

# Annals of the Institute of Statistical Mathematics

## Estimation with Multivariate Outcomes Having Nonignorable Item Nonresponse

--Manuscript Draft--

<b>Manuscript Number:</b>	
<b>Article Type:</b>	Original Paper
<b>Full Title:</b>	Estimation with Multivariate Outcomes Having Nonignorable Item Nonresponse
<b>Short Title:</b>	Estimation with Nonignorable Nonresponse
<b>Order of Authors:</b>	Lyu Ni, PhD
	Jun Shao
<b>Corresponding Author:</b>	Jun Shao University of Wisconsin, Madison UNITED STATES
<b>Abstract:</b>	To estimate unknown population parameters based on $y$ , a vector of multivariate outcomes having nonignorable item nonresponse that directly depends on $y$ , we propose an innovative inverse propensity weighting approach when the joint distribution of $y$ and associated covariate $x$ is nonparametric and the nonresponse probability conditional on $y$ and $x$ has a parametric form. To deal with the identifiability issue we utilize a nonresponse instrument $z$ , an auxiliary variable related to $y$ but not related to the nonresponse probability conditional on $y$ and $x$ . We utilize a modified generalized method of moments to obtain estimators of the parameters in the nonresponse probability. Simulation results are presented and an application is illustrated in a real data set.
<b>Keywords:</b>	Generalized Method of Moments; Item Nonresponse; Inverse Propensity Weighting; Multivariate Outcome; Nonresponse Instrument
<b>Author Comments:</b>	Dear Editor: We would like to submit our paper "Estimation with Multivariate Outcomes Having Nonignorable Item Nonresponse" for publication in the Annals of the Institute of Statistical Mathematics. We look forward to your comments. Thanks. Sincerely The Authors
<b>Additional Information:</b>	
<b>Question</b>	<b>Response</b>
<b>Is your paper accompanied by any supplementary files?</b>	No
The paper should not exceed 30 pages, as described in "Instruction for Authors". Is your paper expected not to exceed 30 pages if you use a style file of AISM? (If not, please move some materials into a supplementary file.)	Yes

# Estimation with Multivariate Outcomes Having Nonignorable Item Nonresponse

Lyu Ni

School of Data Science & Engineering, East China Normal University, Shanghai, China

Jun Shao

School of Statistics, East China Normal University, Shanghai, China

Department of Statistics, University of Wisconsin, Madison, WI, U.S.A.

## Abstract

To estimate unknown population parameters based on  $\mathbf{y}$ , a vector of multivariate outcomes having nonignorable item nonresponse that directly depends on  $\mathbf{y}$ , we propose an innovative inverse propensity weighting approach when the joint distribution of  $\mathbf{y}$  and associated covariate  $\mathbf{x}$  is nonparametric and the nonresponse probability conditional on  $\mathbf{y}$  and  $\mathbf{x}$  has a parametric form. To deal with the identifiability issue we utilize a nonresponse instrument  $\mathbf{z}$ , an auxiliary variable related to  $\mathbf{y}$  but not related to the nonresponse probability conditional on  $\mathbf{y}$  and  $\mathbf{x}$ . We utilize a modified generalized method of moments to obtain estimators of the parameters in the nonresponse probability. Simulation results are presented and an application is illustrated in a real data set.

*Keywords:* Generalized Method of Moments; Item Nonresponse; Inverse Propensity Weighting; Multivariate Outcome; Nonresponse Instrument.

# 1 Introduction

In many statistical applications, multivariate outcomes or responses are collected from every sampled unit in the study. For example, in health studies conducted by the U.S. Centers for Disease Control and Prevention, measurements of total cholesterol, high density lipoprotein cholesterol, body mass index, and average sagittal abdominal diameter etc. may be obtained from each sampled person in the non-institutionalized civilian resident population of the United States. Longitudinal responses are another type of multivariate outcomes, in which each sampled unit is repeatedly measured over several time periods. An example is the AIDS Clinical Trial Group 193A discussed in Section 4 for HIV-AIDS patients with advanced immune suppression.

Unfortunately, item nonresponse is a common phenomena in multivariate responses, i.e., some of the multivariate responses, not necessarily all, may be missing with a pattern varying with sampled unit. Estimation and statistical inference without taking nonresponse into consideration may lead to seriously biased estimators and conclusions.

Throughout this article,  $\mathbf{y}$  denotes a  $k$ -dimensional outcome or response vector of interest that is subject to item nonresponse,  $\mathbf{r}$  denotes the response indicator vector of  $\mathbf{y}$ , i.e., the  $j$ th component of  $\mathbf{r}$  is 1 (or 0) if the  $j$ th component of  $\mathbf{y}$  is observed (or missing),  $j = 1, \dots, k$ , and  $\mathbf{x}$  denotes a  $p$ -dimensional covariate vector associated with  $\mathbf{y}$  that is always observed. Statistical approaches dealing with missing data usually depend on the nonresponse propensity (or mechanism), i.e., the conditional distribution of  $\mathbf{r}$  given  $(\mathbf{y}, \mathbf{x})$ , denoted by  $p(\mathbf{r}|\mathbf{y}, \mathbf{x})$ . If  $p(\mathbf{r}|\mathbf{y}, \mathbf{x}) = p(\mathbf{r}|\mathbf{y}_o, \mathbf{x})$ , where  $\mathbf{y}_o$  is the observed part of  $\mathbf{y}$ , then nonresponse is ignorable (Rubin, 1976; Little and Rubin, 2002). Otherwise, nonresponse is nonignorable. While there is a rich literature for valid inference under ignorable nonresponse (Little and Rubin, 2002), there are serious challenges under nonignorable nonresponse, especially for multivariate  $\mathbf{y}$  with item nonresponse.

Greenlees et al. (1982) proposed to handle nonignorable item nonresponse by maximum likelihood estimation, assuming parametric models on both  $p(\mathbf{r}|\mathbf{y}, \mathbf{x})$  and  $p(\mathbf{y}|\mathbf{x})$ , the conditional density of  $\mathbf{y}$  given  $\mathbf{x}$ . However, a fully parametric approach is sensitive to the parametric model assumptions. Since the population  $p(\mathbf{y}, \mathbf{r}|\mathbf{x}) = p(\mathbf{r}|\mathbf{y}, \mathbf{x})p(\mathbf{y}|\mathbf{x})$  is not identifiable when both  $p(\mathbf{r}|\mathbf{y}, \mathbf{x})$  and  $p(\mathbf{y}|\mathbf{x})$  are nonparametric (Robins and Ritov, 1997), efforts have been made in scenarios where one of  $p(\mathbf{r}|\mathbf{y}, \mathbf{x})$  and  $p(\mathbf{y}|\mathbf{x})$  is parametric or semiparametric. Tang et al. (2003)

and Zhao and Shao (2015) considered the situation where  $p(\mathbf{y}|\mathbf{x})$  is parametric but  $p(\mathbf{r}|\mathbf{y}, \mathbf{x})$  is nonparametric, whereas Wang et al. (2014) and Shao and Wang (2016) studied a univariate response  $y$  ( $k = 1$ ) with a nonparametric  $p(y|\mathbf{x})$  and a parametric or semi-parametric  $p(\mathbf{r}|\mathbf{y}, \mathbf{x})$ . Under a mixed-effect model on  $p(\mathbf{y}|\mathbf{x})$ , Wu and Carroll (1988), Xu and Shao (2009), and Shao and Zhang (2015) obtained some results when the dependence of  $\mathbf{r}$  on  $\mathbf{y}$  is through an unobserved random effect  $\mathbf{b}$ , i.e.,  $p(\mathbf{r}|\mathbf{y}, \mathbf{x}) = p(\mathbf{r}|\mathbf{b}, \mathbf{x})$ .

