

AN ABSTRACT OF THE DISSERTATION OF

Lin Lu for the degree of Doctor of Philosophy in Statistics presented on November 15, 2007.

Title: Unconditional Estimating Equation Approaches for Missing Data.

Abstract approved:

Annie Qu

David S. Birkes

Missing data can lead to biased and inefficient estimation if the missing mechanism is not taken into account in the analysis. In this dissertation we propose two estimators that, under fairly general conditions, are asymptotically unbiased. The first proposed estimator assume the data are missing at random (MAR) and does not require a model for the missing mechanism. The second estimator allows the missingness to be nonignorable and requires a model for the mechanism. Both proposed approaches utilize generalized estimating equations (GEE) based on unconditional models.

One main advantage of the proposed approaches is that they do not require full specification of the likelihood. They only need the first few moments of the response variables and covariates. Another advantage is that they can easily handle arbitrary missing patterns. Using simulation, we investigate the efficiency of the proposed approaches relative to the weighted GEE (WEE) and multiple imputation (MI) estimators. The proposed estimators are as efficient as WEE and MI estimators when the latter two approaches use the correct model to obtain weights or impute missing values.

©Copyright by Lin Lu

November 15, 2007

All Rights Reserved

Unconditional Estimating Equation Approaches
for Missing Data

by
Lin Lu

A DISSERTATION
submitted to
Oregon State University

in partial fulfillment of
the requirements for the
degree of

Doctor of Philosophy

Presented November 15, 2007
Commencement June 2008

Doctor of Philosophy dissertation of Lin Lu presented on November 15, 2007.

APPROVED:

Co-Major Professor, representing Statistics

Co-Major Professor, representing Statistics

Chair of the Department of Statistics

Dean of the Graduate School

I understand that my dissertation will become part of the permanent collection of Oregon State University libraries. My signature below authorizes release of my dissertation to any reader upon request.

Lin Lu, Author

ACKNOWLEDGEMENTS

This dissertation would not have been possible without the help of many people. I would like to thank all the faculty, staff, and students of the Statistics department for their assistance with all facets of this dissertation.

I would like to thank Dr. Annie Qu and Dr. David Birkes, my advisers, for extending their complete encouragement during the preparation of this dissertation. I wish to express my sincere appreciation for their helps. I especially appreciate the helps from Dr. Birkes, without his constant guidance, this dissertation would not have been finished. I greatly thank Dr. Qu for her help and suggestions. It is my honor and a distinct pleasure to work with them.

Special thanks are due to Dr. Dan Schafer and Dr. Bob Smythe, for their insightful suggestions, time spent in reading this dissertation and serving on my committee. I greatly appreciate Dr. Lee for in reviewing this dissertation and serving as my graduate representative.

CONTRIBUTION OF AUTHORS

Dr. Annie Qu proposed the original questions that motivated this research. She also contributed to the development of the methodologies, and edited and reviewed the manuscripts as co-author.

Dr. David Birkes contributed to the development of the methodologies, and edited and reviewed the manuscripts as co-author.

TABLE OF CONTENTS

	<u>Page</u>
1 Unconditional Estimating Equation Approaches for Missing Data	1
1.1 Background	1
1.2 Missingness Mechanisms	1
1.3 Models for Longitudinal Data	3
1.3.1 Generalized Estimating Equations	3
1.3.2 Linear Mixed Model	4
1.3.3 Transition Model	5
1.4 Analysis of Missing Data in Longitudinal Study	6
1.4.1 Maximum Likelihood Approach	6
1.4.2 Weighted Generalized Estimating Equations	7
1.4.3 Imputation	9
1.5 Organization of the Dissertation	10
2 Unconditional Estimating Equation Approach for Ignorable Missingness	11
2.1 Introduction	11
2.2 Models and Notation	13
2.2.1 Generalized estimating equations	13
2.2.2 Estimating Equation Approaches for Missing Data	14
2.3 Unconditional Estimating Equation Approach	15
2.3.1 Example	18
2.4 Application	20
2.4.1 Contracepting Women Data	20
2.4.2 Arbitrary Missing Pattern Data	22
2.4.3 Missing Covariate	23
2.5 Simulations	24
2.5.1 Continuous Response	24

TABLE OF CONTENTS (Continued)

	<u>Page</u>
2.5.2 Binary Response	25
2.5.3 Simulation Based on Contracepting Women Data	26
2.5.4 Simulation Based on Monotone Missing Covariate Example	27
2.5.5 Simulation Based on Arbitrary Missing Pattern	27
2.6 Discussion and Conclusion	28
Reference	30
3 Unconditional Estimating Equation Approach for Nonignorable Missingness . . .	35
3.1 Introduction	36
3.2 Background and Available Approaches	38
3.3 Proposed Method	39
3.3.1 Examples	43
3.4 Application	46
3.4.1 TLI Data	46
3.4.2 Contracepting Women Data	48
3.4.3 Missing Covariate	50
3.5 Simulations	52
3.5.1 Simulation Based on TLI Data	52
3.5.2 Simulation Based on Women Data	53
3.5.3 Simulation Based on Missing covariates Data	54
3.6 Discussion and conclusion	55
Reference	56
4 General Conclusion	62
Bibliography	64
Appendix	67

LIST OF TABLES

<u>Table</u>	<u>Page</u>
1 Estimates and Standard Errors for Contracepting Women Data	32
2 Estimates and Standard Errors for Arbitrary Missing Pattern Data	32
3 Estimates and standard errors for missing covariate	32
4 Simulation results for continuous data based on 250 sample and 1000 simulations	33
5 Simulation results for binary data based on 250 sample size and 1000 simulations	33
6 Estimates and standard error for the simulation based on contracepting women data with sample size 250	33
7 Biases and standard errors for the simulation based on missing covariates with sample size 250	33
8 Simulation results for continuous data based on missing covariates data with sample size of 200 and simulation size of 1000	34
9 Estimates for TLI	59
10 Estimates for Women data	59
11 Estimates for missing covariate	60
12 Biases, square root of mean squared errors (RMSE) and standard errors for simulation based on TLI with sample size 82	60
13 Biases and standard errors for simulation based on TLI with sample size 200 . .	60
14 Biases and standard errors for the simulation based on women data with sample size 250	61
15 Estimates and standard errors for 500 simulation based on missing covariates data with sample size of 200	61
16 TLI data	77

Unconditional Estimating Equation Approaches for Missing Data

1.1 Background

Missing data occurs frequently in longitudinal studies. It is well known that missing data can lead serious bias and inefficient estimation if the missingness is not taken into account in the analysis (Afifi, and Elashoff, 1966, Rubin, 1976). Therefore, it is desirable to develop statistical method for handling missing data problem.

Little and Rubin (2002) summarize statistical methods in four categories: complete-cases analyses (Nie et al.,1975); inversed weighting strategies (Horvitz and Thompson, 1952, Rosenbaum and Rubin, 1983, 1985, Little, 1986, Robins et al., 1994, Rotnitzky et al., 1998, Scharfstein, 1999); imputation methods (Rubin, 1977, Paik, 1997, Xie,1997); and the likelihood-based approaches (Ibrahim et al., 1990, 1996, 1999, 2003, Wang, 2002).

Most existing methods are based on likelihood approaches (Ibrahim et al., 1990, 1996, 1999, Lin, 2002, Wu, 2004, Roy, 2005). These approaches require full specification of a joint distribution of response variables and covariates. In practice, the full likelihood is often unknown, and if the specification of likelihood is not valid, the likelihood-based approaches could still cause biased and inefficient estimators. Weighted estimating equation and imputation methods were developed to handle missing values. These approaches do not require full likelihood, but they still rely on modeling assumptions for missingness mechanism or requires imputing missing values under missing at random.

1.2 Missingness Mechanisms

Missing data can occur in response variables, covariates, or both. There are several missing patterns. For example, data could have univariate missing pattern, with one variable having missing values; or it could have a monotonic missing pattern, such as drop-out; or

it could have arbitrary intermittent missing.

Little and Rubin (1987) distinguished the missing patterns and missingness mechanisms. The missing patterns produce missingness indicator matrix based on observed and missing values, while the missingness mechanisms consider how the missingness is related to observed and missing values, either missing or observed. The missingness mechanisms are important because both estimation and inference for missing data depend on the missingness mechanisms. Rubin (1976) classified missingness mechanisms under three categories: missing completely at random (MCAR), missing at random (MAR), and informative missing (IM). MCAR and MAR are referred to as ignorable missing and IM is referred to as nonignorable missing. The main difference between ignorable and non-ignorable missingness is whether the missingness depends on missing values or not. Little and Rubin (1987, 2002) define missingness mechanism based on likelihood function as follows.

Let $\mathbf{Y} = (y_{ij})$ be the complete data, θ be the unknown parameters, and \mathbf{R} be the missingness indicator matrix, where each element r_{ij} of \mathbf{R} is either 1 if y_{ij} is observed or 0 if y_{ij} is missing. If the missingness is MCAR, where the missingness does not depend on the data Y , we have the conditional distribution of \mathbf{R} given \mathbf{Y} and θ , say $f(R|Y, \theta)$, is given as follows

$$f(R|Y, \theta) = f(R|\theta), \text{ for all } Y, \theta.$$

If the missingness is MAR, where the missingness depends on observed data, we have the conditional distribution of R given Y and θ as

$$f(R|Y, \theta) = f(R|Y_{obs}, \theta), \text{ for all } Y_{mis}, \theta.$$

If the missingness is IM, where the missingness depends on both observed and unobserved data, the conditional distribution of R given Y and θ can be written as

$$f(R|Y, \theta) = f(R|Y, \theta), \text{ for all } Y, \theta.$$

1.3 Models for Longitudinal Data

Longitudinal studies are interested in obtaining treatment effect over time. Subjects are measured over time. In general, estimators obtained by repeated measure design are more efficient than those from cross-sectional designs. However, it is more challenge to analyze longitudinal data than cross-sectional data since the observations within the same subject are no longer independent any more.

Major approaches for longitudinal analysis include: marginal regression models (Liang and Zeger, 1986), and mixed-effect regression models (Laird and Gibbons, 1982). Diggle et al (1994) also provides a detailed review of marginal models, random effect models, and transition models. We provide a brief introduction for these approaches.

1.3.1 Generalized Estimating Equations

Marginal models, which investigate population average have the same interpretation as cross-sectional studies. The marginal models model correlation among the repeated observations for a subject directly through correlation matrix of the error term.

One of the most popular techniques used to handle marginal models is the generalized estimating equations (GEE) approach, proposed by Liang and Zeger (1986). The GEE approach is used to model the marginal expectation of responses as a function of a set of covariates. Let $X_i = (x_{i1}, \dots, x_{iT})$ be a $T \times P$ covariates matrix, where $x_{it} = (x_{it1}, \dots, x_{itP})$, y_{it} be the response variable, $Y_i = (y_{i1}, \dots, y_{iT})$ be a $1 \times T$ vector of responses, and $\mu_{it} = E(y_{it})$, $i = 1, \dots, N$ and $t = 1, \dots, T$. Assume the marginal regression model is given as

$$g(\mu_{it}) = x_{it}\beta, \quad (1.1)$$

where β is the $P \times 1$ regression parameters of interest and $g(\cdot)$ is a link function. Assuming the $T \times T$ covariance matrix for Y_i is

$$V(\alpha) = \phi A_i^{\frac{1}{2}} R(\alpha) A_i^{\frac{1}{2}},$$

where A is a diagonal matrix of variance functions, $R(\alpha)$ is the working correlation matrix of Y , α is the correlation parameter, and ϕ is a dispersion parameter. The GEE estimators

for regression parameters are the solutions of

$$\sum_{i=1}^N D_i V(\alpha)^{-1} (Y_i - \mu_i) = 0,$$

where $D_i = \frac{\partial \mu_i}{\partial \beta}$.

The GEE estimators are consistent if the mean model in (1.1) is correctly specified. Moreover, the GEE estimators are still consistent if the correlation structure is misspecified. Liang and Zeger (1986) referred $V(\alpha)$ as a working covariance matrix because it does not need to be correctly specified. However, a correctly specified working correlation matrix can improve the efficiency of estimators.

An empirical sandwich estimator can be used to estimate the variance of estimators. Another available estimator is the model-based estimator, which is only consistent if both the mean model and the covariance structure are correctly specified. Generally, the empirical variance estimator is preferred when the number of clusters is large. When the number of clusters is small, the model based variance estimator may have better properties (Prentice 1988).

1.3.2 Linear Mixed Model

The linear regression model has the following form

$$E(y_i) = x_i \beta,$$

where the regression coefficients β are fixed for all subjects. Mixed-effect models are appropriate if the subject specific effect is of interest.

Mixed effect model is able to take heteroscedastic variation into account and estimates group trends over time. In addition, it provides information of how different subjects change across time. Compared with marginal models, mixed effects models introduce model correlations through random effects in mean structure.

Laird and Ware (1982) and Verbeke et al (2000) provide more applications of linear mixed models. Here we provide a brief introduction. The generalized linear mixed model

has the following form

$$g(\mu_i) = x_i\beta + z_ib_i,$$

where $\mu_i = E(y_i)$; $g(\cdot)$ is a link function; β is the fixed-effect coefficient; z_i is the $r \times 1$ vector of random effect variables; and b_i is the random-effect coefficients for group i , and assume to have $N_q(0, \psi)$ distribution. The subject varies through the random effect.

The estimators of parameters in the linear mixed model can be solved by the maximum likelihood method. In longitudinal studies, it is often assumed that the $y_{it}|b_i$ in $y_i = (t_{i1}, \dots, y_{iT})$ are independent. Let $f_\theta(y_{it}|\mathbf{x}_{it}, b_i)$ be the conditional density of response variable. The likelihood of a generalized linear mixed model (GLMM) can be expressed as

$$\int \prod_j^{n_i} f_\theta(y_{it}|b_i, \mathbf{x}_i) f_D(b_i) db_i.$$

However, the integration does not have explicit form in general so that the numerical approximation for the likelihood is needed.

1.3.3 Transition Model

Diggle et al (2002) introduced the transition models. The transition model is an extension of the generalized linear model, which is constructed by applying a generalized linear model for the marginal mean and specifying conditional dependence of current outcomes based on past outcomes. For repeated measurements in transition model, past outcome and covariates are modeled as covariates for current outcome. Transition models sometimes are also called Markov models. That is, the conditional mean of y_{it} for subject i at time t depends on covariates and prior responses. The simplest transition model is that the responses are conditionally independent from each other given the previous time. For example, suppose the transition model is a linear function of previous outcomes

$$E(y_{it}|y_{i1}, \dots, y_{i(t-1)}, x_i) = \sum_{r=1}^{t-1} \alpha_r(y_{ir}).$$

For a first-order Markov model, we have

$$f(y_{it}|y_{i1}, \dots, y_{i(t-1)}; \alpha) = f(y_{it}|y_{i(t-1)}),$$

The likelihood contribution from subject i is given by

$$f(y_{i1}, y_{i2}, \dots, y_{iT_i}; \alpha) = f(y_{i1}; \alpha) \prod_{t=2}^{T_i} f(y_{it} | y_{it-1}).$$

In general, the estimators of parameters in transition models can be obtained by the maximum likelihood methods.

Diggle et al (2000) mention that the transition models are intractable in general if many nuisance parameters are involved and need to be estimated. Therefore, it often requires more assumptions in order to specify the entire likelihood.

1.4 Analysis of Missing Data in Longitudinal Study

When missingness is not MCAR, the estimators using complete-case only are biased and not efficient. Therefore, it is important to incorporate missing data information in the data analysis. We provide several existing methods to handle missing data.

1.4.1 Maximum Likelihood Approach

Likelihood based method is the most fundamental approach. Suppose that \mathbf{Y} is the response variable with density function $f(\mathbf{Y}|\theta)$, where θ is the parameters. The inference based on maximum likelihood is a model-based inference about the parameters θ .

Let \mathbf{Y} be observed data and \mathbf{Z} be missing data, \mathbf{Y} and \mathbf{Z} form the complete data. Let f_i be the density function of the i^{th} subject of complete data with parameters θ . The likelihood of the complete data is given by $L(\theta) = \prod_{i=1}^N f_i(\theta)$. By Bayes' Rule and the law of total probability, the probability of missing data given the observed data can be expressed as:

$$p(\mathbf{Z}|\mathbf{Y}, \theta) = \frac{\mathbf{p}(\mathbf{Y}, \mathbf{Z}|\theta)}{\mathbf{p}(\mathbf{Y}|\theta)} = \frac{\mathbf{p}(\mathbf{Y}|\mathbf{Z}, \theta)\mathbf{p}(\mathbf{Z}|\theta)}{\int \mathbf{p}(\mathbf{Y}|\mathbf{Z}, \theta)\mathbf{p}(\mathbf{Z}|\theta)d\mathbf{Z}}$$

This requires the density function of $p(\mathbf{Y}|\mathbf{Z}, \theta)$ and $p(\mathbf{Z}|\theta)$.

The major approach for obtaining maximum likelihood estimator where there is missing data is by applying the expectation and maximization (EM) algorithm. Hartley (1958) proposes the earliest version of EM; Dempster, Laird, and Rubin (1977) formalizes EM

and provides a proof of convergence. EM is particularly useful if the completed likelihood function is easy to formulate. EM is an iterative optimization method to estimate some unknown parameters based on given data and with some hidden variables or missing values, which requires to be integrated out. An EM algorithm updates θ in two steps. The first step is expectation. The expectation of $Q(\theta)$ is given by

$$Q(\theta) = E_{\mathbf{Z}}(\log L(\mathbf{Y}, \mathbf{Z}|\theta)|\mathbf{Y}) = \int \log L(\mathbf{Y}, \mathbf{Z}|\theta) \mathbf{p}(\mathbf{Y}, \mathbf{Z}|\theta) d\mathbf{Z}.$$

Once the parameters, θ , of the Q are known in the first step, Q is fully determined and can be maximized in the second step of an EM algorithm. The second step is maximization, that is,

$$\theta_{n+1} = \max_{\theta} Q(\theta).$$

In other words, θ_{n+1} is the value that maximizes the expectation of the complete data log-likelihood with the observed variables given the previous parameter value.

The EM algorithm is able to handle missing values under MAR and informative missing. Once the joint probability distribution of data and missingness mechanism is obtained, the conditional likelihood given missing data can be determined and the missing values can be integrated out. The estimations of parameters are given by maximizing the conditional log likelihood. However, EM algorithms depend on distributional assumptions so that the estimators could still be biased and inefficient when the distribution assumptions are misspecified. EM could also be extremely slow computationally if the integration of missing information involves high dimension.

1.4.2 Weighted Generalized Estimating Equations

The original idea of weighted approach for handling missing data is from weighting strategies for finite population surveys (Horvitz and Thompson, 1952) to obtain unbiased estimator. The weighted method assigns weight to each observed-case to adjust for sampling bias. Robins et al. extend this idea and propose the weighted estimating equations approach (1994, 1995). The weighted estimating equations approach weights the observed subjects with the inverse probability of being observed. This approach is valid when the

missingness is MAR and given the model for estimating the probability for missingness mechanism is correctly specified. The consistent estimators of β can be obtained by solving

$$\sum_{i=1}^K D_i' V_i^{-1} W_i (Y_i - \mu_i) = 0, \quad (1.2)$$

where $D_i = \frac{\partial \mu_i}{\partial \beta}$ and $V_i = A_i^{\frac{1}{2}} R A_i^{\frac{1}{2}}$ is a $T \times T$ working covariate matrix for Y_i and R is a $T \times T$ working correlation matrix, which are assumed known. The choice of working correlation affects the efficiency of estimator. The missingness is taken into account through specification of a $T \times T$ diagonal weighting matrix of W_i , $W_i = \text{diag}\{R_{i1}w_{i1}, \dots, R_{iT}w_{iT}\}$, and $R_{it} = 1$ if the i^{th} subject is observed at time t , and 0 otherwise. That is, W_i provides weight for the observed visits and 0 for the unobserved visits. The weight w_{it} is the inverse of the probability where the i^{th} subject is observed at the t^{th} visit, w_{it} is often unknown and needs to be estimated.

It requires modeling the missing process in order to obtain the weights w_{it} . We denote $\lambda_{it} = P(R_{it} = 1 | R_{i(t-1)} = 1, X_i, Y_i, \alpha)$ as the probability of a response being observed at time t for the i^{th} subject given the subject is observed at the time $t - 1$. If the missingness is assumed to be MAR, we have

$$\lambda_{it} = P(R_{it} = 1 | R_{i(t-1)} = 1, X_i, Y_{i1}, \dots, Y_{i(t-1)}, \alpha),$$

where the missingness mechanism only depends on observed data and may be specified up to a $q \times 1$ vector of unknown parameters α . Here λ_{it} can be modeled as a logistic regression model of Z_{it} , a vector of predictor, which may include missingness indicator variables, covariates and previous responses.

$$\text{logit } \lambda_{it}(\alpha) = Z_{it}\alpha.$$

Therefore the weight w_{it} , the inverse of the unconditional probability of being observed at time t , can be calculated as,

$$\hat{w}_{it} = \frac{1}{\hat{\lambda}_{i1} \times \dots \times \hat{\lambda}_{it}}, \quad i = 2, \dots, T$$

and $\hat{w}_{i1} = 1$. The weighted GEE estimator by solving equation (1.2) is consistent and the asymptotic variance of β is

$$\left(\sum_{i=1}^K D_i' V_i^{-1} W_i D_i \right)^{-1} \sum_{i=1}^K E_i E_i' \left(\sum_{i=1}^K D_i' V_i^{-1} W_i D_i \right)^{-1}, \quad (1.3)$$

where $E_i = U_i - (\sum_{i=1}^K U_i S_i') (\sum_{i=1}^K S_i S_i') S_i$, $U_i = D_i' V_i^{-1} W_i (Y_i - \pi_i)$, and $S_i = \sum_t R_{it-1} Z_{it} (R_{it} - \lambda_{it})$ is the score component for the i^{th} subject from the missingness mechanism model (Robins et al., 1995). The use of $\sum_{i=1}^K E_i E_i'$ instead of $\sum_{i=1}^K U_i U_i'$ adjusts for the variation of α estimation.

