A two-stage multi-level randomized response technique with proportional odds models and missing covariates

Appendix A: Maximum Likelihood Estimator

To derive the asymptotic properties of the ML estimator $\widehat{\boldsymbol{\Theta}}_F$, the following regularity conditions are required.

(A1) $\mathrm{E}\left[\Psi_1(\boldsymbol{\Theta})\Psi_1^T(\boldsymbol{\Theta})\right]$ is positive definite in a neighborhood of the true value of $\boldsymbol{\Theta}$, where $\Psi_1(\boldsymbol{\Theta}) = \mathcal{X}_1 \mathrm{diag}\left[\boldsymbol{H}_1^{(1)}(\boldsymbol{\Theta})\right] \boldsymbol{W}^T \Sigma_1^{-1}(\boldsymbol{\Theta}) \left\{\boldsymbol{d}_1 - [\boldsymbol{p}_1 + \boldsymbol{W}\boldsymbol{H}_1(\boldsymbol{\Theta})]\right\}$.

(A2) The first derivative of $U_n(\boldsymbol{\Theta})$ with respect to $\boldsymbol{\Theta}$ exists almost surely in a neighborhood of the true value of $\boldsymbol{\Theta}$. Furthermore, in such a neighborhood, the first derivative is bounded above by a function of $(D, \boldsymbol{X}, \boldsymbol{Z})$.

We require condition (A1) for the unique solution of the estimating equations. Condition (A2) is required for the proof of consistency in the estimating equation theory. To show the consistency of the estimator $\widehat{\boldsymbol{\Theta}}_F$, we consider

$$
\begin{aligned}
G_n(\boldsymbol{\Theta}) &= \frac{1}{\sqrt{n}} \left[\frac{-\partial U_n(\boldsymbol{\Theta})}{\partial \boldsymbol{\Theta}^T}\right] \\
&= \frac{1}{n}\sum_{i=1}^n \left[\frac{\partial \boldsymbol{\pi}_i^D(\boldsymbol{\Theta})}{\partial \boldsymbol{\Theta}}\right] \Sigma_i^{-1}(\boldsymbol{\Theta}) \left[\frac{\partial \boldsymbol{\pi}_i^D(\boldsymbol{\Theta})}{\partial \boldsymbol{\Theta}}\right]^T \\
&\quad - \frac{1}{n}\sum_{i=1}^n \frac{\partial}{\partial \boldsymbol{\Theta}^T} \left\{\left[\frac{\partial \boldsymbol{\pi}_i^D(\boldsymbol{\Theta})}{\partial \boldsymbol{\Theta}}\right] \Sigma_i^{-1}(\boldsymbol{\Theta})\right\} \left\{\boldsymbol{d}_i - \boldsymbol{\pi}_i^D(\boldsymbol{\Theta})\right\} \\
&= \frac{1}{n}\sum_{i=1}^n \mathcal{X}_i \mathrm{diag}\left[\boldsymbol{H}_i^{(1)}(\boldsymbol{\Theta})\right] \boldsymbol{W}^T \Sigma_i^{-1}(\boldsymbol{\Theta}) \boldsymbol{W} \mathrm{diag}\left[\boldsymbol{H}_i^{(1)}(\boldsymbol{\Theta})\right] \mathcal{X}_i^T \\
&\quad - \frac{1}{n}\sum_{i=1}^n \frac{\partial}{\partial \boldsymbol{\Theta}^T} \left\{\left[\frac{\partial \boldsymbol{\pi}_i^D(\boldsymbol{\Theta})}{\partial \boldsymbol{\Theta}}\right] \Sigma_i^{-1}(\boldsymbol{\Theta})\right\} \left\{\boldsymbol{d}_i - \boldsymbol{\pi}_i^D(\boldsymbol{\Theta})\right\}.
\end{aligned}
$$

It can then be shown that $G_n(\mathbf{\Theta}) \xrightarrow{p} G(\mathbf{\Theta})$, where $G(\mathbf{\Theta}) = \mathrm{Cov}\left[U_n(\mathbf{\Theta})\right] = \mathrm{E}\left[\Psi_1(\mathbf{\Theta})\Psi_1^T(\mathbf{\Theta})\right]$. By the Inverse Function Theorem of Foutz (1977) and condition (A1), it follows that a unique consistent solution exists for the estimating equations $U_n(\mathbf{\Theta}) = \mathbf{0}$ in a neighborhood of the true value of $\mathbf{\Theta}$. Therefore, $\widehat{\mathbf{\Theta}}_F$ is shown to be a consistent estimator of $\mathbf{\Theta}$.

