Statistica Sinica **28** (2018), 1761-1785
doi:https://doi.org/10.5705/ss.202016.0300

GENERALIZATION OF HECKMAN SELECTION MODEL TO NONIGNORABLE NONRESPONSE USING CALL-BACK INFORMATION

Baojiang Chen, Pengfei Li and Jing Qin

University of Texas Health Science Center at Houston,
University of Waterloo and National Institute of Health

Abstract: Call-back of nonrespondents is common in surveys involving telephone or mail interviews. In general, these call-backs gather information on unobserved responses, so incorporating them can improve the estimation accuracy and efficiency. Call-back studies mainly focus on Alho (1990)'s selection model or the pattern mixture model formulation. In this paper, we generalize the Heckman selection model to nonignorable nonresponses using call-back information. The unknown parameters are then estimated by the maximum likelihood method. The proposed formulation is simpler than Alho's selection model or the pattern mixture model formulation. It can reduce the bias caused by the nonignorably missing mechanism and improve the estimation efficiency by incorporating the call-back information. Further, it provides a marginal interpretation of a covariate effect. Moreover, the regression coefficient of interest is robust to the misspecification of the distribution. Simulation studies are conducted to evaluate the performance of the proposed method. For illustration, we apply the approach to National Health Interview Survey data.

Key words and phrases: Call-back, heckman model, maximum likelihood estimate, nonignorable, nonresponse.

1. Introduction

In survey studies involving phone or mail interviews, the first contact may be unsuccessful, leading to incomplete data. If respondents and nonrespondents tend to give different answers to the same questions, the missingness mechanism is called missing not at random (MNAR) or nonignorable (Little and Rubin (2002)). It is well known that with MNAR or nonignorably missing data, a statistical analysis based solely on the respondents may lead to invalid inference. One popular method for handling nonignorable nonresponse is a selection model, for example the Heckman selection model (Heckman (1976, 1979)). For more discussion of the analysis of nonignorably missing data see Little and Rubin (2002).

To improve the estimation precision and testing power, additional calls are typically made if the first contact fails (e.g., Wood, White and Hotopf (2006); Jackson et al. (2012)). In surveys, information gained from additional calls is paradata, which is defined to be "data about the process by which the survey data were collected" (Groves and Heeringa (2006)). Survey paradata includes the times that interviews were conducted, the length of the interviews, the number of contacts made with each interviewee or the number of attempts to contact the interviewee, the level of reluctance of the interviewee, and the mode of communication (such as phone, internet, email, or in person) (Taylor (2008)). In general, additional calls gather information on the unobserved responses. Appropriately using this information can reduce the estimation bias and improve the estimation efficiency. Therefore, call-backs have been used in many surveys, for example in the Asthma Call-back Survey, sponsored by the National Asthma Control Program of the Centers for Disease Control and Prevention. Call-backs were also used in the National Survey of Family Growth (Grady (1981)), the National Comorbidity Survey (Kessler and Walters (2002)), the American Community Survey (Alexander, Dahl and Weidmann (1997)), and the National Health Interview Survey (NHIS). Motivated by the NHIS example in Section 2, we are interested in incorporating the call-back information to improve the estimation efficiency in regression analysis.

There are two approaches to using call-back information in regression analysis. In the context of selection models, Alho (1990) estimated an informative missing mechanism by modeling the effect of the probability of response at each attempt on the true outcome and related covariates through a logistic regression model. Wood, White and Hotopf (2006) and Jackson, White and Leese (2010); Jackson et al. (2012) further developed this model. In these selection models, the multiple call-backs provide data on the "continuum of resistance" (Lin and Schaeffer (1995); Daniels et al. (2015)). Another commonly used model for missing data is the pattern mixture model (Little (1993, 1995)); it allows for sensitivity analysis (Daniels and Hogan (2000, 2008)). Daniels et al. (2015) proposed a pattern mixture model for the analysis of repeated-attempt designs; it allows the type of sensitivity parameter defined by Daniels and Hogan (2000, 2008).

In parallel to the use of call-back information for regression analysis, there have been developments in its use for other purposes. For example, Potthoff, Manton and Woodbury (1993) proposed a weighting method based on the number of call-backs to reduce nonavailability bias in surveys. Elliott, Little and Lewitzky (2000) showed that using call-back information potentially improves survey efficiency. Gendall and Davis (1993) used call-back data for market research. Starting from Alho (1990)'s model, Qin and Follmann (2014) proposed a semiparametric maximum likelihood method to estimate the mean of the responses using failed contact attempts to adjust for nonignorable nonresponses; this approach is more efficient than Alho's method. Kim and Im (2014) proposed a propensity score adjustment when there are several follow-ups and used the generalized method of moments to estimate the population total. Other pioneering research can be found in Proctor (1977) and Drew and Fuller (1980, 1981).

The Heckman selection model (Heckman (1976, 1979)) has been widely used to reduce bias from nonignorably missing data because it provides a simple formulation of the response and missing-data models. In this model, the missing indicator is assumed to be a manifestation of a latent variable that may be associated with some covariates. The nonignorably missing mechanism is found from the correlation between the response and this latent variable, which is simpler than Alho (1990)'s selection model and the pattern mixture model. Furthermore, the Heckman model provides an estimation of the marginal effect of the covariates on the response, so it is easier to interpret than the pattern mixture model. The estimation of the former model is based on a two-step estimation procedure or the maximum likelihood method.

It is challenging to incorporate information about the multiple attempts made to obtain data to improve the estimation accuracy and efficiency of the Heckman model. Few researchers have explored this problem. In this paper, we propose a model formulation that adapts the Heckman selection model (Heckman (1979)) to incorporate this information. The basic idea is that, in addition to the response and missing-data models, we build a call-back model and assume that the call-back success indicator is a manifestation of a latent variable. We assume a joint distribution of the response and the latent variables from the missing-data and call-back models, and in this way the nonignorably missing mechanism is incorporated. Tunali (1986) proposed a double selection model that is similar to our call-back model, but our proposed model differs in that if an individual responses, we do not observe the call-back information at all. We develop a likelihood-based method for the estimation, and our simulation studies show that it is more efficient than a method based solely on the response and missing-data models. The proposed method is built under a multivariate normality assumption on the joint distribution of the response and latent variables, but we have proved that the estimator of the regression coefficient of interest is robust to the misspecification of the distribution. Our method is more flexible than the method of Alho (1990). In Alho's logistic regression model, the covariate vector and its slope are assumed to be common for all call-backs, but the intercepts are different. This assumption can be too strong because, in some situations, different covariates may affect the probability of different call-backs (see the NHIS example in Section 6), or the effects may be different at different call-backs. Our method weakens these assumptions: it assumes the covariate vector and its slope are different for different call-backs in the call-back models. Furthermore, our method yields a marginal interpretation of the covariate effect. Another advantage is that it can be implemented easily.

The rest of this paper is organized as follows. In Section 2, we introduce the NHIS example. In Section 3, we introduce the Heckman selection model to model the nonignorable nonresponse and the maximum likelihood estimate of the unknown parameters. In Section 4, we discuss the call-back model for a single call-back, derive the maximum likelihood estimate of the unknown parameters, and study the robustness of the estimate. In Section 5, we evaluate the performance of our method via simulation studies. In Section 6, we apply our method to the NHIS data, and in Section 7 we provide some concluding remarks. In the Supplementary Material, we provide the regularity conditions, detailed derivations, and the extension of call-back model to multiple call-backs.

2. National Health Interview Survey

Our work is motivated by the NHIS, a cross-sectional household interview survey initiated in 1957. Its main goals are to monitor the health of the US population, and to track health status, health-care access, and progress toward national health objectives. The sampling and interviewing are continuous, and the data are collected through personal household interviews. The interviewees may refuse to answer the survey or may be unavailable, leading to a low response rate. Since 2006, repeated contacts have been used to obtain extensive information on the nonrespondents (Taylor (2008)). For more information, see `http://www.cdc.gov/nchs/nhis.htm`.

The NHIS data are widely used by the public health research community for epidemiologic and policy analysis. The data are used to characterize those with various health problems, determine barriers to health-care access, and evaluate

Federal health programs (`http://www.cdc.gov/nchs/nhis/about_nhis.htm`). One question of interest is the determination of barriers or predictors that are associated with medical costs. The potential predictors include: family income (divided by $10,000) (FIN), number of family members with limitations (FMAL), family health insurance (FHI) cost (scaled 0–9), and poverty ratio (RAT_CAT2). We use a sample of 2,000 families in the 2011 survey for illustration. Of the 2,000 families, 503 (25.2%) responded the first time, and 756 (37.8%) responded in the first call-back. The high nonrespondent rates may be associated with medical cost; those with higher medical costs may be less likely to provide data, leading to the nonignorably missing mechanism. We propose a generalization of the Heckman selection model to this nonignorable nonresponse problem.

3. Heckman Selection Model for Nonignorable Nonresponse

Consider a sample involving n individuals, $(Y_i, \mathbf{X}_{1i}^\tau)^\tau$, $i = 1, \ldots, n$, where Y_i is an outcome of interest and $\mathbf{X}_{1i}$ is an associated $(p-1) \times 1$ vector of covariates. Consider the linear regression model

$$Y_i = \beta_0 + \mathbf{X}_{1i}^\tau \boldsymbol{\beta}_1 + \sigma \epsilon_{1i}, \tag{3.1}$$

where $\boldsymbol{\beta} = (\beta_0, \boldsymbol{\beta}_1^\tau)^\tau$ is a $p \times 1$ vector of unknown parameters, σ is an unknown parameter, and ϵ_{1i} is a random error term. It is typically assumed that $\epsilon_{1i} \sim N(0,1)$.

In practice, the outcome Y_i may be missing nonrandomly, and we let R_i be the missing indicator of Y_i, 1 if Y_i is observed and 0 if Y_i is missing. The Heckman selection model (Heckman (1979)) assumes that R_i is a manifestation of a latent variable

$$Z_i = \mathbf{X}_{2i}^\tau \boldsymbol{\gamma} + \epsilon_{2i}, \tag{3.2}$$

where $\mathbf{X}_{2i}$ is a $q \times 1$ vector with the first element 1 and the remaining $q-1$ elements covariates associated with Z_i, $\boldsymbol{\gamma}$ is a $q \times 1$ vector of unknown parameters, and $\epsilon_{2i} \sim N(0,1)$. Specifically, we assume that $R_i = I(Z_i > 0)$, where $I(A)$ is an indicator function, 1 if A is true and 0 otherwise. Furthermore, in the Heckman model, it is typically assumed that $Corr(\epsilon_{1i}, \epsilon_{2i}) = \rho_{12}$ and $(\epsilon_{1i}, \epsilon_{2i})^\tau$ follows a bivariate normal distribution.