Under nonparametric conditional density  $p(\mathbf{y}|\mathbf{x})$  and nonparametric marginal density  $p(\mathbf{y})$ , in this paper we propose an innovative inverse propensity weighting approach to construct valid estimators of population parameters in the presence of nonignorable item nonresponse in  $\mathbf{y}$ , assuming the following two assumptions on the propensity:

- (A1) The covariate vector  $\mathbf{x} = (\mathbf{u}, \mathbf{z})$  with a non-constant sub-vector  $\mathbf{z}$  such that  $p(\mathbf{r}|\mathbf{y}, \mathbf{x}) = p(\mathbf{r}|\mathbf{y}, \mathbf{u})$  and  $p(\mathbf{y}|\mathbf{x}) = p(\mathbf{y}|\mathbf{u}, \mathbf{z})$  depends on  $\mathbf{z}$ .
- (A2) Given  $(\mathbf{y}, \mathbf{u})$ , components of  $\mathbf{r}$  are conditionally independent and, for each  $j = 1, \dots, k$ , the probability of observing the  $j$ th component of  $\mathbf{y}$  is  $\pi_j(\mathbf{y}, \mathbf{u}, \boldsymbol{\theta}_j)$ , where  $\boldsymbol{\theta}_j$  is an unknown parameter vector and  $\pi_j$  is a known function of  $(\mathbf{y}, \mathbf{u})$  when  $\boldsymbol{\theta}_j$  is known.

The covariate  $\mathbf{z}$  in (A1) is referred to as a nonresponse instrument (Wang et al., 2014; Zhao and Shao, 2015). The existence of a nonresponse instrument that can be excluded from the propensity is almost necessary for handling nonignorable nonresponse (Wang et al., 2014; Zhao and Shao, 2015; Shao and Wang, 2016). Also, as discussed earlier, the parametric assumption on propensity is needed as  $p(\mathbf{y}|\mathbf{x})$  is nonparametric. Finally, the conditional independence of components of  $\mathbf{r}$  given  $(\mathbf{y}, \mathbf{u})$  in (A2) is actually reasonable in many applications with item nonresponse, as the conditional independence is not the same as the unconditional independence of components of  $\mathbf{r}$ .

Under (A2), conditioned on  $(\mathbf{y}, \mathbf{u})$ , the nonresponse propensity  $\pi_j(\mathbf{y}, \mathbf{u}, \boldsymbol{\theta}_j)$  not only directly depends on the entire  $\mathbf{y}$  and possibly  $\mathbf{u}$ , but also varies with  $j$  (component). No general result is available under this type of item nonresponse in the literature. The closest is Li and Shao (2021), but it assumes that given  $(\mathbf{y}, \mathbf{u})$ , components of  $\mathbf{r}$  are identically distributed, which may not be realistic when components of  $\mathbf{y}$  have different distributions (see the real data example in Section 4).

Our main methodology is introduced in Section 2, followed by some simulation results in Section 3 and one real data example in Section 4.

## 2 Methodology

Let  $(\mathbf{y}_i, \mathbf{x}_i, \mathbf{r}_i)$ ,  $i = 1, \dots, n$ , be identically distributed and independently sampled from the population of  $(\mathbf{y}, \mathbf{x}, \mathbf{r})$ . Values of  $\mathbf{x}_i$  are always observed and components of  $\mathbf{y}_i$  are observed if and only if the corresponding components of  $\mathbf{r}_i$  are equal to one. Under assumptions (A1)-(A2), we propose to estimate population parameters using inverse propensity weighting, based on observed data in  $(\mathbf{y}_i, \mathbf{x}_i, \mathbf{r}_i)$ ,  $i = 1, \dots, n$ .

### 2.1 Estimation when $\boldsymbol{\theta}_j$ 's are known

To illustrate the idea, we consider estimating population mean  $\mu_j = E(y_j)$ , where  $y_j$  is the  $j$ th component of  $\mathbf{y}$  and  $j$  is a fixed integer between 1 and  $k$ . Estimation of other parameters is discussed in the end of this subsection.

In this subsection, we assume that  $\boldsymbol{\theta}_j$ 's in (A2) are known. Estimation of  $\boldsymbol{\theta}_j$ 's is considered in the next subsection. For  $\mathbf{r}_i$  and  $\mathbf{y}_i$ , denote their  $j$ th components by  $r_{ij}$  and  $y_{ij}$ , respectively. The simple inverse propensity weighting estimator,

$$\sum_{i=1}^n \frac{r_{ij} y_{ij}}{\pi_j(\mathbf{y}_i, \mathbf{u}_i, \boldsymbol{\theta}_j)} \bigg/ \sum_{i=1}^n \frac{r_{ij}}{\pi_j(\mathbf{y}_i, \mathbf{u}_i, \boldsymbol{\theta}_j)},$$

which works for the univariate case of  $k = 1$ , does not work because  $\pi_j(\mathbf{y}_i, \mathbf{u}_i, \boldsymbol{\theta}_j)$  cannot be computed when  $\mathbf{y}_i$  has a missing component  $l \neq j$ . Thus, we propose the following estimator of  $\mu_j$  using composite inverse propensity weighting:

$$\hat{\mu}_j(\boldsymbol{\theta}) = \sum_{i=1}^n \frac{(r_{i1} \cdots r_{ik}) y_{ij}}{\pi_1(\mathbf{y}_i, \mathbf{u}_i, \boldsymbol{\theta}_1) \cdots \pi_k(\mathbf{y}_i, \mathbf{u}_i, \boldsymbol{\theta}_k)} \bigg/ \sum_{i=1}^n \frac{r_{i1} \cdots r_{ik}}{\pi_1(\mathbf{y}_i, \mathbf{u}_i, \boldsymbol{\theta}_1) \cdots \pi_k(\mathbf{y}_i, \mathbf{u}_i, \boldsymbol{\theta}_k)}, \quad (1)$$

where  $\boldsymbol{\theta}$  is a vector with  $\boldsymbol{\theta}_1, \dots, \boldsymbol{\theta}_k$  as sub-vectors assumed to be known at this moment. Since the product  $r_{i1} \cdots r_{ik}$  is used, we must use the product  $\pi_1(\mathbf{y}_i, \mathbf{u}_i, \boldsymbol{\theta}_1) \cdots \pi_k(\mathbf{y}_i, \mathbf{u}_i, \boldsymbol{\theta}_k)$  as weights, which can be computed when  $r_{i1} \cdots r_{ik} = 1$ . To see why  $\hat{\mu}_j(\boldsymbol{\theta})$  in (1) is asymptotically valid as  $n \rightarrow \infty$ , note that