To derive the asymptotic distribution of $\sqrt{n}(\widehat{\mathbf{\Theta}}_F - \mathbf{\Theta})$, by a Taylor's series expansion of $U(\widehat{\mathbf{\Theta}}_F)$ at $\mathbf{\Theta}$, we can have

$$\begin{aligned}\mathbf{0} = U_n(\widehat{\mathbf{\Theta}}_F) &= U_n(\mathbf{\Theta}) + \frac{\partial U_n(\mathbf{\Theta})}{\partial \mathbf{\Theta}^T}(\widehat{\mathbf{\Theta}}_F - \mathbf{\Theta}) + o_p(1) \\ &= U_n(\mathbf{\Theta}) - G_n(\mathbf{\Theta})\sqrt{n}(\widehat{\mathbf{\Theta}}_F - \mathbf{\Theta}) + o_p(1).\end{aligned}$$

It can be shown that $\sqrt{n}(\widehat{\mathbf{\Theta}}_F - \mathbf{\Theta}) = G^{-1}(\mathbf{\Theta})U_n(\mathbf{\Theta}) + o_p(1)$ because $G_n(\mathbf{\Theta}) \xrightarrow{p} G(\mathbf{\Theta})$. $\sqrt{n}(\widehat{\mathbf{\Theta}}_F - \mathbf{\Theta})$ is then shown to be asymptotically a normal distribution with mean $\mathbf{0}$ and covariance matrix $\Delta_F = G^{-1}(\mathbf{\Theta})$. Let $\mathcal{G}_0(\widehat{\mathbf{\Theta}}_F) = \frac{1}{n}\sum_{i=1}^{n} \mathcal{X}_i \mathrm{diag}\left[\boldsymbol{H}_i^{(1)}(\widehat{\mathbf{\Theta}}_F)\right] \boldsymbol{W}^T \Sigma_i^{-1}(\widehat{\mathbf{\Theta}}_F) \boldsymbol{W} \mathrm{diag}\left[\boldsymbol{H}_i^{(1)}(\widehat{\mathbf{\Theta}}_F)\right] \mathcal{X}_i^T$. $\widehat{\Delta}_F = \mathcal{G}_0^{-1}(\widehat{\mathbf{\Theta}}_F)$ can then be shown to be a consistent estimator of Δ_F because $\widehat{\mathbf{\Theta}}_F \xrightarrow{p} \mathbf{\Theta}$.

Appendix B: Missing Data in Covariates

We wish to derive the asymptotic properties of the proposed estimators under the assumptions that X is MAR and D, X, and $\boldsymbol{V}$ are discrete. Hence, the following regularity conditions are required.

(B1) The $supp(\boldsymbol{V})$ denotes the support of $\boldsymbol{V}$. For $D_i = 0, 1, 2, \ldots, L+m-2$ and $\boldsymbol{V}_i \in supp(\boldsymbol{V})$, the selection probability $\pi(D_i, \boldsymbol{V}_i) > 0$, $i = 1, 2, \ldots, n$.

(B2) $\mathrm{E}\left[\frac{\Psi_1(\boldsymbol{\Theta})\Psi_1^T(\boldsymbol{\Theta})}{\pi(D_1, \boldsymbol{V}_1)}\right]$ is finite and positive definite in a neighborhood of true value of $\boldsymbol{\Theta}$, where $\Psi_1(\boldsymbol{\Theta}) = \mathcal{X}_1 \mathrm{diag}\left[\boldsymbol{H}_1^{(1)}(\boldsymbol{\Theta})\right] \boldsymbol{W}^T \Sigma_1^{-1}(\boldsymbol{\Theta}) \left\{\boldsymbol{d}_1 - [\boldsymbol{p}_1 + \boldsymbol{W}\boldsymbol{H}_1(\boldsymbol{\Theta})]\right\}$.

(B3) The first derivative of $U_{wn}(\boldsymbol{\Theta}; \boldsymbol{\pi})$ with respect to $\boldsymbol{\Theta}$ exists almost surely in a neighborhood of the true value of $\boldsymbol{\Theta}$. Further, in such a neighborhood, the first derivative is bounded above by a function of $(D, X, \boldsymbol{V})$, whose expectation exists.

Inverse Probability Weighting Method

For any nonsingular matrix B, define $B^{-T} = [B^{-1}]^T$. For vector $\boldsymbol{b}$, define $\boldsymbol{b}^{\otimes 2} = \boldsymbol{b}\boldsymbol{b}^T$. The asymptotic properties of $\widehat{\boldsymbol{\Theta}}_W$ are stated in Theorem 1.