Note that

$$\begin{aligned} P(R_i = 1 | Y_i = y_i, \mathbf{X}_{1i}, \mathbf{X}_{2i}) &= P(Z_i > 0 | Y_i = y_i, \mathbf{X}_{1i}, \mathbf{X}_{2i}) \\ &= P(\epsilon_{2i} > -\mathbf{X}_{2i}^\tau \boldsymbol{\gamma} | Y_i = y_i, \mathbf{X}_{1i}, \mathbf{X}_{2i}) \end{aligned}$$

$$= \Phi\left(\frac{\mathbf{X}_{2i}^{\tau}\boldsymbol{\gamma} + \rho_{12}(y_i - \beta_0 - \mathbf{X}_{1i}^{\tau}\boldsymbol{\beta}_1)/\sigma}{\sqrt{1-\rho_{12}^2}}\right),$$

where $\Phi(x)$ is the cumulative distribution function of the standard normal random variable. This means that the Heckman model leads to a nonignorably missing mechanism when $\rho_{12} \neq 0$, since the missing probability depends on y_i.

Heckman (1979) introduced a two-step procedure to estimate the coefficients in the response and missing-data models (3.1) and (3.2). Alternatively, one can estimate the coefficients using a likelihood-based method. The likelihood function of the unknown parameters is

$$L_M(\boldsymbol{\beta}, \boldsymbol{\gamma}, \sigma, \rho_{12}) = \prod_{i=1}^{n} \left[\{P(R_i=1, Y_i=y_i|\mathbf{X}_{1i}, \mathbf{X}_{2i})\}^{R_i} \{P(R_i=0|\mathbf{X}_{1i}, \mathbf{X}_{2i})\}^{1-R_i}\right],$$

where

$$\begin{aligned} & P(R_i = 1, Y_i = y_i|\mathbf{X}_{1i}, \mathbf{X}_{2i}) \\ & = P(R_i = 1|Y_i = y_i, \mathbf{X}_{1i}, \mathbf{X}_{2i}) P(Y_i = y_i|\mathbf{X}_{1i}, \mathbf{X}_{2i}) \\ & = \Phi\left(\frac{\mathbf{X}_{2i}^{\tau}\boldsymbol{\gamma} + \rho_{12}(y_i - \beta_0 - \mathbf{X}_{1i}^{\tau}\boldsymbol{\beta}_1)/\sigma}{\sqrt{1-\rho_{12}^2}}\right) \\ & \quad \times \sigma^{-1}\phi\left(\frac{y_i - \beta_0 - \mathbf{X}_{1i}^{\tau}\boldsymbol{\beta}_1}{\sigma}\right), \qquad (3.3) \\ & P(R_i = 0|\mathbf{X}_{1i}, \mathbf{X}_{2i}) = P(\epsilon_{2i} < -\mathbf{X}_{2i}^{\tau}\gamma|\mathbf{X}_{2i}) = \Phi(-\mathbf{X}_{2i}^{\tau}\gamma). \end{aligned}$$

Here $\phi(x)$ is the probability density function of the standard normal random variable. Consequently, the log-likelihood of the unknown parameters is

$$\begin{aligned} & \ell_M(\boldsymbol{\beta}, \boldsymbol{\gamma}, \sigma, \rho_{12}) \\ & = \sum_{i=1}^{n} \left[R_i \log\left\{\Phi\left(\frac{\mathbf{X}_{2i}^{\tau}\gamma + \rho_{12}(y_i - \beta_0 - \mathbf{X}_{1i}^{\tau}\boldsymbol{\beta}_1)/\sigma}{\sqrt{1-\rho_{12}^2}}\right)\right\} - R_i \log(\sigma)\right. \\ & \quad \left. + R_i \log\left\{\phi\left(\frac{y_i - \beta_0 - \mathbf{X}_{1i}^{\tau}\boldsymbol{\beta}_1}{\sigma}\right)\right\} + (1 - R_i)\log\{\Phi(-\mathbf{X}_{2i}^{\tau}\gamma)\}\right]. \qquad (3.4) \end{aligned}$$

Maximizing (3.4) with respect to $\boldsymbol{\beta}$, $\boldsymbol{\gamma}$, σ, and ρ_{12}, we obtain the maximum likelihood estimators of the unknown parameters:

$$(\widetilde{\boldsymbol{\beta}}, \widetilde{\boldsymbol{\gamma}}, \tilde{\sigma}, \tilde{\rho}_{12}) = \arg \max_{\boldsymbol{\beta}, \boldsymbol{\gamma}, \sigma, \rho_{12}} \ell_M(\boldsymbol{\beta}, \boldsymbol{\gamma}, \sigma, \rho_{12}).$$

4. Incorporating Call-Back Information by Generalizing the Heckman Selection Model

In this section, we discuss how to incorporate call-back information by generalizing the Heckman selection model. We further study the consistency of the estimator of $\boldsymbol{\beta}_1$ in (3.1) under model misspecification. For convenience of presentation, we assume that there is a single call-back. For multiple call-backs, see the Supplementary Material.

4.1. Call-back model and identifiability

Let $D_i = 1$ if the ith subject is called back, and $D_i = 0$ otherwise. In the spirit of the Heckman model, we assume that the call-back indicator D_i is a manifestation of a latent variable model

$$U_i = \mathbf{X}_{3i}^{\tau}\boldsymbol{\xi} + \epsilon_{3i}, \tag{4.1}$$

and $D_i = I(U_i > 0)$, where $\mathbf{X}_{3i}$ is an $r \times 1$ vector with the first element 1 and the remaining $r-1$ elements covariates associated with U_i. We assume that the error term $\epsilon_{3i} \sim N(0,1)$, $Corr(\epsilon_{1i}, \epsilon_{3i}) = \rho_{13}$, and $Corr(\epsilon_{2i}, \epsilon_{3i}) = \rho_{23}$, and that the joint distribution of $(\epsilon_{1i}, \epsilon_{2i}, \epsilon_{3i})^{\tau}$ is trivariate normal. It is easy to verify that the probability subject i is called back conditional on nonresponse depends on the response Y_i if $\rho_{13} \neq 0$, leading to the nonignorably call-back mechanism.

Let $\boldsymbol{\theta} = (\boldsymbol{\beta}^{\tau}, \boldsymbol{\gamma}^{\tau}, \boldsymbol{\xi}^{\tau}, \sigma, \rho_{12}, \rho_{13}, \rho_{23})^{\tau}$ denote the $p+q+r+4$ unknown parameters in models (3.1), (3.2), and (4.1). Throughout the paper, we assume that the components of $(1, \mathbf{X}_1^{\tau})^{\tau}$, the components of $\mathbf{X}_2$, and the components of $\mathbf{X}_3$ are respectively linearly independent. Here $\mathbf{X}_1$, $\mathbf{X}_2$, and $\mathbf{X}_3$ are the covariates for the response, missing-data, and call-back models, respectively.

Proposition 1. *The parameters* $(\boldsymbol{\beta}^{\tau}, \boldsymbol{\gamma}^{\tau}, \sigma, \rho_{12})^{\tau}$ *in the response and missing-data models* (3.1) *and* (3.2) *are always identifiable. If further* $\mathbf{X}_2$ *contains a continuous covariate which does not appear in* $\mathbf{X}_3$*, the parameters* $(\boldsymbol{\xi}^{\tau}, \rho_{13}, \rho_{23})^{\tau}$ *are identifiable.*

For presentational continuity, we have relegated the proof to the Supplementary Material. Here the $(\boldsymbol{\beta}^{\tau}, \boldsymbol{\gamma}^{\tau}, \sigma, \rho_{12})^{\tau}$ are generally identifiable, but the identifiability of $(\boldsymbol{\xi}^{\tau}, \rho_{13}, \rho_{23})^{\tau}$ is only established under the condition that $\mathbf{X}_2$ contains a continuous covariate that does not appear in $\mathbf{X}_3$. The identifiability of $(\boldsymbol{\xi}^{\tau}, \rho_{13}, \rho_{23})^{\tau}$ under weaker conditions becomes much more complicated. We leave it as future research.

4.2. Maximum likelihood method

We now develop the full likelihood function of $\boldsymbol{\theta}$ based on the observed data. If $R_i = 1$, we observe $(R_i = 1, Y_i = y_i, \mathbf{X}_{1i}, \mathbf{X}_{2i}, \mathbf{X}_{3i})$; if $R_i = 0$ and $D_i = 1$, we observe $(R_i = 0, D_i = 1, Y_i = y_i, \mathbf{X}_{1i}, \mathbf{X}_{2i}, \mathbf{X}_{3i})$; and if $R_i = 0$ and $D_i = 0$, we observe $(R_i = 0, D_i = 0, \mathbf{X}_{1i}, \mathbf{X}_{2i}, \mathbf{X}_{3i})$. Therefore, the likelihood function of $\boldsymbol{\theta}$, conditional on all the covariates, is

$$\begin{aligned}L(\boldsymbol{\theta}) = \prod_{i=1}^{n} \Big[&\big\{P(Y_i = y_i, R_i = 1|\mathbf{X}_{1i}, \mathbf{X}_{2i}, \mathbf{X}_{3i})\big\}^{R_i} \\ &\times \big\{P(Y_i = y_i, R_i = 0, D_i = 1|\mathbf{X}_{1i}, \mathbf{X}_{2i}, \mathbf{X}_{3i})\big\}^{(1-R_i)D_i} \\ &\times \big\{P(R_i = 0, D_i = 0|\mathbf{X}_{1i}, \mathbf{X}_{2i}, \mathbf{X}_{3i})\big\}^{(1-R_i)(1-D_i)}\Big]. \qquad (4.2)\end{aligned}$$

By (3.3), the first term in the likelihood (4.2) is

$$\begin{aligned}&P(R_i = 1, Y_i = y_i|\mathbf{X}_{1i}, \mathbf{X}_{2i}, \mathbf{X}_{3i}) \\ &= P(R_i = 1, Y_i = y_i|\mathbf{X}_{1i}, \mathbf{X}_{2i}) \\ &= \Phi\bigg(\frac{\mathbf{X}_{2i}^{\tau}\boldsymbol{\gamma} + \rho_{12}(y_i - \beta_0 - \mathbf{X}_{1i}^{\tau}\boldsymbol{\beta}_1)/\sigma}{\sqrt{1-\rho_{12}^2}}\bigg) \\ &\quad \times \sigma^{-1}\phi\bigg(\frac{y_i - \beta_0 - \mathbf{X}_{1i}^{\tau}\boldsymbol{\beta}_1}{\sigma}\bigg). \qquad (4.3)\end{aligned}$$

