Consistent Estimation in Incomplete Longitudinal Binary Models

Taslim S. Mallick, Patrick J. Farrell, and Brajendra C. Sutradhar

Abstract It is well known that in the complete longitudinal setup, the so-called working correlation-based generalized estimating equations (GEE) approach may yield less efficient regression estimates as compared to the independence assumption-based method of moments and quasi-likelihood (QL) estimates. In the incomplete longitudinal setup, there exist some studies indicating that the use of the same "working" correlation-based GEE approach may provide inconsistent regression estimates especially when the longitudinal responses are at risk of being missing at random (MAR). In this paper, we revisit this inconsistency issue under a longitudinal binary model and empirically examine the relative performance of the existing weighted (by inverse probability weights for the missing indicator) GEE (WGEE), a fully standardized GQL (FSGQL) and conditional GQL (CGQL) approaches. In the comparative study, we consider both stationary and non-stationary covariates, as well as various degrees of missingness and longitudinal correlation in the data.

1 Introduction

Consider a longitudinal binary data setup where y_{it} is the Bernoulli response for the i-th ($i = 1, \cdots, K$) individual at the t-th time point ($t = 1, \cdots, T$) and $x_{it} = (x_{it1}, \cdots, x_{itu}, \cdots, x_{itp})'$ is the associated p-dimensional covariate vector. When the longitudinal data are complete (that is, there are no missing responses from any of the individuals in the study), an estimating approach such as generalized quasi-

T.S. Mallick • P.J. Farrell (✉)
School of Mathematics and Statistics, Carleton University, Ottawa, ON, Canada
e-mail: pfarrell@math.carleton.ca

B.C. Sutradhar
Memorial University, St. John's, NF, Canada, A1C5S7
e-mail: bsutradh@mun.ca

likelihood (GQL) can be used to obtain an estimate of the regression parameter vector, β, that is both consistent and efficient, provided that the correlation structure associated with the repeated binary responses is known (see Sutradhar 2003). In order to describe the longitudinal correlation in the data, it seems reasonable to assume deterioration in the association between observations on the same individuals that are further apart in time. Thus, to achieve this, we let ρ be a longitudinal correlation parameter and consider a conditional linear binary dynamic (CLBD) model proposed by Zeger et al. (1985) (see also Qaqish 2003), which is given by

$$P(Y_{i1} = 1) = \mu_{i1}, \text{ and}$$

$$P(Y_{it} = 1 \mid y_{i,t-1}) = \mu_{it} + \rho(y_{i,t-1} - \mu_{i,t-1}) = \lambda_{i,t|t-1}(\beta, \rho) = \lambda_{it}, \text{ for } t = 2, \cdots, T \tag{1}$$

with $\mu_{it} = exp(x'_{it}\beta)/[1 + exp(x'_{it}\beta)]$ for $t = 1, \cdots, T$. According to model (1), the marginal means and variances of y_{it} are

$$E(Y_{it}) = \mu_{it} \tag{2}$$

and

$$Var(Y_{it}) = \sigma_{i,tt} = \mu_{it}(1 - \mu_{it}), \tag{3}$$

while the correlations between Y_{it} and $Y_{i,t+l}$ for $l = 1, \cdots, T-1, t = 1, \cdots, T-l$ are given by

$$corr(Y_{it}, Y_{i,t+l}) = \rho^l \left[\frac{\sigma_{i,tt}}{\sigma_{i,t+l,t+l}} \right]^{1/2}. \tag{4}$$

The means, variances, and covariances defined by (2) through (4) are nonstationary, since they are all functions of time-dependent covariates $\{x_{it}\}$. However, if the $\sigma_{i,tt}$ are not extremely different, the correlations given by (4) assume a behavior that is analogous to an autoregressive process of order one, AR(1). Under the present model, the correlation parameter ρ must satisfy the range restriction

$$max\left[-\frac{\mu_{it}}{1-\mu_{i,t-1}}, -\frac{1-\mu_{it}}{\mu_{i,t-1}} \right] \leq \rho \leq min\left[\frac{1-\mu_{it}}{1-\mu_{i,t-1}}, \frac{\mu_{it}}{\mu_{i,t-1}} \right]. \tag{5}$$

Suppose that we let μ_i and $\Sigma_i(\rho)$ represent the mean vector and the covariance matrix of the complete data vector Y_i, where $\mu_i = (\mu_{i1}, \cdots, \mu_{it}, \cdots, \mu_{iT})'$ and $\Sigma_i(\rho) = A_i^{1/2} C_i(\rho) A_i^{1/2}$. Here, $C_i(\rho)$ is the $T \times T$ correlation matrix based on (4), and $A_i = diag(\sigma_{i,11}, \cdots, \sigma_{i,tt}, \cdots, \sigma_{i,TT})$. An estimator for β that is both consistent and highly efficient can be obtained by solving the GQL estimating equation

$$\sum_{i=1}^{K} \frac{\partial \mu_i^{'}}{\partial \beta} [\Sigma_i(\rho)]^{-1} (y_i - \mu_i) = 0, \tag{6}$$

(Sutradhar 2003).

In practice, it is typically the case that some of the responses associated with each of a number of individuals in the study may be missing. To acknowledge this phenomenon during the data collection process, we introduce an indicator variable R_{it}, that takes on a value of one if Y_{it} is observed, and zero otherwise. For purposes of our investigation here, we adopt the not-so-unreasonable assumption that all individuals provide a response at the first time point, so that $R_{i1} = 1$ for all $i = 1, \cdots, K$. We also assume monotonic missingness, suggesting that the R_{it} satisfy the inequality $R_{i1} \geq R_{i2} \geq \cdots \geq R_{it} \geq \cdots \geq R_{iT}$. Thus, if responses are no longer observed for the i-th individual after the j-th time point, for this individual we would have available y_{it} for $t = 1, \cdots, T_i = j$.

Regarding the missing data mechanism, at this time we distinguish between responses that are missing completely at random, MCAR, and those that are missing at random, MAR (see Fitzmaurice et al. 1996; Paik 1997; Rubin 1976). When the responses are MCAR, the indicator variable R_{it} reflecting the presence or absence of Y_{it} does not depend on the previous responses $Y_{i1}, \cdots, Y_{i,t-1}$. In this instance, if we define $R_i = diag(R_{i1}, \cdots, R_{iT})$ and incorporate this matrix into the estimating equation given by (6) to yield

$$\sum_{i=1}^{K} \frac{\partial \mu_i^{'}}{\partial \beta} [\Sigma_i(\rho)]^{-1} R_i (y_i - \mu_i) = 0, \tag{7}$$

it is still possible to obtain an unbiased estimator for β that will be consistent and efficient. Note that $\Sigma_i(\rho)$ is a $T \times T$ matrix with appropriate variance and covariance entries in the first T_i rows and T_i columns and zeroes in the last $T - T_i$ rows and columns. On the other hand, when the missing data mechanism for the responses is assumed to be MAR (implying that R_{it} does depend on the previous responses $Y_{i1}, \cdots, Y_{i,t-1}$), it can be shown that $E[R_{it}(Y_{it} - \mu_{it})] \neq 0$. In this situation, the estimator for β based on (7) will be biased and inconsistent. Upon realizing this to be the case, many studies have attempted to correct for this problem by using a modified inverse probability-weighted distance function

$$w_{it}^{-1} \{H_{i,t-1}(y); \alpha\} [R_{it}(Y_{it} - \mu_{it})], \tag{8}$$

where $H_{i,t-1}(y) \equiv H_{i,t-1} = (Y_{i1}, \cdots, Y_{i,t-1})$, so that the expectation of (8) is zero. Following Robins et al. (1995), for data that are MAR, we can write the probability weight $w_{it} \{H_{i,t-1}(y); \alpha\} = w_{it}$ as a function of past responses as follows. Specifically, imagine that the probability that the i-th individual responds at the j-th time point depends on the past lag q responses, where $q \leq j - 1$. Letting $g_{ij}(y_{i,j-1}, \cdots, y_{i,j-q}; \alpha)$ represent this probability, we can write $g_{ij}(y_{i,j-1}, \cdots, y_{i,j-q}; \alpha) = P(R_{ij} = 1 \mid R_{i1} = 1, \cdots, R_{i,j-1} = 1; y_{i,j-1}, \cdots, y_{i,j-q})$, which can be modeled as

