GENERALIZED UNIFIED APPROACH TO REGRESSION MODELS WITH COVARIATES MISSING IN NONMONOTONE PATTERNS

A THESIS

SUBMITTED TO THE FACULTY OF GRADUATE STUDIES AND RESEARCH

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS

FOR THE DEGREE OF

DOCTOR OF PHILOSIPHY

IN

STATISTICS

UNIVERSITY OF REGINA

By

Meng Liu

Regina, Saskatchewan

May, 2013

UNIVERSITY OF REGINA

FACULTY OF GRADUATE STUDIES AND RESEARCH

SUPERVISORY AND EXAMINING COMMITTEE

Meng Liu, candidate for the degree of Doctor of Philosophy in Statistics**,** has presented a thesis titled, ***Generalized Unified Approach to Regression Models with Covariates Missing in Nonmonotone Patterns,*** in an oral examination held on April 11, 2013. The following committee members have found the thesis acceptable in form and content, and that the candidate demonstrated satisfactory knowledge of the subject material.

External Examiner:	Dr. Bingshu E. Chen, Queen's University
Supervisor:	Dr. Yang Zhao, Department of Mathematics & Statistics
Committee Member:	Dr. Dianliang Deng, Department of Mathematics & Statistics
Committee Member:	Dr. Taehan Bae, Department of Mathematics & Statistics
Committee Member:	Dr. Liming Dai, Industrial Systems Engineering
Chair of Defense:	Dr. Philip Charrier, Department of History

*Not present at defense

Abstract

Complicated designs (eg. partially questionnaire design), which are often used in epidemiologic studies to reduce the cost of data collection while at the same time improving data quality, generate data with nonmonotone missing patterns. This thesis focuses on statistical inference for regression models with nonmonotone missing covariate data under some designs that generate nonmonotone missing data in covariates. Proposed methods in this scenario often depend on additional assumptions about covariates, for example, the covariates need to be categorical or follow a particular semiparametric joint distribution. This thesis describes a generalized unified estimation method for regression models with covariates missing in nonmonotone patterns which use a sequence of working regression models to extract information from incomplete observations. It can deal with both continuous and categorical variables. We consider both cross-sectional and longitudinal studies. The asymptotic theory and variance estimator for the generalized unified estimator are provided. Simulation studies in different settings are used to examine the proposed method. Finally we applied the generalized unified approach to the two real examples. One is a cross-sectional study, and the other is a longitudinal study.

Acknowledgements

First and foremost, I would like to express my deep and sincere gratitude to my supervisor, Dr. Yang Zhao. She has supported me throughout my thesis research with her wide knowledge, great patience, constant encouragement and personal guidance.

I also wish to express my special appreciation to Dr. Bingshu Chen, Dr. Dianliang Deng, Dr. Taehan Bae, and Dr. Liming Dai for their assistance and valuable advices and for serving as thesis committee members.

Many thanks go to all the faculty members, the administrative staff and my fellow graduate students in the department of Mathematics and Statistic at the University of Regina for their help rendered to me during my studies.

Finally, I am forever indebted to my wife, my son and my father-in-law for their understanding, endless encouragement and unconditional love.

Contents

Chapter 1

INTRODUCTION

Missing data frequently occur in epidemiological studies and clinical trials. For example, in epidemiological studies, two-phase sampling designs are used to reduce the cost of data collection. In this design, “cheaper” variables are measured for individuals selected in a phase I sample, then other variables including “expensive” or hard to measure variables, are measured for individuals selected into a subsample, a phase II sample. In clinical trials, missing data occur whenever one or more intended measurements are not taken, lost, or otherwise unavailable. Robins et al. (1994) called this case as “missing by happenstance”.

A naive method for missing data problems is the complete-case analysis. It discards incomplete observations. If the mechanism leading to the missingness is relevant to the response process, the complete-case analysis may produce biased results.

Little and Rubin (2002) defined three missing data mechanisms as (i) missing complete at random (MCAR) if the missing data process does not depend on any data (observed or unobserved); (ii) missing at random (MAR) if the missing data process does not depend on the unobserved data given the observed data; and (iii) not missing at random (NMAR) if the missing data process depends on the unobserved data given the observed data.

The missing-data pattern is another important concept. Little and Rubin (2002) mentioned that "We found it useful to distinguish the missing-data pattern, which describes which values are observed in the data matrix and which values are missing". Many methods described for missing data problems can only deal with simple monotone missingness patterns. When data are missing in arbitrary nonmonotone patterns many methods cannot be applied directly or require intensive computation.

This research focuses on regression models with covariates missing in arbitrary nonmonotone patterns. It deals with the MCAR data and the MAR data separately.

1.1 Literature Review

In epidemiologic studies, complex sampling designs are often used to reduce the cost of data collection while at the same time improving data quality. Complex sampling designs generate data with large proportion of missing values and different missing patterns. Two-phase sampling designs in Zhao and lipsitz (1992), for example, produce data with a simple monotone missing pattern, where the variables measured in phase I have no missing values, and the variables measured in phase II are missing for the subjects selected in the phase I sample but not selected in the phase II subsample. In general, the phase I sample is large whereas the phase II sample is relatively small. In addition, multiphase sampling designs in Holcroft et al. (1997) generate data with general monotone missing patterns, where the subjects selected in the current phase are observed in the previous phases but may not be observed in the future phases. Wacholder et al. (1994) proposed a partial questionnaire design (PQD) for lengthy questionnaires or other burdensome data-collection processes, where subsets of variables are measured for different, but overlapping, groups of subjects to reduce the cost of data collection while at the same time increasing participation and improving data quality. A PQD generates data with nonmonotone missing patterns.

Most of the estimation methods proposed for regression models with data missing by design depend on the assumption of monotone missing patterns (Little and Rubin 2002 ; Zhao and lipsitz 1992; Holcroft et al. 1997; Zhao et al. 2009). However, in regression models it is common that the covariate data are missing in nonmonotone patterns either by design or happenstance. In general estimation methods for monotone missing covariate data may be computationally complex or have difficulties to deal with nonmonotone missing patterns. The double robust estimating equations in Lipsitz and Zhao (1999) and Van der Laan and Robins (2003) may have closed form expressions for monotone missing covariate data but will be difficult to obtain for nonmonotone missing patterns. The semiparametric efficient inference developed by Robins et al. (1994) for semmiparametric regression models and by Robins et al. (1995) for parametric regression models is computationally complex and may be difficult to implement for nonmonotone missing patterns especially for continuous response.

Methods for the analysis of nonmonotone missing data are limited and often require additional assumptions. For example, the maximum likelihood method in Ibrahim et al.

(1990) requires the covariates to be categorical. The consistency of the semiparametric estimator in Chen (2004) for general nonmonotone missing covariates data depends on the correctness of the parametric odds-ratio model. The conditional model in Lipsitz and Ibrahim (1996) depends on parametric assumptions for the joint distribution of the covariates. The three techniques for a PQD described in Chatterjee and Li (2010), including the mean score method, the pseudo-likelihood method, and the full maximum likelihood, are extensions of Reilly and Pepe (1995), Scott and Wild (1998) and Zhao et al. (2009) to a PQD. These methods are based on nonparametic models for the joint distributions of the covariates and auxiliary variables and therefore require certain covariates to be categorical.

The purpose of this research is to develop easily implemented estimation methods for dealing regression models with nonmonotone missing data that obtained from complex designs, which will fill a needed gap in statistical analysis with missing data.

This thesis describes estimation methods for regression models with covariates missing in nonmonotone patterns under a PQD or other designs that generate nonmonotone missing data in covariates. Instead of modeling the distribution of the covariates we propose using a sequence of working regression models to extract information from the incomplete observations. This approach can be easily implemented to deal with both continuous and categorical variables. The initial idea was proposed in Chen and Chen (2000) for two-phase sampling designs based on simple random samples, where the variables observed in phase II are MCAR. In a PQD, the subjects are randomly selected into different, but overlapping, groups, and then different subsets of variables are measured for different groups.

In general, there is information available for all the subjects in the study, and the random selections of subjects into different groups often depend on this fully observed information. If this is the case, then the data are MAR. Motivating examples include (i) a study of occupational risk factors for adult onset asthma using a PQD in Houseman and Milton (2006) and (ii) a case-control study investigating the association of polychlorinated biphenyls with the risk of non-Hodgkin lymphoma (Colt et al. 2005; Deroos et al. 2005). In the latter study, two measurements of polychlorinated biphenyls, one based on home dust samples and the other based on blood plasma levels, were obtained for two distinct but overlapping groups of participants.

1.2 Organization of the Thesis

Chapter 2 describes a generalized unified estimation method for regression models with nonmonotone missing covariates in cross-sectional study. It considers both the MCAR case and the MAR case. It derives the asymptotic theory and variance estimator for the unified generalized estimator in each case. Numerical studies are implemented to examine the finite sample performance of the proposed method.

Chapter 3 extends the generalized unified estimation method for marginal model with nonmonotone missing covariates in longitudinal data analysis. It derives the asymptotic theory and variance estimator for the unified estimator for MCAR data and MAR data respectively. Numerical studies are used to examine the performance of the proposed method in several different settings.

Chapter 4 uses real-data examples in a cross-sectional study and a longitudinal study to illustrate the methods.

Chapter 5 gives a summary and a discussion of future work.

Chapter 2

Generalized Unified Estimation Method

Let Y be a response variable, $\mathbf{X}$ denote a vector of covariates, and $f(\mathbf{X};\beta)$ represent the conditional mean of Y given $\mathbf{X}$, where β is a vector of parameters. For convenience we consider estimating the β parameter in the mean function $f(\mathbf{X};\beta)$, but the procedures readily extend to estimation of the full distribution of Y given $\mathbf{X}$.

According to the finite set of missingness patterns in the observed data we reorder the covariates in $\mathbf{X}$ as $\mathbf{X} = (\mathbf{X}_1^T, \cdots, \mathbf{X}_q^T)^T$ such that each $\mathbf{X}_k$, $k = 1, \cdots, q$, is a vector of covariates with the same missingness pattern, where q is the total number of distinct patterns. We define indicator variables R_k as $R_k = 1$ if $\mathbf{X}_k$ is observed and 0 otherwise for $k = 1, \cdots, q$, and $R_0 = 1$ if $R_1 = \cdots = R_q = 1$ and 0 otherwise. Let N be the total number of individuals in the sample. For $i = 1, \cdots, N$ we define the probabilities of observation to be

$$\pi_{ik} = pr(R_{ik} = 1|Y_i, \mathbf{X}_i) \text{ for } k = 1, \cdots, q, \text{ and } \pi_{i0} = pr(R_{i0} = 1|Y_i, \mathbf{X}_i),$$

where $\pi_{ik} \geq \pi_{i0}$. Throughout we suppose that the selection probabilities are specified values, and that $\pi_{i0} > C > 0$ with probability 1. If R_k, $k = 1, \cdots, q$, can be ordered such that $R_{(1)} \geq R_{(2)} \geq \cdots \geq R_{(q)}$ then the missingness pattern is monotone, otherwise the pattern is nonmonotone. In a PQD, the response variable Y and sometimes certain covariates, without loss of generality say $\mathbf{X}_1$, are available for all the subjects in the study (Chatterjee and Li 2010). Using our notation, we consider that in the PQD the variables are divided into q subsets, $(Y, \mathbf{X}_1^T)^T, (\mathbf{X}_k), k = 2, \cdots, q$, where data on $(Y, \mathbf{X}_1^T)^T$ are fully observed. Then according to the selection probabilities π_{ik} the subjects in the study are selected into (overlapping) groups, G_k, $k = 2, \cdots, q$, based on the fully observed variables $(Y, \mathbf{X}_1)$. That is, the selection probabilities π_{ik} and π_{i0} depend on Y and $\mathbf{X}$ only through $(Y, \mathbf{X}_1)$. Here $R_{ik} = 1$ indicates the ith subject is selected into the group G_k, and $\pi_{ik} = pr(R_{ik} = 1|Y_i, \mathbf{X}_{i1})$, $\pi_{i0} = pr(R_{i0} = 1|Y_i, \mathbf{X}_{i1})$, and the missing covariate data are MAR. In some studies, the missing data probabilities are constants and can be denoted as π_k, and π_0. In this case, the subjects are completely randomly selected into groups and this does not depend on $(Y, \mathbf{X}_1)$, so the missing data are MCAR.

Let $V_0 = \{i : R_{i0} = 1\}$ and $V_k = \{i : R_{ik} = 1\}$, $k = 1, \cdots, q$ be the index set of complete observations and the index set of completely observed $\mathbf{X}_k$ respectively, and let n, n_k be the corresponding number of observations in each set. We see that $n \leq n_k$ and we require $n > 0$. To be complete we denote the complement of V_0 as $\bar{V}_0$. We assume that $(Y_i, \mathbf{X}_{i1}^T, \cdots, \mathbf{X}_{iq}^T, R_{i1}, \cdots, R_{iq})$, $i \in 1, \cdots, N$, are independent and identically distributed.

Next we describe a generalized unified extimation method for MCAR data and MAR data separately.

2.1 MCAR Data

For $k = 1, \cdots, q$, let the parametric function $f_k(\mathbf{X}_k; \gamma_k)$ denote the conditional mean of Y given $\mathbf{X}_k$, where γ_k is a vector of parameters. We call $f_k(\mathbf{X}_k; \gamma_k)$, $k = 1, \cdots, q$, the working regression models or surrogate models and $\gamma = (\gamma_1^T, \cdots, \gamma_q^T)^T$ a vector of surrogate parameters. For convenience we denote the model of interest $f(\mathbf{X}; \beta)$ as $f_0(\mathbf{X}_0; \beta)$ with $\mathbf{X}_0 = \mathbf{X}$.

Assume that $\hat{\beta}$ and $\hat{\gamma}_k$, $k = 1, \cdots, q$, solve the system of estimating equations for β and γ_k given in (2.1) and (2.2) respectively:

$$\sum_{i \in V_0} \mathbf{S}_{i0}(\beta) = \sum_{i \in V_0} \mathbf{w}_0(\mathbf{X}_{i0}) \{Y_i - f_0(\mathbf{X}_{i0}; \beta)\} = 0, \tag{2.1}$$

$$\sum_{i \in V_0} \mathbf{S}_{ik}(\gamma_k) = \sum_{i \in V_0} \mathbf{w}_k(\mathbf{X}_{ik}) \{Y_i - f_k(\mathbf{X}_{ik}; \gamma_k)\} = 0, \text{ for } k = 1, \cdots, q, \tag{2.2}$$

where $\mathbf{w}_0(\mathbf{X}_{i0})$ and $\mathbf{w}_k(\mathbf{X}_{ik})$, using notation similar to that in Chen and Chen (2000), are vectors corresponding to known functions of $\mathbf{X}_{i0}$ and $\mathbf{X}_{ik}$. As a special case, $\mathbf{S}_{i0}(\beta)$ and $\mathbf{S}_{ik}(\gamma_k)$ could be score functions based on some set of distributions. For example, in the case of linear and logistic regression models we could use least squares estimating equations and logistic regression estimating equations respectively. We denote $\mathbf{S}_i(\theta) = (\mathbf{S}_{i0}^T(\beta), \mathbf{S}_{iQ}^T(\gamma))^T$ with $\theta = (\beta^T, \gamma^T)^T$ and $\mathbf{S}_{iQ}(\gamma) = (\mathbf{S}_{i1}^T(\gamma_1), \cdots, \mathbf{S}_{iq}^T(\gamma_q))^T$.

Following Chen and Chen (2000) and Foutz (1977) under regularity conditions we can show that (i) $\hat{\theta} = (\hat{\beta}^T, \hat{\gamma}^T)^T$, with $\hat{\gamma} = (\hat{\gamma}_1^T, \cdots, \hat{\gamma}_q^T)^T$, is consistent for some vector $\theta^* = (\beta^{*T}, \gamma^{*T})^T$; and (ii) $n^{1/2}(\hat{\theta} - \theta^*)$ is asymptotically normal with mean 0 and variance

$D^{-1}CD^{T-1}$ with $D = E\{\partial \mathbf{S}_i(\theta^*)/\partial\theta\}$ and $C = E\{\mathbf{S}_i(\theta^*)\mathbf{S}_i^T(\theta^*)\}$.

We rewrite D as $diag(D_0, D_1)$ with

$$D_0 = E\{\partial \mathbf{S}_{i0}(\beta^*)/\partial\beta\}$$

and

$$D_1 = E\{\partial \mathbf{S}_{iQ}(\gamma^*)/\partial\gamma\}$$

We partition the matrix C as

$$C = \begin{pmatrix} C_{00} & C_{01} \\ C_{01}^T & C_{11} \end{pmatrix},$$

where

$$C_{00} = E\{\mathbf{S}_{i0}(\beta^*)\mathbf{S}_{i0}^T(\beta^*)\},$$

$$C_{01} = E\{\mathbf{S}_{i0}(\beta^*)\mathbf{S}_{iQ}^T(\gamma^*)\},$$

and

$$C_{11} = E\{\mathbf{S}_{iQ}(\gamma^*)\mathbf{S}_{iQ}^T(\gamma^*)\}.$$

According to multivariate normal distribution theory, the conditional distribution of $n^{1/2}(\hat{\beta} - \beta^*)$, given $n^{1/2}(\hat{\gamma} - \gamma^*)$, is asymptotic normal with mean $n^{1/2}D_0^{-1}C_{01}C_{11}^{-1}D_1(\hat{\gamma} - \gamma^*)$, which suggests that the CC estimator $\hat{\beta}$ may be improved by using

$$\bar{\beta} = \hat{\beta} - \hat{D}_0^{-1}\hat{C}_{01}\hat{C}_{11}^{-1}\hat{D}_1(\hat{\gamma} - \bar{\gamma}), \tag{2.3}$$

where

$$\hat{D}_0 = n^{-1}\sum_{i \in V_0}\{\partial \mathbf{S}_{i0}(\hat{\beta})/\partial\beta\},$$

$$\hat{C}_{01} = n^{-1} \sum_{i \in V_0} \{\mathbf{S}_{i0}(\hat{\beta})\mathbf{S}_{iQ}^T(\hat{\gamma})\},$$
$$\hat{C}_{11} = n^{-1} \sum_{i \in V_0} \{\mathbf{S}_{iQ}(\hat{\gamma})\mathbf{S}_{iQ}^T(\hat{\gamma})\},$$
$$\hat{D}_1 = n^{-1} \sum_{i \in V_0} \{\partial \mathbf{S}_{iQ}(\hat{\gamma})/\partial \gamma\},$$

and $\bar{\gamma} = (\bar{\gamma}_1^T, \cdots, \bar{\gamma}_q^T)^T$. Here, $\bar{\gamma}_k$ is an estimate of γ_k^* based on the observations in V_k, that is, $\bar{\gamma}_k$ solves

$$\sum_{i \in V_k} \mathbf{S}_{ik}(\gamma_k) = \sum_{i \in V_k} \mathbf{w}_k(\mathbf{X}_{ik})\{Y_i - f_k(\mathbf{X}_{ik}; \gamma_k)\} = 0,$$

which allows all the observations in V_k to be used in the estimation. We call $\bar{\beta}$ an improved complete-case (ICC) estimator. We expect that the ICC estimator produces efficiency gains when $\hat{\beta}$ and $\hat{\gamma}$ are highly correlated and the sizes of the V_k's are much larger than the size of V_0.

Under the regularity conditions, $\bar{\beta}$ is consistent for β^*, which is the true value of β in the model $f_0(\mathbf{X}_0; \beta)$ when $f_0(\mathbf{X}_0; \beta)$ is correctly specified. The consistency for β^* does not depend on the correctness of the working regression models $f_k(\mathbf{X}_k; \gamma_k)$. In addition $n^{1/2}(\bar{\beta} - \beta^*)$ is asymptotic normal with mean 0 and variance given by

$$D_0^{-1} C_{00} D_0^{T-1} - D_0^{-1} C_{01}(I - C_{11}^{-1} C_{\rho 11}) C_{11}^{-1} C_{01}^T D_0^{T-1}, \tag{2.4}$$

where $C_{\rho_{11}}$ is C_{11} with its khth element c_{kh} replaced by $c_{\rho kh} = (\pi_0 \pi_{kh}/\pi_k \pi_h) c_{kh}$ and $\pi_{kh} = pr(R_k = R_h = 1)$ for $k, h = 1, \cdots, q$. The first term in (2.4) is the variance of $n^{1/2}(\hat{\beta} - \beta^*)$, and the second term represents the improvement of the ICC estimator over the CC estimator. The asymptotic variance in (2.4) can be estimated by

$$\hat{D}_0^{-1} \hat{C}_{00} \hat{D}_0^{T-1} - \hat{D}_0^{-1} \hat{C}_{01}(I - \hat{C}_{11}^{-1} \hat{C}_{\rho 11}) \hat{C}_{11}^{-1} \hat{C}_{01}^T \hat{D}_0^{T-1},$$

where

$$\hat{C}_{00} = n^{-1} \sum_{i \in V_0} \{\mathbf{S}_{i0}(\hat{\beta}) \mathbf{S}_{i0}^T(\hat{\beta})\}$$

and $\hat{C}_{\rho_{11}}$ has khth element $\hat{c}_{\rho kh} = (nn_{kh}/n_k n_h)\hat{c}_{kh}$ and n_{kh} is the total number of observations with $R_{ik} = R_{ih} = 1$ for $k, h = 1, \cdots, q$. A proof and references are given in the Appendix B.

We know that a regular CC analysis for any regression model provides consistent estimates as long as the missing data probability does not depend on the response variable, given the covariates in the model. Therefore, the above method can also be applied in the special MAR case where the missingness does not depend on Y, that is, $\pi_{ik} = pr(R_{ik} = 1|\mathbf{X}_{i1})$ and $\pi_{i0} = pr(R_{i0} = 1|\mathbf{X}_{i1})$. In this special MAR case, to obtain a consistent $\bar{\beta}$ we need to add the fully observed X_1 as covariates in each working regression model so that both $\hat{\gamma}$ and $\bar{\gamma}$ can be consistent for γ^*. We note that the unified estimator of Chen and Chen (2000) is a special case of the generalized unified estimator $\bar{\beta}$ when the covariates follow a simple monotone missing pattern.

2.2 MAR Data with Known Missing Data Probability

In a PQD, it is common that the missingness depends on both the response Y and the fully observed covariates X_1. In this case the ICC estimator will be biased. In this section we extend the generalized unified method of Section 2.1 to deal with general MAR data using inverse probability weighted estimation equations (Horvitz and Thompson 1952).

