Annals of the Institute of Statistical Mathematics manuscript No.
(will be inserted by the editor)

Supplementary material for "Semiparametric inference on general functionals of two semicontinuous populations"

Meng Yuan · Chunlin Wang · Boxi Lin · Pengfei Li

This document supplements the paper entitled "Semiparametric inference on general functionals of two semicontinuous populations." It contains the proofs of Theorem 1, Theorem 2, and Corollary 1, and additional simulation results. Section S1 introduces some notation and preparation. Section S2 presents some useful lemmas. The proofs of Theorem 1, Theorem 2, and Corollary 1 are given in Sections S3, S4, and S5, respectively. Section S6 contains an additional simulation study on the impact of misspecification of the basis function in the density ratio model.

S1 Some preparation

Recall that

$$X_{i1}, \cdots, X_{in_i} \sim F_i(x) = \nu_i I(x \geq 0) + (1-\nu_i) I(x > 0) G_i(x), \quad \text{for} \quad i = 0, 1, \tag{S1.1}$$

where $\nu_i \in (0,1)$, n_i is the sample size for sample i, $I(\cdot)$ is an indicator function, and the $G_i(\cdot)$'s are the cumulative distribution functions (CDFs) of the positive observations in sample i. We link $G_0(x)$ and $G_1(x)$ via a density ratio model (DRM):

$$dG_1(x) = \exp\{\alpha + \boldsymbol{\beta}^\top \boldsymbol{q}(x)\} dG_0(x), \tag{S1.2}$$

where $\boldsymbol{q}(x)$ is a prespecified, nontrivial, d-dimensional basis function.

We are interested in estimating linear functionals of $F_0(x)$ and $F_1(x)$, defined as

$$\boldsymbol{\psi}_0 = \int_0^\infty \mathbf{a}(x) dF_0(x) \quad \text{and} \quad \boldsymbol{\psi}_1 = \int_0^\infty \mathbf{a}(x) dF_1(x) \tag{S1.3}$$

for some given function $\mathbf{a}(x)$. To do that, we consider a class of general functionals $\boldsymbol{\psi}$ of length p, defined as

$$\boldsymbol{\psi} = \int_0^\infty \boldsymbol{u}(x; \boldsymbol{\nu}, \boldsymbol{\theta}) dG_0(x), \tag{S1.4}$$

Meng Yuan
Department of Statistics and Actuarial Science, University of Waterloo, 200 University Avenue West, Waterloo, Ontario, Canada N2L 3G1

Chunlin Wang
Department of Statistics, School of Economics, Wang Yanan Institute for Studies in Economics, MOE Key Lab of Econometrics and Fujian Key Lab of Statistics, Xiamen University, 422 Siming South Road, Xiamen, Fujian, China 361005
Corresponding author. E-mail: wangc@xmu.edu.cn

Boxi Lin
Dalla Lana School of Public Health, University of Toronto, Toronto, 155 College Street, Ontario, Canada M5T 3M7

Pengfei Li
Department of Statistics and Actuarial Science, University of Waterloo, 200 University Avenue West, Waterloo, Ontario, Canada N2L 3G1

where $\boldsymbol{\nu}=(\nu_0,\nu_1)^\top$, $\boldsymbol{\theta}=(\alpha,\boldsymbol{\beta}^\top)^\top$, and $\boldsymbol{u}(x;\boldsymbol{\nu},\boldsymbol{\theta})=(u_1(x;\boldsymbol{\nu},\boldsymbol{\theta}),\ldots,u_p(x;\boldsymbol{\nu},\boldsymbol{\theta}))^\top$ is a given $(p\times 1)$-dimensional function. Note that $\boldsymbol{\psi}$ covers $\boldsymbol{\psi}_0$ and $\boldsymbol{\psi}_1$, defined in (S1.3), as special cases. To see this, let

$$\boldsymbol{u}(X;\boldsymbol{\nu},\boldsymbol{\theta})=\begin{pmatrix}\boldsymbol{u}_0(X;\boldsymbol{\nu},\boldsymbol{\theta})\\ \boldsymbol{u}_1(X;\boldsymbol{\nu},\boldsymbol{\theta})\end{pmatrix}=\begin{pmatrix}(1-\nu_0)\mathbf{a}(x)\\ (1-\nu_1)\mathbf{a}(x)\exp\{\alpha+\boldsymbol{\beta}^\top\boldsymbol{q}(x)\}\end{pmatrix}. \tag{S1.5}$$

Next we argue that $\boldsymbol{\psi}=\left(\boldsymbol{\psi}_0^\top,\boldsymbol{\psi}_1^\top\right)^\top$ under the assumption that $\mathbf{a}(0)=\mathbf{0}$.