1.4.3 Imputation

The imputation approach is first proposed by Rubin (1976) to handle missing values. There are two types of imputation. One is single imputation and another is multiple imputation. Single imputation computes a value for each missing value. There could be the mean estimation, or simply using available data in the same strata, or predicting missing values from a regression model. In general, single imputation does not reflect the uncertainty about the predictions of the missing values; the standard error based on single imputation estimators can be underestimated.

Instead of single imputation, a multiple imputation proposed by Rubin (1978) replaces missing value with a set of imputing values, which also represent the uncertainty about the true value. Paik (1997) and Schafer (1997) provide through overview of the method. The multiple imputed values along with observed data are analyzed as if they are from complete data. The multiple imputation method requires combining results from each imputation.

The way of combining results from multiple imputation is given as follows. Let m be the number of multiple imputations and $\hat{\psi}_i$ be the estimator, and \hat{V}_i be the variance of the estimator from the i^{th} imputation. The combined estimator from the multiple imputations is the average of the estimator from each imputation,

$$\bar{\psi} = \frac{1}{m} \sum_{i=1}^m \hat{\psi}_i.$$

The variance of $\bar{\psi}$ is a combination of within imputation variance and between imputation variance:

$$Var(\bar{\psi}) = \frac{1}{m} \sum_{i=1}^m \hat{V}_i + \frac{m+1}{m} \frac{1}{m} \sum_{i=1}^m (\hat{\psi}_i - \bar{\psi})^2.$$

1.5 Organization of the Dissertation

Chapter 2 introduces an approach based on unconditional generalized estimating equations, which can handle missing values when missingness is ignorable. Contrast to the weighted GEE or imputation methods, when the missingness mechanism is MAR, unconditional generalized estimating equations approach does not require modeling the missing indicator to obtain weights or to impute the missing values. It uses available observed data to estimate the parameters of interest. The main advantage of the proposed approach is that it does not require full specification of likelihood or model assumptions of the missingness mechanism, it only requires the first few moments of the response variables and covariates.

Chapter 3 proposes a different approach based on unconditional estimating equations, which can handle nonignorable missing data. This approach requires modeling assumption for the missing mechanism but does not require fully specification of the joint distribution of response variables, covariates, and missingness mechanism. It also has computational advantages compare to the maximum likelihood approach and multiple imputation approach.

Unconditional Estimating Equation Approach for Ignorable Missingness

Lin Lu, Annie Qu, David Birkes

Abstract

We present an unconditional estimating equation approach to handle missing data that is missing at random (MAR). When generalized estimating equations (GEEs) are used in a regression analysis, one usually regards the covariates as fixed, but in our approach we regard them as random. One advantage of our approach is that it does not require modeling the missingness mechanism. In contrast to the weighted generalized estimating equation (WEE) method, which must model the missingness mechanism in order to obtain weights. Also, it does not require working with the conditional distributions that the multiple imputation method uses to impute missing values. In addition, it does not require fully specifying likelihood function, but only requires the first few moments of response variables and covariates. We use simulation to investigate the finite-sample efficiency of the proposed approach relative to the weighted GEE and multiple imputation approaches. Data examples are also provided for illustration.

2.1 Introduction

Missing data could occur in various studies such as design based study or observational study. It could also occur on response variables, or covariates, or both. Missing data especially occurs often in longitudinal studies where subjects measured over time may dropout early or have intermittent missing observations. If the missingness mechanism is not taken into account in the analysis, missing data can cause serious bias and lead

to inefficient estimation (Afifi, and Elashoff, 1966; Rubin, 1976). Therefore developing efficient method to handle the missing data has become increasingly important.

Rubin (1976) classified missingness mechanisms into three categories: missing completely at random (MCAR), where missingness does not depend on observed data; missing at random (MAR), where missingness only depends on the observed data; and informative missing (IM), where missingness could also depend on unobserved data, and the missing information can not be recovered by known information.

The major approaches to handle missing data can be summarized as follows (Little and Rubin, 2002): complete-cases analysis (Nie et al., 1975); inverse weighting strategies (Horvitz and Thompson, 1952, Rosenbaum and Rubin, 1983, 1985, Little, 1986, Robins et al, 1994, 1995, Rotnitzky et al., 1998, Scharfstein, 1999); imputation methods (Rubin, 1977, Paik, 1997, Xie and Paik, 1997); and the likelihood-based approaches (Ibrahim et al, 1990, 2005). The likelihood approach is consistent if the missingness is MCAR or MAR (Dempster, Laird, and Rubin, 1977, Laird, 1988).

However, the likelihood based approaches require full specification of a joint distribution of response variables and covariates. This assumption might be very restrictive in practice. In addition, if this assumption is violated, the likelihood-based approaches could still lead to biased and inefficient estimation. In contrast, imputation and WEE do not require full likelihood functions, however, these approaches still rely on strong modeling assumptions for modeling missingness mechanism or imputing missing values.

In this paper, we propose a semiparametric approach to handle missing data with missing at random. This new approach uses unconditional modeling for estimating equations under MAR. The proposed approach does not require modeling the missing mechanism to estimate weights as in WEE. It also does not require modeling the relation between missing values and observed values as in imputation approach. In addition, the proposed approach does not require fully specification of the likelihood function, but only needs the first few moments of response variables and covariates.

Our contribution is to construct estimating equation based on unconditional expectation of all possible variables and solve these estimating equation simultaneously. This

approach is able to handle both continuous and discrete response variables. It is also able to handle missing response or missing covariates. In addition, it can incorporate correlation information for clustered data. We show the proposed estimator is a consistent and asymptotically normal.

This paper is organized as follows. We provide model framework and notations in Section 2.2. Then we introduce the unconditional estimating equation method and the asymptotic properties in Section 2.3. The applications of the proposed method and simulation studies are given in Sections 2.4 and 2.5. The discussion and conclusion are given in Section 2.6.

2.2 Models and Notation

We first provide generalized linear models (GLM) and generalized estimating equations (GEE) briefly. The generalized linear model was introduced by McCullagh and Nelder (1983) where the mean of response and covariates have following relation:

$$g(\mu_i) = x_i' \beta.$$

2.2.1 Generalized estimating equations

The generalized estimating equations (Liang and Zeger, 1986) is a widely used marginal approach when the full likelihood function is difficult to specify since it only requires the first two moments of the distribution. Marginal models are appropriate when the inference to population average is of interest. GEE can handle multivariate response variables. In addition, the GEE estimator is consistent when the working correlation is misspecified.

For the longitudinal data setting, suppose that N independent subjects are measured from time 1 to T . For each independent subject i ($i = 1, \dots, N$), y_{it} is a response variable and $\mathbf{X}_i = (\mathbf{x}_{i1}, \dots, \mathbf{x}_{iT})'$ is a $T \times P$ covariates matrix. Let $\mu_{it} = E(y_{it} | \mathbf{X}_i, \beta)$ be the mean of the response y_{it} . The marginal regression model is given as

$$g(\mu_{it}) = \mathbf{x}_{it}' \beta, \tag{2.1}$$

where β is the $P \times 1$ regression parameters of interest and $g(\cdot)$ is a link function. Denote $\mathbf{y}_i = (y_{i1}, \dots, y_{iT})$ as a T dimensional response vector. The covariance of \mathbf{y}_i ($T \times T$ matrix) has the following form:

$$V(\alpha) = \phi_y \mathbf{A}_i^{\frac{1}{2}} \mathbf{R}(\alpha) \mathbf{A}_i^{\frac{1}{2}},$$

where \mathbf{A} is a diagonal marginal variance matrix, $\mathbf{R}(\alpha)$ is the working correlation matrix of \mathbf{Y} , α is the correlation parameters, and ϕ_y is a dispersion parameter. The GEE estimators for the regression parameters are the solutions of

$$\sum_{i=1}^N \mathbf{D}_i' \mathbf{V}_i(\mu_i, \alpha)^{-1} (\mathbf{Y}_i - \mu_i) = 0, \quad (2.2)$$

where $\mathbf{D}_i = \frac{\partial \mu_i}{\partial \beta'}$, and α can be estimated by the method of moments using Pearson residuals. If the missingness is MCAR, the GEE estimator is consistent even when the working correlation is misspecified.

2.2.2 Estimating Equation Approaches for Missing Data

The weighted generalized estimating equations (Robins et al., 1995) and multiple imputation GEE method (Paik, 1997) are effective to handle MAR for longitudinal data. The WEE provides consistent estimators even if the working correlation model is misspecified. However, it requires the model for predicting the probability of missingness to be correctly specified. The WGEE estimator solves

$$\sum_{i=1}^K \mathbf{D}_i' \mathbf{V}_i^{-1} \mathbf{W}_i \{\mathbf{Y}_i - \mu_i(\mathbf{X}_i, \beta)\} = 0, \quad (2.3)$$

where \mathbf{W}_i is a diagonal weighting matrix with the t^{th} diagonal components representing the inverse of the probability being observed at the t^{th} visit for the i^{th} subject. Clearly, any observations with low probability of being observed will have more weights on the estimating equation in (2.3).

The multiple imputation GEE approach is based on filling the missing data with imputed values, and the number of multiple imputation typically ranges between 3 and 10 (Rubin, 1987, p.114). The parameters are estimated by solving the GEE in (2.2)

and replacing the missing response by the imputed response. Rubin (1987, ch4) and Schafer (1997) provide the valid inference from multiple imputation. Specifically, the final estimator and inference are obtained by combining results for multiple imputation. See Section 2.4.1 for more details.

2.3 Unconditional Estimating Equation Approach

Let the response variable $\mathbf{y}_i = (y_{i1}, \dots, y_{iT})'$ be a T -dimensional vector with variance-covariance matrix Σ . We consider two types of covariates in the model. One is time-varying covariate, $\mathbf{X}^*_i = (\mathbf{x}_{i1}^*, \dots, \mathbf{x}_{iT}^*)'$, which is a $T \times P$ matrix. Another is time-invariant covariate, $\mathbf{x}_i^{**} = (\mathbf{x}_{i1}^{**}, \dots, \mathbf{x}_{iQ}^{**})'$, which is a $Q \times 1$ vector. We denote

$$\mathbf{x}_i = \begin{pmatrix} \mathbf{x}_{i1}^* \\ \vdots \\ \mathbf{x}_{iT}^* \\ \mathbf{x}_i^{**} \end{pmatrix},$$

which is a $(T \times P + Q) \times 1$ covariate vector. Missing values could occur both in \mathbf{y}_i and \mathbf{x}_i .

Covariates \mathbf{x}_i are treated as random variables and the mean and variance are given as follows

$$E(\mathbf{x}_i) = \mu_{\mathbf{x}}, \quad \text{Var}(\mathbf{x}_i) = \Sigma_x.$$

Let $\mathbf{u}_i = (\mathbf{y}_i', \mathbf{x}_i')$, here u_i 's are independent identical distributed, and have the first two moments as follows

$$E(\mathbf{u}_i) = \mu_u(\psi), \quad \text{Var}(\mathbf{u}_i) = \mathbf{V}(\mu, \alpha). \quad (2.4)$$

We denote $\psi = (\beta, \phi, \gamma)$, where β is a vector of regression parameters in (2.1), ϕ is a set of joint parameters of \mathbf{y}_i and \mathbf{x}_i , and γ is a vector of parameters for \mathbf{x}_i . Here μ and \mathbf{V} are known functions of ψ . We assume \mathbf{V} is known. If it is unknown, we can estimate it empirically.

The main idea of the unconditional estimating equation is to construct estimating equations simultaneously associated with \mathbf{y}_i and covariates \mathbf{x}_i . The quasiliikelihood equation with respect to ψ is

$$\sum_{i=1}^N \mathbf{D}' \mathbf{V}^{-1} (\mathbf{u}_i - \mu_{\mathbf{u}}) = 0, \quad (2.5)$$

where $\mathbf{D} = \partial \mu_{\mathbf{u}} / \partial \psi'$. Since $\mu_{\mathbf{u}}$ does not depend on a specific subject index i , \mathbf{D} also does not depend on index i . So equation (2.5) can be written as

$$\mathbf{D}' \mathbf{V}^{-1} (\bar{\mathbf{u}} - \mu_{\mathbf{u}}) = 0, \quad (2.6)$$

where $\bar{\mathbf{u}} = \frac{\sum_{i=1}^N \mathbf{u}_i}{N}$ is the mean of \mathbf{u}_i .

We obtain $\hat{\psi}$ by solving estimating equations (2.6). For complete data if we take expectation of equation (2.6),

$$\mathbf{E}(\mathbf{D}' \mathbf{V}^{-1} (\bar{\mathbf{u}} - \mu_{\mathbf{u}})) = \mathbf{D}' \mathbf{V}^{-1} (\mathbf{E}(\bar{\mathbf{u}}) - \mu_{\mathbf{u}}) = 0, \quad (2.7)$$

Therefore, the estimating equation (2.6) is an unbiased estimating equation. We have the following theorem.

Theorem 2.1. *The estimator $\hat{\psi}$ is \sqrt{N} -consistent, and $N^{1/2}(\hat{\psi} - \psi)$ is asymptotically multivariate normal with mean zero and the asymptotic covariance matrix as*

$$(D' V^{-1} D)^{-1}.$$

The matrix $(D' V^{-1} D)^{-1}$ is $k \times k$ where k is the number of parameters that need to be estimated. For this matrix to be invertible, we need the rank of D to be at least k , and in particular, we need the number of rows of D to be at least k . The number of rows of D is equal to the number of variables in \mathbf{u}_i . Above we defined $\mathbf{u}_i = (\mathbf{y}'_i, \mathbf{x}'_i)'$, but if the number of single variables (response variables \mathbf{y} and covariates \mathbf{x}) are not enough, we can include some products of \mathbf{y} and \mathbf{x} .

Suppose some values of response variables and/or some values of covariates are missing and assume that the missingness is MAR. Let R_i be a $N \times G$ missingness indicator matrix where $G = T + T \times P + Q$ (there are T response variables, $T \times P$ time varying covariates, and Q time invariant covariates), where $R_{ig} = 1$ if u_{ig} is observed and $R_{ig} = 0$ if u_{ig} is missing.

Subjects can be grouped by the pattern of R . We denote r as different missing patterns, where $r = 1, \dots, M$. For example, if $G = 3$, there could be $M = 2^3 - 1$ possible different missing patterns. Let U_r be the observed values under missing pattern r and $g(\mathbf{U}, \psi)$ be an estimating function for the hypothetical complete data such that $E(g(\mathbf{U}, \psi)) = 0$. We have following Lemmas

Lemma 2.1. *Under the MAR assumption,*

$$E_\psi(g(\mathbf{U}, \psi)|R = r, U_r = u) = E_\psi(g(\mathbf{U}, \psi)|U_r = u).$$

The conditional expectation $E_\psi(g(\mathbf{U}, \psi)|R = r, U_r = u)$ has information equivalent to $E_\psi(g(\mathbf{U}, \psi)|U_r = u)$, which does not depend on missingness mechanism.

Lemma 2.2. *Under MAR assumption, we have*

$$E\left(\sum_r E_\psi(g(\mathbf{U}, \psi)|R = r, U_r = u)\right) = 0.$$

For each missing pattern r , we have

$$\sum_r \mu'_r V_r^{-1} (U_r - \mu_r) = \sum_{i=1}^N D'_i V_i^{-1} E(u_i - \mu_i | U_i^o) = \sum_{i=1}^N (D'_i)^o V_i^{-1} (u_i^o - \mu_i^o). \quad (2.8)$$

This approach assume that the missing values can be replaced by conditional expectation of $E(U_i^m | U_i^o)$. Therefore, it uses all available information and also provides consistent estimators when missingness is MAR.

For a specified subject i , if it belongs to the missing pattern r , the corresponding quasi-score is

$$D'_{i_r} V_{i_r}^{-1} (U_{i_r} - \mu_{i_r}),$$

where μ_{i_r} consists of the expectation of the response variables and the observed covariates, and V_{i_r} is the corresponding variance-covariance matrix where $R_{i_r} = 1$.

The quasi-score containing subjects with the same missing pattern r is

$$D'_r V_r^{-1} (U_r - \mu_r) = \sum_{i_r} D'_{i_r} V_{i_r}^{-1} (U_{i_r} - \mu_{i_r}).$$

Therefore the quasi-score for all missing patterns has the following form

$$\sum_r \sum_{i_r} D_r' V_r^{-1} (U_r - \mu_r) = 0. \quad (2.9)$$

The parameter ψ is estimated by solving the equation (2.9). Let $p_r = \lim_{N \rightarrow \infty} \frac{n_r}{N}$, where n_r is the number of the subjects for missing pattern r .

Theorem 2.2. *The estimator of ψ by solving the equation (2.9) is an consistent estimator of ψ under MAR, and $N^{1/2}(\hat{\psi} - \psi)$ is asymptotically multivariate normal with mean zero and the covariance matrix as $(\sum_r p_r D_r' V_r^{-1} D_r')^{-1}$.*

2.3.1 Example

In this section, we provide an example to illustrate the proposed method. Although this example only shows how to handle continuous responses and missing one covariate, our approach is not limited and can be applied to discrete responses or missing response.

Suppose that each subject is measured at time $t = 1, 2$. The response variables are y_{i1} , y_{i2} , and the covariate x_i are continuous. We assume x_i has mean μ_x , and considered the following marginal regression model

$$y_{it} = \beta_t(t - 1) + \beta_x x_i + \epsilon_{it}, \quad (2.10)$$

where ϵ_{i1} and ϵ_{i2} are distributed as a bivariate normal with mean $\mathbf{0}$ and the covariance matrix as Σ . Based on the regression model, the unconditional mean of y_{it} is $E(y_{it}) = \beta_t(t - 1) + \beta_x \mu_x$.

We assume that both y_{i1} , y_{i2} are observed but x_i has some missing values. Here $\mathbf{u}_i = (y_{i1}, y_{i2}, x_i)$, and $E(\mathbf{u}_i) = (\mu_{y1}, \mu_{y2}, \mu_x)' = (\beta_x \mu_x, \beta_t + \beta_x \mu_x, \mu_x)'$. There are three parameters $(\beta_t, \beta_x, \mu_x)$, where the first two are the parameters of interest and the last one is the nuisance parameter. The missingness indicator is $R_i = (1, 1, R_{x_i})$, where $R_{x_i} = 1$ if x_i is observed and $R_{x_i} = 0$ if x_i is missing. There are two missing patterns here. One is $r_1 = (1, 1, 1)$ where the data are completed and another is $r_2 = (1, 1, 0)$ for incomplete

cases. Define

$$\dot{\mu}' = \begin{pmatrix} \frac{\partial \mu_{y1}}{\partial \beta_t} & \frac{\partial \mu_{y2}}{\partial \beta_t} & \frac{\partial \mu_x}{\partial \beta_t} \\ \frac{\partial \mu_{y1}}{\partial \beta_x} & \frac{\partial \mu_{y2}}{\partial \beta_x} & \frac{\partial \mu_x}{\partial \beta_x} \\ \frac{\partial \mu_{y1}}{\partial \mu_x} & \frac{\partial \mu_{y2}}{\partial \mu_x} & \frac{\partial \mu_x}{\partial \mu_x} \end{pmatrix}.$$

For the complete data where $r = 1$, we have

$$\begin{aligned} \dot{\mu}'_1 V_1^{-1}(u_1 - \mu_1) &= \sum_{i=1}^{n_1} \dot{\mu}'_{i1} V_{i1}^{-1}(u_{i1} - \mu_{i1}) \\ &= \begin{pmatrix} 0 & 1 & 0 \\ \mu_x & \mu_x & 0 \\ \beta_x & \beta_x & 1 \end{pmatrix} \begin{pmatrix} v_{y1} & v_{y12} & v_{xy1} \\ v_{y12} & v_{y2} & v_{xy2} \\ v_{xy1} & v_{xy2} & v_x \end{pmatrix}^{-1} \sum_{i=1}^{n_1} \begin{pmatrix} y_{i1} - \beta_x \mu_x \\ y_{i2} - (\beta_t + \beta_x \mu_x) \\ x_i - \mu_x \end{pmatrix}. \end{aligned}$$

For the missing pattern $r = 2$ where the covariance is missing, we have

$$\begin{aligned} \dot{\mu}'_2 V_2^{-1}(u_2 - \mu_2) &= \sum_{i=1}^{n_2} \dot{\mu}'_{i2} V_{i2}^{-1}(u_{i2} - \mu_{i2}) \\ &= \begin{pmatrix} 0 & 1 \\ \mu_x & \mu_x \\ \beta_x & \beta_x \end{pmatrix} \begin{pmatrix} v_{y1} & v_{y12} \\ v_{y12} & v_{y2} \end{pmatrix}^{-1} \sum_{i=1}^{n_2} \begin{pmatrix} y_{i1} - \beta_x \mu_x \\ y_{i2} - (\beta_t + \beta_x \mu_x) \end{pmatrix}. \end{aligned}$$

Therefore combining both missing patterns, we have estimating equations:

$$\sum_{r=1}^2 \dot{\mu}'_r V_r^{-1}(u_r - \mu_r) = 0.$$

The estimators of β_t , β_x , and μ are obtained by solving the above equations.

As we mentioned before, we may add cross products into \mathbf{u}_i to estimate all parameters. Here is an example. Let us consider model

$$y_{it} = \beta_0 + \beta_1 t + \beta_2 x_i + \epsilon_i,$$

where t is the time. Suppose there are two time points, and we have two outcomes \mathbf{y}_1 and \mathbf{y}_2 , and one covariate \mathbf{x} . There are four parameters needed to be estimated β_0 , β_1 , β_2 , and μ_x . If \mathbf{u}_i only includes single variables \mathbf{y}_{i1} , \mathbf{y}_{i2} and \mathbf{x}_i , then there are only three independent equations for four parameters. Therefore, we need to add the product of \mathbf{y}_{i1} and \mathbf{x}_i , and the product of \mathbf{y}_{i2} and \mathbf{x}_i to \mathbf{u}_i , which allows us to have enough equations to solve the parameters.

