Modeling the Association Structure in Doubly Robust GEE for Longitudinal Ordinal Missing Data

José Luiz P. da Silva, Enrico A. Colosimo, Fábio N. Demarqui
Departament of Statistics
Universidade Federal de Minas Gerais
June 16, 2015

Abstract

Generalized Estimation Equations (GEE) are a well-known method for the analysis of categorical longitudinal responses. GEE method has computational simplicity and population parameter interpretation. In the presence of missing data it is only valid under the strong assumption of missing completely at random. A doubly robust estimator (DRGEE) for correlated ordinal longitudinal data is a nice approach for handling intermittently missing response and covariate under the MAR mechanism. Independent working correlation is the standard way in DRGEE. However, when covariate is not time stationary, efficiency can be gained using a structured association. The goal of this paper is to extend the DRGEE estimator to allow modeling the association structure by means of either the correlation coefficient or local odds ratio. Simulation results revealed better performance of the local odds ratio parametrization, specially for small samples. The method is applied to a data set related to Rheumatic Mitral Stenosis.

Keywords: Correlation coefficient; Doubly robust estimators; Generalized estimating equations; Local odds ratio; Missing at random.

1 Introduction

Longitudinal data arise when each individual is measured repeatedly through time. These repeated responses form a cluster and it is expected the response within each cluster to be correlated. A popular approach for the analysis of longitudinal data is the Generalized Estimating Equation (GEE) method, proposed by Liang & Zeger (1986). The goal of this procedure is to estimate fixed parameters without specifying a joint distribution for the data (Nooraei et al., 2014). GEE method requires only the correct specification of the response’s mean structure for the parameter estimator to be consistent and asymptotically normal. The attractive feature of GEE is that the association parameters among repeated measures are taken as a ‘nuisance’ parameters and, unlike maximum likelihood methods, mean parameter estimates are not sensible to the specification of the association structure. Furthermore, GEE method allows marginal interpretation of the parameter of interest and it has computational simplicity.
A simple way of analyzing such correlated data is to consider an independence working assumption for the repeated responses. However, when the covariate design is not time-stationary it will lead to inefficient marginal regression estimates (Lipsitz et al., 1994). In the presence of time-varying covariates efficiency can be gained assuming a different correlation structure. Nevertheless, modeling the association in ordinal data is not a simple task. Different approaches have been proposed to estimate the association parameters for ordinal responses. Lipsitz et al. (1994) provided moment estimators for a variety of correlation matrices, while Parsons et al. (2006) proposed an approach which estimates the correlation vector by minimizing the logarithm of the determinant of the covariance matrix of the fixed parameters. Instead of using correlations, Lumley (1996) proposed using a common global odds to reduce the number of association parameters. Heagerty & Zeger (1996) extended the ALR method proposed by Carey et al. (1993) to ordinal responses using a second set of estimating equations for the global odds ratio. Recently, Touloumis et al. (2013) considered a family of association models to estimate local odds ratios as a measure of association. A comparison study of different working association structures can be found in Nooraee et al. (2014).

In the presence of missing data, inferences are valid if the missingness mechanism is missing completely at random (MCAR), as defined by Rubin (1976). When data is MAR, one can adopt multiple imputation GEE (MIGEE) (Little & Rubin, 1987) or a weighted version (WGEE) (Robins et al. 1995). These single robust versions of GEE for incomplete data require the correct specification of the weight model of GEE (WGEE) or the imputation model (MIGEE). Doubly robust estimators (DRGEE) (Carpenter et al. 2006, Tsiatis 2006, Seaman & Copas 2009, Chen & Zhou 2011) combine ideas from these two approaches. For consistency, it requires only the weight or the imputation model to be correctly specified, providing more flexibility for the modeler.

This work was motivated by the Rheumatic Mitral Stenosis study in which a cohort of 164 patients with rheumatic mitral stenosis (a narrowing of the mitral valve in the heart) were referred for treatment at Hospital das Clinicas of the Federal University of Minas Gerais, Brazil. The response of interest was the functional classification (NYHA), major determinant of quality of life and survival of the individual. The main objective of that study was to characterize the improvement of the functional classification over time. This study was characterized by an arbitrary pattern of missing data. Response and a particular covariate (atrial compliance) were missing for some patients and the MAR mechanism seems to be a reasonable assumption for this data.

In the current paper, we consider a doubly robust approach for the analysis of longitudinal ordinal data with intermittently missing response along with a key covariate that is MAR. The cumulative logit model was used for the marginal means. We extend the doubly robust estimator to accommodate two parametrizations of the association structures: the correlation coefficient (Lipsitz et al., 1994) and the local odds ratio (Touloumis et al., 2013). Efficiency and accuracy of the proposed estimator are compared under these association structures.

The paper is organized as follows. In Section 2 the notation for GEE with fully observed data and discuss the two parametrizations of the association structure. Section 3 outlines WGEE and MIGEE approaches. The proposed methodology is established in Section 4. A simulation study is presented in Section 5, in which the finite-sample biases and mainly standard errors are compared for the standard GEE, MIGEE, WGEE.
and doubly robust versions, under both the correlation and local odds parametrizations. Data arising from the Rheumatic Mitral Stenosis study are analyzed in Section 6. Paper ends with a discussion and future directions in Section 7.

2 Notation and GEE for Complete Data

In this section it is introduced the generalized estimating equations for the analysis of fully observed ordinal data. Section 2.1 establishes the model and notation for longitudinal ordinal data. Section 2.2 presents two competing ways of modeling the association structure in GEE.

2.1 GEE for Longitudinal Ordinal Response

Let \(O_{it} \in \{1, 2, \ldots, J\}\) be the ordinal response for \(i\)-th subject \((i = 1, \ldots, n)\) at time \(t\) \((t = 1, \ldots, T_i, T_i \leq T)\). As the response has \(J\) levels it can be defined as \(Y_{itj} = I(O_{it} = j)\) for \(j = 1, \ldots, J\), where \(I(A)\) denotes the indicator function. \(Y_{itj}\) is converted into the equivalent \((J - 1)\)-variate vector \(Y_{it} = (Y_{it1}, \ldots, Y_{it(J-1)})^T\) and let \(Y_i = (Y_{i1}^T, \ldots, Y_{iT_i}^T)^T\) be the stacked response vector. When \(J = 2\) the response is binary and \(Y_{it}\) is a scalar. Let \(X_i\) denotes the time-stationary covariate for the \(i\)-th subject, and \(Z_i = (Z_{i1}^T, \ldots, Z_{iT_i}^T)^T\) a \(T_i \times q\) matrix of explanatory variables.

The marginal distribution of \(Y_{it}\) is assumed to be multinomial \((\sum_{j=1}^{J} Y_{itj} = 1)\), that is

\[
f(Y_{it}|X_i, Z_i, \beta) = \prod_{j=1}^{J} \mu_{itj}^{y_{itj}},
\]

where \(\mu_{itj} = \mu_{itj}(\beta) = E(Y_{itj}|X_i, Z_i, \beta) = Pr(O_{it} = j|X_i, Z_i, \beta)\), is the probability of response \(j\) at time \(t\) and \(\beta\) is a \(p \times 1\) vector of parameters. Two common choices for modeling \(\mu_{itj}\) are the cumulative logit and probit models. In this work it is assumed a cumulative logit link, that is,

\[
\text{logit}[Pr(O_{it} \leq j|X_i, Z_i)] = \beta_{0j} + X_i \beta_z + Z_{it}^T \beta_z, \quad j = 1, \ldots, J - 1.
\]

Formulation in (2) implies a proportional odds model (McCullagh, 1980). In such model the interpretation of \(\beta\) is the same regardless of the number of categories (i.e., it is invariant to combination of categories). A desired feature is that the exponential of the parameters is interpreted as an odds ratio (Agresti, 2013).

The main interest is to make inferences related to the regression parameters \(\beta = (\beta_{01}, \ldots, \beta_{0J-1}, \beta_z^T)^T\) associated to the \((J - 1) \times 1\) marginal probability vectors

\[
E(Y_{it}|X_i, Z_i) = \mu_{it}(\beta) = (\mu_{it1}, \ldots, \mu_{it(J-1)})^T.
\]

\(\mu_{it}\) is grouped to form a vector \(E(Y_i|X_i, Z_i) = \mu_i = (\mu_{i1}^T, \ldots, \mu_{iT_i}^T)^T\) with the same dimension of \(Y_i\).

In order to estimate \(\beta\), generalized estimation equations are used (Liang & Zeger (1986); Lipsitz et al. (1994), Touloumis et al. (2013)), which takes the form

\[
\sum_{i=1}^{n} U_i(\beta, \alpha) = \sum_{i=1}^{n} D_i V_i^{-1}(Y_i - \mu_i) = 0,
\]
where $D_i = \frac{\partial n}{\partial \beta_i}$ and $V_i = V_i(\beta, \alpha)$ is a $T_i(J-1) \times T_i(J-1)$ covariance matrix for $Y_i$. The vector parameter $\alpha$ expresses a ‘working’ assumption about the correlation/association structure.

Under mild regularity conditions, correct specification of the marginal mean model in (2), and provided that a $\sqrt{n}$-consistent of $\alpha$ is available, Liang and Zeger (1986) proved that the estimator $\hat{\beta}$, obtained by solving (3), is consistent and $\sqrt{n}(\hat{\beta} - \beta)$ converges in distribution to a multivariate normal distribution with mean vector $0$ and covariance matrix

$$V_\beta = \lim_{n \to \infty} n\Sigma_0^{-1}\Sigma_1\Sigma_0^{-1},$$

(4)

where $\Sigma_0 = \sum_{i=1}^{n} D_i V_i^{-1} D_i^T$, and $\Sigma_1 = \sum_{i=1}^{n} D_i V_i^{-1} \text{Cov}(Y_i) V_i^{-1} D_i^T$. In practice, the “sandwich” covariance matrix $V_\beta$ in (4) is calculated by ignoring the limit and replacing $(\beta, \alpha)$ and $\text{Cov}(Y_i)$ by $(\hat{\beta}, \hat{\alpha})$ and $(Y_i - \hat{\mu}_i)(Y_i - \hat{\mu}_i)^T$, respectively (Touloumis et al., 2013).

2.2 Estimation of the nuisance parameter vector and covariance matrix

The ‘working’ assumption is ‘independence’ when no correlation is assumed between pairs of the response of each individual, that is, $\alpha = 0$. Independent Estimating Equations (IEE) are proved to be efficient only when covariates are constant over time or if the independence structure is actually true (Lipsitz et al., 1994). In this case, score equations of the ML method for the regression vector $\beta$ are identical to the IEE if all observations are treated as independent. On the other hand, when there exist within-individual association or time-varying covariates, efficiency can be gained by modeling the correlation structure. Hence, a number of proposals have been formulated to estimate $\alpha$. These alternatives differ in the efficiency of estimating the covariance matrix and computational simplicity. Lipsitz et al. (1994) defined $\alpha$ as the correlation coefficient, and suggested the use of the method of moments to estimate a number correlation structures. Parsons et al. (2000) proposed an approach that estimate the correlation vector $\alpha$ by minimizing an objective function $Q(\alpha | \beta, Y)$. Lumley (1996) proposed using a common global odds. Heagerty & Zeger (1996) extended the ALR method proposed by Carey et al. (1993) to ordinal responses by using a second set of estimating equations for the global odds ratio. Finally, Touloumis et al. (2013) identifies $\alpha$ as a ‘nuisance’ parameter vector that contains the marginalized local odds ratios structure. They employed a family of association models in order to develop meaningful structures for the ordinal response.

2.2.1 Correlation Coefficient

Lipsitz et al. (1994) suggested a method that constrains the correlations at different times between two categories of the response. In their approach the weight matrix $V_i$ is decomposed into the form $V_i(\beta, \alpha) = F_i^{1/2}(\beta) C_i(\alpha) F_i^{1/2}(\beta)$, where $F_i$ is a matrix containing marginal variances, $F_i$, given by

$$F_{it} = \text{diag} [\mu_{it1}(1 - \mu_{it1}), \ldots, \mu_{it,J-1}(1 - \mu_{it,J-1})],$$

and $C_i$ is equal to the marginal correlation matrix. The unknown elements of $C_i(\alpha)$ are the elements of the $(J - 1) \times (J - 1)$ matrix $\rho_{it'}(\alpha)$. Thus, $\rho_{it'}(\alpha)$ is parametrized as
\( \rho_{itw} = \text{Corr}(Y_{it}, Y_{it'}) \). The \((J - 1) \times (J - 1)\) diagonal blocks of \( V \) are \( F_{ii}^{-1/2} V_{ii} F_{ii}^{-1/2} \), with \( V_{ii} = \text{diag}(\mu_{it}) - \mu_{it} \mu_{it}' \); and the \((J - 1) \times (J - 1)\) off-diagonal blocks of \( C_i(\alpha) \) are \( \rho_{itw} = \rho_{itw}(\alpha) \), which represents the correlation between \( Y_{it} \) and \( Y_{it'}, t \neq t' \). The vector \( \alpha \) is a parameter vector associated with the model for \( \rho_{itw} \).

Define the Pearson residual vector, \( e_{it} \) as

\[
e_{it} = F_{it}^{-1/2} (Y_{it} - \mu_{it}).
\]

Then it follows that

\[
C_{itw}(\alpha) = \text{Corr}(Y_{it}, Y_{it'}) = E(e_{it} e_{it'}').
\]

In order to reduce the dimension of the correlation vector, \cite{Lipsitz1994} assumed an uniform correlation structure over the individuals. The use of the method of moments was suggested for a variety of correlation matrices such as

- exchangeable: \( \rho_{itw} = \rho \), for all \( t < t' \);
- 1-dependent: \( \rho_{it,t+1} = \rho_t \), for \( t = 1, \ldots, T - 1 \), and \( \rho_{itw} = 0 \) otherwise;
- banded: \( \rho_{itw} = \rho_\tau \), when \( |t' - t| = \tau \), for \( \tau = 1, \ldots, T - 1 \);
- unstructured: \( \rho_{itw} = \rho_{itw} \).

The estimate \( \hat{\alpha} \) is plugged into (3) and a solution is found for \( \beta \). The solution might be obtained by a Fisher-scoring algorithm.

The correlation coefficient parametrization ignores the scale of the response variable and may result in loss of information regarding the correlation between the variables \cite{Lumley1996}. Moreover, \cite{Lipsitz1994} observed that the ‘working’ correlation matrix is not always positive definite, which may result in a breakdown of the Fisher scoring method. This is specially true over unstructured correlations matrices and small sample sizes. When the given model for \( \rho_{itw} \) contains too many parameters, the resulting estimates of \( \beta \) and \( \alpha \) may be highly variable.

