Semiparametric Estimating Equations Inference with Nonignorable Missing Data

Puying Zhao, Niansheng Tang, Annie Qu, and Depeng Jiang

Yunnan University, University of IIIinois at Urbana-Champaign, and University of Manitoba.

Supplementary Material

S1 Technical conditions

Let $f(X)$ be the probability density function of X and the symbol ∂_α represent the partial derivatives taken with respect to parameter α. Define $\pi(X,Y) = \Pr(\delta = 1|X,Y) =: \pi(X,Y,\alpha_0)$, $\pi(X) = \Pr(\delta = 1|X)$, $G(X) = f(X)\{1-\pi(X)\}$, $z(X,Y,\alpha) = \partial_\alpha \text{logit}\{\pi(X,Y,\alpha)\}$, $m^0_\psi(X,\theta,h) = E\{\psi(Y,X,\theta,h)|X,\delta = 0\}$, and $m^0_z(X,\alpha) = E(z(X,Y,\alpha)|X,\delta = 0)$. We equip the space $\mathcal{H}$ with the semi-norm $\|\cdot\|_{\mathcal{H}}$, defined by $\|\cdot\|_{\mathcal{H}} = \sup_{\theta\in\Theta}\|\cdot\|_{L_2(P)} = \sup_\theta\{\int|\cdot|^2dP\}^{1/2}$ with respect to all the θ-arguments. Also, we denote $N(\lambda,\mathcal{H},\|\cdot\|_{\mathcal{H}})$ as the covering number of the class $\mathcal{H}$ with respect to the norm $\|\cdot\|_{\mathcal{H}}$. Define $\Theta_\varrho =: \{\theta\in\Theta : \|\theta-\theta_0\| \le \varrho\}$, $\mathcal{B}_\varrho =: \{\alpha\in\mathcal{B} : \|\alpha-\alpha_0\|\le\varrho\}$ and $\mathcal{H}_\varrho =: \{h\in\mathcal{H} : \|h-h_0\|_{\mathcal{H}}\le\varrho\}$. Let $\widetilde{\mathcal{G}}_n(\theta,h,\alpha) = n^{-1}\sum_{i=1}^n\tilde{\psi}(Y_i,X_i,\theta,h,\alpha)$, where $\tilde{\psi}(Y_i,X_i,\theta,h,\alpha)$ is defined in (2.3), and $\widetilde{\mathcal{Q}}_n(\alpha,\theta,h) = n^{-1}\sum_{i=1}^n(1-\delta_i)\frac{r_i}{\nu}\{\psi(Y_i,X_i,\theta,h)-m^0_\psi(X_i,\theta,h,\alpha)\}$.

Let $\Lambda(\theta,h_0)$ be the partial derivative of $\mathcal{G}(\theta,h_0) = E\{\psi(Y,X,\theta,h_0)\}$ with respect to θ, that is,

$$\Lambda(\theta,h_0) = \frac{\partial}{\partial\theta}\mathcal{G}(\theta,h_0) = \lim_{\kappa\to0}\frac{1}{\kappa}\{\mathcal{G}(\theta+\kappa,h_{0,\theta+\kappa}) - \mathcal{G}(\theta,h_{0\theta})\}.$$

Let $\Gamma(\theta,h_0)(h-h_0)$ be the functional derivative of $\mathcal{G}(\theta,h_0)$ in the direction $h-h_0$, that is,

$$\Gamma(\theta,h_0)(h-h_0) = \lim_{\kappa\to0}\frac{1}{\kappa}\{\mathcal{G}(\theta,h_0+\kappa(h-h_0)) - \mathcal{G}(\theta,h_0)\}.$$

Assumptions for consistency

(A1) For all $\varrho > 0$, there exists $\epsilon > 0$ such that $\inf_{\|\theta-\theta_0\|>\varrho}\|\mathcal{G}(\theta,h_0)\|_W \ge \epsilon > 0$.

(A2) Uniformly for all $\theta \in \Theta$, $\mathcal{G}(\theta, h)$ is continuous (with respect to the metric $\|\cdot\|_{\mathcal{H}}$) in h at $h = h_0$.

(A3) (i) $\|\hat{h} - h_0\|_{\mathcal{H}} = o_p(1)$; (ii) Given estimator $\hat{\alpha}$ (including $\hat{\alpha} = \alpha_0$), for all positive sequences $\varrho_n = o(1)$,

$$\sup_{\theta\in\Theta,\|h-h_0\|_{\mathcal{H}}\le\varrho_n} \|\mathcal{G}_n(\theta, h, \hat{\alpha}) - \widetilde{\mathcal{G}}_n(\theta, h, \alpha_0)\| = o_p(1), \quad \text{and}$$
$$\sup_{\alpha\in\mathcal{B},\theta\in\Theta,\|h-h_0\|_{\mathcal{H}}\le\varrho_n} \|\mathcal{Q}_n(\alpha, \theta, h) - \widetilde{\mathcal{Q}}_n(\alpha, \theta, h)\| = o_p(1).$$

Assumptions for asymptotic normality

(B1) (i) The ordinary derivative $\Lambda(\theta, h_0)$ in θ of $\mathcal{G}(\theta, h_0)$ exists for θ in a neighborhood of θ_0, and is continuous at $\theta = \theta_0$; (ii) the matrix $\Lambda =: \Lambda(\theta_0, h_0)$ is of full rank.

(B2) Let $\psi(Y, X, \theta, h) = (\psi_1(Y, X, \theta, h), \ldots, \psi_q(Y, X, \theta, h))^\top$. $\psi_j(Y, X, \theta, h)$ $(j = 1, \ldots, q)$ is locally uniformly $L_2(P)$ continuous with respect to θ, h in the sense:

$$E\left\{\sup_{(\theta',h'):\|\theta'-\theta\|\le\varrho,\|h'-h\|_{\mathcal{H}}\le\varrho} |\psi_j(Y, X, \theta', h') - \psi_j(Y, X, \theta, h)|^2\right\} \le K\varrho^{2\varsigma}$$

for all $(\theta, h) \in \Theta \times \mathcal{H}$, all small positive values $\varrho = o_p(1)$, and for some constants $0 < K < \infty$ and $0 < \varsigma \le 1$.

(B3) For all $\theta \in \Theta_\varrho$, the pathwise derivative $\Gamma(\theta, h_0)(h - h_0)$ of $\mathcal{G}(\theta, h_0)$ exists in all directions $h - h_0 \in \mathcal{H}$; and for all $(\theta, h) \in \Theta_{\varrho_n} \times \mathcal{H}_{\varrho_n}$ with a positive sequence $\varrho_n = o_p(1)$: (i) $\|\mathcal{G}(\theta, h) - \mathcal{G}(\theta, h_0) - \Gamma(\theta, h_0)(h - h_0)\|_W \le c\|h - h_0\|^2_{\mathcal{H}}$ for some $0 < c < \infty$; (ii) $\|\Gamma(\theta, h_0)(h - h_0) - \Gamma(\theta_0, h_0)(h - h_0)\|_W = o(1)\varrho_n$, and

$$\int_0^\infty \sqrt{\log N(\lambda^{1/\varsigma}, \mathcal{H}, \|\cdot\|_{\mathcal{H}})}d\lambda < \infty,$$

where ς is defined in (B2).

(B4) $\Gamma(\theta_0, h_0)(\hat{h} - h_0) = n^{-1}\sum_{i=1}^n \nabla(X_i, Y_i, \delta_i) + o_p(n^{-1/2})$, where the function $\nabla(X, Y, \delta) = (\nabla_1(X, Y, \delta), \ldots, \nabla_q(X, Y, \delta))^\top$ satisfies $E\{\nabla_j(X, Y, \delta)\} = 0$ and $E\{\nabla_j(X, Y, \delta)\}^2 < \infty$ for $j = 1, \ldots, q$.

(B5) (i) $\hat{h} \in \mathcal{H}$ with probability tending to one, $\|\hat{h} - h_0\|_{\mathcal{H}} = o_p(n^{-1/4})$; ($ii$) Given estimator $\hat{\alpha}$ (including $\hat{\alpha} = \alpha_0$), for all positive values $\varrho_n = o(1)$, $\sup^* \|\mathcal{G}_n(\theta, h, \hat{\alpha}) - \widetilde{\mathcal{G}}_n(\theta, h, \alpha_0) - \mathcal{G}_n(\theta_0, h_0, \hat{\alpha}) + \widetilde{\mathcal{G}}_n(\theta_0, h_0, \alpha_0)\| = o_p(n^{-1/2})$, where $\sup^*$ is the supremum over all $\|\theta - \theta_0\| \leq \varrho_n$ and $\|h - h_0\|_{\mathcal{H}} \leq \varrho_n$.

Regularity assumptions

(C1) The probability density function $f(X)$ is bounded away from ∞ in the support of X and the second partial derivative of $f(X)$ is continuous and bounded.

(C2) (i) The true response probability follows a parametric model given in (2.1), i.e., $\pi(X, Y) = \pi(X, Y, \alpha_0)$; ($ii$) $\pi(X, Y, \alpha_0) \geq c_0 > 0$ a.s. for some positive constant c_0 and $\pi(X) = E\{\pi(X, Y, \alpha_0)|X\} \neq 1$ a.s.; (iii) for $\alpha \in \mathcal{B}_{\varrho}$, $E|\pi(X, Y, \alpha)|^3 < \infty$, and $\partial^2\pi(X, Y, \alpha)/\partial\alpha\partial\alpha^\top$ exists and is bounded by an integrable function.

(C3) For all $x \in \mathcal{X}$ and $\theta \in \Theta$, the function $z \to E\{\psi^l(Y, x, \theta, h_0)|X = x\}$ is uniformly continuous in x for $l = 1$ and 2, and $0 < E|\psi(Y, X, \theta, h_0)|^2 < \infty$. $m_\psi^0(X, \theta, h)$ is twice continuously differentiable in the neighborhood of X.

(C4) $na^s \to \infty$ and $na^{2m} \to 0$ as $n \to \infty$, where m is the order of the kernel function $K(\cdot)$.

Remark. The regularity assumptions (C1)-(C4) are commonly adopted in the missing data and nonparametric literatures. Conditions (A1)-(A3) and (B1)-(B5) are required for the proof of consistency and asymptotic normality of all proposed estimators. In fact, the proofs are totally based on Theorems 1 and 2 in Chen et al. (2003). In these theorems high-level conditions are given under which the proposed estimator is respectively, weakly consistent and asymptotically normal. See Chen et al. (2003) and Chen and Van Keilegom (2013) for discussion on each assumption. It is easily shown that conditions (1.1)-(1.5) and (2.1)-(2.6) in Chen et al. (2003) are valid by combining conditions (A1)-(A3), (B1)-(B5) and the following Lemmas 1-3.

