Statistical Science
2003, Vol. 18, No. 3, 377–393

An Overview on Regression Models for Discrete Longitudinal Responses

Brajendra C. Sutradhar

Abstract. In the longitudinal regression setup, interest may be focused primarily on the regression parameters for the marginal expectations of the longitudinal responses, the longitudinal correlation parameters being of secondary interest. Second, interest may be focused on both the regression and the longitudinal correlation parameters. Under the first setup, there exists a "working" correlation matrix based generalized estimating equation (GEE) approach for the estimation of the regression parameters. Under the second setup, there exist two approaches for the joint estimation of the regression and the longitudinal correlations. In one approach, true longitudinal correlations are modeled and the regression and the true correlation parameters are jointly estimated based on a GEE approach. The second approach avoids the specification of the true longitudinal correlation structure and deals with the joint estimation of the regression and a vector of "working" correlation parameters. In this second approach under the second setup, there again exist two joint estimation methods, one requiring moments up to order 4 and the other somehow using moments up to order 2 for the construction of the estimating equations for the "working" correlation parameters. In this paper, we first provide an outline of the desirable features and drawbacks of each of these four existing approaches. By using a general autocorrelation structure to model the true longitudinal correlations, we then provide an outline of the advantages of three new approaches. In the first new approach, the true longitudinal correlations are estimated by the method of moments, whereas the regression estimates are obtained based on a generalized quasi-likelihood (GQL) estimation approach. The other two new approaches simultaneously estimate the regression and the true longitudinal correlation parameters. It is shown through a simulation study that, among these three new approaches, the first approach performs the best in estimating both the regression and the true correlation parameters, even though the longitudinal correlations are estimated separately by the method of moments.

Key words and phrases: Binary and count responses, repeated measures, time dependence between the responses, marginal models, consistent and efficient estimators.

1. INTRODUCTION

In longitudinal studies, a small number of repeated observations of a response variable and a set of covariates are made on a large number of individuals across occasions. For example, in health care utilization data, the number of visits to the physician by a large number of independent individuals may be recorded over a period of several years. Also, the information on the covariates—gender, number of chronic conditions, education level and age—may be recorded for each individual. Note that as the number of visits to the physician over the years may be treated as repeated

Brajendra C. Sutradhar is Professor of Statistics, Department of Statistics, Memorial University of Newfoundland, St. John's, Newfoundland, Canada A1C 557 (e-mail: bsutradh@math.mun.ca).

measurements made on the same individual, it is likely that these responses will be correlated. The scientific concern is to find the effects of the covariates on the physician visits after taking the longitudinal correlations into account.

Note that as the joint probability model for the discrete responses, such as for the Poisson responses in the above example, is unknown, the estimation of the regression effects, after taking the longitudinal correlation structure into account, has proven to be difficult. In a seminal paper, Liang and Zeger (1986) have bypassed the joint probability model and introduced a "working" correlation structure based generalized estimating equation (GEE) approach to obtain consistent and efficient estimators for the regression parameters that relate the expectation of the response to a set of covariates by some known link functions. To be specific, Liang and Zeger (1986) estimate the working correlation parameters by the method of moments and use these estimates in the GEE for the regression parameters. This GEE approach is usually referred to as the GEE1 approach, which, for convenience, we refer to as the GEE1M approach to distinguish this method of moments based GEE approach from other existing "working" correlation based approaches.

Fitzmaurice, Laird and Rotnitzky [1993, (2)–(4)] discuss a GEE approach following Liang and Zeger (1986) but estimate the "working" correlations through a second set of estimating equations which is quite similar to the set of estimating equations for the regression parameters. Note that, in this approach, the construction of the estimating equations for the "working" correlation parameters requires another "working" correlation matrix consisting of the third- and fourth-order moments of the responses, although Fitzmaurice, Laird and Rotnitzky (1993) use a "working" independence approach to construct such higher-order moments based estimating equations. We refer to this approach as the GEE1J_1 approach, which is a "working" correlation based GEE1 approach but the "working" correlation parameters are jointly estimated along with the regression parameters through an iterative process. For simplicity, Lipsitz, Fitzmaurice, Orav and Laird (1994) introduced a one-step estimator for the regression parameters as opposed to the fully iterated GEE estimator. This approach is, however, a special case of the GEE1J_1 approach discussed by Fitzmaurice, Laird and Rotnitzky (1993), where the "working" correlations are estimated by using a second set of estimating equations.

Note that under the assumption that the cluster correlations of binary responses arise due to a common random effect shared by the individuals of the cluster, Neuhaus (1993) examined the efficiencies of the independence and pairwise "working" correlation based GEE approaches. One of the problems with this type of approach is that the lag correlations of the repeated responses in a cluster cannot be well explained through the mixed model. Thus, the mixed model considered by Neuhaus (1993) appears to be appropriate for the analysis of cluster data with responses collected from different individuals of the cluster as opposed to the cluster data with responses collected repeatedly from an individual.

Similar to Fitzmaurice, Laird and Rotnitzky (1993), Hall and Severini (1998) also estimate the regression and the "working" correlation parameters simultaneously. Hall and Severini (1998) referred to their approach as the extended generalized estimating equation (EGEE) approach. This EGEE approach, unlike the approach of Fitzmaurice, Laird and Rotnitzky, does not require any third- and fourth-order moments based estimating equations for the "working" correlation parameters. It rather uses a second moments based set of estimating equations for the "working" correlation parameters. We refer to the EGEE approach of Hall and Severini as the GEE1J_2 approach as it yields joint estimates for the regression and the "working" correlation parameters.

Zhao and Prentice (1990), Prentice and Zhao (1991) and Zhao, Prentice and Self (1992) have described extensions of the GEE methodology to allow for joint estimation of the regression and the true longitudinal correlation parameters. More specifically, Zhao and Prentice (1990) propose a joint probability model that is based on the "quadratic exponential family," with the three- and higher-way association parameters equal to 0. The "quadratic exponential family" based association parameters are then estimated by using the likelihood estimating or, equivalently, the generalized estimating equation approach. Similarly, a partly exponential model is introduced by Zhao, Prentice and Self (1992) which accommodates the association between the responses and the likelihood, or, equivalently, the GEE approach is used to estimate the mean and the association parameters of the model. These GEE based methods for the joint estimation are referred to as the GEE2 approaches.

In Section 2, we discuss the advantages and drawbacks of each of the above-mentioned GEE1M, GEE1J_1, GEE1J_2 and GEE2 approaches. In Section 3,

we review the performance of the generalized quasi-likelihood (GQL) approach introduced by Sutradhar and Das (1999) for the estimation of the regression parameters, the true longitudinal correlation parameters being of secondary interest. In the same section, we introduce two new approaches for the joint estimation of the regression and the true correlation parameters. One of these two approaches is developed along the lines of the GEE2 approach of Zhao and Prentice (1990) and Prentice and Zhao (1991), whereas the second approach is developed following the EGEE approach of Hall and Severini (1998). These two techniques will be referred to as the general GEE2 (GGEE2) and general EGEE (GEGEE) approaches, respectively. A simulation study in Section 4, however, shows that the GQL approach is superior to the GGEE2 and GEGEE approaches for the estimation of the regression and the correlation parameters of the longitudinal model. The GQL approach is then applied to analyze two sets of longitudinal data as an illustration. Some concluding remarks are made in Section 6.

2. REGRESSION MODELS FOR LONGITUDINAL DATA

Suppose that a scalar response y_{it} and a p-dimensional vector of covariates x_{it} are observed for clusters $i = 1, \ldots, K$ at a time point t, $t = 1, \ldots, n$. For the ith cluster, let $y_i = (y_{i1}, \ldots, y_{it}, \ldots, y_{in})^T$ be the response vector and let $X_i = (x_{i1}, \ldots, x_{it}, \ldots, x_{in})^T$ be the $n \times p$ matrix of covariates. Furthermore, suppose that the marginal density of the response y_{it} is of the exponential family form

$$f(y_{it}) = \exp\big[\{y_{it}\theta_{it} - a(\theta_{it})\}\phi + b(y_{it}, \phi)\big] \tag{2.1}$$

(Liang and Zeger, 1986), where $\theta_{it} = h(\eta_{it})$ with $\eta_{it} = x_{it}^T\beta$, $a(\cdot)$, $b(\cdot)$ and $h(\cdot)$ are of known functional form, ϕ is a possibly unknown scale parameter and β is the $p \times 1$ vector of parameters of interest. In many important situations, for example, for binary and Poisson data, one may use $\phi = 1$. Consequently, for simplicity, we use $\phi = 1$ in (2.1) and write the mean and the variance of y_{it} as

$$E(Y_{it}) = a'(\theta_{it}) \quad \text{and} \quad \operatorname{var}(Y_{it}) = a''(\theta_{it}),$$

where $a'(\theta_{it})$ and $a''(\theta_{it})$ are, respectively, the first and second derivatives of $a(\theta_{it})$ with respect to θ_{it}. For the health care utilization problem introduced in Section 1,

$$a'(\theta_{it}) = a''(\theta_{it}) = \exp(x_{it}^T\beta),$$

where x_{it} is the 4×1 vector of covariates—gender, number of chronic conditions, education level and age—for the ith individual at time t, $t = 1, \ldots, n$.

Furthermore, in the longitudinal setup, the components of the vector y_i are repeated responses, which are likely to be correlated. Let $C(\rho)$ be the $n \times n$ true correlation matrix of y_i, $i = 1, \ldots, K$, which is unknown in practice. Here ρ is, say, an $s \times 1$ vector of correlation parameters which fully characterizes $C(\rho)$. It is of primary interest to estimate β after taking the longitudinal correlation structure $C(\rho)$ into account. For the health care utilization data, this amounts to the estimation of the effects of all four covariates after taking the longitudinal correlations of the individuals into account.

2.1 "Working" Correlation Based GEE Approaches and Their Limitations

For $A_i = \operatorname{diag}[a''(\theta_{i1}), \ldots, a''(\theta_{it}), \ldots, a''(\theta_{in})]$ and for known $C(\rho)$, the quasi-likelihood estimator $\hat{\beta}_Q$ of β under (2.1) is the solution of the score equation

$$\sum_{i=1}^{K} X_i^T A_i \Sigma_i^{-1}(\rho)(y_i - \mu_i) = 0 \tag{2.2}$$

(McCullagh, 1983), where $\mu_i = (a'(\theta_{i1}), \ldots, a'(\theta_{it}), \ldots, a'(\theta_{in}))^T$ and $\Sigma_i(\rho) = A_i^{1/2}C(\rho)A_i^{1/2}$ is the true covariance of y_i. As $C(\rho)$ is unknown in practice, it is impossible to estimate β by solving the estimating equations (2.2). To overcome this problem of unknown $C(\rho)$, Liang and Zeger (1986) introduced a "working" correlation approach, where the estimate of β is obtained by solving the estimating equations

$$\sum_{i=1}^{K} X_i^T A_i^{1/2} R_i^{-1}(\hat{\alpha}) A_i^{-1/2}(y_i - \mu_i) = 0, \tag{2.3}$$

where $R(\alpha)$ is the "working" correlation matrix used for $C(\rho)$ with α as, say, an $s_2 \times 1$ vector of correlation parameters which fully characterizes the $R(\alpha)$ matrix. In this "working" correlation approach, $\hat{\alpha}$ is a moment estimator of α computed based on the Pearson residuals $r_{it} = (y_{it} - a'(\theta_{it}))/\{a''(\theta_{it})\}^{1/2}$. The exact form of the estimator of α, however, depends on the assumed form of $R(\alpha)$.