The second term in the likelihood (4.2) is

$$\begin{aligned}&P(Y_i = y_i, R_i = 0, D_i = 1|\mathbf{X}_{1i}, \mathbf{X}_{2i}, \mathbf{X}_{3i}) \\ &= P(R_i = 0, D_i = 1|Y_i = y_i, \mathbf{X}_{1i}, \mathbf{X}_{2i}, \mathbf{X}_{3i})P(Y_i = y_i|\mathbf{X}_{1i}, \mathbf{X}_{2i}, \mathbf{X}_{3i}) \\ &= P(\epsilon_{2i} < -\mathbf{X}_{2i}^{\tau}\boldsymbol{\gamma}, \epsilon_{3i} > -\mathbf{X}_{3i}^{\tau}\boldsymbol{\xi}|Y_i = y_i, \mathbf{X}_{1i}, \mathbf{X}_{2i}, \mathbf{X}_{3i})P(Y_i = y_i|\mathbf{X}_{1i}) \\ &= \bigg\{\int_{-\infty}^{-\mathbf{X}_{2i}^{\tau}\boldsymbol{\gamma}}\int_{-\mathbf{X}_{3i}^{\tau}\boldsymbol{\xi}}^{\infty} \phi_{23|1}\bigg(t, u; \frac{(y_i - \beta_0 - \mathbf{X}_{1i}^{\tau}\boldsymbol{\beta}_1)}{\sigma}\bigg)dtdu\bigg\} \\ &\quad \times \bigg\{\sigma^{-1}\phi\bigg(\frac{y_i - \beta_0 - \mathbf{X}_{1i}^{\tau}\boldsymbol{\beta}_1}{\sigma}\bigg)\bigg\}, \qquad (4.4)\end{aligned}$$

where $\phi_{23|1}(t, u; s)$ is the conditional probability density function of $(\epsilon_{2i}, \epsilon_{3i})^{\tau}$ given $\epsilon_{1i} = s$. It can be easily verified that $(\epsilon_{2i}, \epsilon_{3i})^{\tau}|\epsilon_{1i} = s \sim N(\boldsymbol{\mu}_{23|1}, \boldsymbol{\Sigma}_{23|1})$ with

$$\boldsymbol{\mu}_{23|1} = \begin{pmatrix}\rho_{12}s \\ \rho_{13}s\end{pmatrix} \quad \text{and} \quad \boldsymbol{\Sigma}_{23|1} = \begin{pmatrix}1-\rho_{12}^2 & \rho_{23}-\rho_{12}\rho_{13} \\ \rho_{23}-\rho_{12}\rho_{13} & 1-\rho_{13}^2\end{pmatrix}. \qquad (4.5)$$

Hence, $\phi_{23|1}(t, u; s)$ is the probability density function of a bivariate normal random vector from $N(\boldsymbol{\mu}_{23|1}, \boldsymbol{\Sigma}_{23|1})$.

The third term in the likelihood (4.2) is

$$\begin{aligned}&P(R_i = 0, D_i = 0|\mathbf{X}_{1i}, \mathbf{X}_{2i}, \mathbf{X}_{3i}) \\ &= P(R_i = 0, D_i = 0|\mathbf{X}_{2i}, \mathbf{X}_{3i}) \qquad (4.6) \\ &= P(\epsilon_{2i} < -\mathbf{X}_{2i}^{\tau}\boldsymbol{\gamma}, \epsilon_{3i} < -\mathbf{X}_{3i}^{\tau}\boldsymbol{\xi}|\mathbf{X}_{2i}, \mathbf{X}_{3i})\end{aligned}$$

$$= \int_{-\infty}^{-\mathbf{X}_{2i}^{\tau}\boldsymbol{\gamma}} \int_{-\infty}^{-\mathbf{X}_{3i}^{\tau}\boldsymbol{\xi}} \phi_{23}(t,u)dtdu,$$

where $\phi_{23}(t,u)$ is the joint probability density function of $(\epsilon_{2i}, \epsilon_{3i})^{\tau}$, a bivariate normal random vector with mean vector $(0,0)^{\tau}$ and covariance matrix

$$\boldsymbol{\Sigma}_{23} = \begin{pmatrix} 1 & \rho_{23} \\ \rho_{23} & 1 \end{pmatrix}. \tag{4.7}$$

Combining (4.2)–(4.6) and taking logarithms of the likelihood function, we get the log-likelihood function of $\boldsymbol{\theta}$:

$$\ell(\boldsymbol{\theta}) = \sum_{i=1}^{n} \ell_i(\boldsymbol{\theta}) = \sum_{i=1}^{n} \{\ell_{1i}(\boldsymbol{\theta}) + \ell_{2i}(\boldsymbol{\theta}) + \ell_{3i}(\boldsymbol{\theta})\}, \tag{4.8}$$

where $\ell_i(\boldsymbol{\theta})$ is the log-likelihood contribution from individual i, and

$$\begin{aligned}
\ell_{1i}(\boldsymbol{\theta}) &= R_i \log\{P(Y_i = y_i, R_i = 1 | \mathbf{X}_{1i}, \mathbf{X}_{2i}, \mathbf{X}_{3i})\} \\
&= R_i \log\left\{\Phi\left(\frac{\mathbf{X}_{2i}^{\tau}\boldsymbol{\gamma} + \rho_{12}\epsilon_{1i}}{\sqrt{1-\rho_{12}^2}}\right)\sigma^{-1}\phi(\epsilon_{1i})\right\}, \\
\ell_{2i}(\boldsymbol{\theta}) &= (1-R_i)D_i \log\{P(Y_i = y_i, R_i = 0, D_i = 1 | \mathbf{X}_{1i}, \mathbf{X}_{2i}, \mathbf{X}_{3i})\} \\
&= (1-R_i)D_i \log\left\{\int_{-\infty}^{-\mathbf{X}_{2i}^{\tau}\boldsymbol{\gamma}} \int_{-\mathbf{X}_{3i}^{\tau}\boldsymbol{\xi}}^{\infty} \sigma^{-1}\phi(\epsilon_{1i})\phi_{23|1}(t,u;\epsilon_{1i})dtdu\right\}, \\
\ell_{3i}(\boldsymbol{\theta}) &= (1-R_i)(1-D_i) \log\{P(R_i = 0, D_i = 0 | \mathbf{X}_{1i}, \mathbf{X}_{2i}, \mathbf{X}_{3i})\} \\
&= (1-R_i)(1-D_i) \log\left\{\int_{-\infty}^{-\mathbf{X}_{2i}^{\tau}\boldsymbol{\gamma}} \int_{-\infty}^{-\mathbf{X}_{3i}^{\tau}\boldsymbol{\xi}} \phi_{23}(t,u)dtdu\right\}.
\end{aligned}$$

In the above presentation, we have used ϵ_{1i} to replace $(y_i - \beta_0 - \mathbf{X}_{1i}^{\tau}\boldsymbol{\beta}_1)/\sigma$ for notational convenience. Importantly, ϵ_{1i} depends on β_0, $\boldsymbol{\beta}_1$, and σ.

With the log-likelihood function $\ell(\boldsymbol{\theta})$ given in (4.8), the maximum likelihood estimator $\widehat{\boldsymbol{\theta}} = \left(\widehat{\beta}_0, \widehat{\boldsymbol{\beta}}_1^{\tau}, \widehat{\boldsymbol{\gamma}}^{\tau}, \widehat{\boldsymbol{\xi}}^{\tau}, \widehat{\sigma}, \widehat{\rho}_{12}, \widehat{\rho}_{13}, \widehat{\rho}_{23}\right)^{\tau}$ of $\boldsymbol{\theta}$ is

$$\widehat{\boldsymbol{\theta}} = \arg\max_{\boldsymbol{\theta}} \ell(\boldsymbol{\theta}). \tag{4.9}$$

Let

$$\boldsymbol{S}_i(\boldsymbol{\theta}) = \frac{\partial \ell_i(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = \frac{\partial \ell_{1i}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} + \frac{\partial \ell_{2i}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} + \frac{\partial \ell_{3i}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}$$

be the score vector contributed by individual i. For continuity, we have relegated the derivation of $\boldsymbol{S}_i(\boldsymbol{\theta})$ to the Supplementary Material. We further define the Fisher information

$$\boldsymbol{J} = E\left\{-\frac{\partial^2 \ell_i(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^{\tau}}\right\} = Var\left\{\boldsymbol{S}_i(\boldsymbol{\theta}_0)\right\},$$

where $\boldsymbol{\theta}_0$ is the true value of $\boldsymbol{\theta}$. From classical maximum likelihood theory (Serfling (1980)), we have that under the conditions in Proposition 1 and Conditions A1–A5 in the Supplementary Material, the maximum likelihood estimator $\widehat{\boldsymbol{\theta}}$ satisfies

$$n^{1/2}(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) \to N\left(0, \boldsymbol{J}^{-1}\right)$$

in distribution as $n \to \infty$. In practice, $\boldsymbol{J}$ can be consistently estimated by

$$\widehat{\boldsymbol{J}} = -n^{-1} \sum_{i=1}^{n} \frac{\partial^2 \ell_i(\widehat{\boldsymbol{\theta}})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^\tau} \quad \text{or} \quad \widehat{\boldsymbol{J}} = n^{-1} \sum_{i=1}^{n} \boldsymbol{S}_i(\widehat{\boldsymbol{\theta}}) \boldsymbol{S}_i^\tau(\widehat{\boldsymbol{\theta}}).$$

4.3. Consistency under misspecification of error distributions

In this subsection, we investigate the effect on the estimation of the regression coefficients $\boldsymbol{\beta}_1$ when the joint distribution of $(\epsilon_{1i}, \epsilon_{2i}, \epsilon_{3i})^\tau$ is misspecified, but the linear regression models for (Y_i, Z_i, U_i) are correct. We show that the maximum likelihood estimator $\widehat{\boldsymbol{\beta}}_1$ of $\boldsymbol{\beta}_1$ is consistent under the condition that $\mathbf{X}_{1i}$ is independent of $\mathbf{X}_{2i}$ and $\mathbf{X}_{3i}$, even when the joint distribution is misspecified.

Suppose the true model for (Y_i, Z_i, U_i) is

$$Y_i = \beta_{T0} + \mathbf{X}_{1i}^\tau \boldsymbol{\beta}_{T1} + \tau w_{1i}, \quad Z_i = \mathbf{X}_{2i}^\tau \boldsymbol{\gamma}_T + w_{2i}, \quad U_i = \mathbf{X}_{3i}^\tau \boldsymbol{\xi}_T + w_{3i}, \qquad (4.10)$$

where the joint cumulative distribution function of $(w_{1i}, w_{2i}, w_{3i})^\tau$ is $H(s, t, u)$.

Instead of using the true model, we consider a working model for (Y_i, Z_i, U_i) as specified in (3.1), (3.2), and (4.1). From the results of White (1982), under Conditions B1–B4 in the Supplementary Material, the maximum likelihood estimator $\widehat{\boldsymbol{\theta}}$, obtained from the working model and defined in (4.9), converges to a unique limit $\boldsymbol{\theta}^* = (\beta_0^*, \boldsymbol{\beta}_1^{*\tau}, \boldsymbol{\gamma}^{*\tau}, \boldsymbol{\xi}^{*\tau}, \sigma^*, \rho_{12}^*, \rho_{13}^*, \rho_{23}^*)^\tau$. Here $\boldsymbol{\theta}^*$ is the unique solution to the equations

$$E_T\left\{\frac{\partial \ell_i(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\right\} = 0,$$

where E_T indicates that the expectation is with respect to the true distribution (4.10) of (Y_i, Z_i, U_i) and $(\mathbf{X}_{i1}, \mathbf{X}_{i2}, \mathbf{X}_{i3})$. In the following, we argue that $\boldsymbol{\beta}_1^* = \boldsymbol{\beta}_{T1}$ when $\mathbf{X}_{1i}$ is independent of $\mathbf{X}_{2i}$ and $\mathbf{X}_{3i}$. We follow the procedures in He and Lawless (2005). The key step in the argument is that when $\boldsymbol{\beta}_1 = \boldsymbol{\beta}_{T1}$,

$$E_T\left\{\frac{\partial \ell_i(\boldsymbol{\theta})}{\partial \boldsymbol{\beta}_1}\right\} = 0, \qquad (4.11)$$

no matter what the values of the other parameters. Without loss of generality, we assume that $\mathbf{X}_{1i}$ has mean $\mathbf{0}$ and that all the expectations below with respect to $(\mathbf{X}_{1i}, \mathbf{X}_{2i}, \mathbf{X}_{3i})$ exist.