$$g_{ij}(y_{i,j-1},\cdots,y_{i,j-q};\alpha) = \frac{exp(1+\sum_{l=1}^{q}\alpha_l y_{i,j-l})}{1+exp(1+\sum_{l=1}^{q}\alpha_l y_{i,j-l})}, \tag{9}$$

where α_l is a parameter that reflects the dependence of R_{ij} on $y_{i,j-l}$ for all $l = 1,\cdots,q$. Robins et al. (1995) set

$$
\begin{aligned}
w_{it} &= P(R_{it}=1, R_{i,t-1}=1,\cdots,R_{i1}=1 \mid H_{i,t-1}) \\
&= P(R_{it}=1 \mid R_{i,t-1}=\cdots=R_{i1}=1; H_{i,t-1}) \times \\
&\quad P(R_{i,t-1}=1 \mid R_{i,t-2}=\cdots=R_{i1}=1; H_{i,t-2}) \times \\
&\quad \cdots \times P(R_{i2}=1 \mid R_{i1}=1; H_{i1})P(R_{i1}=1) \\
&= \prod_{j=1}^{t} g_{ij}(y_{i,j-1},\cdots,y_{i,j-q};\alpha). \tag{10}
\end{aligned}
$$

Since monotonic missingness is assumed

$$E[R_{it}Y_{it} \mid H_{i,t-1}] = P[R_{i1}=1, R_{i2}=1,\cdots,R_{it}=1; Y_{it}=1 \mid H_{i,t-1}], \tag{11}$$

or, alternatively

$$
\begin{aligned}
E[R_{it}Y_{it} \mid H_{i,t-1}] &= P(R_{i1}=1)P[R_{i2}=1 \mid R_{i1}=1; H_{i1}]\cdots \\
&\quad P[R_{it}=1 \mid R_{i1}=1, R_{i2}=1,\cdots,R_{i,t-1}=1; H_{i,t-1}] \\
&\quad P[Y_{it}=1 \mid H_{i,t-1}]. \tag{12}
\end{aligned}
$$

Using model (1) and $g_{ij}(y_{i,j-1},\cdots,y_{i,j-q};\alpha)$ given in (9), (12) becomes

$$
\begin{aligned}
E[R_{it}Y_{it} \mid H_{i,t-1}] &= \prod_{j=1}^{t} g_{ij}(y_{i,j-1},\cdots,y_{i,j-q};\alpha)\lambda_{it} \\
&= w_{it}\lambda_{it}, \tag{13}
\end{aligned}
$$

which implies that

$$E\left[\frac{R_{it}Y_{it}}{w_{it}} \mid H_{i,t-1}\right] = \lambda_{it}, \tag{14}$$

thus giving

$$E_{H_{i,t-1}}E\left[\frac{R_{it}Y_{it}}{w_{it}} \mid H_{i,t-1}\right] = E_{H_{i,t-1}}[\lambda_{it}] = \mu_{it}. \tag{15}$$

Similarly

$$E_{H_{i,t-1}} E\left[\frac{R_{it}\mu_{it}}{w_{it}} \mid H_{i,t-1} \right] = \mu_{it}, \tag{16}$$

suggesting that combining (15) and (16) yields

$$E_{H_{i,t-1}} E\left[\frac{R_{it}(Y_{it} - \mu_{it})}{w_{it}} \mid H_{i,t-1} \right] = 0. \tag{17}$$

This unconditional unbiasedness property of the weighted distance or estimating function $\left[\frac{R_{it}(Y_{it}-\mu_{it})}{w_{it}} \right]$ motivated many researchers to write a weighted generalized estimating equation (WGEE) and solve it for the β involved in those μ_{it}. The WGEE, first developed by Robins et al. (1995), is reproduced in brief, in Sect. 2.1. Note that to construct the WGEE, Robins et al. (1995) suggested the specification of a user-selected covariance matrix of $\{(Y_{it} - \mu_{it}), t = 1, \ldots, T_i\}$ by pretending that as though the data were complete. Recently, Sutradhar and Mallick (2010) have found that this widely used WGEE approach produces highly biased regression estimates, indicating consistency break down. In this paper, specifically in Sect. 3, we carry out an extensive simulation study considering various degrees of missingness and examine further the inconsistency problem encountered by the WGEE approach.

In Sect. 2.2, we consider a simpler version of a fully standardized GQL (FSGQL) approach discussed by Sutradhar (2013, Sect. 3.2.4) by constructing the weight matrix, that is, unconditional covariance matrix of $\{\left[\frac{R_{it}(Y_{it}-\mu_{it})}{w_{it}} \right], t = 1, \ldots, T_i\}$ using longitudinal independence (*i.e.*, $\rho = 0$). We will refer to this as the FSGQL(I) approach. In the simulation study in Sect. 3, we examine the relative performance of this FSGQL(I) approach with the existing WGEE as well as WGEE(I) (independence assumption-based WGEE) approach.

Further note that if the correlation model for the complete data were known through λ_{it} in (14), one could exploit the conditional distance function $\left[\frac{R_{it}(Y_{it}-\lambda_{it})}{w_{it}} \mid H_{i,t-1} \right]$ to construct a conditional-weighted GQL (CWGQL) estimating equation and solve such an equation to obtain consistent regression estimates. We discuss this approach in Sect. 2.3 and include it in the simulation study in Sect. 3 to examine its performance as compared to the aforementioned approaches.

2 Estimation

2.1 WGEE Approach

Robins et al. (1995, Eq. (10), p. 109) used the result in (17) to propose the WGEE

$$\sum_{i=1}^{K} \frac{\partial \mu_i^{'}}{\partial \beta} \left[V_i(\alpha^*) \right]^{-1} \Delta_i(y_i - \mu_i) = 0, \tag{18}$$

(see also Paik 1997, Eq. (1), p. 1321) where $\Delta_i = diag(\delta_{i1}, \delta_{i2}, \cdots, \delta_{iT})$ with $\delta_{it} = R_{it}/w_{it}\{H_{i,t-1}(y); \alpha\} = R_{it}/w_{it}$. The quantity $V_i(\alpha^*)$ is a working covariance matrix of Y_i (see Liang and Zeger 1986) that is used in an effort to increase the efficiency of the estimates. Of note is the fact while Robins et al. (1995) suggested a WGEE, they did not account for the missingness in the data when specifying $V_i(\alpha^*)$; they simply based their working covariance matrix on the complete data formulae. For this reason, this WGEE approach may be referred to as a partially standardized GEE (PSGEE) approach. See the previous article by Sutradhar (2013) in this chapter for details on the use of PSGEE. Note that a user-selected covariance matrix based on complete data that ignores the missing mechanism leads the WGEE to be unstable, in particular, when the proportion of missing data is high, causing breakdown in estimation, i.e., breakdown in consistency. However, this inconsistency issue has not been adequately addressed in the literature including the studies by Robins et al. (1995), Paik (1997), Rotnitzky et al. (1998) and Birmingham et al. (2003). One of the main reasons is that none of the studies used any stochastic correlation structure in conjunction with the missing mechanism to model the binary data in the incomplete longitudinal setup.

In this paper, in order to investigate the effect on the estimates of the regression parameter vector, we propose to replace the working covariance matrix $V_i(\alpha^*)$ in (18) with a proper unconditional covariance matrix that accommodates the missingness in the data. The proposed approach is presented in the next section.