2.1.1 GEE1M *approach.* Let $\hat{\alpha}_{\mathrm{M}}$ be the moment estimator of α and let $\hat{\beta}_{\mathrm{M}}$ be the solution of (2.3) for β by using $\hat{\alpha} = \hat{\alpha}_{\mathrm{M}}$. Recall that this technique for the estimation of β based on the moment estimator $\hat{\alpha}_M$ is referred to as the GEE1M approach. This GEE1M approach has, however, many pitfalls which are discussed by Crowder (1995) and Sutradhar and Das (1999). To

be specific, as demonstrated by Crowder (1995), there may not exist any solutions for $\hat{\alpha}_{\mathrm{M}}$ for various possible reasons, leading to the complete breakdown of the estimation of the regression parameters. Second, even if $\hat{\alpha}_{\mathrm{M}}$ exists, as α is not defined as the correlation parameter of the model, the limiting value of $\hat{\alpha}_{\mathrm{M}}$ will depend on the forms chosen for $R(\alpha)$ and the supplementary estimating equations defining $\hat{\alpha}_{\mathrm{M}}$. Suppose that $\hat{\alpha}_{\mathrm{M}}$ converges in probability to a quantity $\tilde{\alpha}$. In this case, the GEE approach still gives a consistent estimator of the regression parameter β, but this estimator $(\hat{\beta}_{\mathrm{M}})$ is generally less efficient than the regression estimator $\hat{\beta}_{\mathrm{I}}$ obtained based on the independence estimating equation approach. To verify this, by (2.3), one may compute the asymptotic $(K \to \infty)$ covariance matrix of $\hat{\beta}_{\mathrm{M}}$ and $\hat{\beta}_{\mathrm{I}}$ as

$$
\begin{aligned}
V_{\mathrm{M}} = \lim_{K\to\infty} &\left\{\sum_{i=1}^{K} X_i^T A_i^{1/2} R^{-1}(\tilde{\alpha}) A_i^{1/2} X_i\right\}^{-1} \\
&\cdot \left\{\sum_{i=1}^{K} X_i^T A_i^{1/2} R^{-1}(\tilde{\alpha}) C(\rho) R^{-1}(\tilde{\alpha}) A_i^{1/2} X_i\right\} \\
&\cdot \left\{\sum_{i=1}^{K} X_i^T A_i^{1/2} R^{-1}(\tilde{\alpha}) A_i^{1/2} X_i\right\}^{-1}
\end{aligned} \tag{2.4}
$$

and

$$
\begin{aligned}
V_{\mathrm{I}} = \lim_{K\to\infty} &\left\{\sum_{i=1}^{K} X_i^T A_i X_i\right\}^{-1} \\
&\cdot \left\{\sum_{i=1}^{K} X_i^T A_i^{1/2} C(\rho) A_i^{1/2} X_i\right\} \\
&\cdot \left\{\sum_{i=1}^{K} X_i^T A_i X_i\right\}^{-1},
\end{aligned} \tag{2.5}
$$

respectively, and compare their respective diagonal elements.

Note that, for a given true correlation structure $C(\rho)$ and various choices of the "working" correlation matrix $R(\alpha)$, it may be shown that, in most cases, the diagonal elements of the V_{I} matrix are smaller than those of the V_{M} matrix (Sutradhar and Das, 1999). The reverse is true in some cases, especially when $C(\rho)$ has the Gaussian AR(1) structure. Thus, $\hat{\beta}_{\mathrm{M}}$ is generally less efficient than $\hat{\beta}_{\mathrm{I}}$. Consequently, as $\hat{\beta}_{\mathrm{I}}$ is always consistent, and as it is also easier to compute, there is no reason to prefer $\hat{\beta}_{\mathrm{M}}$ over $\hat{\beta}_{\mathrm{I}}$ for the estimation of β. Note, however, that as $\hat{\beta}_{\mathrm{I}}$ is not uniformly more efficient than $\hat{\beta}_{\mathrm{M}}$, in Section 3 we review a generalized quasi-likelihood (GQL) estimator which is consistent and always more efficient than $\hat{\beta}_{\mathrm{I}}$ for β.

2.1.2 GEE1J$_1$ *approach.* In this approach, the regression parameters are estimated as in the GEE1M approach but the "working" correlation parameter α is estimated by using a second set of estimating equations given by

$$
\sum_{i=1}^{K} \frac{\partial \sigma_i^T}{\partial \alpha} \Omega_{is}^{-1}(\alpha)\big(s_i - \sigma_i(\alpha)\big) = 0 \tag{2.6}
$$

[cf. Fitzmaurice, Laird and Rotnitzky, 1993, (4), page 287], where $s_i = [(y_i - \mu_{i1})(y_{i2} - \mu_{i2}), \ldots, (y_{i(n-1)} - \mu_{i(n-1)})(y_{in} - \mu_{in})]'$ is the $\{n(n-1)/2\} \times 1$ vector of distinct products, $\sigma_i(\alpha) = E(S_i)$ and $\Omega_{is}(\alpha) = \mathrm{cov}(S_i)$ under the "working" correlation structure. More specifically,

$$
\begin{aligned}
\sigma_i(\alpha) = \big[&\alpha_{|1-2|}\{a''(\theta_{i1})a''(\theta_{i2})\}^{1/2}, \ldots, \\
&\alpha_{|t-t'|}\{a''(\theta_{it})a''(\theta_{it'})\}^{1/2}, \ldots, \\
&\alpha_{|(n-1)-n|}\{a''(\theta_{i(n-1)})a''(\theta_{in})\}^{1/2}\big]^T
\end{aligned} \tag{2.7}
$$

for $R(\alpha) = (\alpha_{|t-t'|})$ with $\alpha_0 = 1$, but the computation of $\Omega_{is}(\alpha)$ requires the formulas for the fourth-order moments, which are unknown. Fitzmaurice, Laird and Rotnitzky (1993) have used $\Omega_{is}(\alpha) = I_{n(n-1)/2}$, where $I_{n(n-1)/2}$ is the $\{n(n-1)/2\} \times \{n(n-1)/2\}$ identity matrix. Let $\hat{\alpha}_{\mathrm{G}}$ be the solution of (2.6) for this choice of the weight matrix $\Omega_{is}(\alpha)$. Note, however, that as $E(S_i)$ under the true correlation structure is a function of $\rho_{|t-t'|}$, contrary to the claim by Fitzmaurice, Laird and Rotnitzky (1993), $\hat{\alpha}_{\mathrm{G}}$ converges to a quantity different from α. Let this quantity be α^*. Consequently, similar to the GEE1M estimator, the $\hat{\alpha}_{\mathrm{G}}$ based solution of (2.3), say $\hat{\beta}_{\mathrm{G}}$, may also be less efficient than $\hat{\beta}_{\mathrm{I}}$ (Sutradhar and Kumar, 2001). For the computation of the efficiency of $\hat{\beta}_{\mathrm{G}}$ as compared to $\hat{\beta}_{\mathrm{I}}$, we provide the formula for the asymptotic covariance of $\hat{\beta}_{\mathrm{G}}$ as

$$
\begin{aligned}
V_{\mathrm{J}_1} = \lim_{K\to\infty} &\left\{\sum_{i=1}^{K} X_i^T A_i^{1/2} R^{-1}(\alpha^*) A_i^{1/2} X_i\right\}^{-1} \\
&\cdot \left\{\sum_{i=1}^{K} X_i^T A_i^{1/2} R^{-1}(\alpha^*) C(\rho) R^{-1}(\alpha^*) Z_i^{1/2} X_i\right\} \\
&\cdot \left\{\sum_{i=1}^{K} X_i^T A_i^{1/2} R^{-1}(\alpha^*) A_i^{1/2} X_i\right\}^{-1},
\end{aligned} \tag{2.8}
$$

whereas the formula for the covariance matrix of $\hat{\beta}_{\mathrm{I}}$ is given by (2.5).

Note that some authors, for example, Prentice and Zhao (1991), suggest using a normal based pseudo

weight matrix Ω_{is}, which is different from the identity weight matrix used by Fitzmaurice, Laird and Rotnitzky (1993). For this purpose, Prentice and Zhao constructed the fourth-order moments by pretending that y_i follows an n-dimensional normal distribution with mean vector μ_i and covariance matrix $V_i(\alpha) = A_i^{1/2} R(\alpha) A_i^{1/2}$. For example, for $t < t' < l < m$, the formula for the covariance between the distinct corrected products $(y_{it} - \mu_{it})(y_{it'} - \mu_{it'})$ and $(y_{il} - \mu_{il})(y_{im} - \mu_{im})$ is given by

$$\begin{aligned} &\alpha_{|t-l|}\alpha_{|t'-m|}\{a''(\theta_{it})a''(\theta_{il})\}^{1/2} \\ &\quad\cdot\{a''(\theta_{it'})a''(\theta_{im})\}^{1/2} \\ &+\alpha_{|t-m|}\alpha_{|t'-l|}\{a''(\theta_{it})a''(\theta_{im})\}^{1/2} \\ &\quad\cdot\{a''(\theta_{it'})a''(\theta_{il})\}^{1/2}. \end{aligned} \tag{2.9}$$

Once the Ω_{is} is constructed following (2.9), the estimating equation (2.6) is solved for α. The estimate of α is then used in (2.3) to estimate β. Note, however, that as the estimating equation for β still uses the "working" correlation matrix, this approach of Prentice and Zhao (1991), similar to the GEE1M approach, may also produce a less efficient estimator of β than $\hat{\beta}_{\mathrm{I}}$.