Note that

$$E_T\left\{\frac{\partial \ell_i(\boldsymbol{\theta})}{\partial \boldsymbol{\beta}_1}\right\} = E_T\left\{\frac{\partial \ell_{1i}(\boldsymbol{\theta})}{\partial \boldsymbol{\beta}_1} + \frac{\partial \ell_{2i}(\boldsymbol{\theta})}{\partial \boldsymbol{\beta}_1}\right\} = -\frac{1}{\sigma}E_T\left[\mathbf{X}_{1i}\left\{\frac{\partial \ell_{1i}(\boldsymbol{\theta})}{\partial \epsilon_{1i}} + \frac{\partial \ell_{2i}(\boldsymbol{\theta})}{\partial \epsilon_{1i}}\right\}\right].$$

From the true and working models for Y_i, we have

$$Y_i = \beta_{T0} + \mathbf{X}_{1i}^\tau \boldsymbol{\beta}_{T1} + \tau w_{1i} = \beta_0 + \mathbf{X}_{1i}^\tau \boldsymbol{\beta}_1 + \sigma\epsilon_{1i},$$

which implies that

$$\epsilon_{1i} = \frac{1}{\sigma}\{\beta_{T0} - \beta_0 + \mathbf{X}_{1i}^\tau(\boldsymbol{\beta}_{T1} - \boldsymbol{\beta}_1) + \tau w_{1i}\}. \tag{4.12}$$

When $\boldsymbol{\beta}_1 = \boldsymbol{\beta}_{T1}$, (4.12) becomes $\epsilon_{1i} = (\beta_{T0} - \beta_0 + \tau w_{1i})/\sigma$, which does not depend on $\mathbf{X}_{1i}$. By the law of total expectation,

$$\begin{aligned} &E_T\left\{\frac{\partial \ell_i(\boldsymbol{\theta})}{\partial \boldsymbol{\beta}_1}\right\} \\ &= -\frac{1}{\sigma}E_{(\mathbf{X}_{1i},\mathbf{X}_{2i},\mathbf{X}_{3i})}\left[\mathbf{X}_{1i}E_{(Y_i,R_i,D_i)|(\mathbf{X}_{1i},\mathbf{X}_{2i},\mathbf{X}_{3i})}\left\{\frac{\partial \ell_{1i}(\boldsymbol{\theta})}{\partial \epsilon_{1i}} + \frac{\partial \ell_{2i}(\boldsymbol{\theta})}{\partial \epsilon_{1i}}\right\}\right]. \end{aligned} \tag{4.13}$$

Next we argue that when $\boldsymbol{\beta}_1 = \boldsymbol{\beta}_{T1}$, $E_{(Y_i,R_i,D_i)|(\mathbf{X}_{1i},\mathbf{X}_{2i},\mathbf{X}_{3i})}(\partial \ell_{1i}(\boldsymbol{\theta})/\partial\epsilon_{1i} + \partial \ell_{2i}(\boldsymbol{\theta})/\partial\epsilon_{1i})$ depends on $\mathbf{X}_{2i}$ and $\mathbf{X}_{3i}$, but not on $\mathbf{X}_{1i}$. This claim, together with $E(\mathbf{X}_{1i}) = \mathbf{0}$, (4.13), and the condition that $\mathbf{X}_{1i}$ is independent of $\mathbf{X}_{2i}$ and $\mathbf{X}_{3i}$, implies (4.11).

Let

$$\kappa(X_{2i}, \epsilon_{1i}; \boldsymbol{\theta}) = \frac{\partial \log\left\{\Phi\left(\mathbf{X}_{2i}^\tau\boldsymbol{\gamma} + \rho_{12}\epsilon_{1i}/\sqrt{1-\rho_{12}^2}\right)\sigma^{-1}\phi(\epsilon_{1i})\right\}}{\partial \epsilon_{1i}},$$

which does not depend on $\mathbf{X}_{1i}$. Then

$$E_{(Y_i,R_i,D_i)|(\mathbf{X}_{1i},\mathbf{X}_{2i},\mathbf{X}_{3i})}\left(\frac{\partial \ell_{1i}(\boldsymbol{\theta})}{\partial \epsilon_{1i}}\right) = E_{(Y_i,R_i)|(\mathbf{X}_{1i},\mathbf{X}_{2i},\mathbf{X}_{3i})}\left\{R_i\kappa(X_{2i}, \epsilon_{1i}; \boldsymbol{\theta})\right\}.$$

Let $H_{2|1}(t|s)$ be the conditional cumulative distribution function of w_{2i} given $w_{1i} = s$ and $\bar{H}_{2|1}(t|s) = 1 - H_{2|1}(t|s)$. By the law of total expectation, it can be verified that

$$\begin{aligned} &E_{(Y_i,R_i)|(\mathbf{X}_{1i},\mathbf{X}_{2i},\mathbf{X}_{3i})}\{R_i\kappa(X_{2i}, \epsilon_{1i}; \boldsymbol{\theta})\} \\ &= E_{\epsilon_{1i}|(\mathbf{X}_{1i},\mathbf{X}_{2i},\mathbf{X}_{3i})}\left\{\bar{H}_{2|1}(-\mathbf{X}_{2i}^\tau\boldsymbol{\gamma}|w_{1i})\kappa(\mathbf{X}_{2i}, \epsilon_{1i}; \boldsymbol{\theta})\right\}, \end{aligned}$$

where $w_{1i} = \{\beta - \beta_{T0} + \sigma\epsilon_{1i}\}/\tau$. Since $\bar{H}_{2|1}(-\mathbf{X}_{2i}^\tau\boldsymbol{\gamma}|w_{1i})\kappa(\mathbf{X}_{2i}, \epsilon_{1i}; \boldsymbol{\theta})$ depends only on $\mathbf{X}_{2i}$ and ϵ_{1i}, and $\mathbf{X}_{1i}$ is independent of $\mathbf{X}_{2i}$, $\mathbf{X}_{3i}$, and ϵ_{1i}, we have

$$E_{(Y_i,R_i)|(\mathbf{X}_{1i},\mathbf{X}_{2i},\mathbf{X}_{3i})}\{R_i\kappa(X_{2i}, \epsilon_{1i}; \boldsymbol{\theta})\} = E_{\epsilon_{1i}|\mathbf{X}_{2i}}\left\{\bar{H}_{2|1}(-\mathbf{X}_{2i}^\tau\boldsymbol{\gamma}|w_{1i})\kappa(\mathbf{X}_{2i}, \epsilon_{1i}; \boldsymbol{\theta})\right\},$$

which is a function of $\mathbf{X}_{2i}$ only. Hence, $E_{(Y_i,R_i,D_i)|(\mathbf{X}_{1i},\mathbf{X}_{2i},\mathbf{X}_{3i})}\{\partial \ell_{1i}(\boldsymbol{\theta})/\partial\epsilon_{1i}\}$

is a function of $\mathbf{X}_{2i}$ only. Similarly, $E_{(Y_i,R_i,D_i)\big|\left(\mathbf{X}_{1i},\mathbf{X}_{2i},\mathbf{X}_{3i}\right)}\{\partial\ell_{2i}(\boldsymbol{\theta})/\partial\epsilon_{1i}\}$ is a function of $\mathbf{X}_{2i}$ and $\mathbf{X}_{3i}$ only. This completes the proof of (4.11).

From (4.11), $\boldsymbol{\beta}_1 = \boldsymbol{\beta}_{T1}$ is a solution to $E_T\{\partial\ell_i(\boldsymbol{\theta})/\partial\boldsymbol{\beta}_1\} = 0$, no matter the values of the other parameters. Thus, $\boldsymbol{\beta}_1 = \boldsymbol{\beta}_{T1}$ is in the solution of $E_T\{\partial\ell_i(\boldsymbol{\theta})/\partial\boldsymbol{\theta}\} = 0$. By the uniqueness of the solution of $\boldsymbol{\theta}^*$ (White (1982)), we conclude that $\boldsymbol{\beta}_1^* = \boldsymbol{\beta}_{T1}$, which establishes the consistency of $\widehat{\boldsymbol{\beta}}_1$.

This result suggests that the estimator of the regression coefficient $\boldsymbol{\beta}_1$ is robust to the misspecification of the joint distribution of the outcome and latent variables if $\mathbf{X}_{1i}$ is independent of $\mathbf{X}_{2i}$ and $\mathbf{X}_{3i}$. If the dependence between $\mathbf{X}_{1i}$ and $(\mathbf{X}_{2i}, \mathbf{X}_{3i})$ is not too strong, our method does not provide substantially biased results, as we will see in Section 5. For more discussion, see He and Lawless (2005).

We now derive the asymptotic distribution of $\widehat{\boldsymbol{\beta}}$. From the results of White (1982), under Conditions B1–B6 in the Supplementary Material, we have

$$n^{1/2}(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}^*) \to N\big(0, \Gamma_1^{-1}\Gamma_2\Gamma_1^{-1}\big),$$

where

$$\Gamma_1 = E_T\left\{-\frac{\partial^2\ell_i(\boldsymbol{\theta}^*)}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^\tau}\right\} \quad \text{and} \quad \Gamma_2 = Var_T\left\{\boldsymbol{S}_i(\boldsymbol{\theta}^*)\right\}.$$

In practice, we can estimate the covariance matrix $\Gamma_1^{-1}\Gamma_2\Gamma_1^{-1}$ by $\widehat{\Gamma}_1^{-1}\widehat{\Gamma}_2\widehat{\Gamma}_1^{-1}$, where

$$\widehat{\Gamma}_1 = -n^{-1}\sum_{i=1}^n \frac{\partial^2\ell_i(\widehat{\boldsymbol{\theta}})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^\tau} \quad \text{and} \quad \widehat{\Gamma}_2 = n^{-1}\sum_{i=1}^n \boldsymbol{S}_i(\widehat{\boldsymbol{\theta}})\boldsymbol{S}_i^\tau(\widehat{\boldsymbol{\theta}}).$$

The estimated covariance matrix of $\widehat{\boldsymbol{\beta}}$ is the upper $p \times p$ sub-matrix of $\widehat{\Gamma}_1^{-1}\widehat{\Gamma}_2\widehat{\Gamma}_1^{-1}$. Alternatively, the standard errors of $\widehat{\boldsymbol{\beta}}$ can be calculated by using the nonparametric bootstrapping method (Efron (1979)).