Assume that $\hat{\beta}_\pi$, $\hat{\gamma}_\pi = (\hat{\gamma}_{\pi 1}^T, \cdots, \hat{\gamma}_{\pi q}^T)^T$, and $\bar{\gamma}_\pi = (\bar{\gamma}_{\pi 1}^T, \cdots, \bar{\gamma}_{\pi q}^T)^T$ solve the system of weighted estimation equations given in (2.5), (2.6), and (2.7) respectively:

$$\sum_{i=1}^{N} \frac{R_{i0}}{\pi_{i0}} \mathbf{S}_{i0}(\beta) = 0, \tag{2.5}$$

$$\sum_{i=1}^{N} \frac{R_{i0}}{\pi_{i0}} \mathbf{S}_{ik}(\gamma_k) = 0, \text{ for } k = 1, \cdots, q, \tag{2.6}$$

$$\sum_{i=1}^{N} \frac{R_{ik}}{\pi_{ik}} \mathbf{S}_{ik}(\gamma_k) = 0, \text{ for } k = 1, \cdots, q. \tag{2.7}$$

We note that $\hat{\beta}_\pi$ and $\hat{\gamma}_\pi$ are computed based on the complete observations in V_0, while $\bar{\gamma}_\pi$ is computed based on the larger data sets V_k, $k = 1, \cdots, q$. Following a similar development to that in Section 2.1, under regularity conditions we obtain the following results:

(i) $N^{1/2}(\hat{\beta}_\pi - \beta^*)$ given $N^{1/2}(\hat{\gamma}_\pi - \gamma^*)$ is asymptotic normal with mean

$$N^{1/2} D_0^{-1} C_{\pi 01} C_{\pi 11}^{-1} D_1 (\hat{\gamma}_\pi - \gamma^*),$$

where

$$C_{\pi 01} = E[(R_{i0}/\pi_{i0}^2)\mathbf{S}_{i0}(\beta^*)\mathbf{S}_{iQ}^T(\gamma^*)]$$

and

$$C_{\pi 11} = E[(R_{i0}/\pi_{i0}^2)\mathbf{S}_{iQ}(\gamma^*)\mathbf{S}_{iQ}^T(\gamma^*)].$$

(ii) β^* can be consistently estimated by

$$\bar{\beta}_\pi = \hat{\beta}_\pi - \hat{D}_{\pi 0}^{-1} \hat{C}_{\pi 01} \hat{C}_{\pi 11}^{-1} \hat{D}_{\pi 1} (\hat{\gamma}_\pi - \bar{\gamma}_\pi), \tag{2.8}$$

where

$$\hat{D}_{\pi 0} = N^{-1} \sum_{i=1}^{N} (R_{i0}/\pi_{i0}) \partial \mathbf{S}_{i0}(\hat{\beta}_\pi)/\partial \beta,$$

$$\hat{C}_{\pi 01} = N^{-1} \sum_{i=1}^{N} (R_{i0}/\pi_{i0}^2) \mathbf{S}_{i0}(\hat{\beta}_\pi) \mathbf{S}_{iQ}^T(\hat{\gamma}_\pi),$$

$$\hat{C}_{\pi 11} = N^{-1}\sum_{i=1}^{N}(R_{i0}/\pi_{0i}^2)\mathbf{S}_{iQ}(\hat{\gamma}_\pi)\mathbf{S}_{iQ}^T(\hat{\gamma}_\pi),$$

and

$$\hat{D}_{\pi 1} = N^{-1}\sum_{i=1}^{N}(R_{0i}/\pi_{0i})\partial\mathbf{S}_{Qi}(\hat{\gamma}_\pi)/\partial\gamma.$$

The consistency of $\bar{\beta}_\pi$ does not depend on the correctness of the working regression models. We call $\bar{\beta}_\pi$ an improved weighted complete-case (IWCC) estimator.

(iii) $N^{1/2}(\bar{\beta}_\pi - \beta^*)$ is asymptotic normal with mean 0; its variance can be estimated by

$$\begin{aligned}&\hat{D}_{\pi 0}^{-1}\hat{C}_{\pi 00}\hat{D}_{\pi 0}^{T-1} - \hat{D}_{\pi 0}^{-1}[\hat{C}_{\pi 01}\hat{C}_{\pi 11}^{-1}\{(\hat{C}_{\pi 11} - \hat{C}_{\pi 22} + \hat{C}_{\pi 12}^T + \hat{C}_{\pi 12})\hat{C}_{\pi 11}^{-1}\hat{C}_{\pi 01}^T - \hat{C}_{\pi 02}^T\} \\ &- \hat{C}_{\pi 02}\hat{C}_{\pi 11}^{-1}\hat{C}_{\pi 01}^T]\hat{D}_{\pi 0}^{T-1},\end{aligned} \tag{2.9}$$

where

$$\hat{C}_{\pi 00} = N^{-1}\sum_{i=1}^{N}(R_{i0}/\pi_{i0}^2)\mathbf{S}_{0i}(\hat{\beta}_\pi)\mathbf{S}_{i0}^T(\hat{\beta}_\pi),$$

$$\hat{C}_{\pi 22} = N^{-1}\sum_{i=1}^{N}\mathbf{S}_{\pi iQ}(\hat{\gamma}_\pi)\mathbf{S}_{\pi iQ}^T(\hat{\gamma}_\pi),$$

$$\hat{C}_{\pi 12} = N^{-1}\sum_{i=1}^{N}(R_{i0}/\pi_{0i})\mathbf{S}_{iQ}(\hat{\gamma}_\pi)\mathbf{S}_{\pi iQ}^T(\hat{\gamma}_\pi),$$

and

$$\hat{C}_{\pi 02} = N^{-1}\sum_{i=1}^{N}(R_{0i}/\pi_{i0})\mathbf{S}_{i0}(\hat{\beta}_\pi)\mathbf{S}_{\pi iQ}^T(\hat{\gamma}_\pi)$$

with

$$\mathbf{S}_{\pi iQ}(\gamma) = ((R_{i1}/\pi_{i1})\mathbf{S}_{i1}^T(\gamma_1), \cdots, (R_{iq}/\pi_{iq})\mathbf{S}_{iq}^T(\gamma_q))^T.$$

2.3 MAR Data with Estimated Missing Data Probability

In practice, the true missing data probabilities are often unknown when data are MAR. Even if the missing probability is known, the estimation efficiency of the IWCC can be further improved by using estimated missing data probabilities $\hat{\pi}_{ij}$ instead of the true known missing data probabilities (Robins et al. 1994; Lawless et al. 1999; Chatterjee and Breslow 2003; Breslow et al. 2009).

Let $\sum_{i=1}^{N} H_{\pi ik}(\alpha_k)$ be an estimating function for the selection probability π_{ik}, $k = 0, \cdots, q$, which can be correctly specified when data are missing by design. Here α_k, $k = 0, \cdots, q$, are vectors of parameters. We denote the estimated selection probabilities as $\hat{\pi}_{ik} = \pi_{ik}(\hat{\alpha}_k)$.

Let $\hat{\beta}_{\hat{\pi}}$, $\hat{\gamma}_{\hat{\pi}}$, and $\bar{\gamma}_{\hat{\pi}}$ denote the corresponding estimators using estimated selection probabilities. Following Robins et al. (1994) we define

$$Res\{A_i(\beta,\alpha), B_i(\alpha)\} = A_i(\beta,\alpha) - E[\frac{\partial A_i(\beta,\alpha)}{\partial\alpha}]E[\frac{\partial B_i(\alpha)}{\partial\alpha}]^{-1}B_i(\alpha),$$

and

$$\hat{R}es\{A_i(\beta,\alpha), B_i(\alpha)\} = A_i(\beta,\alpha) - \{N^{-1}\sum_i^N \frac{\partial A_i(\beta,\alpha)}{\partial\alpha}\}\{N^{-1}\sum_i^N \frac{\partial B_i(\alpha)}{\partial\alpha}\}^{-1}B_i(\alpha),$$

and denote

$$\begin{aligned}
Res_{i0}(\beta,\alpha_0) &= Res\{\frac{R_{i0}}{\pi_{i0}(\alpha_0)}S_{i0}(\beta), H_{\pi i0}(\alpha_0)\},\\
Res_{i1}(\gamma,\alpha_0) &= Res\{\frac{R_{i0}}{\pi_{i0}(\alpha_0)}S_{iQ}(\gamma), H_{\pi i0}(\alpha_0)\},\\
Res_{i2}(\gamma,\alpha_Q) &= Res\{(\frac{R_{i1}}{\pi_{i1}(\alpha_1)}S_{i1}^T(\gamma_1), \cdots, \frac{R_{iq}}{\pi_{iq}(\alpha_q)}S_{iq}^T(\gamma_q))^T, H_{\pi iQ}(\alpha_Q)\},\\
\hat{R}es_{i0}(\beta,\alpha_0) &= \hat{R}es\{\frac{R_{i0}}{\pi_{i0}(\alpha_0)}S_{i0}(\beta), H_{\pi i0}(\alpha_0)\},\\
\hat{R}es_{i1}(\gamma,\alpha_0) &= \hat{R}es\{\frac{R_{i0}}{\pi_{i0}(\alpha_0)}S_{iQ}(\gamma), H_{\pi i0}(\alpha_0)\},\\
\hat{R}es_{i2}(\gamma,\alpha_Q) &= \hat{R}es\{(\frac{R_{i1}}{\pi_{i1}(\alpha_1)}S_{i1}^T(\gamma_1), \cdots, \frac{R_{iq}}{\pi_{iq}(\alpha_q)}S_{iq}^T(\gamma_q))^T, H_{\pi iQ}(\alpha_Q)\},
\end{aligned}$$

where $\alpha_Q = (\alpha_1^T, \cdots, \alpha_q^T)^T$ and $H_{\pi iQ}(\alpha_Q) = (H_{\pi i1}^T(\alpha_1), \cdots, H_{\pi iq}^T(\alpha_q))^T$.

The IWCC using the estimated selection probabilities $\hat{\pi}_{ji}$'s can be written as

$$\bar{\beta}_{\hat{\pi}} = \hat{\beta}_{\hat{\pi}} - \hat{D}_{\hat{\pi}0}^{-1}\hat{C}_{\hat{\pi}01}\hat{C}_{\hat{\pi}11}^{-1}\hat{D}_{\hat{\pi}1}(\hat{\gamma}_{\hat{\pi}} - \bar{\gamma}_{\hat{\pi}}), \tag{2.10}$$

where

$$\begin{aligned}
\hat{D}_{\hat{\pi}0} &= N^{-1}\sum_{i=1}^{N}(R_{i0}/\hat{\pi}_{i0})\partial\mathbf{S}_{i0}(\hat{\beta}_{\hat{\pi}})/\partial\beta, \\
\hat{C}_{\hat{\pi}01} &= N^{-1}\sum_{i=1}^{N}\hat{R}es_{0i}(\hat{\beta}_{\hat{\pi}}, \hat{\alpha}_0)\hat{R}es_{1i}^T(\hat{\gamma}_{\hat{\pi}}, \hat{\alpha}_0), \\
\hat{C}_{\hat{\pi}11} &= N^{-1}\sum_{i=1}^{N}\hat{R}es_{i1}(\hat{\gamma}_{\hat{\pi}}, \hat{\alpha}_0)\hat{R}es_{1i}^T(\hat{\gamma}_{\hat{\pi}}, \hat{\alpha}_0),
\end{aligned}$$

and

$$\hat{D}_{\hat{\pi}1} = N^{-1}\sum_{i=1}^{N}(R_{i0}/\hat{\pi}_{i0})\partial\mathbf{S}_{iQ}(\hat{\gamma}_{\hat{\pi}})/\partial\gamma.$$

The asymptotic variance of $\bar{\beta}_{\hat{\pi}}$ can be given by

$$D_0^{-1}C_{\hat{\pi}00}D_0^{T-1} - \hat{D}_0^{-1}[C_{\hat{\pi}01}C_{\hat{\pi}11}^{-1}\{(C_{\hat{\pi}11} - C_{\hat{\pi}22} + C_{\hat{\pi}12}^T + C_{\hat{\pi}12})C_{\hat{\pi}11}^{-1}C_{\hat{\pi}01}^T - C_{\hat{\pi}02}^T\}$$
$$- C_{\hat{\pi}02}C_{\hat{\pi}11}^{-1}\hat{C}_{\hat{\pi}01}^T]D_0^{T-1}, \tag{2.11}$$

where

$$C_{\hat{\pi}00} = E[Res_{i0}(\beta^*, \alpha_0^*)Res_{i0}^T(\beta^*, \alpha_0^*)],$$

$$C_{\hat{\pi}01} = E[Res_{i0}(\beta^*, \alpha_0^*)Res_{i1}^T(\beta^*, \alpha_0^*)],$$

$$C_{\hat{\pi}02} = E[Res_{i0}(\beta^*, \alpha_0^*)Res_{i2}^T(\beta^*, \alpha_Q^*)],$$

$$C_{\hat{\pi}11} = E[Res_{i1}(\beta^*, \alpha_0^*)Res_{i1}^T(\beta^*, \alpha_0^*)],$$

$$C_{\hat{\pi}12} = E[Res_{i1}(\beta^*, \alpha_0^*)Res_{i2}^T(\beta^*, \alpha_Q^*)],$$

$$C_{\hat{\pi}22} = E[Res_{i2}(\beta^*, \alpha_Q^*)Res_{i2}^T(\beta^*, \alpha_Q^*)].$$

The asymptotic variance in (2.11) be estimated by

$$\hat{D}_{\hat{\pi}0}^{-1}\hat{C}_{\hat{\pi}00}\hat{D}_{\hat{\pi}0}^{T-1} - \hat{D}_{\hat{\pi}0}^{-1}[\hat{C}_{\hat{\pi}01}\hat{C}_{\hat{\pi}11}^{-1}\{(\hat{C}_{\hat{\pi}11} - \hat{C}_{\hat{\pi}22} + \hat{C}_{\hat{\pi}12}^T + \hat{C}_{\hat{\pi}12})\hat{C}_{\hat{\pi}11}^{-1}\hat{C}_{\hat{\pi}01}^T - \hat{C}_{\hat{\pi}02}^T\}$$
$$- \hat{C}_{\hat{\pi}02}\hat{C}_{\hat{\pi}11}^{-1}\hat{C}_{\hat{\pi}01}^T]\hat{D}_{\hat{\pi}0}^{T-1}, \tag{2.12}$$

where

$$\hat{C}_{\hat{\pi}00} = N^{-1}\sum_{i=1}^{N}\hat{R}es_{i0}(\hat{\beta}_{\hat{\pi}}, \hat{\alpha}_0)\hat{R}es_{i0}^T(\hat{\beta}_{\hat{\pi}}, \hat{\alpha}_0),$$
$$\hat{C}_{\hat{\pi}22} = N^{-1}\sum_{i=1}^{N}\hat{R}es_{i2}(\hat{\gamma}_{\hat{\pi}}, \hat{\alpha}_Q)\hat{R}es_{i2}^T(\hat{\gamma}_{\hat{\pi}}, \hat{\alpha}_Q),$$
$$\hat{C}_{\hat{\pi}12} = N^{-1}\sum_{i=1}^{N}\hat{R}es_{i1}(\hat{\gamma}_{\hat{\pi}}, \hat{\alpha}_0)\hat{R}es_{i2}^T(\hat{\gamma}_{\hat{\pi}}, \hat{\alpha}_Q),$$

and

$$\hat{C}_{\hat{\pi}02} = N^{-1}\sum_{i=1}^{N}\hat{R}es_{i0}(\hat{\beta}_{\hat{\pi}}, \hat{\alpha}_0)\hat{R}es_{i2}^T(\hat{\gamma}_{\hat{\pi}}, \hat{\alpha}_Q).$$

A proof and references are provided in the Appendix B. As in Section 2.1, we see that the first term in (2.9) or (2.12) is an estimate of the asymptotic variance of $N^{1/2}(\hat{\beta}_\pi - \beta^*)$ or $N^{1/2}(\hat{\beta}_{\hat{\pi}} - \beta^*)$, and the second term represents the improvement of the IWCC estimator over the weighted CC estimator using know or estimated π_{ik}'s.

We note that in many studies auxiliary variables are used to increase estimation efficiency (Robins et al.1994; Reilly and Pepe 1995). Both the ICC and the IWCC can deal with the case where auxiliary covariates $\tilde{\mathbf{X}}$ are observed. In this case we reorder $(\mathbf{X}^T, \tilde{\mathbf{X}}^T)^T$

as $(\mathbf{X}_1^T, \cdots, \mathbf{X}_q^T)^T$, and the same procedure can be applied to compute an ICC or IWCC estimate of β.

2.4 Simulation Studies

In this section we use simulation studies to examine the finite sample performance of the ICC and IWCC estimators. We consider a linear regression model $Y = \beta_0+\beta_1 X_1+\beta_2 X_2+\beta_3 X_3 + \epsilon$ and a logistic regression model $logit\{P(Y = 1|X_1, X_2, X_3)\} = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3$, where X_2 is generated from the exponential distribution with mean 1, and X_1, X_3 and ϵ are generated independently from the standard normal distribution. Following a PQD, we assume that $\{Y, X_1\}$ are fully observed but both X_2 and X_3 have missing values, and we consider both the MCAR and MAR cases. We assume that each subject is selected into group G_2 and G_3 with probability π_{i2} and π_{i3} respectively. Then values of X_2 and X_3 are observed for the subjects in G_2 and G_3 respectively. In the MCAR case $\pi_{i2} = \pi_2$ and $\pi_{i3} = \pi_3$. For the MAR case we let the selection probabilities depend on the fully observed response Y and covariates X_1 such that $logit\{\pi_{i2}\} = \alpha_{20} + \alpha_{21} X_{i1} + \alpha_{22} Y_i$ and $logit\{\pi_{i3}\} = \alpha_{30} + \alpha_{31} X_{i1} + \alpha_{32} Y_i$.

We set the sample size $N = 1000$ and for each setting we generate 1000 data sets. We let $\beta^* = (0.1, 1, 1, 1)^T$ in the linear model and $\beta^* = (-1.2, 1, 1, 1)^T$ in the logit model. For the MCAR case we set $\pi_2 = \pi_3 = 0.50$. For the MAR case we let $(\alpha_{20}, \alpha_{21}, \alpha_{22})^T = (\alpha_{30}, \alpha_{31}, \alpha_{32})^T = (0.2, 0.2, 0.2)^T$. Here the number of distinct missing patterns $q = 3$. The number of observations in set V_0, V_j, $j = 1, 2, 3$ is approximately 250, 1000, 500

Table 2.1 Simulation Result

	Linear Model $\beta = (0.1, 1, 1, 1)^T$				Logit Model $\beta = (-1.2, 1, 1, 1)^T$			
	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$
(1) MCAR:								
ICC estimation								
Bias	-0.002	0.0^45^a	-0.0^31	0.003	-0.022	0.019	0.021	0.029
$s.d.^b$	0.079	0.054	0.058	0.057	0.198	0.132	0.175	0.157
$s.e.^c$	0.077	0.054	0.056	0.057	0.192	0.133	0.168	0.154
MSE	0.006	0.003	0.003	0.003	0.040	0.018	0.031	0.025
95%CP	95.0%	94.8%	94.3%	94.8%	94.0%	95.4%	94.1%	94.3%
ARE^d	1.269	1.361	1.180	1.183	1.620	2.205	1.481	1.496
CC estimation								
Bias	-0.004	-0.0^34	-0.0^37	0.003	-0.037	0.027	0.029	0.037
$s.d.$	0.089	0.063	0.063	0.062	0.252	0.196	0.213	0.192
MSE	0.008	0.004	0.004	0.004	0.065	0.039	0.046	0.038
(2) MAR:								
IWCC estimation using estimated π_{ji}'s								
Bias	-0.005	-0.004	0.0^44	-0.0^37	-0.031	0.014	0.026	0.016
$s.d.$	0.072	0.051	0.047	0.052	0.170	0.121	0.157	0.139
$s.e.$	0.067	0.048	0.046	0.051	0.171	0.121	0.153	0.137
MSE	0.005	0.003	0.002	0.003	0.030	0.015	0.025	0.020
95%CP	93.2%	92.6%	94.3%	94.2%	95.7%	95.4%	95.3%	94.7%
ARE	1.494	1.526	1.272	1.331	1.690	1.749	1.404	1.426
IWCC estimation using known π_{ji}'s								
Bias	-0.003	-0.004	0.0^56	0.0^48	-0.032	0.014	0.028	0.016
$s.d.$	0.074	0.052	0.048	0.054	0.180	0.122	0.158	0.140
$s.e.$	0.070	0.048	0.047	0.052	0.179	0.122	0.156	0.139
MSE	0.005	0.003	0.002	0.003	0.033	0.015	0.026	0.020
95%CP	94.0%	92.3%	94.3%	94.1%	95.6%	95.5%	94.6%	94.8%
ARE	1.414	1.468	1.219	1.235	1.507	1.720	1.386	1.406
Weighted CC estimation using known π_{ji}'s								
Bias	0.001	-0.005	0.0^33	-0.002	-0.040	0.014	0.036	0.022
$s.d.$	0.088	0.063	0.053	0.060	0.221	0.160	0.186	0.166
MSE	0.008	0.004	0.003	0.004	0.050	0.026	0.036	0.028

$^a0.0^45 = 0.00005$.

$^bs.d.$ is the empirical standard deviation.

$^cs.e.$ is the simulation mean of the asymptotic standard errors.

$^dARE = (s.d.(\hat{\beta})/s.d.(\bar{\beta}))^2$.

and 500 respectively in the MCAR case, 370, 1000, 600 and 600 respectively in the linear model and 330, 1000, 570 and 570 in the logit model in the MAR case. We use linear regression models and logistic regression models as the working regression models for the linear regression model and the logistic regression model respectively. Logistic regression models are used to estimate the selection probabilities in the MCAR case.

We let $\mathbf{X}_0 = (X_1, X_2, X_3)^T$. The model $f_0(\mathbf{X}_0; \beta)$ is of interest. We note that in the logistic regression case, if $f_0(\mathbf{X}_0; \beta)$ is "correct" then logistic models for Y given X_1, for Y given X_2, and for Y given X_3 are misspecified, but still useful for increasing efficiency.

The simulation results for the ICC and IWCC estimates together with the CC and weighted CC estimates are listed in Table 2.1. We see that (i) the biases of the ICC and IWCC estimates are small; (ii) the means of the standard errors ($s.e.$) based on the asymptotic variance estimator are close to the empirical standard deviations ($s.d.$); (iii) the estimated 95% coverage probabilities are close to the nominal level; and (iv) comparing to the (weighted) CC analysis both the ICC and IWCC estimates have smaller mean square errors (MSE) and empirical standard deviations; (v) comparing to the IWCC estimates using known selection probability the corresponding IWCC estimates using estimated selection probability are slightly more efficient.

Chapter 3

Generalized Unified Approach to Longitudinal Data Analysis

3.1 A Brief Review of Generalized Estimating Equation

Longitudinal data frequently occurs in medical and social studies. In longitudinal study measurements from the same individuals are taken repeatedly through time. A primary goal of longitudinal data analysis lies in characterizing the change in responses over time as well as factors that influence the change.

In the past a few decades, statistical methods for the analysis of longitudinal data have been developed tremendously. One of the popular methods is the generalized estimating equations (GEE) approach proposed by Liang and Zeger (1986). The GEE approach does not require a complete probability model of the response vector, and it only needs the first two moments of the response vector. Liang and Zeger (1986) showed that the consistency of the estimates for regression parameters only depends on the correctness of the mean model, but does not depend on the correctness of the "working" correlation structure of the response vector.

Missing data is a common problem in the longitudinal studies. Andrea et al. (1998) described the maximum likelihood method for non-ignorable and nonmonotone missing data problems, but it encounters a difficult numerical problem; Chen et al. (2008) provided a careful investigation of likelihood methods for missing response and covariate data via the EM algorithm. Alternatively, when data are MCAR, GEE approach yields consistent estimates for the regression parameters (Liang and Zeger 1986). When data are MAR, Robins et al. (1994), Robins et al. (1995) and Schaarfstein et al. (1999) proposed methods to improve the efficiency of the inverse probability weighted generalized estimating equations (IPWGEE). The idea is that adding a zero mean function to the estimating equation to maintain unbiasness, and at the same time to extract the remainder information from the incomplete observations to improve estimation efficiency.

3.2 Notation

Let $Y_i = (y_{i1}, \cdots, y_{ij}, \cdots, y_{iJ})^T$ be a response vector of subject i at time points $t = (t_1, t_2, \cdots, t_J)^T$ and $x_{ij} = (x_{ij1}, \cdots, x_{ijk}, \cdots, x_{ijp})^T$ be the $p \times 1$ covariates vector recorded for subject i at the jth time point, $j = 1, \cdots, J$, $i = 1, \cdots, M$. Let X_i be the $J \times p$ matrix $(x_{i1}, \cdots, x_{iJ})^T$. Here i, j and k is the index of subject, observation and covariate respectively. Let $\mu_{ij} = E(y_{ij}|X_i)$, and $\mu_i = (\mu_{i1}, \cdots, \mu_{ij}, \cdots, \mu_{iJ})^T$. Suppose that the mean structure of y_{ij} depends on the covariate vector of subject i at time j, i.e., $E(y_{ij}|X_i) = E(y_{ij}|x_{ij})$ (Pepe and Anderson 1994 and Robins et al. 1999), we are interested in estimating parameter β in the generalized linear regression models

$$g(\mu_{ij}) = x_{ij}^T\beta, j = 1, \cdots, J,$$

where $g(\centerdot)$ is a monotone differentiable link function.