2.1.3 GEE1J$_2$ *approach.* The construction of the estimating equation (2.6) for the "working" correlation parameter α is complicated. This is because the weight matrix involved in this estimating equation requires the computation of the fourth-order moments of the responses. Hall and Severini (1998) avoided this problem and estimated the "working" correlation parameter by using the estimating equation

$$\begin{aligned} &K^{-1}\sum_{i=1}^{K}[s_{id}^T, W_{i\bar{d}}^T] \\ &\quad\cdot\big[(u_i - v_i)^T, \big(s_i - \sigma_i(\alpha)\big)^T\big] = 0, \end{aligned} \tag{2.10}$$

which requires second-order moments only. Note that, in (2.10), s_i and $\sigma_i(\alpha)$ are as in (2.6), $u_i = [(y_{i1} - \mu_{i1})^2, \ldots, (y_{it} - \mu_{it})^2, \ldots, (y_{in} - \mu_{in})^2]^T$ is the $n \times 1$ vector of corrected squares and $v_i = E(U_i)$ under the "working" correlation model and W_{id} and $W_{i\bar{d}}$ are $n \times 1$ and $\{n(n-1)/2\} \times 1$ vectors consisting of the diagonal and distinct off-diagonal elements of the $W_i(\alpha)$ matrix, respectively, with

$$\begin{aligned} W_i(\alpha) &= \frac{\partial V_i^{-1}(\alpha)}{\partial\alpha} \\ &= -A_i^{-1/2} R^{-1}(\alpha)\frac{\partial R(\alpha)}{\partial\alpha} R^{-1}(\alpha) A_i^{-1/2}, \end{aligned} \tag{2.11}$$

where the specific form of $\partial R(\alpha)/\partial\alpha$ will depend on the structure of the $R(\alpha)$ matrix. Let $\hat{\alpha}_{\mathrm{EG}}$ be the solution of (2.10) for α.

Furthermore, in (2.10), $E(U_i)$ and $E(S_i)$ are computed under the "working" correlation model. But, in reality, these expectations are the functions of the true correlation parameters $\rho_{|t-t'|}$. Consequently, similar to $\hat{\alpha}_{\mathrm{G}}$ under the GEE1J$_1$ approach, $\hat{\alpha}_{\mathrm{EG}}$ obtained from (2.10) may also not converge to α. Let $\hat{\alpha}_{\mathrm{EG}}$ converge to $\bar{\alpha}$. Then the estimator of β, say $\hat{\beta}_{\mathrm{EG}}$, is obtained from (2.3) by putting $\hat{\alpha} = \hat{\alpha}_{\mathrm{EG}}$. Since the estimating equations (2.3) and (2.10) are jointly solved under this approach, under some mild conditions $(\hat{\beta}_{\mathrm{EG}}^T, \hat{\alpha}_{\mathrm{EG}})^T$ has the asymptotic covariance matrix given by

$$\begin{aligned} V_{\mathrm{J}_2}^* = \lim_{K\to\infty}&\left[\sum_{i=1}^{K}\begin{pmatrix} A_{i11} & A_{i12} \\ A_{i21} & A_{i22}\end{pmatrix}\right]^{-1} \\ &\cdot\left[\sum_{i=1}^{K}\begin{pmatrix} M_{i11} & M_{i12} \\ M_{i21} & M_{i22}\end{pmatrix}\right] \\ &\cdot\left[\sum_{i=1}^{K}\begin{pmatrix} A_{i11} & A_{i12} \\ A_{i21} & A_{i22}\end{pmatrix}\right]^{-1}, \end{aligned} \tag{2.12}$$

where

$$\begin{aligned} &A_{i11} = -\tilde{D}_i^T V_i^{-1}(\bar{\alpha})\tilde{D}_i, \quad A_{i12} = 0, \\ &A_{i21} = -\tilde{W}_i^T \lambda'_{i\beta}, \qquad\quad A_{i22} = -\tilde{Q}_i^T \lambda'_{i\alpha}, \end{aligned}$$

with

$$\begin{aligned} &\tilde{D}_i = \frac{\partial\mu_i}{\partial\beta^T}, \quad \tilde{W}_i = \big[W_{id}^T, W_{i\bar{d}}^T\big]^T, \\ &\lambda_i = \big(\nu_i^T, \sigma_i^T(\alpha)\big)^T, \quad \lambda'_{i\beta} = \frac{\partial\lambda_i}{\partial\beta^T}, \quad \lambda'_{i\alpha} = \frac{\partial\lambda_i}{\partial\alpha}, \end{aligned}$$

and where

$$\begin{aligned} &M_{i11} = \tilde{D}_i^T V_i^{-1}(\alpha)\Sigma_i(\rho)V_i^{-1}(\alpha)\tilde{D}_i, \\ &M_{i12} = \tilde{D}_i^T V_i^{-1}(\alpha)\operatorname{cov}(Y_i, F_i)\tilde{W}_i, \\ &M_{i21} = M_{i12}^T, \quad M_{i22} = \tilde{W}_i^T \operatorname{var}(F_i)\tilde{W}_i, \end{aligned}$$

with $f_i = (u_i^T, s_i^T)^T$. The covariance matrices, $\operatorname{cov}(Y_i, F_i)$ and $\operatorname{var}(F_i)$, may be computed following (2.9) by using a pseudo-normal distribution for y_i. The leading $p \times p$ matrix of $V_{\mathrm{J}_2}^*$ (2.12) provides the $\operatorname{cov}(\hat{\beta}_{\mathrm{EG}})$, which is used to compute the efficiency of $\hat{\beta}_{\mathrm{EG}}$ as compared to $\hat{\beta}_{\mathrm{I}}$. By numerical comparisons as in the GEE1J$_1$ approach, it can be shown that $\hat{\beta}_{\mathrm{EG}}$ may be less efficient than $\hat{\beta}_{\mathrm{I}}$ under misspecification of the

"working" correlation structure (Sutradhar and Kumar, 2001). Thus, from an efficiency point of view, none of the three "working" correlation based approaches, namely, GEE1M, GEE1J$_1$ and GEE1J$_2$, performs well when estimating the regression parameter vector β. This is because they may produce inefficient estimators as compared to the "working" independence based estimating equation approach. Furthermore, estimation of β is naturally more complicated under these approaches as compared to the computation of $\hat{\beta}_{\mathrm{I}}$.

Note, however, that although $\hat{\beta}_{\mathrm{I}}$ performs, in general, better than any of the three estimators $\hat{\beta}_{\mathrm{M}}$ or $\hat{\beta}_{\mathrm{G}}$ or $\hat{\beta}_{\mathrm{EG}}$, this is not a uniformly more efficient estimator. The efficiency of $\hat{\beta}_{\mathrm{I}}$ can be considerably low, for example, for the case when the true correlation structure is AR(1) (Sutradhar and Das, 1999, Table 2). This suggests that we seek a better estimator than $\hat{\beta}_{\mathrm{I}}$ in terms of both consistency and efficiency, which we discuss in Section 3.

2.2 GEE2 Approach and Its Limitations

As opposed to the "working" correlation structure based GEE1M, GEE1J$_1$ and GEE1J$_2$ approaches, one may construct a true correlation structure based joint estimating equation approach (Prentice and Zhao, 1991) for the estimation of β and α. For true correlation structure $C(\rho)$, these estimating equations for β and ρ may be written as

$$\sum_{i=1}^{K} X_i^T A_i^{1/2} C^{-1}(\rho) A_i^{-1/2}(y_i - \mu_i) = 0 \tag{2.13}$$

and

$$\sum_{i=1}^{K} \frac{\partial \tilde{\sigma}_i^T(\rho)}{\partial \rho} \tilde{\Omega}_{is}^{-1}(\rho)\big(s_i - \tilde{\sigma}_i(\rho)\big) = 0, \tag{2.14}$$

respectively. Let β_{G}^* and ρ_{G}^* be the solutions of (2.13) and (2.14) for β and ρ, respectively. Note that $\tilde{\sigma}_i(\rho)$ and $\tilde{\Omega}_{is}(\rho)$ in (2.14) are the expectation and the covariance matrix of s_i (2.6) under the true correlation structure $C(\rho)$. Thus, for known $C(\rho)$ and $\tilde{\Omega}_{is}(\rho)$, β_{G}^* and ρ_{G}^* are the well-known quasi-likelihood (QL) estimators of β and ρ, respectively. Consequently, β_{G}^* will be consistent and also more efficient than $\hat{\beta}_{\mathrm{I}}$. Note that even if one can write a suitable structure for the $C(\rho)$ matrix in the longitudinal setup (see Section 3), the construction of the $\tilde{\Omega}_{is}(\rho)$ matrix is extremely difficult. Thus, the estimating equations (2.13) and (2.14) become useless, in general, in the longitudinal setup.

Furthermore, note that some authors have exploited the estimating equations (2.13) and (2.14) under certain special correlation models. For example, Zhao and Prentice (1990) attempt to use these estimating equations by modeling the correlations of the responses arising from a "quadratic exponential family" model. Similarly, Fitzmaurice and Laird (1993) modeled the correlation structure based on a mixed parameter model. Note that these approaches, however, are not able to model the Gaussian-type AR(1), MA(1) and exchangeable correlation structures appropriately for the longitudinal data.

In the next section, the true longitudinal correlations of the data are modeled through a general autocorrelation structure which accommodates the usual AR(1), MA(1) and exchangeable-type correlation patterns. This general correlation structure is then used to obtain a generalized quasi-likelihood (GQL) estimator for the regression vector β, which requires only a moment estimate for the true correlation parameter ρ.

3. GENERAL AUTOCORRELATION STRUCTURE BASED GEE APPROACHES

The "working" correlation structure based GEE estimators $\hat{\beta}_{\mathrm{M}}$, $\hat{\beta}_{\mathrm{G}}$ and $\hat{\beta}_{\mathrm{EG}}$ were originally developed to gain efficiency in β estimation, as compared to the "working" independence based estimator $\hat{\beta}_{\mathrm{I}}$. But, as was demonstrated in the previous section (see also Sutradhar and Das, 1999; Sutradhar and Kumar, 2001), these three estimators are rather less efficient than $\hat{\beta}_{\mathrm{I}}$ in many situations. So, it remains as an important issue to find an estimator of β which will always be more efficient than $\hat{\beta}_{\mathrm{I}}$. In this section, we review the longitudinal model with true correlation structure as suggested by Sutradhar and Das (1999, Section 3) and include three new estimation approaches for the estimation of β under such a longitudinal model.

In the first approach, we follow Sutradhar and Das (1999) and construct a generalized quasi-likelihood (GQL) estimator for β, where the associated longitudinal correlations are estimated by the method of moments. In the second approach, we exploit the true correlation structure as in the first approach but estimate the regression and the correlation parameters jointly following the GEE2 approach of Prentice and Zhao (1991). The third approach is similar to the second approach, but it estimates the regression and the correlation parameters jointly by following the EGEE approach of Hall and Severini (1998). We refer to these approaches as the GQL, GGEE2 and GEGEE, respectively.

