelihood score equation. The performance in terms of bias and efficiency of both of these estimates are very similar, and both show superior performance in terms of bias and efficiency compared to the GEE estimates for small samples. These findings were confirmed by an example that was analyzed, and by subsequent further investigations carried out for small samples.

The bias corrected estimates GEEBc are simpler to obtain than the bias reduced estimates BcGEE estimates. This is Because the former needs the GEE estimates which can be obtained using a standard software. Given these estimates, calculation of the biases of the estimates using the formula (2.3) poses no difficulty as they are non-iterative. So, we recommend the GEEBc estimates.

REFERENCES

Cordeiro, G. M. and Klein, R. (1994), "Bias correction in ARMA models," *Statistics and Probability Letters,* 19, 169–176.

Cox, D. R. and Snell, E. J. (1968), "A general definition of residuals (with discussion)," *Journal of the Royal Statistical Society B,* 30, 248–275.

Crowder, M. (2001), "On repeated measures analysis with misspecified covariance structure," *Journal of the Royal Statistical Society B,* 63, 55–62.

Diggle, P. J., Liang, K. Y. and Zeger, S. L. (1994), *Analysis of Longitudinal Data.* Oxford University Press: Oxford.

Firth, D. (1993), "Bias reduction of maximum likelihood estimates," *Biometrika,* 80, 27–38.

Kauermann, G. and Carroll, R. J. (2001), "A note on the efficiency of sandwich covariance matrix estimation," *Journal of the American Statistical Association,* 96, 1387–1396.

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

Magnus, J. R. and Neudecker, H. (1988), *Matrix Differential Calculus with Applications in Statistics and Econometrics.* John Wiley & Sons: Chichester.

Mancl, L. A. and DeRouen, T. A. (2001), "A covariance estimator for GEE with improved small-sample properties," *Biometrics,* 57, 126–134.

Qaqish, B. F. (2003), "A family of multivariate binary distributions for simulating correlated binary variables with specified marginal means and correlations. *Biometrika,* 92, 455–463.

Sharples, K. and Breslow, N. (1992), "Regression analysis of correlated binary data: some small samples results for the estimating equation approach," *Journal of Statistical Computation and Simulation,* 42, 1–20.

Yahav, I. and Shmueli, G. (2012), "On generating multivariate Poisson data in management science applications," *Applied Stochastic Models in Business and Industry*, 28, 91-102.

Zeger, S. L. and Liang, K.-Y. (1986), "Longitudinal data analysis for discrete and continuous outcomes," *Biometrics,* 42, 121–130.

A. Appendix

A.1 Derivation of κ_{ij}, $\kappa_{ij}^{(l)}$, and κ_{ijl}

As mentioned earlier, we treat the generalized estimating function (2) as if it were a likelihood score function.

By the decoupling method of Crowder (2001) where the working covariance matrix is regarded as a constant matrix with respect to the regression parameters $\boldsymbol{\beta}$, the first derivative by using the chain rule and the product rule in matrix calculus (Magnus and Neudecker, 1988) of $U(\boldsymbol{\beta};\boldsymbol{\alpha},\phi)$ with respect to $\boldsymbol{\beta}$ is

$$\frac{\partial U}{\partial \boldsymbol{\beta}} = \sum_{n=1}^{N}\left[(\mathbf{X}_n^T \otimes \mathbf{y}_n^T \mathbf{W}_n^{-1})\frac{\partial \boldsymbol{\Delta}_n}{\partial \boldsymbol{\mu}_n} - (\boldsymbol{\Delta}_n\mathbf{X}_n)^T\mathbf{W}_n^{-1} - (\mathbf{X}_n^T \otimes \boldsymbol{\mu}_n^T\mathbf{W}_n^{-1})\frac{\partial \boldsymbol{\Delta}_n}{\partial \boldsymbol{\mu}_n}\right]\boldsymbol{\Delta}_n\mathbf{X}_n,$$

where $\frac{\partial \boldsymbol{\Delta}_n}{\partial \boldsymbol{\mu}_n}$ is a $d^2 \times d$ dimensional sparse matrix with non-zero quantities $f^{'}(F^{-1}(\mu_{nj}))(F^{-1})^{'}(\mu_{nj})$ in the $[(j-1)d+j, j]$ term, $j = 1,\dots,d$, $n = 1,\dots,N$.