Let us briefly review the generalized estimating equation and its application to regression analysis. To simplify the introduction, we consider a regression model without missing values. We suppose that $\hat{\beta}^f$ is the solution to the generalized estimating equation in (Liang and Zeger 1986)

$$U^f(\hat{\beta}) = \sum_i U_i^f = 0, \tag{3.1}$$

where the summation is over all M independent subjects and

$$U^f(i) = D_i^T V_i^{-1}(Y_i - \mu_i).$$

Here the super-script f denotes the full data, $D_i = \partial\mu_i/\partial\beta$, and V_i is the covariance matrix for the response Y_i. In actual implementation, a working covariance matrix is used to replace V_i, which is often decomposed as

$$V_i = a(\phi)A_i^{1/2}R_i(\rho)A_i^{1/2},$$

where $a(\centerdot)$ is a known function, ϕ is a scaled parameter, A_i is a $J \times J$ diagonal matrix with elements $v_{ij} = Var(y_{ij})$, $R_i(\rho)$ is a $J \times J$ working correlation matrix that is fully specified up to a vector of parameters ρ.

Under mild regularity conditions, the estimate $\hat{\beta}^f$ from the generalized estimating equation (3.1) converges to its true value β^* in probability. Moreover, by the Central Limit Theory, $M^{1/2}(\hat{\beta} - \beta^*)$ has an asymptotic normal distribution with mean 0 and covariance

$$[E\{\partial U_i^f(\beta^*)/\partial\beta\}]^{-1}E(U_i^f(\beta^*)U_i^{fT}(\beta^*))[E\{\partial U_i^{fT}(\theta^*)/\partial\theta\}]^{-1},$$

which can be consistently estimated by

$$M(\sum_{i=1}^{M}\{\partial U_i^f(\hat{\beta}^f)/\partial\beta\})^{-1}\{\sum_{i=1}^{M}U_i^f(\hat{\beta}^f)U_i^{fT}(\hat{\beta}^f)\}(\sum_{i=1}^{M}\{\partial U_i^{fT}(\hat{\beta}^f)/\partial\beta\})^{-1}.$$

Here we consider the missing covariate problem, and we assume that the response vector is fully observed. According to the missingness in the observed data set we reorder the covariates in x_{ij} as $x_{ij} = (x_{ij}^{(1)T}, \cdots, x_{ij}^{(k)T}, \cdots, x_{ij}^{(q)T})^T$ such that each $x_{ij}^{(k)}, k = 1, \cdots, q$, is a vector of covariates with the same missingness pattern, where q is the total number of distinct missingness patterns. We define $r_{ij}^{(k)}$ as an indicator variable and $r_{ij}^{(k)} = 1$ if $x_{ij}^{(k)}$ is observed and 0 otherwise for $k = 1, \cdots, q$, and $r_{ij}^{(0)} = 1$ if $r_{ij}^{(1)} = \cdots = r_{ij}^{(q)} = 1$ and 0 otherwise. In fact $r_{ij}^{(0)} = 1$ indicates x_{ij} is fully observed. For convenience we denote $x_{ij}^{(0)} = x_{ij}$. For each i, we specify $X_i^{(k)} = (x_{i1}^{(k)}, \cdots, x_{ij}^{(k)}, \cdots, x_{iJ}^{(k)})^T$. We assume that $X_i^{(k)}$ has n_{ik} fully observed $x_{ij}^{(k)}$. We remove all the unobserved elements and obtain observed covariates matrix $\tilde{X}_i^{(k)} = (x_{i1}^{(k)}, \cdots, x_{ij}^{(k)}, \cdots, x_{in_{ik}}^{(k)})^T$ and the corresponding response variable $\tilde{Y}_i^{(k)} = (y_{i1}, \cdots, y_{in_{ik}})^T$. Furthermore, we denote $\tilde{Y}_i^{(k)} = (\tilde{y}_{i1}^{(k)}, \cdots, \tilde{y}_{in_{ik}}^{(k)})^T$ and $\tilde{X}_i^{(k)} = (\tilde{x}_{i1}^{(k)}, \cdots, \tilde{x}_{in_{ik}}^{(k)})^T$.

For $k = 0, \cdots, q$, $j = 1, \cdots, J$, and $i = 1, \cdots, M$, we define the missing data probabilities as

$$\pi_{ij}^{(k)} = Pr(r_{ij}^{(k)} = 1|Y_i, X_i).$$

Under the MCAR missing mechanism, the missing data probability does not depend on any observed or unobserved data, that is $\pi_{ij}^{(k)} = P(r_{ij}^{(k)} = 1|Y_i, X_i) = P(r_{ij}^{(k)} = 1)$. Under the MAR missing mechanism, The missing data probability depends on the observed data, for example, $\pi_{ij}^{(k)} = P(r_{ij}^{(k)} = 1|Y_i, X_i) = P(r_{ij}^{(k)} = 1|Y_i, X_i^{(k)})$.

Let S_0 and S_k, $k = 1, \cdots, q$, denote the index set of the complete observed $x_{ij}^{(0)}$ and $x_{ij}^{(k)}$, $k = 1, \cdots, q$ respectively, Let m_k be the corresponding number of subjects in each index set. We see that $m_0 \leq m_k$ and we require $m_0 > C > 0$.

To give a clear description to the notation, we will give a simple example which will be used through the whole section. Suppose that there are two subjects in the study, each subject has four observations and there are three covariates in the data example. The data is as follows.

$$\begin{pmatrix}
y_{11} & \boxed{x_{111}} & \boxed{x_{112}} & \boxed{x_{113}} \\
y_{12} & x_{121} & \boxed{x_{122}} & x_{123} \\
y_{13} & \boxed{x_{131}} & \boxed{x_{132}} & \boxed{x_{133}} \\
y_{14} & x_{141} & x_{142} & x_{143} \\
y_{21} & \boxed{x_{211}} & x_{212} & \boxed{x_{213}} \\
y_{22} & x_{221} & \boxed{x_{222}} & x_{223} \\
y_{23} & x_{231} & x_{232} & x_{233} \\
y_{24} & \boxed{x_{241}} & x_{242} & \boxed{x_{243}}
\end{pmatrix}$$

The data elements with a box are missing value. In this example x_{ij1} and x_{ij3} have the same missingness pattern. We let $x_{ij}^{(1)} = (x_{ij1}, x_{ij3})$ and $x_{ij}^{(2)} = x_{ij2}$. We note that there are $q = 2$ distinct missingness patterns in the covariates. We reorder the covariates by the missing pattern. The data set will be

$$\begin{pmatrix} y_{11} & \boxed{x_{111}} & \boxed{x_{113}} & \boxed{x_{112}} \\ y_{12} & x_{121} & x_{123} & \boxed{x_{122}} \\ y_{13} & \boxed{x_{131}} & \boxed{x_{133}} & \boxed{x_{132}} \\ y_{14} & x_{141} & x_{143} & x_{142} \\ y_{21} & \boxed{x_{211}} & \boxed{x_{213}} & x_{212} \\ y_{22} & x_{221} & x_{223} & \boxed{x_{222}} \\ y_{23} & x_{231} & x_{233} & x_{232} \\ y_{24} & \boxed{x_{241}} & \boxed{x_{243}} & x_{242} \end{pmatrix}.$$

We denote the index set of complete observations as S_0. The observations in the S_0 are $(\tilde{Y}^{(0)}, \tilde{X}^{(0)}) = \begin{pmatrix} y_{14} & x_{141} & x_{142} & x_{143} \\ y_{23} & x_{231} & x_{232} & x_{233} \end{pmatrix} = \begin{pmatrix} \tilde{y}_{14} & \tilde{x}^{(0)}_{14} \\ \tilde{y}_{23} & \tilde{x}^{(0)}_{23} \end{pmatrix}$.

We denote the index set of complete observed $x^{(1)}_{ij}$ and $x^{(2)}_{ij}$ as S_1 and S_2 respectively. The observation in S_1 and S_2 are $(\tilde{Y}^{(1)}, \tilde{X}^{(1)}) = \begin{pmatrix} y_{12} & x_{121} & x_{123} \\ y_{14} & x_{141} & x_{143} \\ y_{22} & x_{221} & x_{223} \\ y_{23} & x_{231} & x_{233} \end{pmatrix} = \begin{pmatrix} \tilde{y}_{12} & \tilde{x}^{(1)}_{12} \\ \tilde{y}_{14} & \tilde{x}^{(1)}_{14} \\ \tilde{y}_{22} & \tilde{x}^{(1)}_{22} \\ \tilde{y}_{23} & \tilde{x}^{(1)}_{23} \end{pmatrix}$

and

$(\tilde{Y}^{(2)}, \tilde{X}^{(2)}) = \begin{pmatrix} y_{14} & x_{142} \\ y_{21} & x_{212} \\ y_{23} & x_{232} \\ y_{24} & x_{242} \end{pmatrix} = \begin{pmatrix} \tilde{y}_{14} & \tilde{x}^{(2)}_{14} \\ \tilde{y}_{21} & \tilde{x}^{(2)}_{21} \\ \tilde{y}_{23} & \tilde{x}^{(2)}_{23} \\ \tilde{y}_{24} & \tilde{x}^{(2)}_{24} \end{pmatrix}$, respectively.

3.3 MCAR Data

The generalized estimating equation will generate consistent estimator when the missing mechanism is MCAR, so we can apply the generalized estimating equation directly for each data set $S_k, k = 0, \cdots, q$.

For $i = 1, \cdots, M$, $j = 1, \cdots, J$ and $k = 1, \cdots, q$, we define $\mu_{ijk} = E(\tilde{y}_{ij}^{(k)} | \tilde{x}_{ij}^{(k)})$. We consider the generalized linear regression models

$$g_k(\mu_{ijk}) = (\tilde{x}_{ij}^{(k)})^T \gamma_k, j = 1, \cdots, n_{ik},$$

where $g_k(\centerdot)$ is a monotone differentiable link function, and γ_k is a vector of regression parameters. For convenience we denote the model of interest $g(\centerdot)$ as $g_0(\centerdot)$, that is $g_0(\mu_{ij0}) = g_0(E(\tilde{y}_{ij}^{(0)} | \tilde{x}_{ij}^{(0)})) = (\tilde{x}_{ij}^{(0)})^T \beta$. Here β is the parameter vector of interest and $\gamma_k, k = 1, \cdots, q$, are the vectors of surrogate parameters.

Let $\mu_{i0} = (\mu_{i10}, \cdots, \mu_{ij0}, \cdots, \mu_{in_{i0}0})^T$ and $\mu_{ik} = (\mu_{i1k}, \cdots, \mu_{ijk}, \cdots, \mu_{in_{i0}k})^T$, $k = 1, \cdots, q$.

Assume that $\hat{\beta}$ and $\hat{\gamma}_j, j = 1, \cdots, q$, solve the generalized estimating equations for β and γ_j given in (3.2) and (3.3) respectively.

$$\sum_{i \in S_0} U_{i0}(\beta) = \sum_{i=1}^{m_0} D_{i0}^T V_{i0}{}^{-1} (\tilde{Y}_i^{(0)} - \mu_{i0}) = 0 \tag{3.2}$$

$$\sum_{i \in S_0} U_{ik}(\gamma_k) = \sum_{i=1}^{m_0} D_{ik}^T V_{i0}{}^{-1} (\tilde{Y}_i^{(0)} - \mu_{ik}), \quad k = 1, \cdots, q \tag{3.3}$$

where $D_{i0} = \partial \mu_{i0} / \partial \beta$, $D_{ik} = \partial \mu_{ik} / \partial \gamma_k$, and V_{i0} is the covariance matrix for the response

$\tilde{Y}_i^{(k)}$. It is well known that $\hat{\beta}$ is consistent for β^* which is the true parameter value that would be computed if the data from the whole cohort were available, provided that some regularity conditions hold. Similarly, under some regularity conditions, $\hat{\gamma}_k$ is consistent for $\gamma_k{}^*$. We call $g_k(\tilde{X}_i^{(k)}, \gamma_k) = g(\mu_{ij}^{(k)})$ surrogate models and call $\gamma = (\gamma_1^T, \cdots, \gamma_q^T)^T$ a vector of surrogate parameters. We denote $U_i(\theta) = (U_{i0}^T(\beta), U_{iQ}^T(\gamma))^T$ with $\theta = (\beta^T, \gamma^T)^T$ and $U_{iQ}(\gamma) = (U_{i1}^T(\gamma_1), \cdots, U_{iq}^T(\gamma_q))^T$.

Under the regularity conditions in Appendix A, we can show that (i) $\hat{\theta} = (\hat{\beta}^T, \hat{\gamma}^T)^T$, with $\hat{\gamma} = (\hat{\gamma}_1^T, \cdots, \hat{\gamma}_q^T)^T$, is consistent for $\theta^* = (\beta^{*T}, \gamma^{*T})^T$ and (ii) $m_0^{1/2}(\hat{\theta} - \theta^*)$ is asymptotically normal with mean 0 and variance $\Gamma^{-1}\Sigma\Gamma^{-1}$ with $\Gamma = E\{\partial U_i(\theta^*)/\partial\theta\}$ and $\Sigma = E\{U_i(\theta^*)U_i^T(\theta^*)\}$.

We rewrite Γ as $diag(\Gamma_{00}, \Gamma_{11})$ with $\Gamma_{00} = E\{\partial U_{i0}(\beta^*)/\partial\beta\}$ and $\Gamma_{11} = E\{\partial U_{iQ}(\gamma^*)/\partial\gamma\}$. We partition the matrix Σ as $\begin{pmatrix} \Sigma_{00} & \Sigma_{01} \\ \Sigma_{01}^T & \Sigma_{11} \end{pmatrix}$ with $\Sigma_{00} = E\{U_{i0}(\beta^*)U_{i0}^T(\beta^*)\}$, $\Sigma_{01} = E\{U_{i0}(\beta^*)U_{iQ}^T(\gamma^*)\}$, and $\Sigma_{11} = E\{U_{iQ}(\gamma^*)U_{iQ}^T(\gamma^*)\}$. According to the multivariate normal distribution theory, the conditional distribution of $m_0^{1/2}(\hat{\beta} - \beta^*)$ given $m_0^{1/2}(\hat{\gamma} - \gamma^*)$ is asymptotic normal with mean

$$m_0^{1/2}\Gamma_{00}^{-1}\Sigma_{01}\Sigma_{11}^{-1}\Gamma_{11}(\hat{\gamma} - \gamma^*), \tag{3.4}$$

which suggests that the CC estimator $\hat{\beta}$ may be improved by using

$$\bar{\beta} = \hat{\beta} - \hat{\Gamma}_{00}^{-1}\hat{\Sigma}_{01}\hat{\Sigma}_{11}^{-1}\hat{\Gamma}_{11}(\hat{\gamma} - \bar{\gamma}), \tag{3.5}$$

where

$$\hat{\Gamma}_{00} = m_0^{-1} \sum_{i \in S_0} \{\partial U_{i0}(\hat{\beta})/\partial \beta\},$$
$$\hat{\Sigma}_{01} = m_0^{-1} \sum_{i \in S_0} \{U_{i0}(\hat{\beta}) U_{iQ}^T(\hat{\gamma})\},$$
$$\hat{\Sigma}_{11} = m_0^{-1} \sum_{i \in S_0} \{U_{iQ}(\hat{\gamma}) U_{iQ}^T(\hat{\gamma})\},$$
$$\hat{\Gamma}_{11} = m_0^{-1} \sum_{i \in S_0} \{\partial U_{iQ}(\hat{\gamma})/\partial \gamma\},$$

and $\bar{\gamma} = (\bar{\gamma}_1^T, \cdots, \bar{\gamma}_q^T)^T$. Here $\bar{\gamma}_k$ is an estimator based on observations in S_k, that is , $\bar{\gamma}_k$ solves

$$\sum_{i \in S_k} \bar{U}_{ik}(\gamma_k) = \sum_{i=1}^{m_k} \bar{D}_{ik}^T {V_{ik}}^{-1} (\tilde{Y}_i^{(k)} - \bar{\mu}_{ik}),$$

where $\bar{D}_{ik} = \partial \bar{\mu}_{ik} / \partial \gamma_k$, V_{ik} is the covariance matrix for the response $\tilde{Y}_i^{(k)}$ and $\bar{\mu}_{ik} = (\mu_{i1k}, \cdots, \mu_{ijk}, \cdots, \mu_{in_{ik}k})^T$, which allows all the observations in S_k to be used in the estimation.

In our simple example, let $g_k(\mu) = \mu$, the estimating equations for $\hat{\beta}$, $\hat{\gamma}$ and $\bar{\gamma}$ are as follows.

$$\begin{aligned} U_0(\beta) &= \sum_{i=1}^{m_0} U_{i0}(\beta) \\ &= (x_{14}^{(0)}) V_{10}^{-1} \left(y_{14} - \mu_{140} \right) + (x_{23}^{(0)}) V_{20}^{-1} \left(y_{23} - \mu_{130} \right), \end{aligned} \quad (3.6)$$

$$\begin{aligned} U_1(\gamma_1) &= \sum_{i=1}^{m_0} U_{i1}(\gamma_1) \\ &= (x_{14}^{(1)}) V_{10}^{-1} \left(y_{14} - \mu_{141} \right) + (x_{23}^{(1)}) V_{20}^{-1} \left(y_{23} - \mu_{131} \right), \end{aligned} \quad (3.7)$$

$$
\begin{aligned}
U_2(\gamma_2) &= \sum_{i=1}^{m_0} U_{i2}(\gamma_2) \\
&= (x_{14}^{(2)})V_{10}^{-1}\begin{pmatrix} y_{14}-\mu_{142} \end{pmatrix} + (x_{23}^{(2)})V_{20}^{-1}\begin{pmatrix} y_{23}-\mu_{132} \end{pmatrix},
\end{aligned} \tag{3.8}
$$

$$
\begin{aligned}
\bar{U}_1(\gamma_1) &= \sum_{i=1}^{m_1} \bar{U}_{i1}(\gamma_1) \\
&= (x_{12}^{(1)}, x_{14}^{(1)})V_{11}^{-1}\begin{pmatrix} y_{12}-\mu_{121} \\ y_{14}-\mu_{141} \end{pmatrix} + (x_{22}^{(1)}, x_{23}^{(1)})V_{21}^{-1}\begin{pmatrix} y_{22}-\mu_{221} \\ y_{23}-\mu_{231} \end{pmatrix},
\end{aligned} \tag{3.9}
$$

and

$$
\begin{aligned}
\bar{U}_2(\gamma_2) &= \sum_{i=1}^{m_2} \bar{U}_{i2}(\gamma_1) \\
&= (x_{14}^{(2)})V_{12}^{-1}\begin{pmatrix} y_{14}-\mu_{142} \end{pmatrix} + (x_{21}^{(2)}, x_{23}^{(2)}, x_{24}^{(2)})V_{22}^{-1}\begin{pmatrix} y_{21}-\mu_{212} \\ y_{23}-\mu_{232} \\ y_{24}-\mu_{242} \end{pmatrix},
\end{aligned} \tag{3.10}
$$

where equations (3.6), (3.7) and (3.8) are based on index set S_0, equations (3.9) and (3.10) are based on index set S_1 and S_2 respectively.

We call $\bar{\beta}$ an improved complete-case (ICC) estimator. We expect that the ICC estimator produces efficiency gains when $\hat{\beta}$ and $\hat{\gamma}$ are highly correlated and the sizes of the observations in S_k's are much larger than the size of the observations in S_0.

It can be shown that under regularity conditions (i) $\bar{\beta}$ is consistent for β^* and the consistency of $\bar{\beta}$ does not depend on the correctness of the sequence of parametric working models, and (ii) $m_0^{1/2}(\bar{\beta} - \beta^*)$ is asymptotic normal with mean 0 and variance

$$Var(m_0^{1/2}\bar{\beta}) = \Gamma_{00}^{-1}\Sigma_{00}\Gamma_{00}^{-1} - \Gamma_{00}^{-1}\Sigma_{01}(I - \Sigma_{11}^{-1}\Sigma_{\rho 11})\Sigma_{11}^{-1}\Sigma_{01}^{T}\Gamma_{00}^{-1}, \tag{3.11}$$

where $\Sigma_{\rho_{11}}$ is Σ_{11} with its khth element σ_{kh} replaced by $\sigma_{\rho kh} = (m_0 \cdot m_{kh})/(m_k \cdot m_h)\sigma_{kh}$ and m_{kh} is the number of observations in the intersection of S_k and S_h for $k, h = 1, \cdots, q$. The first term in (3.11) is the asymptotic variance of $m_0^{1/2}(\hat{\beta} - \beta^*)$, and the second term represents the improvement of the ICC estimator over the CC estimator. The asymptotic variance in (3.11) can be estimated by

$$\hat{\Gamma}_{00}^{-1}\hat{\Sigma}_{00}\hat{\Gamma}_{00}^{-1} - \hat{\Gamma}_{00}^{-1}\hat{\Sigma}_{01}(I - \hat{\Sigma}_{11}^{-1}\hat{\Sigma}_{\rho 11})\hat{\Sigma}_{11}^{-1}\hat{\Sigma}_{01}^{T}\hat{\Gamma}_{00}^{-1},$$

where $\hat{\Sigma}_{00} = m_0^{-1}\sum_{i \in V}\{U_{i0}(\hat{\beta})U_{i0}^{T}(\hat{\beta})\}$ and $\hat{\Sigma}_{\rho_{11}}$ is $\hat{\Sigma}_{11}$ with its khth element $\hat{\sigma}_{kh}$ replaced by $(m_0 \cdot m_{kh})/(m_k \cdot m_h)\hat{\sigma}_{kh}$ for $k, h = 1, \cdots, q$.

3.4 MAR Data with Known Missing Probability

The consistency of the GEE method requires that the missing mechanism is MCAR. When the missing mechanism is MAR, we can use the weighted generalized estimating equations in Robins et al. (1995) to obtain a consistent estimator for β. Chen et al. (2010) and Chen and Zhou (2011) used a new weight matrix and element-wise product to incorporate general working correlation matrices in longitudinal data analysis with missing covariates. Next we will explain how to extend the ICC approach to MAR data using weighted GEEs.

For the data in S_0, we consider the weighted generalized estimating equations given in (3.12), (3.13) and obtain the $\hat{\beta}_\pi$, $\hat{\gamma}_\pi$ respectively.

$$\sum_{i=1}^{M} U_{\pi i0}(\beta) = \sum_{i=1}^{M} D_{\pi i0}^{T} Z_{i0}(Y_i - \mu_{\pi i0}) = 0, \tag{3.12}$$

$$\sum_{i=1}^{M} U_{\pi ik}(\gamma_k) = \sum_{i=1}^{M} D_{\pi ik}^{T} Z_{i0}(Y_i - \mu_{\pi ik}) = 0, \text{ for } k = 1, \cdots, q. \tag{3.13}$$

For the data S_k, $k = 1, \cdots, q$, we consider the weighted generalized estimating equations given in (3.14) and obtain the $\bar{\gamma}_\pi$.

$$\sum_{i=1}^{M} \bar{U}_{\pi ik}(\gamma_k) = \sum_{i=1}^{M} D_{\pi ik}^{T} Z_{ik}(Y_i - \mu_{\pi ik}) = 0, \text{ for } k = 1, \cdots, q, \tag{3.14}$$

where $D_{\pi i0} = \partial \mu_{\pi i0}/\partial \beta$, $D_{\pi ki} = \partial \mu_{\pi ik}/\partial \gamma_k$, and $Z_{ik} = \alpha(\phi)^{-1} A_i{}^{-1/2}[V_i^{-1} \bullet \Delta_{ik}] A_i{}^{-1/2}$ with

$$\Delta_{ik} = \begin{pmatrix} \frac{I(r_{i1}^{(k)}=1)}{\pi_{i1}^{(k)}} & \frac{I(r_{i1}^{(k)}=1, r_{i2}^{(k)}=1)}{\pi_{i12}^{(k)}} & \cdots & \frac{I(r_{i1}^{(k)}=1, r_{iJ}^{(k)}=1)}{\pi_{i1J}^{(k)}} \\ \frac{I(r_{i2}^{(k)}=1, r_{i1}^{(k)}=1)}{\pi_{i21}^{(k)}} & \frac{I(r_{i2}^{(k)}=1)}{\pi_{i2}^{(k)}} & \cdots & \frac{I(r_{i2}^{(k)}=1, r_{iJ}^{(k)}=1)}{\pi_{i2J}^{(k)}} \\ \cdots & \cdots & \cdots & \cdots \\ \frac{I(r_{iJ}^{(k)}=1)}{\pi_{iJ1}^{(k)}} & \frac{I(r_{iJ}^{(k)}=1, r_{i2}^{(k)}=1)}{\pi_{iJ2}^{(k)}} & \cdots & \frac{I(r_{iJ}^{(k)}=1)}{\pi_{iJ}^{(k)}} \end{pmatrix}$$

and

$$\pi_{ijl}^{(k)} = P(r_{ij}^{(k)} = 1, r_{il}^{(k)} = 1 | Y_i, X_i),$$

for $k = 0, 1, \cdots, q$ and $j, l = 1, \cdots, J$. Here Δ_{i0} and Δ_{ik} are the weight matrix for S_0 and S_k respectively.