5. Simulation Studies

We performed simulation studies to compare five methods for estimating $\boldsymbol{\beta}$ in the response model (3.1): our estimator $\widehat{\boldsymbol{\beta}}$ which uses the information from the response, missing-data, and call-back models (3.1), (3.2), and (4.1), and is referred to as the "proposed" method; the estimate $\widetilde{\boldsymbol{\beta}}$, which uses the information from the response and missing-data models (3.1) and (3.2), referred to as the "Heckman-1" method, that uses only the information for R and not the information for D; the estimate $\widetilde{\boldsymbol{\beta}}$, which uses the information from the response and missing-data models (3.1) and (3.2), the "Heckman-2" method that combines the information for R and D and creates a new missing indicator $K = 1$ if $R = 1$ or

$D = 1$ and 0 otherwise, and we apply K in (3.2) instead of R; the ordinary least square estimate of $\boldsymbol{\beta}$ based on model (3.1) and the complete-case data, referred to as the "cc-OLS-1" method; the ordinary least square estimate of $\boldsymbol{\beta}$ based on model (3.1) and the complete-case data, referred to as the "cc-OLS-2" method.

In the simulations, we posited the covariates

$$(X_{1i}, X_{2i}, X_{3i})^\tau \sim N\left(\begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \rho_{x12} & \rho_{x13} \\ \rho_{x12} & 1 & 0.4 \\ \rho_{x13} & 0.4 & 1 \end{pmatrix}\right).$$

We considered five scenarios with correctly and incorrectly specified models to evaluate the robustness of our method.

Scenario I: correctly specified model. For $i = 1, \ldots, n$, (Y_i, Z_i, U_i) was generated from the models

$$Y_i = \beta_0 + \beta_1 X_{1i} + \epsilon_{1i},\ Z_i = \gamma_0 + \gamma_1 X_{2i} + \epsilon_{2i},\ U_i = \xi_0 + \xi_1 X_{3i} + \epsilon_{3i}. \quad (5.1)$$

Further, we took the $(\epsilon_{1i}, \epsilon_{2i}, \epsilon_{3i})^\tau \sim N(\mathbf{0}, \boldsymbol{\Sigma})$, where

$$\boldsymbol{\Sigma} = \begin{pmatrix} 1 & \rho_{12} & \rho_{13} \\ \rho_{12} & 1 & \rho_{23} \\ \rho_{13} & \rho_{23} & 1 \end{pmatrix}. \quad (5.2)$$

In this scenario, $\rho_{x12} = 0.5$, and $\rho_{x13} = 0.3$.

To study the robustness of our method, we considered four scenarios in which the distribution of the error terms is misspecified:

Scenario II: misspecified distribution for $(\epsilon_{1i}, \epsilon_{2i}, \epsilon_{3i})^\tau$ with $\mathbf{X}_{1i}$ being independent of $(\mathbf{X}_{2i}, \mathbf{X}_{3i})$. For $i = 1, \ldots, n$, (Y_i, Z_i, U_i) was generated from (5.1) with $(\epsilon_{1i}, \epsilon_{2i}, \epsilon_{3i})^\tau$ following a multivariate t-distribution with mean $\mathbf{0}$, the covariance matrix $\boldsymbol{\Sigma}$ in (5.2), and 3 degrees of freedom. Further, we posited that $\mathbf{X}_{1i}$ is independent of $(\mathbf{X}_{2i}, \mathbf{X}_{3i})$.

Scenario III: misspecified distribution with $\mathbf{X}_{1i}$ being dependent on $(\mathbf{X}_{2i}, \mathbf{X}_{3i})$, with $\rho_{x12} = 0.5$ and $\rho_{x13} = 0.3$. For $i = 1, \ldots, n$, (Y_i, Z_i, U_i) was generated from (5.1) with $(\epsilon_{1i}, \epsilon_{2i}, \epsilon_{3i})^\tau$ following a multivariate t-distribution with mean $\mathbf{0}$, covariance matrix $\boldsymbol{\Sigma}$ in (5.2), and 3 degrees of freedom.

Scenario IV: misspecified distribution with $\mathbf{X}_{1i}$ dependent on $(\mathbf{X}_{2i}, \mathbf{X}_{3i})$, assuming $\mathbf{X}_{1i} = \mathbf{X}_{2i}$ ($\mathbf{X}_{1i}$ overlaps with $(\mathbf{X}_{2i}, \mathbf{X}_{3i})$). For $i = 1, \ldots, n$, (Y_i, Z_i, U_i) was generated from

$$Y_i = \beta_0 + \beta_1 X_{1i} + \epsilon_{1i}, \quad Z_i = \gamma_0 + \gamma_1 X_{1i} + \epsilon_{2i}, \quad U_i = \xi_0 + \xi_1 X_{3i} + \epsilon_{3i},$$

with $(\epsilon_{1i}, \epsilon_{2i}, \epsilon_{3i})^\tau$ following a multivariate t-distribution with mean $\mathbf{0}$, covariance matrix $\mathbf{\Sigma}$ in (5.2), and 3 degrees of freedom. Here we took $\rho_{x12} = 0.5$ and $\rho_{x13} = 0.3$.

Scenario V: misspecified distribution with $\mathbf{X}_{1i}$ dependent on $(\mathbf{X}_{2i}, \mathbf{X}_{3i})$, assuming $\mathbf{X}_{1i} = \mathbf{X}_{2i} = \mathbf{X}_{3i}$ ($\mathbf{X}_{1i}$ overlaps with $(\mathbf{X}_{2i}, \mathbf{X}_{3i})$ completely). For $i = 1, \ldots, n$, (Y_i, Z_i, U_i) was generated from

$$Y_i = \beta_0 + \beta_1 X_{1i} + \epsilon_{1i}, \quad Z_i = \gamma_0 + \gamma_1 X_{1i} + \epsilon_{2i}, \quad U_i = \xi_0 + \xi_1 X_{1i} + \epsilon_{3i},$$

with $(\epsilon_{1i}, \epsilon_{2i}, \epsilon_{3i})^\tau$ following a multivariate t-distribution with mean $\mathbf{0}$, the covariance matrix $\mathbf{\Sigma}$ in (5.2), and 3 degrees of freedom.

For each scenario, the missing indicator was determined by $R_i = I(Z_i > 0)$, and the call-back indicator is determined by $D_i = I(U_i > 0)$, and the true values were $\beta_0 = \beta_1 = 1$, $\gamma_1 = 1$, $\xi_1 = 1$. We set $\gamma_0 = \xi_0$ and adjusted the values for different missing proportions. For example, in Scenario I, when $\gamma_0 = \xi_0 = 0$, the response proportion (probability $R_i = 1$) is about 50%, and the call-back success rate (probability $D_i = 1$) is about 15%; when $\gamma_0 = \xi_0 = 1$, the response proportion is about 80%, and the call-back success rate is about 12%. We set $\rho_{12} = \rho_{13} = \rho_{23} = \rho$ in (5.2) and adjusted the value of ρ for the degree of nonignorability.

For each scenario, we considered sample sizes 100 and 200, values for ξ_0 of 0 and 1, and values for ρ of 0.8 and 0.5. Hence, we had 8 combinations of sample size, values of ξ_0, and value of ρ in each scenario. For each combination, we calculated the bias, standard deviation (SD), and mean square error (MSE) for each of five estimates of (β_0, β_1) based on 2,000 repetitions.

The results for Scenario I are summarized in Table 1. In Scenario I the model for $(\epsilon_{1i}, \epsilon_{2i}, \epsilon_{3i})^\tau$ is correctly specified for all five methods. The proposed and two Heckman methods yield consistent estimators, but two cc-OLS methods yield biased estimators. Our estimate is more efficient than both Heckman estimates, and the Heckman-2 estimate is more efficient than the Heckman-1 estimate. As the missing proportion increases, the efficiency gain of our method increases.

The results for Scenario II are summarized in Table 2. In Scenario II the model for $(\epsilon_{1i}, \epsilon_{2i}, \epsilon_{3i})^\tau$ is misspecified for all five methods, and $\mathbf{X}_{1i}$ is independent of $(\mathbf{X}_{2i}, \mathbf{X}_{3i})$: $\mathbf{X}_{1i} = (X_{1i})$, $\mathbf{X}_{2i} = (1, X_{2i})^\tau$, and $\mathbf{X}_{3i} = (1, X_{3i})^\tau$. In this scenario, all methods yield small biases for β_1, but our method yields the smallest MSE in all combinations.

The results for Scenario III are summarized in Table 3. In it the model for $(\epsilon_{1i}, \epsilon_{2i}, \epsilon_{3i})^\tau$ is misspecified for all five methods, and $\mathbf{X}_{1i}$ is dependent on

Table 1. Bias, standard deviation, and mean square error for five estimates of (β_0, β_1) in Scenario I, in which the model for $(\epsilon_{1i}, \epsilon_{2i}, \epsilon_{3i})^\tau$ is correctly specified for all five methods.

				β_0			β_1		
n	γ_0	ρ	Methods	Bias	SD	MSE	Bias	SD	MSE
100	0	0.8	Proposed	0.009	0.144	0.021	0.001	0.110	0.012
			Heckman-1	−0.006	0.231	0.053	−0.004	0.136	0.018
			cc-OLS-1	0.490	0.128	0.257	−0.141	0.132	0.037
			cc-OLS-2	0.367	0.110	0.147	−0.109	0.115	0.025
			Heckman-2	−0.011	0.207	0.043	−0.001	0.127	0.016
100	0	0.5	Proposed	0.005	0.168	0.028	0.006	0.126	0.016
			Heckman-1	0.009	0.286	0.082	−0.012	0.164	0.027
			cc-OLS-1	0.310	0.141	0.116	−0.095	0.146	0.030
			cc-OLS-2	0.224	0.119	0.064	−0.073	0.121	0.020
			Heckman-2	0.011	0.222	0.049	−0.010	0.135	0.018
100	1	0.8	Proposed	−0.008	0.110	0.012	0.008	0.103	0.011
			Heckman-1	0.003	0.153	0.023	−0.002	0.130	0.017
			cc-OLS-1	0.246	0.103	0.071	−0.096	0.115	0.023
			cc-OLS-2	0.151	0.099	0.033	−0.063	0.108	0.016
			Heckman-2	0.001	0.146	0.021	−0.004	0.120	0.014
100	1	0.5	Proposed	0.004	0.119	0.014	−0.001	0.109	0.012
			Heckman-1	−0.001	0.174	0.030	0.000	0.133	0.018
			cc-OLS-1	0.149	0.110	0.034	−0.059	0.118	0.017
			cc-OLS-2	0.081	0.098	0.016	−0.037	0.107	0.013
			Heckman-2	−0.006	0.153	0.023	−0.001	0.124	0.015
200	0	0.8	Proposed	−0.005	0.102	0.010	0.002	0.074	0.005
			Heckman-1	0.010	0.148	0.022	0.000	0.094	0.009
			cc-OLS-1	0.494	0.093	0.252	−0.136	0.096	0.028
			cc-OLS-2	0.367	0.080	0.141	−0.103	0.082	0.017
			Heckman-2	0.014	0.131	0.017	0.000	0.088	0.008
200	0	0.5	Proposed	0.007	0.118	0.014	0.003	0.089	0.008
			Heckman-1	−0.002	0.187	0.035	0.001	0.109	0.012
			cc-OLS-1	0.310	0.101	0.107	−0.087	0.101	0.018
			cc-OLS-2	0.220	0.085	0.055	−0.065	0.086	0.012
			Heckman-2	0.004	0.142	0.020	0.001	0.092	0.008
200	1	0.8	Proposed	−0.002	0.083	0.007	0.000	0.068	0.005
			Heckman-1	0.006	0.101	0.010	0.000	0.079	0.006
			cc-OLS-1	0.243	0.075	0.064	−0.091	0.078	0.014
			cc-OLS-2	0.148	0.069	0.027	−0.058	0.072	0.009
			Heckman-2	0.000	0.098	0.010	0.002	0.080	0.006
200	1	0.5	Proposed	−0.002	0.076	0.006	−0.003	0.074	0.006
			Heckman-1	0.007	0.117	0.014	−0.004	0.089	0.008
			cc-OLS-1	0.155	0.079	0.030	−0.062	0.083	0.011
			cc-OLS-2	0.084	0.072	0.012	−0.038	0.075	0.007
			Heckman-2	0.008	0.102	0.011	−0.003	0.082	0.007