S2 An example

It is difficult to understand what the really use of above high-level conditions in practical situations is. Under MAR setup, Chen and Van Keilegom

(2013) considered an example of a partial linear regression model to illustrate their proposed strategy. Here we try to extend their illustrations to MNAR framework, to verify the above listed conditions and to get a clear understanding of the effect of using $\hat{h}$ in the proposed approach.

Let $Y \in \mathcal{R}$ denote the outcome, $X = (X_1, X_2) \in \mathcal{R}^{d+1}$ denote the covariates, θ denote the d-dimensional vector of regression covariates. Suppose that Y and X are related by a partial linear regression model $Y = X_1^\top \theta + h(X_2) + \varepsilon$, where ε are random errors satisfying $E(\varepsilon|X) = 0$. Let $\{(X_i, Y_i, \delta_i) : i = 1, \ldots, n\}$ be independent and identically distributed realizations of (X, Y, δ), where δ_i is dichotomous taking values of 1 or 0, and Y_i is observed if and only if $\delta_i = 1$. Define estimating functions $\psi(Y, X, \theta, h) = X_1(Y - X_1^\top \theta - h(X_2))$. Let $F(y|X)$ be the conditional distribution of Y given X, let $\pi(X_i, Y_i) = \Pr(\delta_i = 1|X_i, Y_i) =: \pi(X_i, Y_i, \alpha_0)$ be the response probability model, and $O(X_i, Y_i, \alpha_0) = \pi^{-1}(X_i, Y_i, \alpha_0) - 1$. Let $K(\cdot)$ be a $(d+1)$-dimensional kernel function. From (2.2), a kernel estimator of $F(y|X)$ based on the sample $\{(X_i, Y_i, , \delta_i) : i = 1, \ldots, n\}$ is given by

$$\hat{F}_Y^0(y|X; \alpha_0) = \frac{\sum_{j=1}^n \delta_j O(X_j, Y_j, \alpha_0) I(Y_j \leq y) K_a(X - X_j)}{\sum_{j=1}^n \delta_j O(X_j, Y_j, \alpha_0) K_a(X - X_j)}, \tag{S2.1}$$

and a kernel estimator of conditional expectation $m_Y^0(X) = E(Y|X, \delta = 0)$ could be obtained by

$$\hat{m}_Y^0(X, \alpha_0) = \int y d\hat{F}_Y^0(y|X; \alpha_0) = \frac{\sum_{j=1}^n \delta_j O(X_j, Y_j, \alpha_0) K_a(X - X_j) Y_j}{\sum_{j=1}^n \delta_j O(X_j, Y_j, \alpha_0) K_a(X - X_j)}.$$

Let $\hat{Y}_i = \delta_i Y_i + (1-\delta_i)\hat{m}_Y^0(X_i, \alpha_0)$, $m_Y(X) = E(Y|X)$ and $\mathcal{X} = (X_1, \ldots, X_n)$. Using the standard kernel regression theory and under MNAR assumption, we have

$$\begin{aligned} E(\hat{Y}_i|\mathcal{X}) &= E(\delta_i Y_i|X_i) + E(1 - \delta_i|X_i)\frac{E\{\delta_i Y_i O(X_i, Y_i, \alpha_0)|X_i\}}{E\{\delta_i O(X_i, Y_i, \alpha_0)|X_i\}} + O_p(a^m) \\ &= E(\delta_i Y_i|X_i) + \Pr(\delta_i = 0|X_i)\frac{E\{(1 - \delta_i)Y_i|X_i\}}{E\{1 - \delta_i|X_i\}} + O_p(a^m) \\ &= E(Y_i|X_i) + O_p(a^m) = m_Y(X_i) + o_p(n^{-1/4}). \end{aligned}$$

Note that $m_Y^0(X) \neq m_Y(X)$ under MNAR assumption. Consequently, a kernel estimator of nonparametric function $h(x_2)$ is $\hat{h}_\theta(x_2) = \sum_{j=1}^n W_{nj}(x_2)(\hat{Y}_j - X_{1j}^\top\theta)$, where $W_{nj}(x_2) = \mathcal{K}_b(x_2 - X_{2j})/\sum_{j=1}^n \mathcal{K}_b(x_2 - X_{2j})$, $\mathcal{K}_b(u) = b^{-1}\mathcal{K}(u/b)$, $\mathcal{K}(\cdot)$ is an univariate kernel function and b is a bandwidth.

Let $h_0^\theta(x_2) = h_0(x_2) - E(X_1^\top|X_2 = x_2)(\theta - \theta_0) = E(Y - X_1^\top\theta|X_2 = x_2)$. Write $\hat{h}_\theta(x_2) - h_0^\theta(x_2) = (\hat{h}_\theta(x_2) - E\{\hat{h}_\theta(x_2)|\mathcal{X}\}) + (E\{\hat{h}_\theta(x_2)|\mathcal{X}\} - h_0^\theta(x_2)) = (T_1 + T_2)(x_2)$. Using the facts that $na^{4m} \to 0$, $nb^2(\log n)^{-2} \to \infty$, $na^{2(d+1)}(\log n)^{-2} \to \infty$ and $nb^8 \to 0$ as $n \to \infty$, we have

$$\begin{aligned} T_1(x_2) &= \sum_{j=1}^n W_{nj}(x_2)\delta_j\{Y_j - m_Y^0(X_j)\} + \sum_{j=1}^n W_{nj}(x_2)\{m_Y^0(X_j) - m_Y(X_j)\} \\ &\quad + \sum_{j=1}^n W_{nj}(x_2)(1-\delta_j)\{\hat{m}_Y^0(X_j,\alpha_0) - m_Y^0(X_j)\} + O_p(a^m) \\ &= O_p((nb)^{-1/2}(\log n)^{1/2}) + O_p((na^{d+1})^{-1/2}(\log n)^{1/2}) + O_p(a^m) \\ &= o_p(n^{-1/4}); \end{aligned}$$

$$\begin{aligned} T_2(x_2) &= \sum_{j=1}^n W_{nj}(x_2)\{m_Y(X_j) - X_{1j}^\top\theta\} - h_0^\theta(x_2) + O_p(a^m) \\ &= \sum_{j=1}^n W_{nj}(x_2)\{E(Y|X_j) - X_{1j}^\top\theta - E(Y|X_{2j}) + E(X_1|X_{2j})\theta^\top\} \\ &\quad + O_p(b^2) + O_p(a^m) \\ &= O_p((nb)^{-1/2}(\log n)^{1/2}) + O_p(a^m) \\ &= o_p(n^{-1/4}). \end{aligned}$$

Using the arguments of Chen and Van Keilegom (2013), we can verify conditions A3(i), B3(ii) and B5(i). Combining the above derivations with $na^{2m} \to 0$ and $nb^4 \to 0$ implies

$$\begin{aligned} \hat{h}_\theta(x_2) - h_0^\theta(x_2) &= \sum_{j=1}^n W_{nj}(x_2)[\delta_j(Y_j - m_Y^0(X_j)) + m_Y^0(X_j) - m_Y(X_j) \\ &\quad + (1-\delta_j)\{\hat{m}_Y^0(X_j,\alpha_0) - m_Y^0(X_j)\}] + o_p(n^{-1/2}). \end{aligned}$$

Let $f_{X_2}(x_2)$ be the probability density function of X_2. Simple algebraic manipulations show that, uniformly in j

$$E\left\{\frac{X_1}{f_{X_2}(x_2)}\mathcal{K}_b(X_2 - X_{2j})\right\} = E(X_1|X_2 = X_{2j}) + O_p(b^2).$$

Now, replacing $W_{nj}(x_2)$ by $(nb)^{-1}\mathcal{K}(\frac{x_2-X_{2j}}{b})/f_{X_2}(x_2)$ leads to

$$\begin{aligned}\Gamma(\theta_0,h_0)(\hat{h}-h_0) &= -E\{(\hat{h}_\theta(X_2)-h_0^\theta(X_2))X_1\}\\ &= -n^{-1}\sum_{j=1}^n E(X_1|X_{2j})[\delta_j\{Y_j-m_Y^0(X_j,\alpha_0)\}+m_Y^0(X_j)\\ &\quad -m_Y(X_j)]-n^{-1}\sum_{j=1}^n E(X_1|X_{2j})[(1-\delta_j)\{\hat{m}_Y^0(X_j,\alpha_0)\\ &\quad -m_Y^0(X_j)\}]+o_p(n^{-1/2}).\end{aligned}$$

Introducing a U statistic and using the same arguments as I_{n3} in the proof of Lemma 1, we have

$$\begin{aligned}&n^{-1}\sum_{j=1}^n E(X_1|X_{2j})[(1-\delta_j)\{\hat{m}_Y^0(X_j,\alpha_0)-m_Y^0(X_j)\}]\\ &\quad = n^{-1}\sum_{j=1}^n \frac{1-\pi(X_j,Y_j)}{\pi(X_j,Y_j)}E(X_1|X_{2j})\delta_j\{Y_j-m_Y^0(X_j)\}+o_p(n^{-1/2}),\end{aligned}$$

which leads to

$$\begin{aligned}\nabla(X_i,Y_i,\delta_i) &= -\frac{\delta_i}{\pi(X_i,Y_i)}E(X_1|X_{2i})\{Y_i-m_Y^0(X_i)\}\\ &\quad -E(X_1|X_{2i})\{m_Y^0(X_i)-m_Y(X_i)\}\\ &= -E(X_1|X_{2i})\Bigg[\frac{\delta_i}{\pi(X_i,Y_i)}\{Y_i-m_Y^0(X_i)\}\\ &\quad +\{m_Y^0(X_i)-m_Y(X_i)\}\Bigg].\end{aligned}$$

Thus, we verify (B4). Some simple algebraic manipulations show

$$\Lambda = \frac{\partial}{\partial\theta}E\{\psi(Y,X,\theta,h_0)\} = -E\{(X_1-E(X_1|X_2))^\top X_1\} = -E\{\mathrm{Var}(X_1|X_2)\}.$$