3.1 GQL Estimation Approach

In this approach, the quasi-likelihood estimator of β is the root of the score equation

$$\sum_{i=1}^{K} X_i^T A_i \Sigma_i^{-1}(\rho)(y_i - \mu_i) = 0, \tag{3.1}$$

where $\Sigma_i(\rho) = A_i^{1/2} C(\rho) A_i^{1/2}$, with $C(\rho)$ as the true correlation structure, given by

$$C(\rho_1, \ldots, \rho_{n-1}) = \begin{bmatrix} 1 & \rho_1 & \rho_2 & \cdots & \rho_{n-1} \\ \rho_1 & 1 & \rho_1 & \cdots & \rho_{n-2} \\ \vdots & \vdots & \vdots & & \vdots \\ \rho_{n-1} & \rho_{n-2} & \rho_{n-3} & \cdots & 1 \end{bmatrix} \tag{3.2}$$

(Sutradhar and Das, 1999, Section 3). The GQL (generalized quasi-likelihood) estimate of β is then computed by solving the estimating equation

$$\sum_{i=1}^{K} X_i^T A_i \Sigma_i^{-1}(\hat{\rho})(y_i - \mu_i) = 0, \tag{3.3}$$

where $\Sigma_i(\hat{\rho}) = A_i^{1/2} C(\hat{\rho}_1, \ldots, \hat{\rho}_{n-1}) A_i^{1/2}$, and for $l = |t - t'|$, $t \neq t'$, $t, t' = 1, \ldots, n$, the autocorrelation of lag l, ρ_l, is estimated by the method of moments as

$$\hat{\rho}_l = \frac{\sum_{i=1}^{K} \sum_{t=1}^{n-l} \tilde{y}_{it} \tilde{y}_{i,t+l} / K(n-l)}{\sum_{i=1}^{K} \sum_{t=1}^{n} \tilde{y}_{it}^2 / Kn} \tag{3.4}$$

[cf. Sutradhar and Kovacevic, 2000, (2.18)], where $\tilde{y}_{it}$ is the standardized residual, defined as $\tilde{y}_{it} = (y_{it} - \mu_{it})/\{a''(\theta_{it})\}^{1/2}$. Let $\hat{\beta}_{\mathrm{GQL}}$ denote this estimator, which is consistent for β. Under some mild conditions, it can be shown that $\hat{\beta}_{\mathrm{GQL}}$ has the asymptotic covariance matrix V_{G}^* given by

$$V_{\mathrm{G}}^* = \lim_{K \to \infty} \left\{ \sum_{i=1}^{K} X_i^T A_i^{1/2} C^{-1} \cdot (\rho_1, \ldots, \rho_{n-1}) A_i^{1/2} X_i \right\}^{-1}. \tag{3.5}$$

Next, as the $\hat{\beta}_{\mathrm{I}}$ has the asymptotic covariance matrix given by (2.5), similar to the correlated linear model case, a comparison of (3.5) with (2.5) shows that $\hat{\beta}_{\mathrm{GQL}}$ is always more efficient than $\hat{\beta}_{\mathrm{I}}$.

3.2 GGEE2 Approach for Regression and Longitudinal Correlation Parameters

As opposed to the GQL approach, we now estimate β and ρ_l, $l = 1, \ldots, n-1$, by solving two sets of estimating equations. The estimating equation for β remains the same as (3.3), whereas a new set of estimating equations for ρ_l, $l = 1, \ldots, n-1$, is defined as

$$\sum_{i=1}^{K} \frac{\partial \tilde{\sigma}_i^T(\rho_1, \ldots, \rho_{n-1})}{\partial \rho_l} \tilde{\Omega}_{is}^{-1}(\rho_1, \ldots, \rho_{n-1}) \cdot \big(s_i - \tilde{\sigma}_i(\rho_1, \ldots, \rho_{n-1})\big) = 0. \tag{3.6}$$

Note that although the estimating equations (3.6) are similar to the estimating equations (2.14), they are, however, quite different. This is because the estimating equations (2.14) use a "quadratic exponential family" based correlation structure under the GEE2 approach of Prentice and Zhao (1991), whereas (3.6) use a general longitudinal autocorrelation structure which accommodates Gaussian-type AR(1), MA(1) and exchangeable correlations. Furthermore, note that although Prentice and Zhao (1991) solve the estimating equations for β and correlation parameters simultaneously, for simplicity, we solve (3.3) and (3.6) separately (cf. Fitzmaurice, Laird and Rotnitzky, 1993), in a cycle of iterations, to obtain the estimates of β and $\rho_1, \ldots, \rho_{n-1}$.

To be specific, in (3.6) $\tilde{\sigma}_i(\rho_1, \ldots, \rho_{n-1})$ is given by

$$\begin{aligned} &\tilde{\sigma}_i(\rho_1, \ldots, \rho_{n-1}) \\ &\quad = \big[\rho_1\{a''(\theta_{i1})a''(\theta_{i2})\}^{1/2}, \ldots, \\ &\qquad \rho_{|t-t'|}\{a''(\theta_{it})a''(\theta_{it'})\}^{1/2}, \ldots, \\ &\qquad\quad \rho_1\{a''(\theta_{i(n-1)})a''(\theta_{in})\}^{1/2}\big]^T, \end{aligned} \tag{3.7}$$

so that

$$\begin{aligned} &\frac{\partial \tilde{\sigma}_i(\rho_1, \ldots, \rho_{n-1})}{\partial \rho_l} \\ &\quad = \big[\delta_{n-1,l}^T\{a''(\theta_{i1})a''(\theta_{i(l+1)})\}^{1/2}, \ldots, \\ &\qquad \delta_{n-t,l}^T\{a''(\theta_{it})a''(\theta_{i(t+l)})\}^{1/2}, \ldots, \\ &\qquad\quad \delta_{n-(n-1),l}^T\{a''(\theta_{i(n-l)})a''(\theta_{in})\}^{1/2}\big]^T, \end{aligned} \tag{3.8}$$

where $\delta_{n-t,l}^T$ is the $1 \times (n-t)$ vector with 1 at the lth position and 0 elsewhere whenever $l \leq n-t$, for t ranging from 1 to $n-1$. For $l > n-t$, $\delta_{n-t,l}^T$ is a zero vector always. Now, by constructing the covariance matrix $\tilde{\Omega}_{is}(\rho_1, \ldots, \rho_{n-1})$ of s_i, based on the normality assumption of y_i with mean μ_i and covariance matrix $\Sigma_i(\rho_1, \ldots, \rho_{n-1}) = A_i^{1/2} C(\rho_1, \ldots, \rho_{n-1}) A_i^{1/2}$, one obtains the GEE2 estimators of $\rho_1, \ldots, \rho_{n-1}$ by solving the estimating equations (3.6). Let $\hat{\beta}_{\mathrm{GGEE2}}$ and $\hat{\rho}_{1,\mathrm{GGEE2}}, \ldots, \hat{\rho}_{n-1,\mathrm{GGEE2}}$ be the GGEE2 estimators of

β and $\rho_1, \ldots, \rho_{n-1}$, respectively, which are obtained by solving the estimating equations (3.3) and (3.6).

It then follows that $\hat{\beta}_{\mathrm{GGEE2}}$ has the same asymptotic covariance structure as V_{G}^*, the covariance of $\hat{\beta}_{\mathrm{GQL}}$ given in (3.5), but the estimates of V_{G}^* under the two approaches GGEE2 and GQL are generally different. This is because these two approaches yield different estimates for $\rho_1, \ldots, \rho_{n-1}$. Note that as $\tilde{\Omega}_{is}(\rho_1, \ldots, \rho_{n-1})$ is a "working" covariance matrix, an estimate of the $(n-1) \times (n-1)$ covariance matrix of $\hat{\rho}_{\mathrm{GGEE2}} = (\hat{\rho}_{1,\mathrm{GGEE2}}, \ldots, \hat{\rho}_{n-1,\mathrm{GGEE2}})^T$ may be obtained by using a "sandwich"-type formula given by

$$\begin{aligned}\hat{\mathrm{cov}}(\hat{\rho}_{\mathrm{GGEE2}}) = {} & \left[\sum_{i=1}^{K}\left(\frac{\partial\tilde{\sigma}_i}{\partial\rho^T}\right)^T\tilde{\Omega}_{is}^{-1}\left(\frac{\partial\tilde{\sigma}_i}{\partial\rho^T}\right)\right]^{-1} \\ & \cdot\left[\sum_{i=1}^{K}\left(\frac{\partial\tilde{\sigma}_i}{\partial\rho^T}\right)^T\tilde{\Omega}_{is}^{-1}(s_i-\tilde{\sigma}_i)\right. \\ & \left.\cdot(s_i-\tilde{\sigma}_i)^T\tilde{\Omega}_{is}^{-1}\left(\frac{\partial\tilde{\sigma}_i}{\partial\rho^T}\right)^T\right] \\ & \cdot\left[\sum_{i=1}^{K}\left(\frac{\partial\tilde{\sigma}_i}{\partial\rho^T}\right)^T\tilde{\Omega}_{is}^{-1}\left(\frac{\partial\tilde{\sigma}_i}{\partial\rho^T}\right)\right]^{-1},\end{aligned} \tag{3.9}$$

evaluated at $\rho = \hat{\rho}_{\mathrm{GGEE2}}$ and $\beta = \hat{\beta}_{\mathrm{GGEE2}}$.

3.3 GEGEE Approach for Regression and Longitudinal Correlation Parameters

One of the disadvantages of the GGEE2 approach discussed in the previous section is that it requires the formulas for the fourth-order moments to construct the weight matrix $\tilde{\Omega}_{is}$ in (3.6), which are not possible to compute exactly even if the true correlation structure is known. Note that the extended generalized estimating equations (2.3) and (2.10) (Hall and Severini, 1998) for the regression and the correlation parameters, in contrast, avoid the computations of the third- and fourth-order moments. But, as the estimating equation (2.10) is constructed to estimate the so-called "working" correlation parameters, this EGEE approach consequently suffers from various pitfalls (Sutradhar and Kumar, 2001) similar to those of the GEE2 approach. As a remedy, one may still follow the EGEE approach but estimate the regression parameter β by exploiting the true correlation structure. This means that the estimating equation for the regression parameter vector β will be a function of the true correlation parameters instead of the so-called "working" correlation parameters. Consequently, one is required to construct the estimating equations for the true correlation parameters following the same technique as used by Hall and Severini (1998) for the "working" correlation parameters.