2.2 FSGQL Approach

The unconditional unbiasedness property in (17), that is,

$$E_{H_{i,t-1}}E\left[\frac{R_{it}(Y_{it} - \mu_{it})}{w_{it}} \mid H_{i,t-1}\right] = E_{H_{i,t-1}}E\left[\delta_{it}(Y_{it} - \mu_{it}) \mid H_{i,t-1}\right] = 0$$

motivates one to develop a FSGQL estimating equation for β, which requires the computation of the unconditional variance of $\delta_{it}(Y_{it} - \mu_{it})$. Thus, for all $t = 1, \ldots, T_i$, we now compute the unconditional covariance matrix, namely

$$\text{cov}[\Delta_i(y_i - \mu_i)] = \Sigma_i^*(\beta, \rho, \alpha), \text{ (say)},$$

by using the formula

$$\Sigma_i^*(\beta, \rho, \gamma) = E_{H_i(y)}[\text{cov}\{\Delta_i(Y_i - \mu_i(\beta))\} \mid H_i(y)]$$
$$+ \text{cov}_{H_i(y)}[E\{\Delta_i(Y_i - \mu_i(\beta))\} \mid H_i(y)],$$

where $H_i(y)$ denotes the history of responses. For computational details under any specified correlation model, we refer to the previous article by Sutradhar

(2013, Sect. 3.2.4). For the binary AR(1) model in (1), the elements of the $T_i \times T_i$ unconditional covariance matrix $\Sigma_i^*(\beta, \rho, \alpha)$ are given by

$$\text{cov}[\delta_{iu}(y_{iu} - \mu_{iu}), \delta_{it}(y_{it} - \mu_{it})]$$

$$\equiv \begin{cases} \sigma_{i,11}^* = \mu_{i1}[1 - \mu_{i1}] \\ \sigma_{i,tt}^* = E_{H_i(y)}[w_{it}^{-1}\{\mu_{it}(1 - \mu_{it}) + \rho(1 - 2\mu_{it})(y_{i,t-1} - \mu_{i,t-1})\}], (\text{for } t = 2, \ldots, T_i) \\ \sigma*_{i,ut} = \rho\rho^{t-1-u}\mu_{iu}(1 - \mu_{iu}), (\text{for } u = 1 < t) \\ \sigma_{i,ut}^* = \rho^2\rho^{t-u}\mu_{i(u-1)}(1 - \mu_{i(u-1)}), (\text{for } 1 < u < t). \end{cases} \tag{19}$$

Note that the formulas in (19) under the present AR(1) binary model may be verified directly. For example, we compute the t-th diagonal element of the $\Sigma_i^*(\beta, \rho, \alpha)$ matrix as follows. Since $\delta_{it} = R_{it}/w_{it}\{H_{i,t-1}(y); \alpha\} = R_{it}/w_{it}$, we can write

$$Var\left[\frac{R_{it}(Y_{it} - \mu_{it})}{w_{it}}\right] = Var_{H_{i,t-1}}E\left[\frac{R_{it}(Y_{it} - \mu_{it})}{w_{it}} \mid H_{i,t-1}\right]$$
$$+ E_{H_{i,t-1}}Var\left[\frac{R_{it}(Y_{it} - \mu_{it})}{w_{it}} \mid H_{i,t-1}\right], \tag{20}$$

where

$$Var_{H_{i,t-1}}E\left[\frac{R_{it}(Y_{it} - \mu_{it})}{w_{it}} \mid H_{i,t-1}\right] = Var_{H_{i,t-1}}\left[\frac{1}{w_{it}}w_{it}(\lambda_{it} - \mu_{it})\right]$$
$$= E_{H_{i,t-1}}\left[(\lambda_{it} - \mu_{it})^2\right] \tag{21}$$

since $\left[E_{H_{i,t-1}}(\lambda_{it} - \mu_{it})\right]^2 = 0$, and

$$E_{H_{i,t-1}}Var\left[\frac{R_{it}(Y_{it} - \mu_{it})}{w_{it}} \mid H_{i,t-1}\right]$$

$$= E_{H_{i,t-1}}\left[\frac{1}{w_{it}}(1 - w_{it})\{\lambda_{it}(1 - \lambda_{it}) + (\lambda_{it} - \mu_{it})^2\} + \lambda_{it}(1 - \lambda_{it})\right]$$

$$= E_{H_{i,t-1}}\left[\frac{1}{w_{it}}\lambda_i(1 - \lambda_{it}) + \frac{1}{w_{it}}(\lambda_{it} - \mu_{it})^2 - (\lambda_{it} - \mu_{it})^2\right] \tag{22}$$

Substituting (21) and (22) into (20) gives

$$Var\left[\frac{R_{it}(Y_{it} - \mu_{it})}{w_{it}}\right] = \mu_{it}(1 - \mu_{it})E_{H_{i,t-1}}\left(\frac{1}{w_{it}}\right) + \rho(1 - 2\mu_{it}) \times$$

$$\left[E_{H_{i,t-1}}\left(\frac{Y_{i,t-1}}{w_{it}}\right) - \mu_{i,t-1}E_{H_{i,t-1}}\left(\frac{1}{w_{it}}\right)\right] \tag{23}$$

since $\lambda_{it} = \mu_{it} + \rho(y_{i,t-1} - \mu_{i,t-1})$ by (1). The conditional expectations given the response history, $H_{i,t-1}$, in (23) are evaluated as follows:

$$E_{H_{i,t-1}}\left(\frac{1}{w_{it}}\right) = \sum_{y_{i1},y_{i2},\cdots,y_{i,t-1}} \frac{1}{w_{it}} \mu_{i1}^{y_{i1}}(1-\mu_{i1})^{1-y_{i1}} \prod_{j=2}^{t-1}(\lambda_{ij})^{y_{ij}}(1-\lambda_{ij})^{1-y_{ij}} \quad (24)$$

and

$$E_{H_{i,t-1}}\left(\frac{Y_{i,t-1}}{w_{it}}\right) = \sum_{y_{i1},y_{i2},\cdots,y_{i,t-1}} \left(\frac{y_{i,t-1}}{w_{it}}\right) \mu_{i1}^{y_{i1}}(1-\mu_{i1})^{1-y_{i1}} \prod_{j=2}^{t-1}(\lambda_{ij})^{y_{ij}}(1-\lambda_{ij})^{1-y_{ij}}$$
$$(25)$$

2.2.1 FSGQL(I) Approach

Note that in a complete longitudinal setup, one may obtain consistent regression estimates even if longitudinal correlations are ignored in developing the estimating equation but such estimates may not be efficient (Sutradhar 2011, Chap. 7). By this token, to obtain consistent regression estimates in the incomplete longitudinal setup, we may still use the independence assumption (i.e., use $\rho = 0$) but the missing mechanism must be accommodated to formulate the covariance matrix for the construction of the estimating equation. Thus, for simplicity, we now consider a specialized version of the FSGQL approach, namely FSGQL(I) approach, where the GQL estimating equation is developed by using the independence assumption ($\rho = 0$)-based covariance matrix. More specifically, under this approach, the covariance matrix $\Sigma_i^*(\beta, \rho = 0, \alpha)$ has the form

$$\Sigma_i^*(\beta, \rho = 0, \alpha)$$
$$\equiv \begin{cases} \sigma_{i,11}^* = \mu_{i1}[1-\mu_{i1}] \\ \sigma_{i,tt}^* = E_{H_i(y)}[w_{it}^{-1}\{\mu_{it}(1-\mu_{it})\}], (\text{for } t = 2,\ldots,T_i) \\ \sigma*_{i,ut} = 0, (\text{for } u \neq t), \end{cases} \quad (26)$$

where $E_{H_i(y)}[w_{it}^{-1}]$ is computed by (24).

Now by replacing the "working" covariance matrix $V_i(\alpha^*)$ in the WGEE given in (18) with $\Sigma_i^*(\beta, \rho = 0, \alpha)$, one may obtain the FSGQL(I) estimate for β by solving the estimating equation

$$\sum_{i=1}^{K} \frac{\partial \mu_i'}{\partial beta} [\Sigma_i^*(\beta, \rho = 0, \alpha)]^{-1} \Delta_i(y_i - \mu_i) = 0. \quad (27)$$

2.3 CWGQL Approach

Note that instead of using the distance function with unconditional zero mean, one may like to exploit the distance function with zero mean conditionally. This is possible only when the expectation of the binary response conditional on the past history is known. In this case, by replacing μ_{it} with λ_{it} in (17), one may construct the distance function which has mean zero conditional on the past history, that is,

$$E\left[\frac{R_{it}(Y_{it} - \lambda_{it}(\beta,\rho))}{w_{it}} \mid H_{i,t-1}\right] = 0, \tag{28}$$

where for binary AR(1) model, for example, the conditional mean has the form

$$\lambda_{it}(\beta,\rho) = \mu_{it} + \rho(y_{i,t-1} - \mu_{i,t-1}), \tag{29}$$

for $t = 2,\ldots,T$.

