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

Baojiang Chen, Pengfei Li and Jing Qin

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

This is a supplementary document to the corresponding paper submitted to the *Statistica Sinica*. It 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.

1 Proof of Proposition 1

Consider the linear regression model

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

the selection model

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

and the call-back model

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

Let $R_i = I(Z_i > 0)$.

Let $f(y, r, d|\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3; \boldsymbol{\theta})$ be the joint distribution of (Y, R, D) conditional on $\mathbf{X}_1 = \mathbf{x}_1$, $\mathbf{X}_2 = \mathbf{x}_2$, and $\mathbf{X}_3 = \mathbf{x}_3$. Under models (1), (2), and (3),

$$\begin{aligned}
f(y, r, d|\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3; \boldsymbol{\theta}) &= P(Y = y, R = r, D = d|\mathbf{X}_1 = \mathbf{x}_1, \mathbf{X}_2 = \mathbf{x}_2, \mathbf{X}_3 = \mathbf{x}_3) \\
&= \big\{P(Y = y, R = 1|\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3)\big\}^{r} \\
&\quad \times\big\{P(Y = y, R = 0, D = 1|\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3\big\}^{(1-r)d} \\
&\quad \times\big\{P(R = 0, D = 0|\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3)\big\}^{(1-r)(1-d)}.
\end{aligned}$$

The three terms in $f(y, r, d|\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3; \boldsymbol{\theta})$ are discussed in (1.7), (1.8), and (1.10) in the main paper, respectively.

We need to prove that if

$$f(y, r, d|\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3; \boldsymbol{\theta}) = f(y, r, d|\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3; \boldsymbol{\theta}^*) \tag{4}$$

for all possible values of y, r, d, $\mathbf{x}_1$, $\mathbf{x}_2$, and $\mathbf{x}_3$, then we must have $\boldsymbol{\theta} = \boldsymbol{\theta}^*$.

We first consider the identifiability of $(\boldsymbol{\beta}^\tau, \boldsymbol{\gamma}^\tau, \sigma, \rho_{12})^\tau$. When $r = 1$, (4) implies that

$$P(Y = y, R = 1|\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3; \boldsymbol{\beta}, \boldsymbol{\gamma}, \sigma, \rho_{12}) = P(Y = y, R = 1|\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3; \boldsymbol{\beta}^*, \boldsymbol{\gamma}^*, \sigma^*, \rho_{12}^*).$$

By the identifiability of Heckman selection model, see for example, Example 5 of Miao, Ding, and Geng (2016), we have

$$\boldsymbol{\beta} = \boldsymbol{\beta}^*,\ \boldsymbol{\gamma} = \boldsymbol{\gamma}^*,\ \sigma = \sigma^*,\ \rho_{12} = \rho_{12}^*.$$

Hence the parameters $(\boldsymbol{\beta}^\tau, \boldsymbol{\gamma}^\tau, \sigma, \rho_{12})^\tau$ are identifiable. This finishes the proof of the first part of Proposition 1.

Next we consider the identifiability of $(\boldsymbol{\xi}^\tau, \rho_{13}, \rho_{23})^\tau$. When $r = 0$ and $d = 1$, together with the identifiability of $(\boldsymbol{\beta}^\tau, \boldsymbol{\gamma}^\tau, \sigma, \rho_{12})^\tau$, (4) implies that

$$\int_{-\infty}^{-\mathbf{x}_2^\tau \boldsymbol{\gamma}} \int_{-\mathbf{x}_3^\tau \boldsymbol{\xi}}^{\infty} \phi_{23|1}(t, u; s) dt du = \int_{-\infty}^{-\mathbf{x}_2^\tau \boldsymbol{\gamma}} \int_{-\mathbf{x}_3^\tau \boldsymbol{\xi}^*}^{\infty} \phi_{23|1}^*(t, u; s) dt du \tag{5}$$

for all $\mathbf{x}_2$, $\mathbf{x}_3$, s. Here $\phi_{23|1}^*$ is the the density of the bivariate normal with mean vector $(\rho_{12}s, \rho_{13}^* s)^\tau$ and the covariance matrix

$$\begin{pmatrix} 1 - \rho_{12}^2 & \rho_{23}^* - \rho_{12}\rho_{13}^* \\ \rho_{23}^* - \rho_{12}\rho_{13}^* & 1 - (\rho_{13}^*)^2 \end{pmatrix}.$$

From (5), we further get that

$$\int_{-\infty}^{\gamma} \int_{-\infty}^{\xi} \phi_{23|1}(t, u; s) dt du = \int_{-\infty}^{\gamma} \int_{-\infty}^{\xi^*} \phi_{23|1}^*(t, u; s) dt du, \tag{6}$$

where $\gamma = -\mathbf{x}_2^\tau \boldsymbol{\gamma}$, $\xi = -\mathbf{x}_3^\tau \boldsymbol{\xi}$, and $\xi^* = -\mathbf{x}_3^\tau \boldsymbol{\xi}^*$.