To be specific, similar to the GQL and GGEE2 approaches, for known $\rho_1, \ldots, \rho_{n-1}$, the estimating equation for β under the present GEGEE approach is given by

$$\sum_{i=1}^{K} X_i^T A_i \Sigma_i^{-1}(\rho_1, \ldots, \rho_{n-1})(y_i - \mu_i) = 0, \tag{3.10}$$

which is the same as the estimating equation (3.3). For the construction of the estimating equations for $\rho = (\rho_1, \ldots, \rho_{n-1})^T$, we, however, follow Hall and Severini (1998) and, for $l = 1, \ldots, n-1$, construct the $W_{il}(\rho_1, \ldots, \rho_{n-1})$ matrix as

$$\begin{aligned} & W_{il}(\rho_1, \ldots, \rho_{n-1}) \\ & \quad = \frac{\partial \Sigma_i^{-1}(\rho_1, \ldots, \rho_{n-1})}{\partial \rho_l} \\ & \quad = -A_i^{-1/2} C^{-1}(\rho_1, \ldots, \rho_{n-1}) \\ & \qquad \cdot \frac{\partial C(\rho_1, \ldots, \rho_{n-1})}{\partial l} \\ & \qquad \cdot C^{-1}(\rho_1, \ldots, \rho_{n-1}) A_i^{-1/2}, \end{aligned} \tag{3.11}$$

where, unlike the "working" correlation approach, the form of $\partial C(\rho_1, \ldots, \rho_{n-1})/\partial\rho_l$ is completely specified. For example, as the autocorrelation matrix $C(\rho_1, \ldots, \rho_{n-1})$ has ρ_1 only in the first upper and lower diagonals, for $l = 1$, $\partial C(\rho_1, \ldots, \rho_{n-1})/\partial\rho_l$ has the specific form given by

$$\frac{\partial C(\rho_1, \ldots, \rho_{n-1})}{\partial \rho_1} = \begin{bmatrix} 0\,1\,0 \cdots 0 \\ 1\,0\,1 \cdots 0 \\ 0\,1\,0 \cdots 0 \\ \vdots\,\vdots\,\vdots \quad \vdots \\ 0\,0\,0 \cdots 0 \end{bmatrix}. \tag{3.12}$$

Similarly, we can compute the other derivatives with respect to $\rho_2, \ldots, \rho_{n-1}$. Now, following (2.10), one may write the estimating equations for $\rho_1, \ldots, \rho_{n-1}$ as

$$\begin{aligned} & K^{-1}\sum_{i=1}^{K}\left[W_{id}^{*T}, W_{i\bar{d}}^{*T}\right] \\ & \qquad \cdot\left[(u_i - v_i)^T, \left(s_i - \tilde{\sigma}_i(\rho_1, \ldots, \rho_{n-1})\right)^T\right] \\ & \quad = 0, \end{aligned} \tag{3.13}$$

where W_{id}^{*T} and $W_{i\bar{d}}^{*T}$ are the $n \times (n-1)$ and $\{n(n-1)/2\} \times (n-1)$ matrices such that the lth, $l = 1, \ldots, n-1$, column of the W_{id}^{*T} matrix consists of the

diagonal elements of the $W_{il}(\rho_1, \ldots, \rho_{n-1})$ matrix in (3.11), and similarly the lth, $l = 1, \ldots, n-1$, column of the W_{id}^{*T} matrix consists of the distinct off-diagonal elements of the $W_{il}(\rho_1, \ldots, \rho_{n-1})$ matrix. In (3.13), $\tilde{\sigma}_1(\rho_1, \ldots, \rho_{n-1})$ is the $\{n(n-1)/2\} \times 1$ vector defined as in (3.7). Next, the regression vector β and the true correlation parameters $\rho_1, \ldots, \rho_{n-1}$ are estimated by solving the estimating equations (3.10) and (3.13) simultaneously. Let $\hat{\beta}_{\text{GEGEE}}$ and $\hat{\rho}_{\text{GEGEE}} = (\hat{\rho}_{1,\text{GEGEE}}, \ldots, \hat{\rho}_{n-1,\text{GEGEE}})^T$ denote the GEGEE estimator of β and $\rho = (\rho_1, \ldots, \rho_{n-1})^T$, respectively.

Furthermore, it may be shown that, under some mild conditions, $\hat{\xi}_{\text{GEGEE}} = (\hat{\beta}_{\text{GEGEE}}^T, \hat{\rho}_{\text{GEGEE}}^T)^T$ has the asymptotic covariance matrix given by

$$\begin{aligned} \operatorname{cov}(\hat{\xi}_{\text{GEGEE}}) = \lim_{K\to\infty} &\left[\sum_{i=1}^{K}\begin{pmatrix} A_{i11}^* & A_{i12}^* \\ A_{i21}^* & A_{i22}^* \end{pmatrix}\right]^{-1} \\ &\cdot \left[\sum_{i=1}^{K}\begin{pmatrix} M_{i11}^* & M_{i12}^* \\ M_{i21}^* & M_{i22}^* \end{pmatrix}\right] \\ &\cdot \left[\sum_{i=1}^{K}\begin{pmatrix} A_{i11}^* & A_{i12}^* \\ A_{i21}^* & A_{i22}^* \end{pmatrix}\right]^{-1}, \end{aligned} \tag{3.14}$$

where $A_{i11}^* = -\tilde{D}_i^T \Sigma_i^{-1}(\rho_1, \ldots, \rho_{n-1})\tilde{D}_i$, $A_{i12}^* = 0$, $A_{i21}^* = -W_i^{*T}\lambda_{i\beta}^{*\prime}$ and $A_{i22} = -W_i^{*T}\lambda_{i\rho}^{*\prime}$, with $W_i^* = [W_{id}^{*T}, W_{i\bar{d}}^{*T}]^T$, $\lambda_i^* = (\nu_i^T, \tilde{\sigma}_i^{-1}(\rho_1, \ldots, \rho_{n-1}))^T$, $\lambda_{i\beta}^{*\prime} = \partial\lambda_i^*/\partial\beta^T$ and $\lambda_{i\rho}^{*\prime} = [\partial\lambda_i^*/\partial\rho_1, \ldots, \partial\lambda_i^*/\partial\rho_{n-2}]$, and where

$$\begin{aligned} M_{i11}^* &= \tilde{D}_i^T \Sigma_i^{-1}(\rho_1, \ldots, \rho_{n-1})\tilde{D}_i, \\ M_{i12}^* &= \tilde{D}_i^T \Sigma_i^{-1}(\rho_1, \ldots, \rho_{n-1}) \operatorname{cov}(Y_i, F_i) W_i^*, \\ M_{i21}^* &= M_{i12}^{*T}, \quad M_{i22} = W_i^{*T} \operatorname{var}(F_i) W_i^*, \end{aligned}$$

with $f_i = (u_i^T, s_i^T)^T$ as in (2.12). The covariance matrices, $\operatorname{cov}(Y_i, F_i)$ and $\operatorname{var}(F_i)$, are computed by using a pseudo-normal distribution for y_i with mean μ_i and covariance matrix $\Sigma_i\ (\rho_1, \ldots, \rho_{n-1})$.

In the following section, the performance of the GQL, GGEE2 and GEGEE approaches will be examined through a simulation study for the estimation of both the regression and the correlation parameters. More specifically, the purpose of the simulation study will be to examine the performance of the regression and correlation estimators, as well as the performance of the estimators of standard errors of the regression estimates computed from (3.5) by using (3.4) for the GQL approach, from (3.5) by using (3.6) for the GGEE2 approach and from (3.14) for the GEGEE approach.

4. A SIMULATION STUDY

To compare the performance of the GQL, GGEE2 and GEGEE approaches through a simulation study, we generate correlated Poisson data following three widely used AR(1), MA(1) and exchangeable autocorrelation structures. For convenience, we describe these three probability models in brief as follows.

Poisson AR(1) *probability model.* Let the response y_{it} at time t be related to $y_{i,t-1}$ at time $t-1$ as

$$y_{it} = \rho * y_{i,t-1} + d_{it} \tag{4.1}$$

(McKenzie, 1988), where $y_{i,t-1}$ has the Poisson distribution with parameter $\mu_{i\cdot} = \exp(x_{i\cdot}^T\beta)$ with $x_{i\cdot} = x_{it}$ for all $t = 1, \ldots, n$ (i.e., covariates are not time dependent). Let $y_{i,t-1} \sim P(\mu_{i\cdot})$ denote this Poisson distribution. In (4.1), ρ is a constant scale parameter satisfying the range restriction $0 \le \rho \le 1$. Further, for given $y_{i,t-1}$, $\rho * y_{i,t-1}$ in (4.1) is computed through a binomial thinning operation (McKenzie, 1988). To be specific, $\rho * y_{i,t-1}$ is the sum of $y_{i,t-1}$ binary observations, where each observation is generated with probability ρ. In notation,

$$\begin{aligned} \rho * y_{i,t-1} &= \sum_{j=1}^{y_{i,t-1}} b_j(\rho) \\ &= z_{i,t-1}, \quad \text{say,} \end{aligned} \tag{4.2}$$

with $\Pr[b_j(\rho) = 1] = \rho$ and $\Pr[b_j(\rho) = 0] = 1 - \rho$. It then follows that, conditional on $y_{i,t-1}$, $z_{i,t-1}$ has the binomial distribution. Denote this binomial distribution by $B(y_{i,t-1}, \rho)$. Next, by assuming that $d_{it} \sim P(\mu_{i\cdot}(1-\rho))$ and is independent of $z_{i,t-1}$, it may be shown that $y_{it} \sim P(\mu_{i\cdot})$. It also follows that $E(y_{it}, y_{i,t-l}) = \mu_{i\cdot}^2 + \mu_{i\cdot}\rho^l$, yielding the lag l correlation between y_{it} and $y_{i,t-l}$ as ρ^l, which is the same as the lag l correlation under the Gaussian AR(1) autocorrelation structure. Note, however, that ρ in (4.1) satisfies the range restriction $0 \le \rho \le 1$, whereas in the Gaussian AR(1) structure ρ lies in the range $-1 < \rho < 1$.

Poisson MA(1) *probability model.* In this process, the response y_{it} is related to the d_{it}'s as

$$y_{it} = \rho * d_{i,t-1} + d_{it}, \tag{4.3}$$

where $d_{it} \overset{\text{i.i.d.}}{\sim} P(\mu_{i\cdot}/(1+\rho))$ for all $t = 1, \ldots, n$. By similar calculations as in the AR(1) process, one obtains

$$\operatorname{corr}(y_{it}, y_{i,t-l}) = \begin{cases} \rho/(1+\rho), & \text{for } l = 1, \\ 0, & \text{otherwise,} \end{cases} \tag{4.4}$$

which has the same form as in the Gaussian MA(1) correlation structure, except that in the present setup $0 \le \rho \le 1$, whereas under the Gaussian structure $-1 < \rho < 1$.

Poisson equicorrelation probability model. Suppose that y_{i0} is a Poisson variable with mean parameter μ_i.. Also suppose that $d_{it} \overset{\text{i.i.d.}}{\sim} P(\mu_i.(1-\rho))$ for all $t = 1, \ldots, n$. By similar arguments as for the AR(1) and MA(1) processes, one can show that y_{it} given by

$$y_{it} = \rho * y_{i0} + d_{it} \tag{4.5}$$

also follows the Poisson distribution, that is, $y_{it} \sim P(\mu_i.)$. Further, it can be shown that

$$\operatorname{corr}(y_{it}, y_{i,t-k}) = \rho \tag{4.6}$$

for all $l = 1, 2, \ldots$, with $0 \le \rho \le 1$ instead of $-1/(n-1) \le \rho \le 1$ under the Gaussian equicorrelation model.