2.4 Application

In this section, we apply the proposed approach for existing data sets with missing values. The first data studies contracepting women. It has monotone missing responses. The second data set studies hip fracture women. It contains both missing response variables and covariates. We compare the proposed method with multiple imputation method (Paik, 1997, Lipsitz et al., 2000) and the weighted approach (Robins et al., 1995) for these data sets.

2.4.1 Contracepting Women Data

This data is a longitudinal randomized clinical trial studying contracepting women (Machin et al, 1988, Fitzmauris, 2000). In this trial 1151 women receive an injection of either 100 or 150 mg of depot-medroxyprogesterone acetate (DMPA) at the beginning of the study and receive three additional injections at 90-day intervals.

Throughout the study each woman is required to complete a menstrual diary. The diary data is used to generate a sequence of binary responses for each subject according to whether or not she has experienced amenorrhea in the four successive three-month intervals. There is monotone dropout missingness. More than 30% of subjects dropped out before the completion of the trial.

The marginal regression model (Fitzmauris, 2000) is formulated as

$$\text{logit } pr(y_{it} = 1|x_i) = \beta_0 + \beta_1(t-1) + \beta_2(t-1)^2 + \beta_3x_i + \beta_4x_i \times (t-1) + \beta_5x_i \times (t-1)^2, \quad (2.11)$$

where $t = 0, 1, 2, 3$ and x_i is the treatment. We denote $x_i = 0$ for DMPA dose 100mg and $x_i = 1$ for DMPA dose 150mg.

To illustrate our method, we assume covariate \mathbf{x} as random variable. There are seven parameters involved in the model with $\psi = (\beta_0, \beta_1, \beta_2, \beta_3, \beta_4, \beta_5, \mu_x)$. This requires at least seven variables in the \mathbf{u}_i . We use all response variables and covariates, and the cross products of y and x additionally,

$$\mathbf{u}_i = [y_{i1}, y_{i2}, y_{i3}, y_{i4}, x_i, (xy_1)_i, (xy_2)_i, (xy_3)_i, (xy_4)_i]. \quad (2.12)$$

There are four possible different missing patterns including complete case with $R = (1, 0, 0, 0), (1, 1, 0, 0), (1, 1, 1, 0), (1, 1, 1, 1)$. For different missing pattern, we use subset of (2.12) according to the missing pattern.

We compare our approach with the weighted approach and the multiple imputation. We briefly describe the weighted method here. In the weighted approach, the probability of being observed is modeled as

$$Pr(R_{it} = 1 | y_1, \dots, y_{it}, x_i) = \frac{e^{g_i' \gamma}}{1 + e^{g_i' \gamma}},$$

where g_i is a combination of the elements of $(y_{i1}, \dots, y_{it}, x_i)$. Specifically, for this application, we modeled the probability of y_{it} being observed condition on previous responses and covariate x_i as

$$\lambda_{it} = Pr(R_{it} = 1 | y_{i1}, \dots, y_{it-1}, x_i) = \frac{e^{\gamma_0 + \gamma_1 y_{i1} + \dots + \gamma_{t-1} y_{it-1} + \gamma x_i}}{1 + e^{\gamma_0 + \gamma_1 y_{i1} + \dots + \gamma_{t-1} y_{it-1} + \gamma x_i}}.$$

The weights W_i in (2.3) are calculated by taking the inverse of the above estimated probability of y_{it} being observed. The covariance matrix of $\hat{\beta}$ of the weighted GEE (Robins et al., 1995) is given by

$$\Sigma_{\beta} = \left(\sum_{i=1}^K \frac{\partial \mu_i'}{\partial \beta} V_i^{-1} W_i \frac{\partial \mu_i}{\partial \beta} \right)^{-1} \sum_{i=1}^K E_i E_i' \left(\sum_{i=1}^K \frac{\partial \mu_i'}{\partial \beta} V_i^{-1} W_i \frac{\partial \mu_i}{\partial \beta} \right)^{-1},$$

where $E_i = U_i - \left(\sum_{i=1}^K U_i S_i' \right) \left(\sum_{i=1}^K S_i S_i' \right)^{-1} S_i$, $U_i = \frac{\partial \mu_i'}{\partial \beta} V_i^{-1} (Y_i - \mu_i(\hat{\beta}))$, and $S_i = \sum_t R_{it-1} Z_{it} (R_{it} - \lambda_{it})$ is the score component for the i^{th} subject from the missingness mechanism model (Robins et al., 1995, Preisser et al., 2002).

For the multiple imputation, we replace the missing values sequentially based on the imputation using logistic model:

$$Pr(y_{it} = 1 | y_1, \dots, y_{it}, x_i) = \frac{e^{\alpha_0 + \alpha_1 y_{i1} + \dots + \alpha_{t-1} y_{it-1} + \alpha x_i}}{1 + e^{\alpha_0 + \alpha_1 y_{i1} + \dots + \alpha_{t-1} y_{it-1} + \alpha x_i}}.$$

The estimator of multiple imputation is the average of estimators from each imputation. The covariance of estimator (Rubin, 1978) is given by

$$\Sigma(\beta) = \bar{U} + \left(1 + \frac{1}{m}\right) B,$$

where m is the number of the multiple imputations, and $\bar{U} = \frac{1}{m} \sum_{j=1}^m U_j$ is the average of within-imputation variance, where U_j is the variance associated with β_j with

$$U_j = \left(\sum_{i=1}^K \frac{\partial \mu'_i}{\partial \beta} V_i^{-1} \frac{\partial \mu_i}{\partial \beta} \right)^{-1} \sum_{i=1}^K \frac{\partial \mu'_i}{\partial \beta} V_i^{-1} (Y_i - \mu_i(\hat{\beta})) (Y_i - \mu_i(\hat{\beta}))' V_i^{-1} \frac{\partial \mu_i}{\partial \beta} \left(\sum_{i=1}^K \frac{\partial \mu'_i}{\partial \beta} V_i^{-1} \frac{\partial \mu_i}{\partial \beta} \right)^{-1}.$$

Here $B = \frac{1}{m-1} \sum_{j=1}^m (\beta_j - \bar{\beta})^2$ is the variance between imputations.

Table 1 shows that multiple imputation provides highest standard errors among three approaches. The proposed approach and the weighted method provide similar estimates. However, the weighted method provides the lowest standard errors among these three approaches.

2.4.2 Arbitrary Missing Pattern Data

The second example investigates (Allison, 2001) 220 caucasian women, at least 60 years old, who are treated surgically for hip fractures in Philadelphia area (Mossey, Knott, and Craik, 1990). They are interviewed three times after the hospital release: at 2 months, 6 months, and 12 months. The outcome variable is a measurement of depression, on a scale from 0 to 60. The four related covariates are the number of self-care that could be completed without assistance (x_1 , ranges from 0 to 3), the degree of pain experienced by the patient (x_2 , ranges from 0 (none) to 6 (constant)), self-rated health (x_3 , measured on a four-point scale, 1 = poor, 4 = excellent), and whether able to walk without aid at home (x_4 , 0 = no and 1 = yes).

The goal of this study is to investigate the relation between depression and the other four predictors. The linear regression model is given by

$$y_{it} = \beta_1 x_{it1} + \dots + \beta_4 x_{it4} + \beta_5 w_1 + \beta_6 w_2 + \epsilon_{it},$$

where y_{it} is the depression score for the person i at time t , w_1 and w_2 are two indicator variables for time ($w_1 = 1$ if time at 2 months, and $w_1 = 0$ otherwise. $w_2 = 1$ if time at 6 months and $w_1 = 0$ otherwise.), and the error $\epsilon_i = (\epsilon_{i1}, \dots, \epsilon_{it})$ is distributed as mean $\mathbf{0}$ and variance Σ . Here except x_4 is binary variable and the rest of covariates are continuous variables.

For this data example, the missing values have arbitrary missing pattern (intermittent missing) so that the WEE is not applicable. We compared the proposed approach with the multiple imputation (Paik, 1997, Lipsitz et al., 2000). We imputed the missing values for month 2 first then for month 6 and month 12 and within the same time period, we imputed the missing values in covariates first, and then the response variables. We apply the GEE using imputed values along with the observed values.

Table 2 provides the estimates and standard errors for the estimators. The estimate for x_3 is similar for both approaches. The estimates for time effect (coefficients of w_1 and w_2) are significant from both approaches. The number of self-care (x_1) and the degree of pain (x_2) do not have significant effect on depression score based on our proposed approach while they are significant effect based on the multiple imputation approach.

2.4.3 Missing Covariate

We use the subset of the second data example to create missing covariates. We select 138 subjects with complete responses. We consider a relatively simpler model and is given by

$$y_{it} = \beta_0 + \beta_1 x_{it2} + \beta_2 w_1 + \beta_3 w_2 + \epsilon_{it} .$$

We make the missing patterns of x_2 as monotone. There are about 24% of the covariate missing. We compare the proposed approach with the multiple imputation (Paik, 1997, Lipsitz et al., 2000) and the weighted GEE (Robins et al., 1994). For the multiple imputation approach, we imputed missing values for $t = 1$ first then for $t = 2$ and 3. We apply the GEE based on imputed values and observed data.

For the weighted method, the probability of being observed is modeled as

$$Pr(R_{it} = 1 | y_1, \dots, y_{it}, x_{i1}, \dots, x_{it}) = \frac{e^{g'_i \gamma}}{1 + e^{g'_i \gamma}} ,$$

where g_i is a combination of elements in $(y_{i1}, \dots, y_{it-1}, x_{i2})$.

Table 3 provides the estimates and standard errors for all approaches. The estimates of all three approaches are similar and all of them show significance of all effects. The standard error of estimators are similar.

2.5 Simulations

We conduct several simulations to study the finite sample performance of the unconditional estimating equation approach. The first two simulations are artificial and the other two are based on application data. We compare the proposed approach with the WEE and imputation method. For the first two simulations, the complete data are generated from sample size $N = 250$ for each simulation. There is only one covariate x and it had missing values. All responses are observed. The true missingness mechanism is specified such that the missingness depends on the observed response variables. This ensures that the missingness is MAR. The true missingness mechanism follows logistic model

$$\text{logit } pr(R_i = 1|y_{it}, x_i) = \alpha_0 + \alpha_1 y_{i1} + \alpha_2 y_{i2} + \alpha_3 y_{i3} . \quad (2.13)$$

2.5.1 Continuous Response

In this section, we simulate response variables and covariates both as continuous. We generate x_i from normal distribution with $\mu_x = 2$ and $\sigma_x^2 = 1$. The response variables are assumed to follow the linear model

$$y_{it} = \beta_t(t - 1) + \beta_x x_i + \epsilon_{it} , \quad t = 1, 2, 3$$

The error term $(\epsilon_{i1}, \epsilon_{i2}, \epsilon_{i3})$ are generated to be correlated with exchangeable correlation structure (ρ) . We set $\beta_t = 0.5$, $\beta_x = 1$, and $\alpha_1 = -1$, $\alpha_2 = 1$ in (2.13). We study the performance under different correlations. We set $\rho = 0.25$ for low, $\rho = 0.5$ for median, and $\rho = 0.75$ for high correlation. We chose different values of α_0 and α_3 so that the amount of missing data varies. The bias of the estimators and the square root of mean square error (RMSE) are given in Table 4. The standard deviations of the estimators are not in the table because they are very close to RMSE. The missing rates are about 36% and 60% for the two setups.

Table 4 indicates that the proposed method and the multiple imputation method yield nearly unbiased estimators of β_t and β_x with the biases for β_x less than 0.2% and for the bias for β_t always less than 0.4%. Among these approaches, the weighted GEE has the highest RMSE. The RMSE of the proposed approach is smaller than or equal

to that of multiple imputation for both β_t and β_x . In addition, when the correlation among responses increases, the proposed approach tends to have smaller RMSE for β_t and similar RMSE for β_x . When the missing rate increases, the RMSE for all three approaches decreases. Moreover, the RMSE of the proposed approach is slightly smaller than that of multiple imputation for both β_t and β_x .

2.5.2 Binary Response

In this section, we simulated both response variables and covariate as binary. Here x_i 's are generated from a Bernoulli distribution with the probability of success as 0.5. The marginal model for $E(y_{it}|x_i)$ is given as follows

$$\text{logit } pr(y_{it} = 1|x_i) = \beta_t(t - 1) + \beta_x x_i, \quad t = 1, 2$$

The response variables (y_{i1}, y_{i2}) are generated to be correlated with exchangeable correlation structure with correlation ρ . We set $\beta_t = 0.5$, $\beta_x = 1$, $\alpha_1 = 1$, and $\alpha_3 = 0$. Similar to the last simulation, we set $\rho = 0.25$ for low, $\rho = 0.5$ for median, and $\rho = 0.75$ for high correlation, we choose different values of α_0 and α_2 so that the amount of missing data varies.

The bias and square root of mean square error are given in Table 5. The missing rates are around 30% and 60% for the two setups, respectively. Table 5 indicates that the proposed method yields nearly unbiased estimators of β_t and β_x with the biases for β_t are less than 1% and β_x are less than 3%. The WEE and the multiple imputation approach yield similar biases when correlation among response is low or median. When the correlation is high, the biases of the proposed estimators are smaller than those of the other two approaches. Among these approaches, the RMSE of the proposed approach is less than or equal to that of weighted GEE or multiple imputation. When the correlation is high, the ratios of RMSE of the proposed approach relative to the multiple imputation or the weighted approach are nearly 0.5 for $\hat{\beta}_t$ and are less than 0.82 for $\hat{\beta}_x$.

2.5.3 Simulation Based on Contracepting Women Data

We simulate data based on the women contracepting data with the sample size as 250. Both response variables and covariate are binary variables. The response variables given the covariate x follow a logistic model:

$$\text{logit } pr(y_{it} = 1) = \beta_0 + \beta_t(t - 1) + \beta_x x_i \quad t = 1, \dots, 4.$$

Correlated response variables are generated following the procedure by Preisser (2002) and the covariate is generated by Bernoulli distribution with the mean as the sample mean of x from the women contracepting data. The missing values occur in responses only. The true missingness mechanism is specified such that the missingness is monotone and depends on the previous observed outcome. This ensured that the missing is MAR. The missingness indicator R_{it} is 1 if y_{it} is observed and 0 if y_{it} is missing, where $t = 2, 3, 4$. There is no missing in y_1 . The true missingness mechanism follows logistic model and only depends on the previous response y .

$$\text{logit } pr(R_{it} = 1 | y_{i1}, \dots, y_{it}, x_i) = \alpha_0 + \alpha_1 y_{i(t-1)}. \quad (2.14)$$

We choose α such that the missing rate is about 27% and also with another missing rate as 53%).

The bias and standard error of the estimators are provided in Table 6. The multiple imputation and the weighted GEE have similar biases when the model for the missingness mechanism or the model for imputing missing values are correctly specified. The proposed approach gives slightly larger biases. The standard errors for all three approaches are similar. But when the model of missingness mechanism or the model for imputing missing values are misspecified, the proposed approach has smaller biases than the WEE and the multiple imputation estimators. When sample size increases, the variances of estimators from all three approaches are similar.

2.5.4 Simulation Based on Monotone Missing Covariate Example

We simulate data based on hip fracture data with missing covariates. The mean of response (y_1, y_2, y_3) given covariate x_{it3} is

$$E(y_{it}) = \beta_0 + \beta_1 x_{it2} + \beta_2 w_1 + \beta_3 w_2 , \quad (2.15)$$

and the covariance is Σ_y . Three covariates are generated by multinormal distribution with mean $(\mu_{x_1}, \mu_{x_2}, \mu_{x_3})$ and covariance Σ_x . They are truncated according to the range of x_1 , x_2 , and x_3 . We use x_2 as covariate in (2.15) and the other two covariates for modeling missingness mechanism.

The means and covariances of y and covariates are based on the sample mean and variance of the data. The missingness only occurs in the covariate. The true missingness mechanism is specified as monotone missing given by

$$\begin{cases} \text{logit}Pr(R_{it} = 1) = \alpha_0 + \alpha_1 y_{it} + \alpha_2 x_{it1} + \alpha_3 x_{it2}, \\ Pr(R_{it} = 1) = 0 \quad \text{if } R_{it-1} = 0 . \end{cases}$$

All α 's are obtained based on the information of observed data. The sample size for this simulation is 250 and the missing rate are 27% and 41% respectively. The bias and standard error are given in Table 7. The biases of the proposed approach are equal to or smaller than those of WEE and multiple imputation. The standard errors of the proposed approach are equal to or slightly larger than those of WEE and multiple imputation. If the model for the weights or the model for the imputation are misspecified, the proposed approach has smaller biases than the other two approaches. For two missing rates, all three approaches have similar trends, with more biases and larger standard errors of estimators when the missing rate increases.

2.5.5 Simulation Based on Arbitrary Missing Pattern

In this simulation, the data is generated based on the hip fracture example. The response variables (y_1, y_2, y_3) are generated with the mean

$$E(y_{it}) = \beta_1 x_{it3} + \beta_2 w_1 + \beta_3 w_2 ,$$

and a random error with variance $(\sigma_{y_1}^2, \sigma_{y_2}^2, \sigma_{y_3}^2)$ and exchangeable correlation structure. We set $\rho = 0.25, 0.5$, and 0.75 for low, median, and high correlation. Three covariates are generated by multinormal distribution with mean $(\mu_{x_1}, \mu_{x_2}, \mu_{x_3})$ and covariance Σ_x . They are truncated according to the range of x_1, x_2 , and x_3 . We use x_3 as covariate in (2.15) and the other two covariates for modeling missingness mechanism.

The means, variances and covariances of the responses and covariates are based on the sample mean and variance of the data. The missing values occur in both covariates and responses. The true missingness mechanism is modeled as

$$Pr(R_{it} = 1) = \alpha_0 + \alpha_2 x_{it1} + \alpha_3 x_{it2} ,$$

and $R_{x_{t3}}$ and R_{y_t} are generated with the same correlation as from the hip fracture data. All α 's are obtained from the information of the observed data. The sample size for this simulation is 200.

Since the weighted approach can not handle this kind of missing data, Table 8 provides the bias and standard errors for our approach and the imputation method. Our proposed approach has smaller biases and standard errors than multiple imputation for time effect. As for covariate, the proposed estimator has smaller bias than multiple imputation estimator, but the standard error of our estimator is similar to or slightly larger than that of multiple imputation.

2.6 Discussion and Conclusion

The unconditional estimating equation approach is mainly based on unconditional expectation of all possible variables including the responses and covariates and relies on approximation to the multivariate normal distribution. The main advantage of the unconditional estimating equation approach is that it does not require modeling for missingness mechanism or imputing the missing values.

Our simulation studies for different missing situations suggest that the proposed approach has negligible bias and the bias was found to be less than 2% in general. The variance of the proposed estimators is smaller than or equal to that of the weighted and

the imputation approaches in most cases.

Furthermore, the proposed method can be easily applied to arbitrary missing patterns. While the WEE can not be easily extended in this case. The sequential imputation approach (Paik, 1997) can be extended to arbitrary missing patterns. but it does not perform as well as the proposed approach based on the simulation results.

There are some drawbacks of the proposed approach. In particular, it does not perform well for the models when matrix D is close to singular, because there can be convergence problem. But this problem could be solved by including cross products of response variables and covariates in data matrix.

Reference

1. Afifi, A. and Elashoff, R. (1966) Missing observations in multivariate statistics I: Review of the literature. *Journal of the American Statistical Association*, 61, 595-604.
2. Allison, P. (2001) *Missing data*, Thousand Oaks, CA: Sage.
3. Dempster, A. P., Laird, N. M. and Rubin, D. B. (1977) Maximum likelihood from incomplete data via the EM algorithm (with discussion). *Journal of the Royal Statistical Society Series B (Statistical Methodology)*, 39, 1-38.
4. Greene, H. (2000) *Econometric analysis*, 4th ed, Englewood Cliffs, NJ: Prentice Hall.
5. Horvitz, D. and Thompson, D. (1952) A Generalization of Sampling without replacement from a Finite Universe, *J. Amer. Statist. Assoc.*, 47, 663-685.
6. Laird, N. (1988) Missing data in longitudinal studies. *Statistics in Medicine*, 7, 305-315.
7. Lipsitz S., Ibrahim J., Zhao L. (1999) A weighted estimating equation for missing covariate data with properties similar to maximum likelihood, *J. Am. Stat. Assoc.*, 94, 1147-60.
8. Lipsitz S., Molenberghs G., Fitzmaurice, G., and Ibrahim, J. (2000) GEE with Gaussian estimation of the correlations when data are incomplete, *Biometrics*, 56, 528-536.
9. Little, R. and Rubin, D. (2002) *Statistical Analysis with Missing Data*, Wiley Series in Probability and Statistics, 2nd edition.
10. McCullagh, P. and Nelder, J. (1983) *Generalized Linear Models*. Chapman and Hall.
11. Mossey, J., Knotf K., and Crak, R. (1990) Effects of persistent depressive symptoms on hip fracture recovery, *Journal of Gerontology: Medical Sciences*, 45, 163-168.
12. Nie, Hull, Jenkins, Steinbrenner, and Bent, (1975) *SPSS: Statistical Package for the Social Sciences*, 2nd ed, New York, McGraw-Hill.
13. Paik, M. (1997) The generalized estimating equation approach when data are not missing completely at random, *Journal of the American Statistical Association*, 92, 1320-1329.
14. Preisser, J., Lohman, K., and Rathouz, P. (2002) Performance of weighted estimating equations for longitudinal binary data with drop-outs missing at random, *Statistics in Medicine*, 21, 3035-54.
15. Robins, J., Rotnitzky, A. (1995) Semiparametric efficiency in multivariate regression models with missing data, *Journal of the American Statistical Association*, 90, 106 -121.
16. Robins, J., Rotnitzky, A., Zhao, L. (1994) Estimation of regression coefficients when some regressors are not always observed, *J. Am. Stat. Assoc.*, 89, 846-66.
17. Rosenbaum, P. and Rubin, D. (1983) The central role of the propensity score in observational studies for causal effects, *Biometrika*, 70, 41-55.
18. Rosenbaum, P. and Rubin, D. (1985) Constructing a control group using multivariate matched sampling methods that incorporate the propensity score, *American Statistician*, 39, 33-38.
19. Rotnitzky, A., Robins, J., and Scharfstein, D. (1998) Semiparametric regression for repeated outcomes with nonignorable nonresponse, *Journal of the American Statistical Association*, 93, 1321-39.
20. Rubin, D. (1976) *Inference and Missing Data*, *Biometrika*, 63, 581-92.