With the true response model, we define a nonparametric estimator $\hat{\theta}_{NP}$ of θ, which is the solution to $\mathcal{G}_n(\theta,\hat{h}_\theta,\alpha_0)=0$, where

$$\mathcal{G}_n(\theta,\hat{h}_\theta,\alpha) = \frac{1}{n}\sum_{i=1}^n\{\delta_i\psi(Y_i,X_i,\theta,\hat{h}_\theta(X_{2i}))+(1-\delta_i)\hat{m}_\psi^0(X_i,\theta,\hat{h}_\theta,\alpha)\} \tag{S2.2}$$

in which $\hat{m}_\psi^0(X_i,\theta,\hat{h}_\theta,\alpha)$ is defined in equation (2.4). Using the results of Theorem 1, the asymptotic distribution of $\hat{\theta}_{NP}$ is given by

$$n^{1/2}(\hat{\theta}_{NP}-\theta_0)\xrightarrow{\mathcal{L}}\mathcal{N}(0,\Sigma_1),$$

where $\Sigma_1 = \Lambda^{-1}\Gamma_1\Lambda^{-\top}$, $\Gamma_1 = \text{Var}[\delta\{\pi(X,Y)\}^{-1}\{\psi(Y,X,\theta_0,h_0) - m_\psi^0(X,\theta_0, h_0)\} + m_\psi^0(X,\theta_0,h_0) + \nabla(X,Y,\delta)]$. Also, if $m_\psi^0(X_i,\theta_0,h_0) = E\{\psi(Y_i,X_i,\theta_0, h_0)|X_i,\delta_i = 0\} = E\{\psi(Y_i,X_i,\theta_0,h_0)|X_i\}$ and $m_Y^0(X_i) = E(Y|X_i,\delta_i = 0) = E(Y|X_i)$ are true, thus we have

$$\Sigma_1 = \Lambda^{-1}\text{Var}\left[\frac{\delta}{\pi(X,Y)}\{X_1 - E(X_1|X_2)\}\{Y - X_1^\top\theta_0 - h_0(X_2)\}\right]\Lambda^{-\top}.$$

The equality is true because $E\{\psi(Y,X,\theta_0,h_0)|X\} = 0$ and $E(Y|X) = X_1^\top\theta_0 - h_0(X_2)$.

Thus, our presented methods and theories in the paper are fully illustrated for a special case. Using the same arguments as those used in the Chen and Van Keilegom (2013), we can further apply our proposed approach to a single index regression model. To save space, we here omit the details.

S3 Robust estimation via validation sample

To study the robustness of our proposed *Propensity-Score-Based Nonparametric Imputation* procedure, following Kim and Yu (2011), we first consider a simpler setup for estimating θ_0 defined via estimating equations of the form $\psi(Y,X,\theta_0,h_0) = Y - \theta_0$, that is, consider the estimation of the population mean $\theta_0 = E(Y)$ under MNAR assumption. Define $\hat\theta_M = n^{-1}\sum_{i=1}^n\{\delta_iY_i + (1-\delta_i)\hat m_Y^0(X_i,\hat\alpha_{\text{v}})\}$, where $\hat m_Y^0(X,\alpha) = \sum_{i=1}^n \omega_{i0}(X,\alpha_0)Y_i$ with $\omega_{i0}(X,\alpha_0) = \delta_i O(X_i,Y_i,\alpha)K_a(X - X_i)/\{\sum_{k=1}^n \delta_k O(X_k,Y_k,\alpha)K_a(X - X_k)\}$. Here $\hat\alpha_{\text{v}}$ satisfies $\sum_{i=1}^n Q_1(X_i,Y_i,\hat\alpha_{\text{v}}) = \sum_{i=1}^n(1-\delta_i)r_i\{Y_i - \hat m_Y^0(X_i,\hat\alpha_{\text{v}})\} = 0$, where r_i is an indicator, which takes 1 if individual i belongs to the follow-up sample and 0 otherwise. In the following proposition, we study the asymptotic properties of $\hat\theta_M$ in the presence of possible model misspecification.

Proposition S1. *Suppose that the assumptions of Theorem 1 hold, except that the parametric response model (2.1) is misspecified. Assume that the estimators* $\hat\alpha_{\text{v}}$ *satisfies* $\sum_{i=1}^n(1-\delta_i)r_i\{Y_i - \hat m_Y^0(X_i,\hat\alpha_{\text{v}})\} = 0$. *Then, we have* $n^{1/2}(\hat\theta_M - \theta_0) \xrightarrow{\mathcal{L}} \mathcal{N}(0,\sigma_1^2)$, *where* $\sigma_1^2 = \text{Var}(\eta_{1i})$ *in which* $\eta_{1i} = m_Y^0(X_i,\alpha^*) + \{\frac{r_i}{\nu}(1-\delta_i)+\delta_i\}\{Y_i - m_Y^0(X_i,\alpha^*)\}$, $m_Y^0(X_i,\alpha^*) = E\{\delta Y O(X,Y,\alpha^*)|X\}/E\{\delta O(X, Y,\alpha^*)|X\}$, $\nu = E(r|\delta = 0)$, *and* α^* *is the probability limit of* $\hat\alpha_{\text{v}}$.

Next, we consider another simpler setup of estimating the distribution function (i.e., $F(y) = P(Y \le y)$) of the response variable Y. Define $\hat F_Y^0(y|X_i,\alpha) = \sum_{i=1}^n \omega_{i0}(X,\alpha)I(Y_j \le y)$. Then, using our proposed imputation approach, a semiparametric estimator of $F(y)$ under MNAR can be obtained by $\hat{F}_{SP}(y) = n^{-1}\sum_{i=1}^{n}\{\delta_i I(Y_i \le y) + (1-\delta_i)\hat{F}_Y^0(y|X_i, \hat{\alpha}_{\rm v})\}$ in which $\hat{\alpha}_{\rm v}$ satisfies $\sum_{i=1}^{n} Q_2(X_i, Y_i,\ \hat{\alpha}_{\rm v}) = \sum_{i=1}^{n}(1-\delta_i)r_i\{I(Y_i \le y) - \hat{F}_Y^0(y|X_i, \hat{\alpha}_{\rm v})\} = 0$. The following proposition presents some asymptotic properties of the estimator $\hat{F}_{SP}(y)$ in the presence of possible model misspecification.

Proposition S2. *Suppose that the assumptions of Theorem 1 hold, except that the parametric response model (2.1) is misspecified. Assume that the estimators* $\hat{\alpha}_{\rm v}$ *satisfies* $\sum_{i=1}^{n}(1-\delta_i)r_i\{I(Y_i \le y) - \hat{F}_Y^0(y|X_i, \hat{\alpha}_{\rm v})\} = 0$. *For any given* y, *we have* $n^{1/2}\{\hat{F}_{SP}(y) - F(y)\} \xrightarrow{\mathcal{L}} \mathcal{N}(0, \sigma_2^2(y))$, *where* $\sigma_2^2(y) = \mathrm{Var}(\eta_{2i})$, $\eta_{2i} = F_Y^0(y|X_i, \alpha^*) + \{r_i(1-\delta_i)/\nu + \delta_i\}\{I(Y_i \le y) - F_Y^0(y|X_i, \alpha^*)\}$,

$$F_Y^0(y|X_i, \alpha^*) = \frac{E\{\delta I(Y \le y)O(X, Y, \alpha^*)|X\}}{E\{\delta O(X, Y, \alpha^*)|X\}},$$

$\nu = E(r|\delta = 0)$ *and* α^* *is the probability limit of* $\hat{\alpha}_{\rm v}$.

Following the same arguments as discussed in Theorem 3 in Kim and Yu (2011), we can obtain that the validity of estimators $\hat{\theta}_M$ and $\hat{F}_{SP}(y)$ does not depend on the posited parametric response model, and the role of response model (2.1) is to improve the efficiency. However, in evaluating $\hat{\theta}_M$ and $\hat{F}_{SP}(y)$, estimating parameter α in a parametric response model may involve solving some under-identified estimating equations $Q_1(X_i, Y_i, \alpha)$ and $Q_2(X_i, Y_i, \alpha)$ with respect to α, which can lead to infinitely many solutions. To address the issue, we propose estimating response model by solving equations $\sum_{i=1}^{n}\Phi_j(X_i, Y_i, \alpha) = 0$, where $\Phi_j(X_i, Y_i, \alpha) = (Q_j(X_i, Y_i, \alpha), \widetilde{Q}_j^\top(X_i, Y_i, \alpha))^\top$ for $j = 1$ and 2, $\widetilde{Q}_1(X_i, Y_i, \alpha) = (1-\delta_i)r_i\zeta(X_i)\{Y_i - \hat{m}_Y^0(X_i, \hat{\alpha}_{\rm v})\}$, $\widetilde{Q}_2(X_i, Y_i, \alpha) = (1-\delta_i)r_i\zeta(X_i)\{I(Y_i \le y) - \hat{F}_Y^0(y|X_i, \hat{\alpha}_{\rm v})\}$, and $\zeta(\cdot)$ is an arbitrary user-specified vector function such that Φ_j $(j = 1, 2)$ has the same dimension as α.

Proof of proposition S1. Define $\hat{\theta}_M(\alpha) = n^{-1}\sum_{i=1}^{n}\{\delta_i Y_i + (1-\delta_i)\hat{m}_Y^0(X_i, \alpha)\}$ $+n^{-1}\sum_{i=1}^{n}(1-\delta_i)\frac{r_i}{\nu}\{Y_i - \hat{m}_Y^0(X_i, \alpha)\}$. It is easy to show that $\hat{\theta}_M(\hat{\alpha}_{\rm v}) = \hat{\theta}_M$, and

$$E\left\{\frac{\partial}{\partial\alpha}\hat{\theta}_M(\alpha)|\alpha = \alpha^*\right\} = 0.$$

This, together with the arguments of Randles (1982) and the fact $n^{1/2}(\hat{\alpha}_{\rm v} - \alpha^*) = o_p(1)$, implies that

$$\hat{\theta}_M(\hat{\alpha}_{\rm v}) = \hat{\theta}_M(\alpha^*) + o_p(n^{-1/2}).$$

For $\hat{\theta}_M(\alpha^*)$, we have the following decomposition

$$\begin{aligned}\hat{\theta}_M(\alpha^*) &= \frac{1}{n}\sum_{i=1}^{n}\{\delta_i Y_i + (1-\delta_i)\, m_Y^0(X_i,\alpha^*)\} \\ &+\frac{1}{n}\sum_{i=1}^{n}(1-\delta_i)\frac{r_i}{\nu}\{Y_i - m_Y^0(X_i,\alpha^*)\} \\ &+\frac{1}{n}\sum_{i=1}^{n}(1-\delta_i)(1-r_i\nu^{-1})\{\hat{m}_Y^0(X_i,\alpha^*) - m_Y^0(X_i,\alpha^*)\} \\ &=: V_{n1}+V_{n2}+V_{n3}.\end{aligned}$$