In our example, the weighted generalized estimating equation for β is $\sum_{i=1}^{2} U_{\pi i0}(\beta)$, where

$$U_{\pi10}(\beta) = (\boxed{x_{11}^{(0)}}, \boxed{x_{12}^{(0)}}, \boxed{x_{13}^{(0)}}, x_{14}^{(0)})\Phi[V_1^{-1} \bullet \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \frac{1}{\pi_{14}^{(0)}} \end{pmatrix}] \begin{pmatrix} \boxed{y_{11} - \mu_{110}} \\ \boxed{y_{12} - \mu_{120}} \\ \boxed{y_{13} - \mu_{130}} \\ y_{14} - \mu_{140} \end{pmatrix},$$

and

$$U_{\pi20}(\beta) = (\boxed{x_{21}^{(0)}}, \boxed{x_{22}^{(0)}}, x_{23}^{(0)}, \boxed{x_{24}^{(0)}})\Phi[V_2^{-1} \bullet \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & \frac{1}{\pi_{23}^{(0)}} & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}] \begin{pmatrix} \boxed{y_{21} - \mu_{210}} \\ \boxed{y_{22} - \mu_{220}} \\ y_{23} - \mu_{130} \\ \boxed{y_{24} - \mu_{240}} \end{pmatrix}.$$

Here [$\bullet$] is element-wise multiplication.

The weighted generalized estimating equations for γ_1 and γ_2 are $\sum_{i=1}^{2} U_{\pi i1}(\gamma_1)$, $\sum_{i=1}^{2} U_{\pi i2}(\gamma_2)$, $\sum_{i=1}^{2} \bar{U}_{\pi i1}(\gamma_1)$ and $\sum_{i=1}^{2} \bar{U}_{\pi i2}(\gamma_2)$, where

$$U_{\pi11}(\gamma_1) = (\boxed{x_{11}^{(1)}}, \boxed{x_{12}^{(1)}}, \boxed{x_{13}^{(1)}}, x_{14}^{(1)})\Phi[V_1^{-1} \bullet \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \frac{1}{\pi_{14}^{(0)}} \end{pmatrix}] \begin{pmatrix} \boxed{y_{11} - \mu_{111}} \\ \boxed{y_{12} - \mu_{121}} \\ \boxed{y_{13} - \mu_{131}} \\ y_{14} - \mu_{141} \end{pmatrix},$$

$$U_{\pi 21}(\gamma_1) = (\boxed{x_{21}^{(1)}}, \boxed{x_{22}^{(1)}}, x_{23}^{(1)}, \boxed{x_{24}^{(1)}})\Phi[V_2^{-1} \bullet \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & \frac{1}{\pi_{23}^{(0)}} & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}] \begin{pmatrix} \boxed{y_{21} - \mu_{211}} \\ \boxed{y_{22} - \mu_{221}} \\ y_{23} - \mu_{131} \\ \boxed{y_{24} - \mu_{241}} \end{pmatrix},$$

$$U_{\pi 12}(\gamma_2) = (\boxed{x_{11}^{(2)}}, \boxed{x_{12}^{(2)}}, \boxed{x_{13}^{(2)}}, x_{14}^{(2)})\Phi[V_1^{-1} \bullet \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \frac{1}{\pi_{14}^{(0)}} \end{pmatrix}] \begin{pmatrix} \boxed{y_{11} - \mu_{112}} \\ \boxed{y_{12} - \mu_{122}} \\ \boxed{y_{13} - \mu_{132}} \\ y_{14} - \mu_{142} \end{pmatrix},$$

$$U_{\pi 22}(\gamma_2) = (\boxed{x_{21}^{(2)}}, \boxed{x_{22}^{(2)}}, x_{23}^{(2)}, \boxed{x_{24}^{(2)}})\Phi[V_2^{-1} \bullet \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & \frac{1}{\pi_{23}^{(0)}} & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}] \begin{pmatrix} \boxed{y_{21} - \mu_{212}} \\ \boxed{y_{22} - \mu_{222}} \\ y_{23} - \mu_{132} \\ \boxed{y_{24} - \mu_{242}} \end{pmatrix},$$

$$\bar{U}_{\pi 11}(\gamma_1) = (\boxed{x_{11}^{(1)}}, x_{12}^{(1)}, \boxed{x_{13}^{(1)}}, x_{14}^{(1)})\Phi[V_1^{-1} \bullet \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & \frac{1}{\pi_{12}^{(1)}} & 0 & \frac{1}{\pi_{124}^{(1)}} \\ 0 & 0 & 0 & 0 \\ 0 & \frac{1}{\pi_{142}^{(1)}} & 0 & \frac{1}{\pi_{14}^{(1)}} \end{pmatrix}] \begin{pmatrix} \boxed{y_{11} - \mu_{111}} \\ y_{12} - \mu_{121} \\ \boxed{y_{13} - \mu_{131}} \\ y_{14} - \mu_{141} \end{pmatrix},$$

$$\bar{U}_{\pi 21}(\gamma_1) = (\boxed{x_{21}^{(1)}}, x_{22}^{(1)}, x_{23}^{(1)}, \boxed{x_{24}^{(1)}})\Phi[V_2^{-1} \bullet \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & \frac{1}{\pi_{22}^{(1)}} & \frac{1}{\pi_{223}^{(1)}} & 0 \\ 0 & \frac{1}{\pi_{232}^{(1)}} & \frac{1}{\pi_{23}^{(1)}} & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}] \begin{pmatrix} \boxed{y_{21} - \mu_{211}} \\ y_{22} - \mu_{221} \\ y_{23} - \mu_{231} \\ \boxed{y_{24} - \mu_{241}} \end{pmatrix},$$

$$\bar{U}_{\pi 12}(\gamma_2) = (\boxed{x_{11}^{(2)}}, \boxed{x_{12}^{(2)}}, \boxed{x_{13}^{(2)}}, x_{14}^{(2)})\Phi[V_1^{-1} \bullet \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \frac{1}{\pi_{14}^{(2)}} \end{pmatrix}] \begin{pmatrix} \boxed{y_{11} - \mu_{112}} \\ \boxed{y_{12} - \mu_{122}} \\ \boxed{y_{13} - \mu_{132}} \\ y_{14} - \mu_{142} \end{pmatrix},$$

$$\bar{U}_{\pi 22}(\gamma_2) = (x_{21}^{(2)}, \boxed{x_{22}^{(2)}}, x_{23}^{(2)}, x_{24}^{(2)})\Phi[V_2^{-1} \bullet \begin{pmatrix} \frac{1}{\pi_{21}^{(2)}} & 0 & \frac{1}{\pi_{213}^{(2)}} & \frac{1}{\pi_{214}^{(2)}} \\ 0 & 0 & 0 & 0 \\ \frac{1}{\pi_{231}^{(2)}} & 0 & \frac{1}{\pi_{23}^{(2)}} & \frac{1}{\pi_{234}^{(2)}} \\ \frac{1}{\pi_{241}^{(2)}} & 0 & \frac{1}{\pi_{243}^{(2)}} & \frac{1}{\pi_{24}^{(2)}} \end{pmatrix}] \begin{pmatrix} y_{21} - \mu_{212} \\ \boxed{y_{22} - \mu_{222}} \\ y_{23} - \mu_{232} \\ y_{24} - \mu_{242} \end{pmatrix}.$$

We see that $\sum_{i=1}^{2} U_{\pi i1}(\gamma_1)$ and $\sum_{i=1}^{2} U_{\pi i2}(\gamma_2)$ are based on the observations in S_0, and $\sum_{i=1}^{2} \bar{U}_{\pi i1}(\gamma_1)$ and $\sum_{i=1}^{2} \bar{U}_{\pi i2}(\gamma_2)$ are based on observations in S_1 and S_2 respectively.

We note that $\hat{\beta}_\pi$ and $\hat{\gamma}_\pi$ are computed based on observations in S_0, while $\bar{\gamma}_\pi$ is computed based on the larger data sets $S_k, k = 1, \cdots, q$. Following a procedure similar to that in Section 3.2, under regularity conditions we obtain the following results:

(i) $M^{1/2}(\hat{\beta}_\pi - \beta^*)$ given $M^{1/2}(\hat{\gamma}_\pi - \gamma^*)$ is asymptotic normal with mean

$$M^{1/2}\Gamma_{00}^{-1}\Sigma_{\pi 01}\Sigma_{\pi 11}^{-1}\Gamma_{11}(\hat{\gamma}_\pi - \gamma^*), \text{ where}$$

$\Sigma_{\pi 01} = E[U_{\pi i0}(\beta^*)U_{\pi iQ}^T(\gamma^*)], \Sigma_{\pi 11} = E[U_{\pi iQ}(\gamma^*)U_{\pi iQ}^T(\gamma^*)]$ with $U_{\pi iQ}(\gamma) = (U_{\pi i1}(\gamma_1), \cdots, U_{\pi iq}(\gamma_q))^T$

(ii) β can be consistently estimated by

$$\bar{\beta}_\pi = \hat{\beta} - \hat{\Gamma}_{\pi 00}^{-1}\hat{\Sigma}_{\pi 01}\hat{\Sigma}_{\pi 11}^{-1}\hat{\Gamma}_{\pi 11}(\hat{\gamma}_\pi - \bar{\gamma}_\pi), \tag{3.15}$$

where

$$\hat{\Gamma}_{\pi 00} = M^{-1}\sum_{i=1}^{M}\partial U_{\pi i0}(\hat{\beta}_\pi)/\partial\beta,$$
$$\hat{\Sigma}_{\pi 01} = M^{-1}\sum_{i=1}^{M}U_{\pi i0}(\hat{\beta}_\pi)U_{\pi iQ}^T(\hat{\gamma}_\pi),$$
$$\hat{\Sigma}_{\pi 11} = M^{-1}\sum_{i=1}^{M}U_{\pi iQ}(\hat{\gamma}_\pi)U_{\pi iQ}^T(\hat{\gamma}_\pi),$$
$$\hat{\Gamma}_{\pi 11} = M^{-1}\sum_{i=1}^{M}\partial U_{\pi iQ}(\hat{\gamma}_\pi)/\partial\gamma.$$

The consistency of $\bar{\beta}_\pi$ does not depend on the correctness of the working regression models. We call $\bar{\beta}_\pi$ an improved weighted complete-case (IWCC) estimator.

(iii) $M^{1/2}(\bar{\beta}_\pi - \beta^*)$ is asymptotically normal with mean 0 and variance

$$\begin{aligned}&\Gamma_{00}^{-1}\Sigma_{\pi 00}\Gamma_{00}^{-1} - \Gamma_{00}^{-1}\{\Sigma_{\pi 01}\Sigma_{\pi 11}^{-1}[(\Sigma_{\pi 12} - \Sigma_{\pi 22} + \Sigma_{\pi 12}^T)\Sigma_{\pi 11}^{-1}\Sigma_{\pi 01}^T - \Sigma_{\pi 02}^T]+\\&(\Sigma_{\pi 01} - \Sigma_{\pi 02})\Sigma_{\pi 11}^{-1}\Sigma_{\pi 01}^T\}\Gamma_{00}^{-1},\end{aligned} \tag{3.16}$$

where

$$\Sigma_{\pi 00} = E[U_{\pi i0}(\beta^*)U_{\pi i0}^T(\beta^*)],$$

$$\Sigma_{\pi 02} = E[U_{\pi i0}(\beta^*)\bar{U}_{\pi iQ}^T(\gamma^*)],$$

$$\Sigma_{\pi 12} = E[U_{\pi iQ}(\gamma^*)\bar{U}_{\pi iQ}^T(\gamma^*)],$$

$$\Sigma_{\pi 22} = E[\bar{U}_{\pi i0}(\gamma^*)\bar{U}_{\pi i0}^T(\gamma^*)],$$

with

$$\bar{U}_{\pi iQ}(\gamma) = (\bar{U}_{\pi i1}(\gamma_1), \cdots, \bar{U}_{\pi ik}(\gamma_k), \cdots, \bar{U}_{\pi iq}(\gamma_q))^T.$$

The asymptotic variance (3.16) can be estimated by

$$\hat{\Gamma}_{\pi 00}^{-1}\hat{\Sigma}_{\pi 00}\hat{\Gamma}_{\pi 00}^{-1} - \hat{\Gamma}_{\pi 00}^{-1}\{\hat{\Sigma}_{\pi 01}\hat{\Sigma}_{\pi 11}^{-1}[(\hat{\Sigma}_{\pi 12} - \hat{\Sigma}_{\pi 22} + \hat{\Sigma}_{\pi 12}^T)\hat{\Sigma}_{\pi 11}^{-1}\hat{\Sigma}_{\pi 01}^T - \hat{\Sigma}_{\pi 02}^T] + (\hat{\Sigma}_{\pi 01} - \hat{\Sigma}_{\pi 02})\hat{\Sigma}_{\pi 11}^{-1}\hat{\Sigma}_{\pi 01}^T\}\hat{\Gamma}_{\pi 00}^{-1}, \quad (3.17)$$

where

$$\hat{\Gamma}_{\pi 00} = M^{-1}\sum_{i=1}^{M} \partial U_{\pi i0}(\hat{\beta}_\pi)/\partial\beta,$$
$$\hat{\Gamma}_{\pi 11} = M^{-1}\sum_{i=1}^{M} \partial U_{\pi iQ}(\hat{\gamma}_\pi)/\partial\gamma,$$
$$\hat{\Sigma}_{\pi 00} = M^{-1}\sum_{i=1}^{M} U_{\pi i0}(\hat{\beta}_\pi)U_{\pi i0}^T(\hat{\beta}_\pi),$$
$$\hat{\Sigma}_{\pi 01} = M^{-1}\sum_{i=1}^{M} U_{\pi i0}(\hat{\beta}_\pi)U_{\pi iQ}^T(\hat{\gamma}_\pi),$$
$$\hat{\Sigma}_{\pi 02} = M^{-1}\sum_{i=1}^{M} U_{\pi i0}(\hat{\beta}_\pi)\bar{U}_{\pi iQ}^T(\bar{\gamma}_\pi),$$
$$\hat{\Sigma}_{\pi 11} = M^{-1}\sum_{i=1}^{M} U_{\pi iQ}(\hat{\gamma}_\pi)U_{\pi iQ}^T(\hat{\gamma}_\pi),$$
$$\hat{\Sigma}_{\pi 12} = M^{-1}\sum_{i=1}^{M} U_{\pi iQ}(\hat{\gamma}_\pi)\bar{U}_{\pi Q}^T(\bar{\gamma}_\pi),$$
$$\hat{\Sigma}_{\pi 22} = M^{-1}\sum_{i=1}^{M} \bar{U}_{\pi iQ}(\bar{\gamma}_\pi)\bar{U}_{\pi iQ}^T(\bar{\gamma}_\pi).$$

As in Section 3.2, we see that the first term in (3.17) is an estimate of the asymptotic variance of $M^{1/2}(\hat{\beta}_\pi - \beta^*)$, and the second term represents the improvement of the IWCC estimator over the weighted CC estimator.

3.5 MAR Data with Estimated Missing Probability

It is well known that the estimation efficiency of the inverse probability weighted estimates can be further improved by using estimated selection probabilities $\hat{\pi}_{ij}$ instead of the known selection probabilities (Robins et al. 1994; Lawless et al. 1999; Chatterjee and Breslow 2003; Breslow et al. 2009). In practice, MAR data often occurs with unknown missing probabilities where the selection probabilities must be estimated in the weighted estimating equations. Robins et al. (1995) developed a class of inverse probability weighted generalized estimating equations (IPWGEE), which can yield consistent estimators when data are MAR. The weights are obtained from models for the missing data process, and these models must be correctly specified for the resulting estimators to be consistent.

Modeling the missing data process can be very difficult in practice. To illustrate how to use the unified approach in missing by happenstance case, we only consider a simple missing data process. Suppose now that $r_{ij}^{(k)}$ and $r_{il}^{(k)}$ are independent for $j, l = 1, \cdots, J$, $k = 0, 1, \cdots, q$, and $\pi_{ij}^{(k)}$ depends on the fully observed variables which may include our variables in the regression model and other auxiliary variables and the dependence is specified up to a known probability function indexed by a finite number of unknown parameters α_k.

One can estimate $(\beta^{*T}, \gamma^{*T})^T$ by $(\hat{\beta}_{\hat{\pi}}^T, \hat{\gamma}_{\hat{\pi}}^T)^T$ with $\hat{\gamma}_{\hat{\pi}} = (\hat{\gamma}_{\hat{\pi}1}^T, \cdots, \hat{\gamma}_{\hat{\pi}q}^T)^T$ using the weighted estimating equations (3.18) and (3.19), while constructing another estimating equations (3.20) for on $r_{ij}^{(0)}$ to estimate the nuisance parameters α_0.

$$\sum_{i=1}^{M} U_{\hat{\pi}i0}(\beta) = \sum_{i=1}^{M} D_{\pi i0}^{T} \hat{Z}_{i0}(Y_i - \mu_{\pi i0}) \;=\; 0, \tag{3.18}$$

$$\sum_{i=1}^{M} U_{\hat{\pi}ik}(\gamma_k) = \sum_{i=1}^{M} D_{\pi ik}^{T} \hat{Z}_{i0}(Y_i - \mu_{\pi ik}) \;=\; 0, \text{ for } k = 1, \cdots, q, \tag{3.19}$$

$$\sum_{i=1}^{M} H_{i0}(\alpha_0) \;=\; 0, \tag{3.20}$$

where $D_{\pi i0} = \partial \mu_{\pi i0}/\partial \beta$, $D_{\pi ki} = \partial \mu_{\pi ik}/\partial \gamma_k$, $\hat{Z}_{i0} = \alpha(\phi)^{-1} A_i{}^{-1/2}[V_i^{-1} \bullet \hat{\Delta}_{i0}] A_i{}^{-1/2}$ with

$$\hat{\Delta}_{i0} = \begin{pmatrix} \frac{I(r_{i1}^{(0)}=1)}{\hat{\pi}_{i1}^{(0)}} & \frac{I(r_{i1}^{(0)}=1, r_{i2}^{(0)}=1)}{\hat{\pi}_{i12}^{(0)}} & \cdots & \frac{I(r_{i1}^{(0)}=1, r_{iJ}^{(0)}=1)}{\hat{\pi}_{i1J}^{(0)}} \\ \frac{I(r_{i2}^{(0)}=1, r_{i1}^{(0)}=1)}{\hat{\pi}_{i21}^{(0)}} & \frac{I(r_{i2}^{(0)}=1)}{\hat{\pi}_{i2}^{(0)}} & \cdots & \frac{I(r_{i2}^{(0)}=1, r_{iJ}^{(0)}=1)}{\hat{\pi}_{i2J}^{(0)}} \\ \cdots & \cdots & \cdots & \cdots \\ \frac{I(r_{iJ}^{(0)}=1)}{\hat{\pi}_{iJ1}^{(0)}} & \frac{I(r_{iJ}^{(0)}=1, r_{i20}=1)}{\hat{\pi}_{iJ2}^{(0)}} & \cdots & \frac{I(r_{iJ}^{(0)}=1)}{\hat{\pi}_{iJ}^{(0)}} \end{pmatrix}.$$

Let $U_{\hat{\pi}iQ}(\gamma, \alpha_0) = (U_{\hat{\pi}i1}^T, \cdots, U_{\hat{\pi}ik}^T, \cdots, U_{\hat{\pi}iq}^T)^T$. Following the procedure similar to that in Section 3.2, we can show that the conditional distribution of $M^{1/2}(\hat{\beta}_{\hat{\pi}} - \beta^*)$ given $M^{1/2}(\hat{\gamma}_{\hat{\pi}} - \gamma^*)$ is asymptotic normal with mean

$$M^{1/2} \Gamma_{00}^{-1} \Sigma_{\hat{\pi}01} \Sigma_{\hat{\pi}11}^{-1} \Gamma_{11}(\gamma_{\hat{\pi}} - \gamma^*),$$

where

$$\Sigma_{\hat{\pi}01} = E\{Res(U_{\hat{\pi}i0}(\beta^*, \alpha_0^*), H_{i0}(\alpha_0^*)) Res^T(U_{\hat{\pi}iQ}(\gamma^*, \alpha_0^*), H_{i0}(\alpha_0^*))\},$$

and

$$\Sigma_{\hat{\pi}11} = E\{Res(U_{\hat{\pi}iQ}(\gamma^*, \alpha_0^*), H_{i0}(\alpha_0^*)) Res^T(U_{\hat{\pi}iQ}(\gamma^*, \alpha_0^*), H_{i0}(\alpha_0^*))\}.$$

It suggests that the weighted CC estimator $\hat{\beta}_{\hat{\pi}}$ may be improved by using

$$\bar{\beta}_{\hat{\pi}} = \hat{\beta}_{\hat{\pi}} - \hat{\Gamma}_{\hat{\pi}00}^{-1}\hat{\Sigma}_{\hat{\pi}01}\hat{\Sigma}_{\hat{\pi}11}^{-1}\hat{\Gamma}_{\hat{\pi}11}(\hat{\gamma}_{\hat{\pi}} - \bar{\gamma}_{\hat{\pi}}), \tag{3.21}$$

where

$$\begin{aligned}
\hat{\Gamma}_{\hat{\pi}00} &= M^{-1}\sum_{i=1}^{M} \partial U_{\hat{\pi}i0}(\hat{\beta}, \hat{\alpha}_0)/\partial\beta, \\
\hat{\Sigma}_{\hat{\pi}01} &= M^{-1}\sum_{i=1}^{M} \hat{R}es(U_{\hat{\pi}i0}(\hat{\beta}, \hat{\alpha}_0), H_{i0}(\hat{\alpha}_0))\hat{R}es^T(U_{\hat{\pi}iQ}(\hat{\beta}, \hat{\alpha}_0), H_{i0}(\hat{\alpha}_0)), \\
\hat{\Sigma}_{\hat{\pi}11} &= M^{-1}\sum_{i=1}^{M} \hat{R}es(U_{\hat{\pi}iQ}(\hat{\beta}, \hat{\alpha}_0), H_{i0}(\hat{\alpha}_0))\hat{R}es^T(U_{\hat{\pi}iQ}(\hat{\beta}, \hat{\alpha}_0), H_{i0}(\hat{\alpha}_0)), \\
\hat{\Gamma}_{\hat{\pi}11} &= M^{-1}\sum_{i=1}^{M} \partial U_{\hat{\pi}iQ}(\hat{\beta}, \hat{\alpha}_0)/\partial\gamma.
\end{aligned}$$

We note that $\bar{\gamma}_{\hat{\pi}} = (\bar{\gamma}_{\hat{\pi}1}^T, \cdots, \bar{\gamma}_{\hat{\pi}q}^T)^T$ is estimated using the weighted estimating equations (3.22) and (3.23).

$$\sum_{i=1}^{M} \bar{U}_{\pi ik}(\gamma_k) = \sum_{i=1}^{M} D_{\pi ik}^T \hat{Z}_{ik}(Y_i - \mu_{\pi ik}) = 0, \text{ for } k = 1, \cdots, q, \tag{3.22}$$

$$\sum_{i=1}^{N} H_{ik}(\alpha_k) = 0, \text{ for } k = 1, \cdots, q, \tag{3.23}$$

where $D_{\pi ki} = \partial\mu_{\pi ik}/\partial\gamma_k$, $\hat{Z}_{ik} = \alpha(\phi)^{-1}A_i{}^{-1/2}[V_i^{-1} \bullet \hat{\Delta}_{ik}]A_i{}^{-1/2}$ with

$$\hat{\Delta}_{ik} = \begin{pmatrix}
\frac{I(r_{i1}^{(k)}=1)}{\hat{\pi}_{i1}^{(k)}} & \frac{I(r_{i1}^{(k)}=1, r_{i2}^{(k)}=1)}{\hat{\pi}_{i12}^{(k)}} & \cdots & \frac{I(r_{i1}^{(k)}=1, r_{iJ}^{(k)}=1)}{\hat{\pi}_{i1J}^{(k)}} \\
\frac{I(r_{i2}^{(k)}=1, r_{i1}^{(k)}=1)}{\hat{\pi}_{i21}^{(k)}} & \frac{I(r_{i2}^{(k)}=1)}{\hat{\pi}_{i2}^{(k)}} & \cdots & \frac{I(r_{i2}^{(k)}=1, r_{iJ}^{(k)}=1)}{\hat{\pi}_{i2J}^{(k)}} \\
\cdots & \cdots & \cdots & \cdots \\
\frac{I(r_{iJ}^{(k)}=1)}{\hat{\pi}_{iJ1}^{(k)}} & \frac{I(r_{iJ}^{(k)}=1, r_{i2}^{(k)}=1)}{\hat{\pi}_{iJ2}^{(k)}} & \cdots & \frac{I(r_{iJ}^{(k)}=1)}{\hat{\pi}_{iJ}^{(k)}}
\end{pmatrix}.$$

Let $\alpha_Q = (\alpha_1^T, \cdots, \alpha_q^T)^T$, $\hat{\alpha}_Q = (\hat{\alpha}_1^T, \cdots, \hat{\alpha}_q^T)^T$, and $\sum_{i=1}^N H_{ik}(\alpha_k)$ be a system of estimating functions for α_k.