21. Rubin, D. (1977) Formalizing Subjective Notion about the Effect of Nonrespondents in Sample Surveys, *Journal of the American Statistical Association*, 72, 538-43.
22. Rubin, D. (1987) Multiple Imputation in Sample Surveys, *Proc. Survey Res. Meth. Sec.*, Am. Stat. Assoc. 1978, 20-34.
23. Rubin, D. (1987) *Multiple Imputation for Nonresponse in Surveys*, New York: John Wiley & Sons.
24. Rubin, D. (1996) Multiple Imputation after 18+ Years (with discussion) *Journal of the American Statistical Association*, 91, 473-489.
25. Schafer, J. (1997) *Analysis of incomplete multivariate data*, London: Chapman and Hall.
26. Scharfstein, D., Rotnitzky, A. and Robins, J. (1999) Adjusting for nonignorable drop-out using semi-parametric nonresponse models, *Journal of the American Statistical Association*, 94, 1096-1146.
27. Xie, F. and Paik, M. (1997) Generalized estimating equation model for binary outcomes with missing covariates, *Biometrics*, 53, 1458-66.

Table 1: Estimates and Standard Errors for Contracepting Women Data

		estimate	SE	Z	P-value
Int	$\hat{\beta}_{MI}$	-1.5537	0.1074	-14.4665	0.0000
	$\hat{\beta}_w$	-1.4923	0.0967	-15.4323	0.0000
	$\hat{\beta}_{prop}$	-1.4890	0.1053	-14.1406	0.0000
time	$\hat{\beta}_{MI}$	0.6511	0.1894	3.4377	0.0006
	$\hat{\beta}_w$	0.5335	0.1073	4.9720	0.0000
	$\hat{\beta}_{prop}$	0.5139	0.1305	3.9379	0.0001
t^2	$\hat{\beta}_{MI}$	0.0299	0.0604	0.4950	0.6206
	$\hat{\beta}_w$	-0.0058	0.0256	-0.2266	0.8207
	$\hat{\beta}_{prop}$	0.0019	0.0398	0.0477	0.9620
trt	$\hat{\beta}_{MI}$	0.1024	0.1553	0.6594	0.5096
	$\hat{\beta}_w$	0.1057	0.1395	0.7577	0.4486
	$\hat{\beta}_{prop}$	0.1057	0.1412	0.7486	0.4541
trt*t	$\hat{\beta}_{MI}$	0.4734	0.2763	1.7134	0.0866
	$\hat{\beta}_w$	0.4069	0.1559	2.6100	0.0091
	$\hat{\beta}_{prop}$	0.4248	0.1829	2.3226	0.0202
trt*t ²	$\hat{\beta}_{MI}$	-0.1446	0.0877	-1.6488	0.0992
	$\hat{\beta}_w$	-0.1318	0.0374	-3.5241	0.0004
	$\hat{\beta}_{prop}$	-0.1271	0.0559	-2.2737	0.0230

Table 2: Estimates and Standard Errors for Arbitrary Missing Pattern Data

		estimate	SE	Z	P-value
x_1	$\hat{\beta}_{prop}$	0.2241	0.2314	0.9685	0.8336
	$\hat{\beta}_{MI}$	-0.5702	0.2122	-2.6871	0.0036
x_2	$\hat{\beta}_{prop}$	-0.0665	0.2475	-0.2687	0.3941
	$\hat{\beta}_{MI}$	0.4240	0.1316	3.2219	0.0006
x_3	$\hat{\beta}_{prop}$	2.3678	0.2440	9.7041	0.0000
	$\hat{\beta}_{MI}$	2.63289	0.2268	11.6089	0.0000
x_4	$\hat{\beta}_{prop}$	-2.8666	0.7584	-3.7798	0.0001
	$\hat{\beta}_{MI}$	-1.9252	0.4918	-3.9146	0.0001
w_1	$\hat{\beta}_{prop}$	6.6085	0.4521	14.6173	0.0000
	$\hat{\beta}_{MI}$	8.6531	0.7395	11.7013	0.0000
w_2	$\hat{\beta}_{prop}$	5.5184	0.4090	13.4924	0.0000
	$\hat{\beta}_{MI}$	8.0443	0.5180	15.5295	0.0000

Table 3: Estimates and standard errors for missing covariate

		estimate	SE	Z	P-value
Int	$\hat{\beta}_w$	2.3609	0.5036	4.6880	0.0000
	$\hat{\beta}_{MI}$	2.2801	0.4895	4.6580	0.0000
	$\hat{\beta}_{prop}$	2.4715	0.4717	5.2396	0.0000
x_2	$\hat{\beta}_w$	0.9479	0.2153	4.4027	0.0000
	$\hat{\beta}_{MI}$	0.9462	0.2115	4.4738	0.0000
	$\hat{\beta}_{prop}$	0.9143	0.2173	3.3852	0.0003
w_1	$\hat{\beta}_w$	9.4847	0.5410	17.5318	0.0000
	$\hat{\beta}_{MI}$	9.1413	0.5792	15.7826	0.0000
	$\hat{\beta}_{prop}$	9.3067	0.5629	16.5335	0.0000
w_2	$\hat{\beta}_w$	8.4414	0.5180	16.2961	0.0000
	$\hat{\beta}_{MI}$	8.0565	0.5043	15.9756	0.0000
	$\hat{\beta}_{prop}$	8.4380	0.4937	17.0914	0.0000

Table 4: Simulation results for continuous data based on 250 sample and 1000 simulations

$\beta_t=0.5$	$\beta_x=1$	$\rho=0.25$			$\rho=0.5$			$\rho=0.75$		
		$\hat{\beta}_w$	$\hat{\beta}_{MI}$	$\hat{\beta}_p$	$\hat{\beta}_w$	$\hat{\beta}_{MI}$	$\hat{\beta}_p$	$\hat{\beta}_w$	$\hat{\beta}_{MI}$	$\hat{\beta}_p$
$\alpha_0=-1.1, \alpha_3=0$ missr=36 %	$bias_{\beta_t}$	0.0055	0.0010	0.0012	0.0028	0.0018	0.0014	0.0021	0.0002	0.0005
	$RMSE_{\beta_t}$	0.0767	0.0412	0.0398	0.0565	0.0364	0.0320	0.0408	0.0308	0.0222
	$bias_{\beta_x}$	0.0006	0.0000	0.0000	0.0000	0.0002	0.0002	0.0007	0.0001	0.0003
$\alpha_0=-1, \alpha_3=0.67$ missr=60 %	$RMSE_{\beta_x}$	0.0221	0.0147	0.0149	0.0206	0.0155	0.0153	0.0195	0.0163	0.0154
	$bias_{\beta_t}$	0.0039	0.0039	0.0037	0.0004	0.0002	0.0013	0.0003	0.0012	0.0013
	$RMSE_{\beta_t}$	0.0558	0.0390	0.0382	0.0423	0.0354	0.0333	0.0313	0.0287	0.0223
	$bias_{\beta_x}$	0.0005	0.0005	0.0005	0.0005	0.0000	0.0004	0.0002	0.0000	0.0003
	$RMSE_{\beta_x}$	0.0159	0.0130	0.0130	0.0148	0.0138	0.0134	0.0146	0.0143	0.0134

Note: missr represents the missing rate of the data, e.g. 36% means 36% subjects have missing values.

Table 5: Simulation results for binary data based on 250 sample size and 1000 simulations

		$\rho=0.25$			$\rho=0.5$			$\rho=0.75$		
		$\hat{\beta}_w$	$\hat{\beta}_{MI}$	$\hat{\beta}_p$	$\hat{\beta}_w$	$\hat{\beta}_{MI}$	$\hat{\beta}_p$	$\hat{\beta}_w$	$\hat{\beta}_{MI}$	$\hat{\beta}_p$
$\alpha_0=-1.1, \alpha_3=0$ missr=32 %	$bias_{\beta_t}$	0.0109	0.0041	0.0061	0.0158	0.0094	0.0138	0.0045	0.0064	0.0019
	$RMSE_{\beta_t}$	0.1838	0.1722	0.1679	0.1789	0.1722	0.1360	0.1744	0.1755	0.0965
	$bias_{\beta_x}$	0.0137	0.0367	0.0087	0.0413	0.0676	0.0406	0.0540	0.0614	0.0296
$\alpha_0=-1, \alpha_3=0.67$ missr=60 %	$RMSE_{\beta_x}$	0.2868	0.2765	0.3048	0.3346	0.3412	0.3078	0.3599	0.3678	0.2940
	$bias_{\beta_t}$	0.0017	0.0028	0.0049	0.0050	0.0010	0.0075	0.0011	0.0024	0.0002
	$RMSE_{\beta_t}$	0.2025	0.1809	0.1636	0.1978	0.1840	0.1360	0.1842	0.1834	0.0948
	$bias_{\beta_x}$	0.0227	0.0356	0.0095	0.0416	0.0544	0.0261	0.0304	0.0395	0.0183
	$RMSE_{\beta_x}$	0.3140	0.3118	0.2899	0.3619	0.3763	0.3172	0.3755	0.3901	0.3147

Note: missr represents the missing rate of the data, e.g. 32% means 32% subjects have missing values.

Table 6: Estimates and standard error for the simulation based on contracepting women data with sample size 250

	WEE	WEE(Y)	MI	MI(Y)	Prop	true β	
mir =27%	int	-1.5230(0.1644)	-1.4674(0.1640)	-1.5219(0.1632)	-1.4657(0.1641)	-1.5422(0.1656)	-1.5072
	t	0.5103(0.0640)	0.4672(0.0700)	0.5140(0.0562)	0.4720(0.0654)	0.5179(0.0596)	0.5114
	x	0.2599(0.2064)	0.2783(0.1811)	0.2624(0.1801)	0.2777(0.1775)	0.2700(0.1810)	0.2653
mir =53%	int	-1.5165(0.1909)	-1.4675(0.2040)	-1.5172(0.1868)	-1.4287(0.2136)	-1.5120(0.1948)	-1.5072
	t	0.4922(0.0842)	0.4719(0.0806)	0.5123(0.0731)	0.4707(0.0731)	0.4910(0.0720)	0.5114
	x	0.2661(0.2329)	0.2628(0.2443)	0.2705(0.2220)	0.2151(0.2141)	0.2544(0.2417)	0.2653

Note: mir represents the missing rate of the data, e.g. 27% means 27% subjects have missing values.

Table 7: Biases and standard errors for the simulation based on missing covariates with sample size 250

	WEE	WEE(Y)	MI	MI(Y)	Prop	true β	
mir=26 %	int	0.6108(0.3286)	0.7448(0.3534)	0.8226(0.3156)	1.1698(0.2913)	0.5177(0.3433)	2.6419
	x	-0.0836(0.1512)	-0.0918(0.1580)	-0.3099(0.1371)	-0.4886(0.1256)	-0.1111(0.1551)	1.4937
	t1	-0.4100(0.4229)	-0.5316(0.4412)	-0.2995(0.5427)	-0.3562(0.5418)	-0.2965(0.4258)	9.9432
	t2	-0.3286(0.3477)	-0.4848(0.3669)	-0.2871(0.3448)	-0.3356(0.3473)	-0.2472(0.3492)	8.0057
mir=41 %	int	0.6550(0.3447)	0.7988(0.3779)	0.9488(0.3105)	1.0995(0.2939)	0.5314(0.3643)	2.6419
	x	-0.0942(0.1656)	-0.0906(0.1694)	-0.4045(0.1355)	-0.6195(0.1254)	-0.1173(0.1688)	1.4937
	t1	-0.4621(0.3952)	-0.6115(0.4223)	-0.2823(0.3926)	-0.4051(0.3952)	-0.3002(0.4015)	9.9432
	t2	-0.3740(0.3575)	-0.5662(0.3869)	-0.3250(0.3542)	-0.3935(0.3650)	-0.2755(0.3585)	8.0057

Note: mir represents the missing rate of the data, e.g. 26% means 26% subjects have missing values.

Table 8: Simulation results for continuous data based on missing covariates data with sample size of 200 and simulation size of 1000

		$\beta_x=2.3853$	$\beta_{t_1}=8.2119$	$\beta_{t_2}=6.3340$			
		$\rho=0.25$		$\rho=0.5$		$\rho=0.75$	
		$\hat{\beta}_{MI}$	$\hat{\beta}_p$	$\hat{\beta}_{MI}$	$\hat{\beta}_p$	$\hat{\beta}_{MI}$	$\hat{\beta}_p$
missr=37 %	$\hat{\beta}_x$	2.5824	2.5048	2.5906	2.5067	2.5836	2.5025
	std_{β_x}	0.1310	0.1363	0.1334	0.1362	0.1363	0.1361
	$\hat{\beta}_{t_1}$	9.2439	7.9044	9.2418	7.9319	9.2374	7.9562
	$std_{\beta_{t_1}}$	0.6321	0.6036	0.6086	0.5291	0.5599	0.4811
	$\hat{\beta}_{t_2}$	6.9323	6.1496	6.9359	6.1885	6.9263	6.2094
	$std_{\beta_{t_2}}$	0.5866	0.5315	0.5633	0.4504	0.5621	0.3740
missr=53 %	$\hat{\beta}_x$	2.6814	2.4850	2.6947	2.5053	2.6940	2.4966
	std_{β_x}	0.1427	0.1489	0.1439	0.1511	0.1425	0.1477
	$\hat{\beta}_{t_1}$	9.3730	7.9265	9.3702	7.9353	9.3620	7.9322
	$std_{\beta_{t_1}}$	0.6696	0.6042	0.6629	0.5287	0.6217	0.4802
	$\hat{\beta}_{t_2}$	7.2428	6.2161	7.2863	6.2570	7.2445	6.2489
	$std_{\beta_{t_2}}$	0.6550	0.5221	0.5907	0.4631	0.4955	0.3745

Note: missr represents the missing rate of the data, e.g. 37% means 37% subjects have missing values.

Unconditional Estimating Equation Approach for Nonignorable Missingness

Lin Lu, David Birkes, Annie Qu

Abstract

Given regression data from a sample of independent subjects, we propose a new approach to estimating regression coefficients when some data are missing and the missingness mechanism is nonignorable. For a complete data set, with no missing values, an estimate of the vector of regression coefficients can be obtained as the solution to a generalized estimating equation (GEE) based on the difference between the vector of outcomes and its expectation conditional on the covariates. When data are missing, our approach uses an “unconditional” generalized estimating equation, based on the difference between a vector of “extended-sense” outcomes and its unconditional expectation. Extended-sense outcomes include (1) the original outcomes in the complete data set multiplied by their missingness indicators, (2) the covariates in the complete data set multiplied by their missingness indicators, (3) the missingness indicators, and (4) selected products of these variables. The advantage of an estimating-equation approach is that fewer distributional assumptions are needed in comparison to maximum likelihood (ML) and explicitly modeled multiple imputation (MI). Weighted estimating-equation (WEE) estimators have the same advantage of requiring few distributional assumptions, but our estimator can be easier to obtain in some situations. Through simulation we investigate the efficiency of the proposed estimator relative to ML, MI, and WEE estimators. The unconditional estimating-equation (UEE) estimator is as efficient as the other three estimators when the missingness mechanism is correctly specified.

3.1 Introduction

Missing data occur frequently in scientific research, for many reasons, such as if subjects drop out of a longitudinal study before its completion, or if the data is missing by human error. The missing-data mechanism could be ignorable, where missingness only depends on observed information, or it could be nonignorable (or informative), where missingness depends on unobserved information. Considerable research has been done on statistical methods for handling ignorable missingness; see Little and Rubin (2002), Schafer (1997), Allison (2002), Ibrahim, Chen, Lipsitz, and Herring (2005), Robins, Rotnitzky, and Zhao (1994, 1995), Davidian, Tsiatis, and Leon (2005).

The problem of nonignorable missingness is appreciably more difficult to deal with since the missingness mechanism depends on missing values, which of course are unobserved. Existing methods for nonignorable missingness include likelihood methods (Diggle and Kenward 1994; Ibrahim, Lipsitz, and Chen 1999), Bayesian analyses (Huang, Chen, and Ibrahim 2005), multiple imputation methods (Paik 1997), and weighted GEE methods (Rotnitzky, Robins, and Scharfstein 1998). Maximum likelihood (ML) requires specification of a full joint likelihood for the data, including the missingness indicators. A Bayesian approach also requires a full likelihood, plus specification of a prior distribution on the parameters. Multiple imputation (MI) in many settings (Kenward and Carpenter 2007, section 4) involves specification of the conditional distribution of the missing data given the observed data and sometimes the missingness indicators. Some versions of MI (Paik 1997), weighted estimating-equation (WEE) estimation and the unconditional estimating-equation (UEE) we proposed require only that certain moments of the data be specified.

In this paper, we propose a semiparametric approach to handle missing data. The procedure is partially parametric in so far as moments of the data, including the missingness indicators, are expressed in parametric form, and it is partially nonparametric in so far as no further distributional assumptions are made. When no data are missing, a commonly used method for estimating regression coefficients in a generalized linear model is to solve a generalized estimating equation (GEE) (Liang and Zeger 1986). A GEE is based on the difference between the vector of outcomes and its expectation conditional

on the covariates, which are regarded as fixed. In the proposed method, the covariates are regarded as random. Our unconditional estimating equation (UEE) is based on the difference between a vector of “extended-sense” outcomes and its unconditional expectation. Extended-sense outcomes include (1) the original outcomes in the complete data set multiplied by their missingness indicators, (2) the covariates in the complete data set multiplied by their missingness indicators, (3) the missingness indicators, and (4) selected products of these variables.

Treating the covariates as extended-sense outcomes is suggested by Kenward and Carpenter (2007, p. 203). Inclusion of the missingness indicators as extended-sense outcomes is similar to the joint estimating-equation approach of Zhao, Lipsitz, and Lew (1996), but their estimating equation is based on conditional expectations. The idea of an extended-sense outcome vector including an original outcome, a covariate, and a missingness indicator appears in the weighted estimating equation derived by Lipsitz, Ibrahim, and Zhao (1999).

In a typical ML estimation, one must postulate a parametric model for the joint distribution of the data and missingness indicators. In multiple imputation (MI), the imputation is often done by postulating a parametric model for simulating missing values given the observed data. Estimating-equation approaches such as WEE and UEE are semiparametric in that a parametric expression is required only for certain moments of the data and missingness indicators, and no further restrictions are placed on the distribution of the data. To be assured that an ML or explicitly modeled MI estimator is consistent, the postulated distributions should be true. For a WEE or UEE estimator to be consistent, one needs only the weaker condition that the postulated expressions for the moments are true, regardless of the distribution of the “residuals”. Here the moment is the expectation of some quantity, namely, a power of an observation. The corresponding “residual” is the difference between the quantity and the moment.

This article is organized as follows. The GEE, ML, MI, and WEE estimation procedures are discussed further in Section 3.2. In Section 3.3 we give the model and notation and introduce the UEE estimator. Its asymptotic properties are stated in Theorem 3.1.

The proposed method is illustrated in two simple examples. In Section 3.5 UEE estimation is applied to three real data sets. In Section 3.4 three simulation studies are presented, based on models taken from the applications in the preceding section. Discussion and conclusions are given in Section 3.6.

3.2 Background and Available Approaches

The generalized estimating equations (GEE) (Liang and Zeger, 1986) is a widely used when the full likelihood function is difficult to specify and the inference of population average is of interest. The GEE estimator is also consistent when the working correlation is misspecified. However, the GEE estimator is biased if the missing mechanism is not missing completely at random.

The weighted generalized estimating equations (WEE) (Robins et al., 1995) and multiple imputation GEE methods e.g. (Paik, 1997) are effective to remove biases for missing at random (MAR), where missingness depends on observed data only. WEE provides consistent estimators even if the working correlation model is misspecified. However, it requires that the model for predicting the probability of missingness is correctly specified. The WEE estimator solves the weighted GEE equations with weights of the inverse of the probability being observed. The WEE estimator is biased if the missing mechanism is not ignorable.

The multiple imputation GEE approach (Paik, 1997, Xie, 1997, Lipsitz, 2000) fills the missing data with imputed values. The number of multiple imputation is discussed by Rubin (1987, p.114). The parameters are estimated by solving the GEE equations with missing response being substituted by the imputed values. Rubin (1987, ch4) and Schafer (1997, p.109) provide valid inferences for multiple imputation approaches. Multiple imputation is able to handle nonignorable missing if the distribution of missing variables can be modeled by observed variables and missingness indicators. However, this approach requires Bayesian technique for computation.

Likelihood approaches for handling missing data are based on a likelihood function by assuming a specific model assumptions. The main technique for computing maximum

likelihood estimators from incomplete data is the expectation and maximization (EM) algorithm (Hartley, 1958, Dempster, Laird, and Rubin, 1977). EM is particularly useful when the maximum likelihood estimation of a complete data model is not difficult to compute. The EM algorithm is able to handle nonignorable missingness. However, it requires specification of full likelihood and it could be extremely slow when the likelihood function is complicated.

3.3 Proposed Method

Let the response variable $\mathbf{y}_i = (y_{i1}, \dots, y_{iT})'$ be a T -dimensional vector with variance-covariance matrix Σ and mean $E(\mathbf{y}_i)$, which satisfies

$$E(\mathbf{y}_i) = \mu_i = g(\mu_i) = \mathbf{x}_i \beta. \quad (3.1)$$

where β is the $P \times 1$ regression parameters of interest and $g(\cdot)$ is a link function.