Applying the same argument given in Lemma 1 and using the fact that $\nu = E(r|\delta=0,X)$, we have $V_{n3} = o_p(n^{-1/2})$. Then, the proof of Proposition S1 is completed. □

Proof of Proposition S2. Denote

$$\begin{aligned}\hat{F}_{SP}(y,\alpha) &= \frac{1}{n}\sum_{i=1}^{n}\{\delta_i I(Y_i\le y) + (1-\delta_i)\hat{F}_Y^0(y|X_i,\alpha)\} \\ &+\frac{1}{n}\sum_{i=1}^{n}(1-\delta_i)\tfrac{r_i}{\nu}\{I(Y_i\le y) - \hat{F}_Y^0(y|X_i,\alpha)\}.\end{aligned}$$

Some simple algebraic manipulations show

$$\hat{F}_{SP}(y,\hat{\alpha}_{\mathrm{v}}) = \hat{F}_{SP}(y), \text{ and } E\left\{\frac{\partial}{\partial\alpha}\hat{F}_{SP}(y,\alpha)|\alpha=\alpha^*\right\} = 0,$$

where α^* is the probability limit of $\hat{\alpha}_{\mathrm{v}}$. According to Kim and Yu (2011) and Randles (1982), using $n^{1/2}(\hat{\alpha}_{\mathrm{v}} - \alpha^*) = O_p(1)$, we have $\hat{F}_{SP}(y,\hat{\alpha}_{\mathrm{v}}) = \hat{F}_{SP}(y,\alpha^*) + o_p(n^{-1/2})$. For $\hat{F}_{SP}(y,\alpha^*)$, we have

$$\begin{aligned}\hat{F}_{SP}(y,\alpha^*) &= \frac{1}{n}\sum_{i=1}^{n}\{\delta_i I(Y_i\le y) + (1-\delta_i)\, F_0(y|X_i,\alpha^*)\} \\ &+\frac{1}{n}\sum_{i=1}^{n}(1-\delta_i)\frac{r_i}{\nu}\{I(Y_i\le y) - F_Y^0(y|X_i,\alpha^*)\} \\ &+\frac{1}{n}\sum_{i=1}^{n}(1-\delta_i)(1-r_i/\nu)\{\hat{F}_Y^0(y|X_i,\alpha^*) - F_Y^0(y|X_i,\alpha^*)\} \\ &=: V_{n1}(y)+V_{n2}(y)+V_{n3}(y).\end{aligned}$$

According the same argument given in Lemma 1 and using the fact that $\nu = E(r|\delta=0,X)$, we have that $n^{1/2}V_{n3}(y) = o_p(1)$. Hence, the proof of Proposition S2 is completed. □

S4 Lemmas

Lemma 1. *Assume that the regularity conditions (C1)-(C4) hold. Then, we have*

$$\begin{aligned}\mathcal{G}_n(\theta_0, h_0, \alpha_0) &= \frac{1}{n}\sum_{i=1}^{n}\bigg[\frac{\delta_i}{\pi(X_i, Y_i)}\{\psi(Y_i, X_i, \theta_0, h_0) - m_\psi^0(X_i, \theta_0, h_0)\} \\ &\quad + m_\psi^0(X_i, \theta_0, h_0)\bigg] + o_p(n^{-1/2}).\end{aligned}$$

Proof. By the definition of $\mathcal{G}_n(\theta_0, h_0, \alpha_0)$, we have

$$\begin{aligned}\mathcal{G}_n(\theta_0, h_0, \alpha_0) &= \frac{1}{n}\sum_{i=1}^{n}\delta_i\{\psi(Y_i, X_i, \theta_0, h_0) - m_\psi^0(X_i, \theta_0, h_0)\} \\ &\quad + \frac{1}{n}\sum_{i=1}^{n} m_\psi^0(X_i, \theta_0, h_0) \\ &\quad + \frac{1}{n}\sum_{i=1}^{n}(1-\delta_i)\{\hat{m}_\psi^0(X_i, \theta_0, h_0, \alpha_0) - m_\psi^0(X_i, \theta_0, h_0)\} \\ &:= I_{n1} + I_{n2} + I_{n3}.\end{aligned}$$

Define $\pi(X) = E(\delta|X)$ and $G(X) = f(X)(1-\pi(X))$. From $E\{\delta O(X, Y, \alpha_0)|X\} = 1-\pi(X)$, using the kernel regression method yields $\hat{G}(X) = n^{-1}\sum_{j=1}^{n}\mathcal{D}_j(X)$, where $\mathcal{D}_j(X) = \delta_j O(X_j, Y_j, \alpha_0)K_a(X_j - X)$. Then, for I_{n3}, we have

$$\begin{aligned}I_{n3} &= \frac{1}{n}\sum_{i=1}^{n}(1-\delta_i)\frac{n^{-1}\sum_{j=1}^{n}\mathcal{D}_j(X_i)\{\psi(Y_j, X_j, \theta_0, h_0) - m_\psi^0(X_j, \theta_0, h_0)\}}{G(X_i)} \\ &\quad + \frac{1}{n}\sum_{i=1}^{n}(1-\delta_i)\{\hat{m}_\psi(X_i, \theta_0, h_0) - m_\psi^0(X_i, \theta_0, h_0)\}\{1 - \hat{G}(X_i)/G(X_i)\} \\ &\quad + \frac{1}{n}\sum_{i=1}^{n}(1-\delta_i)\frac{n^{-1}\sum_{j=1}^{n}\mathcal{D}_j(X_i)\{m_\psi^0(X_j, \theta_0, h_0) - m_\psi^0(X_i, \theta_0, h_0)\}}{G(X_i)} \\ &:= I_{n31} + I_{n32} + I_{n33}.\end{aligned}$$

We first investigate asymptotic properties of I_{n31} and I_{n33}.

Let $S_j = \{X_j, Y_j, \delta_j\}$, $\mathcal{B}_\psi(X_j, X_i) = \delta_j O(X_j, Y_j, \alpha_0)\{m_\psi^0(X_j, \theta_0, h_0) - m_\psi^0(X_i,\ \theta_0, h_0)\}$, $\mathcal{W}_j = \delta_j O(X_j, Y_j, \alpha_0)\{\psi(Y_j, X_j, \theta_0, h_0) - m_\psi^0(X_j, \theta_0, h_0)\}$, $\mathcal{B}_n(X_i) = n^{-1}\sum_{j=1}^{n}K_a(X_j - X_i)\mathcal{B}_\psi(X_j, X_i)$, and $\varphi_n(X_i) = n^{-1}\sum_{j=1}^{n}K_a(X_j - X_i)\mathcal{W}_j$. Define a kernel function of the U statistic for all pair (i, j):

$$H(S_i, S_j) = \frac{1}{2}K_a(X_i - X_j)\{(1-\delta_j)\mathcal{W}_i/G(X_j) + (1-\delta_i)\mathcal{W}_j/G(X_i)\}.$$

It is easy to show that $E(\mathcal{W}_j) = 0$. Therefore, we have

$$\begin{aligned}
&E\left\{K_a(X_i - X_j)\frac{(1-\delta_i)\mathcal{W}_j}{G(X_i)}\right\} \\
&= E\left\{\frac{1-\delta_i}{G(X_i)}E\{K_a(X_i - X_j)\mathcal{W}_j|S_i\}\right\} \\
&= E\left\{\frac{1-\delta_i}{G(X_i)}E\{K_a(X_i - X_j)\mathcal{W}_j|X_i\}\right\} \\
&= E\left\{\frac{1-\delta_i}{G(X_i)}E\{K_a(X_i - X_j)E(\mathcal{W}_j|X_i)|X_i, X_j\}\right\} \\
&= 0.
\end{aligned}$$

According to the symmetry of U statistic $H(S_i, S_j)$, we have $E\{H(S_i, S_j)\} = 0$.

On the other hand, we have

$$\begin{aligned}
E\{H(S_i, S_j)|S_j\} &= \frac{1-\delta_j}{2G(X_j)}E\left\{K_a(X_i - X_j)\mathcal{W}_i|S_j\right\} \\
&\quad + \frac{\mathcal{W}_j}{2}E\left\{K_a(X_i - X_j)\frac{1-\delta_i}{G(X_i)}|S_j\right\} \\
&:= J_1 + J_2.
\end{aligned}$$

From $E(\mathcal{W}_j|X_j) = 0$, we have

$$\begin{aligned}
J_1 &= \frac{1-\delta_j}{2g(X_j)}E\{K_a(X_i - X_j)\mathcal{W}_i|X_j\} \\
&= \frac{1-\delta_j}{2g(X_j)}E\{K_a(X_i - X_j)E(\mathcal{W}_j|X_i)|X_i, X_j\} = 0.
\end{aligned}$$

For J_2, we have

$$\begin{aligned}
J_2 &= \frac{\mathcal{W}_j}{2}E\{K_a(X_i - X_j)\frac{1-\pi(X_i)}{G(X_i)}|X_j\} \\
&= \frac{\mathcal{W}_j}{2}\int K_a(x - X_j)dx = \mathcal{W}_j/2 + O_p(a^2) \\
&= \frac{1-\pi(X_j, Y_j)}{2\pi(X_j, Y_j)}\delta_j\{\psi(Y_j, X_j, \theta_0, h_0) - m_\psi^0(X_j, \theta_0, h_0)\} + O_p(a^2).
\end{aligned}$$

Combining the above equations yields

$$\begin{aligned}
H_1(S_j) &:= E\{H(S_i, S_j|S_j)\} \\
&= \frac{\{1-\pi(X_j, Y_j)\}\delta_j\{\psi(Y_j, X_j, \theta_0, h_0) - m_\psi^0(X_j, \theta_0)\}}{2\pi(X_j, Y_j)}\{1 + O(a^2)\}.
\end{aligned}$$

Now, we show that

$$\frac{1}{n}\sum_{i=1}^{n}(1-\delta_i)\varphi_n(X_i)/G(X_i) = \frac{2}{n}\sum_{i=1}^{n} H_1(S_i), \tag{S4.3}$$

$$\frac{1}{n}\sum_{i=1}^{n}(1-\delta_i)\mathcal{B}_n(X_i)/G(X_i) = O(a^2). \tag{S4.4}$$

To save space, we only prove (S4.3). (S4.4) can be similarly shown.