Theorem 1 *Under the regularity conditions (B1)-(B3), $\widehat{\boldsymbol{\Theta}}_W$ is a consistent estimator of $\boldsymbol{\Theta}$ and $\sqrt{n}(\widehat{\boldsymbol{\Theta}}_W - \boldsymbol{\Theta})$ has asymptotically a normal distribution with mean $\boldsymbol{0}$ and covariance matrix $\Delta_W = G^{-1}(\boldsymbol{\Theta}, \boldsymbol{\pi}) M(\boldsymbol{\Theta}, \boldsymbol{\pi}) G^{-T}(\boldsymbol{\Theta}, \boldsymbol{\pi})$, where $G(\boldsymbol{\Theta}, \boldsymbol{\pi}) = E\left[-\frac{\partial U_{wn}(\boldsymbol{\Theta}, \boldsymbol{\pi})}{\sqrt{n}\partial \boldsymbol{\Theta}}\right]$, $M(\boldsymbol{\Theta}, \boldsymbol{\pi}) = E\left\{\left[\frac{\delta_1}{\pi_1}\Psi_1(\boldsymbol{\Theta}) + \left(1 - \frac{\delta_1}{\pi_1}\right)\Psi_1^*(\boldsymbol{\Theta})\right]^{\otimes 2}\right\}$, and $\Psi_1^*(\boldsymbol{\Theta}) = E[\Psi_1(\boldsymbol{\Theta}) | D_1, \boldsymbol{V}_1]$.*

Proof of Theorem 1

Under the two-stage MRR technique, the IPW estimator of the vector of parameters of the POM solves the following weighted estimating equations:

$$U_{wn}(\boldsymbol{\Theta}; \widehat{\boldsymbol{\pi}}) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{\delta_i}{\widehat{\pi}_i} \Psi_i(\boldsymbol{\Theta}) = \boldsymbol{0}.$$

We can then have

$$
\begin{aligned}
U_{wn}(\boldsymbol{\Theta};\widehat{\boldsymbol{\pi}}) - U_{wn}(\boldsymbol{\Theta};\boldsymbol{\pi}) &= \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left(\frac{1}{\widehat{\pi}_i} - \frac{1}{\pi_i} \right) \delta_i \Psi_i(\boldsymbol{\Theta}) \\
&= \frac{-1}{\sqrt{n}} \sum_{i=1}^{n} \left\{ \left(\frac{\widehat{\pi}_i - \pi_i}{\pi_i^2} \right) + O_p\left[(\widehat{\pi}_i - \pi_i)^2 \right] \right\} \delta_i \Psi_i(\boldsymbol{\Theta}) \\
&= \frac{-1}{n^{3/2}} \sum_{k=1}^{n} \sum_{i=1}^{n} \left[\frac{(\delta_i - \pi_i)(\delta_k - \pi_i) I(D_k = D_i, \boldsymbol{V}_k = \boldsymbol{V}_i)}{\pi_i^2 P(D = D_i, \boldsymbol{V} = \boldsymbol{V}_i)} \right] \Psi_i(\boldsymbol{\Theta}) \\
&\quad - \frac{1}{n^{3/2}} \sum_{k=1}^{n} \sum_{i=1}^{n} \left[\frac{\pi_i(\delta_k - \pi_i) I(D_k = D_i, \boldsymbol{V}_k = \boldsymbol{V}_i)}{\pi_i^2 P(D = D_i, \boldsymbol{V} = \boldsymbol{V}_i)} \right] \Psi_i(\boldsymbol{\Theta}) + o_p(1) \\
&= -A_{1n} - A_{2n} + o_p(1).
\end{aligned}
$$

Let $\Psi_i^*(\boldsymbol{\Theta}) = \mathrm{E}\left[\Psi_i(\boldsymbol{\Theta}) | D_i, \boldsymbol{V}_i\right]$, $i = 1, \ldots, n$. It is then easily shown that

$$
\begin{aligned}
A_{2n} &= \frac{1}{n^{3/2}} \sum_{k=1}^{n} \sum_{i=1}^{n} \left[\frac{\pi_i(\delta_k - \pi_i) I(D_k = D_i, \boldsymbol{V}_k = \boldsymbol{V}_i)}{\pi_i^2 P(D = D_i, \boldsymbol{V} = \boldsymbol{V}_i)} \right] \Psi_i(\boldsymbol{\Theta}) \\
&= \frac{1}{\sqrt{n}} \sum_{k=1}^{n} \frac{\delta_k - \pi_k}{\pi_k} \Psi_i^*(\boldsymbol{\Theta}) + O_p\left(\frac{1}{\sqrt{n}} \right).
\end{aligned}
$$

In addition, let $l_{ik} = \frac{(\delta_i - \pi_i)(\delta_k - \pi_i) I(D_k = D_i, \boldsymbol{V}_k = \boldsymbol{V}_i)}{\pi_i^2 P(D = D_i, \boldsymbol{V} = \boldsymbol{V}_i)} \Psi_i(\boldsymbol{\Theta})$. We will show $\mathrm{E}(A_{1n}) = O\left(\frac{1}{\sqrt{n}}\right)$ and $\mathrm{Var}(A_{1n}) = O\left(\frac{1}{n}\right)$ to show $A_{1n} = O_p\left(\frac{1}{\sqrt{n}}\right)$. First, note that