With the condition that $\mathbf{X}_2$ contains a continuous covariate which does not appear in $\mathbf{X}_3$, we can find a γ_0 such that for γ in a small neighbourhood of γ_0,

$$\int_{-\infty}^{\gamma} \int_{-\infty}^{\xi} \phi_{23|1}(t, u; s) dt du = \int_{-\infty}^{\gamma} \int_{-\infty}^{\xi^*} \phi_{23|1}^*(t, u; s) dt du,$$

which implies that for γ in a small neighbourhood of γ_0

$$\int_{-\infty}^{\xi} \phi_{23|1}(\gamma, u; s) du = \int_{-\infty}^{\xi^*} \phi_{23|1}^*(\gamma, u; s) du. \tag{7}$$

With some calculus work, we obtain from (7) that

$$\begin{aligned} & \frac{1}{\sqrt{1 - \rho_{12}^2}} \phi\left(\frac{\gamma - \rho_{12}s}{\sqrt{1 - \rho_{12}^2}}\right) \Phi\left(\frac{\xi - \frac{\rho_{23} - \rho_{12}\rho_{13}}{1 - \rho_{12}^2}\gamma - \frac{\rho_{13} - \rho_{12}\rho_{23}}{1 - \rho_{12}^2}s}{\sqrt{1 - \rho_{13}^2 - \frac{(\rho_{23} - \rho_{12}\rho_{13})^2}{1 - \rho_{12}^2}}}\right) \\ = \; & \frac{1}{\sqrt{1 - \rho_{12}^2}} \phi\left(\frac{\gamma - \rho_{12}s}{\sqrt{1 - \rho_{12}^2}}\right) \Phi\left(\frac{\xi^* - \frac{\rho_{23}^* - \rho_{12}\rho_{13}^*}{1 - \rho_{12}^2}\gamma - \frac{\rho_{13}^* - \rho_{12}\rho_{23}^*}{1 - \rho_{12}^2}s}{\sqrt{1 - (\rho_{13}^*)^2 - \frac{(\rho_{23}^* - \rho_{12}\rho_{13}^*)^2}{1 - \rho_{12}^2}}}\right). \end{aligned}$$

Therefore,

$$\frac{\xi - \frac{\rho_{23}-\rho_{12}\rho_{13}}{1-\rho_{12}^2}\gamma - \frac{\rho_{13}-\rho_{12}\rho_{23}}{1-\rho_{12}^2}s}{\sqrt{1-\rho_{13}^2 - \frac{(\rho_{23}-\rho_{12}\rho_{13})^2}{1-\rho_{12}^2}}} = \frac{\xi^* - \frac{\rho_{23}^*-\rho_{12}\rho_{13}^*}{1-\rho_{12}^2}\gamma - \frac{\rho_{13}^*-\rho_{12}\rho_{23}^*}{1-\rho_{12}^2}s}{\sqrt{1-(\rho_{13}^*)^2 - \frac{(\rho_{23}^*-\rho_{12}\rho_{13}^*)^2}{1-\rho_{12}^2}}}$$

for γ in a small neighbourhood of γ_0 and all s. Then we must have

$$\begin{aligned}
\frac{\xi}{\sqrt{(1-\rho_{13}^2)(1-\rho_{12}^2)-(\rho_{23}-\rho_{12}\rho_{13})^2}} &= \frac{\xi^*}{\sqrt{\{1-(\rho_{13}^*)^2\}(1-\rho_{12}^2)-(\rho_{23}^*-\rho_{12}\rho_{13}^*)^2}},\\
\frac{\rho_{23}-\rho_{12}\rho_{13}}{\sqrt{(1-\rho_{13}^2)(1-\rho_{12}^2)-(\rho_{23}-\rho_{12}\rho_{13})^2}} &= \frac{\rho_{23}^*-\rho_{12}\rho_{13}^*}{\sqrt{\{1-(\rho_{13}^*)^2\}(1-\rho_{12}^2)-(\rho_{23}^*-\rho_{12}\rho_{13}^*)^2}},\\
\frac{\rho_{13}-\rho_{12}\rho_{23}}{\sqrt{(1-\rho_{13}^2)(1-\rho_{12}^2)-(\rho_{23}-\rho_{12}\rho_{13})^2}} &= \frac{\rho_{13}^*-\rho_{12}\rho_{23}^*}{\sqrt{\{1-(\rho_{13}^*)^2\}(1-\rho_{12}^2)-(\rho_{23}^*-\rho_{12}\rho_{13}^*)^2}}.
\end{aligned}$$

By solving the above three equations and some algebra work, we further have

$$\xi = \xi^*, \quad \rho_{13} = \rho_{13}^*, \quad \rho_{23} = \rho_{23}^*.$$

Recall that the components of $\mathbf{X}_3$ are linearly independent. Then $\xi = \xi^*$ implies that $\boldsymbol{\xi} = \boldsymbol{\xi}^*$. Hence the parameters $(\boldsymbol{\xi}^\tau, \rho_{13}, \rho_{23})^\tau$ are identifiable. This finishes the proof.

2 Regularity conditions

To ensure the asymptotic normality of $\widehat{\boldsymbol{\theta}}$ under the correctly specified models, we need the following regularity conditions.

A1. Suppose the response, missing-data, and call-back models (1), (2), and (3) are correctly specified for (Y_i, Z_i, U_i). Further, the joint distribution of $(\epsilon_{1i}, \epsilon_{2i}, \epsilon_{3i})^\tau$ is trivariate normal with mean vector $\mathbf{0}$ and covariance matrix

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

A2. The errors $(\epsilon_{1i}, \epsilon_{2i}, \epsilon_{3i})$ are independent from $(\mathbf{X}_{1i}, \mathbf{X}_{2i}, \mathbf{X}_{3i})$.

A3. $E\{|\log f(Y, R, D|\mathbf{X}_1, \mathbf{X}_2, \mathbf{X}_3; \boldsymbol{\theta}_0)|\} < \infty$, where $\boldsymbol{\theta}_0$ is the true value of $\boldsymbol{\theta}$ and the expectation is taken under the assumption that $\boldsymbol{\theta} = \boldsymbol{\theta}_0$.

A4. The Fisher information matrix

$$E\left\{-\frac{\partial^2 \log f(Y, R, D|\mathbf{X}_1, \mathbf{X}_2, \mathbf{X}_3; \boldsymbol{\theta}_0)}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^\tau}\right\}$$

is positive definite.