It is easy to see that

$$\mathbf{I} = \{-\kappa_{ij}\} = -\,\mathrm{E}\left(\frac{\partial U(\boldsymbol{\beta};\boldsymbol{\alpha},\phi)}{\partial \boldsymbol{\beta}}\right) = \sum_{n=1}^{N}(\boldsymbol{\Delta}_n\mathbf{X}_n)^T\mathbf{W}_n^{-1}\boldsymbol{\Delta}_n\mathbf{X}_n \tag{A.1}$$

and

$$\begin{aligned}\left(\{\kappa_{ij}^{(1)}\},\{\kappa_{ij}^{(2)}\},\cdots,\{\kappa_{ij}^{(p)}\}\right)^T &= \frac{\partial}{\partial \boldsymbol{\beta}}\left\{\mathrm{E}\left(\frac{\partial U(\boldsymbol{\beta};\boldsymbol{\alpha},\phi)}{\partial \boldsymbol{\beta}}\right)\right\}\\ = -\sum_{n=1}^{N}(\mathbf{X}_n^T \otimes \mathbf{X}_n^T)&\left[(\boldsymbol{\Delta}_n\mathbf{W}_n^{-1}) \otimes I_d + I_d \otimes (\boldsymbol{\Delta}_n\mathbf{W}_n^{-1})\right]\frac{\partial \boldsymbol{\Delta}_n}{\partial \boldsymbol{\mu}_n},\end{aligned} \tag{A.2}$$

where I_d is a d-dimensional identity matrix. Further, the second derivative of U by using the chain rule, the product rule, and the Kronecker product rule in matrix calculus with respect to $\boldsymbol{\beta}$ is

$$\begin{aligned}
\frac{\partial^2 U}{\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}^T} =& \sum_{n=1}^{N}\left\{\frac{\partial}{\partial\boldsymbol{\mu}_n}\left[(\mathbf{X}_n^T\otimes\mathbf{y}_n^T\mathbf{W}_n^{-1})\frac{\partial\boldsymbol{\Delta}_n}{\partial\boldsymbol{\mu}_n}\boldsymbol{\Delta}_n\mathbf{X}_n\right] - \frac{\partial}{\partial\boldsymbol{\mu}_n}\left[(\boldsymbol{\Delta}_n\mathbf{X}_n)^T\mathbf{W}_n^{-1}\boldsymbol{\Delta}_n\mathbf{X}_n\right]\right.\\
&\left. -\frac{\partial}{\partial\boldsymbol{\mu}_n}\left[(\mathbf{X}_n^T\otimes\boldsymbol{\mu}_n^T\mathbf{W}_n^{-1})\frac{\partial\boldsymbol{\Delta}_n}{\partial\boldsymbol{\mu}_n}\boldsymbol{\Delta}_n\mathbf{X}_n\right]\right\}\boldsymbol{\Delta}_n\mathbf{X}_n\\
=& \sum_{n=1}^{N}\left\{(I_p\otimes\mathbf{X}_n^T\otimes y_n^T\mathbf{W}_n^{-1})\left[(\mathbf{X}_n^T\boldsymbol{\Delta}_n)\otimes I_{d^2}\cdot\frac{\partial^2\boldsymbol{\Delta}_n}{\partial\mu_n^2} + (\mathbf{X}_n^T\otimes\frac{\partial\boldsymbol{\Delta}_n}{\partial\boldsymbol{\mu}_n})\frac{\partial\boldsymbol{\Delta}_n}{\partial\boldsymbol{\mu}_n}\right]\right.\\
&-(\mathbf{X}_n^T\otimes\mathbf{X}_n^T)\Big[(\boldsymbol{\Delta}_n\mathbf{W}_n^{-1})\otimes I_d + I_d\otimes(\boldsymbol{\Delta}_n\mathbf{W}_n^{-1})\Big]\frac{\partial\boldsymbol{\Delta}_n}{\partial\boldsymbol{\mu}_n}\\
&-\left[(\frac{\partial\boldsymbol{\Delta}_n}{\partial\boldsymbol{\mu}_n}\boldsymbol{\Delta}_n\mathbf{X}_n)^T\otimes I_p\right](I_d\otimes K_{dp})(\mathrm{vec}(\mathbf{X}_n^T)\otimes I_d)\cdot\mathbf{W}_n^{-1}\\
&\left.-(I_p\otimes\mathbf{X}_n^T\otimes\mu_n^T\mathbf{W}_n^{-1})\left[((\mathbf{X}_n^T\boldsymbol{\Delta}_n)\otimes I_{d^2})\frac{\partial^2\boldsymbol{\Delta}_n}{\partial\boldsymbol{\mu}_n\boldsymbol{\mu}_n^T} + (\mathbf{X}_n^T\otimes\frac{\partial\boldsymbol{\Delta}_n}{\partial\boldsymbol{\mu}_n})\frac{\partial\boldsymbol{\Delta}_n}{\partial\boldsymbol{\mu}_n}\right]\right\}\boldsymbol{\Delta}_n\mathbf{X}_n,
\end{aligned}$$