### 2.2.2 Local Odds Ratio

Instead of modeling the correlation coefficients, some authors \cite{Lumley1996, Touloumis2013} modeled the off-diagonal \( V_{itw} \) through the joint probability of the responses \( Y_{it} \) and \( Y_{it'} \), for \( t \neq t' \). The covariance of \( Y_{itj} \) and \( Y_{it'j} \) can be written as

\[
\text{Cov}(Y_{itj}, Y_{it'j'}|X_i, Z_{it}, \beta) = E(Y_{itj}, Y_{it'j'}|X_i, Z_{it}, \beta) - E(Y_{itj}|X_i, Z_{it}, \beta)E(Y_{it'j'}|X_i, Z_{it}, \beta) = \mu_{itj} \mu_{it'j'} - \mu_{itj} \mu_{it'j'}.
\]

The product of the first moments can be calculated through the specification of the marginal model \cite{Touloumis2013}. The joint probabilities \( \mu_{itj} \mu_{it'j'} = P(Y_{itj} = 1, Y_{it'j'}|X_i, Z_{it}, \beta) \) can be modeled by an association vector that describes the association structure for \( \forall i = 1, \ldots, n, t \neq t' = 1, \ldots, T_i \) and \( j, j' = 1, \ldots, J \) \cite{Touloumis2013}. Estimation of these joint probabilities can be achieved through the global odds ratio \cite{Heagerty1996} or a local odds ratio \cite{Touloumis2013}.
argue that the local odds ratio provides the best parametrization because it is variation independent to $\beta$ and they produce valid and unique positive joint probabilities.

Let’s introduce the local odds ratio of Touloumis et al. (2013). Denote the $L = T(T - 1)/2$ time pairs as $(1, 2), (1, 3), \ldots, (T - 1, T)$, where $T = \max\{T_1, \ldots, T_n\}$. Let $F_{ijF_{ij'}} = \sum_{i=1}^{n} Y_{it}Y_{it'}$ the observed frequency of the cutpoint $(j, j')$ at the $(t, t')$ time pair of the marginalized table. Define $\theta_{ijF_{ij'}}$ as the local odds, that is,

$$\theta_{ijF_{ij'}} = \frac{F_{ijF_{ij'}}F_{ij+1F_{ij'+1}}}{F_{ij+1F_{ij'}}F_{ijF_{ij'+1}}}$$

for $j, j' = 1, \ldots, J - 1$, and let $\alpha$ be the $L \times (J - 1)^2$ vector consisting of the local odds ratio

$$\alpha = (\theta_{1121}, \ldots, \theta_{112(J-1)}, \ldots, \theta_{(T-1)1T1}, \ldots, \theta_{(T-1)(J-1)T(J-1)})^T.$$ 

The association vector $\alpha$ can be estimated by fitting a loglinear model for the counts $\{F_{ijF_{ij'}}\}$ simultaneously to all possible $L$ marginalized contingency tables and then calculating the implied local odds ratio (Touloumis et al. 2013). For notational reasons, let $A$ and $B$ be the row and and column variable, respectively, and let $G$ be the group variable with levels being the $L$ ordered pairs. Assuming a Poisson sampling scheme to the $L$ sets of $J \times J$ contingency tables, fit the RC type model (Becker & Clogg 1989)

$$\log(f_{ijF_{ij'}}) = \lambda + \lambda_j^A + \lambda_j^B + \lambda^{AG}_{(it)} + \lambda^{BG}_{(it)} + \varphi(t, t')\nu_{ijF_{ij'}}(t, t'),$$

where $\{\nu_{ij}(t, t') : j = 1, \ldots, J\}$ are the score parameters for the $J$ response categories at time pair $(t, t')$, and $\{f_{ijF_{ij'}} : j, j' = 1, \ldots, J\}$ are the expected frequencies. The maximum likelihood estimate of $\alpha$ is obtained by treating the $L$ marginalized contingency tables as independent. By imposing identifiability constraints on the regression parameters in (5), the resulting unrestricted local odds ratio are determined by $\log(\theta_{ijF_{ij'}}) = \varphi(t, t')(\nu_{ijF_{ij'}} - \nu_{ijF_{ij'+1}})\nu_{ijF_{ij'+1}} - \nu_{ijF_{ij'+1}}$, where the intrinsic parameter $\varphi(t, t')$ measures the average association of the marginalized contingency table. To increase parsimony, common unit-spaced score parameters ($\nu_{ijF_{ij'}} = j$) are usually assumed. Main options for the marginalized local odds ratio structures include

- **uniform**: $\log(\theta_{ijF_{ij'}}) = \varphi$, estimates a single parameter;
- **time exchangeability**: $\log(\theta_{ijF_{ij'}}) = \varphi(t, t')$; estimates $L$ parameters;
- **category exchangeability**: $\log(\theta_{ijF_{ij'}}) = \varphi(\nu_j - \nu_{j+1})(\nu_{j'} - \nu_{j'+1})$; estimates $J - 1$ parameters, and
- **unstructured**: $\log(\theta_{ijF_{ij'}}) = \varphi(t, t')(\nu_{ijF_{ij'}} - \nu_{ijF_{ij'+1}})(\nu_{ijF_{ij'+1}} - \nu_{ijF_{ij'+1}})$, that requires $L(J - 1)$ parameters.

Conditional on $\hat{\alpha}$, and the marginal specification (2), the joint probabilities $\mu_{ijF_{ij'}}$ are estimated based on the adopted local odds ratio structure using the IPFP (Iterative Proportional Fitting Procedure). This algorithm, proposed by (Deming & Stephan 1940), is used to obtain $\mu_{ijF_{ij'}}$ through the marginals $\mu_{ij}$ and $\mu_{ijF_{ij'}}$.
proved that the IPFP solution preserves local odds ratios of the initial values as long as they are positive. Hence, it is straightforward to calculate the weight matrix $V_i$ and the estimating equations in (5) can be solved with respect to $\beta$.

An advantage of the local odds ratio parametrization over the correlation is that the local odds ratio and the marginal regression vector are variation independent. This means that $\beta$ estimates are less sensible to a possibly wrong specification of $\alpha$. As opposed to the correlation parametrization, the estimation of the association parameters does not depend on covariates and, as long as it is based on maximum likelihood models, the $\alpha$ estimates are consistent under MAR. Thus, no adjustment need to be done on $\alpha$ obtained with available data.

3 Available Approaches for Missing Data

Section 3.1 presents a series of assumptions related to mechanism causing data to be missing and necessary to be considered in order to build valid estimators. Multiple imputation and weighed generalized estimation equations are two commonly methods available for missing data under MAR mechanism. These methods are presented in Sections 3.2 and 3.3, respectively. They serve as the basis for the construction of the doubly robust estimator, presented in Section 4.

3.1 Missing Data Framework

In this work it will be assumed that the time-stationary covariate $X_t$ may be missing for some subjects whereas the explanatory variables $Z_i$ are fully observed.

For each occasion $t$ it can be defined $R_{it} = 0$ if $O_{it}$ and $X_i$ are missing, $R_{it} = 1$ if $O_{it}$ is missing and $X_i$ is observed, $R_{it} = 2$ if $O_{it}$ is observed and $X_i$ is missing, and $R_{it} = 3$ if $O_{it}$ and $X_i$ are both observed. Let $R_i = (R_{i1}, \ldots, R_{iT_i})^T$, and $\bar{R}_{it} = (R_{i1}, \ldots, R_{i,t-1})$.

The marginal probability $\Pr(R_i = r_i | O_i, Z_i)$ can be obtained through conditional models of the form $\Pr(R_{it} = r_{it} | \bar{R}_{it}, O_i, Z_i)$. This general formulation encompasses MCAR, MAR and MNAR mechanisms. In particular, the MAR mechanism requires

$$\Pr(R_i = r_i | O_i, Z_i) = \Pr(R_i = r_i | O^o_i, Z_i),$$

where $O^o_i$ denotes the observed components of $O_i$. The following natural further assumption is made

$$\Pr(R_{it} = r_{it} | \bar{R}_{it}, O_i, Z_i) = \Pr(R_{it} = r_{it} | \bar{R}_{it}, O^o_i, \bar{Z}^o_{it}),$$

for each time $t$, where $O^o_{it}$ and $\bar{Z}^o_{it}$ are the histories of observed responses and covariates up to time $t - 1$.

Let $\pi_{it} = \Pr(R_{it} = 3 | O_i, Z_i)$ be the marginal probability of observing both $O_i$ and $X_i$ at time $t$, given the entire vectors of responses and covariates. Then, $\pi_{it}$ is expressed by

$$\pi_{it} = \sum_{r_{i1}, \ldots, r_{i,t-1}} \Pr(R_{it} = 3, R_{i,t-1} = r_{i,t-1}, \ldots, R_{i1} = r_{i1} | O_i, Z_i).$$

This marginal probability can be expressed in terms of the conditional probabilities $\Pr(R_{it} = k | \bar{R}_{it}, O_i, Z_i)$, for $k = 0, 1, 2, 3$. Throughout this paper it will be required
the so-called positivity assumption, that is, \( \pi_{it} \) must be bounded away from zero. This condition is needed in order to guarantee the existence of \( \sqrt{n} \)-consistent estimators of \( \beta \) (Robins et al., 1995).

### 3.2 Multiple Imputation Generalized Estimating Equations

A imputation model commonly used to handle intermittently missing response and covariate, is imputation using chained equations (van Buuren et al., 1999, van Buuren, 2007), which is more commonly referred to as full conditional specification (FCS). This approach specifies conditional distributions for each incomplete variable, conditional on all others variables in the imputation model. Starting from an initial imputation, FCS draws imputations by iterating over the conditional densities.

Denote by \( \tilde{\beta}_m \) and \( \tilde{U}_m \), respectively, the estimate of \( \beta \) and its covariance matrix from the GEE analysis of the \( m \)-th completed data set, \((m = 1, \ldots, M)\). Following (Rubin, 1987), the combined point estimate for the parameter of interest \( \beta \) based on MI is simply the average of the \( M \) complete-data point estimates

\[
\hat{\beta}_{MI} = \frac{1}{M} \sum_{m=1}^{M} \tilde{\beta}_m,
\]

and an estimate of the covariance matrix of \( \hat{\beta}_{MI} \) is given by

\[
\hat{U}_{MI} = \tilde{W} + \left( \frac{M + 1}{M} \right) \tilde{B},
\]

where

\[
\tilde{W} = \frac{1}{M} \sum_{m=1}^{M} \tilde{U}_m \quad \text{and} \quad \tilde{B} = \frac{1}{M - 1} \sum_{m=1}^{M} (\tilde{\beta}_m - \hat{\beta}_{MI})(\tilde{\beta}_m - \hat{\beta}_{MI})',
\]

### 3.3 Weighted Generalized Estimating Equations

Robins et al. (1995) proposed a class of weighted estimating equations to allow for MAR mechanism. In binary longitudinal data, Chen & Zhou (2011) extended the method to accommodate arbitrary patterns of both missing response and covariate. Their method was adapted here for longitudinal ordinal responses.

Define a weight matrix \( \Delta_i = [\delta_{it'}]_{T_i,J_i \times T_i,J_i} \), \( t = 1, \ldots, T_i, t' = 1, \ldots, T_i \), where \( \delta_{it'} = \{I(R_{it} = 1, R_{it'} = 3) + I(R_{it} = 3, R_{it'} = 3)\} / \pi_{it'} \) for \( t \neq t' \), \( \delta_{it} = I(R_{it} = 3) / \pi_{it} \), and \( \pi_{it'} = Pr(R_{it} = 1, R_{it'} = 3 | O_i, Z_i) + Pr(R_{it} = 3, R_{it'} = 3 | O_i, Z_i) \). In order to construct the weight matrix \( \Delta_i \) the conditional probabilities \( Pr(R_{it} = k | \bar{R}_{it}, O_i, Z_i) \) are decomposed as the product of two separated logistic models. The first one models the probability of observing the potentially missing covariate \( X_i \) whereas the latter models the probability of observing \( Y_{it} \) conditional on observed responses and covariates up to time \( t - 1 \).

The main idea of weighted generalized estimating equations (WGEE) lays on weighting the individual contribution to the estimating equation by introducing the weight matrix \( \Delta_i \) into the covariance matrix \( V_i \). This task can be accomplished by different ways
depending on the parametrization of the association structure. The general WGEE for $\beta$ are given by

$$\sum_{i=1}^{n} U_i(\beta, \alpha, \psi) = \sum_{i=1}^{n} D_i M_i(Y_i - \mu_i) = 0. \quad (8)$$

When the correlation coefficient is adopted, the $M_i$ matrix is decomposed as $M_i = F_i^{-1/2}(C_i^{-1} \cdot \Delta_i)F_i^{-1/2}$, where $A \cdot B = [a_{it} \cdot b_{it}]$ denotes the Hadamard product of matrices $A = [a_{it}]$ and $B = [b_{it}]$. Under the local odds approach, the matrix $M_i$ is given by $M_i = V_i^{-1} \cdot \Delta_i$.

Conditionally on a consistent estimate of the correlation/association structure $\alpha$, a consistent estimate for $\beta$ can be obtained by solving (8), under the correct specification of the missing data model.

In the presence of missing data a consistent estimate for the correlation parameters can be obtained by defining a weighted observed pair of Pearson residual vector as $e^*_it = e_{it}(I(R_{it} = 3)/\pi_{it})$. Then, a moment-based estimator can be constructed using the weighted pair contribution. For instance, an exchangeable correlation estimate can be obtained through

$$\hat{\rho}_{it} \cdot \sum_{i=1}^{n} T_i(T_i - 1) - p \sum_{i=1}^{n} e^*_it e^*_iT_i \pi_{it} \pi_{it}^{-1} \cdot \pi_{it}^{-1}.$$

It is easy to show that $\hat{\rho}_{it}$ is an unbiased estimator.

4 Doubly Robust GEE for Longitudinal Ordinal Data

Some authors (e.g., Scharfstein et al. (1999), Tsiatis (2006)) noted that adding a term of expectation zero, say $\phi(\cdot)$, to the inverse probability weighted estimators would still result in consistent estimates under the MAR mechanism. The solutions of these augmented estimating equations give rise to the so-called doubly robust estimators.