Suppose that

$$\lambda_i(\beta,\rho) = [\lambda_{i1}(\beta), \lambda_{i2}(H_{i,1}(y),\beta,\rho),\ldots,\lambda_{iT_i}(H_{i,T_i-1}(y),\beta,\rho)]'$$

with $\lambda_{i1}(\beta) = \mu_{i1}(\beta)$. To develop a GQL-type estimating equation in the conditional approach, one minimizes the distance function

$$\sum_{i=1}^{K}[\{\Delta_i(y_i - \lambda_i(\beta,\rho))\}'\{\text{cov}(\Delta_i(y_i - \lambda_i(\beta,\rho)))|H_i(y)\}^{-1}\{\Delta_i(y_i - \lambda_i(\beta,\rho))\}'] \tag{30}$$

with respect to β, the parameter of interest. Given the history, let the conditional covariance matrix $\{\text{cov}(\Delta_i(y_i - \lambda_i(\beta,\rho)))|H_i(y)\}$ be denoted by $\Sigma_{ich}(\beta,\rho)$. Then assuming that β and ρ in $\Sigma_{ich}(\beta,\rho)$ are known, minimizing the quadratic distance function (30) with respect to β is equivalent to solving the equation

$$\sum_{i=1}^{K}\frac{\partial[E\{\Delta_i\lambda_i(\beta,\rho)|H_i(y)\}']}{\partial\beta}\Sigma_{ich}^{-1}(\beta,\rho)\{\Delta_i(y_i - \lambda_i(\beta,\rho))\}$$

$$= \sum_{i=1}^{K}\frac{\partial\lambda_i'(\beta,\rho)}{\partial\beta}\Sigma_{ich}^{-1}(\beta,\rho)\{\Delta_i(y_i - \lambda_i(\beta,\rho))\} = 0. \tag{31}$$

Computational formula for $\Sigma_{ich}(\beta,\rho,\gamma)$

For convenience, we first write

$$\Delta_i = W_i^{-1}R_i, \text{ with}$$

$$W_i = \text{diag}[w_{i1}, w_{i2},\ldots,w_{iT_i}], \text{ and } R_i = \text{diag}[R_{i1},\ldots,R_{iT_i}].$$

It then follows that

$$\Sigma_{ich}(\beta,\rho) = \mathrm{cov}[\{\Delta_i(y_i - \lambda_i(\beta,\rho))\}|H_i(y)]$$
$$= W_i^{-1}\mathrm{cov}[\{R_i(y_i - \lambda_i(\beta,\rho))\}|H_i(y)]W_i^{-1}. \tag{32}$$

Now to compute the covariance matrix in (32), we write

$$R_i(y_i - \lambda_i(\beta,\rho)) = [R_{i1}(y_{i1} - \lambda_{i1}), \ldots, R_{it}(y_{it} - \lambda_{it}), \ldots, R_{iT_i}(y_{iT_i} - \lambda_{iT_i})]'.$$

It then follows that for $u < t$, for example,

$$\mathrm{cov}[\{R_{iu}(y_{iu} - \lambda_{iu}), R_{it}(y_{it} - \lambda_{it})\}|H_{i,t-1}(y)] = 0, \tag{33}$$

and for $t = 1, \ldots, T_i$,

$$\mathrm{var}[R_{it}(y_{it} - \lambda_{it})|H_{i,t-1}(y)] = \mathrm{var}[R_{it}|H_{i,t-1}(y)]\mathrm{var}[y_{it}|H_{i,t-1}(y)]$$
$$+ E^2[R_{it}|H_{i,t-1}(y)]\mathrm{var}[y_{it}|H_{i,t-1}(y)] + \mathrm{var}[R_{it}|H_{i,t-1}(y)]E^2[(y_{it} - \lambda_{it})|H_{i,t-1}(y)]$$
$$= w_{it}(1 - w_{it})\sigma_{ic,tt} + w_{it}^2\sigma_{ic,tt}$$
$$= w_{it}\sigma_{ic,tt}, \tag{34}$$

where $\sigma_{ic,tt}$ is the conditional variance of y_{it} given the history. For example, in the binary case, $\sigma_{ic,tt} = \lambda_{it}(1 - \lambda_{it})$.

2.3.1 CWGQL Estimating Equation

Now by substituting (34) and (33) into (32), one obtains

$$\Sigma_{ich}(\beta,\rho) = W_i^{-1}W_i\Sigma_{ic}W_i^{-1} = W_i^{-1}\Sigma_{ic}, \tag{35}$$

where $\Sigma_{ic} = \mathrm{diag}[\sigma_{ic,11}, \ldots, \sigma_{ic,T_iT_i}]$. Consequently, when this formula for Σ_{ich} from (35) is applied to the conditional GQL (CGQL) estimating equation in (31), one obtains

$$\sum_{i=1}^{K} \frac{\partial\lambda_i'(\beta,\rho)}{\partial\beta}\Sigma_{ic}^{-1}(\beta,\rho)W_i\{\Delta_i(y_i - \lambda_i(\beta,\rho))\} = 0, \tag{36}$$

which is unaffected by the missing MAR mechanism. This is not surprising, as conditional on the history, R_{it} and y_{it} are independent. However, this fully conditional approach requires the modeling of the conditional means of the responses, which is equivalent to modeling the correlation structure.

2.3.2 Conditional Likelihood Estimation

In fact when conditional inference is used, one can obtain likelihood estimates for β and ρ by maximizing the exact likelihood function under the condition that Y_{it} and R_{it} are independent given the history. This is easier for the analysis of longitudinal binary data as compared to the longitudinal analysis for count data subject to MAR.

Since the R_{it}'s satisfy the monotonic restriction given in Sect. 1, and because R_{it} and Y_{it} are independent conditional on the history under the MAR mechanism, the likelihood function for the ith individual may be expressed as

$$
\begin{aligned}
L_i(\beta, \rho, \alpha) &= f_{i1}(y_{i1}) f_{i2|1}\{(y_{i2}, r_{i2} = 1)|r_{i1} = 1, y_{i1}\} \ldots \\
&\quad \times f_{iT_i|T_{i-1}}\{(y_{iT_i}, r_{iT_i} = 1)|r_{i1} = 1, r_{i2} = 1, \ldots, r_{i(T_i-1)} = 1, H_{i,t-1}(y)\} \\
&= \mu_{i1}^{y_{i1}}[1 - \mu_{i1}]^{1-y_{i1}} \Pi_{t=1}^{T_i}[\{g_{it}\}\{\lambda_{it}^{y_{it}}(1 - \lambda_{it})^{1-y_{it}}\}],
\end{aligned} \tag{37}
$$

where, by (9),

$$
g_{it}(\alpha) = P[(R_{it} = 1)|r_{i1} = 1, \ldots, r_{i,t-1} = 1, H_{i,t-1}(y)] = \frac{\exp(1 + \alpha y_{i,t-1})}{1 + \exp(1 + \alpha y_{i,t-1})}.
$$

3 Simulation Study

3.1 Comparison Between WGEE (AR(1)), WGEE(I) and FSGQL(I) Approaches: Multinomial Distribution Based Joint Generation of R and y