$$\begin{aligned} E \left\{ \frac{(r_{i1} \cdots r_{ik}) y_{ij}}{\pi_1(\mathbf{y}_i, \mathbf{u}_i, \boldsymbol{\theta}_1) \cdots \pi_k(\mathbf{y}_i, \mathbf{u}_i, \boldsymbol{\theta}_k)} \right\} &= E \left[ E \left\{ \frac{(r_{i1} \cdots r_{ik}) y_{ij}}{\pi_1(\mathbf{y}_i, \mathbf{u}_i, \boldsymbol{\theta}_1) \cdots \pi_k(\mathbf{y}_i, \mathbf{u}_i, \boldsymbol{\theta}_k)} \middle| \mathbf{y}_i, \mathbf{u}_i \right\} \right] \\ &= E \left[ \frac{y_{ij} E(r_{i1} \cdots r_{ik} | \mathbf{y}_i, \mathbf{u}_i)}{\pi_1(\mathbf{y}_i, \mathbf{u}_i, \boldsymbol{\theta}_1) \cdots \pi_k(\mathbf{y}_i, \mathbf{u}_i, \boldsymbol{\theta}_k)} \right] \\ &= E \left[ \frac{y_{ij} E(r_{i1} | \mathbf{y}_i, \mathbf{u}_i) \cdots E(r_{ik} | \mathbf{y}_i, \mathbf{u}_i)}{\pi_1(\mathbf{y}_i, \mathbf{u}_i, \boldsymbol{\theta}_1) \cdots \pi_k(\mathbf{y}_i, \mathbf{u}_i, \boldsymbol{\theta}_k)} \right] \\ &= E(y_{ij}) = \mu_j, \end{aligned}$$

where the third equality follows from the independence of  $r_{ij}$ 's conditioned on  $(\mathbf{y}_i, \mathbf{u}_i)$  and the last equality follows from  $E(r_{ij}|\mathbf{y}_i, \mathbf{u}_i) = \pi_j(\mathbf{y}_i, \mathbf{u}_i, \boldsymbol{\theta}_j)$ , under (A1)-(A2). The consistency and asymptotic normality of  $\hat{\mu}_j(\boldsymbol{\theta})$  as  $n \rightarrow \infty$  can be established by applying standard arguments and the central limit theorem, under some moment conditions, since the right hand side of (1) is a ratio of sums of independent random variables.

In this way, other population characteristics can be similarly estimated. For example, if we want to estimate the distribution of the  $j$ th component of  $\mathbf{y}$  at a point  $t$ , then we just need to replace  $y_{ij}$  by the indicator of  $y_{ij} \leq t$  in the previous discussion. Quantiles can then be estimated. Estimators of correlation between two components of  $\mathbf{y}$  and between  $\mathbf{y}$  and  $\mathbf{x}$  can be similarly derived. We can also estimate parameters defined by some estimating equations.

## 2.2 Estimation of $\boldsymbol{\theta}$

To complete our proposed methodology we need to remove the assumption that  $\boldsymbol{\theta}$  is known, by constructing an estimator  $\hat{\boldsymbol{\theta}}_j$  of  $\boldsymbol{\theta}_j$  for each  $j$  under (A1)-(A2). To estimate  $\boldsymbol{\theta}_j$ , we follow the approach of generalized method of moments (GMM) in Wang et al. (2014) for the univariate response, but we need to add a novel modification to handle the multivariate  $\mathbf{y}$ .

A brief description of the GMM is as follows. Let  $\boldsymbol{\varphi}$  be the parameter vector to estimate, which is a unique solution to  $E\{\mathbf{g}(\boldsymbol{\varphi})\} = 0$  with an  $l$ -dimensional vector estimating function  $\mathbf{g}$  whose  $t$ th component is  $g_t(\mathbf{y}, \mathbf{x}, \mathbf{r}, \boldsymbol{\varphi})$ ,  $t = 1, \dots, l$ . The functions  $g_1, \dots, g_l$  are chosen so that  $l$  is not less than the dimension of  $\boldsymbol{\varphi}$  and at the true parameter value  $\boldsymbol{\varphi}$ ,  $E\{\partial \mathbf{g}(\boldsymbol{\varphi})/\partial \boldsymbol{\varphi}\}$  is of full rank. Let  $\mathbf{g}_n(\boldsymbol{\varphi})$  be the  $l$ -dimensional vector whose  $t$ th component is the sample average  $n^{-1} \sum_{i=1}^n g_t(\mathbf{y}_i, \mathbf{x}_i, \mathbf{r}_i, \boldsymbol{\varphi})$ ,  $t = 1, \dots, l$ . If  $l$  is the same as the dimension of  $\boldsymbol{\varphi}$ , then we estimate  $\boldsymbol{\varphi}$  by  $\hat{\boldsymbol{\varphi}}$  such that  $\mathbf{g}_n(\hat{\boldsymbol{\varphi}}) = 0$ . If  $l$  is larger than the dimension of  $\boldsymbol{\varphi}$ , we apply the following two-step GMM (Hansen, 1982; Hall, 2005):

1. Obtain  $\tilde{\boldsymbol{\varphi}}$  by minimizing  $\{\mathbf{g}_n(\boldsymbol{\varphi})\}^T \mathbf{g}_n(\boldsymbol{\varphi})$ , where  $\mathbf{a}^T$  is the transpose of column vector  $\mathbf{a}$ .
2. Obtain  $\hat{\boldsymbol{\varphi}}$  by minimizing  $\{\mathbf{g}_n(\boldsymbol{\varphi})\}^T \widehat{\mathbf{W}} \mathbf{g}_n(\boldsymbol{\varphi})$ , where  $\widehat{\mathbf{W}}$  is the inverse of  $l \times l$  matrix whose  $(t, t')$  element is  $n^{-1} \sum_{i=1}^n g_t(\mathbf{y}_i, \mathbf{x}_i, \mathbf{r}_i, \tilde{\boldsymbol{\varphi}}) g_{t'}(\mathbf{y}_i, \mathbf{x}_i, \mathbf{r}_i, \tilde{\boldsymbol{\varphi}})$ .

The optimization can be solved by using the MATLAB or R function `fminsearch`.