Table 2. Bias, standard deviation, and mean square error for five estimates of (β_0, β_1) in Scenario II, in which the model for $(\epsilon_{1i}, \epsilon_{2i}, \epsilon_{3i})^\tau$ is misspecified for all five methods, and $\mathbf{X}_{1i}$ is independent of $(\mathbf{X}_{2i}, \mathbf{X}_{3i})$.

				β_0			β_1		
n	γ_0	ρ	Methods	Bias	SD	MSE	Bias	SD	MSE
100	0	0.8	Proposed	−0.081	0.253	0.071	0.002	0.153	0.023
			Heckman-1	−0.177	0.428	0.214	−0.010	0.202	0.041
			cc-OLS-1	0.722	0.226	0.572	−0.012	0.241	0.058
			cc-OLS-2	0.565	0.192	0.356	−0.007	0.202	0.041
			Heckman-2	−0.087	0.343	0.125	−0.007	0.183	0.034
100	0	0.5	Proposed	−0.044	0.295	0.089	−0.004	0.196	0.038
			Heckman-1	−0.105	0.575	0.341	0.007	0.240	0.058
			cc-OLS-1	0.457	0.226	0.260	0.009	0.245	0.060
			cc-OLS-2	0.335	0.191	0.149	0.005	0.205	0.042
			Heckman-2	−0.030	0.399	0.160	0.005	0.205	0.042
100	1	0.8	Proposed	0.012	0.164	0.027	0.002	0.143	0.020
			Heckman-1	−0.027	0.261	0.069	−0.004	0.164	0.027
			cc-OLS-1	0.422	0.164	0.205	−0.002	0.175	0.031
			cc-OLS-2	0.305	0.153	0.117	−0.002	0.159	0.025
			Heckman-2	0.034	0.242	0.060	−0.004	0.159	0.025
100	1	0.5	Proposed	0.008	0.185	0.034	−0.001	0.165	0.027
			Heckman-1	−0.021	0.326	0.106	−0.005	0.199	0.039
			cc-OLS-1	0.262	0.184	0.103	−0.007	0.174	0.030
			cc-OLS-2	0.172	0.170	0.058	−0.006	0.163	0.027
			Heckman-2	0.025	0.297	0.089	−0.008	0.197	0.039
200	0	0.8	Proposed	−0.103	0.291	0.096	−0.004	0.118	0.014
			Heckman-1	−0.203	0.320	0.144	0.002	0.135	0.018
			cc-OLS-1	0.720	0.152	0.542	−0.001	0.158	0.025
			cc-OLS-2	0.565	0.131	0.336	0.000	0.135	0.018
			Heckman-2	−0.094	0.233	0.063	0.000	0.125	0.016
200	0	0.5	Proposed	−0.076	0.261	0.074	0.007	0.139	0.019
			Heckman-1	−0.130	0.405	0.181	−0.002	0.167	0.028
			cc-OLS-1	0.443	0.164	0.224	−0.001	0.173	0.030
			cc-OLS-2	0.327	0.140	0.127	−0.001	0.142	0.020
			Heckman-2	−0.040	0.289	0.085	0.000	0.140	0.020
200	1	0.8	Proposed	0.002	0.126	0.016	0.005	0.102	0.011
			Heckman-1	−0.055	0.174	0.033	−0.001	0.111	0.012
			cc-OLS-1	0.421	0.126	0.193	0.001	0.122	0.015
			cc-OLS-2	0.303	0.115	0.105	0.001	0.113	0.013
			Heckman-2	0.012	0.165	0.027	0.000	0.111	0.012
200	1	0.5	Proposed	0.005	0.146	0.021	0.000	0.115	0.013
			Heckman-1	−0.046	0.227	0.054	0.000	0.130	0.017
			cc-OLS-1	0.265	0.131	0.087	−0.001	0.135	0.018
			cc-OLS-2	0.174	0.122	0.045	−0.001	0.123	0.015
			Heckman-2	0.024	0.171	0.030	−0.004	0.130	0.017

Table 3. Bias, standard deviation, and mean square error for five estimates of (β_0, β_1) in Scenario III, in which the model for $(\epsilon_{1i}, \epsilon_{2i}, \epsilon_{3i})^\tau$ is misspecified for all five methods, and $\mathbf{X}_{1i}$ is dependent on $(\mathbf{X}_{2i}, \mathbf{X}_{3i})$.

				β_0			β_1		
n	γ_0	ρ	Methods	Bias	SD	MSE	Bias	SD	MSE
100	0	0.8	Proposed	−0.079	0.257	0.072	0.023	0.175	0.031
			Heckman-1	−0.210	0.754	0.612	0.042	0.308	0.096
			cc-OLS-1	0.755	0.233	0.625	−0.191	0.239	0.094
			cc-OLS-2	0.573	0.193	0.366	−0.134	0.201	0.058
			Heckman-2	−0.067	0.381	0.150	0.020	0.220	0.049
100	0	0.5	Proposed	−0.052	0.349	0.125	0.004	0.215	0.046
			Heckman-1	−0.093	0.578	0.343	0.031	0.274	0.076
			cc-OLS-1	0.470	0.261	0.289	−0.106	0.255	0.076
			cc-OLS-2	0.340	0.211	0.160	−0.070	0.210	0.049
			Heckman-2	−0.018	0.409	0.167	0.029	0.244	0.060
100	1	0.8	Proposed	0.011	0.155	0.024	0.008	0.149	0.022
			Heckman-1	−0.015	0.293	0.086	0.012	0.197	0.039
			cc-OLS-1	0.432	0.180	0.219	−0.123	0.180	0.047
			cc-OLS-2	0.303	0.164	0.119	−0.082	0.167	0.035
			Heckman-2	0.038	0.276	0.078	−0.001	0.195	0.038
100	1	0.5	Proposed	0.004	0.211	0.044	0.007	0.166	0.028
			Heckman-1	0.023	0.343	0.118	−0.004	0.225	0.051
			cc-OLS-1	0.277	0.190	0.113	−0.082	0.189	0.042
			cc-OLS-2	0.176	0.171	0.060	−0.056	0.169	0.032
			Heckman-2	0.042	0.283	0.082	−0.009	0.201	0.040
200	0	0.8	Proposed	−0.098	0.192	0.047	0.029	0.130	0.018
			Heckman-1	−0.225	0.379	0.194	0.049	0.159	0.028
			cc-OLS-1	0.761	0.171	0.609	−0.187	0.175	0.066
			cc-OLS-2	0.577	0.138	0.352	−0.130	0.146	0.038
			Heckman-2	−0.109	0.259	0.079	0.033	0.139	0.020
200	0	0.5	Proposed	−0.087	0.263	0.077	0.033	0.149	0.023
			Heckman-1	−0.140	0.495	0.264	0.022	0.191	0.037
			cc-OLS-1	0.476	0.181	0.259	−0.129	0.175	0.047
			cc-OLS-2	0.342	0.143	0.138	−0.096	0.152	0.032
			Heckman-2	−0.037	0.337	0.115	0.008	0.168	0.028
200	1	0.8	Proposed	−0.007	0.115	0.013	0.012	0.112	0.013
			Heckman-1	−0.051	0.186	0.037	0.032	0.127	0.017
			cc-OLS-1	0.439	0.123	0.208	−0.115	0.134	0.031
			cc-OLS-2	0.310	0.114	0.109	−0.073	0.123	0.020
			Heckman-2	0.019	0.181	0.033	0.015	0.126	0.016
200	1	0.5	Proposed	0.000	0.151	0.023	0.002	0.130	0.017
			Heckman-1	−0.052	0.263	0.072	0.018	0.152	0.024
			cc-OLS-1	0.270	0.131	0.090	−0.076	0.137	0.025
			cc-OLS-2	0.172	0.119	0.044	−0.051	0.123	0.018
			Heckman-2	0.032	0.178	0.033	−0.006	0.144	0.021

$(\mathbf{X}_{2i}, \mathbf{X}_{3i})$: $\mathbf{X}_{1i} = (X_{1i})$, $\mathbf{X}_{2i} = (1, X_{2i})^\tau$, $\mathbf{X}_{3i} = (1, X_{3i})^\tau$, $Corr(X_{1i}, X_{2i}) = 0.5$, and $Corr(X_{1i}, X_{3i}) = 0.3$. Our method and two Heckman methods yield smaller biases than two cc-OLS methods, and our method gives the smallest MSE for β_1 in all combinations.

The results for Scenario IV are summarized in Table 4. Here the model for $(\epsilon_{1i}, \epsilon_{2i}, \epsilon_{3i})^\tau$ is misspecified for all five methods, and $\mathbf{X}_{1i}$ is dependent on and overlaps with $(\mathbf{X}_{2i}, \mathbf{X}_{3i})$: $\mathbf{X}_{1i} = (X_{1i})$, $\mathbf{X}_{2i} = (1, X_{1i})^\tau$, and $\mathbf{X}_{3i} = (1, X_{3i})^\tau$. Our method still yields the smallest MSE for β_1 in all combinations.