$$
\mathrm{E}\left\{ \mathrm{E}\left[l_{ik} | X_i, D_k = D_i, \boldsymbol{V}_k = \boldsymbol{V}_i \right] \right\} = \begin{cases} 0, & i \neq k, \\ \mathrm{E}\left[\frac{\pi_i(1 - \pi_i)\Psi_i(\boldsymbol{\Theta}) I(D_k = D_i, \boldsymbol{V}_k = \boldsymbol{V}_i)}{\pi_i^2 P(D = D_i, \boldsymbol{V} = \boldsymbol{V}_i})\right], & i = k, \end{cases}
$$

and

$$
\mathrm{E}\left[\frac{\pi_i(1 - \pi_i)\Psi_i(\boldsymbol{\Theta}) I(D_k = D_i, \boldsymbol{V}_k = \boldsymbol{V}_i)}{\pi_i^2 P(D = D_i, \boldsymbol{V} = \boldsymbol{V}_i)} \right] = \mathrm{E}\left[\frac{(1 - \pi_i)\Psi_i^*(\boldsymbol{\Theta})}{\pi_i} \right], \; i = 1, \ldots, n.
$$

Hence, we have

$$
\mathrm{E}(A_{1n}) = \frac{1}{n^{3/2}} \sum_{i=1}^{n} \mathrm{E}\left[\frac{(1 - \pi_i)\Psi_1^*(\boldsymbol{\Theta})}{\pi_i} \right] = O\left(\frac{1}{\sqrt{n}} \right).
$$

By definition of l_{ik}, we can obtain $\mathrm{Cov}(l_{ik}, l_{sb}) = \mathrm{E}\left[\frac{(1-\pi_i)^2 \Psi_i^*(\boldsymbol{\Theta})[\Psi_i^*(\boldsymbol{\Theta})]^T}{\pi_i^2} \right]$ if $(i,k) = (s,b)$ and $\mathrm{Cov}(l_{ik}, l_{sb}) = 0$ otherwise. As a result, we have

$$
\mathrm{Var}(A_{1n}) = \frac{1}{n^3} \sum_{i,k}^{n} \mathrm{Var}(l_{ik}) = O\left(\frac{1}{n} \right)
$$

to show $A_{1n} = O_p(\frac{1}{\sqrt{n}})$. Therefore, it follows that

$$U_{wn}(\boldsymbol{\Theta};\widehat{\boldsymbol{\pi}}) - U_{wn}(\boldsymbol{\Theta};\boldsymbol{\pi}) = \frac{-1}{\sqrt{n}}\sum_{i=1}^{n}\left[\frac{(\delta_i-\pi_i)\Psi_i^*(\boldsymbol{\Theta})}{\pi_i}\right] + O_p\left(\frac{1}{\sqrt{n}}\right)$$

and

$$U_{wn}(\boldsymbol{\Theta};\widehat{\boldsymbol{\pi}}) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left[\frac{\delta_i\Psi_i(\boldsymbol{\Theta})}{\pi_i} - \frac{(\delta_i-\pi_i)\Psi_i^*(\boldsymbol{\Theta})}{\pi_i}\right] + O_p\left(\frac{1}{\sqrt{n}}\right).$$

Let $G_n(\boldsymbol{\Theta},\widehat{\boldsymbol{\pi}}) = \frac{1}{\sqrt{n}}\left[-\frac{\partial U_{wn}(\boldsymbol{\Theta},\widehat{\boldsymbol{\pi}})}{\partial\boldsymbol{\Theta}}\right]$. It can then be shown that $G_n(\boldsymbol{\Theta},\widehat{\boldsymbol{\pi}}) \xrightarrow{p} G(\boldsymbol{\Theta},\boldsymbol{\pi})$, where $G(\boldsymbol{\Theta},\boldsymbol{\pi}) = \mathrm{E}\left[-\frac{\partial U_{wn}(\boldsymbol{\Theta},\boldsymbol{\pi})}{\sqrt{n}\partial\boldsymbol{\Theta}}\right]$. By condition (B3), the convergence of $G_n(\boldsymbol{\Theta},\widehat{\boldsymbol{\pi}})$ to $G(\boldsymbol{\Theta},\boldsymbol{\pi})$ is uniform in a neighborhood of the true value of $\boldsymbol{\Theta}$. By the Inverse Function Theorem of Foutz (1977) and condition (B2), there exists an unique consistent solution to the estimating equations $U_{wn}(\boldsymbol{\Theta};\widehat{\boldsymbol{\pi}}) = \mathbf{0}$ in a neighborhood of the true value of $\boldsymbol{\Theta}$. It then follows that $\widehat{\boldsymbol{\Theta}}_W$ is a consistent estimator of the $\boldsymbol{\Theta}$.