For our problem, it remains to specify the form of the estimating function  $\mathbf{g}$ . Suppose first that the nonresponse instrument  $\mathbf{z}$  is discrete and has  $s$  categories, say  $\mathbf{z} \in \{z_1, \dots, z_s\}$ . A

straightforward extension of the approach in Wang et al. (2014) (from univariate response to multivariate  $\mathbf{y}$ ) is using

$$\mathbf{g}(\boldsymbol{\theta}) = \left\{ \frac{r_1 \cdots r_k}{\pi_1(\mathbf{y}, \mathbf{u}, \boldsymbol{\theta}_1) \cdots \pi_k(\mathbf{y}, \mathbf{u}, \boldsymbol{\theta}_k)} - 1 \right\} \mathbf{v}, \quad (2)$$

where  $\boldsymbol{\theta} = (\boldsymbol{\theta}_1^T, \dots, \boldsymbol{\theta}_k^T)^T$ ,  $r_j$  is the  $j$ th component of the vector  $\mathbf{r}$  of response indicators and  $\mathbf{v}$  is the  $(s + q)$ -dimensional vector whose first  $s$  components are indicators of  $\mathbf{z} = z_t$ ,  $t = 1, \dots, s$ , and the rest  $q$  components are the  $q$ -dimensional covariate vector  $\mathbf{u}$  in (A1)-(A2). With this choice of  $\mathbf{g}$ ,  $E\{\mathbf{g}(\boldsymbol{\theta})\} = 0$  under (A1)-(A2).

However, there is a problem:  $l = s + q$  may be smaller than the dimension of  $\boldsymbol{\theta}$ . For example, if  $\mathbf{u}$  is continuous and

$$\pi_j(\mathbf{y}, \mathbf{u}, \boldsymbol{\theta}_j) = \{1 + \exp(\alpha_j + \boldsymbol{\beta}_j^T \mathbf{y} + \boldsymbol{\gamma}_j^T \mathbf{u})\}^{-1}, \quad j = 1, \dots, k, \quad (3)$$

where  $\alpha_j$  is univariate,  $\boldsymbol{\beta}_j$  is  $k$ -dimensional,  $\boldsymbol{\gamma}_j$  is  $q$ -dimensional, and  $\boldsymbol{\theta}_j = (\alpha_j, \boldsymbol{\beta}_j^T, \boldsymbol{\gamma}_j^T)^T$  with dimension  $q + k + 1$ , then  $l = s + q \geq k(q + k + 1)$  (the dimension of  $\boldsymbol{\theta}$ ) means that  $s \geq (k - 1)q + k(k + 1)$ , which may be unrealistic. For instance, when  $q = 0$  (there is no  $\mathbf{u}$ ),  $s \geq k(k + 1)$  requires that  $\mathbf{z}$  has at least  $k(k + 1)$  categories.

To overcome this difficulty we consider the following modification. First, we construct  $k$  overlapped subsets  $D_1, \dots, D_k$  of the entire data set, where  $D_h$  contains data from units whose  $y_{ih}$  may be missing but all other components are observed,  $h = 1, \dots, k$ . With the notation  $r_j$  = the  $j$ th component of  $\mathbf{r}$ ,  $D_h = \{r_1 = \cdots = r_{h-1} = r_{h+1} = \cdots = r_k = 1\}$ . Table 1 provides an example of  $D_1, D_2, D_3$  in the case of  $k = 3$  and  $n = 30$ .

Then, we estimate  $\boldsymbol{\theta}_j$  one at a time,  $j = 1, \dots, k$ . For each  $j$ , we use data in  $D_j$  and estimating function

$$\mathbf{g}^{(j)}(\boldsymbol{\theta}_j) = \left\{ \frac{r_j}{\pi_j(\mathbf{y}, \mathbf{u}, \boldsymbol{\theta}_j)} - 1 \right\} \delta_j \mathbf{v}_j, \quad (4)$$

where  $\delta_j$  is the indicator of set  $D_j$ ,  $\mathbf{v}_j$  is the vector whose first  $s + q$  components are the same as those of  $\mathbf{v}$  in (2), the rest  $k - 1$  components are  $y_1, \dots, y_{j-1}$ ,  $y_{j+1}, \dots, y_k$ , and  $y_t$  is the  $t$ th component of  $\mathbf{y}$ . Note that  $\mathbf{g}^{(j)}(\boldsymbol{\theta}_j)$  in (4) can always be computed, since when  $\delta_j = 1$ , all  $y_t$  with  $t \neq j$  are observed.

Besides the fact that the estimating function  $\mathbf{g}$  in (2) involves all  $\boldsymbol{\theta}_j$ 's and the estimating function  $\mathbf{g}^{(j)}$  in (4) involves  $\boldsymbol{\theta}_j$  only, another key difference between  $\mathbf{g}$  and  $\mathbf{g}^{(j)}$  is that the observed components of  $\mathbf{y}$  other than the  $j$ th component are used as ‘‘covariates’’ and included

in the vector  $\mathbf{v}_j$  in (4). In this way, we not only make use of the partially observed responses in  $\mathbf{y}$  (note that  $r_1 \cdots r_k = 1$  if and only if all components of  $\mathbf{y}$  are observed), but also include more components in the estimating function so that  $l = s + q + k - 1$  is typically large enough for our purpose of estimating  $\boldsymbol{\theta}_j$ . For example, in the case of (3), the dimension of  $\boldsymbol{\theta}_j$  is  $q + k + 1$ ; hence,  $l = s + q + k - 1 \geq q + k + 1$  is the same as  $s \geq 2$ , which naturally holds as long as  $\mathbf{z}$  is not a constant. However, if we do not include the last  $k - 1$  components in  $\mathbf{v}_j$ , i.e.,  $\mathbf{v}_j$  in (4) is replaced by  $\mathbf{v}$  defined in (2), then the dimension of  $\mathbf{g}^{(j)}$  is  $s + q$ , which is smaller than the dimension of  $\boldsymbol{\theta}_j$  in the case of (3) unless  $s \geq k + 1$ . Therefore, using  $\mathbf{v}_j$  instead of  $\mathbf{v}$  ensures that our procedure has a larger scope in application.

A GMM estimator  $\widehat{\boldsymbol{\theta}}_j$  of  $\boldsymbol{\theta}_j$  can be computed using the estimating function  $\mathbf{g}^{(j)}$  in (4) and data set  $D_j$ . Since  $D_j$ 's are overlapped, some data are repeatedly used in the estimation of  $\widehat{\boldsymbol{\theta}}_1, \dots, \widehat{\boldsymbol{\theta}}_k$ .

To see why the function  $\mathbf{g}^{(j)}(\boldsymbol{\theta}_j)$  in (4) produces asymptotically valid estimator of  $\boldsymbol{\theta}_j$ , note that

$$\begin{aligned} E\{\mathbf{g}^{(j)}(\boldsymbol{\theta}_j)\} &= E \left[ E \left[ \left\{ \frac{r_j}{\pi_j(\mathbf{y}, \mathbf{u}, \boldsymbol{\theta}_j)} - 1 \right\} \delta_j \mathbf{v}_j \middle| \mathbf{y}, \mathbf{u}, \delta_j \right] \right] \\ &= E \left[ \left\{ \frac{E(r_j | \mathbf{y}, \mathbf{u}, \delta_j)}{\pi_j(\mathbf{y}, \mathbf{u}, \boldsymbol{\theta}_j)} - 1 \right\} \delta_j E(\mathbf{v}_j | \mathbf{y}, \mathbf{u}, \delta_j) \right] \\ &= 0, \end{aligned}$$

where the second equality follows from the independence between  $\mathbf{z}$  and  $r_j$  conditioned on  $(\mathbf{y}, \mathbf{u}, \delta_j)$  and the last equality follows from  $E(r_j | \mathbf{y}, \mathbf{u}, \delta_j) = E(r_j | \mathbf{y}, \mathbf{u}) = \pi_j(\mathbf{y}, \mathbf{u}, \boldsymbol{\theta}_j)$  under (A1)-(A2).