The results for Scenario V are summarized in Table 5. Here the model for $(\epsilon_{1i}, \epsilon_{2i}, \epsilon_{3i})^\tau$ is misspecified for all five methods, and $\mathbf{X}_{1i}$ is dependent on and overlaps with $(\mathbf{X}_{2i}, \mathbf{X}_{3i})$ completely: $\mathbf{X}_{1i} = (X_{1i})$, $\mathbf{X}_{2i} = (1, X_{1i})^\tau$, and $\mathbf{X}_{3i} = (1, X_{1i})^\tau$. Our method and two Heckman methods yields smaller biases than two cc-OLS methods, and our method still produces the smallest MSE for β_1 in all combinations.

In summary, our method can reduce the bias caused by a nonignorably missing mechanism and yield more efficient estimates than the Heckman model. Although our method is built under the normal distribution, the estimate of $\boldsymbol{\beta}_1$ is robust to the misspecification of the distribution, even when the condition that $\mathbf{X}_{1i}$ and $(\mathbf{X}_{2i}, \mathbf{X}_{3i})$ are independent does not hold.

6. Application to NHIS Data

We applied our method to the NHIS data. We conducted a two-step preliminary analysis to select the important covariates in models (3.1), (3.2), and (4.1). In the first step, we fit the Heckman models (3.1) and (3.2) with all four covariates (FIN, FMAL, FHI, RAT_CAT2) in each model using the R function *selection* in the R package *sampleSelection* (Toomet and Henningsen (2008)). In (3.1) and (3.2), the covariates with p-values smaller than 0.1 were kept for further analysis. In the second step, we fit a probit model on all four covariates with D_i treated as the response variable. The covariates with p-values smaller than 0.1 were kept for further analysis in the call-back model. After the preliminary analysis, we included FHI cost and FIN for the response model (3.1), we include FMAL and RAT_CAT2 for the missing-data model (3.2), and we included FMAL and FHI cost for the call-back model.

Next, we fit the regression, missing-data, and call-back models (3.1), (3.2), and (4.1) with the selected covariates using the proposed method. We considered the Heckman-1, cc-OLS-1, Heckman-2 and cc-OLS-2 results for comparison. Ta-

Table 4. Bias, standard deviation, and mean square error for five estimates of (β_0, β_1) in Scenario IV, in which the model for $(\epsilon_{1i}, \epsilon_{2i}, \epsilon_{3i})^\tau$ is misspecified for all five methods, and $\mathbf{X}_{1i}$ is dependent on and overlaps with $(\mathbf{X}_{2i}, \mathbf{X}_{3i})$ partially.

				β_0			β_1		
n	γ_0	ρ	Methods	Bias	SD	MSE	Bias	SD	MSE
100	0	0.8	Proposed	−0.102	0.284	0.091	0.078	0.234	0.061
			Heckman-1	−0.001	1.389	1.929	0.048	0.772	0.598
			cc-OLS-1	0.971	0.283	1.023	−0.500	0.293	0.336
			cc-OLS-2	0.608	0.192	0.406	−0.236	0.207	0.099
			Heckman-2	−0.044	0.557	0.312	0.041	0.307	0.096
100	0	0.5	Proposed	−0.076	0.319	0.108	0.058	0.261	0.071
			Heckman-1	0.006	1.724	2.973	0.040	0.954	0.912
			cc-OLS-1	0.603	0.292	0.449	−0.291	0.296	0.172
			cc-OLS-2	0.358	0.213	0.174	−0.134	0.241	0.076
			Heckman-2	−0.004	0.581	0.338	0.032	0.353	0.125
100	1	0.8	Proposed	−0.018	0.207	0.043	0.061	0.204	0.045
			Heckman-1	0.015	0.575	0.331	0.036	0.392	0.155
			cc-OLS-1	0.479	0.188	0.265	−0.272	0.197	0.113
			cc-OLS-2	0.309	0.164	0.122	−0.129	0.174	0.047
			Heckman-2	0.090	0.359	0.137	−0.018	0.249	0.062
100	1	0.5	Proposed	−0.024	0.216	0.047	0.042	0.226	0.053
			Heckman-1	0.043	0.704	0.498	0.000	0.486	0.236
			cc-OLS-1	0.310	0.177	0.127	−0.177	0.197	0.070
			cc-OLS-2	0.175	0.174	0.061	−0.074	0.218	0.053
			Heckman-2	0.023	0.382	0.146	0.017	0.278	0.078
200	0	0.8	Proposed	−0.130	0.200	0.057	0.089	0.172	0.037
			Heckman-1	−0.181	1.021	1.076	0.130	0.553	0.322
			cc-OLS-1	0.946	0.193	0.933	−0.485	0.196	0.274
			cc-OLS-2	0.614	0.134	0.395	−0.222	0.138	0.068
			Heckman-2	−0.083	0.340	0.122	0.077	0.207	0.049
200	0	0.5	Proposed	−0.078	0.262	0.074	0.058	0.198	0.043
			Heckman-1	−0.029	1.028	1.058	0.034	0.589	0.349
			cc-OLS-1	0.588	0.193	0.383	−0.298	0.204	0.131
			cc-OLS-2	0.357	0.143	0.148	−0.137	0.154	0.043
			Heckman-2	−0.011	0.374	0.140	0.032	0.239	0.058
200	1	0.8	Proposed	0.004	0.129	0.017	0.030	0.134	0.019
			Heckman-1	−0.018	0.442	0.196	0.043	0.301	0.093
			cc-OLS-1	0.491	0.128	0.257	−0.294	0.142	0.107
			cc-OLS-2	0.305	0.115	0.106	−0.134	0.113	0.031
			Heckman-2	0.051	0.235	0.058	−0.003	0.189	0.036
200	1	0.5	Proposed	0.003	0.142	0.020	0.021	0.140	0.020
			Heckman-1	0.042	0.438	0.194	−0.005	0.324	0.105
			cc-OLS-1	0.302	0.129	0.108	−0.177	0.142	0.051
			cc-OLS-2	0.170	0.115	0.042	−0.089	0.129	0.024
			Heckman-2	0.049	0.207	0.045	−0.019	0.182	0.034

Table 5. Bias, standard deviation, and mean square error for five estimates of (β_0, β_1) in Scenario IV, in which the model for $(\epsilon_{1i}, \epsilon_{2i}, \epsilon_{3i})^\tau$ is misspecified for all five methods, and $\mathbf{X}_{1i}$ is dependent on and overlaps with $(\mathbf{X}_{2i}, \mathbf{X}_{3i})$ completely.

				β_0			β_1		
n	γ_0	ρ	Methods	Bias	SD	MSE	Bias	SD	MSE
100	0	0.8	Proposed	−0.144	0.347	0.141	0.095	0.284	0.089
			Heckman-1	−0.088	1.315	1.737	0.095	0.741	0.558
			cc-OLS-1	0.948	0.252	0.962	−0.489	0.260	0.307
			cc-OLS-2	0.817	0.251	0.731	−0.451	0.262	0.272
			Heckman-2	0.032	0.981	0.964	0.026	0.605	0.367
100	0	0.5	Proposed	−0.076	0.431	0.192	0.042	0.345	0.121
			Heckman-1	−0.043	1.619	2.622	0.034	0.868	0.754
			cc-OLS-1	0.602	0.291	0.447	−0.320	0.297	0.191
			cc-OLS-2	0.470	0.247	0.282	−0.269	0.264	0.142
			Heckman-2	0.089	0.909	0.834	−0.026	0.575	0.331
100	1	0.8	Proposed	−0.023	0.185	0.035	0.049	0.224	0.052
			Heckman-1	0.090	0.597	0.364	−0.025	0.409	0.168
			cc-OLS-1	0.496	0.172	0.276	−0.300	0.197	0.129
			cc-OLS-2	0.411	0.178	0.201	−0.262	0.187	0.103
			Heckman-2	0.096	0.466	0.226	−0.023	0.375	0.141
100	1	0.5	Proposed	0.001	0.234	0.055	0.023	0.236	0.056
			Heckman-1	0.015	0.720	0.519	0.014	0.493	0.243
			cc-OLS-1	0.310	0.188	0.132	−0.185	0.219	0.082
			cc-OLS-2	0.217	0.175	0.078	−0.158	0.192	0.062
			Heckman-2	0.046	0.425	0.182	−0.027	0.355	0.127
200	0	0.8	Proposed	−0.186	0.290	0.119	0.075	0.243	0.065
			Heckman-1	−0.171	0.988	1.005	0.129	0.543	0.311
			cc-OLS-1	0.953	0.185	0.942	−0.483	0.198	0.272
			cc-OLS-2	0.818	0.169	0.698	−0.449	0.184	0.236
			Heckman-2	−0.075	0.751	0.570	0.076	0.468	0.225
200	0	0.5	Proposed	−0.099	0.362	0.141	0.048	0.273	0.077
			Heckman-1	−0.066	1.030	1.065	0.043	0.563	0.319
			cc-OLS-1	0.590	0.210	0.392	−0.306	0.204	0.135
			cc-OLS-2	0.471	0.173	0.252	−0.274	0.187	0.110
			Heckman-2	0.116	0.573	0.342	−0.050	0.397	0.160
200	1	0.8	Proposed	−0.025	0.147	0.022	0.036	0.165	0.029
			Heckman-1	−0.001	0.418	0.175	0.034	0.297	0.089
			cc-OLS-1	0.500	0.121	0.265	−0.296	0.140	0.107
			cc-OLS-2	0.402	0.119	0.176	−0.259	0.132	0.084
			Heckman-2	0.033	0.327	0.108	0.013	0.259	0.067
200	1	0.5	Proposed	0.002	0.173	0.030	0.031	0.181	0.034
			Heckman-1	0.000	0.445	0.198	0.031	0.313	0.099
			cc-OLS-1	0.320	0.138	0.121	−0.179	0.146	0.053
			cc-OLS-2	0.221	0.130	0.066	−0.158	0.144	0.046
			Heckman-2	0.080	0.257	0.073	−0.045	0.239	0.059

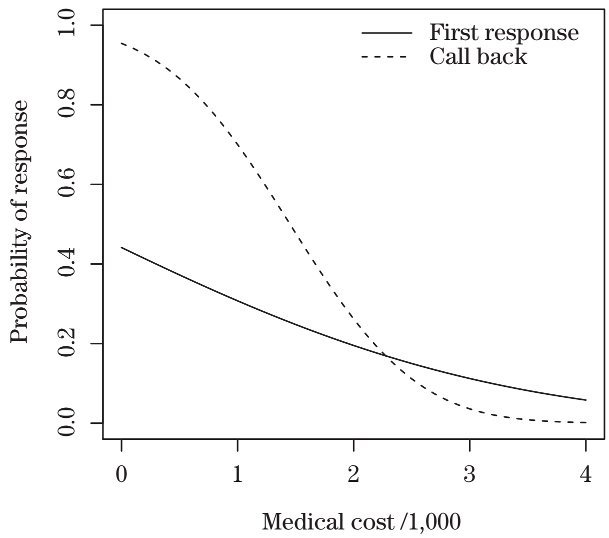

Figure 1. Dependence of the response probabilities on the medical cost for the NHIS data.

Table 6. Application to the NHIS data: response model.