Next, we derive the asymptotic distribution of $\sqrt{n}(\widehat{\boldsymbol{\Theta}}_W - \boldsymbol{\Theta})$. By a Taylor's series expansion of $U_{wn}(\widehat{\boldsymbol{\Theta}}_W,\widehat{\boldsymbol{\pi}})$ at $(\boldsymbol{\Theta},\widehat{\boldsymbol{\pi}})$, we can have

$$\begin{aligned}\mathbf{0} = U_{wn}(\widehat{\boldsymbol{\Theta}}_W,\widehat{\boldsymbol{\pi}}) &= U_{wn}(\boldsymbol{\Theta},\widehat{\boldsymbol{\pi}}) + \frac{\partial U_{wn}(\boldsymbol{\Theta},\widehat{\boldsymbol{\pi}})}{\partial\boldsymbol{\Theta}^T}(\widehat{\boldsymbol{\Theta}}_W - \boldsymbol{\Theta}) + o_p(1)\\ &= U_n(\boldsymbol{\Theta},\widehat{\boldsymbol{\pi}}) - G_n(\boldsymbol{\Theta},\widehat{\boldsymbol{\pi}})\sqrt{n}(\widehat{\boldsymbol{\Theta}}_W - \boldsymbol{\Theta}) + o_p(1).\end{aligned}$$

Because $G_n(\boldsymbol{\Theta},\widehat{\boldsymbol{\pi}}) \xrightarrow{p} G(\boldsymbol{\Theta},\boldsymbol{\pi})$ and condition (B2), we have

$$\begin{aligned}\sqrt{n}(\widehat{\boldsymbol{\Theta}}_W - \boldsymbol{\Theta}) &= G^{-1}(\boldsymbol{\Theta},\boldsymbol{\pi})U_{wn}(\boldsymbol{\Theta};\widehat{\boldsymbol{\pi}}) + o_p(1)\\ &= \frac{1}{\sqrt{n}}G^{-1}(\boldsymbol{\Theta},\boldsymbol{\pi})\sum_{i=1}^{n}\left[\frac{\delta_i\Psi_i(\boldsymbol{\Theta})}{\pi_i} - \frac{(\delta_i-\pi_i)\Psi_i^*(\boldsymbol{\Theta})}{\pi_i}\right] + o_p\left(1\right).\end{aligned}$$

Let $M(\boldsymbol{\Theta},\boldsymbol{\pi}) = \mathrm{Cov}\left[U_{wn}(\boldsymbol{\Theta};\widehat{\boldsymbol{\pi}})\right] = \mathrm{E}\left\{\left[\frac{\delta_1}{\pi_1}\Psi_1(\boldsymbol{\Theta}) + (1-\frac{\delta_1}{\pi_1})\Psi_1^*(\boldsymbol{\Theta})\right]^{\otimes 2}\right\}$. By using the Central Limit Theorem, one can show $\sqrt{n}(\widehat{\boldsymbol{\Theta}}_W - \boldsymbol{\Theta})$ is asymptotically normally distributed with mean $\mathbf{0}$ and covariance matrix $\Delta_W = G^{-1}(\boldsymbol{\Theta},\boldsymbol{\pi})M(\boldsymbol{\Theta},\boldsymbol{\pi})G^{-T}(\boldsymbol{\Theta},\boldsymbol{\pi})$. The proof is completed.

Finally, the covariance matrix of the IPW estimator $\widehat{\boldsymbol{\Theta}}_W$ may be obtained by using a simple sandwich estimator. Let

$$\mathcal{G}(\widehat{\boldsymbol{\Theta}}_W) = \frac{1}{n}\sum_{i=1}^{n}\frac{\delta_i}{\widehat{\pi}_i}\mathcal{X}_i\mathrm{diag}\left[\boldsymbol{H}_i^{(1)}(\widehat{\boldsymbol{\Theta}}_W)\right]\boldsymbol{W}^T\Sigma_i^{-1}(\widehat{\boldsymbol{\Theta}}_W)\boldsymbol{W}\mathrm{diag}\left[\boldsymbol{H}_i^{(1)}(\widehat{\boldsymbol{\Theta}}_W)\right]\mathcal{X}_i^T.$$

and $\widehat{\Psi}_i^*(\widehat{\mathbf{\Theta}}_W) = \frac{\sum_{k=1}^n \delta_k \Psi_k(\widehat{\mathbf{\Theta}}_W) \boldsymbol{I}(D_k=D_i, \boldsymbol{V}_k=\boldsymbol{V}_i)}{\sum_{r=1}^n \delta_r \boldsymbol{I}(D_r=D_i, \boldsymbol{V}_r=\boldsymbol{V}_i)}$. The covariance matrix of $\widehat{\mathbf{\Theta}}_W$ can then be consistently estimated by

$$\mathcal{G}^{-1}(\widehat{\mathbf{\Theta}}_W) \left\{ \sum_{i=1}^n \frac{1}{n} \left[\frac{\delta_i}{\widehat{\pi}_i} \Psi_i(\widehat{\mathbf{\Theta}}_W) + (1 - \frac{\delta_i}{\widehat{\pi}_i}) \widehat{\Psi}_i^*(\widehat{\mathbf{\Theta}}_W) \right]^{\otimes 2} \right\} \mathcal{G}^{-T}(\widehat{\mathbf{\Theta}}_W)/n.$$

Multiple Imputation Method

The following theorem shows that the two MI estimators and the IPW estimator are all asymptotically equivalent.