A5. There exists a function $B(y, r, d, \mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3)$, possible depending on $\boldsymbol{\theta}_0$, such that for $\boldsymbol{\theta}$ in a neighborhood of $\boldsymbol{\theta}_0$,

$$\left| \frac{\partial^3 \log f(y, r, d|\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3; \boldsymbol{\theta})}{\partial\theta_i\theta_j\theta_k} \right| \le B(y, r, d, \mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3)$$

for all $(y, r, d, \mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3)$ and $i, j, k = 1, \ldots, p + q + r + 4$, and

$$E\{B(Y, R, D, \mathbf{X}_1, \mathbf{X}_2, \mathbf{X}_3)\} < \infty.$$

Here θ_i denotes the ith element of $\boldsymbol{\theta}$.

To ensure the consistency of $\widehat{\boldsymbol{\theta}}$ under the misspecified models, we need a new set of regularity conditions.

B1. Suppose the true model for (Y_i, Z_i, U_i) is (1.14) in the main paper and the joint cumulative distribution function of $(w_{1i}, w_{2i}, w_{3i})^\tau$ is $H(s, t, u)$.

B2. The errors (w_{1i}, w_{2i}, w_{3i}) are independent from $(\mathbf{X}_{1i}, \mathbf{X}_{2i}, \mathbf{X}_{3i})$.

B3. There exists a function $C_1(y, r, d, \mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3)$ such that for all $\boldsymbol{\theta}$

$$|\log f(y, r, d|\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3; \boldsymbol{\theta})| \le C_1(y, r, d, \mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3)$$

and

$$E_T\{C_1(Y, R, D, \mathbf{X}_1, \mathbf{X}_2, \mathbf{X}_3)\} < \infty.$$

Here E_T means that the expectation is taken under the true model specified in B1.

B4. $E_T\{\log f(Y, R, D|\mathbf{X}_1, \mathbf{X}_2, \mathbf{X}_3; \boldsymbol{\theta})\}$ is uniquely maximized at $\boldsymbol{\theta} = \boldsymbol{\theta}^*$.

B5. There exists a function $C_2(y, r, d, \mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3)$, possible depending on $\boldsymbol{\theta}^*$, such that for $\boldsymbol{\theta}$ in a neighborhood of $\boldsymbol{\theta}^*$,

$$\left| \frac{\partial^3 \log f(y, r, d|\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3; \boldsymbol{\theta})}{\partial\theta_i\theta_j\theta_k} \right| \le C_2(y, r, d, \mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3)$$

for all $(y, r, d, \mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3)$ and $i, j, k = 1, \ldots, p + q + r + 4$, and

$$E_T\{C_2(Y, R, D, \mathbf{X}_1, \mathbf{X}_2, \mathbf{X}_3)\} < \infty.$$

B6. The two matrices

$$E_T\left\{-\frac{\partial^2 \log f(Y, R, D|\mathbf{X}_1, \mathbf{X}_2, \mathbf{X}_3; \boldsymbol{\theta}^*)}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^\tau}\right\}$$

and

$$Var_T\left\{\frac{\partial \log f(Y, R, D|\mathbf{X}_1, \mathbf{X}_2, \mathbf{X}_3; \boldsymbol{\theta}^*)}{\partial\boldsymbol{\theta}}\right\}$$

are positive definite.

3 Derivation of score functions

Some preparation

Recall that $\epsilon_{1i} = (y_i - \beta_0 - \mathbf{X}_{1i}^{\tau}\boldsymbol{\beta}_1)/\sigma$, $\phi_{23|1}(t,u|s)$ is the density of the bivariate normal with mean vector $\boldsymbol{\mu}_{23|1}$ and the covariance matrix $\boldsymbol{\Sigma}_{23|1}$ specified in (1.9) in the main paper, and $\phi_{23}(t,u)$ is the density of the bivariate normal with mean vector $\mathbf{0}$ and the covariance matrix $\boldsymbol{\Sigma}_{23}$ specified in (1.11) in the main paper. Then

$$\phi_{23|1}(t,u|\epsilon_{1i}) = \frac{1}{2\pi|\boldsymbol{\Sigma}_{23|1}|^{1/2}}\exp\Big\{-\frac{1}{2}(t-\rho_{12}\epsilon_{1i}, u-\rho_{13}\epsilon_{1i})\boldsymbol{\Sigma}_{23|1}^{-1}(t-\rho_{12}\epsilon_{1i}, u-\rho_{13}\epsilon_{1i})^{\tau}\Big\}$$

and

$$\phi_{23}(t,u) = \frac{1}{2\pi|\boldsymbol{\Sigma}_{23}|^{1/2}}\exp\Big\{-\frac{1}{2}(t,u)\boldsymbol{\Sigma}_{23}^{-1}(t,u)^{\tau}\Big\}.$$

When deriving the form of $\boldsymbol{S}_i(\boldsymbol{\theta})$, we need the derivatives of $\phi_{23|1}(t,u|\epsilon_{1i})$ with respect to $\boldsymbol{\beta}$, σ, ρ_{12}, ρ_{13}, and ρ_{23}, and the derivative of $\phi_{23}(t,u)$ with respect to ρ_{23}. We first summarize them.