In the simulation, we consider large values of $\rho = 0.6$ and 0.8 under the AR(1) process, $\rho = 0.2$ and 0.4 under the MA(1) process and $\rho = 0.6$ and 0.8 for the equicorrelation process. Irrespective of the correlation processes, we consider $p = 2$ with $\beta_1 = \beta_2 = 0$. Further, we consider $K = 100$ clusters each with $n = 4$ repeated Poisson observations generated following each of the above three correlation processes. As far as the selection of covariates is concerned, we consider two design matrices. The two covariates under the first design (D_1) were chosen as

$$x_{ij1} = \begin{cases} -1, & \text{for } j = 1, \ldots, n; \\ & i = 1, \ldots, K/4, \\ 0, & \text{for } j = 1, \ldots, n; \\ & i = (K/4)+1, \ldots, K/2, \\ 0, & \text{for } j = 1, \ldots, n; \\ & i = (K/2)+1, \ldots, 3K/4, \\ 1, & \text{for } j = 1, \ldots, n; \\ & i = (3K/4)+1, \ldots, K \end{cases}$$

and

$$x_{ij2} = z_i^* \quad \text{for } j = 1, \ldots, n;\ i = 1, \ldots, K,$$

where z_i^* is a standard normal value. Under the second design (D_2), the second covariate was chosen to be the same as in the first design D_1, but the first covariate was chosen to be cluster as well as time dependent. More specifically,

$$x_{ij1} = \begin{cases} -j/n, & \text{for } j = 1, \ldots, n; \\ & i = 1, \ldots, K/4, \\ j, & \text{for } j = 1, \ldots, n; \\ & i = (K/4)+1, \ldots, K/2, \\ j-(n+1)/2, & \text{for } j = 1, \ldots, n; \\ & i = (K/2)+1, \ldots, 3K/4, \\ j/n, & \text{for } j = 1, \ldots, n; \\ & i = (3K/4)+1, \ldots, K. \end{cases}$$

Based on the above designs, we then generated four correlated Poisson observations under the ith ($i = 1, \ldots, K$) cluster, following (4.1) for the AR(1) process, (4.3) for the MA(1) process, and (4.5) for the equicorrelation process.

Note that as we have generated a proper discrete correlated data set under each of the three correlation processes, we may now apply the GQL, GGEE2 and GEGEE estimation approaches discussed in Section 3 to examine their performances in estimating ρ_1, ρ_2, ρ_3 and β_1, β_2. Further, note that although we have generated the count data under AR(1), MA(1) and equicorrelation structures, the correlation model is, however, not known in practice except that we may use the form of the correlation structure given by (3.2) for the purpose of estimation. We have used 500 simulations and it was found that the GEGEE approach of Hall and Severini (1998), in general, has serious convergence problems for the estimation of ρ_1, ρ_2 and ρ_3, yielding inconsistent estimates for β_1 and β_2. Consequently, we do not report the simulation results for this approach, but explain the behavior of the GQL approach of Sutradhar and Das (1999) and the GGEE2 approach of Zhao and Prentice (1990). Note that the GQL approach uses the sample autocorrelation formula (3.4) to estimate its population counterpart, whereas the GGEE2 approach uses the estimating equation (3.6) to estimate ρ_l, $l = 1, 2, 3$, which requires the construction of a normality based "working" fourth-order covariance matrix with general elements given by

$$\begin{aligned} &\operatorname{cov}\big((Y_{it} - \mu_{it})(Y_{it'} - \mu_{it'}), (Y_{il} - \mu_{il})(Y_{im} - \mu_{im})\big) \\ &\quad = \rho_{|t-l|}\rho_{|t'-m|}\{a''(\theta_{it})a''(\theta_{il})\}^{1/2} \\ &\qquad \cdot \{a''(\theta_{it'})a''(\theta_{im})\}^{1/2} \\ &\qquad + \rho_{|t-m|}\rho_{|t'-l|}\{a''(\theta_{it})a''(\theta_{im})\}^{1/2} \\ &\qquad \cdot \{a''(\theta_{it'})a''(\theta_{il})\}^{1/2}, \end{aligned}$$

where, for example, $a''(\theta_{it}) = \mu_{it} = \exp(x_{it}'\beta)$ under the present Poisson model. The GQL and GGEE2 approaches, however, use the same estimating equation (3.3) to obtain the regression estimates. Thus, the regression estimates under these two approaches would be different only because of different correlation estimates used in (3.3) under the two approaches. The simulated mean (SM) and simulated standard error (SSE) are computed for each of the two regression estimates, as well as for the estimates of all three lag correlations. The estimated standard errors of the regression estimates are also computed by using the estimate of the covariance matrix of regression estimates given

TABLE 1

Simulated means (SM), simulated standard errors (SSE) and estimated standard errors (ESE) of the GQL *and* GGEE2 *estimates for regression coefficients and autocorrelation for selected values of the true correlation parameter for the Poisson* AR(1) *process with* $n = 4$, $K = 100$, $\beta_1 = \beta_2 = 0$, *based on* 500 *simulations*

Design	Method	AR(1) correlation parameter (ρ)	Number of convergent simulations	Statistic	Estimate $\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\rho}_1$	$\hat{\rho}_2$	$\hat{\rho}_3$
D_1	GQL	0.6	500	SM	−0.003	−0.001	0.595	0.352	0.203
				SSE	0.085	0.049	0.061	0.088	0.108
				ESE	0.086	0.050			
			486	SM	−0.002	0.000	0.584	0.348	0.205
				SSE	0.087	0.050	0.060	0.085	0.109
				ESE	0.085	0.049			
		0.8	500	SM	0.000	0.003	0.791	0.626	0.496
				SSE	0.096	0.056	0.043	0.070	0.098
				ESE	0.098	0.057			
			425	SM	0.000	0.000	0.784	0.616	0.483
				SSE	0.100	0.056	0.041	0.067	0.093
				ESE	0.098	0.056			
	GGEE2	0.6	486	SM	−0.004	0.000	0.592	0.348	0.199
				SSE	0.085	0.050	0.058	0.085	0.106
				ESE	0.086	0.050			
		0.8	425	SM	0.001	0.000	0.734	0.580	0.459
				SSE	0.100	0.056	0.113	0.110	0.129
				ESE	0.096	0.056			
D_2	GQL	0.6	500	SM	−0.004	−0.004	0.592	0.349	0.208
				SSE	0.044	0.087	0.051	0.078	0.105
				ESE	0.040	0.083			
			495	SM	−0.004	0.000	0.591	0.348	0.208
				SSE	0.044	0.087	0.050	0.077	0.105
				ESE	0.040	0.083			
		0.8	500	SM	−0.004	−0.005	0.794	0.636	0.510
				SSE	0.033	0.089	0.036	0.055	0.082
				ESE	0.037	0.096			
			435	SM	−0.004	−0.002	0.788	0.630	0.503
				SSE	0.029	0.090	0.032	0.051	0.079
				ESE	0.037	0.096			
	GGEE2	0.6	495	SM	−0.005	0.000	0.587	0.351	0.214
				SSE	0.046	0.087	0.107	0.105	0.129
				ESE	0.040	0.083			
		0.8	435	SM	−0.005	−0.002	0.737	0.594	0.482
				SSE	0.032	0.092	0.108	0.108	0.124
				ESE	0.037	0.094			

in (3.5). The simulation results based on 500 simulations are reported in Tables 1–3 for the AR(1), MA(1) and equicorrelation processes, respectively.

The results in Tables 1–3 show that while there is no convergence problem in the GQL approach to estimate ρ_l by $\hat{\rho}_l$ (3.4), the estimating equations (3.6) of the GGEE2 approach, however, yielded $\hat{\rho}_l$ greater than 1 in many simulations under the AR(1) and equicorrelation processes. The problem becomes serious for larger ρ, ρ being the correlation parameter of a given process. For example, for $\rho = 0.8$, the GGEE2 approach yielded correlation estimates within the permissible range only in 425 out of 500 simulations. For the GGEE2 approach, the selection of the design matrix does not appear to have a significant effect on the convergence for correlation estimates. For example, for the equicorrelation process with $\rho = 0.8$, it is clear from Table 3 that the convergence was achieved in 471 simulations with design D_1, and in 465 simulations with design D_2, showing a slight change only. The number of conver-

TABLE 2

Simulated means (SM), simulated standard errors (SSE) and estimated standard errors (ESE) of the GQL *and* GGEE2 *estimates for regression coefficients and autocorrelation for selected values of the true correlation parameter for the Poisson* MA(1) *process with* $n = 4$, $K = 100$, $\beta_1 = \beta_2 = 0$, *based on* 500 *simulations*

Design	Method	MA(1) correlation parameter (ρ)	Number of convergent simulations	Estimate					
				Statistic	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\rho}_1$	$\hat{\rho}_2$	$\hat{\rho}_3$
D_1	GQL	0.2	500	SM	0.002	0.002	0.191	−0.006	0.004
				SSE	0.083	0.063	0.058	0.073	0.100
				ESE	0.081	0.063			
		0.4	500	SM	−0.004	−0.004	0.396	−0.005	−0.004
				SSE	0.085	0.069	0.059	0.074	0.097
				ESE	0.088	0.070			
	GGEE2	0.2	500	SM	0.002	0.002	0.184	−0.006	0.005
				SSE	0.083	0.063	0.063	0.070	0.098
				ESE	0.080	0.063			
		0.4	500	SM	−0.004	−0.004	0.381	−0.002	−0.004
				SSE	0.085	0.069	0.075	0.074	0.107
				ESE	0.088	0.069			
D_2	GQL	0.2	500	SM	0.000	0.000	0.192	−0.004	0.007
				SSE	0.038	0.062	0.058	0.073	0.100
				ESE	0.035	0.063			
		0.4	500	SM	0.000	−0.001	0.397	−0.004	−0.001
				SSE	0.038	0.068	0.050	0.075	0.098
				ESE	0.038	0.069			
	GGEE2	0.2	500	SM	0.000	0.000	0.185	−0.004	0.007
				SSE	0.038	0.062	0.063	0.071	0.098
				ESE	0.035	0.062			
		0.4	500	SM	0.000	0.000	0.382	−0.001	−0.001
				SSE	0.038	0.068	0.073	0.074	0.108
				ESE	0.037	0.068			

gent simulations mainly for the GGEE2 approach are shown in column 4 in each of the three tables, for different ρ values. Note that, for the GQL process, we have reported two different simulation results under the AR(1) and equicorrelation processes. First, the simulation results based on all 500 convergent simulations are reported. Next, we have also reported the simulation results for this GQL approach based on those simulations which yielded correlation estimates under the GGEE2 approach.