When  $\mathbf{z}$  has continuous components, we can apply the method by discretizing  $\mathbf{z}$  into  $s$  categories or use  $s$  moments of  $\mathbf{z}$  as components of  $\mathbf{v}$ . If  $\mathbf{z} = z$  is a univariate continuous covariate, then typically we use  $s = 2$  with the first and second components of  $\mathbf{v}_j$  being 1 and  $z$ , respectively.

Once  $\widehat{\boldsymbol{\theta}}_1, \dots, \widehat{\boldsymbol{\theta}}_k$  are obtained, we estimate  $\mu_j$  by  $\widehat{\mu}_j(\widehat{\boldsymbol{\theta}})$ , obtained by substituting  $\boldsymbol{\theta}$  in  $\widehat{\mu}_j(\boldsymbol{\theta})$  in (1) with  $\widehat{\boldsymbol{\theta}} = (\widehat{\boldsymbol{\theta}}_1^T, \dots, \widehat{\boldsymbol{\theta}}_k^T)^T$ . Note that the entire  $\widehat{\boldsymbol{\theta}}$ , not just  $\widehat{\boldsymbol{\theta}}_j$ , should be used in  $\widehat{\mu}_j(\widehat{\boldsymbol{\theta}})$  according to (1).



## 2.3 Asymptotic theory

Under the same regularity conditions assumed in Wang et al. (2014), consistency and asymptotic normality of  $\widehat{\boldsymbol{\theta}}_j$  can be established and details are omitted. For the point estimator  $\widehat{\mu}_j(\widehat{\boldsymbol{\theta}})$ , its consistency and asymptotic normality can be established. We provide the main argument below and omit the details of proof. Define

$$\psi(\mathbf{y}_i, \mathbf{u}_i, \boldsymbol{\theta}) = \frac{1}{\pi_1(\mathbf{y}_i, \mathbf{u}_i, \boldsymbol{\theta}_1) \cdots \pi_k(\mathbf{y}_i, \mathbf{u}_i, \boldsymbol{\theta}_k)},$$

$$\tau(\boldsymbol{\theta}) = \frac{1}{n} \sum_{i=1}^n (r_{i1} \cdots r_{ik}) \psi(\mathbf{y}_i, \mathbf{u}_i, \boldsymbol{\theta}),$$

and

$$\zeta_j(\boldsymbol{\theta}) = \frac{1}{n} \sum_{i=1}^n (r_{i1} \cdots r_{ik}) y_{ij} \psi(\mathbf{y}_i, \mathbf{u}_i, \boldsymbol{\theta}).$$

Then, by (1),  $\widehat{\mu}_j(\widehat{\boldsymbol{\theta}}) = \zeta_j(\widehat{\boldsymbol{\theta}})/\tau(\widehat{\boldsymbol{\theta}})$  and

$$\sqrt{n}\{\widehat{\mu}_j(\widehat{\boldsymbol{\theta}}) - \mu_j\} = \frac{1}{\tau(\widehat{\boldsymbol{\theta}})} \left[ \sqrt{n}\{\zeta_j(\widehat{\boldsymbol{\theta}}) - \mu_j\} - \mu_j \sqrt{n}\{\tau(\widehat{\boldsymbol{\theta}}) - 1\} \right].$$

Assume that  $\nabla\psi(\mathbf{y}, \mathbf{u}, \boldsymbol{\theta}) = \partial\psi(\mathbf{y}, \mathbf{u}, \boldsymbol{\theta})/\partial\boldsymbol{\theta}$  exists and each component of  $\nabla\psi(\mathbf{y}, \mathbf{u}, \boldsymbol{\theta}) - \nabla\psi(\mathbf{y}, \mathbf{u}, \boldsymbol{\theta})$  is bounded in absolute value by  $H(\mathbf{y}, \mathbf{u})\|\boldsymbol{\theta} - \boldsymbol{\theta}\|$  with  $E\{H(\mathbf{y}, \mathbf{u})\} < \infty$ , where  $\|\cdot\|$  is the  $L_2$  norm. This assumption holds if  $\pi_j(\mathbf{y}, \mathbf{u}, \boldsymbol{\theta}_j)$ 's are given by (3). Then, by the consistency of the GMM estimator  $\widehat{\boldsymbol{\theta}}$ ,

$$\begin{aligned} \sqrt{n}\{\zeta_j(\widehat{\boldsymbol{\theta}}) - \mu_j\} &= \sqrt{n}\{\zeta_j(\boldsymbol{\theta}) - \mu_j\} + \sqrt{n}\{\zeta_j(\widehat{\boldsymbol{\theta}}) - \zeta_j(\boldsymbol{\theta})\} \\ &= \sqrt{n}\{\zeta_j(\boldsymbol{\theta}) - \mu_j\} + \frac{1}{\sqrt{n}} \sum_{i=1}^n (r_{i1} \cdots r_{ik}) y_{ij} \left\{ \psi(\mathbf{y}_i, \mathbf{u}_i, \widehat{\boldsymbol{\theta}}) - \psi(\mathbf{y}_i, \mathbf{u}_i, \boldsymbol{\theta}) \right\} \\ &= \sqrt{n}\{\zeta_j(\boldsymbol{\theta}) - \mu_j\} + \frac{1}{\sqrt{n}} \sum_{i=1}^n (r_{i1} \cdots r_{ik}) y_{ij} \nabla\psi(\mathbf{y}_i, \mathbf{u}_i, \boldsymbol{\theta})(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}) + o_p(1) \\ &= \sqrt{n}\{\zeta_j(\boldsymbol{\theta}) - \mu_j\} + \left\{ \frac{1}{n} \sum_{i=1}^n (r_{i1} \cdots r_{ik}) y_{ij} \nabla\psi(\mathbf{y}_i, \mathbf{u}_i, \boldsymbol{\theta}) \right\} \sqrt{n}(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}) + o_p(1) \\ &= \sqrt{n} \left\{ \frac{1}{n} \sum_{i=1}^n (r_{i1} \cdots r_{ik}) y_{ij} \psi(\mathbf{y}_i, \mathbf{u}_i, \boldsymbol{\theta}) - \mu_j \right\} + A(\boldsymbol{\theta}) \sqrt{n}(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}) + o_p(1), \end{aligned}$$

where  $A(\boldsymbol{\theta}) = E\{(r_{i1} \cdots r_{ik}) y_{ij} \nabla\psi(\mathbf{y}_i, \mathbf{u}_i, \boldsymbol{\theta})\}$ ,  $o_p(1)$  denotes a term converging to 0 in probability as  $n \rightarrow \infty$ , and the last equality follows from the law of large numbers and the definition