Let $\mathbf{X}_{1i}^{*} = (1, \mathbf{X}_{1i}^{\tau})^{\tau}$ and

$$\begin{aligned} h_{23|1}(t,u;s) &= -0.5(t-\rho_{12}s, u-\rho_{13}s)\boldsymbol{\Sigma}_{23|1}^{-1}(t-\rho_{12}s, u-\rho_{13}s)^{\tau} \\ &= -0.5|\boldsymbol{\Sigma}_{23|1}|^{-1}\Big\{(1-\rho_{13})^2(t-\rho_{12}s)^2 + 2(\rho_{12}\rho_{13}-\rho_{23})(t-\rho_{12}s)(u-\rho_{13}s) \\ &\qquad + (1-\rho_{12})^2(u-\rho_{13}s)^2\Big\}. \end{aligned}$$

It can be verified that

$$\frac{\partial\phi_{23|1}(t,u|\epsilon_{1i})}{\partial\boldsymbol{\beta}} = -\sigma^{-1}\phi_{23|1}(t,u|\epsilon_{1i})(t-\rho_{12}\epsilon_{1i}, u-\rho_{13}\epsilon_{1i})\boldsymbol{\Sigma}_{23|1}^{-1}(\rho_{12},\rho_{13})^{\tau}\mathbf{X}_{1i}^{*}, \tag{8}$$

$$\frac{\partial\phi_{23|1}(t,u|\epsilon_{1i})}{\partial\sigma} = -\sigma^{-1}\phi_{23|1}(t,u|\epsilon_{1i})(t-\rho_{12}\epsilon_{1i}, u-\rho_{13}\epsilon_{1i})\boldsymbol{\Sigma}_{23|1}^{-1}(\rho_{12},\rho_{13})^{\tau}\epsilon_{1i}, \tag{9}$$

$$\frac{\partial\phi_{23|1}(t,u|\epsilon_{1i})}{\partial\rho_{12}} = \phi_{23|1}(t,u|\epsilon_{1i})\left\{-0.5|\boldsymbol{\Sigma}_{23|1}|^{-1}\frac{\partial|\boldsymbol{\Sigma}_{23|1}|}{\partial\rho_{12}} + \frac{\partial h_{23|1}(t,u;\epsilon_{1i})}{\partial\rho_{12}}\right\}, \tag{10}$$

$$\frac{\partial\phi_{23|1}(t,u|\epsilon_{1i})}{\partial\rho_{13}} = \phi_{23|1}(t,u|\epsilon_{1i})\left\{-0.5|\boldsymbol{\Sigma}_{23|1}|^{-1}\frac{\partial|\boldsymbol{\Sigma}_{23|1}|}{\partial\rho_{13}} + \frac{\partial h_{23|1}(t,u;\epsilon_{1i})}{\partial\rho_{13}}\right\}, \tag{11}$$

$$\frac{\partial\phi_{23|1}(t,u|\epsilon_{1i})}{\partial\rho_{23}} = \phi_{23|1}(t,u|\epsilon_{1i})\left\{-0.5|\boldsymbol{\Sigma}_{23|1}|^{-1}\frac{\partial|\boldsymbol{\Sigma}_{23|1}|}{\partial\rho_{23}} + \frac{\partial h_{23|1}(t,u;\epsilon_{1i})}{\partial\rho_{23}}\right\}. \tag{12}$$

Here $|\boldsymbol{\Sigma}_{23|1}| = (1-\rho_{12}^2)(1-\rho_{13}^2) - (\rho_{23}-\rho_{12}\rho_{13})^2$ and

$$\begin{aligned} \frac{\partial|\boldsymbol{\Sigma}_{23|1}|}{\partial\rho_{12}} &= -2(\rho_{12}-\rho_{13}\rho_{23}), \\ \frac{\partial|\boldsymbol{\Sigma}_{23|1}|}{\partial\rho_{13}} &= -2(\rho_{13}-\rho_{12}\rho_{23}), \\ \frac{\partial|\boldsymbol{\Sigma}_{23|1}|}{\partial\rho_{23}} &= -2(\rho_{23}-\rho_{12}\rho_{13}). \end{aligned}$$

After some calculus work, we have that

$$\begin{aligned}\frac{\partial}{\partial\rho_{12}}h_{23|1}(t,u;\epsilon_{1i}) &= 2|\mathbf{\Sigma}_{23|1}|^{-1}(\rho_{12}-\rho_{13}\rho_{23})h_{23|1}(t,u|\epsilon_{1i})\\ &\quad -0.5|\mathbf{\Sigma}_{23|1}|^{-1}\Big\{-2\epsilon_{1i}(1-\rho_{13})^2(t-\rho_{12}\epsilon_{1i})+2\rho_{13}(t-\rho_{12}\epsilon_{1i})(u-\rho_{13}\epsilon_{1i})\\ &\quad -2\epsilon_{1i}(\rho_{12}\rho_{13}-\rho_{23})(u-\rho_{13}\epsilon_{1i})-2(1-\rho_{12})(u-\rho_{13}\epsilon_{1i})^2\Big\}.\end{aligned}$$

Similarly,

$$\begin{aligned}\frac{\partial}{\partial\rho_{13}}h_{23|1}(t,u;\epsilon_{1i}) &= 2|\mathbf{\Sigma}_{23|1}|^{-1}(\rho_{13}-\rho_{12}\rho_{23})h_{23|1}(t,u|\epsilon_{1i})\\ &\quad -0.5|\mathbf{\Sigma}_{23|1}|^{-1}\Big\{-2(1-\rho_{13})(t-\rho_{12}\epsilon_{1i})^2+2\rho_{12}(t-\rho_{12}\epsilon_{1i})(u-\rho_{13}\epsilon_{1i})\\ &\quad -2\epsilon_{1i}(\rho_{12}\rho_{13}-\rho_{23})(t-\rho_{12}\epsilon_{1i})-2\epsilon_{1i}(1-\rho_{12})^2(u-\rho_{13}\epsilon_{1i})\Big\}\end{aligned}$$

and

$$\frac{\partial}{\partial\rho_{23}}h_{23|1}(t,u;\epsilon_{1i}) = 2|\mathbf{\Sigma}_{23|1}|^{-1}(\rho_{23}-\rho_{12}\rho_{13})h_{23|1}(t,u|\epsilon_{1i})+|\mathbf{\Sigma}_{23|1}|^{-1}(t-\rho_{12}\epsilon_{1i})(u-\rho_{13}\epsilon_{1i}).$$

Combining the above terms, we get the derivatives of $\phi_{23|1}(t,u|\epsilon_{1i})$ with respect to $\boldsymbol{\beta}$, σ, ρ_{12}, ρ_{13}, and ρ_{23}.