We see that $\bar{\gamma}_{\hat{\pi}k}$ is estimated based on observations in S_k, which allows all the information in S_k to be used to increase the estimation efficiency. We call $\bar{\beta}_{\hat{\pi}}$ an improved weighted complete-case (IWCC) estimator using estimated π.

Under some regularity conditions, we can obtain that $M^{1/2}(\bar{\beta}_{\hat{\pi}} - \beta^*)$ is asymptotic normal with mean 0 and variance given by

$$\begin{aligned}&\Gamma_{00}^{-1}\Sigma_{\hat{\pi}00}\Gamma_{00}^{-1} - \Gamma_{00}^{-1}\{\Sigma_{\hat{\pi}01}\Sigma_{\hat{\pi}11}^{-1}[(\Sigma_{\hat{\pi}12} - \Sigma_{\hat{\pi}22} + \Sigma_{\hat{\pi}12}^T)\Sigma_{\hat{\pi}11}^{-1}\Sigma_{\hat{\pi}01}^T - \Sigma_{\hat{\pi}02}^T] + \\ &(\Sigma_{\hat{\pi}01} - \Sigma_{\hat{\pi}02})\Sigma_{\hat{\pi}11}^{-1}\Sigma_{\hat{\pi}01}^T\}\Gamma_{00}^{-1},\end{aligned} \tag{3.24}$$

where

$$\begin{aligned}\Sigma_{\hat{\pi}00} &= E\{Res(U_{\hat{\pi}i0}(\beta^*, \alpha_0^*), H_{i0}(\alpha_0^*))Res^T(U_{\hat{\pi}i0}(\beta^*, \alpha_0^*), H_{i0}(\alpha_0^*))\},\\ \Sigma_{\hat{\pi}02} &= E\{Res(U_{\hat{\pi}i0}(\beta^*, \alpha_0^*), H_{i0}(\alpha_0^*))Res^T(\bar{U}_{\hat{\pi}iQ}(\gamma^*, \alpha_Q^*), H_{iQ}(\alpha_Q^*))\},\\ \Sigma_{\hat{\pi}12} &= E\{Res(U_{Qi}(\gamma^*, \alpha_0^*), H_{i0}(\alpha_0^*))Res^T(\bar{U}_{\hat{\pi}iQ}(\gamma^*, \alpha_Q^*), H_{iQ}(\alpha_Q^*))\},\\ \Sigma_{\hat{\pi}22} &= E\{Res(\bar{U}_{\hat{\pi}iQ}(\gamma^*, \alpha_Q^*), H_{iQ}(\alpha_Q^*))Res^T(\bar{U}_{\hat{\pi}iQ}(\gamma^*, \alpha_Q^*), H_{iQ}(\alpha_Q^*))\}\end{aligned}$$

with $\bar{U}_{\hat{\pi}iQ}(\gamma, \alpha_Q) = (\bar{U}_{\hat{\pi}i1}^T, \cdots, \bar{U}_{\hat{\pi}iq}^T)^T$, and $H_{iQ}(\alpha_Q) = (H_{i1}^T(\alpha_1), \cdots, H_{iq}^T(\alpha_q))^T$.

The asymptotic variance in (3.24) can be estimated by

$$\begin{aligned}&\hat{\Gamma}_{\hat{\pi}00}^{-1}\hat{\Sigma}_{\hat{\pi}00}\hat{\Gamma}_{\hat{\pi}00}^{-1} - \hat{\Gamma}_{\hat{\pi}00}^{-1}\{\hat{\Sigma}_{\hat{\pi}01}\hat{\Sigma}_{\hat{\pi}11}^{-1}[(\hat{\Sigma}_{\hat{\pi}12} - \hat{\Sigma}_{\hat{\pi}22} + \hat{\Sigma}_{\hat{\pi}12}^T)\hat{\Sigma}_{\hat{\pi}11}^{-1}\hat{\Sigma}_{\hat{\pi}01}^T - \hat{\Sigma}_{\hat{\pi}02}^T] + \\ &(\hat{\Sigma}_{\hat{\pi}01} - \hat{\Sigma}_{\hat{\pi}02})\hat{\Sigma}_{\hat{\pi}11}^{-1}\hat{\Sigma}_{\hat{\pi}01}^T\}\hat{\Gamma}_{\hat{\pi}00}^{-1},\end{aligned} \tag{3.25}$$

where

$$\hat{\Sigma}_{\hat{\pi}00} = N^{-1}\sum_{i=1}^{N}\hat{R}es(U_{\hat{\pi}i0}(\hat{\beta},\hat{\alpha}_0),H_{i0}(\hat{\alpha}_0))\hat{R}es^T(U_{\hat{\pi}i0}(\hat{\beta},\hat{\alpha}_0),H_{i0}(\hat{\alpha}_0)),$$

$$\hat{\Sigma}_{\hat{\pi}02} = N^{-1}\sum_{i=1}^{N}\hat{R}es(U_{\hat{\pi}i0}(\hat{\beta},\hat{\alpha}_0),H_{i0}(\hat{\alpha}_0))\hat{R}es^T(\bar{U}_{\hat{\pi}iQ}(\bar{\gamma},\hat{\alpha}),H_{iQ}(\hat{\alpha}_Q)),$$

$$\hat{\Sigma}_{\hat{\pi}12} = N^{-1}\sum_{i=1}^{N}\hat{R}es(U_{\hat{\pi}iQ}(\hat{\beta},\hat{\alpha}_0),H_{i0}(\hat{\alpha}_0))\hat{R}es^T(\bar{U}_{\hat{\pi}iQ}(\bar{\gamma},\hat{\alpha}),H_{iQ}(\hat{\alpha}_Q)),$$

$$\hat{\Sigma}_{\hat{\pi}22} = N^{-1}\sum_{i=1}^{N}\hat{R}es(\bar{U}_{\hat{\pi}iQ}(\bar{\gamma},\hat{\alpha}_Q),H_{iQ}(\hat{\alpha}_Q))\hat{R}es^T(\bar{U}_{\hat{\pi}iQ}(\bar{\gamma},\hat{\alpha}),H_{iQ}(\hat{\alpha}_Q)).$$

For the IPWGEE, to obtain a consistent estimator we need to “correct“ models for the missing data process and also need “correct” models for the response process given the covariates, but we do not need to model the distribution of the missing covariates. If the missing data process models are misspecified, both the $\bar{\beta}_{\hat{\pi}}$ and $\hat{\beta}_{\hat{\pi}}$ can be biased.

3.6 Simulation Studies

In this section we use simulation studies to examine the finite sample performance of the ICC and the IWCC estimators. We consider the linear regression model,

$$y_{ij} = \mu_{ij} + \epsilon_{ij} = \beta_0 + \beta_1 * x_{ij1} + \beta_2 * x_{ij2} + \epsilon_{ij}$$

and logistic regression model,

$$logit(\mu_{ij}) = logit(Pr(y_{ij} = 1|x_{ij})) = \beta_0 + \beta_1 * x_{ij1} + \beta_2 * x_{ij2},$$

where x_{ij1} and x_{ij2} are time-dependent continuous covariates. We consider two correlation structures (i) exchangeable and (ii) Ar(1) with parameter $\rho = 0.3$. We let $J = 3$, $\beta^* =$

Table 3.1 Linear Regression Model

	MCAR			MAR π			MAR $\hat{\pi}$		
	β_0	β_1	β_2	$\beta_{\pi 0}$	$\beta_{\pi 1}$	$\beta_{\pi 2}$	$\beta_{\hat{\pi} 0}$	$\beta_{\hat{\pi} 1}$	$\beta_{\hat{\pi} 2}$
Exchangeable Correlation $\rho = 0.3$									
ICC or IWCC estimation									
Bias	0.001	$0.0^3 3$	0.002	-0.003	-0.002	-0.001	-0.007	$-0.0^3 2$	$0.0^3 8$
$s.d.$	0.043	0.045	0.045	0.054	0.045	0.044	0.060	0.042	0.046
$s.e.$	0.046	0.044	0.044	0.057	0.044	0.044	0.074	0.045	0.045
MSE	0.002	0.002	0.002	0.003	0.002	0.002	0.004	0.002	0.002
95%CP	95.6%	95.2%	94.8%	95.8%	94.4%	97.0%	97.8%	95.0%	93.8%
ARE	1.519	1.284	1.235	1.449	1.138	1.291	1.480	1.474	1.328
CC or weighted CC estimation									
Bias	0.002	0.002	$0.0^3 3$	$0.0^3 4$	-0.005	$-0.0^3 2$	0.003	-0.002	-0.002
$s.d.$	0.053	0.051	0.050	0.065	0.048	0.050	0.073	0.051	0.053
MSE	0.003	0.003	0.003	0.004	0.002	0.002	0.005	0.003	0.003
Ar(1) $\rho = 0.3$									
ICC or IWCC estimation									
Bias	-0.002	-0.001	-0.001	$-0.0^3 6$	0.005	0.002	$-0.0^3 9$	$-0.0^3 6$	-0.002
$s.d.$	0.043	0.043	0.045	0.041	0.041	0.041	0.042	0.042	0.042
$s.e.$	0.041	0.043	0.044	0.040	0.041	0.041	0.043	0.041	0.042
MSE	0.002	0.002	0.002	0.002	0.002	0.002	0.002	0.002	0.002
95%CP	94.4%	95.4%	93.2%	93.4%	94.2%	94.6%	96.2%	94.0%	94.2%
ARE	1.462	1.299	1.186	1.563	1.205	1.259	1.361	1.252	1.252
CC or weighted CC estimation									
Bias	-0.004	-0.001	$-0.0^3 5$	0.002	0.003	0.001	0.006	-0.005	-0.004
$s.d.$	0.052	0.049	0.049	0.050	0.045	0.046	0.049	0.047	0.047
MSE	0.003	0.002	0.002	0.002	0.002	0.002	0.002	0.002	0.002

[a] $0.0^3 2 = 0.0002$.

Table 3.2 Logistic Regression Model

	MCAR			MAR π			MAR $\hat{\pi}$		
	β_0	β_1	β_2	$\beta_{\pi 0}$	$\beta_{\pi 1}$	$\beta_{\pi 2}$	$\beta_{\hat{\pi} 0}$	$\beta_{\hat{\pi} 1}$	$\beta_{\hat{\pi} 2}$
Exchangeable Correlation $\rho = 0.3$									
ICC or IWCC estimation									
Bias	-0.002	-0.002	-0.003	0.003	-0.013	0.005	0.005	-0.005	0.001
$s.d.$	0.140	0.417	0.363	0.096	0.121	0.096	0.088	0.118	0.095
$s.e.$	0.139	0.424	0.373	0.096	0.125	0.093	0.091	0.120	0.091
MSE	0.020	0.173	0.131	0.009	0.015	0.009	0.008	0.014	0.009
95%CP	95.8%	96.2%	96.0%	95.8%	97.4%	94.8%	96.4%	96.2%	94.8%
ARE	2.050	1.014	1.154	2.219	1.537	1.668	1.768	1.638	1.596
CC or weighted CC estimation									
Bias	-0.002	0.002	-0.012	0.005	-0.011	0.001	0.005	-0.005	$-0.0^{4}6$
$s.d.$	0.199	0.420	0.390	0.143	0.150	0.124	0.117	0.151	0.120
MSE	0.039	0.176	0.152	0.020	0.023	0.015	0.014	0.023	0.014
ar(1) $\rho = 0.3$									
ICC or IWCC estimation									
Bias	-0.011	0.008	-0.003	-0.016	0.003	0.006	-0.012	$0.0^{3}3$	0.001
$s.d.$	0.098	0.134	0.095	0.099	0.119	0.094	0.102	0.121	0.087
$s.e.$	0.105	0.133	0.096	0.106	0.124	0.092	0.109	0.122	0.089
MSE	0.010	0.018	0.009	0.010	0.014	0.009	0.011	0.015	0.008
95%CP	95.8%	95.0%	95.8%	95.8%	95.8%	94.2%	96.4%	95.8%	95.6%
ARE	2.129	1.572	1.844	3.197	2.089	2.583	2.681	2.116	2.165
CC or weighted CC estimation									
Bias	-0.008	0.014	-0.012	-0.024	0.006	0.007	-0.010	0.002	0.005
$s.d.$	0.143	0.168	0.129	0.177	0.172	0.135	0.167	0.176	0.128
MSE	0.020	0.028	0.017	0.032	0.029	0.018	0.028	0.031	0.017

[a] $0.0^{4}6 = 0.00006$.

$(0.5, 1.0, 1.0)^T$ in the linear regression model, and $\beta^* = (-0.7, 0.1, 0.1)^T$ in the logistic regression model. The data generation procedures are provided in the Appendix D.

For the missing covariates process, we assume that y_{ij} are fully observed and x_{ij1} and x_{ij2} are missing independently. We consider both the MCAR and the MAR cases. We assume that x_{ij1} and x_{ij2} are observed with probability $\pi_{ij}^{(1)}$ and $\pi_{ij}^{(2)}$ respectively. In the MCAR case we let $\pi_{ij}^{(1)} = \pi_1$ and $\pi_{ij}^{(2)} = \pi_2$. For the MAR case, we let observed probability depend on the fully observed response y_{ij}, and we consider two settings: (i) we let $(\pi_{ij}^{(1)}, \pi_{ij}^{(2)}) = (\pi_{1y1}, \pi_{2y1})$ if $Y \geq 0$ (in the linear regression model) or $Y = 1$ (in the logistic regression model) and $(\pi_{ij}^{(1)}, \pi_{ij}^{(2)}) = (\pi_{1y0}, \pi_{2y0})$ otherwise. (ii) we let the observed probabilities depend on the response Y such that $logit(\pi_{ij}^{(1)}) = \alpha_{01} + \alpha_{11} y_{ij}$ and $logit(\pi_{ij}^{(2)}) = \alpha_{02} + \alpha_{12} y_{ij}$.

We set the sample size $m = 500$ and for each setting we generate 500 data sets. For the MCAR case we set $\pi_1 = \pi_2 = 0.50$, for the MAR case we let $(\pi_{1y1}, \pi_{2y1}, \pi_{1y0}, \pi_{2y0}) = (0.5, 0.5, 0.4, 0.4)$ and $(\alpha_{01}, \alpha_{11}) = (\alpha_{02}, \alpha_{12}) = (0.2, 0.2)$; Here the number of distinct missing patterns $q = 2$. We use linear regression models and logistic regression models as surrogate models for the linear model and the logistic model respectively.

The simulation results for the ICC and the IWCC estimates together with the CC and the weighted CC estimates are given in Table 3.1 and Table 3.2 respectively. We see that (i) the biases of the ICC and the IWCC estimates are small; (ii) the means of the standard errors $(s.e)$ calculated based on the asymptotic variance estimator are close to the empirical standard deviations $(s.d.)$; (iii) the estimated 95% coverage probabilities are close to the nominal level; and (iv) compared to the (weighted) CC analysis both the ICC and the IWCC estimates have smaller mean square errors (MSE) and empirical standard deviations.

Chapter 4

Examples

In this section, we will use the generalized unified approach to analysis two real data examples. One is a cross-sectional data, and the other one is a longitudinal data.

4.1 A Case-Control Study of Risk Factors of Hip Fractures

We consider a case-control study of risk factors of hip fractures among male veterans. The study was carried out at the University of Illinois at Chicago College of Medicine (Barengolts et al. 2001; Chen 2004), where a case was matched with a control on age and race, and 25 potential risk factors in addition to age and race were recorded. One major analysis is fitting a logistic regression model with nine potentially important risk factors identified in preliminary exploratory analysis. There are 436 subjects in the study and $q = 9$ distinct missingness patterns in the covariates (each risk factor has a unique missingness pattern). The number of observations in V_0 is 237 and the overall missing percentage is 10.81%.

Table 4.1 Analysis of hip fracture data

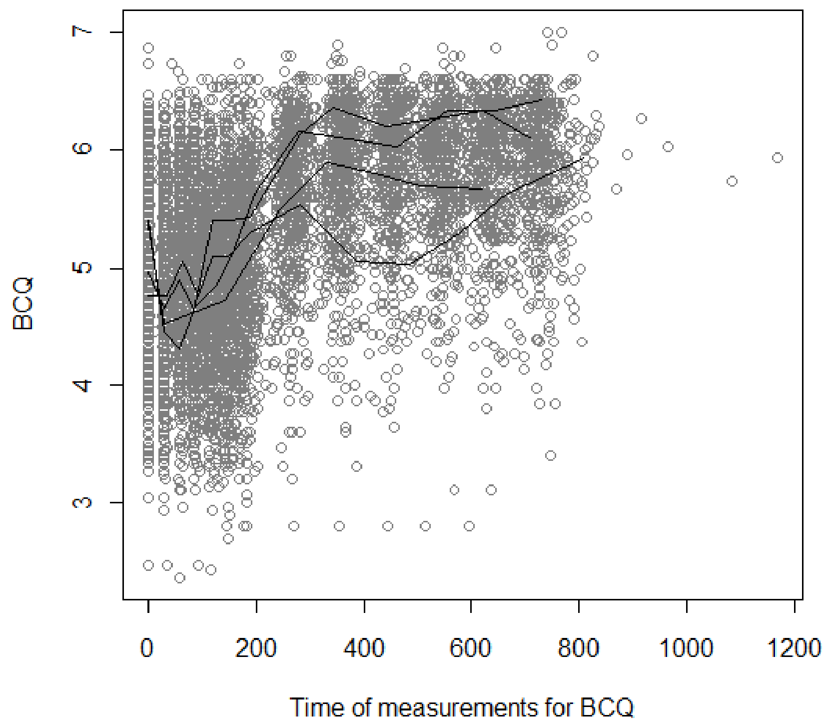
Variable	Weighted CC $\hat{\beta}_{\hat{\pi}}$	Weighted CC $s.e.(\hat{\beta}_{\hat{\pi}})$	IWCC $\tilde{\beta}_{\hat{\pi}}$	IWCC $s.e.(\tilde{\beta}_{\hat{\pi}})$	ARE^*
Etoh	1.380	0.401	1.232	0.351	1.305
Smoke	0.936	0.385	0.799	0.333	1.337
Dementia	2.506	0.672	2.017	0.539	1.554
AntiSeiz	3.275	0.914	3.144	0.786	1.352
LevoT4	1.875	0.734	1.539	0.657	1.248
AntiChol	-1.803	0.727	-2.032	0.669	1.181
BMI	-0.103	0.040	-0.093	0.034	1.384
log(HGB)	-2.618	1.268	-3.429	1.134	1.250
Albumin	-0.904	0.371	-0.792	0.325	1.303

*: $ARE = (s.e.(\hat{\beta}_{\hat{\pi}})/s.e.(\tilde{\beta}_{\hat{\pi}}))^2$

Following Chen (2004) we assume that the covariates are MAR. We estimate the missing data probabilities, π_j, $j = 0, 1, \ldots, 9$, using logistic regression models with hip fracture (the binary outcome variable), age and race as predictors. We report the results of the weighted CC analysis and the IWCC analysis in Table 4.1. We use logistic regression models as the working regression models in the IWCC analysis. We see that the weighted CC estimates and the IWCC estimates are close but the IWCC estimates have relatively smaller $s.e.$'s than the weighted CC estimates.

4.2 A Clinical Study of Breast Cancer

The quality of life is a question of interest in many clinical studies. A Breast Cancer Chemotherapy Questionnaire (BCQ) has been designed for women with stage II breast cancer. The questions selected for this questionnaire were based on common problems and experiences of women undergoing adjutant chemotherapy. The BCQ consists of 30 questions that focus on loss of attractiveness, fatigue, physical symptoms, inconvenience, Fig 4.1 Plot of bcq VS. qol_time



emotional distress, and feelings of hope and support from others. Longitudinal data of 715 patients in NCIC Clinical Trail Group were collected to study the relationship between BCQ and other physical variables. The following are the variables collected in this study.

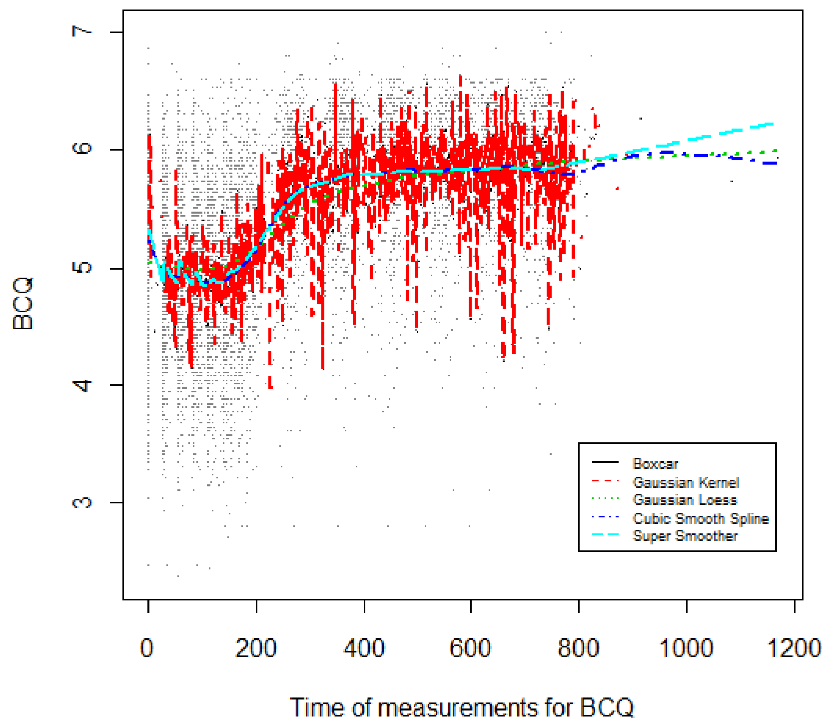
id: Patient identification;

bcq: Average of 30 bcq questions (from 0 to 7);

qol_time: Time (from randomization) of measurements for bcq.

surg_typ: Type of surgery for breast cancer with T=total mastectomy and P=partial mastectomy;

Fig 4.2 Smooth Curves



est_recp: Estrogen receptor status which is a continuous variables with some observations missing;

node_pos: Number of positive nodes;

pth_tcls: Size of the tumor with some observations missing;

all01_co: Treatment group with E=CEF and M=CMF;

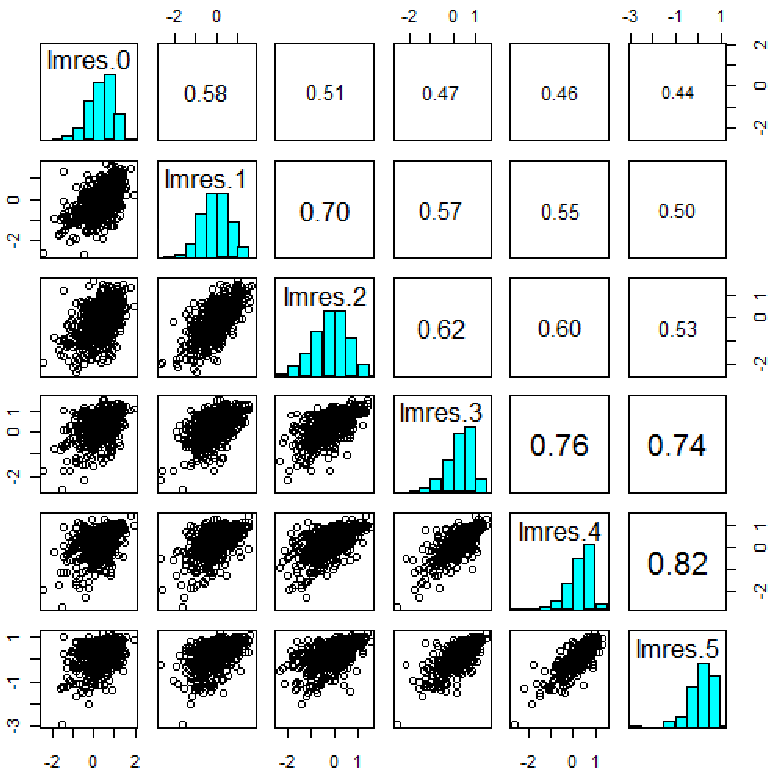
dead: Death indicator with D=dead and A=Alive;

age: Age of patients (in year);

survival: Survival or censoring time (in days);

progress: Relapse-free survival time (in days);

Fig 4.3 Correlation Structure



recurr: Whether patients recurred (Y=yes, N=No).