It can be shown that

$$n^{-1}\sum_{i=1}^{n}(1-\delta_i)\varphi_n(X_i)/G(X_i) = n^{-2}\sum_{i=1}^{n} H(S_i,S_i) + U_n,$$

where $U_n = 2n^{-2}\sum_{i=1}^{n}\sum_{i<j}^{n} H(S_i,S_j)$. By the definition of $H(S_i,S_i)$, it is easy to obtain that $E\{H(S_i,S_i)\} = 0$. By the law of large numbers, we obtain that $n^{-2}\sum_{i=1}^{n} H(S_i,S_i)$ approximates $o(n^{-1})$ in probability, which indicates that it suffices to only consider statistic U_n in statistic $n^{-1}\sum_{i=1}^{n}(1-\delta_i)\varphi_n(X_i)/G(X_i)$. It is easy to show that $\zeta_1 := \mathrm{Var}\{H(S_j)\} = E[\{4\pi(X_i,Y_i)\}^{-1}\{1-\pi(X_i,Y_i)\}\{\psi(Y_i,\, X_i,\theta_0,h_0) - m_\psi^0(X_i,\theta_0,h_0)\}^{\otimes 2}](1+O(a^2))$. Similarly, we have

$$\begin{aligned}
\zeta_2 &:= \mathrm{Var}\{H(S_i,S_j)\} \\
&= \frac{1}{4}E\left\{K_a^2(X_i-X_j)\frac{(1-\delta_j)\mathcal{W}_i^{\otimes 2}}{G^2(X_j)}\right\} + \frac{1}{4}E\left\{K_a^2(X_i-X_j)\frac{(1-\delta_i)\mathcal{W}_j^{\otimes 2}}{G^2(X_i)}\right\} \\
&:= K_1 + K_2.
\end{aligned}$$

For K_1, we have

$$\begin{aligned}
K_1 &= \frac{1}{4}E\left\{E\{K_a^2(X_i-X_j)\frac{(1-\delta_j)\mathcal{W}_i^{\otimes 2}}{G^2(X_j)}|X_i,Y_i,\delta_i\}\right\} \\
&= \frac{1}{4}E\left\{\mathcal{W}_i^{\otimes 2}E\{E(K_a^2(X_i-X_j)\frac{(1-\delta_j)}{G^2(X_j)}|X_i,X_j,Y_i)|X_i,Y_i\}\right\} \\
&= \frac{1}{4}E\left\{\mathcal{W}_i^{\otimes 2}E\{\frac{K_a^2(X_i-X_j)}{G^2(X_j)}(1-\pi(X_j))|X_i\}\right\} \\
&= \frac{1}{4a}E\left\{\mathcal{W}_i^{\otimes 2}\frac{[1-\pi(X_i)]f(X_i)}{G^2(X_i)}\int K^2(u)du\right\} + o_p(1) \\
&= \frac{1}{4a}E\left\{\frac{(1-\pi(X_i,Y_i))^2(\psi(Y_i,X_i,\theta_0,h_0) - m_\psi^0(X,\theta_0,h_0))^{\otimes 2}}{\pi(X_i,Y_i)(1-\pi(X))f(X)}\int K^2(u)du\right\} + o_p(1).
\end{aligned}$$

Similarly, for K_2, it can be shown that

$$K_2 = \frac{1}{4a} E\left\{ \frac{(1-\pi(X,Y))^2(\psi(Y_i,X_i,\theta_0,h_0) - m_\psi^0(X,\theta_0,h_0))^{\otimes 2}}{\pi(X,Y)(1-\pi(X))f(X)} \int K^2(u)du \right\} + o_p(1).$$

Combining the above two equations yields

$$\zeta_2 = \frac{1}{2a} E\left\{ \frac{[1-\pi(X,Y)]^2[\psi(Y_i,X_i,\theta_0,h_0) - m_\psi^0(X,\theta_0,h_0)]^{\otimes 2}}{\pi(X,Y)[1-\pi(X)]f(X)} \int K^2(u)du \right\} + o_p(1).$$

Define $\hat{U}_n = \frac{2}{n}\sum_{i=1}^n H_1(S_i)$ from which it can be shown that $E\hat{U}_n^2 = 4\zeta_1/n$. Also, by the definition of U_n, we have $E(U_n^2) = 4(n-2)\zeta_1/\{n(n-1)\} + 2\zeta_2/\{n(n-1)\}$. Combining the above equations yields $E(U_n - \hat{U}_n)^2 = 2\zeta_2/\{n(n-1)\} + O(n^{-2})$. Thus, we obtain

$$\begin{aligned} U_n &= \hat{U}_n + \left\{ \frac{2\zeta_2}{n(n-1)} + O(n^{-2}) \right\}^{\frac{1}{2}} \\ &= \frac{1}{n}\sum_{i=1}^n \frac{\delta_i(\psi(Y_i,X_i,\theta_0,h_0) - m_\psi^0(X_i,\theta_0,h_0))(1-\pi(X_i,Y_i))}{\pi(X_i,Y_i)}\{1+O(a^2)\} \\ &\quad + O((n^2a)^{-\frac{1}{2}}). \end{aligned}$$

This completes the proof of (S4.3). Using the same arguments, we can prove (S4.4).

It follows from (S4.3) and (S4.4) that $I_{n33} = o_p(n^{-1/2})$ and

$$I_{n31} = \frac{1}{n}\sum_{i=1}^n \frac{(1-\pi(X_i,Y_i))\delta_i(\psi(Y_i,X_i,\theta_0,h_0) - m_\psi^0(X_i,\theta_0,h_0))}{\pi(X_i,Y_i)} + o_p(n^{-1/2}).$$

By a standard argument, we have $I_{32} = o_p(n^{-1/2})$. Combining the above results leads to

$$\begin{aligned} \mathcal{G}_n(\theta_0,h_0,\alpha_0) &= \frac{1}{n}\sum_{i=1}^n \delta_i\{\psi(Y_i,X_i,\theta_0,h_0) - m_\psi^0(X_i,\theta_0,h_0)\} \\ &\quad + \frac{1}{n}\sum_{i=1}^n \frac{(1-\pi(X_i,Y_i))\delta_i(\psi(Y_i,X_i,\theta_0,h_0) - m_\psi^0(X_i,\theta_0,h_0))}{\pi(X_i,Y_i)} \\ &\quad + \frac{1}{n}\sum_{i=1}^n m_\psi^0(X_i,\theta_0,h_0) + o_p(n^{-1/2}) \\ &= \frac{1}{n}\sum_{i=1}^n \frac{\delta_i}{\pi(X_i,Y_i)}\{\psi(Y_i,X_i,\theta_0,h_0) - m_\psi^0(X_i,\theta_0,h_0)\} \\ &\quad + \frac{1}{n}\sum_{i=1}^n m_\psi^0(X_i,\theta_0,h_0) + o_p(n^{-1/2}). \end{aligned}$$

Therefore, we have

$$\begin{aligned}\mathcal{G}_n(\theta_0, h_0, \alpha_0) &= \frac{1}{n}\sum_{i=1}^{n}\frac{\delta_i}{\pi(X_i, Y_i)}\{\psi(Y_i, X_i, \theta_0, h_0) - m_\psi^0(X_i, \theta_0, h_0)\}\\ &\quad + \frac{1}{n}\sum_{i=1}^{n} m_\psi^0(X_i, \theta_0, h_0) + o_p(n^{-1/2}).\end{aligned}$$

This completes the proof. □

Lemma 2. *Assume that the regularity conditions (C1)-(C4) hold. Then, for any estimator $\hat{\alpha}$ of α, we have*

$$\mathcal{G}_n(\theta_0, h_0, \hat{\alpha}) = \mathcal{G}_n(\theta_0, h_0, \alpha_0) - H \times (\hat{\alpha} - \alpha_0) + o_p(n^{-1/2}),$$

where $H = E[(1-\delta)\{\psi(Y, X, \theta_0, h_0) - m_\psi^0(X, \theta_0, h_0, \alpha_0)\}\{z(X, Y, \alpha_0) - m_z^0(X,\ \alpha_0)\}^\top]$ with $z(X, Y, \alpha) = \partial \mathrm{logit}\{\pi(X, Y, \alpha)\}/\partial\alpha$ and $m_z^0(X, \alpha) = E\{z(X, Y, \alpha)|X,\ \delta = 0\}$, and $\mathrm{logit}(p) = \log\{p/(1-p)\}$.

Proof. By the definition of $\mathcal{G}_n(\theta_0, h_0, \hat{\alpha})$, we have the following decomposition

$$\begin{aligned}&\mathcal{G}_n(\theta_0, h_0, \hat{\alpha})\\ &\quad = \frac{1}{n}\sum_{i=1}^{n}\delta_i\{\psi(Y_i, X_i, \theta_0, h_0) - m_\psi^0(X_i, \theta_0, h_0)\} + \frac{1}{n}\sum_{i=1}^{n} m_\psi^0(X_i, \theta_0, h_0)\\ &\qquad + \frac{1}{n}\sum_{i=1}^{n}(1-\delta_i)\{\hat{m}_\psi^0(X_i, \theta_0, h_0, \alpha_0) - m_\psi^0(X_i, \theta_0, h_0)\}\\ &\qquad + \frac{1}{n}\sum_{i=1}^{n}(1-\delta_i)\{\hat{m}_\psi^0(X_i, \theta_0, h_0, \hat{\alpha}) - \hat{m}_\psi^0(X_i, \theta_0, h_0, \alpha_0)\}\\ &\quad := I_{n1} + I_{n2} + I_{n3} + I_{n4},\end{aligned}$$

where I_{n1}, I_{n2}, I_{n3} are defined in Lemma 1. We now consider I_{n4}. By the first-order Taylor expansion of $\hat{m}_\psi^0(X_i, \theta_0, h_0, \hat{\alpha})$ at α_0, we have

$$I_{n4} = \frac{1}{n}\sum_{i=1}^{n}(1-\delta_i)\frac{\partial \hat{m}_\psi^0(X_i, \theta_0, h_0, \alpha_0)}{\partial \alpha^\top} \times (\hat{\alpha} - \alpha_0) := W \times (\hat{\alpha} - \alpha_0),$$

where

$$\frac{\partial \hat{m}_\psi^0(X_i, \theta_0, h_0, \alpha_0)}{\partial \alpha^\top} = \frac{\sum_{j=1}^{n}\partial\mathcal{F}_{ij}(\alpha_0)\psi_j}{\sum_{j=1}^{n}\mathcal{F}_{ij}(\alpha_0)} - \frac{\sum_{j=1}^{n}\mathcal{F}_{ij}(\alpha_0)\psi_j\sum_{j=1}^{n}\partial\mathcal{F}_{ij}(\alpha_0)Y_j}{\{\sum_{j=1}^{n}\mathcal{F}_{ij}(\alpha_0)\}^2}$$

in which $\psi_j = \psi(Y_j, X_j, \theta_0, h_0)$, $\mathcal{F}_{ij}(\alpha_0) = \delta_j O_j(\alpha_0) K_a(X_j - X_i)$, $\partial\mathcal{F}_{ij}(\alpha_0) = \delta_j \partial_\alpha O_j(\alpha_0) K_a(X_j - X_i)$, $O_j(\alpha_0) = O(X_j, Y_j, \alpha_0)$ and $\partial_\alpha O_j(\alpha_0) = \partial O_j(\alpha_0)/\partial\alpha$.