As a final piece of preparation, we provide the form of $\partial\phi_{23}(t,u)/\partial\rho_{23}$. Note that $\phi_{23}(t,u)$ can be rewritten as

$$\phi_{23}(t,u)=\frac{1}{2\pi\sqrt{1-\rho_{23}^2}}\exp\left\{-\frac{1}{2(1-\rho_{23}^2)}(t^2-2\rho_{23}tu+u^2)\right\}.$$

Hence,

$$\frac{\partial\phi_{23}(t,u)}{\partial\rho_{23}}=\phi_{23}(t,u)\left\{\frac{\rho_{23}}{1-\rho_{23}^2}-\frac{\rho_{23}}{(1-\rho_{23}^2)^2}(t^2-2\rho_{23}tu+u^2)+\frac{tu}{1-\rho_{23}^2}\right\}. \tag{13}$$

Form of $\boldsymbol{S}_i(\boldsymbol{\theta})$

For ease of expression, we denote $g(u)=\phi(u)/\Phi(u)$ and use the result that $\phi'(u)=-u\phi(u)$. Recall that $\boldsymbol{S}_i(\boldsymbol{\theta})=\partial\ell_i(\boldsymbol{\theta})/\partial\boldsymbol{\theta}$. Next we find each term in $\boldsymbol{S}_i(\boldsymbol{\theta})$.

For $\partial\ell_i(\boldsymbol{\theta})/\partial\boldsymbol{\beta}$, we have that

$$\begin{aligned}\frac{\partial\ell_i(\boldsymbol{\theta})}{\partial\boldsymbol{\beta}} &= \frac{\partial\ell_{1i}(\boldsymbol{\theta})}{\partial\boldsymbol{\beta}}+\frac{\partial\ell_{2i}(\boldsymbol{\theta})}{\partial\boldsymbol{\beta}}\\ &= R_i\sigma^{-1}\left\{\epsilon_{1i}-g\Big(\frac{\mathbf{X}_{2i}^{\tau}\boldsymbol{\gamma}+\rho_{12}\epsilon_{1i}}{\sqrt{1-\rho_{12}^2}}\Big)\frac{\rho_{12}}{\sqrt{(1-\rho_{12}^2)}}\mathbf{X}_{1i}^*\right\}\\ &\quad +D_i(1-R_i)\left\{\frac{\int_{-\infty}^{-\mathbf{X}_{2i}^{\tau}\boldsymbol{\gamma}}\int_{-\mathbf{X}_{3i}^{\tau}\boldsymbol{\xi}}^{\infty}\frac{\partial}{\partial\boldsymbol{\beta}}\phi_{23|1}(t,u|\epsilon_{1i})dtdu}{\int_{-\infty}^{-\mathbf{X}_{2i}^{\tau}\boldsymbol{\gamma}}\int_{-\mathbf{X}_{3i}^{\tau}\boldsymbol{\xi}}^{\infty}\phi_{23|1}(t,u|\epsilon_{1i})dtdu}+\sigma^{-1}\epsilon_{1i}\mathbf{X}_{1i}^*\right\},\end{aligned}$$

where $\partial\phi_{23|1}(t,u|\epsilon_{1i})/\partial\boldsymbol{\beta}$ is given in (8).

For $\partial\ell_i(\boldsymbol{\theta})/\partial\boldsymbol{\gamma}$, we have that

$$\begin{aligned}
\frac{\partial\ell_i(\boldsymbol{\theta})}{\partial\boldsymbol{\gamma}} &= \frac{\partial\ell_{1i}(\boldsymbol{\theta})}{\partial\boldsymbol{\gamma}} + \frac{\partial\ell_{2i}(\boldsymbol{\theta})}{\partial\boldsymbol{\gamma}} + \frac{\partial\ell_{3i}(\boldsymbol{\theta})}{\partial\boldsymbol{\gamma}} \\
&= R_i\left\{g\Big(\frac{\mathbf{X}_{2i}^\tau\gamma+\rho_{12}\epsilon_{1i}}{\sqrt{1-\rho_{12}^2}}\Big)\frac{1}{\sqrt{(1-\rho_{12}^2)}}\mathbf{X}_{2i}\right\} \\
&\quad -D_i(1-R_i)\left\{\frac{\int_{-\mathbf{X}_{3i}^\tau\boldsymbol{\xi}}^{\infty}\phi_{23|1}(-\mathbf{X}_{2i}^\tau\boldsymbol{\gamma},u|\epsilon_{1i})du}{\int_{-\infty}^{-\mathbf{X}_{2i}^\tau\boldsymbol{\gamma}}\int_{-\mathbf{X}_{3i}^\tau\boldsymbol{\xi}}^{\infty}\phi_{23|1}(t,u|\epsilon_{1i})dtdu}\mathbf{X}_{2i}\right\} \\
&\quad -(1-R_i)(1-D_i)\left\{\frac{\int_{-\infty}^{-\mathbf{X}_{3i}^\tau\boldsymbol{\xi}}\phi_{23}(-\mathbf{X}_{2i}^\tau\boldsymbol{\gamma},u)du}{\int_{-\infty}^{-\mathbf{X}_{2i}^\tau\boldsymbol{\gamma}}\int_{-\infty}^{-\mathbf{X}_{3i}^\tau\boldsymbol{\xi}}\phi_{23}(t,u)dtdu}\mathbf{X}_{2i}\right\}.
\end{aligned}$$