Following Chen & Zhou (2011), the optimal $\phi_{opt}$ for missing response and covariate is given by $\phi_{opt} = E_{Y^m_i, X^m_i |Y^t_i, X^t_i, Z_i, R_i} \{D_i N_i(Y_i - \mu_i)\}$, where $Y^m_i$ and $X^m_i$ denote the missing components of $Y_i$ and $X_i$, respectively. When the correlation coefficient parametrization is adopted, $N_i$ is defined as $N_i = F_i^{-1/2} \{C_i^{-1} \cdot (11^T - \Delta_i)\} F_i^{-1/2}$, where $1$ is a vector of $1$’s of length $T_i(J - 1)$. With local odds $N_i$ can be defined as $N_i = V_i^{-1} \cdot (11^T - \Delta_i)$.

Conditionally on a consistent estimate of the correlation/association structure $\alpha$, an improved estimate for $\beta$ can then be obtained by solving the estimating equations

$$\sum_{i=1}^{n} [D_i M_i(Y_i - \mu_i) + E_{Y^m_i, X^m_i |Y^t_i, X^t_i, Z_i, R_i} \{D_i N_i(Y_i - \mu_i)\}] = 0. \quad (9)$$

The estimator for $\beta$ in (9) is doubly-robust in the sense that it is consistent if at least one of the missing data model or the covariate model is correctly specified (Chen & Zhou, 2011).
The referred expectation in the second part of (9) is over the conditional distribution of \((Y^m_i, X^m_i|Y^o_i, X^o_i, Z_i, R_i)\), which can be written as

\[
P(Y^m_i = y^m_i, X^m_i = x^m_i|Y^o_i, X^o_i, Z_i, R_i; \beta^*, \gamma) = P(Y^m_i = y^m_i, X^m_i = x^m_i|Y^o_i, X^o_i, Z_i, R_i; \beta^*) \times P(Y^o_i = y^o_i, X^o_i = x^o_i|Y^o_i, X^o_i, Z_i; \gamma).
\]

The multivariate distribution \(P(Y^m_i = y^m_i, Y^o_i, X_i = x_i, Z_i; \beta^*)\) is expressed by a product of univariate ordinal models. When \(X\) is discrete, the second term in (9) can be written as

\[
E_{(Y^m_i, X^m_i|Y^o_i, X^o_i, Z_i, R_i)} \{ D_iN_i(Y_i - \mu_i) \} \sum_{(y^m_i, x^m_i)} w_{ixy} \{ D_iN_i(Y_i - \mu_i) \},
\]

where the weight \(w_{ixy}\) is given by

\[
w_{ixy} = P(Y^m_i = y^m_i|Y^o_i, X_i = x_i, Z_i; \beta^*) \times P(X^m_i = x^m_i|Y^o_i, X^o_i, Z_i; \gamma).
\]

In the case of \(X\) continuous, the second term in (9) takes the form

\[
E_{(Y^m_i, X^m_i|Y^o_i, X^o_i, Z_i, R_i)} \{ D_iN_i(Y_i - \mu_i) \} = \int_{(Y^m_i, X^m_i)} w_{ixy} \{ D_iN_i(Y_i - \mu_i) \} dY^m_i X^m_i,
\]

This expectation can be cumbersome, depending on the missing data pattern. In such case, instead of using numerical integration, a Monte Carlo method can be applied to approximate the corresponding integral.

Inspired by doubly robust ideas, we constructed the following estimator for the correlation structure

\[
\hat{\rho}_{it} = \frac{\omega}{n^t} \sum_{i=1}^{n^t} \sum_{t' > t} e_{it}^T e_{it'} \frac{\pi_{it} \pi_{it'}}{\pi_{it't}} + \frac{(1 - \omega)}{n^t} \sum_{i=1}^{n^t} \sum_{t' > t} \left[ \sum_{(y^m_i, x^m_i)} w_{ixy} e_{it}^T e_{it'} \right],
\]

where \(n^t = T_i(T_i - 1) - p\) and \(0 \leq \omega \leq 1\).

A sandwich estimator for the standard error of \(\hat{\beta}\) is given in Appendix.

## 5 Simulation Study

In this section a simulation study is presented in order to investigate the performance of the proposed method under the two parametrizations of the association vector as well as its robustness to misspecification of the predictive models. It is considered a study with \(T_i = T = 3\) repeated ordinal measures (with three categories) and two covariates (one quantitative and other qualitative). The true marginal model is

\[
\text{logit} \ Pr(O_{it} \leq j|X_{it}, Z_{it}) = \beta_{0j} + \beta_1 X_i + \beta_2 Z_{it}, \quad j = 1, 2.
\]

(10)

where \(Z_{it} \sim N(0, 1/2)\) for \(t = 1, 2, 3\).

The binary covariate \(X_{it}\) may be missing for some subjects and is generated according to

\[
\text{logit} \ Pr(X_i = 1|Z_{i1}) = \gamma_0 + \gamma_1 Z_{i1}.
\]

(11)
It is assumed that $\beta_{01} = -0.4$, $\beta_{02} = 1.2$, $\beta_1 = -0.35$, $\beta_2 = 0.35$, $\gamma_0 = \log(1)$, $\gamma_1 = 2$. The correlated ordinal responses were generated using the NORTA method (Anestis Touloumi, 2013) with constant correlation between the latent vectors set equal to $\rho = 0.7$.

In order to model $R_{it}$ two new indicators were defined. Let $R^*_t$ the indicator of observing $X_i$ and $R^c_{it}$ the indicator of observing $O_{it}$. The response variable in the first time occasion was allowed to be fully observed. The model for $R^*_t$ was defined as

$$
\log \left( \frac{Pr(R^*_t = 1)}{Pr(R^*_t = 0)} \right) = \psi^* + \psi^*_1 O_{i1} + \psi^*_2 Z_{i1},
$$

(12)

and the model for $R^c_{it}$ was taken as

$$
\log \left( \frac{Pr(R^c_{it} = 1)}{Pr(R^c_{it} = 0)} \right) = \psi^c_0 + \psi^c_1 O^*_{i,t-1} + \psi^c_2 I(R^c_{i,t-1} = 1) + \psi^c_3 Z_{it}, \quad t = 2, 3,
$$

(13)

where $O^*_{i,t-1} = O_{i,t-1}$, if $O_{i,t-1}$ is observed and 0 otherwise. The true values are taken as $\psi^*_1 = 1.2$, $\psi^*_1 = -1.5$, $\psi^*_2 = -1.5$, $\psi^c_0 = 0.6$, $\psi^c_1 = -1.5$, $\psi^c_2 = 2.5$, and $\psi^c_3 = -1.3$. It was observed about 30% of missing observations under this setup.

For comparison purposes, it was considered ordinary GEE for the complete and available data, respectively, weighted GEE (WGEE), multiple imputation (MIGEE) by chained equations (van Buuren & Groothuis-Oudshoorn, 2011) with $M = 10$ multiple imputations, and the doubly robust versions (DRGEE). The primary goal of this simulation was to compare the performance of the above mentioned methods under the correlation and local odds ratio parametrizations. Three correlation structures (independent – ind, exchangeable – exch, and unstructured – unst) and four local odds ratio structures (uniform – unif, category exchangeability – cat.exch, time exchangeability – time.exch, and unstructured – RC) were to be compared. Under independence, the estimates from the two parametrization are identical.

In order to investigate robustness of these methods, the predicted models were also misspecified by omitting the covariate $Z_1$ from the covariate model (11) or the missing data model (12).

Let $S = 1000$ the total number of Monte Carlo replications. Whenever, in a given iteration, a working association structure (C) failed to converge a new sample data were generated. Denote by $\hat{\beta}^C_\ell$ the corresponding GEE estimator at the $r$-th Monte Carlo replication and let $\bar{\beta}^C_\ell$ be the arithmetic mean, $\bar{\beta}^C = 1/S \sum_{r=1}^S \hat{\beta}^C_r$. To evaluate the consistency of the competing methods the relative bias, defined as $100 \times (\hat{\beta}^C - \beta)/\beta$, was calculated for each parameter. Interest is in quantifying the gain in efficiency by the working association structures over the independence structure. The Monte Carlo relative efficiency was defined as $\sum_{r=1}^S \hat{E}P(\hat{\beta}^C_r) / \sum_{r=1}^S \hat{E}P(\bar{\beta}^C)$, where $\hat{E}P(\hat{\beta}^C)$ is the standard error of $\hat{\beta}^C$ based on the estimated robust covariance matrix under the (C) working association structure. The estimated coverage probability for a nominal 95% level based on the asymptotic normality of the GEE estimators is also reported. Simulation was conducted for sample sizes $n = 50, 150, 300, 600$. Results are presented for sample size $n = 300$ subjects.

Table I presents the simulation results associated with ordinary GEE for available data and incorrectly specified methods, those in which the covariate $Z_1$ was omitted from their predictive models. For each method the first three lines refer to correlation
Table 1: Evaluation criteria for misspecified models. Results for $n = 300$ and $S = 1000$ simulations.

| Structure       | Relative Bias | Relative Efficiency | Empirical Coverage |
|-----------------|---------------|---------------------|--------------------|
|                 | $\beta_{01}$ | $\beta_{02}$ | $X$ | $Z$ |                 | $\beta_{01}$ | $\beta_{02}$ | $X$ | $Z$ |                 |
| Available       |               |                     | 1.00 | 1.00 | 1.00 | 0.15 | 0.36 | 0.93 | 0.73 |
| ind             | 121.1         | -31.5               | -23.9 | 60.2 | 1.02 | 1.01 | 1.01 | 1.20 | 0.35 | 0.55 | 0.94 | 0.86 |
| exch            | 94.0          | -25.2               | -26.4 | 32.3 | 1.03 | 1.01 | 1.02 | 1.19 | 0.31 | 0.48 | 0.90 | 0.90 |
| unst            | 94.3          | -25.9               | -31.4 | 26.7 | 1.02 | 0.99 | 1.01 | 1.23 | 0.33 | 0.57 | 0.92 | 0.88 |
| unif            | 93.7          | -23.9               | -27.8 | 26.4 | 1.02 | 0.99 | 1.01 | 1.24 | 0.33 | 0.57 | 0.92 | 0.89 |
| cat.exch        | 93.8          | -23.9               | -28.0 | 25.5 | 1.02 | 0.99 | 1.01 | 1.24 | 0.36 | 0.56 | 0.93 | 0.89 |
| time.exch       | 93.7          | -24.0               | -27.9 | 26.3 | 1.02 | 0.99 | 1.01 | 1.24 | 0.34 | 0.56 | 0.92 | 0.91 |
| RC              | 95.1          | -24.2               | -30.8 | 22.5 | 1.02 | 0.99 | 1.01 | 1.24 | 0.94 | 0.93 | 0.92 | 0.94 |
| Available       |               |                     | 1.00 | 1.00 | 1.00 | 0.94 | 0.93 | 0.92 | 0.89 |
| WGE($r^{-}$)    |               |                     | 1.00 | 1.00 | 1.00 | 0.94 | 0.94 | 0.94 | 0.91 |
| ind             | 19.9          | -5.4                | -28.9 | 32.8 | 1.00 | 0.99 | 1.00 | 1.17 | 0.93 | 0.93 | 0.92 | 0.94 |
| exch            | 17.9          | -5.1                | -29.9 | 22.2 | 1.01 | 0.99 | 1.01 | 1.16 | 0.93 | 0.92 | 0.91 | 0.94 |
| unst            | 21.0          | -6.5                | -35.0 | 15.2 | 1.01 | 0.99 | 1.01 | 1.16 | 0.95 | 0.94 | 0.92 | 0.94 |
| unif            | 18.8          | -5.1                | -28.6 | 22.7 | 1.00 | 0.99 | 1.00 | 1.16 | 0.95 | 0.94 | 0.92 | 0.94 |
| cat.exch        | 19.0          | -5.2                | -28.8 | 22.0 | 1.00 | 0.99 | 1.00 | 1.17 | 0.95 | 0.94 | 0.92 | 0.94 |
| time.exch       | 18.3          | -5.0                | -27.7 | 22.8 | 1.00 | 0.99 | 1.00 | 1.17 | 0.95 | 0.93 | 0.93 | 0.94 |
| RC              | 18.7          | -5.6                | -31.4 | 19.8 | 1.01 | 0.99 | 1.00 | 1.17 | 0.94 | 0.93 | 0.92 | 0.94 |
| Available       |               |                     | 1.00 | 1.00 | 1.00 | 0.93 | 0.93 | 0.92 | 0.94 |
| MIGE($x^{-}$)   |               |                     | 1.00 | 1.00 | 1.00 | 0.94 | 0.94 | 0.95 | 0.95 |
| ind             | 20.9          | -7.1                | -48.0 | -8.3 | 1.00 | 0.99 | 1.00 | 1.12 | 0.94 | 0.94 | 0.93 | 0.95 |
| exch            | 19.9          | -6.8                | -47.9 | -1.6 | 0.99 | 0.99 | 1.00 | 1.12 | 0.94 | 0.94 | 0.93 | 0.95 |
| unst            | 21.5          | -7.6                | -51.9 | -5.9 | 0.99 | 1.00 | 1.00 | 1.12 | 0.94 | 0.94 | 0.93 | 0.95 |
| unif            | 19.7          | -6.7                | -46.8 | -2.2 | 1.00 | 1.00 | 1.00 | 1.12 | 0.95 | 0.94 | 0.93 | 0.94 |
| cat.exch        | 19.9          | -6.8                | -47.0 | -3.1 | 1.00 | 1.00 | 1.00 | 1.12 | 0.95 | 0.94 | 0.93 | 0.94 |
| time.exch       | 20.0          | -6.8                | -46.6 | -2.8 | 1.00 | 1.00 | 1.00 | 1.12 | 0.94 | 0.94 | 0.94 | 0.94 |
| RC              | 20.1          | -7.3                | -49.7 | -5.8 | 1.00 | 1.01 | 1.00 | 1.12 | 0.94 | 0.93 | 0.92 | 0.93 |
| Available       |               |                     | 1.00 | 1.00 | 1.00 | 0.93 | 0.92 | 0.89 | 0.94 |
| DRGE($x^{+},r^{-}$) |           |                     | 1.00 | 1.00 | 1.00 | 0.94 | 0.94 | 0.91 | 0.94 |
| ind             | 17.0          | -5.5                | -41.4 | -8.9 | 1.00 | 1.00 | 1.00 | 1.14 | 0.93 | 0.92 | 0.89 | 0.94 |
| exch            | 15.7          | -5.1                | -40.7 | -2.4 | 0.99 | 0.99 | 1.00 | 1.14 | 0.94 | 0.94 | 0.91 | 0.94 |
| unst            | 18.5          | -6.3                | -45.5 | -8.5 | 1.00 | 1.00 | 1.01 | 1.14 | 0.93 | 0.92 | 0.89 | 0.95 |
| unif            | 15.5          | -5.0                | -39.1 | -3.1 | 1.00 | 0.99 | 1.00 | 1.14 | 0.94 | 0.94 | 0.91 | 0.94 |
| cat.exch        | 15.6          | -5.1                | -39.3 | -3.5 | 1.00 | 0.99 | 1.00 | 1.15 | 0.95 | 0.94 | 0.91 | 0.94 |
| time.exch       | 15.0          | -4.8                | -38.1 | -2.5 | 1.00 | 0.99 | 1.00 | 1.14 | 0.94 | 0.92 | 0.91 | 0.94 |
| RC              | 15.5          | -5.5                | -41.1 | -5.2 | 1.00 | 0.99 | 1.00 | 1.14 | 0.94 | 0.93 | 0.91 | 0.95 |

"\^\text{+}^\text{+}" indicates correctly specified model and "\^\text{-}^\text{-}^\text{-}\" indicates misspecified model omitting the $Z_{1}$ predictor.
structures and the last four ones refer to local odds ratio structures. These seven structures are distinguished in terms of the number of parameters being estimated as well as the restrictions placed on associations/correlations between the response indicators at different time pairs.