Note that the GQL approach performs almost the same even if the simulation estimates are obtained based on fewer simulations. For example, for the AR(1) process with $\rho = 0.8$ under D_1, the GQL approach produces lag correlation estimates of 0.791, 0.626 and 0.496 based on all 500 simulations, whereas these correlation estimates are 0.784, 0.616 and 0.483 based on 425 simulations. When these estimates are compared with the GGEE2 of estimates 0.734, 0.580 and 0.459, it is clear that the GQL approach produces less biased estimates for $\rho_l = \rho^l$ with $\rho = 0.8$. The GQL approach continues to perform better than the GGEE2 approach in estimating lag correlations under the MA(1) and equicorrelation models, too. The simulated standard errors (SSEs) of the correlation estimates are, in general, larger for the GGEE2 approach as compared to the GQL approach. For example, under the equicorrelation process with design D_2 and $\rho = 0.8$, the GQL approach produces SMs of 0.793, 0.793 and 0.794 and corresponding SSEs of 0.041, 0.039 and 0.058 for the three lag correlation estimates, whereas the GGEE2 approach produces SMs of 0.734, 0.733 and 0.732 and corresponding standard errors of 0.115, 0.116 and 0.116. This leads to relative mean squared error efficiencies of 16%, 15.46% and 16.41% for the GGEE2 approach as compared to the GQL approach.

For the estimation of the regression parameters, the GQL and GGEE2 approaches appear to produce unbiased estimates, the true regression parameters

TABLE 3

Simulated means (SM), simulated standard errors (SSE) and estimated standard errors (ESE) of the GQL *and* GGEE2 *estimates for regression coefficients and autocorrelation for selected values of the true correlation parameter for the Poisson equicorrelation process with* $n = 4$, $K = 100$, $\beta_1 = \beta_2 = 0$, *based on* 500 *simulations*

Design	Method	Equicorrelation parameter (ρ)	Number of convergent simulations	Estimate					
				Statistic	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\rho}_1$	$\hat{\rho}_2$	$\hat{\rho}_3$
D_1	GQL	0.6	500	SM	−0.006	−0.005	0.587	0.587	0.587
				SSE	0.119	0.096	0.064	0.065	0.088
				ESE	0.118	0.093			
			497	SM	−0.005	−0.002	0.586	0.586	0.586
				SSE	0.119	0.096	0.064	0.065	0.087
				ESE	0.116	0.093			
		0.8	500	SM	−0.009	−0.009	0.790	0.790	0.789
				SSE	0.131	0.101	0.043	0.041	0.059
				ESE	0.130	0.103			
			471	SM	−0.008	0.002	0.787	0.786	0.786
				SSE	0.132	0.101	0.041	0.039	0.058
				ESE	0.130	0.103			
	GGEE2	0.6	497	SM	−0.005	−0.002	0.562	0.561	0.560
				SSE	0.119	0.096	0.120	0.122	0.131
				ESE	0.115	0.091			
		0.8	471	SM	−0.008	0.002	0.728	0.727	0.725
				SSE	0.131	0.101	0.117	0.118	0.117
				ESE	0.126	0.099			
D_2	GQL	0.6	500	SM	−0.002	−0.002	0.591	0.591	0.592
				SSE	0.033	0.094	0.064	0.065	0.088
				ESE	0.033	0.092			
			496	SM	−0.002	−0.004	0.590	0.590	0.590
				SSE	0.033	0.095	0.064	0.065	0.087
				ESE	0.033	0.092			
		0.8	500	SM	−0.002	−0.001	0.793	0.793	0.794
				SSE	0.027	0.097	0.042	0.041	0.058
				ESE	0.026	0.102			
			465	SM	−0.001	0.000	0.789	0.789	0.790
				SSE	0.027	0.098	0.041	0.039	0.058
				ESE	0.026	0.102			
	GGEE2	0.6	496	SM	−0.003	−0.004	0.570	0.569	0.568
				SSE	0.033	0.095	0.122	0.123	0.132
				ESE	0.034	0.090			
		0.8	465	SM	−0.003	0.000	0.734	0.733	0.732
				SSE	0.028	0.099	0.115	0.116	0.116
				ESE	0.029	0.098			

being $\beta_1 = \beta_2 = 0$. This is evident from all three tables, as the simulated means (SMs) appear to take values within the range from −0.009 (for design D_1 under the equicorrelation process with $\rho = 0.8$) to 0.003 [for D_1 under the AR(1) process with $\rho = 0.6$]. Also, the SSEs of the regression estimates appear to be the same under both GQL and GGEE2. For instance, the D_2 based GQL and GGEE2 approaches produce SSE's of 0.027, 0.097 and 0.028 and 0.099 for the estimates of β_1 and β_2, respectively, under the equicorrelation process with $\rho = 0.8$. The corresponding standard errors appear to be very close to each other. Next, the estimates of the standard errors (ESEs) computed by (3.5) appear to perform extremely well under both GQL and GGEE2. This is because all ESEs from the three tables appear to be very close to the corresponding SSEs for the regression estimates. With regard to the design effect on the regression estimation, both GQL and GGEE2 appear to perform the same irrespective of the selection of the design matrix. For a given method, the SSEs

of the regression estimates, however, appear to vary in magnitude from one process to the other. For example, for the AR(1) count model, D_1 based SSEs of the estimates of β_1 are much larger (in the range from 0.085 to 0.100) than D_2 based estimates (in the range from 0.032 to 0.044), whereas the reverse happens for the estimates of β_2. This behavior appears to be true under both GQL and GGEE2. This estimating behavior that both GQL and GGEE2 perform equally in estimating regression effects is not surprising as both approaches use the same GEE for β (3.3) based on their own correlation estimates, which are consistent for their parameters under both approaches.

In summary, the limited simulation study conducted in the paper indicates the superiority of the GQL approach over the GGEE2 approach in estimating the parameters of the longitudinal models. The specific reasons for this are as follows. First, based on the same number of convergent simulations, the GQL approach produces estimates for the longitudinal correlations with smaller bias as well as smaller standard errors, as compared to the GGEE2 approach. Second, although both of these approaches perform the same in estimating the regression parameters, the GQL approach is relatively much simpler. This is because the estimation of the correlations by (3.4) under the GQL approach is straightforward as compared to the estimation of correlations by (3.6) under the GGEE2 approach. Moreover, if the GGEE2 approach encounters convergence problems in estimating the longitudinal correlation parameters for a data set in practice, the procedure will be subsequently useless for the estimation of the regression parameters. We therefore recommend the use of the proposed general autocorrelation structure based GQL approach in estimating the parameters of the longitudinal models for the discrete data.

To implement the computational formulas for the GQL estimates of the regression and longitudinal correlation parameters, one may follow the two-step procedure given below. First, one solves the estimating equation for β (3.3) iteratively, using starting values 0 for the longitudinal correlations and small positive or negative values for the regression parameters. This interim solution of (3.3) is then used in (3.4) to obtain the estimates of the autocorrelations, which are used in turn in (3.3) to compute new β estimates. This cycle of iterations continues until convergence.

5. APPLICATIONS OF THE GQL APPROACH

Recall that the simulation study in the previous section indicates that among the three new iterative approaches—GQL, GGEE2 and GEGEE—the GEGEE approach has serious convergence problems. To be specific, the GEGEE approach quite often produces estimates for the correlation parameters beyond the permissible range, yielding inconsistent estimates for the regression parameters. The GGEE2 approach also has convergence problems. This is because it was found in the simulation study that, under this approach, the iterations in some simulations did not converge for the estimation of the correlation parameters. The GQL approach, however, never encounters any such convergence problems. Moreover, the GQL approach is the simplest among the three approaches in estimating both the regression and the correlation parameters. Consequently, the GQL approach is recommended in practice in analyzing discrete (such as count and binary) longitudinal data. In view of this recommendation, in this section we illustrate the use of the GQL approach only, first to analyze a Poisson longitudinal data set and then a binary longitudinal data set.

5.1 Example 1: Analyzing Poisson Longitudinal Data

Recall from Section 1 that in a health care utilization data analysis, one may be interested in finding the effects of the related covariates on the physician visits over the years, after taking the longitudinal correlations of the data into account. We now consider a real-life data set on the health care utilization problem collected by the General Hospital of the city of St. John's, Newfoundland, Canada. The data contain the complete records for 144 individuals for four years ($n = 4$) from 1985 to 1988. To be specific, the number of visits to a physician by each individual during a given year was recorded as the response, and this was repeated for four years. Also, the information on four covariates, namely, gender, number of chronic conditions in 1985, education level in 1985 and age, was recorded for each individual. Note that as the responses are counts, it is appropriate to assume that the response variable marginally follows the Poisson distribution, and the repeated counts recorded for four years will be longitudinally correlated. It is of scientific interest to take the longitudinal correlations into account and examine the effects of the above four covariates on the physician visits.

Following the notation used in Section 3, the four covariates for the ith, $i = 1, \ldots, K = 144$, individual at time t, $t = 1, \ldots, 4$, are denoted by $x_{it1}, x_{it2}, x_{it3}$ and x_{it4}. The first covariate, sex, was coded as 0 for female and 1 for male. Thus, at any time t, $x_{it1} = 0$ if

the ith individual is female; otherwise, $x_{it1} = 1$. Similarly, the number of chronic diseases was coded as $x_{it2} = 0$ for the absence of chronic disease for the ith individual in 1985 and $x_{it2} = 1$ if the ith individual had one or more chronic diseases in 1985. The third covariate, education level, x_{it3}, was coded as 1 for less than high school and 0 for high school or more education. The last covariate, x_{it4}, was taken as the deviation age from 50. For example, an individual with age 30 has the covariate value -20 and so on. The effects of these covariates are denoted by $\beta = (\beta_1, \beta_2, \beta_3, \beta_4)'$, so that the mean of the count responses for the ith individual at time t is given by

$$\mu_{it} = \exp(x_{it}'\beta), \tag{5.1}$$

where $x_{it} = (x_{it1}, x_{it2}, x_{it3}, x_{it4})'$. Furthermore, it is assumed that the repeated responses $y_{i1}, \ldots, y_{i4}$ have an autocorrelation structure as in (3.2).

Now, by applying the two-step approach introduced through (3.3) and (3.4), we obtain in five cycles of iterations the GQL based regression estimates as

$$\begin{aligned}
\hat{\beta}_1(\text{effect of sex}) &= -0.200,\\
\hat{\beta}_2(\text{effect of chronic conditions}) &= 0.398,\\
\hat{\beta}_3(\text{effect of education level}) &= -0.116,\\
\hat{\beta}_4(\text{effect of age}) &= 0.027,
\end{aligned}$$

along with the moment estimates for the three autocorrelations as

$$\hat{\rho}_1 = 0.538, \quad \hat{\rho}_2 = 0.488, \quad \hat{\rho}_3 = 0.440.$$

The autocorrelation values appear to be large, indicating high longitudinal correlations. Next, the standard errors of the regression estimates are computed using the formula for the asymptotic covariance of $\hat{\beta}_{\text{GQL}}$ given by (3.5). These standard errors are

$$\begin{aligned}
\text{s.e.}(\hat{\beta}_1) &= 0.066, \quad \text{s.e.}(\hat{\beta}_2) = 0.079,\\
\text{s.e.}(\hat{\beta}_3) &= 0.063, \quad \text{s.e.}(\hat{\beta}_4) = 0.001.
\end{aligned}$$

Note that as the standard errors are quite small as compared to the corresponding values of the regression estimates, all four covariates appear to have significant effects on the physician visits.