Fig 4.1 is a plot of bcq and qol_time, and lines of randomly selected four patients. In Fig 4.2, we highlight the average changes in BCQ over time. The scatter plot in Fig 4.3 indicates that (i) the correlation is weaker for observations far away from each other; and (ii) there is some hint that the correlation between observations at time t_i and t_j primarily depends on $|t_i - t_j|$.

We note that variables, est_recp and pth_tcls, have missing values, and they do not have the same missingness pattern. We let $r_{ij}^{(1)}$ and $r_{ij}^{(2)}$ indicate the missingnesses for est_recp and pth_tcls respectively, and let $r_{ij}^{(0)}$ indicate both est_recp and pth_tcls missing.

Table 4.2 Clinical study of breast cancer

	intercept	surg_typ	est_recp	node_pos	pth_tcls	allo1_co	qol_time
				ICC estimates			
$\bar{\beta}$	5.055	-0.005	$-0.0^{3}2$	0.004	0.037	-0.073	0.001
$s.e.$	0.063	0.040	$0.0^{3}2$	0.005	0.031	0.039	$0.0^{4}4$
				CC estimates			
$\hat{\beta}$	5.078	$0.0^{3}4$	$-0.0^{3}2$	0.006	0.021	-0.075	0.001
$s.e.$	0.064	0.041	$0.0^{3}2$	0.005	0.032	0.039	$0.0^{4}4$

[a] $0.0^{3}2 = 0.0002$.

In order to apply our unified method, we must test which missing mechanism $r_{ij}^{(1)}$, $r_{ij}^{(2)}$ and $r_{ij}^{(0)}$ follow. We will first test the null hypothesis that the probability of the missingness for each covariate is independent of the response variable bcq. We want to construct a "score" variable H_{ij} of bcq such that for each $j = 1, \cdots, J$, $H_{ij}(y_{i1}, \cdots, y_{ij})$ is a "score" of the responses up to that time. Following Diggle et al. (2002), we let

$$H_{ij} = H_{ij}(y_{i1}, \cdots, y_{ij}) = \sum_{t=1}^{j} w_t * y_{it}, \text{with} \sum_{t=1}^{j} w_t = 1.$$

The choice of weights, w_ts, reflects analysts' knowledge or judgment about how the past measurement history influences missingness, Some examples are as follows.

(i) Missing influenced immediately by an abnormally high/low measurement:

$$H_{ij} = H_{ij}(y_{i1}, \cdots, y_{ij}) = y_{ij}.$$

(ii) Missing influenced by a sustained sequence of higher/lower measurements:

$$H_{ij} = H_{ij}(y_{i1}, \cdots, y_{ij}) = \frac{1}{j} \sum_{t=1}^{j} y_{it}.$$

We assume

$$logit(\pi_{ij}^{(k)}) = \alpha_k + \beta_k * H_{ij},$$

and we need to test the hypothesis that $\beta_k = 0$, $k = 0, 1, 2$. The corresponding p-values are 0.519, 0.978, and 0.721 respectively, which indicate the MCAR mechanism may be reasonable, thus the generalize unified method for the MCAR case may be applicable.

In this study, we note that there are two covariates est_recp and pth_tcls with missing values, and three distinct missing patterns. The number of subjects in S_0, $S_j, j = 1, 2, 3$ is 620, 626, 704, and 713 respectively. We use the ICC method to estimate the regression parameters. The model of interest is

$$bcq_{ij} = \beta_0 + \beta_1 * surg\_typ_{ij} + \beta_2 * est\_recp_{ij} + \beta_3 * node\_pos_{ij}$$

$$+\beta_4 * pth\_tcls_{ij} + \beta_5 * allo1_{ij} + \beta_6 * qol\_time_{ij} + \epsilon_{ij},$$

and some preliminary analysis indicate that an $Ar(1)$ model may be considered.

Our surrogate models are

$$bcq_{ij} = \gamma_0 + \gamma_1 * surg\_typ_{ij} + \gamma_2 * pth\_tcls_{ij} + \gamma_3 * node\_pos_{ij} + \gamma_4 * allo1_{ij} + \gamma_5 * qol\_time_{ij} + \epsilon_{ij}$$

and

$$bcq_{ij} = \eta_0 + \eta_1 * surg\_typ_{ij} + \eta_2 * est\_recp_{ij} + \eta_3 * node\_pos_{ij} + \eta_4 * qol\_time_{ij} + \epsilon_{ij}$$

respectively, and the correlation structure for each surrogate model is $Ar(1)$.

Table 4.2 lists the results of ICC estimates and the CC estimates. We see that (i) the ICC estimates are close to the CC estimates, and (ii) the standard errors $(s.e.)$ of the ICC estimators are consistently smaller than the corresponding $s.e.$'s of the CC estimators.

Chapter 5

Discussion and Future Research

The proposed generalized unified parametric methods, the ICC and the IWCC, provide convenient estimation procedures for regression models with covariates missing in arbitrary nonmonotone patterns. It uses all the observed data to compute estimates which are more efficient than the (weighted) CC analysis. It is computationally simple and does not require an iteration procedure. When the covariates have a simple monotone missingness pattern Chen and Chen (2000) showed that the unified estimation method can be as efficient as the semiparametric efficient method of Robins et al. (1994). We note that the estimation efficiency of the generalized unified estimation methods depend on the selected working parametric regression models. Further investigations on selecting 'optimal' working parametric regression models is one of my future research topics.

A limitation of the generalized unified parametric method is that it requires MCAR data or MAR data with known selection probabilities or with known true models for the selection probabilities. One exception is the case where the selection probabilities only depend on the fully observed covariates X_1 but do not depend on the response variable Y, then the proposed ICC will be consistent if we include X_1 in each working regression model. Extending the IWCC method to MAR data with unknown selection probabilities or unknown true models for the selection probabilities requires constructing sufficient models to estimate the selection probabilities. Zhao et al. (1996) gave several recommendations for modeling the selection probabilities. For example, one can use some 'stable' 'saturated' models or consider nonparametric estimates for categorical and/or continuous variables. The semiparametric weighted estimation with selection probabilities estimated by kernel smoothers (Wang et al. 1997) may be considered to achieve more general applications of the IWCC method for MAR data which is another topic in my future research.

The missingness in the longitudinal data may be caused by many factors, for example some covariates or historical response data. Models for the missing data probability are very complex. In the future, I will investigate how to obtain robust models for the missing probability in longitudinal data.

I am also interested in extending the generalized unified methods to deal with other statistical models, for example, partial linear model and Cox proportional hazard model, with arbitrary nonmonotone missing covariate data.

R is a free software environment for statistical computing and graphics which include many packages for statistical analysis. However, there are few packages for the missing data problems. In the future I will develop an R package based on the proposed generalized unified approach so that the generalized unified approach can be widely used in statistical analysis with missing data.

Appendix

Appendix A: Regularity Conditions

Let ξ be a vector of the parameters, including the parameters of interest and nuisance parameters, ξ^* be the true value of ξ, and $U_i(\xi)$ be the estimating functions. The regularity conditions are as follows.

(a) ξ^* exists and lies in the interior of a compact parameter space;

(b) $U_i(\xi)$ has zero mean only at true value ξ^*;

(c) There is a neighborhood of ξ^*, $N_\delta(\xi^*)$, such that $\mathbb{E}\{\sup_{\xi\in \mathrm{N}_\delta(\xi^*)}||\mathrm{U_i}(\xi)||\}$, $\mathbb{E}\{\sup_{\xi\in \mathrm{N}_\delta(\xi^*)}||\partial\mathrm{U_i}(\xi)/\partial\xi||\}$ and $\mathbb{E}\{\sup_{\xi\in \mathrm{N}_\delta(\xi^*)}||\mathrm{U_i}(\xi)\mathrm{U_i^T}(\xi)||\}$ are all finite, where $||M|| = (\Sigma_{ij} m_{ij}^2)^{1/2}$ for any matrix M with elements m_{ij};

(d) $var(U_i(\xi))$ is finite and positive definite, and $E[\partial U_i(\xi)/\partial\xi]$ exists and is invertible.

Appendix B: Asymptotic Properties

Cross-Sectional Study

For the MCAR case, following Foutz (1977) , Chen and Chen (2000) proved the consistency and the asymptotic normality of $\bar{\beta}$ under the regularity conditions (a)-(d) in Appendix A. The proof can be directly extended to the generalized unified estimator by letting $U(\theta) = (S_0^T(\beta), S_Q^T(\gamma))^T$ and assuming the conditions (a)-(d) in Appendix A. hold for $U(\theta)$.

In the MAR case, there are two sets of estimation equations for γ with different weights (see equations (2.6) and (2.7)). We let $U(\theta) = (R_0/\pi_0)(S_0^T(\beta), S_Q^T(\gamma))^T$ and $\bar{U}(\gamma) = ((R_1/\pi_1)S_1^T(\gamma_1), \cdots, (R_q/\pi_q)S_q^T(\gamma_q))^T$. Following Chen and Chen (2000), when $U(\theta)$ satisfies conditions (a)-(c) in Appendix A. we can obtain the consistency of $\hat{\beta}_\pi$ and $\hat{\gamma}_\pi$ by the uniform law of large numbers and the inverse function theory. Similarly, we can derive that $\bar{\gamma}_\pi$ is a consistent estimator for γ^* when $\bar{U}(\gamma)$ satisfies conditions (a)-(c). Then under conditions (c) and (d) for $U(\theta)$ it can be shown that $\hat{D}_{\pi 0}^{-1}\hat{C}_{\pi 01}\hat{C}_{\pi 11}^{-1}\hat{D}_{\pi 1}$ converges uniformly to the finite matrix $D_0^{-1}C_{\pi 01}C_{\pi 11}^{-1}D_1$ with probability going to 1. Therefore $\bar{\beta}_\pi$ is consistent for β^*. Finally, by the central limit theorem and Slutsky's theorem, $(\hat{\beta}_\pi, \hat{\gamma}_\pi)$ and $\bar{\gamma}_\pi$ are asymptotically normal, and furthermore $\bar{\beta}_\pi$ is asymptotically normal. For the IWCC using the estimated selection probabilities the asymptotic normality and consistency can be derived similarly by combining $S_{\pi 0}(\alpha_0)$ with $U(\theta)$ and $(S_{\pi i1}^T(\alpha_1), \cdots, S_{\pi iq}^T(\alpha_q))^T$ with $\bar{U}(\gamma)$.

Let $\mathcal{V}(.)$ and $\mathcal{C}(.,.)$ denote the asymptotic variance and covariance respectively. We note that

$$n^{1/2}(\hat{\beta}-\beta^*) = n^{-1/2}E[\partial \mathbf{S}_{i0}(\beta^*)/\partial\beta]^{-1}\sum_{i\in V_0}\{\mathbf{S}_{i0}(\beta^*)\} + o_p(1),$$

$$n^{1/2}(\hat{\gamma}_k-\gamma_k^*) = n^{-1/2}E[\partial \mathbf{S}_{ik}(\gamma_k^*)/\partial\gamma]^{-1}\sum_{i\in V_0}\{\mathbf{S}_{ik}(\gamma_k^*)\} + o_p(1),$$

and

$$n_k^{1/2}(\bar{\gamma}_k-\gamma_k^*) = n_k^{-1/2}E[\partial \mathbf{S}_{ik}(\gamma^*)/\partial\gamma]^{-1}\sum_{i\in V_k}\{\mathbf{S}_{ik}(\gamma_k^*)\} + o_p(1).$$

Let $D_1 = diag\{d_{11},\cdots,d_{qq}\}$, $C_{01} = (c_{01}, c_{02}, ..., c_{0q})$, and

$$C_{11} = \begin{pmatrix} c_{11} & \cdots & c_{1q} \\ ... & & \\ c_{q1} & \cdots & c_{qq} \end{pmatrix}.$$

We can get

$$\mathcal{V}(n^{1/2}\hat{\beta}) = D_0^{-1}C_{00}D_0^{-1T},$$

$$\mathcal{V}(n^{1/2}\bar{\gamma}_k) = \frac{\pi_0}{\pi_k}\mathcal{V}(n^{1/2}\hat{\gamma}_k) = \frac{\pi_0}{\pi_k}d_{kk}^{-1}c_{kk}d_{kk}^{T-1},$$

$$\mathcal{C}(n^{1/2}\hat{\beta}, n^{1/2}\bar{\gamma}_k) = \frac{\pi_0}{\pi_k}\mathcal{C}(n^{1/2}\hat{\beta}, n^{1/2}\hat{\gamma}_k) = \frac{\pi_0}{\pi_k}D_0^{-1}c_{0k}d_{kk}^{-1T},$$

$$\mathcal{C}(n^{1/2}\hat{\gamma}_k, n^{1/2}\bar{\gamma}_h) = \frac{\pi_0}{\pi_h}\mathcal{C}(n^{1/2}\hat{\gamma}_k, n^{1/2}\hat{\gamma}_h) = \frac{\pi_0}{\pi_h}d_{kk}^{-1}c_{kh}d_{hh}^{-1T},$$

$$\mathcal{C}(n^{1/2}\bar{\gamma}_k, n^{1/2}\bar{\gamma}_h) = \frac{\pi_0\pi_{kh}}{\pi_k\pi_h}d_{kk}^{-1}c_{kh}d_{hh}^{-1T}.$$

The asymptotic variance of $\bar{\beta}$ thus follows.

We note that

$$N^{1/2}(\hat{\beta}_\pi-\beta^*) = N^{-1/2}E[\partial\frac{R_{i0}}{\pi_{i0}}S_{i0}(\beta^*)/\partial\beta]^{-1}\sum_{i=1}^{N}\{\frac{R_{i0}}{\pi_{i0}}S_{i0}(\beta^*)\} + o_p(1),$$

$$N^{1/2}(\hat{\gamma}_\pi - \gamma^*) = N^{-1/2}E[\partial \frac{R_{i0}}{\pi_{i0}} S_{iQ}(\gamma^*)/\partial\gamma]^{-1} \sum_{i=1}^{N} \{\frac{R_{i0}}{\pi_{i0}} S_{iQ}(\gamma^*)\} + o_p(1),$$

and

$$N^{1/2}(\bar{\gamma}_\pi - \gamma^*) = N^{-1/2}E[\partial S_{\pi iQ}(\gamma^*)/\partial\gamma]^{-1} \sum_{i=1}^{N} \{S_{\pi iQ}(\gamma^*)\} + o_p(1).$$

We can get

$$\begin{aligned}
&\mathcal{V}(N^{1/2}\hat{\beta}_\pi)\\
&= E[\partial \frac{R_{i0}}{\pi_{i0}} S_{i0}(\beta^*)/\partial\beta]^{-1} E[\frac{R_{i0}}{\pi_{i0}^2} S_{i0}(\beta^*) S_{i0}^T(\beta^*)] E[\partial \frac{R_{i0}}{\pi_{i0}} S_{i0}^T(\beta^*)/\partial\beta]^{-1}\\
&= D_0^{-1} C_{\pi 00} D_0^{T-1},\\
&\mathcal{V}(N^{1/2}\hat{\gamma}_\pi)\\
&= E[\partial \frac{R_{i0}}{\pi_{i0}} S_{iQ}(\gamma^*)/\partial\gamma]^{-1} E[\frac{R_{i0}}{\pi_{i0}^2} S_{iQ}(\gamma^*) S_{iQ}^T(\gamma^*)] E[\partial \frac{R_{i0}}{\pi_{i0}} S_{iQ}^T(\gamma^*)/\partial\gamma]^{-1}\\
&= D_1^{-1} C_{\pi 11} D_1^{T-1},\\
&\mathcal{V}(N^{1/2}\bar{\gamma}_\pi)\\
&= E[\partial S_{\pi iQ}(\gamma^*)/\partial\gamma]^{-1} E[S_{\pi iQ}(\gamma^*) S_{\pi iQ}^T(\gamma^*)] E[\partial S_{\pi iQ}^T(\gamma^*)/\partial\gamma]^{-1}\\
&= D_1^{-1} C_{\pi 22} D_1^{T-1},\\
&\mathcal{C}(N^{1/2}\hat{\beta}_\pi, N^{1/2}\hat{\gamma}_\pi)\\
&= E[\partial \frac{R_{i0}}{\pi_{i0}} S_{i0}(\beta^*)/\partial\beta]^{-1} E[\frac{R_{i0}}{\pi_{i0}^2} S_{i0}(\beta^*) S_{iQ}^T(\gamma^*)] E[\partial \frac{R_{i0}}{\pi_{i0}} S_{iQ}^T(\gamma^*)/\partial\gamma]^{-1}\\
&= D_0^{-1} C_{\pi 01} D_1^{T-1},\\
&\mathcal{C}(N^{1/2}\hat{\beta}_\pi, N^{1/2}\bar{\gamma}_\pi)\\
&= E[\partial \frac{R_{i0}}{\pi_{i0}} S_{i0}(\beta^*)/\partial\beta]^{-1} E[\frac{R_{i0}}{\pi_{i0}} S_{i0}(\beta^*) S_{\pi iQ}^T(\gamma^*)] E[\partial S_{\pi iQ}^T(\gamma^*)/\partial\gamma]^{-1}\\
&= D_0^{-1} C_{\pi 02} D_1^{T-1},
\end{aligned}$$

$$
\begin{aligned}
&\mathcal{C}(N^{1/2}\hat{\gamma}_\pi, N^{1/2}\bar{\gamma}_\pi) \\
&= E[\partial \frac{R_{i0}}{\pi_{i0}} S_{i0}(\beta^*)/\partial\beta]^{-1} E[\frac{R_{i0}}{\pi_{i0}} S_{iQ}(\gamma^*) S_{\pi iQ}^T(\gamma^*)] E[\partial S_{\pi iQ}^T(\gamma^*)/\partial\gamma]^{-1} \\
&= D_1^{-1} C_{\pi 12} D_1^{T-1},
\end{aligned}
$$

where $E[\partial \frac{R_{i0}}{\pi_{i0}} S_{iQ}(\gamma^*)/\partial\gamma] = E[\partial S_{\pi iQ}(\gamma^*)/\partial\gamma] = D_1$. The asymptotic variance of $\bar{\beta}_\pi$ thus follows.

We note that

$$
N^{1/2}(\hat{\beta}_{\hat{\pi}} - \beta^*) = N^{-1/2} E[\partial \frac{R_{i0}}{\pi_{i0}} S_{i0}(\beta^*)/\partial\beta]^{-1} \sum_{i=1}^{N} Res(\frac{R_{i0}}{\pi_{i0}} S_{i0}(\beta^*), H_{\pi i0}(\alpha_0^*)) + o_p(1),
$$

$$
N^{1/2}(\hat{\gamma}_{\hat{\pi}} - \gamma^*) = N^{-1/2} E[\partial \frac{R_{i0}}{\pi_{i0}} S_{iQ}(\gamma^*)/\partial\gamma]^{-1} \sum_{i=1}^{N} Res(\frac{R_{i0}}{\pi_{i0}} S_{iQ}(\gamma^*), H_{\pi i0}(\alpha_0^*)) + o_p(1),
$$

and

$$
N^{1/2}(\bar{\gamma}_{\hat{\pi}} - \gamma^*) = N^{-1/2} E[\partial S_{\pi iQ}(\gamma^*, \alpha^*)/\partial\gamma]^{-1} \sum_{i=1}^{N} Res(S_{\pi iQ}(\gamma^*), H_{\pi iQ}(\alpha_Q^*)) + o_p(1).
$$

We can get

$$
\begin{aligned}
&\mathcal{V}(N^{1/2}\hat{\beta}_{\hat{\pi}}) \\
&= E[\partial \frac{R_{i0}}{\pi_{i0}} S_{i0}(\beta^*)/\partial\beta]^{-1} \\
&E[Res(\frac{R_{i0}}{\pi_{i0}} S_{i0}(\beta^*), H_{\pi i0}(\alpha_0^*)) Res^T(\frac{R_{i0}}{\pi_{i0}} S_{i0}(\beta^*), H_{\pi i0}(\alpha_0^*))] E[\partial \frac{R_{i0}}{\pi_{i0}} S_{i0}^T(\beta^*)/\partial\beta]^{-1} \\
&= D_0^{-1} C_{\hat{\pi}00} D_0^{T-1},
\end{aligned}
$$

$\mathcal{V}(N^{1/2}\hat{\gamma}_{\hat{\pi}})$

$= E[\partial \frac{R_{i0}}{\pi_{i0}} S_{iQ}(\gamma^*)/\partial\gamma]^{-1}$

$E[Res(\frac{R_{i0}}{\pi_{i0}} S_{iQ}(\gamma^*), H_{\pi i0}(\alpha_0^*)) Res^T(\frac{R_{i0}}{\pi_{i0}} S_{iQ}(\gamma^*), H_{\pi i0}(\alpha_0^*))] E[\partial \frac{R_{i0}}{\pi_{i0}} S_{iQ}^T(\gamma^*)/\partial\gamma]^{-1}$

$= D_1^{-1} C_{\hat{\pi}11} D_1^{T-1},$

$\mathcal{V}(N^{1/2}\bar{\gamma}_{\hat{\pi}})$

$= E[\partial S_{\pi iQ}(\gamma^*)/\partial\gamma]^{-1}$

$E[Res(S_{\pi iQ}(\gamma^*), H_{\pi iQ}(\alpha_Q^*)) Res^T(S_{\pi iQ}(\gamma^*), H_{\pi iQ}(\alpha_Q^*))] E[\partial S_{\pi iQ}(\gamma^*)/\partial\gamma]^{-1}$

$= D_1^{-1} C_{\hat{\pi}22} D_1^{T-1},$

$\mathcal{C}(N^{1/2}\hat{\beta}_{\hat{\pi}}, N^{1/2}\hat{\gamma}_{\hat{\pi}})$

$= E[\partial \frac{R_{i0}}{\pi_{i0}} S_{i0}(\beta^*)/\partial\beta]^{-1}$

$E[Res(\frac{R_{i0}}{\pi_{i0}} S_{i0}(\beta^*), H_{\pi i0}(\alpha_0^*)) Res^T(\frac{R_{i0}}{\pi_{i0}} S_{iQ}(\gamma^*), H_{\pi i0}(\alpha_0^*))] E[\partial \frac{R_{i0}}{\pi_{i0}} S_{iQ}^T(\gamma^*)/\partial\gamma]^{-1}$

$= D_0^{-1} C_{\hat{\pi}01} D_1^{T-1},$

$\mathcal{C}(N^{1/2}\hat{\beta}_{\hat{\pi}}, N^{1/2}\bar{\gamma}_{\hat{\pi}})$

$= E[\partial \frac{R_{i0}}{\pi_{i0}} S_{i0}(\beta^*)/\partial\beta]^{-1} E[Res(\frac{R_{i0}}{\pi_{i0}} S_{i0}(\beta^*), H_{\pi i0}(\alpha_0^*)) Res^T(S_{\pi iQ}(\gamma^*), H_{\pi iQ}(\alpha_Q^*))]$

$E[\partial S_{\pi iQ}(\gamma^*)/\partial\gamma]^{-1}$

$= D_0^{-1} C_{\hat{\pi}02} D_1^{T-1},$

$\mathcal{C}(N^{1/2}\hat{\gamma}_{\hat{\pi}}, N^{1/2}\bar{\gamma}_{\hat{\pi}})$

$= E[\partial \frac{R_{i0}}{\pi_{i0}} S_{i0}(\beta^*)/\partial\beta]^{-1}$

$E[Res(\frac{R_{i0}}{\pi_{i0}} S_{iQ}(\gamma^*), H_{\pi i0}(\alpha_0^*)) Res^T(S_{\pi iQ}(\gamma^*), H_{\pi iQ}(\alpha_Q^*))] E[\partial S_{\pi iQ}(\gamma^*)/\partial\gamma]^{-1}$

$= D_1^{-1} C_{\hat{\pi}12} D_1^{T-1},$

where $E[\partial \frac{R_{i0}}{\pi_{i0}} S_{iQ}(\gamma^*)/\partial\gamma] = E[\partial S_{\pi iQ}(\gamma^*)/\partial\gamma] = D_1$. The asymptotic variance of $\hat{\beta}_{\hat{\pi}}$ thus follows.