Regarding the GEE for available data, the missing data impact on bias can be noticed for all parameters, the higher biases being observed in the first intercept, followed by the parameter associated with the covariate $Z$. The impact of the bias on parameter estimates was also clearly noticed by the low coverage rates. Still considering the available data it is worth noting that the bias for the parameter associated with covariate $Z$ was reduced by more than half when the association structure is modeled. All methods being compared exhibited bias when their predictive models are incorrectly specified, although the bias for the two intercepts was considerably reduced. For the parameters associated with the covariates, the bias of WGEE method was the same magnitude as those provided by ordinary GEE, and the performance of DRGEE was slightly superior to MIGEE.

Independent estimating equations are efficient for the intercept parameters and regression coefficient associated with the baseline covariate $X$. As expected, the gain in efficiency by modeling the association structure occurs only for the time-varying covariate $Z$. Comparing to the independence structure, the gain in efficiency ranged from 14% on average for DRGEE, 23% for multiple imputation and about 17% for WGEE.

All misspecified methods presented empirical coverage rates below nominal level although they are somewhat close to the expected value in some cases, particularly for MIGEE.

Table 2 presents simulation results for complete data in addition to correctly specified methods. In terms of bias there is no clear distinction between the two approaches, except when an unstructured matrix is chosen, which causes an increase in bias, especially for WGEE and DRGEE when only the weight model is correctly specified. For the scenario under consideration, the results suggest that the performances of both WGEE and DRGEE are slightly superior to multiple imputation, especially for estimates associated with covariate $X$.

Comparing to independence structure, all other association structures presented smaller standard errors for the time-dependent covariate $Z$. Although they are very similar within each method, the largest gain in efficiency, around 23%, is obtained for the MIGEE method, followed by WGEE with 19%. For DRGEE estimators it ranged from 14% to 18%.

All correctly specified methods showed coverage rates close to the nominal levels for all association structures. In terms of empirical bias, relative efficiency and empirical coverage there was no clear distinction between the correlation and local odds parametrizations for $n = 300$ subjects. For the scenario under consideration, an exchangeable correlation structure or an uniform local odds structure resulted in good marginal mean estimates.

Simulation results for $n = 50, 150$ and 600 are presented in the Appendix. With sample sizes $n = 50$ or $n = 150$ subjects, simulation results suggest that the local odds parametrization outperforms the correlation coefficient parametrization in terms of bias and convergence issues.
Table 2: Evaluation criteria for correctly specified models. Results for $n = 300$ and $S = 1000$ simulations.

| Structure | $β_{01}$ | $β_{02}$ | $X$ | $Z$ | $β_{01}$ | $β_{02}$ | $X$ | $Z$ |
|-----------|----------|----------|-----|-----|----------|----------|-----|-----|
| Complete  |          |          |     |     |          |          |     |     |
| ind       | 2.1      | 0.1      | -1.1| -0.5| 1.00     | 1.00     | 1.00| 1.00|
| exch      | 0.3      | 0.6      | -0.2| 1.3 | 0.99     | 0.99     | 0.99| 1.25|
| unst      | 0.9      | 0.0      | -3.4| -1.7| 1.00     | 1.00     | 1.01| 1.25|
| unif      | 0.1      | 0.7      | 0.7 | 0.3 | 1.00     | 1.00     | 1.00| 1.27|
| cat.exch  | 0.2      | 0.7      | 0.6 | -0.2| 1.00     | 1.00     | 1.00| 1.27|
| time.exch | 0.2      | 0.3      | 0.1 | 0.3 | 1.00     | 1.00     | 1.00| 1.27|
| RC        | 0.2      | 0.3      | -0.8| -1.9| 1.00     | 1.00     | 1.00| 1.27|
| WGEE($r^+$)|         |          |     |     |          |          |     |     |
| ind       | 3.6      | -0.4     | -1.8| -0.7| 1.00     | 1.00     | 1.00| 1.00|
| exch      | 1.3      | 0.3      | 0.2 | 0.6 | 0.99     | 0.98     | 1.00| 1.19|
| unst      | 6.2      | -1.6     | -6.8| -4.1| 1.00     | 1.00     | 1.01| 1.25|
| unif      | 2.4      | -0.1     | -0.3| -0.4| 0.99     | 0.99     | 1.00| 1.18|
| cat.exch  | 2.6      | -0.2     | -0.5| -1.0| 0.99     | 0.99     | 1.00| 1.18|
| time.exch | 1.4      | 0.3      | 1.8 | 1.1 | 0.99     | 0.99     | 1.00| 1.18|
| RC        | 2.5      | -0.5     | -2.5| -1.6| 0.99     | 0.99     | 1.00| 1.18|
| MIGEE($x^+$)|         |          |     |     |          |          |     |     |
| ind       | 5.3      | -1.3     | -7.1| -0.4| 1.00     | 1.00     | 1.00| 1.00|
| exch      | 3.5      | -0.8     | -5.6| -0.7| 0.99     | 0.99     | 1.00| 1.22|
| unst      | 4.6      | -1.5     | -9.4| -3.6| 0.99     | 0.99     | 1.00| 1.21|
| unif      | 2.4      | -0.1     | -0.3| -0.4| 0.99     | 0.99     | 1.00| 1.22|
| cat.exch  | 4.2      | -0.9     | -5.7| -0.7| 0.99     | 0.99     | 1.00| 1.23|
| time.exch | 4.2      | -1.0     | -5.2| -0.2| 0.99     | 0.99     | 1.00| 1.23|
| RC        | 3.7      | -1.3     | -8.0| -3.1| 0.99     | 0.99     | 1.00| 1.23|
| DRGEE($x^+, r^+$)| | | | | | | | |
| ind       | 1.0      | 0.4      | -1.7| -1.4| 1.00     | 1.00     | 1.00| 1.00|
| exch      | -0.8     | 0.9      | 0.3 | 1.4 | 0.99     | 0.99     | 1.00| 1.16|
| unst      | 4.6      | -1.5     | -9.4| -3.7| 0.99     | 0.99     | 1.00| 1.21|
| unif      | 0.0      | 0.7      | -0.6| -0.1| 1.00     | 1.00     | 1.00| 1.16|
| cat.exch  | 0.1      | 0.6      | -0.8| -0.6| 1.00     | 1.00     | 1.00| 1.17|
| time.exch | -0.9     | 1.2      | 2.0 | 1.4 | 0.99     | 0.99     | 1.00| 1.17|
| RC        | -0.2     | 0.3      | -2.3| -1.8| 1.00     | 1.00     | 1.00| 1.16|
| DRGEE($x^+, r^+$)| | | | | | | | |
| ind       | 2.2      | 0.0      | -4.7| -1.5| 1.00     | 1.00     | 1.00| 1.00|
| exch      | 0.2      | 0.6      | -2.5| 1.3 | 0.98     | 0.99     | 1.00| 1.17|
| unst      | 4.7      | -1.3     | -11.1| -3.7| 1.00     | 1.01     | 1.03| 1.18|
| unif      | 1.2      | 0.3      | -3.3| -0.1| 1.00     | 1.00     | 1.00| 1.17|
| cat.exch  | 1.3      | 0.3      | -3.5| -0.5| 1.00     | 1.00     | 1.00| 1.18|
| time.exch | 0.3      | 0.7      | -1.2| 1.3 | 1.00     | 1.00     | 1.00| 1.18|
| RC        | 1.2      | -0.1     | -5.5| -1.6| 1.00     | 1.00     | 1.00| 1.17|
| DRGEE($x^-, r^+$)| | | | | | | | |
| ind       | 1.0      | 0.2      | -2.5| -1.8| 1.00     | 1.00     | 1.00| 1.00|
| exch      | -0.4     | 0.6      | -1.4| 0.5 | 0.99     | 0.99     | 1.00| 1.15|
| unst      | 2.1      | -0.5     | -5.7| -5.6| 1.00     | 1.00     | 1.01| 1.14|
| unif      | -0.2     | 0.7      | -0.6| 0.0 | 1.00     | 1.00     | 1.00| 1.15|
| cat.exch  | -0.1     | 0.6      | -0.8| -0.5| 1.00     | 1.00     | 1.00| 1.15|
| time.exch | -0.4     | 0.8      | 0.1 | 0.5 | 1.00     | 1.00     | 1.00| 1.14|
| RC        | 0.0      | 0.1      | -3.3| -2.2| 1.00     | 1.00     | 1.00| 1.14|

**+** indicates correctly specified model and **-** indicates misspecified model omitting the $Z_1$ predictor.

Relative Bias, Relative Efficiency, and Empirical Coverage are all given in terms of the empirical coverage of the confidence intervals for $β_{01}$ and $β_{02}$. The $X$ and $Z$ columns indicate the presence of the respective predictors in the model.
6 Data Analysis: Functional Classification in Rheumatic Mitral Stenosis

A cohort of 164 patients with rheumatic mitral stenosis who were referred for treatment at Hospital das Clinicas of the Federal University of Minas Gerais, Brazil, was selected for a mitral valve intervention. Patients were included before intervention and then followed up in the outpatient clinic every 4 months according to their clinical status. The first three measurements were available for analysis.

Mitral stenosis is a narrowing of the mitral valve in the heart caused by rheumatic disease, which restricts the flow of blood through the valve. The main clinical manifestation of this disease is shortness of breath, classified in four categories based on how much the patients are limited during physical activity. The response of interest is The New York Heart Association (NYHA) Functional Classification, that provides a simple way of classifying the extent of shortness of breath. Patients with no symptoms and no limitation in ordinary physical activity were classified in class I; slight limitation of physical activity in class II; marked limitation of physical activity in class III; and patients with severe limitations resulting in inability to carry on any physical activity without discomfort in class IV. Only one patient were classified into class IV in the followup evaluation and hence class IV was combined with class III in the analysis. Thus, the ordinal response was defined as (1: if class I, 2: if class II, 3 if class III or class IV).

Percutaneous mitral valvuloplasty (PMV) is an effective treatment for stretching the stenosed mitral valve. This procedure is carried out by inserting a catheter with a balloon at its tip to open the narrowed mitral valve. This procedure causes improvement of the functional class in the majority of the patients. A number of patient’s characteristics were measured at baseline, such as atrial compliance (Cn: defined as 1, if ≤ 4, and 0 otherwise), cardiac rhythm, morphological features of the mitral valve expressed as an echocardiographic score, mitral valve area, pressure transmitrval gradients, and pulmonary artery pressure. Some variables measured after the procedure include the success of the procedure to open the mitral valve without complications, long-termevent-free survival, mitral valve area, pressure transmitral gradients, and pulmonary artery pressure at the follow-up appointment.

This study was characterized by an arbitrary pattern of missing data. The response were fully observed for 125 patients, 29 had only the first measurement, 1 patient had only the first data collected, and for 9 patients the second occasion was missing. There was no missing data in response at baseline. Some collected variables, such as success of the procedure and long term events were responsible for the missingness at followup. A baseline covariate of particular interest, atrial compliance, were missing for 54(32.9%) of the patients. Among the reasons for not observing such predictor it can be included morphological characteristics of the mitral valve and valve calcification. Therefore a MAR mechanism seems to be a reasonable assumption for this data set.

Now, the models used for analysis are described. For the ordinal response it was used the following proportional odds model

\[
\text{logit } Pr(NYHA_{itj} \leq j | u_{it}) = \beta_{0j} + u_{it}^T \beta, \quad j = 1, 2, \quad t = 1, 2, 3, \quad (14)
\]

where \( u_{it} \) is the covariate vector at time \( t \), and it is formed by time, atrial compliance, echocardiographic score, and success of the procedure.
When using WGEE or DRGGE it is necessary to correctly model the missingness mechanism in order to obtain consistent estimates of $\beta$. For the missing data process, $R^y_{it}$ was defined as the indicator of observing the response $NYHA_{it}$ and $R^x_i$ was defined as the indicator of observing the baseline atrial compliance. The probability of observing the response was modeled as

$$\text{logit} \ Pr(R^y_{it} = 1) = \psi^y_0 + w^y_{it} \psi^y, \ t = 2, 3,$$

where the vector $w^y_{it}$ included success of the procedure, long term events, and the previous indicator of missing data. The model for $R^x_i$ was specified as

$$\text{logit} \ Pr(R^x_i = 1) = \psi^x_0 + w^x_{it} \psi^x,$$

where the vector $w^x_{it}$ included the baseline response, echocardiographic score, valve calcification and ECG rhythm.

For the covariate, the following model was built

$$\text{logit} \ Pr(Cn = 1) = v^T_i \gamma,$$

where $v^T_i$ included the baseline right ventricular systolic pressure, mean gradient and mitral valve area. A imputation model for NYHA was also specified that included the covariates in (14) in addition to the baseline right ventricular systolic pressure and the response history.

Here, four methods were used to analyze the data. Results are shown in Table 3. The first method is the usual GEE method using the available data; the second is the weighted method (WGEE) using models (15) and (16) for the weights; the third is the multiple imputation by chained equation (MIGEE); and the fourth, labeled DRGEE, is the doubly robust method using (15) and (16) for the missing data process and (17) for the covariate models. In order to account for the dependence structure of the repeated measures the correlation coefficient and the local odds ratio parametrization were both initially applied with the same association structures presented in the simulation. Convergence issues were observed for unstructured correlation matrix. Motivated by simulation results that suggested better performance of the local odds ratio structures, results are shown only for the independence, uniform and category exchangeability local odds ratio structures. The independence structure estimates no parameters, while the uniform represents the dependence structure by a single parameter. The category exchangeability assumes constant odds ratio among the levels of the response, but different odds ratio between time pairs. Thus, the association is explained by three parameters in this structure. Similar results were obtained for the other dependence structures they are shown in Appendix.