We consider two types of covariates in the model. One is time-varying covariate, $\mathbf{X}_i^* = (\mathbf{x}_{i1}^*, \dots, \mathbf{x}_{iT}^*)'$, and is a $T \times P$ matrix. The other is time-invariant covariate, $\mathbf{x}_i^{**} = (\mathbf{x}_{i1}^{**}, \dots, \mathbf{x}_{iQ}^{**})'$, which is a $Q \times 1$ vector. We denote

$$\mathbf{x}_i = \begin{pmatrix} \mathbf{x}_{i1}^* \\ \vdots \\ \mathbf{x}_{iT}^* \\ \mathbf{x}_i^{**} \end{pmatrix},$$

which is a $(T \times P + Q) \times 1$ covariate vector. Missing values could occur in \mathbf{y}_i or \mathbf{x}_i , or both.

Covariates \mathbf{x}_i are treated as random variables and the mean and variance are given as follows

$$E(\mathbf{x}_i) = \mu_{\mathbf{x}}, \quad \text{var}(\mathbf{x}_i) = \Sigma_x.$$

Let $\mathbf{U}_i = (\mathbf{y}_i', \mathbf{x}_i')$, \mathbf{U}_i 's are independent identical distributed, which have the following first two moments,

$$E(\mathbf{U}_i) = \mu_u(\psi), \quad \text{Var}(\mathbf{U}_i) = \mathbf{V}(\mu_u, \alpha). \quad (3.2)$$

We denote $\psi = (\beta, \phi, \gamma)$, where β is a vector of regression parameters in (3.1), ϕ is a set of joint parameters of \mathbf{y}_i and \mathbf{x}_i , and γ is a vector of parameters for \mathbf{x}_i . Here μ and \mathbf{V} are known functions of ψ . We assume \mathbf{V} is known. If it is unknown, we can estimate it empirically.

Suppose some of responses Y_i or covariates X_i are missing and the missingness is either MAR or informative missingness. Let \mathbf{R} be a $N \times G$ missingness indicator matrix where G is the number of the variables with missing values. Let $R_{ig} = 1$ if \mathbf{U}_{ig} is observed and $R_{ig} = 0$ if \mathbf{U}_{ig} is missing.

The joint probability of missingness indicators can be expressed as

$$p(R_{i1}, \dots, R_{ig} | \mathbf{u}_i, \tau) = p(R_{ig} | R_{i1}, \dots, R_{ig-1}, \mathbf{u}_i, \tau_g) \\ \times p(R_{ig-1} | R_{i1}, \dots, R_{ig-2}, \mathbf{u}_i, \tau_{g-1}) \times \dots \times p(R_{i1} | \mathbf{u}_i, \tau_1),$$

where $\tau = (\tau'_1, \dots, \tau'_{g-1}, \tau'_g)'$ is a vector of parameters for missingness indicator. The number of parameters can be reduced by using this sequential probability (Lipsitz et al., 2000).

Since the missingness indicators are always binary outcome, they can be modeled as logistic regression. For any missingness indicator, we model it as

$$\mu_{R_{ig} | R_{i1}, \dots, R_{ig-1}, \mathbf{U}_i, \tau} = \frac{\exp(\mathbf{U}_i \tau^u + \sum_{j=1}^{g-1} R_{ij} \tau_j^r)}{1 + \exp(\mathbf{U}_i \tau^u + \sum_{j=1}^{g-1} R_{ij} \tau_j^r)},$$

where τ^u are coefficients for \mathbf{U} and τ^r are coefficients for missingness indicators. Therefore, a joint probability of missingness indicators can be modeled as a sequence of logistic models.

Let $\mathbf{w}_i = (\mathbf{U}_i^o, \mathbf{R}_i \odot \mathbf{U}_i^m, \mathbf{R}_i)'$ and define 0 multiply by a missing-values equals 0, where \mathbf{w}_i is a vector of data, \mathbf{U}_i^o includes variables without any missing values, \mathbf{U}_i^m includes variables with missing values, and \odot is element-wise multiplication. Therefore, the “extended-sense” outcome \mathbf{w}_i does not include any missing values. All missing values are replaced by zeros, which incorporates the effect of missingness indicators.

We have \mathbf{w}_i as i.i.d, with first two moments as follows

$$\mathbb{E}(\mathbf{w}_i) = \mu_w(\psi_w, \phi), \quad \text{var}(\mathbf{w}_i | S) = V(\mu_w, \phi), \quad (3.3)$$

We denote $\psi = (\beta, \phi, \gamma, \alpha)$, where β is a vector of regression parameters in (3.1), γ is a vector of parameters for \mathbf{x}_i , α is a vector of parameters for missingness indicator, and ϕ is a set of joint parameters of \mathbf{y}_i , \mathbf{x}_i , and \mathbf{R}_i . Here μ_w and \mathbf{V} are known functions of ψ . We assume \mathbf{V} is known. If it is unknown, we can estimate it empirically.

The main idea of this approach is to construct unconditional estimating equations for \mathbf{w} . The quasilielihood equation with respect to ψ_w is

$$\sum_{i=1}^N D'_w V_w^{-1} (w_i - \mu_w) = D'_w V_w^{-1} \sum_{i=1}^N (w_i - \mu_w) = 0, \quad (3.4)$$

where $\mathbf{D}_w = \partial \mu_w / \partial \psi'$. Since μ_w does not depend on a specific subject index i , \mathbf{D} also does not depend on index i . So equation (3.4) can be written as

$$\mathbf{D}'_w \mathbf{V}^{-1} (\bar{\mathbf{w}} - \mu_w) = 0, \quad (3.5)$$

where $\bar{\mathbf{u}} = \frac{\sum_{i=1}^N \mathbf{w}_i}{N}$ is the mean of \mathbf{w}_i .

We obtain $\hat{\psi}$ by solving estimating equations (3.5). If we take expectation of equation (3.5),

$$\mathbf{E}(\mathbf{D}'_w \mathbf{V}^{-1} (\bar{\mathbf{w}} - \mu_w)) = \mathbf{D}'_w \mathbf{V}^{-1} (\mathbf{E}(\bar{\mathbf{w}}) - \mu_w) = 0, \quad (3.6)$$

Therefore, the estimating equation (3.5) is an unbiased estimating equation. We have the following theorem for consistency and asymptotical normality of $\hat{\psi}_w$.

Theorem 3.1. *The estimator $\hat{\psi}_w$ is \sqrt{N} -consistent, $N^{1/2}(\hat{\psi}_w - \psi_w)$ is asymptotically multivariate normal with mean zero, with the asymptotic covariance matrix as*

$$(\mathbf{D}'_w V_w^{-1} \mathbf{D}_w)^{-1}.$$

The matrix $(\mathbf{D}'_w V_w^{-1} \mathbf{D}_w)^{-1}$ is $k \times k$ where k is the number of parameters that need to be estimated. For this matrix to be invertible, we need the rank of \mathbf{D} to be at least k , and in particular, we need the number of rows of \mathbf{D} to be at least k . The number of rows of \mathbf{D} is equal to the number of variables in \mathbf{w}_i . Above we defined $\mathbf{w}_i = (\mathbf{U}_i^o, \mathbf{R}_i \odot \mathbf{U}_i^m, \mathbf{R}_i)'$, but if the number of single variables $(\mathbf{U}^o, \mathbf{R} \odot \mathbf{U}^m, \mathbf{R})$ are not enough, we can include some products of \mathbf{U}^o , $\mathbf{R} \odot \mathbf{U}^m$, and \mathbf{R} .

In order to apply GEE for \mathbf{w}_i , we need to implement the expectation of \mathbf{w}_i . For example, considering missingness indicator R_{ig} , the expectation of R_{ig} is

$$\sum_{R_{ig}} R_{ig} p(R_{ig}) = p(R_{ig}) = p(R_{ig} = 1 | R_{i1}, \dots, R_{ig-1}, w_i) p(R_{i1}, \dots, R_{ig-1}, w_i). \quad (3.7)$$

Equation (3.7) can be written as

$$\int_{w_i} \sum_{R_{i1}, \dots, R_{ig-1}} p(R_{ig} = 1 | R_{i1}, \dots, R_{ig-1}, w_i, \alpha) \times p(R_{i1}, \dots, R_{ig-1} | w_i, \alpha) p(w_i).$$

To evaluate the integral above, the full specification of the distribution of $(R_{i1}, \dots, R_{ig}, w_i)$ is required. To avoid specifying the joint distribution of $(R_{i1}, \dots, R_{ig}, w_i)$, we use the multivariate normal approximation (Whittle, 1961, Crowder, 1985, Lipsitz et al., 2000) to evaluate the joint distribution. For the multinormal distribution, the conditional distribution of \mathbf{W}_1 given $\mathbf{W}_2 = \mathbf{w}_2$ is multivariate normal with mean

$$\mu_1 + V_{12} V_2^{-1} (w_2 - \mu_2),$$

and the covariance

$$V_1 - V_{12} V_2^{-1} V_{12}'.$$

For binary variables, the conditional distribution of one variable $W_1 = 1$ given $\mathbf{W}_2 = \mathbf{w}_2$ is

$$p(W_1 = 1 | \mathbf{W}_2 = \mathbf{w}_2) = E(W_1 | \mathbf{W}_2 = \mathbf{w}_2). \quad (3.8)$$

By Taylor expansion, equation (3.8) becomes

$$p(W_1 = 1 | \mathbf{W}_2 = \mathbf{w}_2) \approx \mu_1 + V_{12} V_2^{-1} (\mathbf{w}_2 - \mu_2).$$

However, this proposed approach requires to model missingness indicator correctly. If the model of missingness is misspecified, equation (3.4) may be biased, also the proposed approach would not be as efficient as the model of the missingness is correctly specified.

For informative missingness, the choice of variables in the model of missingness indicator is important. If the model for the missingness indicator can not be determined with certainty, sensitivity analysis is needed to decide the model for R_i . Sensitivity analysis,

which fit a range of different nonignorable missing patterns by changing a subset of parameters. Discussions about sensitivity analysis can be found in Rubin (1977), Zhu and Lee (2001), Verbeke et al. (2001), Molenberghs et al. (2001), and Kenward (1998).

3.3.1 Examples

In this section, we provide two simple examples to illustrate the proposed method. The first example shows how to handle missing covariates under ignorable missingness for continuous variables. The second example shows how to handle the missing response under nonignorable missingness for binary response variables. Although these examples show how to handle missing responses or covariates, the proposed approach is able to handle both missing response and covariates.

Example 1 Consider a generalized linear model

$$y_i = \beta x_i + \epsilon_i, \quad (3.9)$$

where ϵ_i has a normal distribution with mean 0 and variance σ^2 , and x_i has $N(\mu_x, \sigma_x^2)$.

Suppose y_i is observed but x_i has some missing values. The missingness indicator $r_i = 1$ for x_i being observed and $r_i = 0$ for x_i being missing. The true probability of x_i being observed is

$$Pr(r_i = 1|y_i, x_i) = \frac{\exp(\alpha y_i)}{1 + \exp(\alpha y_i)}. \quad (3.10)$$

We model the probability of x_i being observed as the true given in (3.10). There are three parameters β , α , and μ_x that need to be estimated. Let $W_i = (r_i, y_i, r_i * x_i)$. We solve following equation to obtain estimators for the three parameters.

$$D'_w V_w^{-1} \sum_{i=1}^N (W_i - \mu_w) = 0, \quad (3.11)$$

The unconditional mean of y_i is $E(y) = \mu_y = \beta \mu_x$, and let μ_r be the mean of r_i and μ_{rx} be the mean of $r_i * x_i$, the unconditional mean of W_i can be expressed as

$$\mu_w = E(W_i) = \begin{pmatrix} E(r) \\ E(y) \\ E(rx) \end{pmatrix} = \begin{pmatrix} \mu_r \\ \mu_y \\ \mu_{rx} \end{pmatrix} = \begin{pmatrix} \int_y \frac{e^{\alpha y}}{1+e^{\alpha y}} f(y) dy \\ \beta \mu_x \\ \int_x \int_y \frac{e^{\alpha y}}{1+e^{\alpha y}} x f(y|x) f(x) dx dy \end{pmatrix},$$

where $f(y)$ is the density of y , $f(y|x)$ is the density of y given x , and $f(x)$ is the density of x , which are given by

$$\begin{aligned} f(y) &= \frac{\exp(-(y - \mu_y)^2 / (2\sigma^2))}{\sqrt{2\pi\sigma^2}} \\ f(y|x) &= \frac{\exp(-(y - \beta x)^2 / (2\sigma^2))}{\sqrt{2\pi\sigma^2}} \\ f(x) &= \frac{\exp(-(x - \mu_x)^2 / (2\sigma_x^2))}{\sqrt{2\pi\sigma_x^2}}. \end{aligned}$$

Let $A_1 = Pr(r = 1|y) = \frac{e^{\alpha y}}{1 + e^{\alpha y}}$, $A_2 = \frac{\partial Pr(r=1|y)}{\partial \alpha} = \frac{ye^{\alpha y}}{(1 + e^{\alpha y})^2}$, D can be written as

$$\begin{aligned} D' &= \begin{pmatrix} \frac{\partial \mu_r}{\partial \beta} & \frac{\partial \mu_y}{\partial \beta} & \frac{\partial \mu_{rx}}{\partial \beta} \\ \frac{\partial \mu_r}{\partial \alpha} & \frac{\partial \mu_y}{\partial \alpha} & \frac{\partial \mu_{rx}}{\partial \alpha} \\ \frac{\partial \mu_r}{\partial \mu_x} & \frac{\partial \mu_y}{\partial \mu_x} & \frac{\partial \mu_{rx}}{\partial \mu_x} \end{pmatrix} \\ &= \begin{pmatrix} \int_y A_1 \frac{\partial f(y)}{\partial \beta} dy & \mu_x & \int_x \int_y A_1 x f(x) \frac{\partial f(y|x)}{\partial \beta} dx dy \\ \int_y A_2 f(y) dy & 0 & \int_x \int_y A_2 x f(y|x) f(x) dx dy \\ \int_y A_1 \frac{\partial f(y)}{\partial \mu_x} dy & \beta & \int_x \int_y A_1 x f(y|x) \frac{\partial f(x)}{\partial \mu_x} dx dy \end{pmatrix}, \end{aligned}$$

Since D' does not have an explicit form, the numerical integration is needed. In addition, V matrix in equation (3.11) can be estimated empirically. we solve $\hat{\beta}$, $\hat{\alpha}$, and $\hat{\mu}_x$ from equation (3.11).

Example 2 Suppose that each subject is measured at time $t = 1, 2$. The response variables are y_{i1} and y_{i2} , and the covariate x_i is binary. We assume that x_i has a Bernoulli distribution with mean μ_x . Considered y_{it} follows the logistic regression model

$$\text{logit } y_{it} = \beta_t(t - 1) + \beta x_i. \quad (3.12)$$

Based on the regression model, the unconditional mean of y_{it} is

$$\mu_t = E(y_t) = \frac{e^{\beta_t(t-1)+\beta}}{1 + e^{\beta_t(t-1)+\beta}} \mu_x + \frac{e^{\beta_t(t-1)}}{1 + e^{\beta_t(t-1)}} (1 - \mu_x).$$

We assume that both y_{i1} , x_i are observed but y_{i2} has some missing values. The missingness indicator for y_{i2} is r_i . The missingness indicator is

$$\text{logit } r_i = \alpha_0 + \alpha_1 y_{i2}. \quad (3.13)$$

We model the missingness indicator following the true missingness indicator. There are five parameters $(\beta_t, \beta, \alpha_0, \alpha_1, \mu_x)$ that need to be estimated. If \mathbf{w}_i only includes single variables y_{i1} , $r_i * y_{i2}$, x_i , and r_i , then there are only four independent equations for five parameters. Therefore, we need to add at least one cross product from y_{i1} , $r_i * y_{i2}$, x_i , and r_i to obtain enough equations to solve the parameters. Let $W_i = (r_i, y_{i1}, x_i * y_{i1}, r_i * y_{i2}, x_i)$, we solve the following equation to obtain estimators for the parameters.

$$D'_w V_w^{-1} \sum_{i=1}^N (W_i - \mu_w) = 0, \quad (3.14)$$

The expectation of W_i is given as follows

$$\mu_w = E(W_i) = \begin{pmatrix} \mu_r \\ \mu_{y_1} \\ \mu_{xy_1} \\ \mu_{ry_2} \\ \mu_x \end{pmatrix} = \begin{pmatrix} \frac{e^{\alpha_0 + \alpha_1}}{1 + e^{\alpha_0 + \alpha_1}} \mu_2 + \frac{e^{\alpha_0}}{1 + e^{\alpha_0}} (1 - \mu_2) \\ \mu_1 \\ \frac{e^\beta}{1 + e^\beta} \mu_x \\ \frac{e^{\alpha_0 + \alpha_1}}{1 + e^{\alpha_0 + \alpha_1}} \mu_2 \\ \mu_x \end{pmatrix}.$$

Let

$$\begin{aligned} A_1 &= \frac{e^{\alpha_0 + \alpha_1}}{1 + e^{\alpha_0 + \alpha_1}} \\ A_2 &= \frac{e^{\alpha_0}}{1 + e^{\alpha_0}} \\ B_1 &= \frac{\partial A_1}{\partial \alpha_0} = \frac{\partial A_1}{\partial \alpha_1} = \frac{e^{\alpha_0 + \alpha_1}}{(1 + e^{\alpha_0 + \alpha_1})^2} \\ B_2 &= \frac{\partial A_2}{\partial \alpha_0} = \frac{e^{\alpha_0}}{(1 + e^{\alpha_0})^2} \\ \dot{\mu}_{1\beta} &= \frac{e^{\beta_t + \beta}}{(1 + e^{\beta_t + \beta})^2} \\ \dot{\mu}_{2\beta} &= \frac{e^{\beta_t}}{(1 + e^{\beta_t})^2} \end{aligned}$$

D' can be written as

$$D' = \begin{pmatrix} \frac{\partial \mu_r}{\partial \beta_t} & \frac{\partial \mu_{y1}}{\partial \beta_t} & \frac{\partial \mu_{xy1}}{\partial \beta_t} & \frac{\partial \mu_{ry2}}{\partial \beta_t} & \frac{\partial \mu_x}{\partial \beta_t} \\ \frac{\partial \mu_r}{\partial \beta_x} & \frac{\partial \mu_{y1}}{\partial \beta_x} & \frac{\partial \mu_{xy1}}{\partial \beta_x} & \frac{\partial \mu_{ry2}}{\partial \beta_x} & \frac{\partial \mu_x}{\partial \beta_x} \\ \frac{\partial \mu_r}{\partial \alpha_0} & \frac{\partial \mu_{y1}}{\partial \alpha_0} & \frac{\partial \mu_{xy1}}{\partial \alpha_0} & \frac{\partial \mu_{ry2}}{\partial \alpha_0} & \frac{\partial \mu_x}{\partial \alpha_0} \\ \frac{\partial \mu_r}{\partial \alpha_1} & \frac{\partial \mu_{y1}}{\partial \alpha_1} & \frac{\partial \mu_{xy1}}{\partial \alpha_1} & \frac{\partial \mu_{ry2}}{\partial \alpha_1} & \frac{\partial \mu_x}{\partial \alpha_1} \\ \frac{\partial \mu_r}{\partial \mu_x} & \frac{\partial \mu_{y1}}{\partial \mu_x} & \frac{\partial \mu_{xy1}}{\partial \mu_x} & \frac{\partial \mu_{ry2}}{\partial \mu_x} & \frac{\partial \mu_x}{\partial \mu_x} \end{pmatrix}$$

$$= \begin{pmatrix} (A_1 - A_2)\dot{\mu}_{1\beta}\mu_x + \dot{\mu}_{2\beta}(1 - \mu_x) & 0 & 0 & A_1\dot{\mu}_{1\beta}\mu_x + \dot{\mu}_{2\beta}(1 - \mu_x) & 0 \\ (A_1 - A_2)\dot{\mu}_{1\beta}\mu_x & \frac{e^\beta}{(1+e^\beta)^2} & \frac{e^\beta}{(1+e^\beta)^2} & A_1\dot{\mu}_{1\beta}\mu_x & 0 \\ B_1\mu_2 + B_2(1 - \mu_2) & 0 & 0 & B_1\mu_2 & 0 \\ B_1\mu_2 & 0 & 0 & B_1\mu_2 & 0 \\ (A_1 - A_2)(\frac{e^{\beta_t+\beta}}{1+e^{\beta_t+\beta}} - \frac{e^{\beta_t}}{1+e^{\beta_t}}) & \frac{e^\beta}{1+e^\beta} - \frac{1}{2} & \frac{e^\beta}{1+e^\beta} & A_1(\frac{e^{\beta_t+\beta}}{1+e^{\beta_t+\beta}} - \frac{e^{\beta_t}}{1+e^{\beta_t}}) & 1 \end{pmatrix}.$$

All estimates of parameters are obtained by solving equation (3.14).

3.4 Application

In this section, we apply the proposed approach for existing data sets with missing values. The first data studies translaryngeal intubation. It is a special case of longitudinal model-GLM with some missing covariate values. The second data studies contracepting women. It has monotone missing responses. The third data set studies hip fracture women. It contains both missing response variables and covariates. We compare the proposed method with multiple imputation method (Paik, 1997, Lipsitz et al., 2000), the weighted approach (Robins et al., 1995), and maximum likelihood approach for these data sets.

3.4.1 TLI Data

The data is based on a translaryngeal intubation study (Ibrahim, 1990 and Colice, Stukel, and Dain, 1989). The data set involves a study of 82 patients who experienced translaryngeal intubation (TLI) for more than four days and are prospectively evaluated for laryngeal complications. The study is designed to identify a group of patients experiencing prolonged TLI (more than four days) and to prospectively evaluate the incidence and

the type of laryngeal complications they might suffer. The data are collected from the patients during the period of TLI. Three covariates (serum albumin x_1 , Serum creatinine x_2 , and the ratio of laryngeal size to tracheal tube size x_3) are incomplete. The response variable is dichotomized to 0 or 1, with 0 representing no damage of the larynx at baseline, and 1 otherwise. All covariates are dichotomized too. More details of this data is given by Ibrahim (1990).