Define $\pi_j(\alpha_0) := \pi(X_j, Y_j, \alpha_0)$ and $z_j(\alpha_0) := \partial \text{logit}\{\pi_j(\alpha_0)\}/\partial\alpha$. By simple calculation, we obtain $\partial\pi_j(\alpha_0)/\partial\alpha = \pi_j(\alpha_0)(1-\pi_j(\alpha_0))z_j(\alpha_0)$, which leads to $\partial O_j(\alpha_0)/\partial\alpha = (1-\pi_j^{-1}(\alpha_0))z_j(\alpha_0) = -O_j(\alpha_0)z_j(\alpha_0)$. Let $m_z^0(X,\alpha_0) = E(z(X,Y,\alpha_0)|X,\delta=0)$ and $m_{z\psi}^0(X,\alpha_0) = E(z(X,Y,\alpha_0)\psi(Y,X,\theta_0,h_0)|Z,\delta = 0)$. Using the kernel regression method yields

$$\hat{m}_z^0(X_i,\alpha_0) = \frac{\sum_{j=1}^n \mathcal{F}_{ij}(\alpha_0)z_j(\alpha_0)}{\sum_{j=1}^n \mathcal{F}_{ij}(\alpha_0)}, \quad \hat{m}_{z\psi}^0(X_i,\alpha_0) = \frac{\sum_{j=1}^n \mathcal{F}_{ij}(\alpha_0)z_j(\alpha_0)\psi_j}{\sum_{j=1}^n \mathcal{F}_{ij}(\alpha_0)}.$$

Combining the above equations leads to

$$\frac{\partial \hat{m}_\psi^0(X_i,\theta_0,h_0,\alpha_0)}{\partial\alpha} = \hat{m}_\psi^0(X_i,\theta_0,h_0,\alpha_0)\hat{m}_z^0(X_i,\alpha_0) - \hat{m}_{z\psi}^0(X_i,\alpha_0).$$

Let $\Delta_n(X_i) = \hat{G}(X_i) - G(X_i)$. For the sake of notation simplicity, we temporarily denote $m_\psi^0(X) = m_\psi^0(X_i,\theta_0,h_0)$ and $\hat{m}_\psi^0(X_i) = \hat{m}_\psi^0(X_i,\theta_0,h_0,\alpha_0)$.

Taking further decomposition for W, we have

$$\begin{aligned}
W &= \frac{1}{n}\sum_{i=1}^n(1-\delta_i)\frac{\partial \hat{m}_\psi^0(X_i,\theta_0,h_0,\alpha_0)}{\partial\alpha} \\
&= \frac{1}{n}\sum_{i=1}^n(1-\delta_i)\{\hat{m}_\psi^0(X_i)\hat{m}_z^0(X_i,\alpha_0) - \hat{m}_{z\psi}^0(X_i,\alpha_0)\} \\
&= \frac{1}{n}\sum_{i=1}^n(1-\delta_i)\{\hat{m}_\psi^0(X_i)\hat{m}_z^0(X_i,\alpha_0) - m_\psi^0(X_i)m_z^0(X_i,\alpha_0)\} \\
&\quad -\frac{1}{n}\sum_{i=1}^n(1-\delta_i)\{\hat{m}_{z\psi}^0(X_i,\alpha_0) - m_\psi^0(X_i)m_z^0(X_i,\alpha_0)\} \\
&:= W_1 - W_2
\end{aligned}$$

For W_1, we have

$$\begin{aligned}
W_1 &= \frac{1}{n}\sum_{i=1}^n(1-\delta_i)m_z^0(X_i,\alpha_0)\{\hat{m}_\psi^0(X_i) - m_\psi^0(X_i)\} \\
&\quad +\frac{1}{n}\sum_{i=1}^n(1-\delta_i)\{\hat{m}_z^0(X,\alpha_0) - m_z^0(X_i,\alpha_0)\}\{\hat{m}_\psi^0(X_i) - m_\psi^0(X_i)\} \\
&\quad +\frac{1}{n}\sum_{i=1}^n(1-\delta_i)m_\psi^0(X_i)\{\hat{m}_z^0(X_i,\alpha_0) - m_z^0(X_i,\alpha_0)\} \\
&:= W_{11} + W_{12} + W_{13}.
\end{aligned}$$

For W_2, we have

$$\begin{aligned} W_2 &= \frac{1}{n}\sum_{i=1}^{n}(1-\delta_i)\{\hat{m}^0_{z\psi}(X_i,\alpha_0) - m^0_{\psi}(X_i)m^0_z(X_i,\alpha_0)\} \\ &= \frac{1}{n}\sum_{i=1}^{n}(1-\delta_i)\{\hat{m}^0_{z\psi}(X_i,\alpha_0) - m^0_{z\psi}(X,\alpha_0)\} \\ &\quad + \frac{1}{n}\sum_{i=1}^{n}(1-\delta_i)\{m^0_{z\psi}(X,\alpha_0) - m^0_{\psi}(X_i)m^0_z(X_i,\alpha_0)\} \\ &:= W_{21} + W_{22}. \end{aligned}$$

It follows from standard arguments that $W_{12} = o_p(n^{-1/2})$. Using the similar arguments given in proof of Lemma 1 for I_3, we have

$$\begin{aligned} W_{11} &= \frac{1}{n}\sum_{j=1}^{n}\delta_j\frac{1-\pi_j(\alpha_0)}{\pi_j(\alpha_0)}\{\psi_j - m^0_{\psi}(X_j)\}m^0_z(X_i) + o_p(n^{-1/2}), \\ W_{13} &= \frac{1}{n}\sum_{j=1}^{n}\delta_j\frac{1-\pi_j(\alpha_0)}{\pi_j(\alpha_0)}\{z_j(\alpha_0) - m^0_z(X_j)\}m^0_{\psi}(X_j,\theta_0,h_0) + o_p(n^{-1/2}), \\ W_{21} &= \frac{1}{n}\sum_{j=1}^{n}\delta_j\frac{1-\pi_j(\alpha_0)}{\pi_j(\alpha_0)}\{z_j(\alpha_0)\psi_j - m^0_{z\psi}(X,\alpha_0)\} + o_p(n^{-1/2}). \end{aligned}$$

By the law of large number,

$$\begin{aligned} W_{11} &= E[(1-\delta)\{\psi(Y,X,\theta_0,h_0) - m^0_{\psi}(X,\theta_0,h_0,\alpha_0)\}m^0_z(X,\alpha_0)] + o_p(1), \\ W_{13} &= E[(1-\delta)\{z(\alpha_0) - m^0_z(X)\}m^0_{\psi}(X,\theta_0,h_0) + o_p(1), \\ W_{21} &= E[(1-\delta)\{z(\alpha_0)\psi - m^0_{z\psi}(X,\alpha_0)\} + o_p(1), \\ W_{22} &= E[(1-\delta)\{m^0_{z\psi}(X,\alpha_0) - m^0_{\psi}(X)m^0_z(X,\alpha_0)\} + o_p(1). \end{aligned}$$

This leads to

$$W = -E[(1-\delta)\{\psi(Y,X,\theta_0,h_0) - m^0_{\psi}(X,\theta_0,h_0,\alpha_0)\}\{z(X,Y,\alpha_0) - m^0_z(X,\alpha_0)\}^{\top}] + o_p(1).$$

It follows from Lemma 1 and the above arguments that

$$\mathcal{G}_n(\theta_0,h_0,\hat{\alpha}) = \mathcal{G}_n(\theta_0,h_0,\alpha_0) - H\times(\hat{\alpha}-\alpha_0) + o_p(n^{-1/2}).$$

This completes the proof. □

Lemma 3. *Suppose that the regularity conditions (C1)-(C4) hold. Then, we have*

$$\begin{aligned} \mathcal{Q}_n(\alpha_0,\theta_0,h_0) &= \frac{1}{n}\sum_{i=1}^{n}\mathcal{D}_i(\theta_0,h_0,\alpha_0) + o_p(n^{-1/2}), \\ \frac{\partial\mathcal{Q}_n(\alpha_0,\theta_0,h)}{\partial\alpha^{\top}} &= H + o_p(1), \end{aligned}$$

where $\mathcal{D}_i(\theta_0,h_0,\alpha_0) = [(1-\delta_i)r_i/\nu - \delta_i\{\pi^{-1}(X_i,Y_i,\alpha_0)-1\}][\psi(Y_i,X_i,\theta_0,h_0) - m^0_{\psi}(X_i,\theta_0,h_0,\alpha_0)\}$*, and* H *is defined in Lemma 2.*

Proof. For $\mathcal{Q}_n(\alpha,\theta,h)$, we have

$$\begin{aligned}\mathcal{Q}_n(\alpha,\theta,h) &= \frac{1}{n}\sum_{i=1}^{n}(1-\delta_i)\tfrac{r_i}{\nu}\{\psi(Y_i,X_i,\theta,h)-m_\psi^0(X_i,\theta,h,\alpha)\}\\ &\quad -\frac{1}{n}\sum_{i=1}^{n}(1-\delta_i)\{\hat{m}_\psi^0(X_i,\theta,h,\alpha)-m_\psi^0(X_i,\theta,h,\alpha)\}\\ &\quad +\frac{1}{n}\sum_{i=1}^{n}(1-\delta_i)\{1-\tfrac{r_i}{\nu}\}\{\hat{m}_\psi^0(X_i,\theta,h,\alpha)-m_\psi^0(X_i,\theta,h,\alpha)\}.\end{aligned}$$

Using the similar arguments as given in proof of Lemma 1 for I_3, we have $n^{-1}\sum_{i=1}^n(1-\delta_i)\{\hat{m}_\psi^0(X_i,\theta_0,h_0,\alpha_0)-m_\psi^0(X_i,\theta_0,h_0,\alpha_0)\} = n^{-1}\sum_{i=1}^n\delta_i(1-\pi_i(\alpha_0))\ \{\pi_i(\alpha_0)\}^{-1}\{\psi(Y_i,X_i,\theta_0,h_0)-m_\psi^0(X_i,\theta_0,h_0,\alpha_0)\}+o_p(n^{-1/2})$, and $n^{-1}\sum_{i=1}^n(1-\delta_i)(1-r_i/\nu)\{\hat{m}_\psi^0(X_i,\theta_0,h_0,\alpha_0)-m_\psi^0(X_i,\theta_0,h_0,\alpha_0)\} = o_p(n^{-1/2})$.