The significance of time effect indicates the effectiveness of the intervention and improvement of functional class over time. Similar coefficients for the two followup occasions suggest the major change in functional class occurs right after the valvuloplasty intervention. Total score was non significant for all the four methods, although a marginally significant ($p = 0.060$) was noted for MIGEE when the category exchangeability odds ratio structure was adopted, as an indication that higher scores may be related to cardiac insufficiency. All methods provide the same conclusion for effects of the missing covariate Cn. It can be noticed that estimates for success of the procedure effect goes from non significant in the standard GEE to significant for all methods, except WGEE, regardless
Table 3: Results for Rheumatic Mitral Stenosis study under independence, uniform and category exchangeability association structures

| Parameter  | Independence | Uniform | Cat. Exchangeability |
|------------|--------------|---------|----------------------|
|            | Est. | SE | p       | Est. | SE | p       | Est. | SE | p       |
| Available  |      |    |         |      |    |         |      |    |         |
| β_{01}     | -0.800 | 0.765 | 0.295  | -0.838 | 0.735 | 0.254  | -0.768 | 0.751 | 0.307  |
| β_{02}     | 0.997  | 0.745 | 0.181  | 0.954  | 0.726 | 0.189  | 1.037  | 0.746 | 0.164  |
| Success    | 0.733  | 0.441 | 0.096  | 0.662  | 0.391 | 0.090  | 0.692  | 0.400 | 0.083  |
| Total Score| -0.147 | 0.104 | 0.158  | -0.142 | 0.101 | 0.161  | -0.156 | 0.103 | 0.130  |
| Cn         | 0.533  | 0.328 | 0.104  | 0.498  | 0.315 | 0.113  | 0.499  | 0.318 | 0.116  |
| time=2     | 1.003  | 0.413 | 0.015  | 1.084  | 0.389 | 0.005  | 1.126  | 0.390 | 0.004  |
| time=3     | 1.216  | 0.420 | 0.004  | 1.214  | 0.390 | 0.002  | 1.198  | 0.394 | 0.002  |
|            |      |    |         |      |    |         |      |    |         |
|            |      |    |         |      |    |         |      |    |         |
|            |      |    |         |      |    |         |      |    |         |
|            |      |    |         |      |    |         |      |    |         |
| β_{01}     | -0.852 | 0.786 | 0.278  | -0.883 | 0.763 | 0.247  | -0.792 | 0.776 | 0.307  |
| β_{02}     | 0.985  | 0.768 | 0.244  | 0.882  | 0.748 | 0.238  | 0.974  | 0.767 | 0.204  |
| Success    | 0.750  | 0.449 | 0.094  | 0.732  | 0.405 | 0.071  | 0.749  | 0.416 | 0.072  |
| Total Score| -0.139 | 0.107 | 0.192  | -0.136 | 0.104 | 0.190  | -0.153 | 0.106 | 0.148  |
| Cn         | 0.524  | 0.336 | 0.118  | 0.464  | 0.325 | 0.153  | 0.472  | 0.327 | 0.148  |
| time=2     | 1.030  | 0.413 | 0.013  | 1.039  | 0.401 | 0.010  | 1.085  | 0.406 | 0.007  |
| time=3     | 1.269  | 0.419 | 0.002  | 1.352  | 0.404 | 0.001  | 1.331  | 0.410 | 0.001  |
|            |      |    |         |      |    |         |      |    |         |
|            |      |    |         |      |    |         |      |    |         |
|            |      |    |         |      |    |         |      |    |         |
|            |      |    |         |      |    |         |      |    |         |
| β_{01}     | -0.856 | 0.582 | 0.141  | -0.805 | 0.580 | 0.165  | -0.788 | 0.572 | 0.169  |
| β_{02}     | 0.988  | 0.582 | 0.090  | 1.019  | 0.574 | 0.076  | 1.035  | 0.572 | 0.070  |
| Success    | 0.938  | 0.367 | 0.010  | 0.922  | 0.333 | 0.006  | 0.980  | 0.338 | 0.004  |
| Total Score| -0.141 | 0.080 | 0.077  | -0.140 | 0.079 | 0.075  | -0.147 | 0.078 | 0.060  |
| Cn         | 0.425  | 0.305 | 0.163  | 0.307  | 0.261 | 0.239  | 0.343  | 0.267 | 0.199  |
| time=2     | 0.995  | 0.334 | 0.003  | 1.106  | 0.308 | 0.000  | 1.127  | 0.330 | 0.001  |
| time=3     | 0.969  | 0.338 | 0.004  | 1.068  | 0.304 | 0.000  | 1.079  | 0.333 | 0.001  |
|            |      |    |         |      |    |         |      |    |         |
|            |      |    |         |      |    |         |      |    |         |
|            |      |    |         |      |    |         |      |    |         |
|            |      |    |         |      |    |         |      |    |         |
| β_{01}     | -0.793 | 0.730 | 0.277  | -0.830 | 0.711 | 0.243  | -0.726 | 0.711 | 0.308  |
| β_{02}     | 0.952  | 0.722 | 0.187  | 0.934  | 0.707 | 0.186  | 1.039  | 0.712 | 0.145  |
| Success    | 0.825  | 0.422 | 0.050  | 0.806  | 0.382 | 0.035  | 0.841  | 0.388 | 0.030  |
| Total Score| -0.150 | 0.099 | 0.131  | -0.147 | 0.097 | 0.132  | -0.165 | 0.098 | 0.092  |
| Cn         | 0.571  | 0.341 | 0.094  | 0.520  | 0.329 | 0.113  | 0.523  | 0.333 | 0.116  |
| time=2     | 1.013  | 0.499 | 0.042  | 1.005  | 0.512 | 0.050  | 1.020  | 0.525 | 0.052  |
| time=3     | 1.135  | 0.476 | 0.017  | 1.185  | 0.473 | 0.012  | 1.145  | 0.491 | 0.020  |
the adopted association structure. It is interesting to note that the estimated effect are increased for all methods, especially for the multiple imputation procedure. So, for example, for the uniform structure, the estimated odds of response in class I for subjects with success of the procedure compared to patients with suboptimal results were \( e^{0.806} = 2.24 \) for DRGEE and \( e^{0.922} = 2.51 \) for MIGEE. Regarding the association structures it can be noticed that the uniform and category exchangeability choices presented some gain in efficiency, specially for the success of procedure effect. Similar conclusions were reached for the other association structures.

7 Discussion

In this paper it was considered a doubly robust estimator for analysis of longitudinal data when missingness can occur in a baseline covariate or intermittently in the ordinal response. The main objective was to compare the performance of the proposed method in terms of bias and efficiency for two different approaches to modeling the covariance matrix of the longitudinal outcome, namely, the correlation coefficient proposed by Lipsitz et al. (1994) and local odds ratio proposed by Touloumis et al. (2013). Although a complete data comparison between these two distinct approaches had already appeared elsewhere (Nooraee et al., 2014) such a comparison with MAR missing data was still in need of investigation.

The covariate design plays an important role in the efficiency of GEE estimators. The working independence structure is expected to be efficient for time-stationary covariates. This is no longer true for time-varying covariates and/or missing data (Lipsitz et al., 1994). Simulation results agreed with the literature, that is, the gain in efficiency by adopting more complicated association structures was noticed for the time-varying covariate and it somewhat varied through the different methods. For complete data, the standard error for the independence was, on average, about 27% larger compared to a uniform structure that estimates a single association parameter. The gain in efficiency reached 23% for MIGEE, 19% for WGEE and 18% for DRGEE.

As Liang & Zeger (1986) pointed out when the assumed correlation is the true correlation, the missing completely at random assumption can be unnecessary. However, this is not true when missingness occurs in a covariate that is MAR given the response. In this case even likelihood methods are biased (Carpenter & Kenward, 2013) and the missingness mechanism must be modeled in order to obtain valid estimates for the regression parameters.

The dependence structures compared here differ in terms of number of parameters and restrictions imposed on the correlation/association between levels of ordinal response at different time pairs. Although the same definition can be used for an independence, exchangeable and unstructured association matrix, this does not imply that identical associations are fit (Nooraee et al., 2014). Under the independence working assumption all off-diagonal blocks of the covariance matrix are constant and equal to zero. Exchangeability over time indicates that the association between \( Y_{ij} \) and \( Y_{i'j'} \) is independent of time, but it depends on the levels \( j \) and \( j' \). For unstructured associations there are no restrictions implied. With moderate to large number of subjects (at least 300 subjects) the simulation results did not suggest a very clear distinction in terms of bias, relative efficiency and empirical coverage between the correlation or local odds ratio parametriza-
tions, although some bias was observed for unstructured matrices. On the other side, for small sample sizes (n = 50 or 150) simulation results did suggest that the local odds ratio outperforms the correlation coefficient parametrization in terms of bias and convergence issues. It seems that local odds structures work fine for small samples sizes and the correlation coefficient needs relatively more subjects to achieve the same reduction in bias.

It is noticeable two important differences between the local odds ratio and the correlation coefficient parametrizations. Firstly, the local odds ratios does not depend on the marginal specification, that is, $\beta$ and $\alpha$ are variation independent (Touloumis et al., 2013). Unlike correlations, the local odds estimates for the association structure does not depend on the values of the observed covariates and thus these estimates do not need to be obtained at each step of the modified Fisher Scoring algorithm. As a consequence, the iterative procedure converges faster. Secondly, the correlation estimates are obtained through a moment based estimator, thus some work was necessary to ensure their validity in the presence of data that is MAR. The estimates of local odds are based on maximum likelihood methods, assuming independent Poisson sampling for the observed counts in the marginalized contingency tables. As it considers only the observed responses the resulting estimates are valid under a MAR mechanism.

In some situations (small samples, complex patterns of missing data), it is possible that the algorithm will not converge because the estimated correlation matrix is not guaranteed to be positive definite (Lipsitz et al., 1994). For small sample sizes there may not be enough data to estimate both the regression parameters and a correlation matrix that is highly unstructured. Simulation results for unstructured correlation matrices and small sample sizes presented very low convergence rates (in addition to increased bias) and thus an exchangeable correlation matrix is preferred. There were no serious convergence issues with the local odds ratio parametrization, except for an unstructured matrix applied to $n = 50$.

Although the latent vectors have been generated from an exchangeable correlation, the correlation coefficient between binary variables $Y_{itj}$ and $Y_{it'j'}$ also depends on the linear predictor in times $t$ and $t'$ (Nooraei et al., 2014). That is, the correlation between the latent vectors was exchangeable, but the correlation between the binary variables was not exchangeable due to the dependent mean in the marginal model. Nevertheless, it is expected the GEE method remain valid even assuming a misspecified correlation matrix provided the marginal mean is correctly specified (Molenberghs & Kenward, 2010).

Overall, the doubly robust method performed well for all working association structures under comparison except when applied to a very small number of subjects (i.e., $n = 50$) and an unstructured matrix was adopted. The coverage probabilities were relatively close to the nominal level of 95%.

The doubly robust estimator considered here is restricted to missingness in the response and a baseline covariate. A natural extension is to consider an intermittently missing time-varying covariate and/or allow multiple missing covariates. In the proposed approach the marginal means were modeled by cumulative logits. This implies a proportional odds model that in some cases may not be valid. Another possible extension of the proposed model is, therefore, to allow non-proportional odds for a subset of the explanatory variables (Peterson & Harrell Jr, 1990).
References

Agresti, Alan. 2013. *Categorical data analysis*. John Wiley & Sons.

Anestis Touloumis. 2013. *SimCorMultRes: Simulating Correlated Multinomial Responses*. Version 1.1.

Becker, Mark P, & Clogg, Clifford C. 1989. Analysis of sets of two-way contingency tables using association models. *Journal of the American Statistical Association, 84*(405), 142–151.

Carey, Vincent, Zeger, Scott L, & Diggle, Peter. 1993. Modelling multivariate binary data with alternating logistic regressions. *Biometrika, 80*(3), 517–526.

Carpenter, James R., & Kenward, Michael G. 2013. *Multiple Imputation and Its Application*. Wiley & Sons.

Carpenter, James R, Kenward, Michael G, & Vansteelandt, Stijn. 2006. A comparison of multiple imputation and doubly robust estimation for analyses with missing data. *Journal of the Royal Statistical Society: Series A (Statistics in Society), 169*(3), 571–584.

Chen, Baojiang, & Zhou, Xiao-Hua. 2011. Doubly robust estimates for binary longitudinal data analysis with missing response and missing covariates. *Biometrics, 67*(3), 830–842.

Deming, W Edwards, & Stephan, Frederick F. 1940. On a least squares adjustment of a sampled frequency table when the expected marginal totals are known. *The Annals of Mathematical Statistics, 11*(4), 427–444.

Heagerty, Patrick J, & Zeger, Scott L. 1996. Marginal regression models for clustered ordinal measurements. *Journal of the American Statistical Association, 91*(435), 1024–1036.

Liang, Kung-Yee, & Zeger, Scott L. 1986. Longitudinal data analysis using generalized linear models. *Biometrika, 73*(1), 13–22.

Lipsitz, Stuart R, Kim, Kyungmann, & Zhao, Lueping. 1994. Analysis of repeated categorical data using generalized estimating equations. *Statistics in medicine, 13*(11), 1149–1163.

Little, Roderick JA, & Rubin, Donald B. 1987. *Statistical analysis with missing data*. Vol. 539. Wiley New York.

Lumley, Thomas. 1996. Generalized estimating equations for ordinal data: a note on working correlation structures. *Biometrics, 354–361.*

McCullagh, Peter. 1980. Regression models for ordinal data. *Journal of the royal statistical society. Series B (Methodological), 109–142.*

Molenberghs, Geert, & Kenward, Michael G. 2010. Semi-parametric marginal models for hierarchical data and their corresponding full models. *Computational Statistics & Data Analysis, 54*(2), 585–597.
Nooraee, Nazanin, Molenberghs, Geert, & van den Heuvel, Edwin R. 2014. GEE for longitudinal ordinal data: Comparing R-geepack, R-multgee, R-repolr, SAS-GENMOD, SPSS-GENLIN. *Computational Statistics & Data Analysis*, **77**, 70–83.