	Intercept			FHI cost			FIN		
Method	Estimate	Se	p-value	Estimate	Se	p-value	Estimate	Se	p-value
Proposed	−0.542	0.023	<0.001	1.078	0.007	<0.001	−0.017	0.003	<0.001
Heckman-1	−0.493	0.088	<0.001	1.077	0.011	<0.001	−0.016	0.005	<0.001
cc-OLS-1	−0.578	0.029	<0.001	1.077	0.011	<0.001	−0.018	0.005	<0.001
cc-OLS-2	−0.598	0.018	<0.001	1.070	0.007	<0.001	−0.014	0.003	<0.001
Heckman-2	−0.567	0.027	<0.001	1.069	0.007	<0.001	−0.012	0.003	<0.001

bles 6 and 7 report the response models, missing data and call-back models. The significance of ρ_{12} and ρ_{13} indicates that the nonignorably missing mechanism is reasonable. Our method, the Heckman-2 and cc-OLS-2 methods yield similar estimates for the response model. This may because that the degree of nonignorable missingness is not too strong (the estimates of ρ_{12} and ρ_{13} are small), which is consistent with the observations of the simulation studies. All methods indicate that FHI cost is positively associated with medical costs, while family income is negatively associated with medical costs.

The covariate vectors for the missing-data (first response) and call-back models are different, indicating that the method of Alho (1990) is not appropriate for this data analysis. Although both the missing-data and call-back models indicate that the nonignorably missing mechanism is reasonable, the dependence of the response probabilities on the outcome is different. Figure 1 plots the de-

Table 7. Application to NHIS data: missing data and call-back model

	Proposed			Heckman-1			Heckman-2		
Parameter	Estimate	Se	p-value	Estimate	Se	p-value	Estimate	Se	p-value
Missing-data model:									
Intercept	−0.461	0.071	<0.001	−0.510	0.071	<0.001	0.548	0.069	<0.001
FMAL	0.097	0.053	0.066	0.097	0.053	0.065	0.155	0.053	0.004
RAT_CAT2	−0.023	0.006	<0.001	−0.019	0.006	0.001	−0.029	0.006	<0.001
Call-back model:									
Intercept	0.092	0.207	0.658						
FMAL	0.189	0.067	0.005						
FHI cost	−0.095	0.019	<0.001						
Error terms:									
σ	0.314	0.035	<0.001	0.320	0.015	<0.001	0.309	0.008	<0.001
ρ_{12}	−0.111	0.056	0.045	−0.164	0.254	0.519	−0.193	0.117	0.097
ρ_{13}	−0.343	0.103	0.001						
ρ_{23}	−0.030	0.474	0.949						

pendence of the response probabilities on the outcome for the first-response and call-back models; the other covariate values are replaced by their sample means. Both plots indicate that the response probability decreases as the medical cost increases. When the medical cost is not extremely high (for example, below \$3,000), the rate of decrease is lower for the probability of first response and higher for the probability of call-back success. This also indicates that the method of Alho (1990) is not appropriate, since Alho's method assumes a common effect of the outcome on the response probability.

In the missing-data model, the poverty ratio is negatively associated with the probability of first response; and the number of family members with limitations is positively associated with the probability of first response, but the significance is moderate. In the call-back model, the number of family members with limitations is positively associated with the probability of call-back success, while FHI cost is negatively associated with the probability of call-back success.

7. Conclusions and Discussion

We have proposed a likelihood-based method that incorporates call-back information and reduces the bias caused by the nonignorably missing mechanism. It is based on an adapted Heckman selection model. The missing-data and call-back indicators are assumed to be manifestations of latent variables, and the nonignorably missing mechanism is incorporated via correlations among these latent variables. The proposed method has a simple formulation, but it can reduce the bias and improve the estimation efficiency. We have proved that, under some conditions, the coefficient estimator of the response model is robust to the misspecification of the error distribution. Simulation studies have demonstrated that the method performs well under different scenarios.

In the Heckman selection model, the response and latent variables are assumed to follow a multivariate normal distribution. Marchenko and Genton (2012) extended the normality assumption to the t-distribution. The derivation is tedious, but our method can easily be extended to the t-distribution by assuming that the response and latent variables follow a multivariate t-distribution. We leave this to future research.

In this paper, we mainly discussed how to incorporate single call-back information by generalizing the Heckman selection model. In applications, there may be multiple call-backs. Our methods can be easily extended to this situation. We refer to the *Supplementary Materials* for more details.

Supplementary Materials

The *Supplementary Material* contains proof of Proposition 1, regularity conditions, derivation of score functions, and the extension of the proposed method in main paper to multiple call-backs.

Acknowledgment

The authors thank the Editor, an associate editor, and two referees for constructive comments and suggestions that led to significant improvements in the paper. Dr. Li's research is supported in part by NSERC Grant RGPIN-2015-06592.

References

Alexander, C. H., Dahl, S. and Weidmann, L. (1997). Making estimates from the American Community Survey. Paper presented to the Annual Meeting of the American Statistical Association (ASA), Anaheim, CA.

Alho, J. M. (1990). Adjusting for nonresponse bias using logistic regression. *Biometrika* **77**, 617–624.

Daniels, M. J., Jackson, D., Feng, W. and White, I. R. (2015). Pattern mixture models for the analysis of repeated attempt designs. *Biometrics* **71**, 1160–1167.

Daniels, M. J. and Hogan, J. W. (2000). Reparameterizing the pattern mixture model for sensitivity analyses under informative dropout. *Biometrics* **56**, 1241–1248.

Daniels, M. J. and Hogan, J. W. (2008). Missing data in longitudinal studies: strategies for bayesian modeling and sensitivity analysis, In *Monographs on Statistics and Applied Probability* **109**. Boca Raton, FL: Chapman & Hall/CRC.

Drew, J. H. and Fuller, W. A. (1980). Modeling nonresponse in surveys with callbacks. In *Proceedings of the Section on Survey Research Methods of the American Statistical Association*, 639–642.

Drew, J. H. and Fuller, W. A. (1981). Nonresponse in complex multiphase surveys. In *Proceedings of the Section on Survey Research Methods of the American Statistical Association*, 623–628.

Elliott, M. R., Little, R. J. A. and Lewitzky, S. (2000). Subsampling callbacks to improve survey efficiency. *Journal of the American Statistical Association* **95**, 730–738.

Gendall, P. and Davis, P. (1993). Are callbacks a waste of time? *Marketing Bulletin* **4**, 53–57.

Efron, B. (1979). Bootstrap methods: another look at the jackknife. *The Annals of Statistics* **7**, 1–26.

Grady, W. R. (1981). National Survey of Family Growth, Cycle II: Sample design, estimation procedures, and variance estimation. *Vital and Health Statistics Series 2* **87**, 1–36.

Groves, R. M. and Heeringa, S. G. (2006). Responsive design for household surveys: Tools for actively controlling survey errors and costs. *Journal of the Royal Statistical Society: Series A* (Statistics in Society) **169**, 439–457.

He, W. and Lawless, J. F. (2005). Bivariate location–scale models for regression analysis, with applications to lifetime data. *Journal of the Royal Statistical Society: Series B* (Statistical Methodology) **67**, 63–78.

Heckman, J. J. (1976). The common structure of statistical models of truncation, sample selection and limited dependent variables and a simple estimator for such models. *The Annals of Economic and Social Measurement* **5**, 475–492.

Heckman, J. J. (1979). Sample selection bias as a specification error. *Econometrica* **47**, 153–161.

Jackson, D., Mason, D., White, I. R. and Sutton, S. (2012). An exploration of the missing data mechanism in an internet based smoking cessation trial. *BMC Medical Research Methodology* **12**, 157.

Jackson, D., White, I. R. and Leese, M. (2010). How much can we learn about missing data? An exploration of a clinical trial in psychiatry. *Journal of the Royal Statistical Society: Series A* (Statistics in Society) **173**, 593–612.

Kessler, R. C. and Walters, E. E. (2002). The national comorbidity survey. *Textbook in Psychiatric Epidemiology.* (Edited by M. T. Tsaung, M. Tohen) 343–362. John Wiley and Sons, New York.

Kim, J. K. and Im, J. (2014). Propensity score adjustment with several follow-ups. *Biometrika* **101**, 439–448.

Lin, I. F. and Schaeffer, N. C. (1995). Using survey participants to estimate the impact of nonparticipants. *Public Opinion Quarterly* **59**, 236–258.

Little, R. J. (1993). Pattern-mixture models for multivariate incomplete data. *Journal of the American Statistical Association* **88**, 125–134.

Little, R. J. (1995). Modeling the drop-out mechanism in repeated-measures studies. *Journal of the American Statistical Association* **90**, 1112–1121.

Little, R. J. A. and Rubin, D. B. (2002). *Statistical Analysis with Missing Data* 2 Edition. John Wiley, New York.

Marchenko, Y. V. and Genton, M. G. (2012). A Heckman selection-t model. *Journal of the*

American Statistical Association **107**, 304–317.

Potthoff, R. F., Manton, K. G. and Woodbury, M. A. (1993). Correcting for nonavailability bias in surveys by weighting based on number of callbacks. *Journal of the American Statistical Association* **88**, 1197–1207.

Proctor, C. (1977). Two direct approaches to survey nonresponse: Estimating a proportion with callbacks and allocating effort to raise the response rate. In *Proceedings of the Social Statistics Secttion, American Statistical Association*, 284–290.

Qin, J. and Follmann, D. A. (2014). Semiparametric maximum likelihood inference by using failed contact attempts to adjust for the nonignorable nonresponse. *Biometrika* **101**, 985–991.

Serfling, R. J. (1980). *Approximation Theorems of Mathematical Statistics*. John Wiley & Sons.

Taylor, B. L. (2008). The 2006 National Health Interview Survey (NHIS) paradata file: Overview and applications. In *JSM Proceedings* 2008, *Survey Research Methods Section*. American Statistical Association, Alexandria, VA.

Toomet, O. and Henningsen, A. (2008). Sample selection models in R: Package sampleSelection. *Journal of Statistical Software* **27**, 1–23.

Tunali, I. (1986). A general structure for models of double-selection and an application to a joint-migration/earnings process with remigration. *Research in Labor Economics* **8**, 235–282.

White, H. (1982). Maximum likelihood estimation of misspecified models. *Econometrica* **50**, 1–25.

Wood, A. M., White, I. R. and Hotopf, M. (2006). Using number of failed contact attempts to adjust for non-ignorable nonresponse. *Journal of the Royal Statistical Society: Series A* (Statistics in Society) **169**, 525–542.

Department of Biostatistics, School of Public Health in Austin, University of Texas Health Science Center at Houston, Austin, TX 78701, USA.

E-mail: baojiang.chen@uth.tmc.edu

Department of Statistics and Actuarial Sciences, University of Waterloo, Waterloo, ON, Canada, N2L 3G1

E-mail: pengfei.li@uwaterloo.ca

National Institute of Allergy and Infectious Diseases, National Institute of Health, Bethesda, MD 20892, USA.

E-mail: jingqin@niaid.nih.gov

(Received June 2016; accepted March 2017)