For $\partial\ell_i(\boldsymbol{\theta})/\partial\boldsymbol{\xi}$, we have that

$$\begin{aligned}
\frac{\partial\ell_i(\boldsymbol{\theta})}{\partial\boldsymbol{\xi}} &= \frac{\partial\ell_{2i}(\boldsymbol{\theta})}{\partial\boldsymbol{\xi}} + \frac{\partial\ell_{3i}(\boldsymbol{\theta})}{\partial\boldsymbol{\xi}} \\
&= D_i(1-R_i)\left\{\frac{\int_{-\infty}^{-\mathbf{X}_{2i}^\tau\boldsymbol{\gamma}}\phi_{23|1}(t,-\mathbf{X}_{3i}^\tau\boldsymbol{\xi}|\epsilon_{1i})du}{\int_{-\infty}^{-\mathbf{X}_{2i}^\tau\boldsymbol{\gamma}}\int_{-\mathbf{X}_{3i}^\tau\boldsymbol{\xi}}^{\infty}\phi_{23|1}(t,u|\epsilon_{1i})dtdu}\mathbf{X}_{3i}\right\} \\
&\quad -(1-R_i)(1-D_i)\left\{\frac{\int_{-\infty}^{-\mathbf{X}_{2i}^\tau\boldsymbol{\gamma}}\phi_{23}(t,-\mathbf{X}_{3i}^\tau\boldsymbol{\xi})du}{\int_{-\infty}^{-\mathbf{X}_{2i}^\tau\boldsymbol{\gamma}}\int_{-\infty}^{-\mathbf{X}_{3i}^\tau\boldsymbol{\xi}}\phi_{23}(t,u)dtdu}\mathbf{X}_{3i}\right\}.
\end{aligned}$$

For $\partial\ell_i(\boldsymbol{\theta})/\partial\sigma$, we have that

$$\begin{aligned}
\frac{\partial\ell_i(\boldsymbol{\theta})}{\partial\sigma} &= \frac{\partial\ell_{1i}(\boldsymbol{\theta})}{\partial\sigma} + \frac{\partial\ell_{2i}(\boldsymbol{\theta})}{\partial\sigma} \\
&= R_i\sigma^{-1}\Big\{\epsilon_{1i}^2 - g\Big(\frac{\mathbf{X}_{2i}^\tau\gamma+\rho_{12}\epsilon_{1i}}{\sqrt{1-\rho_{12}^2}}\Big)\frac{\rho_{12}\epsilon_{1i}}{\sqrt{1-\rho_{12}^2}} - 1\Big\} \\
&\quad +D_i(1-R_i)\left\{\frac{\int_{-\infty}^{-\mathbf{X}_{2i}^\tau\boldsymbol{\gamma}}\int_{-\mathbf{X}_{3i}^\tau\boldsymbol{\xi}}^{\infty}\frac{\partial}{\partial\sigma}\phi_{23|1}(t,u|\epsilon_{1i})dtdu}{\int_{-\infty}^{-\mathbf{X}_{2i}^\tau\boldsymbol{\gamma}}\int_{-\mathbf{X}_{3i}^\tau\boldsymbol{\xi}}^{\infty}\phi_{23|1}(t,u|\epsilon_{1i})dtdu} - \sigma^{-1} + \sigma^{-1}\epsilon_{1i}^2\right\},
\end{aligned}$$

where $\partial\phi_{23|1}(t,u|\epsilon_{1i})/\partial\sigma$ is given in (9).

For $\partial\ell_i(\boldsymbol{\theta})/\partial\rho_{12}$, we have that

$$\begin{aligned}
\frac{\partial\ell_i(\boldsymbol{\theta})}{\partial\rho_{12}} &= \frac{\partial\ell_{1i}(\boldsymbol{\theta})}{\partial\rho_{12}} + \frac{\partial\ell_{2i}(\boldsymbol{\theta})}{\partial\rho_{12}} \\
&= R_i\left\{g\Big(\frac{\mathbf{X}_{2i}^\tau\boldsymbol{\gamma}+\rho_{12}\epsilon_{1i}}{\sqrt{1-\rho_{12}^2}}\Big)\frac{\epsilon_{1i}+\rho_{12}\mathbf{X}_{2i}^\tau\boldsymbol{\gamma}}{(1-\rho_{12}^2)^{3/2}}\right\} \\
&\quad +D_i(1-R_i)\left\{\frac{\int_{-\infty}^{-\mathbf{X}_{2i}^\tau\boldsymbol{\gamma}}\int_{-\mathbf{X}_{3i}^\tau\boldsymbol{\xi}}^{\infty}\frac{\partial}{\partial\rho_{12}}\phi_{23|1}(t,u|\epsilon_{1i})dtdu}{\int_{-\infty}^{-\mathbf{X}_{2i}^\tau\boldsymbol{\gamma}}\int_{-\mathbf{X}_{3i}^\tau\boldsymbol{\xi}}^{\infty}\phi_{23|1}(t,u|\epsilon_{1i})dtdu}\right\},
\end{aligned}$$

where $\partial\phi_{23|1}(t,u|\epsilon_{1i})/\partial\rho_{12}$ is given in (10).