Parsons, Nicholas R, Edmondson, RN, & Gilmour, SG. 2006. A generalized estimating equation method for fitting autocorrelated ordinal score data with an application in horticultural research. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, **55**(4), 507–524.

Peterson, Bercedis, & Harrell Jr, Frank E. 1990. Partial proportional odds models for ordinal response variables. *Applied Statistics*, 205–217.

Pierce, Donald A. 1982. The asymptotic effect of substituting estimators for parameters in certain types of statistics. *The Annals of Statistics*, 475–478.

Robins, James M, Rotnitzky, Andrea, & Zhao, Lue Ping. 1995. Analysis of semiparametric regression models for repeated outcomes in the presence of missing data. *Journal of the American Statistical Association*, **90**(429), 106–121.

Rubin, Donald B. 1976. Inference and missing data. *Biometrika*, **63**, 581–592.

Rubin, Donald B. 1987. *Multiple Imputation for Nonresponse in Surveys*. John Wiley.

Scharfstein, Daniel O, Rotnitzky, Andrea, & Robins, James M. 1999. Adjusting for non-ignorable drop-out using semiparametric nonresponse models. *Journal of the American Statistical Association*, **94**(448), 1096–1120.

Seaman, Shaun, & Copas, Andrew. 2009. Doubly robust generalized estimating equations for longitudinal data. *Statistics in medicine*, **28**(6), 937–955.

Touloumis, Anestis, Agresti, Alan, & Kateri, Maria. 2013. GEE for Multinomial Responses Using a Local Odds Ratios Parameterization. *Biometrics*, 1–8.

Tsiatis, Anastasios A. 2006. *Semiparametric theory and missing data*. Springer.

van Buuren, S., Boshuizen, H. C., & Knook, D. L. 1999. Multiple Imputation of Missing Blood Pressure Covariates in Survival Analysis. *Statistics in Medicine*, **18**, 681–694.

van Buuren, Stef. 2007. Multiple imputation of discrete and continuous data by fully conditional specification. *Statistical methods in medical research*, **16**(3), 219–242.

van Buuren, Stef, & Groothuis-Oudshoorn, Karin. 2011. mice: Multivariate Imputation by Chained Equations in R. *Journal of Statistical Software*, **45**(3), 1–67.
A Appendix

A.1 Asymptotic Variance

To state the asymptotic properties of \( \hat{\beta} \), let \( S_{1i}(\beta, \psi, \gamma) \) be the individual’s contribution to the estimating equations for \( \beta \), \( S_{2i}(\psi) \) be the individual’s contribution to the estimating equations for \( \psi \), and \( S_{3i}(\gamma) \) be the individual’s contribution to the estimating equations for \( \gamma \).

Define \( \Gamma(\beta, \psi, \gamma) = E \{ \partial S_{1i}(\beta, \psi, \gamma)/\partial \beta^T \} \), \( I_{12}(\beta, \psi, \gamma) = E \{ \partial S_{1i}(\beta, \psi, \gamma)/\partial \gamma^T \} \), \( I_{2}(\psi) = E \{ \partial S_{2i}(\psi)/\partial \psi^T \} \), \( I_{3}(\gamma) = E \{ \partial S_{3i}(\gamma)/\partial \gamma^T \} \), and \( Q_{i}(\beta, \psi, \gamma) = S_{1i}(\beta, \psi, \gamma) - I_{12}(\beta, \psi, \gamma) I_{2}^{-1}(\psi) S_{2i}(\psi) - I_{13}(\beta, \psi, \gamma) I_{3}^{-1}(\gamma) S_{3i}(\gamma) \).

**Theorem 1** If either the missing data model or the covariate model is correctly specified, then

\[
n^{1/2}(\hat{\beta} - \beta_0) \longrightarrow N(0, \Gamma^{-1}(\beta_0, \psi_0, \gamma_0) \Sigma \{ \Gamma^{-1}(\beta_0, \psi_0, \gamma_0) \}^T),
\]

where \( \beta_0 \) is the true value of \( \beta \), \( \psi_0 \) and \( \gamma_0 \) are the probability limits of \( \hat{\psi} \) and \( \hat{\gamma} \), and \( \Sigma = E \{ Q_{i}(\beta_0, \psi_0, \gamma_0) Q_{i}^T(\beta_0, \psi_0, \gamma_0) \} \).

Inferences for \( \beta \) follows by replacing the unknown quantities in (18) by its consistent estimators. We make use of “generalized information equality” (Pierce, 1982) that

\[
E \{ \partial S_{1i}(\beta, \psi, \gamma)/\partial \psi^T \} = -E \{ S_{1i}(\beta, \psi, \gamma) S_{2i}^T(\psi) \},
\]

and

\[
E \{ \partial S_{1i}(\beta, \psi, \gamma)/\partial \gamma^T \} = -E \{ S_{1i}(\beta, \psi, \gamma) S_{3i}^T(\gamma) \}.\]

Similarly (Robins et al., 1995),

\[
E \{ \partial S_{2i}(\psi)/\partial \psi^T \} = -\text{Var} \{ S_{2i}(\psi) \}, \quad \text{and} \quad E \{ \partial S_{3i}(\gamma)/\partial \gamma^T \} = -\text{Var} \{ S_{3i}(\gamma) \}.
\]

The matrix \( \Gamma \) is replaced by \( \hat{\Gamma} = n^{-1} \sum_{i=1}^{n} \{ \partial S_{1i}(\hat{\theta})/\partial \beta^T \} \), and \( \Sigma \) by \( \hat{\Sigma} = n^{-1} \sum_{i=1}^{n} \{ \hat{Q}_{i}, \hat{Q}_{i}^T \} \),

\[
\hat{Q}_{i} = S_{1i}(\hat{\theta}) - \hat{I}_{12}(\hat{\theta}) \hat{I}_{2}^{-1}(\hat{\psi}) S_{2i}(\hat{\psi}) - \hat{I}_{13}(\hat{\theta}) \hat{I}_{3}^{-1}(\hat{\gamma}) S_{3i}(\hat{\gamma}), \quad \hat{I}_{12}(\hat{\theta}) = n^{-1} \sum_{i=1}^{n} \{ \partial S_{1i}(\hat{\theta})/\partial \beta^T \},
\]

\[
\hat{I}_{13}(\hat{\theta}) = n^{-1} \sum_{i=1}^{n} \{ \partial S_{1i}(\hat{\theta})/\partial \gamma^T \}, \quad \hat{I}_{2}(\hat{\psi}) = n^{-1} \sum_{i=1}^{n} \{ \partial S_{2i}(\hat{\psi})/\partial \psi^T \},
\]

\[
\hat{I}_{3}(\hat{\gamma}) = n^{-1} \sum_{i=1}^{n} \{ \partial S_{3i}(\hat{\gamma})/\partial \gamma^T \}.
\]

The proof is similar to Chen & Zhou (2011) and is omitted here.
### A.2 Additional Simulation Results

Table 4: Evaluation criteria for misspecified models. Results for \( n = 50 \) and \( S = 1000 \) simulations.

| Structure | Relative Bias | Relative Efficiency | Empirical Coverage |
|-----------|---------------|---------------------|--------------------|
|           | \( \beta_0 \) | \( \beta_0 \) | Available | \( \beta_0 \) | \( \beta_0 \) | X | Z | \( \beta_0 \) | \( \beta_0 \) | X | Z | \( \beta_0 \) | \( \beta_0 \) | X | Z |
| ind       | 101.9        | -25.2              | -10.0              | 48.4              | 1.00 | 1.00 | 1.00 | 1.00 | 0.85 | 0.85 | 1.00 | 0.95 |
| exch      | 96.2         | -22.6              | -36.0              | 19.5              | 1.02 | 1.03 | 1.03 | 1.15 | 0.87 | 0.88 | 1.00 | 0.95 |
| unst      | 91.0         | -21.8              | -44.1              | -1.3              | 1.08 | 1.10 | 1.10 | 1.17 | 0.86 | 0.90 | 1.00 | 0.96 |
| unif      | 80.2         | -18.6              | -14.8              | 26.5              | 1.01 | 0.99 | 1.00 | 1.19 | 0.91 | 0.90 | 1.00 | 0.94 |
| cat.exch  | 80.9         | -19.0              | -16.8              | 19.5              | 1.01 | 1.00 | 1.01 | 1.23 | 0.90 | 0.89 | 1.00 | 0.94 |
| time.exch | 82.8         | -18.2              | -16.9              | 25.1              | 1.00 | 0.99 | 1.01 | 1.20 | 0.88 | 0.90 | 1.00 | 0.93 |
| RC        | 83.6         | -20.0              | -19.2              | 9.1               | 1.01 | 1.00 | 1.01 | 1.26 | 0.95 | 0.93 | 1.00 | 0.96 |
|           | 24.5         | -0.7               | -13.5              | 30.3              | 1.00 | 1.00 | 1.00 | 1.00 | 0.94 | 0.93 | 1.00 | 0.96 |
| exch      | 36.3         | -5.2               | -29.7              | 14.2              | 1.01 | 1.02 | 1.03 | 1.14 | 0.94 | 0.94 | 1.00 | 0.95 |
| unst      | 43.3         | -7.0               | -44.7              | -3.0              | 1.07 | 1.11 | 1.09 | 1.17 | 0.93 | 0.93 | 1.00 | 0.96 |
| unif      | 17.9         | 0.6                | -12.6              | 24.8              | 1.00 | 0.99 | 0.99 | 1.14 | 0.96 | 0.94 | 1.00 | 0.96 |
| cat.exch  | 19.2         | 0.0                | -11.6              | 19.0              | 1.00 | 0.99 | 0.99 | 1.15 | 0.95 | 0.93 | 1.00 | 0.95 |
| time.exch | 18.1         | 0.6                | -12.6              | 23.6              | 0.98 | 0.98 | 0.99 | 1.15 | 0.96 | 0.95 | 1.00 | 0.94 |
| RC        | 22.3         | -1.4               | -12.7              | 9.9               | 0.99 | 0.99 | 0.99 | 1.17 | 0.97 | 0.95 | 1.00 | 0.95 |
|           | 25.6         | -5.0               | -34.7              | -2.2              | 1.00 | 1.00 | 1.00 | 1.00 | 0.96 | 0.95 | 1.00 | 0.96 |
| exch      | 35.4         | -7.8               | -53.5              | -8.3              | 1.00 | 1.01 | 1.00 | 1.18 | 0.95 | 0.97 | 1.00 | 0.94 |
| unst      | 35.2         | -6.7               | -59.9              | -28.8             | 1.03 | 1.06 | 1.05 | 1.16 | 0.96 | 0.97 | 1.00 | 0.95 |
| unif      | 23.0         | -5.1               | -36.4              | 2.4               | 1.00 | 0.99 | 1.00 | 1.19 | 0.97 | 0.96 | 1.00 | 0.94 |
| cat.exch  | 24.8         | -5.8               | -35.3              | -3.5              | 0.99 | 0.99 | 1.00 | 1.24 | 0.96 | 0.96 | 1.00 | 0.95 |
| time.exch | 22.9         | -4.0               | -36.0              | -1.2              | 0.98 | 0.99 | 0.99 | 1.21 | 0.97 | 0.96 | 1.00 | 0.92 |
| RC        | 25.5         | -5.5               | -35.8              | -16.3             | 1.00 | 0.99 | 1.00 | 1.28 | 0.95 | 0.94 | 1.00 | 0.93 |
|           | 21.1         | -1.4               | -19.0              | -4.0              | 1.00 | 1.00 | 1.00 | 1.00 | 0.94 | 0.93 | 1.00 | 0.95 |
| exch      | 33.8         | -6.0               | -35.9              | -14.4             | 1.02 | 1.03 | 1.04 | 1.18 | 0.93 | 0.94 | 1.00 | 0.92 |
| unst      | 38.2         | -7.8               | -55.3              | -34.2             | 1.08 | 1.13 | 1.13 | 1.20 | 0.93 | 0.93 | 1.00 | 0.93 |
| unif      | 15.4         | -0.1               | -18.2              | -1.0              | 1.00 | 0.98 | 0.99 | 1.15 | 0.96 | 0.94 | 1.00 | 0.94 |
| cat.exch  | 16.3         | -0.7               | -18.9              | -4.8              | 1.00 | 0.98 | 0.99 | 1.17 | 0.95 | 0.93 | 1.00 | 0.94 |
| time.exch | 15.5         | -0.1               | -17.1              | -2.5              | 0.99 | 0.98 | 1.00 | 1.17 | 0.96 | 0.94 | 1.00 | 0.92 |
| RC        | 20.1         | -1.7               | -20.1              | -15.2             | 0.99 | 0.98 | 0.99 | 1.21 | 0.95 | 0.94 | 1.00 | 0.91 |

"\( + \)" indicates correctly specified model and "\( - \)" indicates misspecified model omitting the \( Z_1 \) predictor.
Table 5: Evaluation criteria for correctly specified models. Results for $n = 50$ and $S = 1000$ simulations.