Theorem 2 *Under the regularity conditions (B1)-(B3), both $\sqrt{n}(\widehat{\mathbf{\Theta}}_{M_1} - \widehat{\mathbf{\Theta}}_{M_2})$ and $\sqrt{n}(\widehat{\mathbf{\Theta}}_W - \widehat{\mathbf{\Theta}}_{M_1})$ converge in probability to $\mathbf{0}$ as the number of replications M goes to infinity.*

Proof of Theorem 2

Based on the empirical conditional distribution

$$\widehat{F}(x|D_i, \boldsymbol{V}_i) = \frac{\sum\limits_{r=1}^n \delta_r I(D_r = D_i, \boldsymbol{V}_r = \boldsymbol{V}_i) I(X_r \leq x)}{\sum\limits_{k=1}^n \delta_k I(D_k = D_i, \boldsymbol{V}_k = \boldsymbol{V}_i)},$$

we can obtain

$$\mathrm{E}_{\widehat{F}} \left[\widetilde{\Psi}_{qi}(\mathbf{\Theta}) | D_i, \boldsymbol{V}_i \right] = \int \Psi_i(\mathbf{\Theta}) d\widehat{F}(x|D_i, \boldsymbol{V}_i) = \sum_{r=1}^n \frac{\delta_r I(D_r = D_i, \boldsymbol{V}_r = \boldsymbol{V}_i) \Psi_r(\mathbf{\Theta})}{\sum\limits_{k=1}^n \delta_k I(D_k = D_i, \boldsymbol{V}_k = \boldsymbol{V}_i)}.$$

Let $\mathcal{O}$ denote the observed data. We can then have

$$\begin{aligned}
\mathrm{E}_{\widehat{F}} \left[\widetilde{U}_{qn}(\mathbf{\Theta}) | \mathcal{O} \right] &= \frac{1}{\sqrt{n}} \sum_{i=1}^n \left\{ \delta_i \mathrm{E}_{\widehat{F}} \left[\Psi_i(\mathbf{\Theta}) | D_i, X_i, Z_i \right] + (1 - \delta_i) \mathrm{E}_{\widehat{F}} \left[\widetilde{\Psi}_{qi}(\mathbf{\Theta}) | D_i, Z_i \right] \right\} \\
&= \frac{1}{\sqrt{n}} \sum_{i=1}^n \delta_i \Psi_i(\mathbf{\Theta}) + \frac{1}{\sqrt{n}} \sum_{i=1}^n (1 - \delta_i) \sum_{r=1}^n \frac{\delta_r I(D_r = D_i \boldsymbol{V}_r = \boldsymbol{V}_i) \Psi_r(\mathbf{\Theta})}{\sum\limits_{k=1}^n \delta_k I(D_k = D_i, \boldsymbol{V}_k = \boldsymbol{V}_i)} \\
&= \frac{1}{\sqrt{n}} \sum_{i=1}^n \delta_i \Psi_i(\mathbf{\Theta}) + \frac{1}{\sqrt{n}} \sum_{r=1}^n \delta_r \Psi_r(\mathbf{\Theta}) \left(\frac{1}{\widehat{\pi}_r} - 1 \right) \\
&= \frac{1}{\sqrt{n}} \sum_{i=1}^n \frac{\delta_i}{\widehat{\pi}_i} \Psi_i(\mathbf{\Theta}) = U_{wn}(\mathbf{\Theta}, \widehat{\boldsymbol{\pi}}).
\end{aligned}$$

Similarly, it can be obtained that $\mathrm{E}_{\widehat{F}}\left[\frac{\partial \widetilde{U}_{qn}(\mathbf{\Theta})}{\partial \mathbf{\Theta}}|\mathcal{O}\right] = \frac{\partial U_{wn}(\mathbf{\Theta},\widehat{\boldsymbol{\pi}})}{\partial \mathbf{\Theta}}$ and $\mathrm{E}\left[\frac{\partial \widetilde{U}_{qn}(\mathbf{\Theta})}{\partial \mathbf{\Theta}}\right] = \mathrm{E}\left[\frac{\partial U_{wn}(\mathbf{\Theta},\widehat{\boldsymbol{\pi}})}{\partial \mathbf{\Theta}}\right]$.