In this section, we describe and report the results of a simulation study that centers on a comparison of the WGEE approach of Robins et al. (1995) for estimating the regression parameter vector with the proposed FSGQL approach. Recall that the WGEE in (18) was constructed by using a "working" covariance matrix $V_i(\alpha^*) = A_i^{\frac{1}{2}} R_i^*(\alpha^*) A_i^{\frac{1}{2}}$, of the response vector y_i. Note that this weight matrix was chosen ignoring the missing mechanism. Furthermore, there is no guideline to choose the "working" correlation matrix $R_i^*(\alpha^*)$. In the simulation study, we will consider a non-stationary longitudinal binary AR(1) model with true correlation structure $C_i(\rho)$ given by (4), for the responses subject to MAR. To examine the performance of the WGEE approach (18), we choose the best possible stationary AR(1) correlation form, namely,

$$
R_i^*(\alpha^*) = (r_{ut}^*(\alpha^*)) = (\alpha^{*|t-u|}),
$$

as compared to using MA(1) and EQC-based "working" correlation matrices. We will refer to this WGEE as the WGEE(AR(1)). Also we will consider the simplest version of the WGEE approach, namely WGEE(I), which is obtained based on the independence assumption by using $\alpha^* = 0$ in the "working" correlation matrix $R_i^*(\alpha^*)$. These two versions of the WGEE approach will be compared with the FSGQL(I) approach in (27) which was constructed by accommodating missing mechanism but by using longitudinal independence assumption, i.e., $\rho = 0$ or $C_i(\rho) = I_{T_i}$. For simplicity, in the present simulation study, we do not consider the true complete covariance matrix $\Sigma_i^*(\beta, \rho, \alpha)$-based FSGQL approach in (19).

3.1.1 Joint Generation of (R and y) Incomplete Binary Data: Multinomial Distribution Based

In order to generate an incomplete longitudinal binary data set subject to MAR, we follow the approach of Sutradhar and Mallick (2010). Specifically, the procedure initially assumes that every individual provides a response at time $t = 1$. Thus, since $R_{i1} = 1$ for all $i = 1, \cdots, K$, a binary response y_{i1} is generated with marginal probability μ_{i1}. Subsequently, y_{it} is only observed for the i-th individual ($i = 1, \cdots, K$) at time t ($t = 2, \cdots, T$) when $R_{it} = 1$ conditional on having observed the previous $t - 1$ responses for that individual; in other words, conditional on $R_{i1} = 1, \cdots, R_{i,t-1} = 1$. Therefore, at time t ($t = 2, \cdots, T$), both Y_{it} and R_{it} are random variables conditional on the observed history up to time $t - 1$, and, as such, one of the following three events occurs:

$$E_1 : [R_{it} = 1, Y_{it} = 1 \mid R_{i1} = \cdots = R_{i,t-1} = 1, H_{i,t-1}(y)],$$
$$E_2 : [R_{it} = 1, Y_{it} = 0 \mid R_{i1} = \cdots = R_{i,t-1} = 1, H_{i,t-1}(y)],$$
$$\text{or } E_3 : [R_{it} = 0 \mid R_{i1} = \cdots = R_{i,t-1} = 1, H_{i,t-1}(y)], \text{ which implies that } y_{it}$$

is not observed.

Let $z_{its} = 1$ for any $s = 1, 2, 3$ indicate that E_s has occurred. Then, for $l \neq s$, $z_{itl} = 0$, and it must be the case that $\sum_{s=1}^{3} z_{its} = 1$. Let $p_{its} = P(z_{its} = 1)$ for $s = 1, 2, 3$. If we set $q = 1$ in (9), and use the resulting equation in conjunction with model (1), the p_{its} may be expressed as

$$p_{it1} = P(z_{it1} = 1) = P[R_{it} = 1, Y_{it} = 1 \mid R_{i1} = 1, \cdots, R_{i,t-1} = 1; H_{i,t-1}(y)],$$
$$p_{it2} = P(z_{it2} = 1) = P[R_{it} = 1, Y_{it} = 0 \mid R_{i1} = 1, \cdots, R_{i,t-1} = 1; H_{i,t-1}(y)],$$

and

$$p_{it3} = P(z_{it3} = 1) = P[R_{it} = 0 \mid R_{i1} = 1, \cdots, R_{i,t-1} = 1; H_{i,t-1}(y)],$$

which can be written as

$$p_{it1} = P[R_{it} = 1 \mid R_{i1} = 1, \cdots, R_{i,t-1} = 1; H_{i,t-1}(y)] P[Y_{it} = 1 \mid H_{i,t-1}(y)]$$
$$= g_{it}(y_{i,t-1}; \alpha) \lambda_{it}, \tag{38}$$

$$p_{it2} = P[R_{it} = 1 \mid R_{i1} = 1, \cdots, R_{i,t-1} = 1; H_{i,t-1}(y)] P[Y_{it} = 0 \mid H_{i,t-1}(y)]$$
$$= g_{it}(y_{i,t-1}; \alpha)(1 - \lambda_{it}), \tag{39}$$

and

$$p_{it3} = P[R_{it} = 0 \mid R_{i1} = 1, \cdots, R_{i,t-1} = 1; H_{i,t-1}(y)] = 1 - g_{it}(y_{i,t-1}; \alpha), \tag{40}$$

where

$$g_{it}(y_{i,t-1}; \alpha) = exp(1 + \alpha y_{i,t-1})/[1 + exp(1 + \alpha y_{i,t-1})]. \tag{41}$$

Thus, Sutradhar and Mallick (2010) summarize the data generation routine for the i-th individual, $i = 1, \cdots, K$, as follows:

1. Generate y_{i1} from a Bernoulli distribution with parameter μ_{i1}.
2. For any $t > 1$, the values of z_{its} for $s = 1, 2, 3$ are realized according to the multinomial probability distribution

$$P(z_{it1}, z_{it2}, z_{it3}) = \frac{1!}{z_{it1}! z_{it2}! z_{it3}!} p_{it1}^{z_{it1}} p_{it2}^{z_{it2}} p_{it3}^{z_{it3}}$$

   with $\sum_{s=1}^{3} z_{its} = 1$. For $z_{its} = 1$, allocate the response y_{it} following E_s.
3. If $z_{its} = 1$, stop generating y_{it} for this individual; otherwise repeat steps (1) and (2) for $t \leq T$.

3.1.2 Comparison Under Various Designs

Regarding the simulation study, for each of four designs, we set $K = 100$ and $T = 4$ and performed 1,000 replications. We considered three different values of longitudinal correlation parameter, setting $\rho = 0.2, 0.5,$ and 0.8 in turn. In order to investigate the effect of the degree of missingness on the estimates of the regression parameter vector, for $\Delta_i = diag(\delta_{i1}, \delta_{i2}, \cdots, \delta_{iT})$ in (18) with $\delta_{it} = R_{it}/w_{it}\{H_{i,t-1}(y); \alpha\} = R_{it}/w_{it}$, we set $w_{it} = \prod_{j=1}^{t} g_{ij}(y_{i,j-1}; \alpha)$ according to (10) with $q = 1$. We then studied two levels for α, namely $\alpha = 1$, and $\alpha = -3$ (We assume throughout the simulation study that both ρ and α are known; hence, we do not concern ourselves with estimating these quantities). Note that, according to (41), $P[R_{it} = 0 \mid y_{i,t-1} = 1] = 0.12$ and $P[R_{it} = 0 \mid y_{i,t-1} = 0] = 0.27$ for $\alpha = 1$, while $P[R_{it} = 0 \mid y_{i,t-1} = 1] = 0.88$ and $P[R_{it} = 0 \mid y_{i,t-1} = 0] = 0.27$ for $\alpha = -3$. Thus, when $\alpha = -3$, the rate of missingness is extremely high, as expected.