| Structure | $\beta_0$ | $\beta_2$ | $X$ | $Z$ | Relative Bias | Relative Efficiency | Empirical Coverage |
|-----------|-----------|-----------|-----|-----|---------------|---------------------|-------------------|
|           | $\beta_0$ | $\beta_2$ | $X$ | $Z$ |               |                     |                   |
| Complete  |           |           |     |     |               |                     |                   |
| ind       | 3.4       | 4.4       | 3.3 | 3.5 | 1.00          | 1.00                | 0.95              |
| exch      | 13.6      | 1.3       | -12.4 | -8.0 | 0.99          | 1.01                | 1.01              |
| unif      | 17.0      | 2.3       | -25.5 | -18.9 | 1.03          | 1.06                | 1.05              |
| cat.exch  | 0.9       | 4.2       | 3.6 | -1.8 | 1.00          | 1.00                | 1.00              |
| time.exch | 3.3       | 4.8       | -2.6 | 1.6 | 0.99          | 0.99                | 1.00              |
| RC        | 5.1       | 3.0       | -0.6 | -8.1 | 1.00          | 1.00                | 1.00              |

|           | $\beta_0$ | $\beta_2$ | $X$ | $Z$ |               |                     |                   |
| Complete  |           |           |     |     |               |                     |                   |
| ind       | 14.8      | 3.2       | 1.8 | 9.1 | 1.00          | 1.00                | 0.95              |
| exch      | 27.7      | -1.7      | -16.3 | -1.6 | 1.01          | 1.03                | 1.03              |
| unif      | 37.5      | -4.7      | -29.3 | -13.5 | 1.08          | 1.12                | 1.11              |
| cat.exch  | 8.1       | 5.1       | 4.7 | 10.1 | 1.00          | 0.98                | 0.99              |
| time.exch | 6.9       | 5.7       | 8.6 | 5.7 | 0.97          | 0.99                | 0.99              |
| RC        | 12.4      | 2.9       | 0.8 | -5.9 | 0.98          | 0.99                | 0.99              |

|           | $\beta_0$ | $\beta_2$ | $X$ | $Z$ |               |                     |                   |
| Complete  |           |           |     |     |               |                     |                   |
| ind       | 12.1      | 0.6       | -5.1 | 6.5 | 1.00          | 1.00                | 0.96              |
| exch      | 23.7      | -2.9      | -20.3 | -7.3 | 0.99          | 1.01                | 1.01              |
| unif      | 26.8      | -3.3      | -32.4 | -27.0 | 1.02          | 1.06                | 1.04              |
| cat.exch  | 7.5       | 0.9       | 0.0 | 5.1 | 1.00          | 1.00                | 1.00              |
| time.exch | 7.8       | 1.6       | -2.1 | 2.1 | 0.99          | 0.99                | 1.00              |
| RC        | 13.4      | -0.1      | -4.9 | -14.0 | 1.00          | 0.99                | 0.99              |

|           | $\beta_0$ | $\beta_2$ | $X$ | $Z$ |               |                     |                   |
| Complete  |           |           |     |     |               |                     |                   |
| ind       | 7.0       | 4.1       | 0.1 | 2.4 | 1.00          | 1.00                | 0.94              |
| exch      | 19.3      | -0.5      | -13.9 | -10.0 | 1.01          | 1.03                | 1.04              |
| unif      | 25.0      | -2.5      | -28.3 | -23.3 | 1.08          | 1.13                | 1.13              |
| cat.exch  | 3.9       | 5.5       | 3.7 | 4.5 | 0.99          | 0.98                | 0.99              |
| time.exch | 2.7       | 5.7       | 10.5 | 0.8 | 0.97          | 0.97                | 0.99              |
| RC        | 9.0       | 3.2       | 1.0 | -5.9 | 0.97          | 0.97                | 0.98              |

|           | $\beta_0$ | $\beta_2$ | $X$ | $Z$ |               |                     |                   |
| Complete  |           |           |     |     |               |                     |                   |
| ind       | 11.6      | 3.3       | -2.4 | 2.3 | 1.00          | 1.00                | 0.94              |
| exch      | 25.8      | -2.0      | -22.2 | -9.0 | 1.02          | 1.06                | 1.06              |
| unif      | 31.8      | -5.0      | -39.5 | -26.1 | 1.09          | 1.18                | 1.18              |
| cat.exch  | 5.4       | 5.1       | -1.3 | 3.2 | 1.00          | 0.98                | 1.01              |
| time.exch | 7.3       | 4.5       | -2.7 | -1.3 | 1.00          | 0.98                | 1.01              |
| RC        | 11.2      | 3.1       | -4.7 | -12.5 | 0.98          | 0.97                | 0.99              |

|           | $\beta_0$ | $\beta_2$ | $X$ | $Z$ |               |                     |                   |
| Complete  |           |           |     |     |               |                     |                   |
| ind       | 8.9       | 3.9       | -2.0 | 2.0 | 1.00          | 1.00                | 0.93              |
| exch      | 18.3      | -0.5      | -14.8 | -12.9 | 1.01          | 1.03                | 1.03              |
| unif      | 23.7      | -2.3      | -28.5 | -20.3 | 1.06          | 1.11                | 1.11              |
| cat.exch  | 5.7       | 4.2       | 2.4 | -2.4 | 0.99          | 0.98                | 0.99              |
| time.exch | 3.3       | 5.3       | 4.4 | 1.3 | 0.98          | 0.98                | 0.99              |
| RC        | 10.5      | 3.4       | -2.2 | -11.9 | 0.99          | 0.98                | 0.99              |

**+** indicates correctly specified model and **−** indicates misspecified model omitting the $Z_1$ predictor.
Table 6: Evaluation criteria for misspecified models. Results for $n = 150$ and $S = 1000$ simulations.

| Structure | $\beta_0$ | $\beta_0$ | $X$ | $Z$ | $\beta_0$ | $\beta_0$ | $X$ | $Z$ | $\beta_0$ | $\beta_0$ | $X$ | $Z$ |
|-----------|-----------|-----------|-----|-----|-----------|-----------|-----|-----|-----------|-----------|-----|-----|
| Available |           |           |     |     |           |           |     |     |           |           |     |     |
| ind       | 117.0     | -29.9     | -20.9 | 60.5 | 1.00      | 1.00      | 1.00 | 1.00 | 0.46      | 0.66      | 0.94 | 0.84 |
| exch      | 96.4      | -25.5     | -32.2 | 30.2 | 1.02      | 1.01      | 1.02 | 1.20 | 0.58      | 0.7       | 0.92 | 0.92 |
| unst      | 97.4      | -26.3     | -40.0 | 17.4 | 1.04      | 1.03      | 1.04 | 1.21 | 0.59      | 0.71      | 0.94 | 0.94 |
| unif      | 92.5      | -22.6     | -25.9 | 26.8 | 1.02      | 1.00      | 1.01 | 1.23 | 0.63      | 0.78      | 0.93 | 0.91 |
| cat.exch  | 92.6      | -22.7     | -26.4 | 24.8 | 1.02      | 1.00      | 1.01 | 1.24 | 0.64      | 0.77      | 0.93 | 0.92 |
| time.exch | 93.6      | -22.8     | -25.6 | 30.0 | 1.02      | 1.00      | 1.01 | 1.23 | 0.61      | 0.76      | 0.94 | 0.91 |
| RC        | 93.2      | -22.4     | -25.1 | 24.1 | 1.02      | 1.00      | 1.01 | 1.25 | 0.62      | 0.78      | 0.94 | 0.92 |
| WGEE($r^-$) |           |           |     |     |           |           |     |     |           |           |     |     |
| ind       | 19.0      | -3.4      | -22.9 | 37.5 | 1.00      | 1.00      | 1.00 | 1.00 | 0.94      | 0.94      | 0.94 | 0.9 |
| exch      | 21.5      | -5.5      | -32.9 | 22.2 | 1.00      | 1.00      | 1.00 | 1.16 | 0.94      | 0.93      | 0.94 | 0.94 |
| unst      | 27.3      | -7.4      | -44.2 | 10.0 | 1.02      | 1.02      | 1.02 | 1.17 | 0.94      | 0.94      | 0.95 | 0.94 |
| unif      | 18.7      | -3.4      | -24.0 | 24.7 | 1.00      | 1.00      | 1.00 | 1.16 | 0.94      | 0.94      | 0.94 | 0.93 |
| cat.exch  | 19.1      | -3.6      | -24.1 | 23.0 | 1.00      | 1.00      | 1.00 | 1.17 | 0.94      | 0.94      | 0.94 | 0.93 |
| time.exch | 19.0      | -4.0      | -25.3 | 26.4 | 1.00      | 1.00      | 1.00 | 1.16 | 0.94      | 0.94      | 0.94 | 0.92 |
| RC        | 20.3      | -4.0      | -24.8 | 20.9 | 1.00      | 1.00      | 1.00 | 1.16 | 0.94      | 0.96      | 0.94 | 0.94 |
| MIGEE($x^-$) |           |           |     |     |           |           |     |     |           |           |     |     |
| ind       | 21.2      | -5.9      | -46.4 | -7.4 | 1.00      | 1.00      | 1.00 | 1.00 | 0.94      | 0.95      | 0.94 | 0.95 |
| exch      | 22.7      | -7.5      | -53.4 | -2.2 | 0.98      | 0.99      | 0.99 | 1.21 | 0.95      | 0.95      | 0.95 | 0.96 |
| unst      | 24.9      | -8.1      | -57.2 | -10.0 | 0.99      | 1.00      | 1.00 | 1.21 | 0.95      | 0.96      | 0.96 | 0.94 |
| unif      | 20.8      | -5.9      | -45.9 | -1.8 | 1.00      | 1.00      | 1.00 | 1.22 | 0.94      | 0.95      | 0.94 | 0.93 |
| cat.exch  | 21.3      | -6.1      | -46.4 | -3.7 | 1.00      | 1.00      | 1.00 | 1.24 | 0.94      | 0.95      | 0.94 | 0.93 |
| time.exch | 20.1      | -6.2      | -44.8 | 0.4  | 1.00      | 1.00      | 1.00 | 1.22 | 0.95      | 0.96      | 0.96 | 0.94 |
| RC        | 20.8      | -6.1      | -45.3 | -3.8 | 1.00      | 1.00      | 1.00 | 1.25 | 0.95      | 0.96      | 0.96 | 0.94 |
| DRGEE($x^-, r^-$) |           |           |     |     |           |           |     |     |           |           |     |     |
| ind       | 16.6      | -3.5      | -32.9 | -5.4 | 1.00      | 1.00      | 1.00 | 1.00 | 0.93      | 0.94      | 0.94 | 0.94 |
| exch      | 19.4      | -5.5      | -42.8 | -4.9 | 1.00      | 1.00      | 1.01 | 1.13 | 0.93      | 0.93      | 0.93 | 0.95 |
| unst      | 24.7      | -7.4      | -54.6 | -16.1 | 1.01      | 1.02      | 1.03 | 1.16 | 0.93      | 0.93      | 0.93 | 0.93 |
| unif      | 16.0      | -3.3      | -31.9 | -1.7 | 1.00      | 1.00      | 0.99 | 1.14 | 0.93      | 0.94      | 0.95 | 0.92 |
| cat.exch  | 16.4      | -3.5      | -32.2 | -2.9 | 1.00      | 1.00      | 0.99 | 1.16 | 0.94      | 0.94      | 0.94 | 0.92 |
| time.exch | 16.0      | -3.8      | -34.2 | 0.6  | 1.00      | 1.00      | 1.00 | 1.16 | 0.93      | 0.94      | 0.93 | 0.95 |
| RC        | 17.0      | -4.0      | -35.3 | -5.7 | 1.00      | 1.00      | 1.00 | 1.17 | 0.94      | 0.95      | 0.94 | 0.94 |

"++" indicates correctly specified model and "---" indicates misspecified model omitting the $Z_1$ predictor.
Table 7: Evaluation criteria for correctly specified models. Results for n = 150 and S = 1000 simulations.

| Structure | $\beta_0$ | $\beta_1$ | $\beta_2$ | $\beta_0$ | $\beta_1$ | $\beta_2$ | $\beta_0$ | $\beta_1$ | $\beta_2$ |
|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|
| Complete  |           |           |           | Complete  |           |           | Complete  |           |           |
| ind       | 0.2       | 1.9       | 2.9       | 1.0       | 1.0       | 1.0       | 0.94      | 0.94      | 0.95      |
| exch      | 0.9       | 0.8       | -3.8      | 0.7       | 0.99      | 1.00      | 0.93      | 0.93      | 0.95      |
| unst      | 3.4       | 0.4       | -8.5      | -6.1      | 1.01      | 1.01      | 0.95      | 0.95      | 0.97      |
| cat.exch  | -0.1      | 2.0       | 3.2       | 1.6       | 1.00      | 1.00      | 0.94      | 0.94      | 0.95      |
| time.exch | 0.3       | 1.6       | 3.3       | 2.9       | 1.00      | 1.00      | 0.95      | 0.95      | 0.94      |
| RC        | 0.7       | 1.8       | 1.3       | -0.1      | 1.01      | 1.01      | 0.95      | 0.95      | 0.96      |

WGEE($r^+$) |           |           |           |           | Complete  |           |           |           |           |
| ind       | 3.3       | 1.5       | 3.9       | 6.0       | 1.00      | 1.00      | 0.96      | 0.95      | 0.95      |
| exch      | 5.7       | -0.5      | -7.1      | 1.2       | 0.99      | 1.00      | 0.94      | 0.94      | 0.96      |
| unst      | 13.3      | -2.8      | -18.2     | -7.9      | 1.02      | 1.02      | 0.96      | 0.97      | 0.95      |
| cat.exch  | 2.6       | 1.6       | 4.0       | 3.7       | 0.99      | 0.99      | 0.95      | 0.96      | 0.94      |
| time.exch | 3.0       | 1.4       | 3.9       | 2.5       | 0.99      | 0.99      | 0.95      | 0.96      | 0.94      |
| RC        | 4.4       | 0.9       | 1.5       | -0.3      | 1.00      | 0.99      | 0.95      | 0.96      | 0.95      |

MIGEE($) |           |           |           |           | Complete  |           |           |           |           |
| ind       | 5.4       | 0.1       | -3.3      | 0.9       | 1.00      | 1.00      | 0.96      | 0.95      | 0.95      |
| exch      | 7.7       | -1.8      | -14.5     | 0.3       | 0.99      | 1.00      | 0.94      | 0.94      | 0.96      |
| unst      | 9.6       | -2.6      | -18.9     | -7.5      | 1.02      | 1.02      | 0.96      | 0.97      | 0.95      |
| cat.exch  | 5.0       | 0.1       | -3.0      | 1.4       | 0.99      | 0.99      | 0.95      | 0.95      | 0.93      |
| time.exch | 4.7       | -0.3      | -4.1      | 3.3       | 1.00      | 1.00      | 0.95      | 0.95      | 0.94      |
| RC        | 5.3       | 0.0       | -3.7      | -0.8      | 1.00      | 1.00      | 0.95      | 0.96      | 0.94      |

DRGEE($) |           |           |           |           | Complete  |           |           |           |           |
| ind       | 0.3       | 2.5       | 5.0       | 1.9       | 1.00      | 1.00      | 0.95      | 0.94      | 0.95      |
| exch      | 3.0       | 0.3       | -5.7      | -0.1      | 0.98      | 0.99      | 0.94      | 0.94      | 0.96      |
| unst      | 8.9       | -1.7      | -16.5     | -9.6      | 1.00      | 1.01      | 0.95      | 0.95      | 0.97      |
| cat.exch  | -0.2      | 2.5       | 5.5       | 1.9       | 1.00      | 1.00      | 0.95      | 0.94      | 0.95      |
| time.exch | 4.7       | -0.3      | -4.1      | 3.3       | 1.00      | 1.00      | 0.95      | 0.95      | 0.94      |
| RC        | 5.3       | 0.0       | -3.7      | -0.8      | 1.00      | 1.00      | 0.95      | 0.96      | 0.94      |

+ indicates correctly specified model and - indicates misspecified model omitting the $Z_1$ predictor.
Table 8: Evaluation criteria for misspecified models. Results for $n = 600$ and $S = 1000$ simulations.