We first prove the MI_1 and MI_2 estimators are asymptotically equivalent. Define $G(\mathbf{\Theta},\boldsymbol{\pi}) = \mathrm{E}\left[-\frac{\partial U_{wn}(\mathbf{\Theta},\boldsymbol{\pi})}{\sqrt{n}\partial \mathbf{\Theta}}\right]$ and let $\widehat{\mathbf{\Theta}}_q$ be the solution of the equations $\widetilde{U}_{qn}(\mathbf{\Theta}) = \mathbf{0}$. By a Taylor's series expansion of $U_{qn}(\widehat{\mathbf{\Theta}}_q)$ at $\mathbf{\Theta}$, we can then have

$$\begin{aligned}
\mathbf{0} = \widetilde{U}_{qn}(\widehat{\mathbf{\Theta}}_q) &= \widetilde{U}_{qn}(\mathbf{\Theta}) + \frac{\partial \widetilde{U}_{qn}(\mathbf{\Theta})}{\partial \mathbf{\Theta}}(\widehat{\mathbf{\Theta}}_q - \mathbf{\Theta}) + o_p(1)\\
&= \widetilde{U}_{qn}(\mathbf{\Theta}) - G(\mathbf{\Theta},\boldsymbol{\pi})\sqrt{n}(\widehat{\mathbf{\Theta}}_q - \mathbf{\Theta}) + o_p(1)
\end{aligned}$$

to obtain $\sqrt{n}(\widehat{\mathbf{\Theta}}_q - \mathbf{\Theta}) = G^{-1}(\mathbf{\Theta},\boldsymbol{\pi})\widetilde{U}_{qn}(\mathbf{\Theta}) + o_p(1)$. Because $\widehat{\mathbf{\Theta}}_{M_1} = \frac{1}{M}\sum_{q=1}^{M}\widehat{\mathbf{\Theta}}_q$, by using the above result, it follows that

$$\sqrt{n}(\widehat{\mathbf{\Theta}}_{M_1} - \mathbf{\Theta}) = G^{-1}(\mathbf{\Theta},\boldsymbol{\pi})\left[\frac{1}{M}\sum_{q=1}^{M}\widetilde{U}_{qn}(\mathbf{\Theta})\right] + o_p(1).$$

Because $\widehat{\mathbf{\Theta}}_{M_2}$ is the solution of $U_{mn}(\mathbf{\Theta}) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left[\delta_i\Psi_i(\mathbf{\Theta}) + (1-\delta_i)\bar{\Psi}_i(\mathbf{\Theta})\right] = \mathbf{0}$, where $\bar{\Psi}_i(\mathbf{\Theta}) = \frac{\sum_{q=1}^{M}\widetilde{\Psi}_{iq}(\mathbf{\Theta})}{M}$, by a Taylor's series expansion of $U_{mn}(\widehat{\mathbf{\Theta}}_{M_2})$ at $\mathbf{\Theta}$, we can have

$$\begin{aligned}
\mathbf{0} = U_{mn}(\widehat{\mathbf{\Theta}}_{M_2}) &= \left[\frac{1}{M}\sum_{q=1}^{M}\widetilde{U}_{qn}(\widehat{\mathbf{\Theta}}_{M_2})\right]\\
&= \left[\frac{1}{M}\sum_{q=1}^{M}\widetilde{U}_{qn}(\mathbf{\Theta})\right] - G(\mathbf{\Theta},\boldsymbol{\pi})\sqrt{n}(\widehat{\mathbf{\Theta}}_{M_2} - \mathbf{\Theta}) + o_p(1).
\end{aligned}$$

It then follows that $\sqrt{n}(\widehat{\mathbf{\Theta}}_{M_2} - \widehat{\mathbf{\Theta}}_{M_1}) = o_p(1)$. This implies that $\sqrt{n}(\widehat{\mathbf{\Theta}}_{M_2} - \widehat{\mathbf{\Theta}}_{M_1})$ converges in probability to $\mathbf{0}$.

We now show the second MI estimator is asymptotically equivalent to the IPW estimator. Because

$$U_{mn}(\mathbf{\Theta}) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left[\delta_i\Psi_i(\mathbf{\Theta}) + (1-\delta_i)\bar{\Psi}_i(\mathbf{\Theta})\right],$$

we can rewrite $U_{mn}(\mathbf{\Theta})$ as

$$U_{mn}(\mathbf{\Theta}) = U_{wn}(\mathbf{\Theta},\widehat{\boldsymbol{\pi}}) + \frac{1}{\sqrt{n}}\sum_{i=1}^{n}(1-\delta_i)\left\{\bar{\Psi}_i(\mathbf{\Theta}) - \mathrm{E}_{\widehat{F}}\left[\widetilde{\Psi}_{1i}(\mathbf{\Theta})|D_i,\boldsymbol{V}_i\right]\right\}.$$