Table 1 (Based on data using joint generation approach) Simulated means (SM) and standard errors (SSE) based on 1,000 simulations for $\beta = 0.5$ and selected values of ρ and α, selected design

Design	α	ρ	Statistic	WGEE(AR(1)) $\hat{\beta}$	#	WGEE(I) $\hat{\beta}$	#	FSGQL(I) $\hat{\beta}$	#
No	−3	0.2	SM	0.574	231	0.471	174	0.501	1,000
covariate			SSE	0.284		0.293		0.196	
		0.5	SM	0.551	361	0.250	233	0.502	1,000
			SSE	0.295		0.311		0.204	
		0.8	SM	0.556	500	−0.052	345	0.503	1,000
			SSE	0.270		0.286		0.219	
	1	0.2	SM	0.478	1,000	0.502	1,000	0.501	1,000
			SSE	0.144		0.146		0.141	
		0.5	SM	0.434	1,000	0.503	1,000	0.504	1,000
			SSE	0.164		0.172		0.169	
		0.8	SM	0.372	1,000	0.499	1,000	0.500	1,000
			SSE	0.190		0.206		0.200	
$p = 1$	−3	0.2	SM	0.503	440	0.514	383	0.499	1,000
stationary			SSE	0.348		0.358		0.292	
		0.5	SM	0.504	593	0.508	464	0.521	999
			SSE	0.351		0.380		0.294	
		0.8	SM	0.509	663	0.518	559	0.529	999
			SSE	0.361		0.408		0.325	
	1	0.2	SM	0.500	1,000	0.501	1,000	0.500	1,000
			SSE	0.190		0.192		0.185	
		0.5	SM	0.496	1,000	0.499	1,000	0.498	1,000
			SSE	0.234		0.241		0.233	
		0.8	SM	0.516	1,000	0.514	1,000	0.515	1,000
			SSE	0.271		0.291		0.282	

Note that # refers to the number of simulations where convergence is achieved

Initially, we compared the WGEE(I) and FSGQL(I) approaches using a stationary design that essentially contained no covariates. For this design, we simply had a single $\beta_1 = 0.5$, while the associated $x_{it1} = 1$ for all $i = 1, \cdots, 100$ and $t = 1, \cdots, 4$. Table 1 presents the means and standard errors of the WGEE and FSGQL(I) estimates over the 1,000 replications for each of the six combinations of ρ and α. The number of replications that converged is also reported. When the degree of missingness is not overly severe ($\alpha = 1$), there is little difference in the WGEE(I) and FSGQL(I) estimates. Both approaches produce essentially unbiased estimates, and all replications converge. However, when the degree of missingness is more pronounced ($\alpha = -3$), the WGEE(I) estimates are significantly biased. In addition, regardless of the value of ρ, more than half of the replications did not converge. On the other hand, the FSGQL(I) estimates are still unbiased, and all replications continue to converge. We investigated the WGEE approach further by considering an AR(1) type "working" correlation structure instead of an independence assumption-based "working" correlation matrix. This WGEE

Table 2 (Based on data using joint generation approach) Simulated means (SM) and standard errors (SSE) based on 1,000 simulations for $\beta_1 = \beta_2 = 0.5$ and selected values of ρ and α, selected design (D) with $p = 2$ stationary (S) and non-stationary (NS) covariates

D	α	ρ	Statistic	WGEE(AR(1))			WGEE(I)			FSGQL(I)		
				$\hat{\beta}_1$	$\hat{\beta}_2$	#	$\hat{\beta}_1$	$\hat{\beta}_2$	#	$\hat{\beta}_1$	$\hat{\beta}_2$	#
S	−3	0.2	SM	0.598	0.598	122	0.506	0.593	85	0.512	0.520	999
			SSE	0.284	0.389		0.301	0.380		0.200	0.287	
		0.5	SM	0.550	0.510	277	0.243	0.485	146	0.510	0.518	1,000
			SSE	0.276	0.376		0.290	0.359		0.213	0.299	
		0.8	SM	0.544	0.527	403	−0.046	0.515	239	0.504	0.516	999
			SSE	0.273	0.342		0.301	0.383		0.219	0.317	
	1	0.2	SM	0.477	0.508	1,000	0.501	0.509	1,000	0.503	0.509	1,000
			SSE	0.139	0.202		0.141	0.203		0.137	0.199	
		0.5	SM	0.431	0.500	1,000	0.500	0.500	1,000	0.499	0.499	1,000
			SSE	0.166	0.246		0.171	0.258		0.166	0.248	
		0.8	SM	0.386	0.516	1,000	0.516	0.520	1,000	0.515	0.519	1,000
			SSE	0.205	0.297		0.222	0.322		0.216	0.311	
NS	−3	0.2	SM	0.497	0.284	19	0.449	0.045	20	0.514	0.428	955
			SSE	0.509	1.234		0.503	1.082		0.453	1.406	
		0.5	SM	0.635	−0.693	31	0.174	−0.789	36	0.507	0.520	962
			SSE	0.512	0.886		0.495	1.151		0.459	1.388	
		0.8	SM	0.844	−1.122	120	0.021	−3.764	51	0.491	0.477	948
			SSE	0.545	0.665		0.623	1.387		0.468	1.447	
	1	0.2	SM	0.481	0.453	999	0.503	0.511	999	0.504	0.506	1,000
			SSE	0.390	0.598		0.391	0.607		0.384	0.600	
		0.5	SM	0.439	0.324	1,000	0.509	0.495	1,000	0.509	0.492	1,000
			SSE	0.374	0.562		0.406	0.640		0.402	0.631	
		0.8	SM	0.386	0.201	999	0.502	0.537	996	0.499	0.532	1,000
			SSE	0.290	0.442		0.408	0.693		0.417	0.686	

Table 3 (Based on data using conditional approach) Simulated means (SM) and standard errors (SSE) based on 1,000 simulations for $\beta = 0.5$ and selected values of ρ and α, selected design

Design	α	ρ	Statistic	WGEE(AR(1))		WGEE(I)		FSGQL(I)	
				$\hat{\beta}$	#	$\hat{\beta}$	#	$\hat{\beta}$	#
No	−3	0.2	SM	–	0	0.565	912	0.495	1,000
covariate			SSE	–		0.238		0.200	
		0.5	SM	–	0	0.550	910	0.495	1,000
			SSE	–		0.250		0.206	
		0.8	SM	–	0	0.523	907	0.495	1,000
			SSE	–		0.261		0.218	
	1	0.2	SM	0.402	6	0.471	1,000	0.495	1,000
			SSE	0.293		0.151		0.139	
		0.5	SM	0.323	4	0.478	1,000	0.505	1,000
			SSE	0.374		0.170		0.166	
		0.8	SM	0.391	4	0.466	1,000	0.493	1,000
			SSE	0.201		0.199		0.205	
$p = 1$	−3	0.2	SM	0.472	415	0.479	375	0.497	998
stationary			SSE	0.585		0.598		0.275	
		0.5	SM	0.511	571	0.528	458	0.505	999
			SSE	0.621		0.647		0.287	
		0.8	SM	0.489	696	0.503	574	0.515	1,000
			SSE	0.606		0.643		0.330	
	1	0.2	SM	0.498	1,000	0.499	1,000	0.499	1,000
			SSE	0.537		0.538		0.198	
		0.5	SM	0.513	1,000	0.512	1,000	0.509	1,000
			SSE	0.568		0.571		0.245	
		0.8	SM	0.509	1,000	0.514	1,000	0.514	1,000
			SSE	0.579		0.595		0.290	

Note that # refers to the number of simulations where convergence is achieved

approach is referred to as the WGEE(AR(1)) approach. Specifically, we set $V_i(\alpha^*) = A_i^{1/2} R_i^*(\alpha^*) A_i^{1/2}$, where $R_i^*(\alpha^*)$ is a $T \times T$ correlation matrix with $corr(Y_{it}, Y_{i,t+l}) = \alpha^{*l}$ and $A_i = diag(\sigma_{i,11}, \cdots, \sigma_{i,tt}, \cdots, \sigma_{i,T_iT_i}, 0, \cdots, 0)$ with $\sigma_{i,tt} = \mu_{it}(1 - \mu_{it})$. To avoid estimation of α^* we have used $\alpha^* = \rho$. The results obtained for each of the six combinations of ρ and α (the missing dependence parameter) are also presented in Table 1. For $\alpha = -3$, the WGEE(AR(1)) estimates based on an AR(1) type structure are significantly better than those based on independence, and the number of replications that converged is also notably higher. Nonetheless, the independent FSGQL(I) estimates are still noticeably better than either of the WGEE estimates. Also of note is the fact that the WGEE(AR(1)) estimates based on the AR(1) structure for $\alpha = 1$ are outperformed by their independent covariance structure counterparts.