The marginal regression model for this data is

$$\text{logit } pr(y_i = 1 | x_{1i}, x_{2i}, x_{3i}) = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_3 x_{3i}. \quad (3.15)$$

For the proposed approach, we treat response variable and covariates as random variables. The “extended-sense” outcomes include y , x and cross products of y and x , namely,

$$[y, R_1 x_1, R_2 x_2, R_3 x_3, R_1 y x_1, R_2 y x_2, R_3 y x_3], \quad (3.16)$$

where R_1 , R_2 , and R_3 are missingness indicators for x_1 , x_2 , and x_3 and $R_{ki} \times x_{ki}$ is x_{ki} if $R_{ki} = 1$ and 0 if $R_{ki} = 0$, where $k = 1, 2, 3$. The missingness indicators are modeled as

$$Pr(R_{ki} = 1 | y_i, x_{1i}, x_{2i}, x_{3i}) = \frac{e^{g_i' \alpha}}{1 + e^{g_i' \alpha}},$$

where g_i is a combination of the elements of $(y_i, x_{1i}, x_{2i}, x_{3i})$. Because the missingness indicator for TLI data is assumed as MAR, we only use variable y in g .

We compare the proposed method with multiple imputation methods (Paik, 1997, Lipsitz et al., 2000) and the maximum likelihood approach (Ibrahim, 1990). For multiple imputation, conditional models are used to impute the missing outcome data. We impute missing values in x_1 , x_2 , x_3 by drawing a random sample from a Bernoulli distribution with probability of success

$$Pr(x_{ki} = 1 | y_i, x_{k-1i}) = \frac{e^{m_i' \gamma}}{1 + e^{m_i' \gamma}}, \quad k = 1, 2, 3$$

where m_i is some combination of the elements of (y_i, x_{obs}) . In particular, $m_1 = (y)$, $m_2 = (x_1, y)$, and $m_3 = (x_1, x_2, y)$. The imputed data along with observed data are analyzed by GLM. Five imputations are generated.

For the maximum likelihood approach, it is followed by Ibrahim (1990). Since all covariates are binary variables, we specify a joint discrete distribution on $X' = (x_1, x_2, x_3)$ as

$$f(\mathbf{x}_i|\gamma) = \prod_{j=1}^8 \gamma_j^{I(X_i)_j},$$

where $I(X_i)$ is an indicator function. For each missing X_i , we fill missing values by the possible values (0 and 1) multiplying by the corresponding probability $Pr(X_i^m|X_i^o, y_i)$. Then obtain parameters by solving weighted generalized linear model.

The results are given in table 9. Table 9 indicates that all estimations are not significantly different from 0. All three methods yield similar estimates and variances.

3.4.2 Contracepting Women Data

This data is a longitudinal randomized clinical trial studying contracepting women (Machin et al, 1988, Fitzmauris, 2000). In this trial 1151 women receive an injection of either 100 or 150 mg of depot-medroxyprogesterone acetate (DMPA) at the beginning of the study and receive three additional injections at 90-day intervals.

Throughout the study each woman is required to complete a menstrual diary. The diary data is used to generate a sequence of binary responses for each subject according to whether or not she has experienced amenorrhea in the four successive three-month intervals. There is monotone dropout missingness. More than 30% of subjects dropped out before the completion of the trial.

The regression model (Fitzmauris, 2000) is formulated as

$$\text{logit } pr(y_{it} = 1|x_i) = \beta_0 + \beta_1(t-1) + \beta_2(t-1)^2 + \beta_3x_i + \beta_4x_i \times (t-1) + \beta_5x_i \times (t-1)^2, \quad (3.17)$$

where $t = 0, 1, 2, 3$ and x_i is the treatment. We denote $x_i = 0$ for DMPA dose 100mg and $x_i = 1$ for DMPA dose 150mg.

For the proposed approach, we treat response variables, missingness indicators, and covariates as random variables. Since the missingness indicator is MAR and missing pattern is monotone, the missingness indicators are modeled as

$$Pr(R_{it} = 1|y_{i1}, \dots, y_{i4}, x_i) = \frac{e^{g'_i\alpha}}{1 + e^{g'_i\alpha}},$$

where the missingness indicator $R_{it} = 1$ if y_{it+1} is observed and $R_{it} = 0$ if y_{it+1} is missing and g_i is a combination of the elements of $(y_{i1}, \dots, y_{i4}, x_i)$. In this example, g_1 includes y_1 , g_2 includes y_2 , and g_3 includes y_3 based on data. There are 13 parameters (six β 's, μ_x , six α 's) need to be estimated. The “extended-sense” outcomes include all response variables, covariate, missingness indicators, and cross products of y and x , R , namely,

$$[y_1, xy_1, R_1y_2, R_1xy_2, R_2y_3, R_2xy_3, R_4y_4, R_3xy_4, x, R_1, R_2, R_3, R_1y_1, R_2y_2, R_3y_3]. \quad (3.18)$$

We apply unconditional GEE to \mathbf{w} to solve all parameters.

We compare our approaches with the weighted, multiple imputation and our proposed approach in chapter 2. For our proposed approach in chapter 2, we use all response variables, covariates and the cross products of y and x in data matrix,

$$\mathbf{U} = [y_1, y_2, y_3, y_4, xy_1, xy_2, xy_3, xy_4, x].$$

We briefly describe the weighted method here. In the weighted approach, the probability of an observation being observed is modeled as

$$Pr(R_{it} = 1|y_1, \dots, y_{it}, x_i) = \frac{e^{g'_i\gamma}}{1 + e^{g'_i\gamma}},$$

where g_i is a combination of the elements of $(y_{i1}, \dots, y_{it}, x_i)$. Specifically, for this application, we modeled the probability of y_{it} being observed condition on previous responses and covariate x_i as

$$\lambda_{it} = Pr(R_{it} = 1|y_{i1}, \dots, y_{it-1}, x_i) = \frac{e^{\gamma_0 + \gamma_1 y_{i1} + \dots + \gamma_{t-1} y_{it-1} + \gamma x_i}}{1 + e^{\gamma_0 + \gamma_1 y_{i1} + \dots + \gamma_{t-1} y_{it-1} + \gamma x_i}}.$$

The weights W_i in (2.3) are calculated by taking the inverse of the above estimated probability of y_{it} being observed. The covariance matrix of $\hat{\beta}$ of the weighted GEE (Robins et al., 1995) is given by

$$\Sigma_{\beta} = \left(\sum_{i=1}^K \frac{\partial \mu'_i}{\partial \beta} V_i^{-1} W_i \frac{\partial \mu_i}{\partial \beta} \right)^{-1} \sum_{i=1}^K E_i E'_i \left(\sum_{i=1}^K \frac{\partial \mu'_i}{\partial \beta} V_i^{-1} W_i \frac{\partial \mu_i}{\partial \beta} \right)^{-1},$$

where $E_i = U_i - \left(\sum_{i=1}^K U_i S'_i \right) \left(\sum_{i=1}^K S_i S'_i \right)^{-1} S_i$, $U_i = \frac{\partial \mu'_i}{\partial \beta} V_i^{-1} (Y_i - \mu_i(\hat{\beta}))$, and $S_i = \sum_t R_{it-1} Z_{it} (R_{it} - \lambda_{it})$ is the score component for the i^{th} subject from the missingness indicator model (Robins et al., 1995, Preisser et al., 2002).

For the multiple imputation, we replace the missing values sequentially based on the imputation using logistic model:

$$Pr(y_{it} = 1 | y_1, \dots, y_{it}, x_i) = \frac{e^{\alpha_0 + \alpha_1 y_{i1} + \dots + \alpha_{t-1} y_{it-1} + \alpha x_i}}{1 + e^{\alpha_0 + \alpha_1 y_{i1} + \dots + \alpha_{t-1} y_{it-1} + \alpha x_i}} .$$

The estimator of multiple imputation is the average of estimators from each imputation. The covariance of estimator (Rubin, 1978) is given by

$$\Sigma(\beta) = \bar{U} + (1 + \frac{1}{m})B,$$

where m is the number of the multiple imputations, and $\bar{U} = \frac{1}{m} \sum_{j=1}^m U_j$ is the average of within-imputation variance, where U_j is the variance associated with β_j with

$$U_j = \left(\sum_{i=1}^K \frac{\partial \mu'_i}{\partial \beta} V_i^{-1} \frac{\partial \mu_i}{\partial \beta} \right)^{-1} \sum_{i=1}^K \frac{\partial \mu'_i}{\partial \beta} V_i^{-1} (Y_i - \mu_i(\hat{\beta})) (Y_i - \mu_i(\hat{\beta}))' V_i^{-1} \frac{\partial \mu_i}{\partial \beta} \left(\sum_{i=1}^K \frac{\partial \mu'_i}{\partial \beta} V_i^{-1} \frac{\partial \mu_i}{\partial \beta} \right)^{-1} .$$

Here $B = \frac{1}{m-1} \sum_{j=1}^m (\beta_j - \bar{\beta})^2$ is the variance between imputations.

The results are given in table 10. The proposed approach I and II and the weighted method give similar standard errors for most of the regressors. The only exception is the effect of *time*², the proposed approach II shows that the effect is significant but the other three approaches show no significance. Multiple imputation provides similar estimations to the other three approaches but has highest standard errors among all approaches.

3.4.3 Missing Covariate

This example investigates (Allison, 2001) 220 caucasian women, at least 60 years old, who are treated surgically for hip fractures in Philadelphia area (Mossey, Knott, and Craik, 1990). They are interviewed three times after the hospital release: at 2 months, 6 months, and 12 months. The outcome variable is a measurement of depression, on a scale from 0 to 60. The four related covariates are the number of self-care that could be completed without assistance (x_1 , ranges from 0 to 3), the degree of pain experienced by the patient (x_2 , ranges from 0 (none) to 6 (constant)), self-rated health (x_3 , measured on a four-point scale, 1 = poor, 4 = excellent), and whether able to walk without aid at home (x_4 , 0 = no and 1 = yes).

We use the subset of the second data example to create missing covariates. We select 138 subjects with complete responses. We consider a relatively simpler model and is given by

$$y_{it} = \beta_1 x_{it2} + \beta_2 w_1 + \beta_3 w_2 + \epsilon_{it} .$$

We make the missing patterns of x_2 as monotone. There are about 24% of the covariate missing. We compare the proposed approach with the multiple imputation (Paik, 1997, Lipsitz et al., 2000) and the weighted GEE (Robins et al., 1994). For the multiple imputation approach, we imputed missing values for $t = 1$ first then for $t = 2$ and 3. We apply the GEE based on imputed values and observed data.

For the proposed approach, since the missingness indicator is MAR and missing pattern is monotone, the missingness indicators are modeled as

$$Pr(R_{it} = 1 | y_{i1}, \dots, y_{i3}, x_{i1}, \dots, x_{i3}) = \frac{e^{g_i' \alpha}}{1 + e^{g_i' \alpha}},$$

where the missingness indicator $R_{it} = 1$ if x_{it} is observed and $R_{it} = 0$ if x_{it} is missing and g_i is a combination of the elements of $(y_{i1}, \dots, y_{i3}, x_{i1}, \dots, x_{i3})$. In this example, g_1 includes y_1 , g_2 includes y_2 , and g_3 includes y_3 based on data. There are 12 parameters (three β 's, three μ_x , six α 's) need to be estimated. The “extended-sense” outcomes include all response variables, covariate, missingness indicators, and cross products of y and x , R , namely,

$$[y_1, R_1 x_1, R_1 x_1 y_1, y_2, R_2 x_2, R_2 x_2 y_2, y_3, R_3 x_3, R_3 x_3 y_3, R_1, R_2, R_3, R_1 y_1, R_2 y_2, R_3 y_3]. \quad (3.19)$$

We apply unconditional GEE to \mathbf{w} to solve all parameters.

For the weighted method, the probability of being observed is modeled as

$$Pr(R_{it} = 1 | y_1, \dots, y_{it}, x_{i1}, \dots, x_{it}) = \frac{e^{g_i' \gamma}}{1 + e^{g_i' \gamma}},$$

where g_i is a combination of elements in $(y_{i1}, \dots, y_{it-1}, x_{i2})$.

For our proposed approach in chapter 2, we use

$$\mathbf{U} = [y_1, y_2, y_3, x_1, x_2, x_3, x_1 y_1, x_2 y_2, x_3 y_3].$$

The results are given in table 11. The results show that all four approaches give similar estimations. Two proposed approaches provide slight larger standard errors.

3.5 Simulations

We conduct several simulations to study the finite sample performance of the proposed approach for ignorable and informative missingness. Parameters in the simulations are based on the example data.

3.5.1 Simulation Based on TLI Data

Simulations are conducted based on the TLI data (Ibrahim, 1990). There is one response variable and three covariates. The missing values occur on the covariate. The response variable is fully observed. The covariates are generated as correlated binary variables based on the sample means and covariance of observed x_1 , x_2 , and x_3 . The response variable is generated with the logistic model

$$\text{logit } pr(y_i = 1|x_{1i}, x_{2i}, x_{3i}) = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_3 x_{3i}. \quad (3.20)$$

All β 's in (3.20) are estimated from data by using GLM based on complete data.

The true missingness indicator is specified such that the missingness depends on the observed outcome data. This ensures that the missing is MAR. The missingness indicator R_{ki} is 1 if x_{ki} is observed and 0 if x_{ki} is missing, $k = 1, 2, 3$. The probability of x_{ki} being observed is given as

$$\text{logit } pr(R_{ki} = 1|y_i, x_{ki}) = \alpha_0 + \alpha_1 y_i. \quad (3.21)$$

All α 's are modified based on data to increase the missing rate. The missing rate is about 32 %. The simulation uses same sample size 82.

The second simulation based on this data modifies β_3 and increases sample size to 200. Also α are changed to obtain different missing rate.

We compare our proposed approach with multiple imputation (Paik, 1997) and maximum likelihood approach (Ibrahim, 1990) based on the simulation size of 1000. The biases and standard errors for the first simulation are given in Table 12 and for the second are given in 13.

Table 12 indicates that the proposed approach and the maximum likelihood approach yield similar biases and the biases are smaller than the multiple imputation estimates.

The standard errors of the proposed approach and the maximum likelihood approach are similar but they are larger than those of the multiple imputation estimates.

Table 13 indicates that the proposed approach and the multiple imputation have similar biases and standard errors, although the maximum likelihood approach yields smaller biases and standard errors for the intercept and x_1 , it has larger biases and standard errors on x_2 and x_3 . When missing rate increases, the biases increase for all three approaches.

3.5.2 Simulation Based on Women Data

We simulate data based on the women contracepting data with the sample size as 250. Both response variables and covariate are binary variables. The response variables given the covariate x follow a logistic model:

$$\text{logit } pr(y_{it} = 1) = \beta_0 + \beta_t(t - 1) + \beta_x x_i \quad t = 1, \dots, 4.$$

Correlated response variables are generated following the procedure by Preisser (2002) and the covariate is generated by Bernoulli distribution with the mean as the sample mean of x from the women contracepting data. The missing values occur in responses only. The true missingness indicator is specified such that the missingness is monotone and depends on the previous observed outcome. This ensured that the missing is MAR. The missingness indicator R_{it} is 1 if y_{it} is observed and 0 if y_{it} is missing, where $t = 2, 3, 4$. There is no missing in y_1 . The true missingness indicator follows logistic model and only depends on the previous response y .

$$\text{logit } pr(R_{it} = 1 | y_{i1}, \dots, y_{it}, x_i) = \alpha_0 + \alpha_1 y_{i(t-1)}. \quad (3.22)$$

We choose α such that the missing rate is about 27% and also with another missing rate as 53%.

The bias and standard error of the estimators are provided in Table 14. When the missingness indicators are correctly specified, the proposed approach has smallest biases for most of the estimations and has similar standard error for most of the estimators to multiple imputation. The standard errors of these two approaches are smaller than those

of the weighted GEE estimators (with correctly specified missingness indicator) and our first proposed approach.

3.5.3 Simulation Based on Missing covariates Data

We simulate data based on hip fracture data with missing covariates. The means of response (y_1, y_2, y_3) given covariate x_{it2} are

$$E(y_{it}) = \beta_1 x_{it2} + \beta_2 w_1 + \beta_3 w_2 , \quad (3.23)$$

and the covariance is Σ_y . Three covariates are generated by multinormal distribution with mean μ_{x_2} and covariance Σ_{x_2} . They are truncated according to the range of x_2 .

The means and covariances of y and covariates are based on the sample mean and variance of the data. The missingness only occurs in the covariate. The true missingness indicator is specified such that the missingness is informative.

$$Pr(R_{it} = 1) = \alpha_0 + \alpha_1 y_{it} + \alpha_2 x_{it} .$$

The parameters α_0 and α_1 are based on the estimates from the data, but α_2 is artificial. We set $\alpha_2 = 0.5$. We compare our proposed approach with maximum likelihood approach (EM approach). For both approach , we model the missingness indicator the same as true missingness indicator. We also show results for weighted method and multiple imputation for reference.

The biases and standard errors are provided in Table 15. The results are based on 500 simulations with sample size 200. The table shows that our proposed approach and maximum likelihood approach having similar biases and standard errors for covariate effects, but the proposed approach has slightly smaller biases and standard errors than maximum likelihood approach for time effects. Since the missingness indicator is informative, the weighted approach and multiple imputation do not give consistent estimations. Table 15 show that multiple imputation gives more biases and larger standard errors than weighted method, while weighted method gives slightly larger biases and standard errors than our proposed approach and the maximum likelihood approach.

3.6 Discussion and conclusion

The proposed approach is based on unconditional expectation of all possible variables and an approximation to the multivariate normal distribution. This approach avoids complete specifying the full joint distribution of all variables. Although it requires some assumptions about the distributions of covariates and associations among them, in general they are easy to satisfy.

Similar to the weighted method, our approach requires model assumptions for missingness indicator, but it can easily handle nonignorable missingness without full specification of the joint distribution of covariates, responses, and missingness indicator. In addition, our proposed approach can easily extend to arbitrary missing pattern while the weighted method and multiple imputation cannot.

The simulation studies suggest that the proposed approach has negligible bias. The simulation results also indicate that the bias and the variances of the proposed estimators are comparable to the multiple imputation, weighted method, or maximum likelihood. But since there are many possible missing data configurations, it is difficult to draw definitive conclusions from limited simulations.

There are some drawbacks of the proposed approach. In particular, it does not perform well for the models when matrix D is close to singular, because there can be convergence problem. But this problem could be solved by including cross products of variables in response variables, covariates, and missingness indicators.

Reference

1. Allison, P. (2001) Missing data, Thousand Oaks, CA, Sage.
2. Barnard, J. and Meng, X.-L. (1999) Applications of multiple imputation in medical studies: from AIDS to NHANES. *Statistical Methods in Medical Research*, 8, 7-36.
3. Bernaards, C. and Sijtsma, K. (2000) Influence of Imputation and EM Methods on Factor Analysis When Item Nonresponse in Questionnaires is Nonignorable. *Multivariate Behavioral Research*. 35 (3) 321-364.
4. Colice, G., Stukel, T., and Dain, B. (1989), Laryngeal Complications of Prolonged Intubation, *Chest*, 96, 877-884.
5. Diggle P. and Kenward M. (1994) Informative dropout in longitudinal data analysis (with discussion), *Applied Statistics*, 43, 49-93.
6. Dempster, A., Laird, M., and Rubin, D. (1977) Maximum likelihood from incomplete data via the EM algorithm. *J. R. Statist. Soc. B*, 39, 185-197.
7. Green, P. and Park, T. (2003) A Bayesian Hierarchical Model for Categorical Data for Nonignorable Nonresponse, *Biometrics*, 59, 886-896.
8. Hartley, H. (1958) Maximum Likelihood Estimation from Incomplete Data *Biometrics*, 14(2), 174-194.
9. Horvitz, D. and Thompson, D. (1952) A Generalization of Sampling without replacement from a Finite Universe, *Journal of the American Statistical Association*, 47, 663-85.
10. Huang, L., Chen, M., and Ibrahim, J. (2005) Bayesian analysis for generalized linear models with nonignorably missing covariates, *Biometrics*, 61, 767-78.
11. Ibrahim, J. (1990) Incomplete Data in Generalized Linear Models, *Journal of the American Statistical Association*, 85, 765-69.
12. Ibrahim, J., Lipsitz, S., and Chen, M. (1999) Missing covariates in generalized linear models when the missing data mechanism is nonignorable, *Journal of the Royal Statistical Society, Series B*, 61, 173-190.
13. Ibrahim, J., Chen, M., Lipsitz, S., and Herring, M. (2005) Missing Data Methods for Generalized Linear Models: A Comparative Review, *Journal of the American Statistical Association*, 100, 332-46.
14. Kenward, M. (1998) Selection Models for Repeated Measurements with Non-Random Dropout: An Illustration of Sensitivity, *Statistics in Medicine*, 17, 2723-32.
15. Kenward M., Carpenter J. (2007) Multiple imputation: current perspectives, *Statistical Methods in Medical Research*, 16, 199-218.
16. Laird, N. (1988) Missing data in longitudinal studies, *Statistics in Medicine*, 7, 305-15.
17. Lipsitz, S., Ibrahim, J., and Zhao, L. (1999) A weighted estimating equation for missing covariate data with properties similar to maximum likelihood, *Journal of the American Statistical Association*, 94, 1147-60.
18. Lipsitz, S., Molenberghs, G., Fitzmaurice, G., and Ibrahim, J. (2000) GEE with Gaussian estimation of the correlations when data are incomplete, *Biometrics*, 56, 528-536.
19. Little, R. (1986) Survey nonresponse adjustments. *International Statistical Review*, 54, 139-157.
20. Little, R. and Rubin, D. (2002) *Statistical Analysis with Missing Data*, Wiley Series in Probability and Statistics, 2nd edition.