It follows that $\sqrt{M}\left[\bar{\Psi}_i(\mathbf{\Theta}) - \mathrm{E}_{\widehat{F}}\left(\widetilde{\Psi}_{1i}(\mathbf{\Theta})|D_i,\boldsymbol{V}_i\right)\right]$ converges in distribution to a normal random vector as $M \to \infty$. Therefore, we have $\bar{\Psi}_i(\mathbf{\Theta}) - \mathrm{E}_{\widehat{F}}\left[\widetilde{\Psi}_{1i}(\mathbf{\Theta})|D_i,\boldsymbol{V}_i\right] = O_p\left(\frac{1}{\sqrt{M}}\right)$ and

$$U_{mn}(\mathbf{\Theta}) - U_{wn}(\mathbf{\Theta},\widehat{\boldsymbol{\pi}}) = \frac{1}{\sqrt{n}\sqrt{M}}\sum_{i=1}^{n}(1-\delta_i)\sqrt{M}\left\{\bar{\Psi}_i(\mathbf{\Theta}) - \mathrm{E}_{\widehat{F}}\left[\widetilde{\Psi}_{1i}(\mathbf{\Theta})|D_i,\boldsymbol{V}_i\right]\right\} = O_p\left(\frac{1}{\sqrt{M}}\right).$$

Now let $\widehat{\mathbf{\Theta}}_{M_2}^{(M)}$ be the solution of $U_{mn}^{(M)}(\mathbf{\Theta}) = \mathbf{0}$. It can be obtained that

$$\mathbf{0} = U_{mn}^{(M)}(\widehat{\mathbf{\Theta}}_{M_2}^{(M)}) = U_{mn}^{(M)}(\mathbf{\Theta}) - G(\mathbf{\Theta}, \boldsymbol{\pi})\sqrt{n}(\widehat{\mathbf{\Theta}}_{M_2}^{(M)} - \widehat{\mathbf{\Theta}}) + o_p(1).$$

By using the above equations and $\mathbf{0} = U_{wn}(\widehat{\mathbf{\Theta}}_W, \widehat{\boldsymbol{\pi}}) = U_{wn}(\mathbf{\Theta}, \widehat{\boldsymbol{\pi}}) - G(\mathbf{\Theta}, \boldsymbol{\pi})\sqrt{n}(\widehat{\mathbf{\Theta}}_W - \widehat{\mathbf{\Theta}}) + o_p(1)$, we can have

$$\begin{aligned}\sqrt{n}(\widehat{\mathbf{\Theta}}_{M_2}^{(M)} - \widehat{\mathbf{\Theta}}_W) &= G^{-1}(\mathbf{\Theta}, \boldsymbol{\pi})\left[U_{mn}^{(M)}(\mathbf{\Theta}) - U_{wn}(\mathbf{\Theta}, \widehat{\boldsymbol{\pi}})\right] + o_p(1) \\ &= o_p(1) + O_p\left(\frac{1}{\sqrt{M}}\right).\end{aligned}$$

Therefore, it follows that $\sqrt{n}(\widehat{\mathbf{\Theta}}_{M_2}^{(M)} - \widehat{\mathbf{\Theta}}_W)$ converges in probability to $\mathbf{0}$ as M and n go to infinity. The proof is completed.

Finally, the covariance matrix of $\widehat{\mathbf{\Theta}}_{M_1}$ (Rubin 1987) is estimated by

$$\widehat{\mathrm{Var}}(\widehat{\mathbf{\Theta}}_{M_1}) = \frac{1}{M}\sum_{q=1}^{M}\widehat{V}_q + \left(1 + \frac{1}{M}\right)\frac{\sum_{q=1}^{M}(\widehat{\mathbf{\Theta}}_q - \widehat{\mathbf{\Theta}}_{M_1})(\widehat{\mathbf{\Theta}}_q - \widehat{\mathbf{\Theta}}_{M_1})^T}{M-1}.$$

According to Theorem 2 and using a common linearization technique, we may estimate the covariance matrix of $\widehat{\mathbf{\Theta}}_{M_2}$, denoted by $\widehat{\mathrm{Var}}(\widehat{\mathbf{\Theta}}_{M_2})$, expressed as follows:

$$\mathcal{G}_*^{-1}(\widehat{\mathbf{\Theta}}_{M_2})\left[\frac{1}{Mn}\sum_{q=1}^{M}\sum_{i=1}^{n}U_{qi}(\widehat{\mathbf{\Theta}}_{M_2})U_{qi}^T(\widehat{\mathbf{\Theta}}_{M_2}) + \left(1 + \frac{1}{M}\right)\frac{\sum_{q=1}^{M}\widetilde{U}_{qn}(\widehat{\mathbf{\Theta}}_{M_2})\widetilde{U}_{qn}^T(\widehat{\mathbf{\Theta}}_{M_2})}{M-1}\right]\mathcal{G}_*^{-T}(\widehat{\mathbf{\Theta}}_{M_2}),$$

where $\mathcal{G}_*(\widehat{\mathbf{\Theta}}_{M_2}) = \frac{1}{M}\sum_{q=1}^{M}\frac{-\partial\widetilde{U}_{qn}(\mathbf{\Theta})}{\partial\mathbf{\Theta}}|_{\mathbf{\Theta}=\widehat{\mathbf{\Theta}}_{M_2}}$.

References

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

[2] Rubin, D. B. 1987. *Multiple Imputation for Nonresponse in Surveys.* New York: Wiley.