For $\partial\ell_i(\boldsymbol{\theta})/\partial\rho_{13}$, we have that

$$\frac{\partial\ell_i(\boldsymbol{\theta})}{\partial\rho_{13}} = \frac{\partial\ell_{2i}(\boldsymbol{\theta})}{\partial\rho_{13}} = D_i(1-R_i)\left\{\frac{\int_{-\infty}^{-\mathbf{X}_{2i}^\tau\boldsymbol{\gamma}}\int_{-\mathbf{X}_{3i}^\tau\boldsymbol{\xi}}^{\infty}\frac{\partial}{\partial\rho_{13}}\phi_{23|1}(t,u|\epsilon_{1i})dtdu}{\int_{-\infty}^{-\mathbf{X}_{2i}^\tau\boldsymbol{\gamma}}\int_{-\mathbf{X}_{3i}^\tau\boldsymbol{\xi}}^{\infty}\phi_{23|1}(t,u|\epsilon_{1i})dtdu}\right\},$$

where $\partial\phi_{23|1}(t,u|\epsilon_{1i})/\partial\rho_{13}$ is given in (11).

For $\partial\ell_i(\boldsymbol{\theta})/\partial\rho_{23}$, we have that

$$\begin{aligned}\frac{\partial\ell_i(\boldsymbol{\theta})}{\partial\rho_{23}} &= \frac{\partial\ell_{2i}(\boldsymbol{\theta})}{\partial\rho_{23}}+\frac{\partial\ell_{3i}(\boldsymbol{\theta})}{\partial\rho_{23}}\\ &= D_i(1-R_i)\left\{\frac{\int_{-\infty}^{-\mathbf{X}_{2i}^{\tau}\boldsymbol{\gamma}}\int_{-\mathbf{X}_{3i}^{\tau}\boldsymbol{\xi}}^{\infty}\frac{\partial}{\partial\rho_{23}}\phi_{23|1}(t,u|\epsilon_{1i})dtdu}{\int_{-\infty}^{-\mathbf{X}_{2i}^{\tau}\boldsymbol{\gamma}}\int_{-\mathbf{X}_{3i}^{\tau}\boldsymbol{\xi}}^{\infty}\phi_{23|1}(t,u|\epsilon_{1i})dtdu}\right\}\\ &\quad +(1-R_i)(1-D_i)\left\{\frac{\int_{-\infty}^{-\mathbf{X}_{2i}^{\tau}\boldsymbol{\gamma}}\int_{-\infty}^{-\mathbf{X}_{3i}^{\tau}\boldsymbol{\xi}}\frac{\partial}{\partial\rho_{23}}\phi_{23}(t,u)dtdu}{\int_{-\infty}^{-\mathbf{X}_{2i}^{\tau}\boldsymbol{\gamma}}\int_{-\infty}^{-\mathbf{X}_{3i}^{\tau}\boldsymbol{\xi}}\phi_{23}(t,u)dtdu}\right\},\end{aligned}$$

where $\partial\phi_{23|1}(t,u|\epsilon_{1i})/\partial\rho_{23}$ is given in (12) and $\partial\phi_{23}(t,u)/\partial\rho_{23}$ is given in (13).

4 Extension to multiple call-backs

The proposed method in Section 4 of the main paper can easily be extended to multiple call-backs.

Suppose there are K call-backs, and let $D_{ik}=1$ if the ith subject is called back, and 0 otherwise, $k=1,\ldots,K$. We again assume that D_{ik} is a manifestation of a latent variable U_{ik}, which is from the multivariate regression model

$$U_{ik}=\mathbf{X}_{3ik}^{\tau}\boldsymbol{\xi}_k+\epsilon_{3ik}, \tag{14}$$

$k=1,\ldots,K$, where $\mathbf{X}_{3ik}$ is an $r_k\times 1$ vector with the first element being 1 and the remaining r_k-1 elements being covariates associated with U_{ik}. We further assume that $\epsilon_{3ik}\sim N(0,1)$, $k=1,\ldots,K$, and $(\epsilon_{1i},\epsilon_{2i},\epsilon_{3i1},\ldots,\epsilon_{3iK})^{\tau}$ follows a multivariate normal distribution with the covariance matrix $\boldsymbol{\Sigma}$. The diagonal elements of $\boldsymbol{\Sigma}$ are all equal to 1 and the off-diagonal elements of $\boldsymbol{\Sigma}$ are unknown. Let $\mathbf{X}_i=(\mathbf{X}_{1i}^{\tau},\mathbf{X}_{2i}^{\tau},\mathbf{X}_{3i1}^{\tau},\ldots,\mathbf{X}_{3iK}^{\tau})^{\tau}$.

We now derive the likelihood function. Let $\boldsymbol{\theta}$ be the vector of unknown parameters in models (1), (2), and (14). For ease of expression, we denote $R_i=D_{i0}$. When $D_{i0}=1$, we observe $(Y_i=y_i,D_{i0}=1,\mathbf{X}_i)$; when $D_{ik}=1$, we observe $(Y_i=y_i,D_{i0}=0,\ldots,D_{i,k-1}=0,D_{ik}=1,\mathbf{X}_i)$ for $k\le K$; when $D_{iK}=0$, we observe $(D_{i0}=0,\ldots,D_{iK}=0,\mathbf{X}_i)$. Therefore, the likelihood function of $\boldsymbol{\theta}$ is

$$\begin{aligned}L(\boldsymbol{\theta}) &= \prod_{i=1}^{n}\Big[\{P(Y_i=y_i,D_{i0}=1|\mathbf{X}_i)\}^{D_{i0}}\\ &\quad\times\prod_{k=1}^{K}\{P(Y_i=y_i,D_{i0}=0,\ldots,D_{i,k-1}=0,D_{ik}=1|\mathbf{X}_i)\}^{(1-D_{i0})\cdots(1-D_{i,k-1})D_{ik}}\\ &\quad\times\{P(D_{i0}=0,\ldots,D_{iK}=0|\mathbf{X}_i)\}^{(1-D_{i0})\cdots(1-D_{iK})}\Big].\end{aligned}$$