Using the similar arguments as given in proof of Lemma 2 for I_{n4}, together with $E\{r|\delta=0,X\}=\nu$, we have

$$\begin{aligned}\frac{\partial\mathcal{Q}_n(\alpha_0,\theta_0,h)}{\partial\alpha^\top} &= -\frac{1}{n}\sum_{i=1}^{n}(1-\delta_i)\frac{r_i}{\nu}\frac{\partial\hat{m}_\psi^0(X_i,\theta,h,\alpha)}{\partial\alpha^\top}\\ &= -\frac{1}{n}\sum_{i=1}^{n}(1-\delta_i)\frac{\partial\hat{m}_\psi^0(X_i,\theta,h,\alpha)}{\partial\alpha^\top}\\ &\quad +\frac{1}{n}\sum_{i=1}^{n}(1-\delta_i)(1-r_i/\nu)\frac{\partial\hat{m}_\psi^0(X_i,\theta,h,\alpha)}{\partial\alpha^\top}\\ &= H+o_p(1).\end{aligned}$$

Lemma 3 is proved. □

S5 Proofs of Theorems 1 - 5

Proofs of Theorem 1 and Theorem 2. We only prove Theorem 2. Theorem 1 can be proved in the same way by setting $\hat{\alpha}=\alpha_0$ and using Lemma 1. Define $\widetilde{\mathcal{G}}(\theta,h,\alpha)=E\{\tilde{\psi}(Y_i,X_i,\theta,h,\alpha)\}$, where $\tilde{\psi}(Y_i,X_i,\theta,h,\alpha)$ is defined in equation (2.3). It is easily shown that $\|\mathcal{G}_n(\theta,h,\hat{\alpha})-\widetilde{\mathcal{G}}(\theta,h,\alpha_0)\| \le \|\mathcal{G}_n(\theta,h,\hat{\alpha})-\widetilde{\mathcal{G}}_n(\theta,h,\alpha_0)\|+\|\widetilde{\mathcal{G}}_n(\theta,h,\alpha_0)-\widetilde{\mathcal{G}}(\theta,h,\alpha_0)\|$, where $\widetilde{\mathcal{G}}_n(\theta,h,\alpha) = n^{-1}\sum_{i=1}^n\tilde{\psi}(Y_i,X_i,\theta,h,\alpha)$. From condition (A3), we have, for all sequences $\varrho_n=o_p(1)$,

$$\sup_{\theta\in\Theta,\|h-h_0\|\le\varrho_n}\|\mathcal{G}_n(\theta,h,\hat{\alpha})-\widetilde{\mathcal{G}}_n(\theta,h,\alpha_0)\| = o_p(1).$$

It is easily shown that $\sup_{\theta\in\Theta,\|h-h_0\|\le\varrho_n}\|\widetilde{\mathcal{G}}_n(\theta,h,\alpha_0)-\widetilde{\mathcal{G}}(\theta,h,\alpha_0)\|=o_p(1)$. Therefore, we obtain

$$\sup_{\theta\in\Theta,\|h-h_0\|\le\varrho_n}\|\mathcal{G}_n(\theta,h,\hat{\alpha})-\widetilde{\mathcal{G}}(\theta,h,\alpha_0)\|=o_p(1).$$

Under NMAR assumption, and when the response model (2.1) is correctly specified, we obtain $E\{\widetilde{\mathcal{G}}_n(\theta_0,h_0,\alpha_0)\}=\widetilde{\mathcal{G}}(\theta_0,h_0,\alpha_0)=\mathcal{G}(\theta_0,h_0)=0$. This, together with the construction of estimator $\hat{\theta}_{SP}$, conditions (A1), (A2), and applying Theorem 1 in Chen et al. (2003), implies the uniform consistency of $\hat{\theta}_{SP}$.

Next, we investigate the asymptotic normality $\hat{\theta}_{SP}$. Note that

$$\begin{aligned}
&\|\mathcal{G}_n(\theta,h,\hat{\alpha})-\widetilde{\mathcal{G}}(\theta,h,\alpha_0)-\mathcal{G}_n(\theta_0,h_0,\hat{\alpha})+\widetilde{\mathcal{G}}(\theta_0,h_0,\alpha_0)\|\\
&\quad\le\|\mathcal{G}_n(\theta,h,\hat{\alpha})-\widetilde{\mathcal{G}}_n(\theta,h,\alpha_0)-\mathcal{G}_n(\theta_0,h_0,\hat{\alpha})+\widetilde{\mathcal{G}}_n(\theta_0,h_0,\alpha_0)\|\\
&\qquad+\|\widetilde{\mathcal{G}}_n(\theta,h,\alpha_0)-\widetilde{\mathcal{G}}(\theta,h,\alpha_0)-\widetilde{\mathcal{G}}_n(\theta_0,h_0,\alpha_0)+\widetilde{\mathcal{G}}(\theta_0,h_0,\alpha_0)\|.
\end{aligned}$$

Using condition (B5) yields

$$\sup{}^*\|\mathcal{G}_n(\theta,h,\hat{\alpha})-\widetilde{\mathcal{G}}_n(\theta,h,\alpha_0)-\mathcal{G}_n(\theta_0,h_0,\hat{\alpha})+\widetilde{\mathcal{G}}_n(\theta_0,h_0,\alpha_0)\|=o_p(n^{-1/2}),$$

where $\sup^*$ is the supremum over all $\|\theta-\theta\|\le\varrho_n$ and $\|h-h\|_{\mathcal{H}}\le\varrho_n$, with $\varrho_n=o(1)$. Under NMAR assumption, and when the response model (2.1) is correctly specified, we have that (i) $\tilde{\psi}(Y_i,X_i,\theta,h,\alpha_0)=\delta_i\psi(Y_i,X_i,\theta,h)+(1-\delta_i)E\{\psi(Y_i,X_i,\theta,h)|X_i,\delta_i=0\}$, (ii) $\widetilde{\mathcal{G}}(\theta,h,\alpha_0)=\mathcal{G}(\theta,h)$ and $\widetilde{\mathcal{G}}(\theta_0,h_0,\alpha_0)=\mathcal{G}(\theta_0,h_0)=0$. Therefore, by condition (B2), for $j=1,\cdots,q$ and for fixed $(\theta,h)\in\Theta\times\mathcal{H}$,

$$\begin{aligned}
&E\Bigg[\sup{}^{**}|\tilde{\psi}_j(Y_i,X_i,\theta',h',\alpha_0)-\tilde{\psi}_j(Y_i,X_i,\theta,h,\alpha_0)|\\
&=E\Bigg[\sup{}^{**}|\delta\psi_j(Y,X,\theta',h')-\delta\psi_j(Y,X,\theta,h)\\
&\quad+(1-\delta)E\{\psi_j(Y,X,\theta',h')|X,\delta=0\}-(1-\delta)E\{\psi_j(Y,X,\theta,h)|X,\delta=0\}|\Bigg]\\
&\quad\le 2E\{\sup{}^{**}|\psi_j(Y,X,\theta',h')-\psi_j(Y,X,\theta,h)|\},
\end{aligned}$$

where $\sup^{**}$ is the supremum over all $\|\theta'-\theta\|\le\varrho_n$ and $\|h'-h\|_{\mathcal{H}}\le\varrho_n$, with $\varrho_n=o(1)$. Following the argument of Theorem 3 in Chen et al. (2003) and using condition (B3), we have $\|\widetilde{\mathcal{G}}_n(\theta,h,\alpha_0)-\widetilde{\mathcal{G}}(\theta,h,\alpha_0)-\widetilde{\mathcal{G}}_n(\theta_0,h_0,\alpha_0)+\widetilde{\mathcal{G}}(\theta_0,h_0,\alpha_0)\|=o_p(n^{-1/2})$. Therefore, we obtain

$$\sup{}^*\|\mathcal{G}_n(\theta,h,\hat{\alpha})-\mathcal{G}(\theta,h)-\mathcal{G}_n(\theta_0,h_0,\hat{\alpha})+\mathcal{G}(\theta_0,h_0)\|=o_p(n^{-1/2}).$$

Now we define the linearization $\mathcal{L}_n(\theta) = \mathcal{G}_n(\theta_0, h_0, \hat{\alpha}) + \Lambda(\theta_0, h_0)(\theta - \theta_0) + \Gamma(\theta_0, h_0)[\hat{h} - h_0]$. By conditions (B1)-(B5) and using the arguments of Chen et al. (2003), we obtain

$$\begin{aligned}
&\|\mathcal{G}_n(\hat{\theta}_{SP}, \hat{h}, \hat{\alpha}) - \mathcal{L}_n(\hat{\theta}_{SP})\| \\
&\quad \le \|\mathcal{G}_n(\hat{\theta}_{SP}, \hat{h}, \hat{\alpha}) - \mathcal{G}(\hat{\theta}_{SP}, \hat{h}) - \mathcal{G}_n(\theta_0, h_0, \hat{\alpha}) + \mathcal{G}(\theta_0, h_0)\| \\
&\qquad + \|\mathcal{G}(\hat{\theta}_{SP}, \hat{h}) - \mathcal{G}(\hat{\theta}_{SP}, h_0) - \Gamma(\theta_0, h_0)[\hat{h} - h_0]\| \\
&\qquad + \|\mathcal{G}(\hat{\theta}_{SP}, h_0) - \Lambda(\theta_0, h_0)(\hat{\theta}_{SP} - \theta_0)\| = o_p(n^{-1/2}).
\end{aligned}$$

Using the arguments of Pakes and Pollard (1989) and Chen et al. (2003) again, we have that $\|\mathcal{G}_n(\bar{\theta}, \hat{h}) - \mathcal{L}_n(\bar{\theta})\| = o_p(n^{-1/2})$, where

$$n^{1/2}(\bar{\theta} - \theta_0) = -(\Lambda^\top W \Lambda)^{-1} \Lambda^\top W n^{1/2} \{\mathcal{G}_n(\theta_0, h_0, \hat{\alpha}) + \Gamma(\theta_0, h_0)[\hat{h} - h_0]\},$$

is the minimizer of $\mathcal{L}_n(\theta)$, and $n^{1/2}(\hat{\theta}_{SP} - \bar{\theta}) = o_p(1)$. This, together with Lemma 2 and condition (B4), implies that

$$\hat{\theta}_{SP} - \theta_0 = -n^{-1} \sum_{i=1}^{n} (\Lambda^\top W \Lambda)^{-1} \Lambda^\top W \{\mathcal{S}(X_i, Y_i \theta_0, h_0) - H \times (\hat{\alpha} - \alpha_0)\} + o_p(n^{-1/2}).$$