We also considered a stationary design consisting of one covariate with associated parameter $\beta_1 = 0.5$. Specifically, for all $t = 1, \cdots, 4$, we set $x_{it1} = -1$ for $i = 1, \cdots, K/4$, $x_{it1} = 0$ for $i = (K/4) + 1, \cdots, 3K/4$, and $x_{it1} = 1$ for $i = (3K/4) +$

Table 4 (Based on data using conditional approach) Simulated means (SM) and standard errors (SSE) based on 1,000 simulations for $\beta_1 = \beta_2 = 0.5$ and selected values of ρ and α, selected design (D) with $p = 2$ stationary (S) and non-stationary (NS) covariates

D	α	ρ	Statistic	WGEE(AR(1)) $\hat{\beta}_1$	$\hat{\beta}_2$	#	WGEE(I) $\hat{\beta}_1$	$\hat{\beta}_2$	#	FSGQL(I) $\hat{\beta}_1$	$\hat{\beta}_2$	#
S	−3	0.2	SM	0.550	0.483	126	0.453	0.512	101	0.505	0.511	1,000
			SSE	0.277	0.400		0.269	0.395		0.207	0.297	
		0.5	SM	0.594	0.493	279	0.279	0.469	150	0.514	0.510	998
			SSE	0.292	0.400		0.293	0.424		0.208	0.309	
		0.8	SM	0.561	0.575	385	−0.039	0.557	231	0.514	0.551	1,000
			SSE	0.262	0.379		0.307	0.415		0.222	0.325	
	1	0.2	SM	0.478	0.501	1,000	0.503	0.500	1,000	0.502	0.502	1,000
			SSE	0.143	0.205		0.144	0.207		0.141	0.202	
		0.5	SM	0.439	0.506	1,000	0.509	0.506	1,000	0.509	0.504	1,000
			SSE	0.172	0.240		0.179	0.251		0.174	0.242	
		0.8	SM	0.386	0.522	1,000	0.520	0.526	1,000	0.517	0.526	1,000
			SSE	0.200	0.291		0.218	0.317		0.211	0.307	
NS	−3	0.2	SM	0.678	−0.095	27	0.504	−0.185	25	0.497	0.538	952
			SSE	0.529	1.346		0.572	1.268		0.458	1.358	
		0.5	SM	0.654	−1.098	45	0.216	−1.614	26	0.486	0.421	946
			SSE	0.521	0.914		0.634	1.117		0.469	1.411	
		0.8	SM	0.841	−1.261	134	0.086	−3.630	59	0.470	0.389	949
			SSE	0.562	0.851		0.565	1.374		0.471	1.451	
	1	0.2	SM	0.488	0.436	1,000	0.507	0.502	1,000	0.506	0.498	1,000
			SSE	0.372	0.582		0.376	0.595		0.370	0.584	
		0.5	SM	0.450	0.317	1,000	0.512	0.500	999	0.510	0.498	1,000
			SSE	0.354	0.557		0.393	0.644		0.397	0.634	
		0.8	SM	0.372	0.200	1,000	0.506	0.539	998	0.505	0.536	1,000
			SSE	0.291	0.435		0.413	0.720		0.423	0.711	

$1, \cdots, K$. The simulation results associated with this design are presented in Table 1 for each combination of ρ and α. When $\alpha = 1$, the performance of WGEE(AR(1)), WGEE(I) and FSGQL(I) are very similar. It is also important to note that when $\alpha = -3$, despite the fact that the average estimates for the regression parameters are better for WGEE under relatively higher longitudinal correlations of $\rho = 0.5$ and 0.8, the estimated standard errors are significantly smaller for the proposed FSGQL(I) technique. In addition, WGEE experiences convergence problems on a significant number of simulation replications; when an independent covariance structure is assumed, convergence rates ranged between 40% and 60%, approximately, and were only slightly better when an AR(1) structure was specified.

Two designs consisting of two covariates with associated regression parameters $\beta_1 = \beta_2 = 0.5$ were also studied; one consisted of two stationary covariates, the other of nonstationary ones. For the design consisting of two stationary covariates, we set $x_{it1} = 1$ for all $i = 1, \cdots, 100$ and $t = 1, \cdots, 4$ as in the design with no covariate, and x_{it2} according to the values specified for the single covariate design described above. The two covariates in the nonstationary design were set as follows:

$$
x_{it1} = \begin{cases}
\frac{1}{2}, & \text{for } i = 1, \cdots, \frac{K}{4}; t = 1,2 \\
0, & \text{for } i = 1, \cdots, \frac{K}{4}; t = 3,4 \\
-\frac{1}{2}, & \text{for } i = \frac{K}{4}+1, \cdots, \frac{3K}{4}; t = 1 \\
0, & \text{for } i = \frac{K}{4}+1, \cdots, \frac{3K}{4}; t = 2,3 \\
\frac{1}{2}, & \text{for } i = \frac{K}{4}+1, \cdots, \frac{3K}{4}; t = 4 \\
\frac{t}{2T}, & \text{for } i = \frac{3K}{4}+1, \cdots, K; t = 1, \cdots, 4
\end{cases}
$$

and

$$
x_{it2} = \begin{cases}
\frac{t-2.5}{2T}, & \text{for } i = 1, \cdots, \frac{K}{2}; t = 1, \cdots, 4 \\
0, & \text{for } i = \frac{K}{2}+1, \cdots, K; t = 1,2 \\
\frac{1}{2}, & \text{for } i = \frac{K}{2}+1, \cdots, K; t = 3,4
\end{cases}
$$

The results for both the stationary and non-stationary two-covariate designs are presented in Table 2. For both designs, when $\alpha = 1$, the performance of WGEE(I) under an independent covariance structure and FSGQL(I) is very similar. The estimates obtained using WGEE(AR(1)) with an AR(1) structure appear to be biased. When $\alpha = -3$, and there is a significantly higher degree of missingness, the estimates obtained under WGEE are biased regardless of the assumed covariance structure and the level of longitudinal correlation; this is particularly the case in the nonstationary design. Also of note is the fact that the convergence rates under the WGEE approach are very poor, with the majority under 5% for the nonstationary design.

3.2 Comparison of WGEE(AR(1)), WGEE(I), and FSGQL(I) Approaches: Generating R and y conditionally

Because R_{it} and y_{it} are independent conditional on the history $H_{i,t-1}(y)$, instead of generating them by using a multinomial distribution discussed in Sect. 3.1.1, one may generate them by using a conditional approach as follows:

1. Generate y_{i1} from $bin(\mu_{i1})$ for all $i, i = 1, \ldots, K$.
2. For i-th individual, generate R_{i2} from $bin(g_{i2})$, where g_{it} is given by (9) for $t = 2, \ldots, T$.
3. If $R_{i2} = 0$, consider $R_{ij} = 0$ and stop generating y_{ij} $(j = 2, \cdots, T)$.
4. If $R_{i2} = 1$, generate y_{i2} from $bin(\lambda_{i2})$, where λ_{it} is the mean of Y_{it} conditional on $y_{i,t-1}$ for $t = 2, \ldots, T$, as given by (1).
5. Repeat from step 2 for $j = 3, \cdots, T$.

The estimates for the same designs are obtained as in Sect. 3.1.2, and the simulation results are reported in Tables 3 and 4. The results are similar to those of Tables 1 and 2, except that WGEE approaches appear to encounter more convergence problems especially when proportion of missing values is large.