| Structure | Relative Bias $\beta_01$ | Relative Efficiency $\beta_02$ | Empirical Coverage $\beta_01$ | Empirical Coverage $\beta_02$ |
|-----------|--------------------------|-------------------------------|-------------------------------|-------------------------------|
|           | $X$ | $Z$ | $X$ | $Z$ | $X$ | $Z$ |
| Available | 1.00 | 1.00 | 1.00 | 1.00 | 0.01 | 0.08 | 0.92 | 0.50 |
| ind       | 119.9 | -31.8 | -22.5 | 61.8 | 1.02 | 1.01 | 1.01 | 1.21 | 0.08 | 0.27 | 0.89 | 0.76 |
| exch      | 95.2 | -24.9 | -25.1 | 30.8 | 1.03 | 1.02 | 1.02 | 1.21 | 0.09 | 0.20 | 0.88 | 0.79 |
| unst      | 94.3 | -26.2 | -27.9 | 29.8 | 1.02 | 1.00 | 1.01 | 1.24 | 0.08 | 0.26 | 0.90 | 0.82 |
| unif      | 93.5 | -24.5 | -28.5 | 25.3 | 1.02 | 1.00 | 1.01 | 1.24 | 0.08 | 0.26 | 0.90 | 0.83 |
| cat.exch  | 93.5 | -24.5 | -28.5 | 24.7 | 1.02 | 1.00 | 1.01 | 1.24 | 0.08 | 0.26 | 0.90 | 0.83 |
| time.exch | 94.9 | -24.7 | -29.4 | 25.3 | 1.02 | 1.00 | 1.01 | 1.25 | 0.07 | 0.28 | 0.89 | 0.83 |
| RC        | 93.4 | -24.0 | -26.9 | 26.0 | 1.02 | 1.00 | 1.01 | 1.24 | 0.08 | 0.31 | 0.89 | 0.82 |
|           |     |     |     |     |     |     |     |     |     |     |     |     |
| WGEE($p^-$) | 19.1 | -5.6 | -28.1 | 34.8 | 1.00 | 1.00 | 1.00 | 1.00 | 0.92 | 0.92 | 0.91 | 0.82 |
|           |     |     |     |     |     |     |     |     |     |     |     |     |
| MIGEE($x^-$) | 19.9 | -7.2 | -47.9 | -8.1 | 1.00 | 1.00 | 1.00 | 1.00 | 0.92 | 0.91 | 0.86 | 0.92 |
|           |     |     |     |     |     |     |     |     |     |     |     |     |
| DRGEE($x^+, r^-$) | 18.7 | -5.2 | -27.9 | 22.1 | 1.00 | 0.99 | 0.99 | 1.18 | 0.92 | 0.92 | 0.89 | 0.88 |
|           |     |     |     |     |     |     |     |     |     |     |     |     |
| RC       | 15.8 | -5.9 | -41.1 | -6.7 | 1.00 | 1.00 | 1.00 | 1.00 | 0.92 | 0.91 | 0.86 | 0.92 |
|           |     |     |     |     |     |     |     |     |     |     |     |     |
| RC       | 15.2 | -5.2 | -38.6 | -2.0 | 1.00 | 0.99 | 0.99 | 1.14 | 0.92 | 0.91 | 0.86 | 0.95 |

$++$ indicates correctly specified model and $-$ indicates misspecified model omitting the $Z_1$ predictor.
Table 9: Evaluation criteria for correctly specified models. Results for \( n = 600 \) and \( S = 1000 \) simulations.

| Structure | \( \beta_{01} \) | \( \beta_{02} \) | \( \beta_{11} \) | \( \beta_{12} \) | \( \beta_{01} \) | \( \beta_{02} \) | \( \beta_{11} \) | \( \beta_{12} \) | Empirical Coverage |
|-----------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|
| **Complete** | | | | | | | | | |
| **ind** | 0.9 | 0.1 | 1.0 | 1.0 | 0.96 | 0.96 | 0.95 | 0.95 | 0.96 |
| **exch** | 0.9 | 0.1 | 1.0 | 1.0 | 0.97 | 0.96 | 0.95 | 0.95 | 0.96 |
| **unif** | 0.9 | 0.1 | 1.0 | 1.0 | 0.97 | 0.96 | 0.95 | 0.95 | 0.96 |
| **cat.exch** | 0.9 | 0.1 | 1.0 | 1.0 | 0.97 | 0.96 | 0.95 | 0.95 | 0.96 |
| **time.exch** | 0.9 | 0.1 | 1.0 | 1.0 | 0.97 | 0.96 | 0.95 | 0.95 | 0.96 |
| **RC** | 1.0 | 0.1 | 1.0 | 1.0 | 0.97 | 0.96 | 0.95 | 0.95 | 0.96 |
| **WGEE(\( r^+ \))** | | | | | | | | | |
| **ind** | 2.2 | 0.1 | 1.0 | 1.0 | 0.98 | 0.97 | 0.97 | 0.97 | 0.95 |
| **exch** | 2.2 | 0.1 | 1.0 | 1.0 | 0.97 | 0.96 | 0.95 | 0.95 | 0.96 |
| **unif** | 2.2 | 0.1 | 1.0 | 1.0 | 0.97 | 0.96 | 0.95 | 0.95 | 0.96 |
| **cat.exch** | 2.2 | 0.1 | 1.0 | 1.0 | 0.97 | 0.96 | 0.95 | 0.95 | 0.96 |
| **time.exch** | 2.2 | 0.1 | 1.0 | 1.0 | 0.97 | 0.96 | 0.95 | 0.95 | 0.96 |
| **RC** | 2.4 | 0.1 | 1.0 | 1.0 | 0.97 | 0.96 | 0.95 | 0.95 | 0.96 |
| **MIGEE(\( x^+ \))** | | | | | | | | | |
| **ind** | 3.6 | 0.1 | 1.0 | 1.0 | 0.97 | 0.96 | 0.95 | 0.95 | 0.96 |
| **exch** | 3.6 | 0.1 | 1.0 | 1.0 | 0.97 | 0.96 | 0.95 | 0.95 | 0.96 |
| **unif** | 3.6 | 0.1 | 1.0 | 1.0 | 0.97 | 0.96 | 0.95 | 0.95 | 0.96 |
| **cat.exch** | 3.6 | 0.1 | 1.0 | 1.0 | 0.97 | 0.96 | 0.95 | 0.95 | 0.96 |
| **time.exch** | 3.6 | 0.1 | 1.0 | 1.0 | 0.97 | 0.96 | 0.95 | 0.95 | 0.96 |
| **RC** | 3.6 | 0.1 | 1.0 | 1.0 | 0.97 | 0.96 | 0.95 | 0.95 | 0.96 |
| **DRGEE(\( x^+ \), \( r^+ \))** | | | | | | | | | |
| **ind** | 0.9 | 0.1 | 1.0 | 1.0 | 0.97 | 0.96 | 0.95 | 0.95 | 0.96 |
| **exch** | 0.9 | 0.1 | 1.0 | 1.0 | 0.97 | 0.96 | 0.95 | 0.95 | 0.96 |
| **unif** | 0.9 | 0.1 | 1.0 | 1.0 | 0.97 | 0.96 | 0.95 | 0.95 | 0.96 |
| **cat.exch** | 0.9 | 0.1 | 1.0 | 1.0 | 0.97 | 0.96 | 0.95 | 0.95 | 0.96 |
| **time.exch** | 0.9 | 0.1 | 1.0 | 1.0 | 0.97 | 0.96 | 0.95 | 0.95 | 0.96 |
| **RC** | 0.9 | 0.1 | 1.0 | 1.0 | 0.97 | 0.96 | 0.95 | 0.95 | 0.96 |
| **DRGEE(\( x^+ \), \( r^- \))** | | | | | | | | | |
| **ind** | 0.9 | 0.1 | 1.0 | 1.0 | 0.97 | 0.96 | 0.95 | 0.95 | 0.96 |
| **exch** | 0.9 | 0.1 | 1.0 | 1.0 | 0.97 | 0.96 | 0.95 | 0.95 | 0.96 |
| **unif** | 0.9 | 0.1 | 1.0 | 1.0 | 0.97 | 0.96 | 0.95 | 0.95 | 0.96 |
| **cat.exch** | 0.9 | 0.1 | 1.0 | 1.0 | 0.97 | 0.96 | 0.95 | 0.95 | 0.96 |
| **time.exch** | 0.9 | 0.1 | 1.0 | 1.0 | 0.97 | 0.96 | 0.95 | 0.95 | 0.96 |
| **RC** | 0.9 | 0.1 | 1.0 | 1.0 | 0.97 | 0.96 | 0.95 | 0.95 | 0.96 |

**"++" indicates correctly specified model and "\(-\)" indicates misspecified model omitting the \( Z_1 \) predictor.**
Table 10: Convergence Rate (CR) for the Simulation Study

| sample size | 50  | 150 | 300 | 600 |
|-------------|-----|-----|-----|-----|
| ind         | 0.97| 1.00| 1.00| 1.00|
| exch        | 0.59| 0.89| 0.96| 1.00|
| unst        | 0.08| 0.59| 0.89| 0.99|
| unif        | 0.97| 1.00| 1.00| 1.00|
| cat.exch    | 0.97| 1.00| 1.00| 1.00|
| time.exch   | 0.96| 1.00| 1.00| 1.00|
| RC          | 0.67| 0.99| 1.00| 1.00|

Table 10 shows the convergence rate obtained from the simulation results for seven association structures and four sample sizes. The convergence rate for the working association structure (C) was calculated as

\[
CR^C = \frac{1000}{\text{total number of simulations}}.
\]

The local odds ratio parametrization presents a clear advantage over the correlation coefficient in terms of convergence issues, specially for unstructured association matrices and small sample sizes.
A.3 Additional Real Data Results
Table 11: Results for Rheumatic Mitral Stenosis study under exchangeability, time exchangeability and RC association structures

| Parameter | Exchangeable | Time Exchangeability | RC |
|-----------|--------------|----------------------|-----|
|           | Est. | SE  | p  | Est. | SE  | p  | Est. | SE  | p  |
| Available |      |     |    |      |     |    |      |     |    |
| $\beta_{01}$ | -1.064 | 0.782 | 0.174 | -0.909 | 0.730 | 0.213 | -0.881 | 0.744 | 0.236 |
| $\beta_{02}$ | 0.721 | 0.776 | 0.353 | 0.890 | 0.724 | 0.219 | 0.924 | 0.738 | 0.210 |
| Success   | 0.680 | 0.397 | 0.086 | 0.626 | 0.387 | 0.105 | 0.641 | 0.391 | 0.101 |
| Total Score | -0.108 | 0.107 | 0.313 | -0.130 | 0.101 | 0.197 | -0.137 | 0.102 | 0.179 |
| Cn        | 0.468 | 0.339 | 0.168 | 0.477 | 0.314 | 0.129 | 0.496 | 0.317 | 0.118 |
| time=2    | 1.083 | 0.398 | 0.007 | 1.022 | 0.383 | 0.008 | 1.081 | 0.390 | 0.006 |
| time=3    | 1.208 | 0.402 | 0.003 | 1.120 | 0.380 | 0.003 | 1.214 | 0.392 | 0.002 |
| WGEE      |      |     |    |      |     |    |      |     |    |
| $\beta_{01}$ | -1.178 | 0.807 | 0.144 | -0.934 | 0.756 | 0.217 | -0.892 | 0.771 | 0.248 |
| $\beta_{02}$ | 0.611 | 0.797 | 0.443 | 0.848 | 0.741 | 0.252 | 0.898 | 0.757 | 0.236 |
| Success   | 0.744 | 0.410 | 0.070 | 0.623 | 0.400 | 0.119 | 0.645 | 0.404 | 0.111 |
| Total Score | -0.101 | 0.109 | 0.355 | -0.127 | 0.103 | 0.216 | -0.136 | 0.105 | 0.195 |
| Cn        | 0.490 | 0.351 | 0.162 | 0.433 | 0.322 | 0.179 | 0.449 | 0.325 | 0.167 |
| time=2    | 1.037 | 0.414 | 0.012 | 1.013 | 0.398 | 0.011 | 1.060 | 0.404 | 0.009 |
| time=3    | 1.395 | 0.418 | 0.001 | 1.363 | 0.405 | 0.001 | 1.437 | 0.415 | 0.001 |
| MIGEE     |      |     |    |      |     |    |      |     |    |
| $\beta_{01}$ | -1.100 | 0.618 | 0.075 | -0.865 | 0.573 | 0.131 | -0.908 | 0.567 | 0.109 |
| $\beta_{02}$ | 0.743 | 0.621 | 0.231 | 0.973 | 0.575 | 0.091 | 0.934 | 0.572 | 0.102 |
| Success   | 1.079 | 0.334 | 0.001 | 0.834 | 0.322 | 0.010 | 0.887 | 0.324 | 0.006 |
| Total Score | -0.098 | 0.083 | 0.239 | -0.131 | 0.078 | 0.094 | -0.131 | 0.078 | 0.093 |
| Cn        | 0.237 | 0.280 | 0.397 | 0.281 | 0.265 | 0.288 | 0.350 | 0.272 | 0.198 |
| time=2    | 1.095 | 0.317 | 0.001 | 1.075 | 0.320 | 0.001 | 1.033 | 0.323 | 0.001 |
| time=3    | 1.071 | 0.316 | 0.001 | 1.091 | 0.308 | 0.000 | 1.072 | 0.313 | 0.001 |
| DRGEE     |      |     |    |      |     |    |      |     |    |
| $\beta_{01}$ | -1.199 | 0.781 | 0.125 | -0.896 | 0.692 | 0.195 | -0.852 | 0.700 | 0.224 |
| $\beta_{02}$ | 0.591 | 0.785 | 0.451 | 0.888 | 0.686 | 0.195 | 0.934 | 0.696 | 0.180 |
| Success   | 0.808 | 0.380 | 0.034 | 0.727 | 0.368 | 0.048 | 0.748 | 0.371 | 0.044 |
| Total Score | -0.097 | 0.108 | 0.368 | -0.137 | 0.094 | 0.148 | -0.143 | 0.096 | 0.135 |
| Cn        | 0.475 | 0.373 | 0.202 | 0.490 | 0.326 | 0.133 | 0.495 | 0.330 | 0.134 |
| time=2    | 0.937 | 0.630 | 0.137 | 0.985 | 0.534 | 0.065 | 0.983 | 0.540 | 0.069 |
| time=3    | 1.203 | 0.601 | 0.046 | 1.189 | 0.498 | 0.017 | 1.208 | 0.507 | 0.017 |