The first term in the likelihood is

$$\begin{aligned}P(D_{i0}=1,Y_i=y_i|\mathbf{X}_i) &= P(R_i=1|Y_i=y_i,\mathbf{X}_i)P(Y_i=y_i|\mathbf{X}_i)\\ &= \Phi\Big(\frac{\mathbf{X}_{2i}^{\tau}\boldsymbol{\gamma}+\rho_{12}\epsilon_{1i}}{\sqrt{1-\rho_{12}^2}}\Big)\sigma^{-1}\phi(\epsilon_{1i}),\end{aligned}$$

where $\epsilon_{1i} = (y_i - \beta_0 - \mathbf{X}_{1i}^\tau \boldsymbol{\beta}_1)/\sigma$.

The second term in the likelihood is

$$\begin{aligned} & P(Y_i = y_i, D_{i0} = 0, \ldots, D_{i,k-1} = 0, D_{ik} = 1 | \mathbf{X}_i) \\ = \; & P(D_{i0} = 0, \ldots, D_{i,k-1} = 0, D_{ik} = 1 | Y_i = y_i, \mathbf{X}_i) P(Y_i = y_i | \mathbf{X}_i) \\ = \; & P(\epsilon_{2i} < -\mathbf{X}_{2i}^\tau \boldsymbol{\gamma}, \epsilon_{3i1} < -\mathbf{X}_{3i1}^\tau \boldsymbol{\xi}_1, \ldots, \epsilon_{3i\,k-1} < -\mathbf{X}_{3i\,k-1}^\tau \boldsymbol{\xi}_{k-1}, \epsilon_{3ik} > -\mathbf{X}_{3ik}^\tau \boldsymbol{\xi}_k | Y_i = y_i, \mathbf{X}_i) \\ & \times P(Y_i = y_i | \mathbf{X}_i) \\ = \; & \int_{-\infty}^{-\mathbf{X}_{2i}^\tau \gamma} \int_{-\infty}^{-\mathbf{X}_{3i1}^\tau \xi_1} \cdots \int_{-\infty}^{-\mathbf{X}_{3i\,k-1}^\tau \xi_{k-1}} \int_{-\mathbf{X}_{3ik}^\tau \xi_k}^{\infty} \phi_{2,31,\ldots,3k|1}\big(t, u_1, \ldots, u_k | \epsilon_{1i}\big) dt du_1 \cdots du_k \\ & \times \sigma^{-1} \phi\Big(\frac{y_i - \beta_0 - \mathbf{X}_{1i}^\tau \boldsymbol{\beta}_1}{\sigma}\Big), \end{aligned}$$

where $\phi_{2,31,\ldots,3k|1}\big(t, u_1, \ldots, u_k | s)$ is the density function of $(\epsilon_{2i}, \epsilon_{3i1}, \ldots, \epsilon_{3ik})^\tau$ conditional on $\epsilon_{1i} = s$.

The third term in the likelihood is

$$\begin{aligned} & P(D_{i0} = 0, \ldots, D_{iK} = 0 | \mathbf{X}_i) \\ = \; & P(\epsilon_{2i} < -\mathbf{X}_{2i}^\tau \boldsymbol{\gamma}, \epsilon_{3i1} < -\mathbf{X}_{3i1}^\tau \boldsymbol{\xi}_1, \ldots, \epsilon_{3iK} < -\mathbf{X}_{3iK}^\tau \boldsymbol{\xi}_K | \mathbf{X}_i) \\ = \; & \int_{-\infty}^{-\mathbf{X}_{2i}^\tau \boldsymbol{\gamma}} \int_{-\infty}^{-\mathbf{X}_{3i1}^\tau \boldsymbol{\xi}_1} \cdots \int_{-\infty}^{-\mathbf{X}_{3iK}^\tau \boldsymbol{\xi}_K} \phi_{2,31,\ldots,3k}\big(t, u_1, \ldots, u_K\big) dt du_1 \cdots du_K, \end{aligned}$$

where $\phi_{2,31,\ldots,3k}\big(t, u_1, \ldots, u_K\big)$ is the density function for $(\epsilon_{2i}, \epsilon_{3i1}, \ldots, \epsilon_{3iK})^\tau$.

Let

$$\ell(\boldsymbol{\theta}) = \log L(\boldsymbol{\theta}) = \sum_{i=1}^{n} \ell_i(\boldsymbol{\theta}) \tag{15}$$

be the log-likelihood, where $l_i(\boldsymbol{\theta})$ is the log-likelihood contribution from individual i. Maximizing (15) with respect to $\boldsymbol{\theta}$, we obtain the maximum likelihood estimator, $\widehat{\boldsymbol{\theta}}$. Similarly, we can show that the maximum likelihood estimate $\widehat{\boldsymbol{\theta}}$ satisfies

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

in distribution as $n \to \infty$, where $\boldsymbol{J} = -E[\partial^2 \ell_i(\boldsymbol{\theta}_0)/\{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^\tau\}]$.

References

Miao, W., Ding, P., and Geng, Z. (2016). Identifiability of normal and normal mixture models with nonignorable missing data. *Journal of the American Statistical Association* 111, 1673–1683.