3.3 Performance of CWGQL Approach: Multinomial Distribution-Based Joint Generation of R and y

As opposed to the marginal approach where the unconditional mean function $\mu_{it}(\beta)$ is modeled, in the longitudinal setup it is more appropriate to model the conditional regression (mean) function. When complete longitudinal binary data follow an AR(1)-type correlation model, as pointed out in (1), the conditional regression function may be modeled as

$$\lambda_{it}(\beta, \rho) = \mu_{it} + \rho(y_{i,t-1} - \mu_{i,t-1}), \text{for } t = 2, \ldots, T.$$

Furthermore, as pointed out in (28), because in the incomplete longitudinal setup with MAR mechanism one finds

$$E\left[\frac{R_{it}(Y_{it} - \lambda_{it}(\beta, \rho))}{w_{it}} \mid H_{i,t-1}\right] = 0,$$

the regression parameter β in $\mu_{it}(\beta)$ modeled through $\lambda_{it}(\beta, \rho)$ can be estimated by solving the CWGQL estimating (36). For the same design parameters used in Sect. 3.1.2 for Tables 1 and 2, by generating incomplete data using the multinomial distribution discussed in Sect. 3.1.1, we have obtained the CWGQL estimates for β under different scenarios as for the results shown in Tables 1 and 2. The CWGQL estimates along with their standard errors are reported in Tables 5 and 6.

Table 5 (Based on data using joint generation approach) Simulated means (SM) and standard errors (SSE) for *CWGQL* approach with $\beta = 0.5$ and selected values of ρ and α, based on 1,000 simulations

			No covariate		$p = 1$ (stationary)	
α	ρ	Statistic	$\hat{\beta}$	#	$\hat{\beta}$	#
-3	0.2	SM	0.505	1,000	0.494	1,000
		SSE	0.178		0.247	
	0.5	SM	0.505	1,000	0.509	1,000
		SSE	0.188		0.256	
	0.8	SM	0.507	1,000	0.517	1,000
		SSE	0.204		0.291	
1	0.2	SM	0.500	1,000	0.499	1,000
		SSE	0.138		0.183	
	0.5	SM	0.501	1,000	0.493	1,000
		SSE	0.158		0.224	
	0.8	SM	0.497	1,000	0.514	1,000
		SSE	0.183		0.265	

Note that # refers to the number of simulations where convergence is achieved

Table 6 (Based on data using joint generation approach) Simulated means (SM) and standard errors (SSE) for *CWGQL* approach with $\beta_1 = \beta_2 = 0.5$ and selected values of ρ and α, based on 1,000 simulations

			Stationary covariates			Nonstationary covariates		
α	ρ	Statistic	$\hat{\beta}_1$	$\hat{\beta}_2$	#	$\hat{\beta}_1$	$\hat{\beta}_2$	#
-3	0.2	SM	0.522	0.513	1,000	0.517	0.515	789
		SSE	0.185	0.260		0.429	0.992	
	0.5	SM	0.516	0.510	1,000	0.507	0.645	770
		SSE	0.201	0.278		0.410	0.903	
	0.8	SM	0.506	0.508	1,000	0.497	0.730	597
		SSE	0.203	0.298		0.370	0.837	
1	0.2	SM	0.501	0.507	1,000	0.503	0.503	1,000
		SSE	0.133	0.193		0.372	0.581	
	0.5	SM	0.498	0.500	1,000	0.502	0.471	901
		SSE	0.160	0.232		0.350	0.515	
	0.8	SM	0.506	0.509	1,000	0.449	0.423	302
		SSE	0.195	0.281		0.254	0.455	

Note that # refers to the number of simulations where convergence is achieved

In order to examine the relative performance of the CWGQL approach with those of WGEE(AR(1)), WGEE(I), and FSGQL(I), it is sufficient to compare the CWGQL approach with the FSGQL(I) approach only. This is because it was found from the results in Tables 1 and 2 that the WGEE approaches may encounter serious convergence problems (showing consistency breakdown) and also may produce highly biased estimates, where the FSGQL(I) approach, in general,

does not encounter such convergence problems and produces almost unbiased estimates even if a large proportion of values are missing. Now as compared to the FSGQL(I) approach, the CWGQL approach appears to produce slightly more efficient estimates than the FSGQL(I) approach. For example, when $\alpha = 1$ (moderate missing) and $\rho = 0.5$, in the no-covariate case, the FSGQL(I) approach (Table 1) produces an average estimate of $\beta = 0.5$ as 0.504 with standard error 0.169, whereas the CWGQL approach (Table 5) produces β estimate as 0.501 with standard error 0.158. Similarly when $\rho = 0.8$ and $\alpha = -3$ (high missing), in the stationary one-covariate case, FSGQL(I) produces an estimate with standard error 0.325 as compared to 0.291 for CWGQL. Similar results are found for the stationary two-covariate case. Also in these stationary cases, the CWGQL approach does not encounter any convergence problems even if the proportion of missing is high. In the non-stationary cases however, the CWGQL approach encounters some convergence problems when the proportion of missing is high, but the problem is less serious than the WGEE and WGEE(I) approaches.

4 Conclusion and Discussion

It was found that the existing WGEE (Robins et al. 1995) and WGEE(I) approaches in general encounter convergence problems when the proportion of missing is high, and the WGEE approach may produce highly biased estimates even when the proportion of missing is moderate or low. These results agree with the recent study reported by Sutradhar and Mallick (2010). The WGEE(I) approach, however, produces almost unbiased estimates and consequently this approach produces consistent estimates when the proportion of missing is moderate or low. However, it can be inefficient. The proposed FSGQL(I) approach does not appear to encounter any serious convergence problems even when the proportion of missing is high and the covariates are non-stationary. Also, it produces unbiased estimates similar to the WGEE(I) approach but with smaller standard errors, showing that FSGQL(I) is more efficient as expected than the WGEE(I) approach. Thus even with high proportion of missing, one may reliably use the proposed FSGQL(I) approach for regression estimation whether the covariates are stationary or time dependent. The general FSGQL approach is supposed to increase the efficiency as compared to the FSGQL(I) approach when correlations are large, but this will be studied in the future.

We have also reported some results on the performance of a conditional estimating equation, namely CWGQL estimating equation approach. This approach was found to produce regression estimates with more efficiency than the FSGQL(I) approach. However as compared to the FSGQL(I) approach it encounters convergence problems when covariates are time dependent and the proportion of missing is high. However, it experiences less convergence problems than the WGEE(I) approach.

References

Birmingham, J., Rotnitzky, A., Fitzmaurice, G.M.: Pattern-mixture and selection models for analysing longitudinal data with monotone missing patterns. J. R. Stat. Soc. B **65**, 275–297 (2003)

Fitzmaurice, G.M., Laird, N.M., Zahner, G.E.P.: Multivariate logistic models for incomplete binary responses. J. Am. Stat. Assoc. **91**, 99–108 (1996)

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

Paik, M.C.: The generalized estimating equation approach when data are not missing completely at random. J. Am. Stat. Assoc. **92**, 1320–1329 (1997)

Qaqish, B.F.: A family of multivariate binary distributions for simulating correlated binary variables with specified marginal means and correlations. Biometrika **90**, 455–463 (2003)

Robins, J.M., Rotnitzky, A., Zhao, L.P.: Analysis of semiparametric regression models for repeated outcomes in the presence of missing data. J. Am. Stat. Assoc. **90**, 106–121 (1995)

Rotnitzky, A., Robins, M., Scharfstein, D.O.: Semi-parametric regression for repeated outcomes with nonignorable nonresponse. J. Am. Stat. Assoc. **93**, 1321–1339 (1998)

Rubin, D.B.: Inference and missing data. Biometrika **63**, 581–592 (1976)

Sutradhar, B.C.: An overview on regression models for discrete longitudinal responses. Stat. Sci. **18**, 377–93 (2003)

Sutradhar, B.C.: Dynamic Mixed Models for Familial Longitudinal Data. Springer, New York (2011)

Sutradhar, B.C.: Inference progress in missing data analysis from independent to longitudinal setup. In: Sutradhar, B.C. (ed.) ISS-2012 Proceedings volume on longitudinal data analysis subject to measurement errors, missing values, and/or outliers, Springer Lecture Notes Series, pp. 101–123 (2013)

Sutradhar, B.C., Mallick, T.S.: Modified weights based generalized quasilikelihood inferences in incomplete longitudinal binary models. Can. J. Stat. **38**, 217–231 (2010)

Zeger, S.L., Liang, K.-Y., Self, S.G.: The analysis of binary longitudinal data with time-independent covariates. Biometrika **72**, 31–38 (1985)