of  $\zeta_j(\boldsymbol{\theta})$ . Similarly,

$$\begin{aligned}
\sqrt{n}\{\tau(\hat{\boldsymbol{\theta}}) - 1\} &= \sqrt{n}\{\tau(\boldsymbol{\theta}) - 1\} + \sqrt{n}\{\tau(\hat{\boldsymbol{\theta}}) - \tau(\boldsymbol{\theta})\} \\
&= \sqrt{n}\{\tau(\boldsymbol{\theta}) - 1\} + \frac{1}{\sqrt{n}} \sum_{i=1}^n (r_{i1} \cdots r_{ik}) \left\{ \psi(\mathbf{y}_i, \mathbf{u}_i, \hat{\boldsymbol{\theta}}) - \psi(\mathbf{y}_i, \mathbf{u}_i, \boldsymbol{\theta}) \right\} \\
&= \sqrt{n}\{\tau(\boldsymbol{\theta}) - 1\} + \left\{ \frac{1}{n} \sum_{i=1}^n (r_{i1} \cdots r_{ik}) \nabla \psi(\mathbf{y}_i, \mathbf{u}_i, \boldsymbol{\theta}) \right\} \sqrt{n}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}) + o_p(1) \\
&= \sqrt{n} \left\{ \frac{1}{n} \sum_{i=1}^n (r_{i1} \cdots r_{ik}) \psi(\mathbf{y}_i, \mathbf{u}_i, \boldsymbol{\theta}) - 1 \right\} + B(\boldsymbol{\theta}) \sqrt{n}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}) + o_p(1),
\end{aligned}$$

where  $B(\boldsymbol{\theta}) = E\{(r_{i1} \cdots r_{ik}) \nabla \psi(\mathbf{y}_i, \mathbf{u}_i, \boldsymbol{\theta})\}$ . From the theory in Wang et al. (2014), the GMM estimator  $\hat{\boldsymbol{\theta}}$  has the property that

$$\sqrt{n}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}) = \frac{1}{\sqrt{n}} \sum_{i=1}^n \boldsymbol{\phi}(\mathbf{y}_i, \mathbf{x}_i) + o_p(1), \quad (5)$$

where  $\boldsymbol{\phi}$  is an unknown vector function with  $E\{\boldsymbol{\phi}(\mathbf{y}, \mathbf{x})\} = 0$  and a finite positive definite matrix  $E\{\boldsymbol{\phi}(\mathbf{y}, \mathbf{x})\boldsymbol{\phi}(\mathbf{y}, \mathbf{x})^T\}$ . Then the asymptotic normality of  $\sqrt{n}\{\hat{\mu}_j(\hat{\boldsymbol{\theta}}) - \mu_j\}$  with asymptotic mean 0 follows from the joint asymptotic normality of the following vector,

$$\frac{1}{\sqrt{n}} \sum_{i=1}^n \begin{pmatrix} (r_{i1} \cdots r_{ik}) y_{ij} \psi(\mathbf{y}_i, \mathbf{u}_i, \boldsymbol{\theta}) - \mu_j \\ (r_{i1} \cdots r_{ik}) \psi(\mathbf{y}_i, \mathbf{u}_i, \boldsymbol{\theta}) - 1 \\ \boldsymbol{\phi}(\mathbf{y}_i, \mathbf{x}_i) \end{pmatrix} \quad (6)$$

However, the asymptotic variance of  $\sqrt{n}\{\hat{\mu}_j(\hat{\boldsymbol{\theta}}) - \mu_j\}$  is very complicated, because it involves not only the asymptotic variances of the three components in (6), but also their asymptotic covariances, and the form of function  $\boldsymbol{\phi}$  in (5) is complicated (Wang et al., 2014). Thus, we do not try to obtain an explicit form of the asymptotic variance of  $\hat{\mu}_j(\hat{\boldsymbol{\theta}})$ . Instead, we recommend the bootstrap method for variance estimation or inference. Using the previous arguments, we can show that the bootstrap analog  $\hat{\mu}_j^*(\hat{\boldsymbol{\theta}}^*)$  is asymptotically normal and the general bootstrap theory (Shao and Tu, 1995) ensures that the bootstrap variance estimators are consistent. Applying the bootstrap effectively avoids the complicated derivation of asymptotic variances, at the expense of a large amount of computations. In Section 3, the performance of bootstrap standard error (squared root of variance estimator) is evaluated by simulations.

### 3 Simulation Results

We carry out a simulation study to investigate the finite sample performance of our proposed  $\hat{\mu}_j(\hat{\boldsymbol{\theta}})$  given by (1) as an estimator of the marginal population mean  $\mu_j = E(y_j)$ ,  $j = 1, \dots, k$ , with  $\boldsymbol{\theta}$  estimated by the GMM estimator  $\hat{\boldsymbol{\theta}}$  derived in Section 2.2.

We consider a panel size  $k = 4$  and sample size  $n = 1,200$ , reflexing the panel and sample sizes in the real data AIDS Clinical Trial Group 193A example presented in Section 4. A univariate and continuous covariate  $z$  is considered with  $\log z \sim N(2.9, 1)$ . Given  $z$ ,  $y_j$ 's are conditionally independent,  $\log y_1 \sim N(0.4 + 0.9 \log z, 0.8^2)$ ,  $\log y_2 \sim N(0.6 + 0.8 \log z, 0.8^2)$ ,  $\log y_3 \sim N(0.8 + 0.7 \log z, 0.8^2)$ , and  $\log y_4 \sim N(0.9 + 0.6 \log z, 0.8^2)$ . The true marginal means are  $\mu_1 = 41.89$ ,  $\mu_2 = 35.16$ ,  $\mu_3 = 29.81$ , and  $\mu_4 = 23.10$ . These  $\mu_j$ 's are chosen to match the estimated values in the real data example considered in Section 4.

The nonresponse propensity is given by (3) with  $\mathbf{u} = 0$ ,  $\alpha_j$ 's and  $\beta_j$ 's shown in Tables 2-3 for settings 1-2, respectively. The parameter values  $\alpha_j$  and  $\beta_j$  are chosen so that the unconditional nonresponse probability matches the observed proportion in the real data example for every  $j$ . The difference in two settings is that all coefficients in front of  $y_j$ 's in the propensity (3) are positive in setting 1 so that larger values of  $y_j$  have a higher probability to be nonresponse, whereas in setting 2, the coefficients may be positive or negative. The covariate  $z$  in the real data example is the baseline response and is used as nonresponse instrument in the estimation.

To evaluate the performance, we include two other estimators, the naive estimator = the sample mean of observed values of  $y_j$  and the sample mean of  $y_j$  with full data (no nonresponse) available in the simulation as nonresponse is constructed. The naive estimator is theoretically biased due to nonignorable nonresponse and is included to see the effect of bias; the full data sample mean is used as a standard.

Based on 1,000 simulation runs, Tables 2-3 report, for settings 1-2 respectively, simulation average of estimates of  $\mu_j$ , bias, bias in percentage, standard deviation of the estimate, average of the standard error obtained by bootstrap with size 200, and coverage probability of the approximate 95% confidence interval with limits = estimate  $\pm 1.96$ (bootstrap standard error). Results are given for  $j = 1, 2, 3, 4$  and three estimators, based on the proposed, naive, and full data methods. In the calculation of the proposed estimator given by (1), the GMM estimator  $\hat{\boldsymbol{\theta}}_j$  is calculated using the MATLAB or R function `fminsearch` with initial value  $\boldsymbol{\theta}_j = 0$ . Since  $z$  is continuous, we use  $(1, z)$  as the first two components of  $\mathbf{v}_j$  in (4).