Longitudinal Study

For the MCAR case, under the regularity conditions (a)-(c) in Appendix A, the consistence of $(\hat{\beta}, \hat{\gamma})$ and $\bar{\gamma}$ can be obtained by Theorem 2.6 of Newey (1994). Under the conditions (c) and (d), with probability going to 1, $\hat{\Gamma}_{00}^{-1}\hat{\Sigma}_{01}\hat{\Sigma}_{22}^{-1}\hat{\Gamma}_{11}$ converges uniformly to the finite matrix $\Gamma_{00}^{-1}\Sigma_{01}\Sigma_{22}^{-1}\Gamma_{11}$ by Theorem 4.5 of Newey (1994). The consistency of $\bar{\beta}$ thus follows. The asymptotic normality of $\hat{\beta}$, $\hat{\gamma}$ and $\bar{\gamma}$ can be obtained under the conditions (a)-(d) by the Theorem 3.4 of Newey (1994). The asymptotic normality of $\bar{\beta}$ will follow by Slutsky's theorem.

For the MAR case, there are two sets estimation equations for γ with different weights (see equation (3.13) and (3.14)). When $(U_{\pi i0}^T(\beta), U_{\pi iQ}^T(\gamma))^T$ satisfies conditions (a)-(c), we can obtain the consistency of $\hat{\beta}_\pi$ and $\hat{\gamma}_\pi$ by Theorem 2.6 of Newey (1994). Similarly, we can drive that $\bar{\gamma}_\pi$ is a consistent estimator for γ^* when $\bar{U}_{\pi iQ}(\gamma)$ satisfies conditions (a)-(c). Then under conditions (c) and (d) for $(U_{\pi i0}^T(\beta), U_{\pi iQ}^T(\gamma))^T$ it can be shown that $\hat{\Gamma}_{\pi 00}^{-1}\hat{\Sigma}_{\pi 01}\hat{\Sigma}_{\pi 11}^{-1}\hat{\Gamma}_{\pi 11}$ converges uniformly to the finite matrix $\Gamma_{00}^{-1}\Sigma_{\pi 01}\Sigma_{\pi 11}^{-1}\Gamma_{11}$ with probability going to 1. Therefore $\bar{\beta}_\pi$ is consistent for β^*. Finally, by the central limit theorem and Slutsky's theorem, $\hat{\beta}_\pi$, $\hat{\gamma}_\pi$ and $\bar{\gamma}_\pi$ are asymptotically normal, and furthermore $\bar{\beta}_\pi$ is asymptotically normal. For the IWCC using the estimated selection probabilities the asymptotic normality and consistency can be derived similarly by combining $H_{i0}(\alpha_0)$

with$(U^T_{\hat{\pi}i0}(\beta), U^T_{\hat{\pi}iQ}(\gamma))^T$ and $H_{iQ}(\alpha)$ with $\bar{U}_{\hat{\pi}iQ}(\gamma)$.

To derive the asymptotic variance of $\bar{\beta}$ in (3.11), we only need to consider the asymptotic variance and covariance between $\hat{\beta}$, $\hat{\gamma}$ and $\bar{\gamma}$.

We have

$$m_0^{1/2}(\hat{\beta} - \beta^*) = m_0^{-1/2} E[\partial U_{i0}(\beta^*)/\partial \beta] \sum_{i=1}^{m_0} \{U_{i0}(\beta^*)\} + o_p(1),$$
$$m_0^{1/2}(\hat{\gamma}_k - \gamma_k^*) = m_0^{-1/2} E[\partial U_{ik}(\gamma_k^*)/\partial \gamma_k] \sum_{i=1}^{m_0} \{U_{ik}(\gamma_k^*)\} + o_p(1),$$

and

$$m_k^{1/2}(\bar{\gamma}_k - \gamma_k^*) = m_k^{-1/2} E[\partial \bar{U}_{ik}(\gamma_k^*)/\partial \gamma_k] \sum_{i=1}^{m_k} \{\bar{U}_{ik}(\gamma_k^*)\} + o_p(1).$$

So we can obtain

$$\mathcal{V}(m_0^{1/2}(\hat{\beta})) = E[\partial U_{i0}(\beta^*)/\partial \beta] E[U_{i0}(\beta^*) U_{i0}^T(\beta^*)] E[\partial U_{i0}^T(\beta^*)/\partial \beta],$$

$$\mathcal{V}(m_0^{1/2}(\hat{\gamma}_k)) = E[\partial U_{ik}(\gamma_k^*)/\partial \gamma_k] E[U_{ik}(\gamma_k^*) U_{ik}^T(\gamma_k^*)] E[\partial U_{ik}^T(\gamma_k^*)/\partial \gamma_k],$$

$$\mathcal{V}(m_0^{1/2}(\bar{\gamma}_k)) = \frac{m_0}{m_k} E[\partial \bar{U}_{ik}(\gamma_k^*)/\partial \gamma_k] E[\bar{U}_{ik}(\gamma_k^*) U_{ik}^T(\gamma_k^*)] E[\partial \bar{U}_{ik}^T(\gamma_k^*)/\partial \gamma_k],$$

$$\mathcal{C}(m_0^{1/2}\hat{\beta}, m_0^{1/2}\hat{\gamma}_k)) = E[\partial U_{i0}(\beta^*)/\partial \beta] E[U_{i0}(\beta^*) U_{ik}^T(\gamma_k^*)] E[\partial U_{ik}^T(\gamma_k^*)/\partial \gamma_k],$$

$$\mathcal{C}(m_0^{1/2}\hat{\beta}, m_0^{1/2}\bar{\gamma}_k)) = \frac{m_0}{m_k} E[\partial U_{i0}(\beta^*)/\partial \beta] E[U_{i0}(\beta^*) \bar{U}_{ik}^T(\gamma_k^*)] E[\partial \bar{U}_{ik}^T(\gamma_k^*)/\partial \gamma_k],$$

$$Cov(m_0^{1/2}\hat{\gamma}_k, m_0^{1/2}\hat{\gamma}_h)) = E[\partial U_{ik}(\gamma_k^*)/\partial \gamma_k] E[U_{ik}(\gamma_k^*) U_{ih}^T(\gamma_h^*)] E[\partial U_{ih}^T(\gamma_h^*)/\partial \gamma_h],$$

$$\mathcal{C}(m_0^{1/2}\hat{\gamma}_k, m_0^{1/2}\bar{\gamma}_h)) = \frac{m_0}{m_h} E[\partial U_{ik}(\gamma_k^*)/\partial \gamma_k] E[U_{ik}(\gamma_k^*) \bar{U}_{ih}^T(\gamma_h^*)] E[\partial \bar{U}_{ih}^T(\gamma_h^*)/\partial \gamma_h],$$

$$\mathcal{C}(m_0^{1/2}\bar{\gamma}_k, m_0^{1/2}\bar{\gamma}_h)) = \frac{m_0 * m_{kh}}{m_k * m_h} E[\partial \bar{U}_{ik}(\gamma_k^*)/\partial \gamma_k] E[\bar{U}_{ik}(\gamma_k^*) \bar{U}_{ih}^T(\gamma_h^*)] E[\partial \bar{U}_{ih}^T(\gamma_h^*)/\partial \gamma_h].$$

In fact U_{i0} and U_i^f are the same main models, and U_{ik}, $\bar{U}_{ik}$, and U_{ik}^f are the same surrogate models, where super-script f denotes the regression model without missing data, so we can obtain that

$$\mathcal{V}(m_0^{1/2}(\hat{\beta})) = E[\partial U_i^f(\beta^*)/\partial\beta]E[U_i^f(\beta^*)U_i^{fT}(\beta^*)]E[\partial U_i^{fT}(\beta^*)/\partial\beta],$$

$$\mathcal{V}(m_0^{1/2}(\hat{\gamma}_k)) = E[\partial U_{ik}^f(\gamma_k^*)/\partial\gamma_k]E[U_{ik}^f(\gamma_k^*)U_{ik}^{fT}(\gamma_k^*)]E[\partial U_{ik}^{fT}(\gamma_k^*)/\partial\gamma_k],$$

$$\mathcal{V}(m_0^{1/2}(\bar{\gamma}_k)) = \frac{m_0}{m_k}E[\partial U_{ik}^f(\gamma_k^*)/\partial\gamma_k]E[U_{ik}^f(\gamma_k^*)U_{ik}^{fT}(\gamma_k^*)]E[\partial U_{ik}^{fT}(\gamma_k^*)/\partial\gamma_k],$$

$$\mathcal{C}(m_0^{1/2}\hat{\beta}, m_0^{1/2}\hat{\gamma}_k)) = E[\partial U_i^f(\beta^*)/\partial\beta]E[U_i^f(\beta^*)U_{ik}^{fT}(\gamma_k^*)]E[\partial U_{ik}^{fT}(\gamma_k^*)/\partial\gamma_k],$$

$$\mathcal{C}(m_0^{1/2}\hat{\beta}, m_0^{1/2}\bar{\gamma}_k)) = \frac{m_0}{m_k}E[\partial U_i^f(\beta^*)/\partial\beta]E[U_i^f(\beta^*)U_{ik}^{fT}(\gamma_k^*)]E[\partial U_{ik}^{fT}(\gamma_k^*)/\partial\gamma_k],$$

$$\mathcal{C}(m_0^{1/2}\hat{\gamma}_k, m_0^{1/2}\bar{\gamma}_h)) = \frac{m_0}{m_h}E[\partial U_{ik}^f(\gamma_k^*)/\partial\gamma_k]E[U_{ik}^f(\gamma_k^*)U_{ih}^{fT}(\gamma_h^*)]E[\partial U_{ih}^{fT}(\gamma_h^*)/\partial\gamma_h],$$

$$\mathcal{C}(m_0^{1/2}\bar{\gamma}_k, m_0^{1/2}\bar{\gamma}_h)) = \frac{m_0 * m_{kh}}{m_k * m_h}E[\partial U_{ik}^f(\gamma_k^*)/\partial\gamma_k]E[U_{ik}^f(\gamma_k^*)U_{ih}^{fT}(\gamma_h^*)]E[\partial U_{ih}^{fT}(\gamma_h^*)/\partial\gamma_h]$$

since the missingness is MCAR.

Furthermore, we can obtain that

$$\mathcal{V}(m_0^{1/2}(\bar{\gamma}_k)) = \frac{m_0}{m_k}\mathcal{V}(m_0^{1/2}(\hat{\gamma}_{\pi k})),$$

$$\mathcal{C}(m_0^{1/2}\hat{\beta}, m_0^{1/2}\bar{\gamma}_k)) = \frac{m_0}{m_k}\mathcal{C}(m_0^{1/2}\hat{\beta}_\pi, m_0^{1/2}\hat{\gamma}_{\pi k})),$$

$$\mathcal{C}(m_0^{1/2}\bar{\gamma}_k, m_0^{1/2}\bar{\gamma}_h)) = \frac{m_0 * m_{kh}}{m_k * m_h}\mathcal{C}(m_0^{1/2}\hat{\gamma}_{\pi k}, m_0^{1/2}\hat{\gamma}_{\pi h})).$$

Finally following the procedure similar to that of the MCAR case, we can derive the variance of $\bar{\beta}$.

The proof of the asymptotic variance in (3.16) is as follows.

We know

$$M^{1/2}(\hat{\beta}_\pi - \beta^*) = M^{-1/2} E[\partial U_{\pi i0}(\beta^*)/\partial\beta] \sum_{i=1}^{M} \{U_{\pi 0i}(\beta^*)\} + o_p(1),$$

$$M^{1/2}(\hat{\gamma}_\pi - \gamma^*) = M^{-1/2} E[\partial U_{\pi iQ}(\gamma^*)/\partial\gamma] \sum_{i=1}^{M} \{U_{\pi iQ}(\gamma^*)\} + o_p(1),$$

and

$$M^{1/2}(\bar{\gamma}_\pi - \gamma^*) = M^{-1/2} E[\partial \bar{U}_{\pi iQ}(\gamma^*)/\partial\gamma] \sum_{i=1}^{M} \{\bar{U}_{\pi iQ}(\gamma^*)\} + o_p(1).$$

So we can get

$$\begin{aligned}
\mathcal{V}(M^{1/2}\hat{\beta}) = &\quad E[\partial U_{\pi i0}(\beta^*)/\partial\beta] E[U_{\pi 0i}(\beta^*) U_{\pi i0}^T(\beta^*)] E[\partial U_{\pi i0}^T(\beta^*)/\partial\beta] = \Gamma_{00}^{-1}\Sigma_{\pi 00}\Gamma_{00}^{T-1},\\
\mathcal{V}(M^{1/2}\hat{\gamma}) = &\quad E[\partial U_{\pi iQ}(\gamma^*)/\partial\gamma] E[U_{\pi iQ}(\gamma^*) U_{\pi iQ}^T(\gamma^*)] E[\partial U_{\pi iQ}^T(\gamma^*)/\partial\gamma] = \Gamma_{11}^{-1}\Sigma_{\pi 11}\Gamma_{11}^{T-1},\\
\mathcal{V}(M^{1/2}\bar{\gamma}) = &\quad E[\partial \bar{U}_{\pi iQ}(\gamma^*)/\partial\gamma] E[\bar{U}_{\pi iQ}(\gamma^*) \bar{U}_{\pi iQ}^T(\gamma^*)] E[\partial \bar{U}_{\pi iQ}(\gamma^*)/\partial\gamma] = \Gamma_{11}^{-1}\Sigma_{\pi 22}\Gamma_{11}^{T-1},\\
\mathcal{C}(M^{1/2}\hat{\beta}, M^{1/2}\hat{\gamma}) = &\quad E[\partial U_{\pi i0}(\beta^*)/\partial\beta] E[U_{\pi i0}(\beta^*) U_{\pi iQ}^T(\gamma^*)] E[\partial U_{\pi iQ}^T(\gamma^*)/\partial\gamma] = \Gamma_{00}^{-1}\Sigma_{\pi 01}\Gamma_{11}^{T-1},\\
\mathcal{C}(M^{1/2}\hat{\beta}, M^{1/2}\bar{\gamma}) = &\quad E[\partial U_{\pi i0}(\beta^*)/\partial\beta] E[U_{\pi i0}(\beta^*) \bar{U}_{iQ}^T(\gamma^*)] E[\partial \bar{U}_{\pi iQ}^T(\gamma^*)/\partial\gamma] = \Gamma_{00}^{-1}\Sigma_{\pi 02}\Gamma_{11}^{T-1},\\
\mathcal{C}(M^{1/2}\hat{\gamma}, M^{1/2}\bar{\gamma}) = &\quad E[\partial U_{\pi iQ}(\gamma^*)/\partial\gamma] E[U_{\pi iQ}(\gamma^*) \bar{U}_{\pi iQ}^T(\gamma^*)] E[\partial \bar{U}_{\pi iQ}^T(\gamma^*)/\partial\gamma] = \Gamma_{11}^{-1}\Sigma_{\pi 12}\Gamma_{11}^{T-1},
\end{aligned}$$

where $E[\partial U_{\pi iQ}(\gamma^*)/\partial\gamma] = E[\partial \bar{U}_{\pi iQ}(\gamma^*)/\partial\gamma] = \Gamma_{11}$.

The following is the proof of the asymptotic variance in (3.24).

We have

$$M^{1/2}(\hat{\beta}_{\hat{\pi}} - \beta^*) = M^{-1/2} E[\partial U_{\hat{\pi} i0}(\beta^*)/\partial\beta] \sum_{i=1}^{M} \{Res(U_{\hat{\pi} i0}(\beta^*, \alpha_0^*), H_{i0}(\alpha_0^*))\} + o_p(1),$$

$$M^{1/2}(\hat{\gamma}_{\hat{\pi}} - \gamma^*) = M^{-1/2} E[\partial U_{\hat{\pi} iQ}(\gamma^*)/\partial\gamma] \sum_{i=1}^{M} \{Res(U_{\hat{\pi} iQ}(\gamma^*, \alpha_0^*), H_{i0}(\alpha_0^*))\} + o_p(1).$$

and

$$M^{1/2}(\bar{\gamma}_{\hat{\pi}} - \gamma^*) = M^{-1/2} E[\partial \bar{U}_{\hat{\pi} iQ}(\gamma^*)/\partial \gamma] \sum_{i=1}^{M} \{Res(\bar{U}_{\hat{\pi} iQ}(\gamma^*, \alpha^*), H_{iQ}(\alpha_Q^*))\} + o_p(1).$$

So we can get

$$\mathcal{V}(M^{1/2}\hat{\beta}_{\hat{\pi}})$$

$$= E[\partial U_{\hat{\pi} i0}(\beta^*)/\partial \beta]^{-1}$$

$$E[Res(U_{\hat{\pi} i0}(\beta^*, \alpha_0^*), H_{i0}(\alpha_0^*)) Res^T(U_{\hat{\pi} i0}(\beta^*, \alpha_0^*), H_{i0}(\alpha_0^*))] E[\partial U_{\hat{\pi} i0}^T(\beta^*)/\partial \beta]^{-1}$$

$$= \Gamma_{00}^{-1} \Sigma_{\hat{\pi} 00} \Gamma_{00}^{T-1},$$

$$\mathcal{V}(M^{1/2}\hat{\gamma}_{\hat{\pi}})$$

$$= E[\partial U_{\hat{\pi} iQ}(\gamma^*)/\partial \gamma]^{-1}$$

$$E[Res(U_{\hat{\pi} iQ}(\gamma^*, \alpha_0^*), H_{i0}(\alpha_0^*)) Res^T(U_{\hat{\pi} iQ}(\gamma^*, \alpha_0^*), H_{i0}(\alpha_0^*))] E[\partial U_{\hat{\pi} iQ}^T(\gamma^*)/\partial \gamma]^{-1}$$

$$= \Gamma_{11}^{-1} \Sigma_{\hat{\pi} 11} \Gamma_{11}^{T-1},$$

$$\mathcal{V}(M^{1/2}\bar{\gamma}_{\hat{\pi}})$$

$$= E[\partial \bar{U}_{\hat{\pi} iQ}(\gamma^*)/\partial \gamma]^{-1}$$

$$E[Res(\bar{U}_{\hat{\pi} iQ}(\gamma^*, \alpha_Q^*), H_{iQ}(\alpha_Q^*)) Res^T(\bar{U}_{\hat{\pi} iQ}(\gamma^*, \alpha_Q^*), H_{iQ}(\alpha_Q^*))] E[\partial \bar{U}_{\hat{\pi} iQ}(\gamma^*)/\partial \gamma]^{-1}$$

$$= \Gamma_{11}^{-1} \Sigma_{\hat{\pi} 22} \Gamma_{11}^{T-1},$$

$$\mathcal{C}(M^{1/2}\hat{\beta}, M^{1/2}\hat{\gamma})$$

$$= E[\partial U_{\hat{\pi}i0}(\beta^*)/\partial\beta]^{-1}$$

$$E[Res(U_{\hat{\pi}i0}(\beta^*, \alpha_0^*), H_{i0}(\alpha_0^*))Res^T(U_{\hat{\pi}iQ}(\gamma^*, \alpha_0^*), H_{i0}(\alpha_0^*))]E[\partial U_{\hat{\pi}iQ}^T(\gamma^*)/\partial\gamma]^{-1}$$

$$= \Gamma_{00}^{-1}\Sigma_{\hat{\pi}01}\Gamma_{11}^{T-1},$$

$$\mathcal{C}(M^{1/2}\hat{\beta}, M^{1/2}\bar{\gamma})$$

$$= E[\partial U_{\hat{\pi}i0}(\beta^*)/\partial\beta]^{-1}$$

$$E[Res(U_{\hat{\pi}i0}(\beta^*, \alpha_0^*), H_{i0}(\alpha_0^*))Res^T(\bar{U}_{\hat{\pi}iQ}(\gamma^*, \alpha_Q^*), H_{iQ}(\alpha_Q^*))]E[\partial \bar{U}_{\hat{\pi}iQ}^T(\gamma^*)/\partial\gamma]^{-1}$$

$$= \Gamma_{00}^{-1}\Sigma_{\hat{\pi}02}\Gamma_{11}^{T-1},$$

$$\mathcal{C}(M^{1/2}\hat{\gamma}, M^{1/2}\bar{\gamma})$$

$$= E[\partial U_{\hat{\pi}iQ}(\gamma^*)/\partial\gamma]^{-1}$$

$$E[Res(U_{\hat{\pi}iQ}(\gamma^*, \alpha_0^*), H_{i0}(\alpha_0^*))Res^T(\bar{U}_{\hat{\pi}iQ}(\gamma^*, \alpha_Q^*), H_{iQ}(\alpha_Q^*))]E[\partial \bar{U}_{\hat{\pi}iQ}^T(\gamma^*)/\partial\gamma]^{-1}$$

$$= \Gamma_{11}^{-1}\Sigma_{\hat{\pi}12}\Gamma_{11}^{T-1},$$

where $E[\partial U_{\hat{\pi}iQ}(\gamma^*)/\partial\gamma] = E[\partial \bar{U}_{\hat{\pi}iQ}(\gamma^*)/\partial\gamma] = \Gamma_{11}$.

Appendix C: Relationship to existing approaches

Robins et al. (1994) proposed a general class of estimators, which includes all regular asymptotic linear estimators. The relationship between the estimator $\bar{\beta}$, which is also asymptotic linear, and those of Robins et al. (1994) has been discussed in Chen and Chen (2000) for MCAR case with simple monotone missing pattern. We will consider the other cases in our approaches.

Adding a function with zero expectation to the estimating function, Robins et al. (1994) maintains an unbiased estimating function. A suitable choice of this added estimation function may improve the estimation efficiency. The serial of surrogate models in our generalized approach takes the role of the function with zero expectations. We will show that our generalized unified estimator corresponds to a member in Robins et al. (1994), and it is more efficient than $\hat{\beta}$ using other perspective.

Recall that the estimators of Robins et al. (1994), which make essentially the same assumptions as our proposal, are asymptotic linear with influence function of the form $D_{00}^{-1}R(w,\kappa)$, where

$$R(w,\kappa) = \delta S(\beta^*)/\pi - (\delta - \pi)\kappa/\pi$$

with $\delta = 1$ if an observation belong to complete case sample and $\delta = 0$ otherwise, $\pi = P(\delta = 1|y,x)$ and $\kappa = \kappa(y,x)$ being a function of (y,x). We note that $E[(\delta-\pi)\kappa/\pi] = 0$ does not depend on κ.