where K_{dp} is a $dp\times dp$ commutation matrix and $\frac{\partial^2\boldsymbol{\Delta}_n}{\partial\boldsymbol{\mu}_n\boldsymbol{\mu}_n^T}$ is a $d^3\times d$ dimensional sparse matrix with non-zero quantities $f''(F^{-1}(\mu_{nj}))\left[(F^{-1})'(\mu_{nj})\right]^2 + f'(F^{-1}(\mu_{nj}))(F^{-1})''(\mu_{nj})$ in the $[d(d+1)(j-1)+j, j]$ term, $j=1,\ldots,d$, $n=1,\ldots,N$. Then, after a few steps of algebra, we obtain

$$\begin{aligned}
&(\{\kappa_{ij1}\},\{\kappa_{ij2}\},\cdots,\{\kappa_{ijp}\})^T = \mathrm{E}\left(\frac{\partial^2 U(\boldsymbol{\beta};\boldsymbol{\alpha},\phi)}{\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}^T}\right)\\
=&-\sum_{n=1}^{N}\left\{(\mathbf{X}_n^T\otimes\mathbf{X}_n^T)\Big[(\boldsymbol{\Delta}_n\mathbf{W}_n^{-1})\otimes I_d + I_d\otimes(\boldsymbol{\Delta}_n\mathbf{W}_n^{-1})\Big]\frac{\partial\boldsymbol{\Delta}_n}{\partial\boldsymbol{\mu}_n}\right. \qquad (A.3)\\
&+\left.\left[(\frac{\partial\boldsymbol{\Delta}_n}{\partial\boldsymbol{\mu}_n}\boldsymbol{\Delta}_n\mathbf{X}_n)^T\otimes I_p\right](I_d\otimes K_{dp})(\mathrm{vec}(\mathbf{X}_n^T)\otimes I_d)\cdot\mathbf{W}_n^{-1}\right\}\boldsymbol{\Delta}_n\mathbf{X}_n.
\end{aligned}$$

A.2 An iterative method for obtaining the estimates β^*

Following the GEE method, an iterative procedure for obtaining $\boldsymbol{\beta}^*$ can be described as in what follows.

Step 1: Choose an initial estimate $\boldsymbol{\beta}^{(0)}$ of $\boldsymbol{\beta}$. For example, $\boldsymbol{\beta}^{(0)}$ is obtained by the GEE method with an independence working correlation matrix.