From the simulation results in Tables 2-3, the performance of proposed estimator (1) can be summarized as follows. It has negligible bias: the largest biases are 3.1% and 2.4% and the rest of biases are all smaller than 2%. The coverage probability of the related confidence interval is close to 95%; the worst cases are in setting 1: 0.932 when  $j = 1$  and 0.934 when  $j = 2$ , but even the full data approach may also have coverage probabilities 0.932 and 0.936. The bootstrap standard error for the proposed method performs well in general, and is sometimes a little bit conservative, which results in slightly conservative coverage probability of the confidence interval.

In setting 1 where larger  $y_j$  values have higher probability to be missing data, the naive estimator has a negative bias. Although the bias is around  $-5\%$  only, it still affects considerably the coverage probability of the related confidence interval. In setting 2, when smaller  $y_j$  values have higher probability to be missing data ( $j = 2$  or  $4$ ), the naive estimator has a small positive bias  $= 2.3\%$  and  $3.7\%$  so that its coverage probability is acceptable. This appears by luck but cannot support the naive approach of ignoring nonignorable nonresponse.

## 4 A Real Data Example

For illustration, we apply our proposed estimation method to the AIDS Clinical Trial Group 193A data set, which can be found at <https://www.hsph.harvard.edu/fitzmaur/ala/cd4.txt>. Longitudinal responses, the CD4 cell counts, were collected from HIV-AIDS patients with advanced immune suppression. After removing some patients with abnormal data, we focus on 1,271 patients with responses in four time intervals,  $(4,12]$ ,  $(12,20]$ ,  $(20,28]$ ,  $(28,36]$ , denoted as  $y_1, y_2, y_3, y_4$ .

The longitudinal response  $\mathbf{y} = (y_1, \dots, y_4)^T$  has item nonresponse, as summarized in Table 4. The item nonresponse is due to adverse events, low-grade toxic reactions, the desire to seek other therapies, death, and some other reasons. Previous experiences from doctors and Cho et al. (2016) found that a steep decline in the CD4 cell count indicates the disease progression, and patients with low CD4 cell counts are more likely to miss the scheduled study visits as compared to patients with normal CD4. Therefore, nonresponse of the CD4 cell count is likely related to itself and is nonignorable (Cho et al., 2016; Yuan and Yin, 2010).

We apply our proposed method in Section 2 to estimate  $\mu_j = E(y_j)$ ,  $j = 1, \dots, 4$ , with

the always observed  $z$  = the baseline CD4 measurement as the instrument described in (A1). Since  $z$  is the baseline CD4 cell count and  $\mathbf{y}$  is the after-baseline CD4 cell count vector, based on the reason of nonresponse described previously, it is reasonable to assume that the item nonresponse of  $\mathbf{y}$  is unrelated with the baseline  $z$  once we conditioned on  $\mathbf{y}$ , i.e., (A1) holds with  $z$  as an instrument. To apply the proposed method, we assume model (3) with  $\mathbf{x} = z$ , i.e., there is no other covariate.

The proposed estimates for  $j = 1, \dots, 4$  are shown in Table 5, together with their bootstrap standard errors with bootstrap size 200. For comparison, we also include in Table 5 the sample mean of observed values of  $y_j$  (naive estimate ignoring nonresponse), the differences between the proposed and naive estimates, and the bootstrap standard errors for differences.

From Table 5, the proposed estimates show a more serious decline in CD4 cell count over the time than naive estimates, although naive estimates also indicate the decline. Compared with 2 times the standard error, the difference between the proposed and naive estimates is significant at  $j = 4$ .

## Acknowledgement

The authors' research was partially supported by the National Natural Science Foundation of China grant 11831008 and the U.S. National Science Foundation grant DMS-1914411.

## References

- Cho, H., Hong, H.G., and Kim, M.O. (2016). Efficient quantile marginal regression for longitudinal data with dropouts. *Biostatistics*, **17**, 561–575.
- Greenlees, J.S., Reece, W.S. and Zieschang, K.D. (1982). Imputation of missing values when the probability of response depends on the variable being imputed. *Journal of the American Statistical Association*, **77**, 251–261.
- Hansen, L. (1982). Large sample properties of generalized method of moments estimators. *Econometrica*, **50**, 1029–1054.
- Hall, A. R. (2005). *Generalized Method of Moments*. Oxford University Press, New York.

- Li, S. and Shao, J. (2021). Nonignorable item nonresponse in panel data. *Statistical Theory and Related Fields*, **6**, to appear.
- Little, R.J.A. and Rubin, D.B. (2002). *Statistical Analysis with Missing Data*, second edition. Wiley, New York.
- Rubin D.B. (1976). Inference with Missing Data. *Biometrika*, **63**, 581–592.
- Robins, J.M. and Rotiv, Y. (1997). Toward a curse of dimensionality appropriate (CODA) asymptotic theory for semiparametric models. *Statistics in Medicine*, **16**, 285–319.
- Scharfstein, D., Rotnitzky, A. and Robins, J. (1999). Point processes and multivariate extreme values. *Journal of the American Statistical Association*, **94**, 1096–1120.
- Shao, J. and Tu, D. (1995). *The Jackknife and Bootstrap*. Springer-Verlag, New York.
- Shao, J. and Wang, L. (2016). Semiparametric inverse propensity weighting for nonignorable missing data. *Biometrika*, **103**, 175–187.
- Shao, J. and Zhang, J. (2015). A transformation approach in linear mixed-effect models with informative missing responses. *Biometrika*, **102**, 107–119.
- Tang, G., Little, R.J.A. and Raghunathan, T.E. (2003). Analysis of multivariate missing data with nonignorable nonresponse. *Biometrika*, **90**, 747–764.
- Wu, M.C. and Carroll, R.J. (1988). Estimation and Comparison of changes in the presence of informative right censoring by modeling the censoring process, *Biometrics*, **44**, 175–188.
- Wang, S., Shao, J. and Kim, J.K. (2014). An instrumental variable approach for identification and estimation with nonignorable nonresponse. *Statistica Sinica*, **24**, 1097–1116.
- Xu, L. and Shao, J. (2009). Estimation in longitudinal or panel data models with random-effect-based missing responses. *Biometrics*, **65**, 1175–1183.
- Yuan, Y. and Yin, G. (2010). Bayesian quantile regression for longitudinal studies with nonignorable missing data. *Biometrics*, **66**, 105–114.

Zhao, J. and Shao, J. (2015). Semiparametric pseudo likelihoods in generalized linear models with nonignorable missing data. *Journal of American Statistical Association*, **110**, 1577–1590.