First we consider the MCAR case with q distinct missingness pattern. We have

$$\bar{\beta} = \hat{\beta} - D_{00}^{-1}C_{12}C_{22}^{-1}D_{11}(\hat{\gamma} - \bar{\gamma}),$$

$$\bar{\beta} - \beta^* = \hat{\beta} - \beta^* - D_{00}^{-1}C_{12}C_{22}^{-1}D_{11}((\hat{\gamma} - \gamma^*) - (\bar{\gamma} - \gamma^*)), \text{ and}$$

$$N^{1/2}(\bar{\beta} - \beta^*) = N^{1/2}(\hat{\beta} - \beta^*) - D_{00}^{-1}C_{12}C_{22}^{-1}D_{11}(N^{1/2}(\hat{\gamma} - \gamma^*) - N^{1/2}(\bar{\gamma} - \gamma^*)),$$

where

$$N^{1/2}(\hat{\beta} - \beta^*) = N^{-1/2}D_{00}^{-1}\sum_{i=1}^{N}\{(R_{i0}/\pi_0)S_{i0}\} + o_p(1),$$

$$N^{1/2}(\hat{\gamma} - \gamma^*) = N^{-1/2}D_{11}^{-1}\sum_{i=1}^{N}\{(R_{i0}/\pi_0)S_{iQ}\} + o_p(1)$$

and

$$N^{1/2}(\bar{\gamma} - \gamma^*) = N^{-1/2} D_{11}^{-1} \sum_{i=1}^{N} \{\Delta_i S_{iQ}\} + o_p(1).$$

So

$$N^{1/2}(\bar{\beta} - \beta^*) = N^{-1/2} D_{00}^{-1} \sum_{i=1}^{N} \{(R_{i0}/\pi_0)(S_{i0} - BS_{iQ}) + B\Delta_i S_{iQ}\} + o_p(1),$$

$$N^{1/2}(\bar{\beta} - \beta^*) = N^{-1/2} D_{00}^{-1} \sum_{i=1}^{N} \{(R_{i0}/\pi_0)(S_{i0}) - ((R_{i0}/\pi_0)BS_{iQ} - B\Delta_i S_{iQ})\} + o_p(1),$$

where $B = C_{01}C_{11}^{-1} = cov(S_{i0}, S_{iQ})var(S_{iQ})^{-1}$ and $\Delta_i = diag(R_{i1} * I_1/\pi_1, \cdots, R_{iq} * I_q/\pi_q)$. We note that $E[B(R_{i0}/\pi_0)S_{iQ} - B\Delta_i S_{iQ}] = 0$ does not depending on S_{iQ}, the estimator $\bar{\beta}$ corresponds to a member of the class of estimators in Robins et al. (1994) by replacing $(\delta - \pi)\kappa/\pi$ with $((R_{i0}/\pi)BS_{iQ} - B\Delta_i S_{iQ})$.

The estimator $\bar{\beta}_\pi$ in the MAR case with known missing probability is similar to that in the MCAR case. We note that

$$N^{1/2}(\bar{\beta}_\pi - \beta^*) = N^{-1/2} D_{00}^{-1} \sum_{i=1}^{N} \{R_{i0}(S_{i0} - BS_{iQ})/\pi_{i0} + B\Delta_i S_{iQ}\} + o_p(1),$$

$$N^{1/2}(\bar{\beta}_\pi - \beta^*) = N^{-1/2} D_{00}^{-1} \sum_{i=1}^{N} \{\frac{R_{i0}}{\pi_{i0}} S_{i0} - (B(R_{i0}/\pi_{i0})S_{iQ} - B\Delta_i S_{iQ})\} + o_p(1),$$

where $B = C_{\pi 01}C_{\pi 11}^{-1} = cov(S_{i0}, S_{iQ})var(S_{iQ})^{-1}$ and $\Delta_i = diag(R_{ik} * I_k/\pi_{ik}), k = 1, \cdots, q$. Here $E[B(R_{i0}/\pi_{i0})S_{iQ} - B\Delta_i S_{iQ}] = 0$ still does not depend on S_{iQ}, Thus it can be seen that the estimator $\bar{\beta}$ corresponds to a member of the class of estimators in Robins et al. (1994) by replacing $(\delta - \pi)\kappa/\pi$ with $((R_{i0}/\pi_{i0})BS_{iQ} - B\Delta_i S_{iQ})$.

Finally, we consider the MAR case with estimated missing probability. We note that

$$\bar{\beta}_{\hat{\pi}} = \hat{\beta}_{\hat{\pi}} - D_0^{-1} C_{\hat{\pi}12} C_{\hat{\pi}22}^{-1} D_1 (\hat{\gamma}_{\hat{\pi}} - \bar{\gamma}_{\hat{\pi}}),$$

$$\bar{\beta}_{\hat{\pi}} - \beta^* = \hat{\beta}_{\hat{\pi}} - \beta^* - D_0^{-1} C_{\hat{\pi}12} C_{\hat{\pi}22}^{-1} D_1((\hat{\gamma}_{\hat{\pi}} - \gamma^*) - (\bar{\gamma}_{\hat{\pi}} - \gamma^*)), \text{and}$$

$$N^{1/2}(\bar{\beta}_{\hat{\pi}} - \beta^*) = N^{1/2}(\hat{\beta}_{\hat{\pi}} - \beta^*) - D_0^{-1} C_{\hat{\pi}12} C_{\hat{\pi}22}^{-1} D_1(N^{1/2}(\hat{\gamma}_{\hat{\pi}} - \gamma^*) - N^{1/2}(\bar{\gamma}_{\hat{\pi}} - \gamma^*)),$$

where

$$N^{1/2}(\hat{\beta}_{\hat{\pi}} - \beta^*) = N^{-1/2} D_0^{-1} \sum_{i=1}^{N} Res(\frac{R_{i0}}{\pi_{i0}} S_{i0}(\beta^*), H_{\pi i0}(\alpha_0^*)) + o_p(1),$$

$$N^{1/2}(\hat{\gamma}_{\hat{\pi}} - \gamma^*) = N^{-1/2} D_1^{-1} \sum_{i=1}^{N} Res(\frac{R_{i0}}{\pi_{i0}} S_{iQ}(\gamma^*), H_{\pi i0}(\alpha_0^*)) + o_p(1),$$

$$N^{1/2}(\bar{\gamma}_{\hat{\pi}} - \gamma^*) = N^{-1/2} D_1^{-1} \sum_{i=1}^{N} Res(S_{\pi iQ}(\gamma^*, \alpha_Q^*), H_\pi iQ(\alpha_Q^*)) + o_p(1).$$

So we have

$$\begin{aligned} N^{1/2}(\bar{\beta}_{\hat{\pi}} - \beta^*) =& N^{-1/2} D_0^{-1} \sum_{i=1}^{N} \{Res(U_{\hat{\pi} i0}(\beta^*, \alpha_0^*), S_{\pi i0}(\alpha_0^*)) \\ & - B(Res(\frac{R_{i0}}{\pi_{i0}} S_{iQ}(\gamma^*), S_{\pi i0}(\alpha_0^*)) - Res(S_{\pi iQ}(\gamma^*, \alpha_Q^*), H_{\pi iQ}(\alpha_Q^*)))\} + o_p(1), \end{aligned}$$

where

$$B = Cov\{Res(\frac{R_{i0}}{\pi_{i0}} S_{i0}(\beta^*), H_{i0}(\alpha_0^*)) Res(\frac{R_{i0}}{\pi_{i0}} S_{iQ}(\gamma^*), H_{i0}(\alpha_0^*))\} Var^{-1}[Res(\frac{R_{i0}}{\pi_{i0}} S_{iQ}(\gamma^*), H_{i0}(\alpha_0^*))].$$

Under a correctly specified parametric models for the missing data probability, we can show that

$$E[B(Res(\frac{R_{i0}}{\pi_{i0}} S_{iQ}(\gamma^*), S_{\pi i0}(\alpha_0^*)) - Res(S_{\pi iQ}(\gamma^*, \alpha_Q^*), H_{\pi iQ}(\alpha_Q^*)))] = 0$$

does not depend on the surrogate models. Thus it can be seen that the estimator $\bar{\beta}$ corresponds to a member of the class of estimations in Robins et al. (1994).

Appendix D: Generate Correlated Random Number

In this appendix, we will introduce how to generate the correlated response y_{ij} according to the exchangeable correlation and the $Ar(1)$ correlation respectively. We start with the simple case where y_{ij} is continuous (Peter Diggle et al. 2002), then we give the details for the complex case where y_{ij} is a binary variable (Preisser et al. 2002). For each case, we will give the model first, and then provide the algorithm steps.

Continuous Variables

For the exchangeable structure, suppose that y_{ij} follows the model

$$y_{ij} = \mu_{ij} + U_i + Z_{ij}, \qquad i = 1, ..., m, \qquad j = 1, ..., n,$$

where $\mu_{ij} = E(y_{ij})$, the U_i are mutually independent $N(0, \nu^2)$ random variables, the Z_{ij} are mutually independent $N(0, \tau^2)$ random variables, and U_i and Z_{ij} are independent. Then, the covariance structure of the data y_{ij} is $\rho = \nu^2/(\nu^2 + \tau^2)$ and $\sigma^2 = \nu^2 + \tau^2$. The steps that generate the random data x_{ij1}, x_{ij2} and y_{ij} such that the correlation structure between y_{ij} is exchangeable are as follows.

(1) generate x_{ij1} and x_{ij2} such that x_{ij1} follows $N(0.1 * j, 1)$ and X_{ij2} follows $N(0.01 * j^2, 1)$, $i = 1, ..., m, j = 1, ..., n;$

(2) generate U_i such that U_i follows $N(0, \nu^2)$, $i = 1, ..., m;$

(3) generate Z_{ij} such that Z_{ij} follows $N(0, \tau^2)$, $i = 1, ..., m, j = 1, ..., n;$

(4) generate μ_{ij} such that $\mu_{ij} = \beta_0 + \beta_1 * x_{ij1} + \beta_2 * x_{ij2}$, $i = 1, ..., m, j = 1, ..., n;$

(5) generate y_{ij} such that $y_{ij} = \mu_{ij} + U_i + Z_{ij}, i = 1, ..., m, j = 1, ..., n.$

For exponential structure, suppose that y_{ij} satisfies the model

$$y_{ij} = \mu_{ij} + W_{ij}, i = 1, ..., m, j = 1, ..., n,$$

where $W_{ij} = \rho * W_{ij-1} + Z_{ij}$ and the Z_{ij} are mutually independent $N(0, \sigma^2 * (1 - \rho^2))$ random variables. Then $v_{ij} = Cov(Y_{ij}, Y_{ik}) = \sigma^2 \rho^{|j-k|}$. The steps that generate the random data x_{ij1}, x_{ij2} and y_{ij} such that the correlation structure between y_{ij} is $Ar(1)$ are as follows.

(1) generate x_{ij1} and x_{ij2} such that x_{ij1} follows $N(0.1 * j, 1)$ and x_{ij2} follows $N(0.01 * j^2, 1)$, $i = 1, ..., m, j = 1, ..., n$;

(2) generate Z_{ij} such that Z_{ij} follows $N(0, \sigma^2 * (1 - \rho^2))$, $i = 1, ..., m, j = 1, ..., n$;

(3) generate W_{ij} such that $W_{ij} = \rho * W_{ij-1} + Z_{ij}$, $i = 1, ..., m, j = 1, ..., n$;

(4) generate μ_{ij} such that $\mu_{ij} = \beta_0 + \beta_1 * x_{ij1} + \beta_2 * x_{ij2}$, $i = 1, ..., m, j = 1, ..., n$;

(5) generate y_{ij} such that $y_{ij} = \mu_{ij} + W_{ij}, i = 1, ..., m, j = 1, ..., n$.

Binary Variables

Suppose we wish to simulate $Y_i, i = 1, ..., m$, a J-vector of Bernoulli variates with mean vector π_i and covariance matrix V_i. For $j = 2, ..., J$, define $Z_{ij} = (y_{i1}, ..., y_{ij-1})^T$, $\mu_{ij} = E(Z_{ij}), G_{ij} = Cov(Z_{ij})$, and $s_{ij} = Cov(Z_{ij}, y_{ij})$. Note that G_{ij} and s_{ij} are determined from V_i. For a given (π_i, V_i), a $(j - 1)$ vector b_{ij} is defined as $b_{ij} = G_{ij}^{-1} s_{ij} (j = 2, ..., J)$. The conditional probability is defined by

$$\nu_{ij} = \nu_{ij}(z_{ij}; \pi_i, V_i) = P(y_{ij} | Z_{ij} = z_{ij}) = \pi_{ij} + b_{ij}^T (z_{ij} - \mu_{ij}) = \pi_{ij} + \sum_{k=1}^{j-1} b_{ijk}(y_{ik} - \pi_{ik}).$$

The simulation algorithm proceeds as follows. First, simulate y_{i1} as Bernoulli random variable with mean π_{i1}, then for $j = 2, ..., n$, simulate y_{ij} as Bernoulli random variable with conditional mean ν_{ij}. It then follows that $E(Y_i) = \pi_i$ and for $1 < j \leq n$, $Cov(Z_{ij}, y_{ij}) = Cov(Z_{ij}, b_{ij}^T Z_{ij}) = G_{ij} b_{ij} = s_{ij}$. The vector Y_i thus obtained has the required mean, π_i, and covariance V_i.

For the exchangeable structure, we have

$$b_{ijk} = (\frac{\rho}{1+(j-2)\rho})(\frac{\pi_{ij}(1-\pi_{ij})}{\pi_{ik}(1-\pi_{ik})})^{\frac{1}{2}},$$

and

$$\nu_{ij} = \pi_{ij} + \sum_{k=1}^{j-1} b_{ijk}(y_{ik} - \pi_{ik}), (j = 2, ..., J).$$

For the $Ar(1)$ structure, we have

$$\nu_{ij} = \pi_{ij} + \rho(y_{ij-1} - \pi_{ij-1})(\frac{\pi_{ij}(1-\pi_{ij})}{\pi_{ij-1}(1-\pi_{ij-1})})^{\frac{1}{2}}.$$

The steps that generate the random data x_{ij1}, x_{ij2} and y_{ij} such that the correlation structure between y_{ij} is exchangeable are as follows.

(1) generate x_{ij1} and x_{ij2} such that $x_{ij1} = (j + N(0,1))/(n-1)$ and $x_{ij2} = (j + N(0,1)) * (j + N(0,1))/(n-1)^2$ $i = 1, ..., m, j = 1, ..., n$;

(2) generate π_{ij} such that $\pi_{ij} = exp(\beta_1 * x_{ij1} + \beta_2 * x_{ij2})/(1 + exp(\beta_1 * x_{ij1} + \beta_2 * x_{ij2}))$;

(3) generate b_{ijk} and ν_{ij} such that $b_{ijk} = (\frac{\rho}{1+(j-2)\rho})(\frac{\pi_{ij}(1-\pi_{ij})}{\pi_{ik}(1-\pi_{ik})})^{\frac{1}{2}}$ and $\nu_{ij} = \pi_{ij} + \sum_{k=1}^{j-1} b_{ijk}(y_{ik} - \pi_{ik})$;

(4) generate y_{ij} according to the conditional probability ν_{ij};

(5) repeat step 3 and 4 to generate y_{ij} iteratively.

The steps that generate the random data x_{ij1}, x_{ij2} and y_{ij} such that the correlation structure between y_{ij} is $Ar(1)$ are as follows.

(1) generate x_{ij1} and x_{ij2} such that $x_{ij1} = (j + N(0,1))/(n-1)$ and $x_{ij2} = (j + N(0,1)) * (j + N(0,1))/(n-1)^2$ $i = 1, ..., m, j = 1, ..., n$;

(2) generate π_{ij} such that $\pi_{ij} = exp(\beta_1 * x_{ij1} + \beta_2 * x_{ij2})/(1 + exp(\beta_1 * x_{ij1} + \beta_2 * x_{ij2}))$;

(3) generate ν_{ij} such that $\nu_{ij} = \pi_{ij} + \rho(y_{ij-1} - \pi_{ij-1})(\frac{\pi_{ij}(1-\pi_{ij})}{\pi_{ij-1}(1-\pi_{ij-1})})^{\frac{1}{2}}$;

(4) generate y_{ij} according to the conditional probability ν_{ij};

(5) repeat steps 3 and 4 to generate y_{ij} iteratively.

Bibliography

[1] Andrea B. Troxel, David P. Harrington and Stuart R. Lipsitz. (1998). Analysis of longitudinal data with non-ignorable non-monotone missing values. *Journal of the Royal Statistical Society. Series C (Applied Statistics)* 47, 3, 425-438.

[2] Breslow, N. E., Lumley, T., Ballantyne, C. M., Chambless, L. E. and Kulich, M. (2009). Improved Horvitz-Thompson estimation of model parameters from two-phase stratified samples: applications in epidemiology. *Statistics in Biosciences* 1, 32-49.

[3] Chatterjee, N., Chen, Y. and Breslow, N. E. (2003). A pseudo-score estimator for regression problems with two-phase sampling. *Journal of the American Statistical Association* 98, 158-168.

[4] Chatterjee, N. and Li, Y. (2010). Inference in semiparametric regression models under partial questionnaire design and non-monotone missing data. *Journal of the American Statistical Association* 105, 1176-1189.

[5] Yi-Hau Chen and Hung Chen. (2000). A unified approach to regression analysis under double-sampling designs. *Journal of the Royal Statistical Society. Series B (Statistical Methodology)* 2, 3, 449-460.

[6] Chen, H. Y. (2004). Nonparametric and semiparametric models for missing covariates in parametric regression. *Journal of the American Statistical Association* 99, 787-797.

[7] Chen, Q., Ibrahim, J. G., Chen, M. and Senchaudhuri, P. (2008). Theory and inference for regression models with missing responses and covariates. *Journal of Multivariate Analysis* 99, 13021331.

[8] Chen, B., Yi, G.Y. and Cook, R.J. (2010). Weighted generalized estimating functions for incomplete longitudinal response and covariate data that are missing at random. *Journal of the American Statistical Association* 105, 336-353.

[9] Chen, Baojiang. and Zhou, XiaoHua. (2011). Doubly robust estimates for binary longitudinal data analysis with missing response and missing covariates. *Biometrics* 67, 830842.

[10] Colt, J.S., Severson, R.K., Lubin, J.H. and et al. (2005). Organochlorines in carpet dust and non-hodgkin lymphoma. *Epidemiology* 16, 516-525.

[11] DeRoos, A.J., Hartge, P., Lubin, J.H. and et al. (2005). Persistent organochlorine chemicals in plasma and risk of nonhodgkin's lymphoma. *Cancer Research* 65, 11214-11226.

[12] Diggle, P.J. , Heagerty, P., Liang, K.Y. and Zeger, S.L. (2002). *Analysis of longitudinal data*, Oxford University Press 2nd edition.

[13] Foutz, R.V. (1977). On the unique consistent solution to the likelihood equations. *Journal of the American Statistical Association* 72, 147-148.

[14] Holcroft, C.A., Rotnitzky, A. and Robins, J.M. (1997). Efficient estimation of regression parameters from multistage studies with validation of outcome and covariates. *Journal of Statistical Planning and Inference* 65, 349-374.

[15] Hosmer, D.W. and Lemehow, S. (2000). *Applied Logistic Regression*. New York: John Wiley and Sons.

[16] Horvitz, D. G. and Thompson, D. J. (1952). A generalization of sampling without replacement from a finite universe. *Journal of the American Statistical Association* 47, 663-685.

[17] Houseman, E.A. and Milton, D.K. (2006). The partial questionnaire design, questionnaire non-response, and attributable fraction: application to adult onset asthma. *Statistics in Medicine* 25, 1499-1519.

[18] Ibrahim, J.G. (1990). Incomplete data in generalized linear models. *Journal of the American Statistical Association* 85, 765-769.

[19] Lawless, J. F., Kalbfleisch, J. D. and Wild, C. J. (1999). Semiparametric methods for response-selective and missing data problems in regression. *Journal of Royal Statistical Society Series B (Statistical Methodology)* 61, 413-438.

[20] Kung-Yee Liang and Scott L. Zeger. (1986) Longitudinal data analysis using generalized linear models. *Biometrika* 73, 13-22.

[21] Lipsitz, S.R. and Joseph G.IBRAHIM. (1996). A conditional model for incomplete covariates in parametric regression models. *Biometrika* 83, 916-922.

[22] Lipsitz, S.R., I.J.G. and Zhao, L. (1999). A new weighted estimating equation for missing covariate data with properties similar to maximum likelihood. *Journal of the American Statistical Association* 94, 1147-1160.

[23] Little,R.J.A. and Rubin, D.B. (2002). *Statistical Analysis with Missing Data*. Wiley Series in Probability and Statistics.

[24] Newey, W.K. and Mcfadden, D. (1994). *Handbook of Econometrics, Volume IV, Edited by R. F. Engle and D. L. McFadden.* Elsevier Science B. V.

[25] Pepe, M.S. and Anderson, G.L.(1994). A cautionary note on inference for marginal regression models with longitudinal data and general correlated response data. *Communications in Statistics, Simulation and Computation* 23, 939-951.

[26] Peter Diggle, Patrick Heagerty, Kung-Yee Liang and Scott Zeger. (2002). *Analysis of Longitudinal Data*, Oxford University Press, USA.

[27] Preisser, J. S., Lohman, K. K. and Rathouz, P.J. (2002). Performance of weighted estimating equations for longitudinal binary data with drop-outs missing at random. *Statistics in Medicine* 21, 3035-3054.

[28] Reilly, M. and Pepe, M. (1995). A mean score method for missing and auxiliary covariate data in regression models. *Biometrika* 82, 299-314.

[29] Robins,J.M., Rotnitzky, A. and Zhao, L.P.(1994). Estimation of regression coefficients when some regressors are not always observed. *Journal of the American Statistical Association* 89, 846-866.

[30] Robins,J.M., Rotnitzky, A. and Zhao, L.P.(1995). Analysis of semiparametric regression models for repeated outcomes in the presence of missing data. *Journal of the American Statistical Association* 90, 106-121.

[31] Robins,J.M., GreenLand, S. and Hu, F.C.(1999). Estimation of the causal effect of a time varying exposure on the marginal mean of a repeated binary outcome (with discussion). *Journal of the American Statistical Association* 94, 687-712.

[32] Rubin, D.B. (1976). Inference and missing data. *Biometrika* 63, 581-592.

[33] Scott, A.J. and Wild, C.J. (1998). Maximum likelihood for generalized case-control studies. *Preprint, Department of Statistics, University of Auckland.*

[34] Scharfstein, D.O., Rotnitzky, A. and Robins, J.M. (1999). Adjusting for nonignorable drop-out using semiparametric nonresponse models (with discussion), *Journal of the Ameriacan Statistical Association* 94, 1096-1120.

[35] van der Lann, M.J. and Robins, J.M. (2003). Unified methods for censored longitudinal data and causality. *New York: Springer-Verlag.*

[36] Wacholder, S., Carroll, R.J., Pee, D. and Gail, M.G.(1994). The partial questionnaire design for case-control studies. *Statistics in Medicine* 13, 623-634.

[37] Wang, C. Y., Wang, S. J., Zhao, L. P. and T., O. S. (1997). Weighted semiparametric estimation in regression analysis with missing covariate data. *Journal of the American Statistical Association* 92, 512-525.

[38] Zhao, L. P. and Lipsitz, S. (1992). Designs and analysis of two-stage studies. *Statistics in Medicine* 11, 769-782.

[39] Zhao, L.P., Lipsitz, S.R. and Lew, D. (1996). Regression analysis with missing covariate data using estimating equations, *Biometrics* 52, 1165-1182.

[40] Zhao, Y., Lawless, J.F. and McLeish, D.L. (2009). Likelihood methods for regression models with expensive variables missing by design. *Biometrical Journal* 51, 123-136.