Step 2: Given $\boldsymbol{\beta}^*$ (at the first iteration $\boldsymbol{\beta}^* = \boldsymbol{\beta}^{(0)}$), the moment estimate of the overdispersion parameter is given by

$$\phi^* = \frac{1}{Nd}\sum_{n=1}^{N}\mathbf{Z}_n^{*T}\mathbf{Z}_n^{*T},\ \text{where } \mathbf{Z}_n^{*T} = \mathbf{A}_n^{-1/2}(\boldsymbol{\beta}^*)(\mathbf{y}_n - \boldsymbol{\mu}_n(\boldsymbol{\beta}^*)).$$

Step 3: Given $\boldsymbol{\beta}^*$ and ϕ^* obtained in Steps 1 and 2, calculate the moment estimates $\boldsymbol{\alpha}^*$ of $\boldsymbol{\alpha}$ of the working correlation matrix $R(\boldsymbol{\alpha})$ (see Liang and Zeger, 1986 and Wang and Carey, 2003). For example, if the working correlation matrix is exchangeable, then the exchangeable correlation parameter α is estimated by

$$\alpha^* = \frac{\sum_{n=1}^{N}\sum_{j\neq k} y_{nj}^* y_{nk}^*}{\phi^*(d-1)\sum_{n=1}^{N}\sum_{j=1}^{d} {y_{nj}^*}^2},\ \text{where } y_{nj}^* = (y_{nj} - \mu_{nj}(\boldsymbol{\beta}^*))/\sqrt{v(\mu_{nj}(\boldsymbol{\beta}^*))}$$

(see Wang and Carey, 2003). If the working correlation matrix is AR(1), then the AR(1) correlation parameter α is estimated by

$$\alpha^* = \frac{\sum_{n=1}^{N}\sum_{j=2}^{d} y_{nj}^* y_{n,j-1}^*}{\sum_{n=1}^{N}\{\sum_{j=2}^{d-1} {y_{nj}^*}^2 + ({y_{n1}^*}^2 + {y_{nd}^*}^2)/2\}}, \text{ where } y_{nj}^* = (y_{nj} - \mu_{nj}(\boldsymbol{\beta}^*))/\sqrt{v(\mu_{nj}(\boldsymbol{\beta}^*))}$$

(see Wang and Carey, 2003).

Step 4: Given the working correlation matrix $R(\boldsymbol{\alpha}^*)$ obtained in Step 3, the estimate of $\boldsymbol{\beta}$ is updated according to the modified Fisher scoring formula

$$\boldsymbol{\beta}^{(l+1)} = \boldsymbol{\beta}^{(l)} + \left\{\sum_{n=1}^{N} \mathbf{D}_n^T \mathbf{W}_n^{-1} \mathbf{D}_n\right\}^{-1} \Big|_{\boldsymbol{\beta}=\boldsymbol{\beta}^{(l)}} \left\{\sum_{n=1}^{N} \mathbf{D}_n^T \mathbf{W}_n^{-1} (\mathbf{y}_n - \boldsymbol{\mu}_n(\boldsymbol{\beta})) - \mathbf{I}b(\boldsymbol{\beta})\right\} \Big|_{\boldsymbol{\beta}=\boldsymbol{\beta}^{(l)}},$$

where $\mathbf{D}_n = \partial \boldsymbol{\mu}_n / \partial \boldsymbol{\beta}$ and $\mathbf{W}_n = \mathbf{A}_n(\boldsymbol{\beta}) R(\boldsymbol{\alpha}^*) \mathbf{A}_n(\boldsymbol{\beta})$.

Step 5: Iterate between steps 2 to 4 until a desired convergence criteria (for example, max$|\boldsymbol{\beta}^{(l+1)} - \boldsymbol{\beta}^{(l)}| < 0.001$) is satisfied. At convergence, the estimate of $\boldsymbol{\beta}$ is denoted by $\boldsymbol{\beta}^*$ and the final estimates of $\boldsymbol{\alpha}$ and ϕ are given by $\boldsymbol{\alpha}^*$ and ϕ^* used in the last step of the iteration.