Table 1. Example of  $D_1, D_2, D_3$  when  $k = 3$  and  $n = 30$   
( $r_j$  is the indicator of whether  $y_j$  is observed)

entire dataset				$D_1$				$D_2$				$D_3$			
unit	$r_1$	$r_2$	$r_3$	unit	$r_1$	$r_2$	$r_3$	unit	$r_1$	$r_2$	$r_3$	unit	$r_1$	$r_2$	$r_3$
1	0	0	0	2	1	1	1	2	1	1	1	2	1	1	1
2	1	1	1	3	0	1	1	5	1	1	1	5	1	1	1
3	0	1	1	5	1	1	1	8	1	1	1	7	1	1	0
4	1	0	0	8	1	1	1	11	1	0	1	8	1	1	1
5	1	1	1	12	1	1	1	12	1	1	1	10	1	1	0
6	0	0	1	15	1	1	1	14	1	0	1	12	1	1	1
7	1	1	0	16	0	1	1	15	1	1	1	15	1	1	1
8	1	1	1	17	1	1	1	17	1	1	1	17	1	1	1
9	0	0	1	21	1	1	1	21	1	1	1	18	1	1	0
10	1	1	0	23	1	1	1	22	1	0	1	20	1	1	0
11	1	0	1	24	0	1	1	23	1	1	1	21	1	1	1
12	1	1	1	28	1	1	1	27	1	0	1	23	1	1	1
13	0	1	0					28	1	1	1	28	1	1	1
14	1	0	1					30	1	0	1	29	1	1	0
15	1	1	1												
16	0	1	1												
17	1	1	1												
18	1	1	0												
19	0	0	0												
20	1	1	0												
21	1	1	1												
22	1	0	1												
23	1	1	1												
24	0	1	1												
25	0	0	0												
26	0	1	0												
27	1	0	1												
28	1	1	1												
29	1	1	0												
30	1	0	1												



Table 2. Simulation results for the estimation of  $\mu_j$  in setting 1  
( $n = 1, 200$ , bootstrap size = 200, simulation runs = 1,000)

$j$	% of missing	method	estimate	bias	bias %	standard deviation	standard error	coverage prob.
1	30.75	proposed	42.44	0.552	0.013	7.705	7.896	0.932
		naive	39.76	-2.129	-0.051	2.448	2.441	0.784
		full data	41.92	0.034	0.001	2.180	2.150	0.936
2	24.57	proposed	35.76	0.599	0.017	6.246	6.223	0.934
		naive	33.92	-1.241	-0.035	1.836	1.792	0.827
		full data	35.21	0.048	0.001	1.636	1.628	0.943
3	46.66	proposed	30.75	0.933	0.031	6.035	5.911	0.956
		naive	28.09	-1.720	-0.058	1.582	1.583	0.747
		full data	29.91	0.095	0.003	1.236	1.245	0.948
4	39.03	proposed	23.65	0.543	0.024	4.355	4.317	0.956
		naive	22.06	-1.040	-0.045	1.079	1.053	0.780
		full data	23.06	-0.048	-0.002	0.908	0.861	0.932

$$\pi_j(\mathbf{y}, \boldsymbol{\theta}_j) = \{1 + \exp(\alpha_j + \boldsymbol{\beta}_j^T \mathbf{y})\}^{-1}, j = 1, \dots, 4$$

$$\alpha_1 = -1.2, \boldsymbol{\beta}_1^T = (0.1, 0.01, 0.01, 0.01)$$

$$\alpha_2 = -1.5, \boldsymbol{\beta}_2^T = (0.01, 0.1, 0.01, 0.01)$$

$$\alpha_3 = -0.5, \boldsymbol{\beta}_3^T = (0.01, 0.01, 0.1, 0.01)$$

$$\alpha_4 = -0.8, \boldsymbol{\beta}_4^T = (0.01, 0.01, 0.01, 0.1)$$

Table 3. Simulation results for the estimation of  $\mu_j$  in setting 2  
( $n = 1, 200$ , bootstrap size = 200, simulation runs = 1,000)

$j$	% of missing	method	estimate	bias	bias %	standard deviation	standard error	coverage prob.
1	30.47	proposed	42.26	0.375	0.009	6.979	7.176	0.941
		naive	39.58	-2.310	-0.055	2.464	2.382	0.769
		full data	41.98	0.090	0.002	2.138	2.145	0.943
2	22.79	proposed	35.63	0.471	0.013	4.891	5.498	0.961
		naive	35.96	0.796	0.023	1.916	1.857	0.946
		full data	35.18	0.017	0.001	1.682	1.604	0.929
3	48.19	proposed	30.14	0.324	0.011	4.699	5.292	0.960
		naive	28.40	-1.413	-0.047	1.656	1.598	0.797
		full data	29.87	0.052	0.002	1.264	1.227	0.935
4	40.08	proposed	23.53	0.426	0.018	3.734	3.989	0.951
		naive	22.96	0.854	0.037	1.154	1.150	0.926
		full data	23.12	0.019	0.001	0.879	0.862	0.946

$$\pi_j(\mathbf{y}, \boldsymbol{\theta}_j) = \{1 + \exp(\alpha_j + \boldsymbol{\beta}_j^T \mathbf{y})\}^{-1}, j = 1, \dots, 4$$

$$\alpha_1 = -1.3, \boldsymbol{\beta}_1^T = (0.1, 0.02, 0.02, 0.02)$$

$$\alpha_2 = -1.1, \boldsymbol{\beta}_2^T = (0.02, -0.1, 0.02, 0.02)$$

$$\alpha_3 = -0.3, \boldsymbol{\beta}_3^T = (-0.02, 0.02, 0.1, -0.02)$$

$$\alpha_4 = -0.2, \boldsymbol{\beta}_4^T = (0.02, 0.02, -0.02, -0.1)$$

Table 4. Nonresponse pattern of  $\mathbf{y}$  in the example

Nonresponse pattern					Number of observations
$r_1$	$r_2$	$r_3$	$r_4$	$s^\dagger$	
0	0	0	0	0	121
1	0	0	0	1	82
0	1	0	0	1	57
0	0	1	0	1	11
0	0	0	1	1	7
1	1	0	0	2	69
1	0	1	0	2	37
1	0	0	1	2	11
0	1	1	0	2	41
0	1	0	1	2	119
0	0	1	1	2	2
1	1	1	0	3	94
1	1	0	1	3	118
1	0	1	1	3	34
0	1	1	1	3	31
1	1	1	1	4	437
Total					1271

$$^\dagger s = r_1 + r_2 + r_3 + r_4$$

Table 5. Estimates and standard errors of  $\mu_j$ 's in the example

$j$	% of missing	method	estimate	standard error
1	30.61	proposed	38.09	2.670
		naive	35.70	1.416
		difference <sup>†</sup>	2.395	2.158
2	24.00	proposed	32.21	2.132
		naive	34.68	1.490
		difference	-2.468	1.779
3	45.95	proposed	27.44	2.653
		naive	27.68	1.366
		difference	-0.236	2.540
4	40.28	proposed	24.42	1.687
		naive	28.51	1.435
		difference	-4.085	1.667

<sup>†</sup> difference = proposed - naive