21. Molenberghs, G., Kenward, M., and Goetghebeur, E. (2001) Sensitivity analysis for incomplete contingency tables: the Slovenian plebiscite case, *Appl. Statist.*, 50, 15-29.
22. Nie, Hull, Jenkins, Steinbrenner, and Bent (1975). *SPSS: Statistical Package for the Social Sciences*. 2nd ed. New York, McGraw-Hill.
23. Paik, M. (1997) The generalized estimating equation approach when data are not missing completely at random, *Journal of the American Statistical Association*, 92, 1320-29.
24. Paik, M. (2004) Nonignorable Missingness in Matched Case-Control Data Analysis, *Biometrics*, 60, 306-14.
25. Preisser, J., Lohman, K., and Rathouz, P. (2002) Performance of weighted estimating equations for longitudinal binary data with drop-outs missing at random, *Statistics in Medicine*, 21, 3035-54.
26. Robins, J., Rotnitzky A., and Zhao L. (1994) Estimation of regression coefficients when some regressors are not always observed, *Journal of the American Statistical Association*, 89, 846-66.
27. Robins J. and Rotnitzky, A. (1995) Semiparametric efficiency in multivariate regression models with missing data, *Journal of the American Statistical Association*, 90, 106-121.
28. Rosenbaum, P. and Rubin, D. (1983) The central role of the propensity score in observational studies for causal effects. *Biometrika*, 70, 41-55.
29. Rosenbaum, P. and Rubin, D. (1985) Constructing a control group using multivariate matched sampling methods that incorporate the propensity score, *American Statistician* 39, 33-38.
30. Rotnitzky, A., Robins, J., and Scharfstein, D. (1998) Semiparametric regression for repeated outcomes with nonignorable nonresponse, *Journal of the American Statistical Association*, 93, 1321-39.
31. Rubin, D. B. (1976) Inference and Missing Data, *Biometrika* 63, 581-592.
32. Rubin, D. B. (1977) Formalizing Subjective Notion about the Effect of Nonrespondents in Sample Surveys, *Journal of the American Statistical Association*, 72, 538-543.
33. Rubin, D. (1987) *Multiple Imputation for Nonresponse in Surveys*. New York, John Wiley & Sons.
34. Rubin, D. (1996) Multiple Imputation after 18+ Years (with discussion) *Journal of the American Statistical Association*, 91, 473-489.
35. Schafer, J. (1997) *Analysis of incomplete multivariate data*, London, Chapman and Hall.
36. Schafer J. (1999) Multiple imputation: a primer, *Statistical Methods in Medical Research*, 8, 3-15.
37. Scharfstein, D., Rotnitzky, A. and Robins, J. (1999) Adjusting for nonignorable drop-out using semi-parametric nonresponse models (with comments), *Journal of the American Statistical Association*, 94, 1096-1146.
38. Sijsma, K., and van der, A. (2003) Investigation and treatment of missing scores in test and questionnaire data, *Multivariate Behavioral Research*, 38, 505-528.
39. Troxel, A., Harrington, D., and Lipsitz, S. (1998) *Analysis of Longitudinal Data with*

- Non-ignorable Non-monotone Missing Values, *Journal of Applied Statistics*, 47(3), 425-438.
- 40. Verbeke, G., Molenberghs, G., Thijs, H., Lesaffre, E. and Kenward, M. (2001) Sensitivity analysis for non-random dropout: a local influence approach, *Biometrics*, 57, 7-14.
 - 41. Wang, Q. and Rao, J. (2002) Empirical Likelihood-Based Inference under Imputation for Missing Response Data, *The Annals of Statistics*, 30(3), 896-924.
 - 42. Xie, F., Paik, M. (1997) Generalized estimating equation model for binary outcomes with missing covariates, *Biometrics*, 53, 1458-66.
 - 43. Zhao LP, Lipsitz S, and Lew D. (1996) Regression analysis with missing covariate data using estimating equations, *Biometrics*, 52, 1165-1182.
 - 44. Zhu, H. and Lee, S. (2001) A Bayesian analysis of finite mixtures in the LISREL model, *Psychometrika*, 66, 133-52.

Table 9: Estimates for TLI

		estimate	SE	Z	P-value
Int	$\hat{\beta}_{MI}$	-0.0838	0.3900	-0.2149	0.8299
	$\hat{\beta}_{prop}$	0.0385	0.3929	0.0980	0.9219
	$\hat{\beta}_{ml}$	0.0604	0.3769	0.1603	0.8966
x_1	$\hat{\beta}_{MI}$	-0.0079	0.5290	-0.0149	0.9881
	$\hat{\beta}_{prop}$	-0.0297	0.5327	-0.0558	0.9555
	$\hat{\beta}_{ml}$	0.0894	0.5043	0.1773	0.8850
x_2	$\hat{\beta}_{MI}$	0.3913	0.4983	0.7853	0.4323
	$\hat{\beta}_{prop}$	0.3855	0.4944	0.7797	0.4356
	$\hat{\beta}_{ml}$	0.2876	0.4822	0.5964	0.5361
x_3	$\hat{\beta}_{MI}$	0.1307	0.4873	0.2682	0.7885
	$\hat{\beta}_{prop}$	-0.0941	0.4950	-0.1901	0.8492
	$\hat{\beta}_{ml}$	-0.0991	0.4442	-0.2231	0.8200

Table 10: Estimates for Women data

		estimate	SE	Z	P-value
Int	$\hat{\beta}_{MI}$	-1.5537	0.1074	-14.4665	0.0000
	$\hat{\beta}_w$	-1.4923	0.0967	-15.4323	0.0000
	$\hat{\beta}_{propI}$	-1.4890	0.1053	-14.1406	0.0000
	$\hat{\beta}_{propII}$	-1.5788	0.1083	-14.5780	0.0000
time	$\hat{\beta}_{MI}$	0.6511	0.1894	3.4377	0.0006
	$\hat{\beta}_w$	0.5335	0.1073	4.9720	0.0000
	$\hat{\beta}_{propI}$	0.5139	0.1305	3.9379	0.0001
	$\hat{\beta}_{propII}$	0.7195	0.1420	5.0669	0.0000
t^2	$\hat{\beta}_{MI}$	0.0299	0.0604	0.4950	0.6206
	$\hat{\beta}_w$	-0.0058	0.0256	-0.2266	0.8207
	$\hat{\beta}_{propI}$	0.0019	0.0398	0.0477	0.9620
	$\hat{\beta}_{propII}$	-0.0971	0.0368	2.6386	0.0083
trt	$\hat{\beta}_{MI}$	0.1024	0.1553	0.6594	0.5096
	$\hat{\beta}_w$	0.1057	0.1395	0.7577	0.4486
	$\hat{\beta}_{propI}$	0.1057	0.1412	0.7486	0.4541
	$\hat{\beta}_{propII}$	0.1141	0.1545	0.7385	0.4602
trt*t	$\hat{\beta}_{MI}$	0.4734	0.2763	1.7134	0.0866
	$\hat{\beta}_w$	0.4069	0.1559	2.6100	0.0091
	$\hat{\beta}_{propI}$	0.4248	0.1829	2.3226	0.0202
	$\hat{\beta}_{propII}$	0.4272	0.1791	2.3853	0.0171
trt*t ²	$\hat{\beta}_{MI}$	-0.1446	0.0877	-1.6488	0.0992
	$\hat{\beta}_w$	-0.1318	0.0374	-3.5241	0.0004
	$\hat{\beta}_{propI}$	-0.1271	0.0559	-2.2737	0.0230
	$\hat{\beta}_{propII}$	-0.1341	0.0383	-3.5031	0.0004

Table 11: Estimates for missing covariate

		estimate	SE	Z	P-value
x_1	$\hat{\beta}_{MI}$	1.4949	0.1716	8.7115	0.0000
	$\hat{\beta}_w$	1.4080	0.1744	8.0734	0.0000
	$\hat{\beta}_{propI}$	1.8694	0.2485	7.5227	0.0000
	$\hat{\beta}_{prop}$	1.8887	0.2669	7.0764	0.0000
w_1	$\hat{\beta}_{MI}$	10.5166	0.6187	16.9979	0.0000
	$\hat{\beta}_w$	11.2639	0.4926	22.8662	0.0000
	$\hat{\beta}_{propI}$	10.1055	0.6629	15.2444	0.0000
	$\hat{\beta}_{prop}$	10.7647	0.7816	13.7726	0.0000
w_2	$\hat{\beta}_{MI}$	9.4779	0.5283	17.9404	0.0000
	$\hat{\beta}_w$	10.3455	0.5273	19.6198	0.0000
	$\hat{\beta}_{propI}$	9.3242	0.5500	16.9531	0.0000
	$\hat{\beta}_{prop}$	10.0871	0.6579	15.3323	0.0000

Table 12: Biases, square root of mean squared errors (RMSE) and standard errors for simulation based on TLI with sample size 82

		true β	ML	MI	Prop
Est	Int	-0.1642	-0.0364	0.0610	-0.0356
RMSE(std)			0.3416 (0.3397)	0.2701(0.2631)	0.3457(0.3439)
Est	x_1	-0.4013	0.0182	0.2063	-0.0201
RMSE(std)			0.4743(0.4740)	0.4422(0.3911)	0.4752(0.4748)
Est	x_2	0.3554	0.0932	-0.1297	0.0839
RMSE(std)			0.5859(0.5784)	0.4864(0.4688)	0.5861(0.5801)
Est	x_3	0.1093	0.0590	-0.0133	-0.0058
RMSE(std)			0.8084(0.8062)	0.6008(0.6007)	0.8108(0.8108)

Table 13: Biases and standard errors for simulation based on TLI with sample size 200

		beta	ML	MI	Prop	
mir=40 %	Bias	Int	-0.1642	-0.0044	0.0078	-0.0090
	std			0.2792	0.2801	0.2803
	Bias	x_1	-0.4013	-0.0164	-0.0313	-0.0297
	std			0.4520	0.4558	0.4650
	Bias	x_2	0.3554	0.1540	0.0105	0.0200
	std			0.5131	0.4530	0.4612
mir=70 %	Bias	x_3	1.6093	-0.1030	0.0720	-0.0612
	std			0.5801	0.4301	0.4290
	Bias	Int	-0.1642	-0.0531	-0.0006	0.0143
	std			0.2891	0.3178	3201
	Bias	x_1	-0.4013	0.0233	0.0545	-0.0380
	std			0.5081	0.6001	0.6585
	Bias	x_2	0.3554	-0.1312	-0.0557	0.0607
	std			0.5383	0.5810	0.6012
	Bias	x_3	1.6093	-0.2011	-0.1001	-0.1200
	std			0.5889	0.5427	0.5431

Note: mir represents the missing rate of the data, e.g. 40% means 40% subjects have missing values.

Table 14: Biases and standard errors for the simulation based on women data with sample size 250

			true β	WEE	MI	PropI	PropII
mir=27%	Est	Int	-1.5072	-0.0158	-0.0133	-0.0350	-0.0072
	std			0.1644	0.1641	0.1686	0.1640
	Est	time	0.5114	-0.0115	0.0026	0.0065	0.0004
	std			0.0640	0.0562	0.0654	0.0622
	Est	x	0.2653	-0.0058	-0.0029	0.0124	-0.0010
	std			0.2064	0.1801	0.1810	0.1800
mir=53%	Est	Int	-1.5072	-0.0058	-0.0100	-0.0048	-0.0072
	std			0.1909	0.1868	0.1948	0.1910
	Est	time	0.5114	-0.0129	0.0009	0.0204	0.0021
	std			0.0842	0.0731	0.0860	0.0738
	Est	x	0.2653	0.0008	0.0152	-0.0108	-0.0100
	std			0.2329	0.2220	0.2417	0.2310

Note: mir represents the missing rate of the data, e.g. 27% means 27% subjects have missing values.

Table 15: Estimates and standard errors for 500 simulation based on missing covariates data with sample size of 200

		true β	WEE	MI	ML	Prop
missingrate	25%	n=200				
Est	Int	2.3853	2.4374	2.2538	2.3983	2.4046
std			0.1116	0.1315	0.0990	0.1035
Est	time	8.2119	7.9560	9.2176	8.1063	8.1205
std			0.5891	0.6776	0.5192	0.5134
Est	x	6.3340	6.0858	7.3361	6.1968	6.2011
std			0.4032	0.8326	0.4008	0.3730

General Conclusion

This dissertation proposed a class of semiparametric estimators, which is consistent when missingness mechanism is either ignorable or nonignorable. The proposed estimators are based on unconditional model, therefore, we have certain assumptions on covariates and associations among covariates. But they do not require full specification of the distributions, only the first few moments of covariates are necessary, and generally only the first and second moments, which makes the assumptions are easy to meet. This is a significant advantage over likelihood-based approaches.

When missingness mechanism is MAR, The proposed approaches does not require additional model assumptions for missingness mechanism, only use available cases to obtain consistent estimators. This is a significant advantage over weighted and multiple imputation methods.

The results of the simulation study suggest that the proposed approaches have negligible bias either under MAR or nonignorable missingness. The simulation results also indicate that the bias of the estimators from the proposed approach is comparable to the imputation, weighted method, and maximum likelihood approach in most of the situations. The variances of the proposed estimators are comparable to those of the weighted and imputation approaches for missing covariates or missing both. Under missing response situation, the first propose approach has larger variances than the weighted GEE and multiple imputation. But the second propose approach has similar or smaller variances than the weighted GEE and multiple imputation when missingness is correctly specified.

The proposed approach can easily handle nonignorable missingness. But weighted method cannot be extended to nonignorable missingness easily, while the multiple imputation has to use Bayesian imputation approach.

Furthermore, the proposed method can easily be applied to arbitrary missing patterns. The weighted method cannot be easily extended to arbitrary missing patterns. Sequential imputation approach (Paik, 1997) could be extended to arbitrary missing patterns, but it

does perform as well as the proposed approach based on the simulation results.

Although currently we have no extensive simulation study results available, we conjecture that the proposed method will perform well for other types of data sets, it will be very interesting to study the results of applying the approach to other types of data sets (such as count data). If the future research could overcome the numerical difficulties in solving unconditional equations, we could apply these approach more easily and widely.

Since we have numerical difficulties in solving unconditional equations and lots of nuisance parameters sometime, we are considering using conditional model or partial conditional model to handle missing values. We are developing methodology based on conditional model, which has similar idea to the approach proposed by Lipsitz et al. (2000), the approach based on conditional model could reduce the number of parameters. We think partial conditional model is another way to handle missing values. The partial conditional model we mentioned here means we treat fully observed covariates not as random variables but treat covariates with missing values as random variables. By this way, we could reduce the number of parameters and make the equations more easily to solve.

Bibliography

1. Afifi, A. and Elashoff, R. (1966) Missing observations in multivariate statistics I: Review of the literature. *Journal of the American Statistical Association*, 61, 595-604.
2. Allison, P. (2001) *Missing data*, Thousand Oaks, CA, Sage.
3. Barnard, J. and Meng, X.-L. (1999) Applications of multiple imputation in medical studies: from AIDS to NHANES. *Statistical Methods in Medical Research*, 8, 7-36.
4. Bernaards, C. and Sijtsma, K. (2000) Influence of Imputation and EM Methods on Factor Analysis When Item Nonresponse in Questionnaires is Nonignorable. *Multivariate Behavioral Research*. 35 (3) 321-364.
5. Colice, G., Stukel, T., and Dain, B. (1989), *Laryngeal Complications of Prolonged Intubation*, Chest, 96, 877-884.
6. Diggle P. and Kenward M. (1994) Informative dropout in longitudinal data analysis (with discussion), *Applied Statistics*, 43, 49-93.
7. Dempster, A., Laird, M., and Rubin, D. (1977) Maximum likelihood from incomplete data via the EM algorithm. *J. R. Statist. Soc. B*, 39, 185-197.
8. Greene, H. (2000) *Econometric analysis*, 4th ed, Englewood Cliffs, NJ: Prentice Hall.
9. Green, P. and Park, T. (2003) A Bayesian Hierarchical Model for Categorical Data for Nonignorable Nonresponse, *Biometrics*, 59, 886-896.
10. Hartley, H. (1958) Maximum Likelihood Estimation from Incomplete Data *Biometrics*, 14(2), 174-194.
11. Horvitz, D. and Thompson, D. (1952) A Generalization of Sampling without replacement from a Finite Universe, *Journal of the American Statistical Association*, 47, 663-85.
12. Huang, L., Chen, M., and Ibrahim, J. (2005) Bayesian analysis for generalized linear models with nonignorably missing covariates, *Biometrics* 61, 767-78.
13. Ibrahim, J. (1990) Incomplete Data in Generalized Linear Models, *Journal of the American Statistical Association*, 85, 765-69.
14. Ibrahim, J., Lipsitz, S., and Chen, M. (1999) Missing covariates in generalized linear models when the missing data mechanism is nonignorable, *Journal of the Royal Statistical Society, Series B* 61, 173-190.
15. Ibrahim, J., Chen, M., Lipsitz, S., and Herring, M. (2005) Missing Data Methods for Generalized Linear Models: A Comparative Review, *Journal of the American Statistical Association*, 100, 332-46.
16. Kenward, M. (1998) Selection Models for Repeated Measurements with Non-Random Dropout: An Illustration of Sensitivity, *Statistics in Medicine*, 17, 2723-32.
17. Kenward M., Carpenter J. (2007) Multiple imputation: current perspectives, *Statistical Methods in Medical Research*, 16, 199-218.
18. Laird, N. (1988) Missing data in longitudinal studies, *Statistics in Medicine*, 7, 305-15.
19. Lipsitz, S., Ibrahim, J., and Zhao, L. (1999) A weighted estimating equation for missing covariate data with properties similar to maximum likelihood, *Journal of the American Statistical Association*, 94, 1147-60.
20. Lipsitz, S., Molenberghs, G., Fitzmaurice, G., and Ibrahim, J. (2000) GEE with Gaussian estimation of the correlations when data are incomplete, *Biometrics*, 56, 528-536.

21. Little, R. (1986) Survey nonresponse adjustments. *International Statistical Review*, 54, 139-157.
22. Little, R. and Rubin, D. (2002) *Statistical Analysis with Missing Data*, Wiley Series in Probability and Statistics, 2nd edition.
23. McCullagh, P. and Nelder, J. (1983) *Generalized Linear Models*. Chapman and Hall.
24. Molenberghs, G., Kenward, M., and Goetghebeur, E. (2001) Sensitivity analysis for incomplete contingency tables: the Slovenian plebiscite case, *Appl. Statist.*, 50, 15-29.
25. Mossey, J., Knotf K., and Crak, R. (1990) Effects of persistent depressive symptoms on hip fracture recovery, *Journal of Gerontology: Medical Sciences*, 45, 163-168.
26. Nie, Hull, Jenkins, Steinbrenner, and Bent (1975). *SPSS: Statistical Package for the Social Sciences*. 2nd ed. New York, McGraw-Hill.
27. Paik, M. (1997) The generalized estimating equation approach when data are not missing completely at random, *Journal of the American Statistical Association*, 92, 1320-29.
28. Paik, M. (2004) Nonignorable Missingness in Matched Case-Control Data Analysis, *Biometrics*, 60, 306-14.
29. Preisser, J., Lohman, K., and Rathouz, P. (2002) Performance of weighted estimating equations for longitudinal binary data with drop-outs missing at random, *Statistics in Medicine*, 21, 3035-54.
30. Robins, J., Rotnitzky A., and Zhao L. (1994) Estimation of regression coefficients when some regressors are not always observed, *Journal of the American Statistical Association*, 89, 846-66.
31. Robins J. and Rotnitzky, A. (1995) Semiparametric efficiency in multivariate regression models with missing data, *Journal of the American Statistical Association*, 90, 106-121.
32. Rosenbaum, P. and Rubin, D. (1983) The central role of the propensity score in observational studies for causal effects. *Biometrika*, 70, 41-55.
33. Rosenbaum, P. and Rubin, D. (1985) Constructing a control group using multivariate matched sampling methods that incorporate the propensity score, *American Statistician* 39, 33-38.
34. Rotnitzky, A., Robins, J., and Scharfstein, D. (1998) Semiparametric regression for repeated outcomes with nonignorable nonresponse, *Journal of the American Statistical Association*, 93, 1321-39.
35. Rubin, D. B. (1976) Inference and Missing Data, *Biometrika* 63, 581-592.
36. Rubin, D. B. (1977) Formalizing Subjective Notion about the Effect of Nonrespondents in Sample Surveys, *Journal of the American Statistical Association*, 72, 538-543.
37. Rubin, D. (1987) *Multiple Imputation for Nonresponse in Surveys*. New York, John Wiley & Sons.
38. Rubin, D. (1987) Multiple Imputation in Sample Surveys, *Proc. Survey Res. Meth. Sec., Am. Stat. Assoc.* 1978, 20-34.
39. Rubin, D. (1996) Multiple Imputation after 18+ Years (with discussion) *Journal of the American Statistical Association*, 91, 473-489.
40. Schafer, J. (1997) *Analysis of incomplete multivariate data*, London, Chapman and Hall.

41. Schafer J. (1999) Multiple imputation: a primer, *Statistical Methods in Medical Research*, 8, 3-15.
42. Scharfstein, D., Rotnitzky, A. and Robins, J. (1999) Adjusting for nonignorable drop-out using semi-parametric nonresponse models (with comments), *Journal of the American Statistical Association*, 94, 1096-1146.
43. Sijtsma, K., and van der, A. (2003) Investigation and treatment of missing scores in test and questionnaire data, *Multivariate Behavioral Research*, 38, 505-528.
44. Troxel, A., Harrington, D., and Lipsitz, S. (1998) Analysis of Longitudinal Data with Non-ignorable Non-monotone Missing Values, *Journal of Applied Statistics*, 47(3), 425-438.
45. Verbeke, G., Molenberghs, G., Thijs, H., Lesaffre, E. and Kenward, M. (2001) Sensitivity analysis for non-random dropout: a local influence approach, *Biometrics*, 57, 7-14.
46. Wang, Q. and Rao, J. (2002) Empirical Likelihood-Based Inference under Imputation for Missing Response Data, *The Annals of Statistics*, 30(3), 896-924.
47. Xie, F., Paik, M. (1997) Generalized estimating equation model for binary outcomes with missing covariates, *Biometrics*, 53, 1458-66.
48. Zhao LP, Lipsitz S, and Lew D. (1996) Regression analysis with missing covariate data using estimating equations, *Biometrics*, 52, 1165-1182.
49. Zhu, H. and Lee, S. (2001) A Bayesian analysis of finite mixtures in the LISREL model, *Psychometrika*, 66, 133-52.