We now interpret the effects of the covariates on the physician visits as follows. As the first covariate, sex, was coded as 1 for male and 0 for female, it follows from (5.1) that the negative value of $\hat{\beta}_1 = -0.20$ suggests that females made more visits to the physician as compared to males. The positive values of $\hat{\beta}_2 = 0.398$ and $\hat{\beta}_4 = 0.027$ suggest that individuals having one or more chronic diseases or individuals belonging to the higher age group paid more visits to the physician, as expected. The third covariate, education level, was coded as 1 for less than high school and 0 for high school or more. The effect of the education level on the physician visits, however, appears to be intriguing. The negative value of $\hat{\beta}_3 = -0.117$ shows that highly educated individuals paid more visits as compared to individuals with low education level. One of the reasons for this type of behavior of this covariate may be that individuals with high education level (high school or more) are more concerned about their health as compared to individuals with low education level.

5.2 Example 2: Analyzing Binary Longitudinal Data

In this example, we reanalyze a binary longitudinal data set analyzed earlier by Zeger, Liang and Albert (1988), among others. This data set is a subset of data from the Six Cities study, a longitudinal study of the health effects of air pollution. The data set contains complete records of 537 children from Steubenville, Ohio, each of whom was examined annually at ages 7–10. The repeated response is the wheezing status (1 = yes, 0 = no) of a child on each occasion. Maternal smoking status was considered as a covariate and it was recorded as 1 if the mother smoked regularly and 0 otherwise. It is clear that the responses are binary by nature, and, consequently, one may use a longitudinal binary model to analyze such data. Here, the scientific interest is to examine the effect of smoking by a mother on the wheeze status of her child. Thus, in our notation, for this particular data set, $K = 537$, $n = 4$ and $\beta = (\beta_1, \beta_2)'$, with β_1 as the intercept and β_2 as the effect of smoking by mother on her child's wheezing status. As the responses for each child are repeatedly collected over a period of four years, it is likely that these binary responses will be longitudinally correlated. It is of interest to estimate β after taking the longitudinal correlations into account.

Fitzmaurice and Laird (1993), among others, considered the time (here it is age) as a specific factor and found the regression effects of maternal smoking, age and their interaction on the binary responses. Further, because repeated observations are made on the same individual, the response variables will usually be correlated. Fitzmaurice and Laird modeled these associations in terms of conditional log-odds ratios and, following Zhao and Prentice (1990), applied a likelihood-based method to compute the regression effects. This likelihood approach, however, does not accommodate the autocorrelation structures, such as

AR(1) and MA(1) structures, appropriate for repeated binary data.

To model the data, it is assumed that the four binary responses collected on four occasions are longitudinally correlated with correlation structure given as in (3.2) (Sutradhar and Das, 1999). The mean response function for the binary data may be written as

$$\mu_{it} = \frac{\exp(x_{it}^T\beta)}{1+\exp(x_{it}^T\beta)} \tag{5.2}$$

for the ith, $i = 1, \ldots, 537$, child at time t, $t = 1, \ldots, 4$. Now, by applying the GQL approach discussed in Section 3, we obtain the GQL estimate of β, in four iterations, as

$$\hat{\beta}_1(\text{intercept}) = -1.820,$$
$$\hat{\beta}_2(\text{maternal smoking effect}) = 0.263$$

along with the estimates of the standard errors (3.5) of the regression estimates given by

$$\text{s.e.}(\hat{\beta}_1) = 0.111, \quad \text{s.e.}(\hat{\beta}_2) = 0.177.$$

The three longitudinal correlation estimates are

$$\hat{\rho}_1 = 0.397, \quad \hat{\rho}_2 = 0.310, \quad \hat{\rho}_3 = 0.297.$$

These correlation values appear to be moderately large. Thus, ignoring these correlations will result in inefficient regression estimates. Note that as the values of $\hat{\rho}_2$ and $\hat{\rho}_3$ are almost the same, the wheezing status does not appear to change within a short span of time such as in three or four years.

With regard to the effect of the main covariate x_{it2} (i.e., maternal smoking), as $x_{it2} = 1$ for the ith child whose mother is a regular smoker, the large positive value of $\hat{\beta}_2 = 0.263$, by (5.2), indicates an increase in the rate of wheeze for children of mothers who smoke.

6. CONCLUDING REMARKS

In longitudinal data analysis, estimating the effect of covariates on a response variable is often of interest, while longitudinal correlations are typically considered nuisance parameters. As the so-called "working" correlation based generalized estimating equation (GEE) approach may not yield efficient regression estimates as compared to the "working" independence assumption based estimating equation (IEE) approach (cf. Sutradhar and Das, 1999), in this paper we have discussed the generalized quasi-likelihood (GQL) approach which produces consistent as well as more efficient regression estimates as compared to the IEE based estimates. This GQL approach assumes a known longitudinal correlation structure, the correlation parameters being unknown. When the correlation structure is known, there also exist GEE2 and EGEE approaches where both the regression and the correlation parameters are jointly estimated. One of the main shortcomings of these two approaches is that there is no unique way to specify the correlation structure. Consequently, some authors such as Prentice and Zhao (1991) use an "exponential family quadratic model" to define the correlation structure in their GEE2 approach. In the EGEE approach, Hall and Severini (1998) use a "working" correlation model which has pitfalls similar to the GEE approach. In this paper, we have used a general autocorrelation structure for the longitudinal correlations as in Sutradhar and Das (1999) and reviewed the feasible features and drawbacks of the GGEE2 and GEGEE approaches. Although, unlike GEGEE, the GGEE2 approach requires fourth-order moments, the simulation study indicated that the GGEE2 approach is better than the GEGEE approach in estimating the autocorrelation structure based correlation parameters as well as the regression effects. Next, the simulation study also indicated that although the GQL approach uses a moment method to compute the correlations, it is a better approach than the GGEE2 approach in estimating the correlation parameters of the model. One of the main reasons for the poor performance of the GGEE2 approach is that it uses the normality based "working" fourth-order moments even though the data are binary or counts by nature. In fact, in some cases, the GGEE2 approach may yield highly unstable estimates for the correlations causing convergence problems in the estimation of the regression parameters. This was evident in the simulation studies, where the GGEE2 approach failed to converge in a large number of simulations. The GQL approach, however, does not encounter any such convergence problems. Moreover, the GQL approach is much simpler as compared to the GGEE2 and GEGEE approaches for the estimation of both the regression and the longitudinal correlations. Consequently, the GQL approach is recommended in practice to analyze discrete such as (binary and count) longitudinal data. The GQL approach was illustrated in the paper by analyzing two sets of real-life data.

In the present paper, we have reviewed the regression analyses of cluster data with cluster level covariates as in the papers by Liang and Zeger (1986), Fitzmaurice, Laird and Rotnitzky (1993), Hall and Severini (1998) and Sutradhar and Das (1999), for example. Some authors have studied the efficiency aspects of the GEE

approach dealing with longitudinal data with within-cluster covariates. We refer to Fitzmaurice (1995), Mancl and Leroux (1996) and Sutradhar and Das (2000), among others, for such studies. It is shown in these papers that for the models with within-cluster covariates, the generalized GEE approach (based on a suitable correlation matrix) in general has higher efficiencies than the independence based GEE approach. In the spirit of these studies, we have included a within-cluster covariate (the second covariate) as a part of our simulation designs, and examined the performance of the GQL and GGEE2 approaches in estimating the effect of such within-cluster covariates. The results of the simulation study suggest that these approaches have performed almost the same in estimating the effect of the within-cluster covariate.

ACKNOWLEDGMENTS

The author highly appreciates the helpful comments and suggestions made by the referees, the Editor and the Executive Editor. This research was supported in part by a grant from the Natural Sciences and Engineering Research Council of Canada. This work has benefitted from the author's other joint work with Vandna Jowaheer in the area of longitudinal count data analysis.

REFERENCES

CROWDER, M. (1995). On the use of a working correlation matrix in using generalised linear models for repeated measures. *Biometrika* **82** 407–410.

FITZMAURICE, G. M. (1995). A caveat concerning independence estimating equations with multivariate binary data. *Biometrics* **51** 309–317.

FITZMAURICE, G. M. and LAIRD, N. M. (1993). A likelihood-based method for analysing longitudinal binary responses. *Biometrika* **80** 141–151.

FITZMAURICE, G. M., LAIRD, N. M. and ROTNITZKY, A. G. (1993). Regression models for discrete longitudinal responses (with discussion). *Statist. Sci.* **8** 284–309.

HALL, D. B. and SEVERINI, T. A. (1998). Extended generalized estimating equations for clustered data. *J. Amer. Statist. Assoc.* **93** 1365–1375.

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

LIPSITZ, S. R., FITZMAURICE, G. M., ORAV, E. J. and LAIRD, N. M. (1994). Performance of generalized estimating equations in practical situations. *Biometrics* **50** 270–278.

MANCL, L. A. and LEROUX, B. G. (1996). Efficiency of regression estimates for clustered data. *Biometrics* **52** 500–511.

MCCULLAGH, P. (1983). Quasilikelihood functions. *Ann. Statist.* **11** 59–67.

MCKENZIE, E. (1988). Some ARMA models for dependent sequences of Poisson counts. *Adv. in Appl. Probab.* **20** 822–835.

NEUHAUS, J. M. (1993). Estimation efficiency and tests of covariate effects with clustered binary data. *Biometrics* **49** 989–996.

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.

SUTRADHAR, B. C. and DAS, K. (1999). On the efficiency of regression estimators in generalised linear models for longitudinal data. *Biometrika* **86** 459–465.

SUTRADHAR, B. C. and DAS, K. (2000). On the accuracy of efficiency of estimating equation approach. *Biometrics* **56** 622–625.

SUTRADHAR, B. C. and KOVACEVIC, M. (2000). Analysing ordinal longitudinal survey data: Generalized estimating equations approach. *Biometrika* **87** 837–848.

SUTRADHAR, B. C. and KUMAR, P. (2001). On the efficiency of extended generalized estimating equation approaches. *Statist. Probab. Lett.* **55** 53–61.

ZEGER, S. L., LIANG, K.-Y. and ALBERT, P. S. (1988). Models for longitudinal data: A generalized estimating equation approach. *Biometrics* **44** 1049–1060.

ZHAO, L. P. and PRENTICE, R. L. (1990). Correlated binary regression using a quadratic exponential model. *Biometrika* **77** 642–648.

ZHAO, L. P., PRENTICE, R. L. and SELF, S. G. (1992). Multivariate mean parameter estimation by using a partly exponential model. *J. Roy. Statist. Soc. Ser. B* **54** 805–811.