Since $\hat{\alpha}$ is independent of $\hat{\psi}(Y_i, X_i, \theta, h, \alpha)$, it follows from the above arguments that $n^{1/2}(\hat{\theta}_{SP} - \theta_0) \xrightarrow{\mathcal{L}} \mathcal{N}(0, \Sigma_2)$. Theorem 2 is proved. □

Proof of Proposition 1. Let $\widetilde{\mathcal{Q}}_n(\alpha, \theta, h) = n^{-1} \sum_{i=1}^{n} (1 - \delta_i) \frac{r_i}{\nu} \{\psi(Y_i, X_i, \theta, h) - m_\psi^0(X_i, \theta, h, \alpha)\}$. Note that, under condition (A3), we have

$$\begin{aligned}
&\sup{}^{***} \|\mathcal{Q}_n(\alpha, \theta, h) - \mathcal{Q}(\alpha, \theta, h)\| \\
&\le \sup{}^{***} \|\mathcal{Q}_n(\alpha, \theta, h) - \widetilde{\mathcal{Q}}_n(\alpha, \theta, h)\| + \sup{}^{***} \|\widetilde{\mathcal{Q}}_n(\alpha, \theta, h) - \mathcal{Q}(\alpha, \theta, h)\| \\
&= o_p(1),
\end{aligned}$$

where $\sup^{***}$ is the supremum over all $\alpha \in \mathcal{B}, \theta \in \Theta$, and $\|h - h_0\|_{\mathcal{H}} \le \varrho_n$ for all positive values $\varrho_n = o(1)$. Using the argument of Theorem 1 in Chen et al. (2003) implies the uniform consistency of $\hat{\alpha}_{\mathrm{v}}$.

By $\mathcal{Q}(\alpha_0, \theta, h) = 0$ for any $\theta \in \Theta$ and $h \in \mathcal{H}$, we obtain that estimations of nuisance parameters θ and h in (3.2) do not affect the asymptotic properties of α. That is, we could consider the asymptotic properties of $\hat{\alpha}_{\mathrm{v}}$ for some fixed θ and h in (3.2). Therefore, the asymptotic linear expansion for $\hat{\alpha}_{\mathrm{v}}$ is given by

$$\hat{\alpha}_{\mathrm{v}} - \alpha_0 = -(H^\top \widetilde{W} H)^{-1} H^\top \widetilde{W} \frac{1}{n} \sum_{i=1}^{n} \mathcal{D}_i(\theta_0, h_0, \alpha_0) + o_p(n^{-1/2}).$$

The proof is completed. □

Proof of Theorem 3. To prove Theorem 3, we should first consider the asymptotic prosperities of $\mathcal{G}_n(\theta_0, h_0, \hat{\alpha}_{\rm v})$ when $\hat{\alpha}_{\rm v}$ is computed from the validation sample. According to Lemma 2, together with proposition 1, we have

$$\begin{aligned}\mathcal{G}_n(\theta_0, h_0, \hat{\alpha}_{\rm v}) &= \frac{1}{n}\sum_{i=1}^{n}\left\{\frac{\delta_i}{\pi(X_i, Y_i)}\{\psi_i - m_\psi^0(X_i)\} + \frac{1}{n}\sum_{i=1}^{n} m_\psi^0(X_i)\right.\\ &\quad \left. + H(H^\top \widetilde{W} H)^{-1} H^\top \widetilde{W} \mathcal{D}_i(\theta_0, h_0, \alpha_0)\right\} + o_p(n^{-1/2}).\end{aligned}$$

The rest of the proof of consistency and normality is similar to that given in Theorem 2. □

Proof of Proposition 2. Following Qin et al. (2002), we define some notations: $\zeta = \lambda_1(1-\omega)$, $\eta = (\alpha, \omega, \zeta)^\top$, $\eta_0 = (\alpha_0, \omega_0, 0)^\top$ and $a_n = nn_1^{-1} - \omega_0^{-1}$, where (α_0, ω_0) denotes the true value of (α, ω). Denote E and Var as the expectation and variance with respect to $F(X, Y)$ and E_C as expectation with respect to the conditional distribution $\pi(X, Y, \alpha_0)dF(X, Y)/\omega_0$. Let

$$\begin{aligned}\phi_{i1}(\eta_0) &= 0_{\dim(\alpha)\times 1},\\ \phi_{i2}(\eta_0) &= \frac{\omega_0}{1-\omega_0}\pi_i^{-1}(\alpha_0)(\pi_i(\alpha_0) - \omega_0),\\ \phi_{i3}(\eta_0) &= \frac{\omega_0}{1-\omega_0}\pi_i^{-1}(\alpha_0) g(X_i),\\ \mathcal{I}_1 &= \frac{\omega_0^2}{1-\omega_0} E_C\{\pi_i^{-1}(\alpha_0)\partial_\alpha \log \pi_i(\alpha_0)\},\\ \mathcal{I}_2 &= -\frac{\omega_0^3}{(1-\omega_0)^2} E_C\{\pi_i^{-2}(\alpha_0)(\pi_i(\alpha_0) - \omega_0)^2\},\\ \mathcal{I}_3 &= -\frac{\omega_0^3}{(1-\omega_0)^2} E_C\{\pi_i^{-2}(\alpha_0) g(X_i)(\pi_i(\alpha_0) - \omega_0)\}\end{aligned}$$

and

$$M(\eta_0) = \begin{pmatrix} 0 & \dfrac{\partial_\alpha \pi_i(\alpha_0)}{(1-\omega_0)\pi_i^2(\alpha_0)} & \dfrac{\omega_0 g(X_i)\partial_\alpha \pi_i(\alpha_0)}{(1-\omega_0)\pi_i^2(\alpha_0)} \\ \dfrac{\omega_0^2 \partial_\alpha \pi_i(\alpha_0)}{(1-\omega_0)\pi_i^2(\alpha_0)} & \dfrac{\omega_0^2(\pi_i(\alpha_0) - \omega_0)}{(1-\omega_0)^2\pi_i^2(\alpha_0)} & \dfrac{\omega_0^2 g(X_i)(\pi_i(\alpha_0) - \omega_0)}{(1-\omega_0)^2\pi_i^2(\alpha_0)} \\ \dfrac{\omega_0 g(X_i)\partial_\alpha \pi_i(\alpha_0)}{(1-\omega_0)\pi_i^2(\alpha_0)} & -\dfrac{\omega_0 g(X_i)}{(1-\omega_0)^2\pi_i^2(\alpha_0)} & \dfrac{\omega_0^2 g(U_i) g^\top(X_i)}{(1-\omega_0)\pi_i^2(\alpha_0)}, \end{pmatrix}.$$

Then, using the arguments of Qin et al. (2002), we obtain

$$n^{1/2}(\hat{\eta} - \eta_0) = U^{-1} n^{-1/2} \sum_{i=1}^{n} \phi_i(\eta_0) + o_p(1) \xrightarrow{\mathcal{L}} \mathcal{N}(0, U^{-1} V (U^{-1})^\top),$$

where $U = -\omega_0 E_C\{M(\eta_0)\}$, $V = \text{Var}\{\phi_i(\eta_0)\}$ and $\phi_i(\eta_0) = \delta_i(\phi_{i1}(\eta_0)^\top, \phi_{i2}(\eta_0)^\top, \phi_{i3}(\eta_0)^\top)^\top + \mathcal{T}(1/\omega_0 - \delta_i/\omega_0^2)$ with $\mathcal{T} = (\mathcal{I}_1^\top, \mathcal{I}_2^\top, \mathcal{I}_3^\top)^\top$. Consequently, the asymptotic expansion for $\hat{\alpha}_s$ could be given by

$$n^{1/2}(\hat{\alpha}_{\rm s} - \alpha_0) \quad = \quad n^{-1/2}\sum_{i=1}^{n} \Psi_i(\alpha_0) + o_p(1),$$

where $\Psi_i(\alpha)$ is the influence function, which is given by the appropriate submatrix of the matrix $U^{-1}\phi_i(\eta_0)$. □

Proof of Theorem 4. According to Lemma 2 together with Proposition 2, we have $\mathcal{G}_n(\theta_0, h_0, \hat{\alpha}_{\rm s}) = n^{-1}\sum_{i=1}^n\{(\psi_i - m_\psi^0(X_i))/\pi(X_i, Y_i) + m_\psi^0(X_i) - H\Psi_i(\alpha_0)\} + o_p(n^{-1/2})$. Proof of consistency and asymptotic normality is similar to those given in Theorem 2 and Theorem 3. □

Proof of Theorem 5. The proof of Theorem 5 can be easily obtained by using the same argument as given in the proof of Theorems 1-4, and hence is omitted here. □

References

Chen, X., Linton, O. B. and, Van Keilegom, I. (2003). Estimation of semiparametric models when the criterion function is not smooth. *Econometrica.* **71**, 1591–1608.

Chen, S. X., and Van Keilegom, I. (2013). Estimation in semiparametric models with missing data. *Annals of the Institute of Statistical Mathematics.* **65**, 785–805.

Kim, J. K. and Yu, C. L. (2011). A semiparametric estimation of mean functionals with nonignorable missing data. *Journal of the American Statistical Association* **106**, 157–165.

Pakes, A., Pollard, D. (1989). Simulation and the asymptotics of optimization estimators. *Econometrica.* **57,** 1027–1057.

Qin, J., Leung, D., and, Shao, J. (2002). Estimation with survey data under nonignorable nonresponse or informative sampling. *Journal of the American Statistical Association.* **97**, 193–200.

Randles, R. H. (1982). On the asymptotic normality of statistics with estimated parameters. *The Annals of Statistics* **10**, 462–474.

Puying Zhao
Department of Statistics, Yunnan University,
Kunming 650091, China
E-mail: pyzhao@live.cn

Niansheng Tang
Department of Statistics, Yunnan University,
Kunming 650091, China
E-mail: nstang@ynu.edu.cn

Annie Qu
Department of Statistics
University of Illinois at Urbana-Champaign
725 S. Wright Street Champaign, IL 61820 USA
E-mail: anniequ@illinois.edu

Depeng Jiang
Department of Community Health Sciences
University of Manitoba
Winnipeg, Manitoba, R3E 0W3 Canada
E-mail: depeng.jiang@umanitoba.ca