APPENDIX

Proofs of the Theorems

Proofs in Chapter 2

Proof of Lemma 2.1: We write

$$\begin{aligned}
 E_\psi(g(\mathbf{U}, \psi)|R = r, U_r = u) &= \sum_u g(\mathbf{U}; \psi) P_\psi(U = u|R = r, U_r = u) \\
 &= \sum_u g(\mathbf{U}; \psi) \frac{P_\psi(U = u, R = r, U_r = u)}{P_\psi(R = r, U_r = u)} \\
 &= \sum_u g(\mathbf{U}; \psi) \frac{P_\psi(R = r|U = u, U_r = u) P_\psi(U = u, U_r = u)}{P_\psi(R = r|U_r = u) P_\psi(U_r = u)}.
 \end{aligned}$$

Under the MAR assumption, we have

$$P_\psi(R = r|U = u, U_r = u) = P_\psi(R = r|U_r = u).$$

Therefore,

$$\begin{aligned}
 &\sum_u g(\mathbf{U}; \psi) \frac{P_\psi(R = r|U = u, U_r = u) P_\psi(U = u, U_r = u)}{P_\psi(R = r|U_r = u) P_\psi(U_r = u)} \\
 &= \sum_u g(\mathbf{U}; \psi) \frac{P_\psi(U = u, U_r = u)}{P_\psi(U_r = u)} \\
 &= \sum_u g(\mathbf{U}; \psi) P_\psi(U = u|U_r = u).
 \end{aligned}$$

Proof of Lemma 2.2: The estimating function $g(\mathbf{U}, \psi) = \sum_{i=1}^N D'V^{-1}(U_i - \mu_i)$, where $D = \partial\mu/\partial\psi$.

$$\sum_r E_\psi g(\mathbf{U}, \psi)|R_r, U_r = u = \sum_r E_\psi g(\mathbf{U}, \psi)|U_r = u.$$

We write U_i as two parts

$$U_i = \begin{pmatrix} U_i^o \\ U_i^m \end{pmatrix}.$$

Therefore, the covariance matrix V_i can be partitioned as

$$V_i = \begin{pmatrix} V_i^O & V_i^{mo} \\ V_i^{mo} & V_i^m \end{pmatrix}.$$

Using the conditional expectation. We have

$$E(U_i|U_i^o) = E\left(\begin{pmatrix} U_i^o \\ U_i^m \end{pmatrix} | U_i^o\right) = \begin{pmatrix} E(U_i^o|U_i^o) \\ E(U_i^m|U_i^o) \end{pmatrix} = \begin{pmatrix} U_i^o \\ E(U_i^m|U_i^o) \end{pmatrix}.$$

If U_i has a multivariate normal distribution, we have

$$\begin{pmatrix} U_i^o - \mu_i \\ E(U_i^m - \mu_i|U_i^o) \end{pmatrix} = \begin{pmatrix} U_i^o - \mu_i \\ V_i^{mo}(V_i^o)^{-1}(U_i^o - \mu_i) \end{pmatrix} = \begin{pmatrix} I \\ V_i^{mo}(V_i^o)^{-1} \end{pmatrix} (U_i^o - \mu_i). \quad (4.1)$$

If U_i does not have a multivariate normal distribution, by Taylor expansion, equation (4.1) becomes

$$\begin{pmatrix} U_i^o - \mu_i \\ E(U_i^m - \mu_i|U_i^o) \end{pmatrix} \approx \begin{pmatrix} I \\ V_i^{mo}(V_i^o)^{-1} \end{pmatrix} (U_i^o - \mu_i). \quad (4.2)$$

Let $H = V_i^o - V_i^{om}(V_i^m)^{-1}(V_i^{mo})'$, $D = V_i^m$, and $B = V_i^{om} = (V_i^{mo})'$, we have

$$V^{-1} \begin{pmatrix} U_i^o - \mu_i \\ E(U_i^m - \mu_i|U_i^o) \end{pmatrix}. \quad (4.3)$$

Since $BD^{-1}B' = V_i^o - H$, we have

$$H^{-1} - H^{-1}BD^{-1}B'(V_i^o)^{-1} = H^{-1} - H^{-1}(V_i^o - H)(V_i^o)^{-1} = (V_i^o)^{-1},$$

and

$$\begin{aligned} & -D^{-1}B'H^{-1} + D^{-1}B'(V_i^o)^{-1} + D^{-1}B'H^{-1}BD^{-1}B'(V_i^o)^{-1} \\ & = -D^{-1}B'H^{-1} + D^{-1}B'(V_i^o)^{-1} + D^{-1}B'H^{-1}(V_i^o - H)(V_i^o)^{-1} = 0. \end{aligned}$$

Therefore, (4.3) becomes

$$\begin{aligned}
& \begin{pmatrix} H^{-1} & -H^{-1}BD^{-1} \\ -D^{-1}B'H^{-1} & D^{-1} + D^{-1}B'H^{-1}BD^{-1} \end{pmatrix} \begin{pmatrix} I \\ V_i^{mo}(V_i^o)^{-1} \end{pmatrix} (U_i^o - \mu_i) \\
&= \begin{pmatrix} H^{-1} - H^{-1}BD^{-1}V_i^{mo}(V_i^o)^{-1} \\ -D^{-1}B'H^{-1} + D^{-1} + D^{-1}B'H^{-1}BD^{-1}V_i^{mo}(V_i^o)^{-1} \end{pmatrix} (U_i^o - \mu_i) \\
&= \begin{pmatrix} (V_i^o)^{-1} \\ 0 \end{pmatrix} (U_i^o - \mu_i).
\end{aligned}$$

For a specific pattern s , we write D as $\begin{pmatrix} D_o \\ D^m \end{pmatrix}$, the quasi-score is

$$E_\psi\{g(\mathbf{U}, \psi)|R_r, U_r = u\} = \sum_{i=1}^{n_i} D'(V^o)^{-1}(U_i^o - \mu_i^o) = \sum_{i=1}^{n_i} (D^o)'(V^o)^{-1}(U_i^o - \mu_i^o) = 0.$$

Therefore,

$$\sum_r E_\psi\{g(\mathbf{U}, \psi)|R_r, U_r = u\} = \sum_r \sum_{i=1}^{n_r} (D^o)'(V^o)^{-1}(U_i^o - \mu_i^o) = 0.$$

Proof of consistency 2.1: We assume that all subjects are i.i.d. The log quasiliikelihood for the i^{th} subject is $V^{-1}(\mathbf{u}_i - \mu)$, so that the log likelihood for all subjects is $\sum_{i=1}^N V^{-1}(\mathbf{u}_i - \mu)$. The estimators $\hat{\psi}$ maximize the log quasiliikelihood,

$$Q_n(\psi) = \frac{1}{N} \sum_{i=1}^N V^{-1}(u_i - \mu).$$

The log quasiliikelihood function $Q_n(\psi)$ depends on data. Define

$$Q(\psi) = E_\psi V^{-1}(u_i - \mu).$$

where E_ψ is the expectation with respect to the true parameter ψ_0 . By law of large numbers, for any ψ ,

$$Q_n(\psi) \rightarrow E_{\psi_0} V^{-1}(u_i - \mu) = Q(\psi).$$

Note that $Q(\psi)$ only depends on ψ .

Lemma For any $Q(\psi)$, $Q(\psi_0)$ unless

$$P[Q(\psi) = Q(\psi_0)] = 1.$$

Proof: Let us consider the difference

$$\begin{aligned} Q(\psi) - Q(\psi_0) &= E_{\psi_0}[V^{-1}(u_i - \mu_\psi) - V^{-1}(u_i - \mu_{\psi_0})] \\ &\leq E_{\psi_0}\left[\frac{F(u_i, \psi)}{F(u_i, \psi_0)} - 1\right] = 0, \end{aligned}$$

where $F(u_i, \psi)$ and $F(u_i, \psi_0)$ are quasiliikelihood functions.

We know ψ maximizes $Q_n(\psi)$, and ψ_0 maximizes $Q(\psi)$ and for all ψ , we have $Q_n(\psi) \rightarrow Q(\psi)$ by LLN. Therefore, since two functions $Q_n(\psi)$ and $Q(\psi)$ are getting closer, the points of maximum should also get closer which exactly means the consistency.

Proof of Theorem 2.1: We assume that all subjects are i.i.d. The log quasiliikelihood for the i^{th} subject is $V^{-1}(\mathbf{u}_i - \mu)$, so that the log likelihood for all subjects is $\sum_{i=1}^N V^{-1}(\mathbf{u}_i - \mu)$.

Proof of consistency: We assume that all subjects are i.i.d. The log quasiliikelihood for the i^{th} subject is $V^{-1}(\mathbf{u}_i - \mu)$, so that the log likelihood for all subjects is $\sum_{i=1}^N V^{-1}(\mathbf{u}_i - \mu)$. The estimators $\hat{\psi}$ maximize the log quasiliikelihood,

$$Q_n(\psi) = \frac{1}{N} \sum_{i=1}^N V^{-1}(u_i - \mu).$$

The log quasiliikelihood function $Q_n(\psi)$ depends on data. Define

$$Q(\psi) = E_\psi V^{-1}(u_i - \mu).$$

where E_ψ is the expectation with respect to the true parameter ψ_0 . By law of large numbers, for any ψ ,

$$Q_n(\psi) \rightarrow E_{\psi_0} V^{-1}(u_i - \mu) = Q(\psi).$$

Note that $Q(\psi)$ only depends on ψ .

Lemma For any $Q(\psi)$, $Q(\psi_0)$ unless

$$P[Q(\psi) = Q(\psi_0)] = 1.$$

Proof: Let us consider the difference

$$\begin{aligned} Q(\psi) - Q(\psi_0) &= E_{\psi_0}[V^{-1}(u_i - \mu_\psi) - V^{-1}(u_i - \mu_{\psi_0})] \\ &\leq E_{\psi_0}\left[\frac{F(u_i, \psi)}{F(u_i, \psi_0)} - 1\right] = 0, \end{aligned}$$

where $F(u_i, \psi)$ and $F(u_i, \psi_0)$ are quasiliikelihood functions.

We know ψ maximizes $Q_n(\psi)$, and ψ_0 maximizes $Q(\psi)$ and for all ψ , we have $Q_n(\psi) \rightarrow Q(\psi)$ by LLN. Therefore, since two functions $Q_n(\psi)$ and $Q(\psi)$ are getting closer, the points of maximum should also get closer which exactly means the consistency.

The estimators $\hat{\psi}$ maximize the log quasiliikelihood,

$$Q_n(\psi) = \frac{1}{N} \sum_{i=1}^N V^{-1}(u_i - \mu).$$

We have

$$Q'_n(\psi) = \frac{1}{N} \sum_{i=1}^N D'V^{-1}(u_i - \mu) = 0,$$

where $D = \partial\mu/\partial\psi$.

Use first-order Taylor expansion, we have

$$f(a) = f(a_0) + \nabla f^T(a_1)(a - a_0), \quad (4.4)$$

with $f(\psi) = Q'_n(\psi)$, and $a_0 = \psi_0$, where ψ_0 is the true parameter. Let $Q''_n(\psi) = \partial Q'/\partial\psi$, by (4.4) we can write

$$0 = Q'_n(\hat{\psi}) = Q'_n(\psi_0) + Q''_n(\psi_1)(\hat{\psi} - \psi_0) \text{ for } \psi_1 \in [\hat{\psi}, \psi_0],$$

and therefore

$$(\hat{\psi} - \psi_0) = (Q''_n(\psi_1))^{-1}(Q'_n(\hat{\psi}) - Q'_n(\psi_0)) = -(Q''_n(\psi_1))^{-1}Q'_n(\psi_0). \quad (4.5)$$

The mean of $-Q'_n(\psi_0)$ is

$$E(Q'_n(\psi_0)) = \frac{1}{N} \sum_{i=1}^N D' V^{-1} (E(u_i) - \mu) = 0,$$

and the variance of $-Q'_n(\psi_0)$ is

$$\begin{aligned} \text{var}[Q'_n(\psi_0)] &= \frac{1}{N} D' V^{-1} \text{var}\left\{\sum_{i=1}^N (u_i - \mu)\right\} V^{-1} D \\ &= \frac{1}{N} D' V^{-1} \sum_{i=1}^N \{\text{var}(u_i - \mu)\} V^{-1} D' = D' V^{-1} D. \end{aligned}$$

By the central limit theorem, $Q'_n(\psi_0)$ converges to $N(\mathbf{0}, D' V^{-1} D)$ in distribution. By the weak law of large numbers, $Q''_n(\psi_1)$ converges to the mean of $Q''_n(\psi_1)$

$$\begin{aligned} E(Q''_n(\psi_1)) &= E\left[\frac{1}{N} \sum_{i=1}^N \partial D' / \partial \psi V^{-1} (u_i - \mu) - D' V^{-1} D\right] \\ &= -D' V^{-1} D. \end{aligned}$$

Therefore, we have

$$-Q''_n(\psi_1)^{-1} Q'_n(\psi_0) \rightarrow^d N(\mathbf{0}, (D' V^{-1} D)^{-1}).$$

Therefore, we have

$$\sqrt{N}(\hat{\psi} - \psi_0) = \sqrt{N}(-Q''_n(\psi_1)^{-1} Q'_n(\psi_0)) \rightarrow^d N(\mathbf{0}, (D' V^{-1} D)^{-1}).$$

Proof of Theorem 2.2: We assume that all subjects are i.i.d. The log quasiliikelihood for I^{th} subject in s pattern is $V_r^{-1}(\mathbf{u}_{r_i} - \mu_{\mathbf{r}})$, so that the log likelihood for all subjects is $\sum_{i=1}^N V^{-1}(u_i - \mu) = \sum_r \sum_{i=1}^{n_r} V_r^{-1}(\mathbf{u}_{r_i} - \mu_{\mathbf{r}})$. The estimator $\hat{\psi}$ from the proposed approach maximizes the log likelihood,

$$Q_n(\psi) = \frac{1}{N} \sum_r \sum_{i=1}^{n_r} V_r^{-1}(\mathbf{u}_{r_i} - \mu_{\mathbf{r}}).$$

By Lemma 2.2, we have

$$Q'_n(\psi) = \frac{1}{N} \sum_r \sum_{i=1}^{n_r} D'_r V_r^{-1}(\mathbf{u}_{r_i} - \mu_{\mathbf{r}}) = 0,$$

where $D_r = \partial\mu_r/\partial\psi$.

By first-order Taylor expansion,

$$f(a) = f(a_0) + \nabla f^T(a_1)(a - a_0), \quad (4.6)$$

with $f(\psi) = Q'_n(\psi)$, $a_0 = \psi_0$, where ψ_0 is the true parameter. Let $Q''_n(\psi) = \partial Q'/\partial\psi$, by (4.6) we can write

$$0 = Q'_n(\hat{\psi}) = Q'_n(\psi_0) + Q''_n(\psi_1)(\hat{\psi} - \psi_0) \text{ for } \psi_1 \in [\hat{\psi}, \psi_0],$$

and therefore

$$(\hat{\psi} - \psi_0) = (Q''_n(\psi_1))^{-1}(Q'_n(\hat{\psi}) - Q'_n(\psi_0)) = -(Q''_n(\psi_1))^{-1}Q'_n(\psi_0). \quad (4.7)$$

The mean of $-Q'_n(\psi_0)$ is

$$E(Q'_n(\psi_0)) = \frac{1}{N} \sum_r \sum_{i=1}^{n_r} D'_r V_r^{-1} (E(u_{r_i}) - \mu_r) = 0,$$

and the variance of $-Q'_n(\psi_0)$ is

$$\begin{aligned} \text{var}[Q'_n(\psi_0)] &= \frac{1}{N} D'_r V_r^{-1} \text{var}[\sum_r \sum_{i=1}^{n_r} (u_{r_i} - \mu_r)] V_r^{-1} D_r \\ &= \frac{1}{N} D'_r V_r^{-1} \sum_r \sum_{i=1}^{n_r} (\text{var}(u_{r_i} - \mu_r)) V_r^{-1} D_r = \sum_r \sum_{i=1}^{n_r} D'_r V_r^{-1} D_r. \end{aligned}$$

By the central limit theorem, the numerator in (4.7) converges to $N(\mathbf{0}, \sum_r \sum_{i=1}^{n_r} D'_r V_r^{-1} D_r)$ in distribution. By the weak law of large numbers, the denominator in (4.7) converges to the mean of $Q''_n(\psi_1)$

$$\begin{aligned} E(Q''_n(\psi_1)) &= E[\frac{1}{N} \sum_r \sum_{i=1}^{n_r} D'_r V_r^{-1} D_r \partial D'_r / \partial \psi V_r^{-1} (u_{r_i} - \mu_r) - \sum_r \sum_{i=1}^{n_r} D'_r V_r^{-1} D_r] \\ &= - \sum_r \sum_{i=1}^{n_r} D'_r V_r^{-1} D_r. \end{aligned}$$

Therefore, we have

$$-Q''_n(\psi_1)^{-1} Q'_n(\psi_0) \rightarrow^d N(\mathbf{0}, (\sum_r \sum_{i=1}^{n_r} D'_r V_r^{-1} D_r)^{-1}).$$

Therefore, we have

$$\sqrt{N}(\hat{\psi} - \psi_0) = \sqrt{N}(-Q''_n(\psi_1)^{-1} Q'_n(\psi_0)) \rightarrow^d N(\mathbf{0}, (\sum_r \sum_{i=1}^{n_r} D'_r V_r^{-1} D_r)^{-1}).$$

Proofs in Chapter 3

Proof of Theorem 3.1: We assume that all subjects are i.i.d. The log quasilikelihood for the i^{th} subject is $V^{-1}(\mathbf{w}_i - \mu)$, so that the log likelihood for all subjects is $\sum_{i=1}^N V^{-1}(\mathbf{w}_i - \mu)$. The estimators $\hat{\psi}$ maximize the log quasilikelihood,

$$Q_n(\psi) = \frac{1}{N} \sum_{i=1}^N V^{-1}(w_i - \mu).$$

We have

$$Q'_n(\psi) = \frac{1}{N} \sum_{i=1}^N D' V^{-1}(w_i - \mu) = 0,$$

where $D = \partial\mu/\partial\psi$.

Use first-order Taylor expansion, we have

$$f(a) = f(a_0) + \nabla f^T(a_1)(a - a_0), \quad (4.8)$$

with $f(\psi) = Q'_n(\psi)$, and $a_0 = \psi_0$, where ψ_0 is the true parameter. Let $Q''_n(\psi) = \partial Q'/\partial\psi$, by (4.8) we can write

$$0 = Q'_n(\hat{\psi}) = Q'_n(\psi_0) + Q''_n(\psi_1)(\hat{\psi} - \psi_0) \quad \text{for } \psi_1 \in [\hat{\psi}, \psi_0],$$

and therefore

$$(\hat{\psi} - \psi_0) = (Q''_n(\psi_1))^{-1}(Q'_n(\hat{\psi}) - Q'_n(\psi_0)) = -(Q''_n(\psi_1))^{-1}Q'_n(\psi_0). \quad (4.9)$$

The mean of $-Q'_n(\psi_0)$ is

$$E(Q'_n(\psi_0)) = \frac{1}{N} \sum_{i=1}^N D' V^{-1}(E(w_i) - \mu) = 0,$$

and the variance of $-Q'_n(\psi_0)$ is

$$\begin{aligned} \text{var}[Q'_n(\psi_0)] &= \frac{1}{N} D' V^{-1} \text{var}\left\{\sum_{i=1}^N (w_i - \mu)\right\} V^{-1} D \\ &= \frac{1}{N} D' V^{-1} \sum_{i=1}^N \{\text{var}(w_i - \mu)\} V^{-1} D = D' V^{-1} D. \end{aligned}$$

By the central limit theorem, $Q'_n(\psi_0)$ converges to $N(\mathbf{0}, D'V^{-1}D)$ in distribution. By the weak law of large numbers, $Q''_n(\psi_1)$ converges to the mean of $Q''_n(\psi_1)$

$$\begin{aligned} E(Q''_n(\psi_1)) &= E\left[\frac{1}{N} \sum_{i=1}^N \partial D' / \partial \psi V^{-1}(w_i - \mu) - D'V^{-1}D\right] \\ &= -D'V^{-1}D. \end{aligned}$$

Therefore, we have

$$-(Q''_n(\psi_1))^{-1}Q'_n(\psi_0) \rightarrow^d N(\mathbf{0}, (D'V^{-1}D)^{-1}).$$

Therefore, we have

$$\sqrt{N}(\hat{\psi} - \psi_0) = \sqrt{N}(-(Q''_n(\psi_1))^{-1}Q'_n(\psi_0)) \rightarrow^d N(\mathbf{0}, (D'V^{-1}D)^{-1}).$$

Table 16: TLI data

x1	x2	x3	y	Frequency
.	0	0	0	1
0	0	0	0	10
0	0	1	0	10
0	1	0	0	6
0	1	.	0	1
0	1	1	0	2
1	0	1	0	7
1	0	0	0	1
1	1	.	0	1
.	.	.	1	1
.	0	.	1	1
0	.	0	1	1
0	0	0	1	9
0	0	1	1	10
0	0	.	1	2
0	1	0	1	2
0	1	1	1	7
1	0	0	1	3
1	0	.	1	1
1	0	1	1	1
1	1	.	1	1
1	1	0	1	1
1	1	.	1	3