From model (S1.1), we have $F_i(x)=\nu_i I(x\geq 0)+(1-\nu_i)I(x>0)G_i(x)$ for $i=0,1$. This, together with (S1.3), imply that, for $i=0,1$,

$$\begin{aligned}\boldsymbol{\psi}_i&=\int_0^\infty \mathbf{a}(x)d\{\nu_i I(x\geq 0)+(1-\nu_i)I(x>0)G_i(x)\}\\&=\nu_i\mathbf{a}(0)+(1-\nu_i)\int_0^\infty \mathbf{a}(x)dG_i(x)\\&=(1-\nu_i)\int_0^\infty \mathbf{a}(x)dG_i(x),\end{aligned}$$

where the last step uses the assumption that $\mathbf{a}(0)=\mathbf{0}$. Under the density ratio model in (S1.2), we further have

$$\boldsymbol{\psi}_0=\int_0^\infty(1-\nu_0)\mathbf{a}(x)dG_0(x)\quad\text{and}\quad\boldsymbol{\psi}_1=\int_0^\infty(1-\nu_1)\mathbf{a}(x)\exp\{\alpha+\boldsymbol{\beta}^\top\boldsymbol{q}(x)\}dG_0(x).$$

Hence

$$\boldsymbol{\psi}_0=\int_0^\infty \boldsymbol{u}_0(X;\boldsymbol{\nu},\boldsymbol{\theta})dG_0(x)\quad\text{and}\quad\boldsymbol{\psi}_1=\int_0^\infty \boldsymbol{u}_1(X;\boldsymbol{\nu},\boldsymbol{\theta})dG_0(x),$$

as claimed.

Recall that we let n_{i0} and n_{i1} be the (random) numbers of zero observations and positive observations, respectively, in each sample $i=0,1$. Clearly, $n_i=n_{i0}+n_{i1}$, for $i=0,1$. Without loss of generality, we assume that the first n_{i1} observations in group i, $X_{i1},\cdots,X_{in_{i1}}$, are positive, and the remaining n_{i0} observations are 0. Let n be the total (fixed) sample size, i.e., $n=n_0+n_1$.

The maximum empirical likelihood estimators (MELEs) of $\boldsymbol{\nu}$ and $\boldsymbol{\theta}$ respectively maximize $\ell_0(\boldsymbol{\nu})$ and $\ell_1(\boldsymbol{\theta})$, where

$$\ell_0(\boldsymbol{\nu})=\sum_{i=0}^{1}\log\{\nu_i^{n_{i0}}(1-\nu_i)^{n_{i1}}\}$$

and

$$\ell_1(\boldsymbol{\theta})=-\sum_{i=0}^{1}\sum_{j=1}^{n_{i1}}\log\left\{1+\hat\rho[\exp\{\alpha+\boldsymbol{\beta}^\top\boldsymbol{q}(X_{ij})\}-1]\right\}+\sum_{j=1}^{n_{11}}\{\alpha+\boldsymbol{\beta}^\top\boldsymbol{q}(X_{1j})\}$$

with $\hat\rho=n_{11}/(n_{01}+n_{11})$ being a random variable. That is,

$$\hat\nu=\arg\max_{\boldsymbol{\nu}}\ell_0(\boldsymbol{\nu})\quad\text{and}\quad\hat{\boldsymbol{\theta}}=\arg\max_{\boldsymbol{\theta}}\ell_1(\boldsymbol{\theta}). \tag{S1.6}$$

Note that

$$\sum_{i=0}^{1}\sum_{j=1}^{n_{i1}}\frac{1}{n_{01}+n_{11}}\frac{1}{1+\hat\rho[\exp\{\hat\alpha+\hat{\boldsymbol{\beta}}^\top\boldsymbol{q}(X_{ij})\}-1]}=1, \tag{S1.7}$$

which ensures that the MELE of $G_0(x)$ is a CDF.

For convenience of presentation, we recall and introduce some notation. We use $\boldsymbol{\nu}^*$ and $\boldsymbol{\theta}^*$ to denote the true values of $\boldsymbol{\nu}$ and $\boldsymbol{\theta}$, respectively. Let $\boldsymbol{Q}(x)=(1,\boldsymbol{q}(x)^\top)^\top$ and

$$\begin{gathered}w=n_0/n,\ \Delta^*=w(1-\nu_0^*)+(1-w)(1-\nu_1^*),\ \rho^*=\tfrac{(1-w)(1-\nu_1^*)}{\Delta^*},\\ \omega(x;\boldsymbol{\theta})=\exp\{\boldsymbol{\theta}^\top\boldsymbol{Q}(x)\},\ \omega(x)=\omega(x;\boldsymbol{\theta}^*),\\ h(x)=1+\rho^*\{\omega(x)-1\},\ h_1(x)=\rho^*\omega(x)/h(x),\ h_0(x)=(1-\rho^*)/h(x).\end{gathered}$$

Note that $\omega(\cdot)$, $h(\cdot)$, $h_0(\cdot)$, and $h_1(\cdot)$ depend on $\boldsymbol{\theta}^*$ and/or ρ^* and $h_0(x)+h_1(x)=1$. Henceforth, we use $\sum_{ij}$ to denote summation over the full range of data.

Further, define $\hat{\boldsymbol{\eta}}=(\hat{\boldsymbol{\nu}}^\top,\hat{\rho},\hat{\boldsymbol{\theta}}^\top)^\top$ and $\boldsymbol{\eta}^*=(\boldsymbol{\nu}^{*\top},\rho^*,\boldsymbol{\theta}^{*\top})^\top$. To derive the asymptotic properties, we define an expanded function:

$$
\begin{aligned}
H(\boldsymbol{\nu},\rho,\boldsymbol{\theta}) &= n_{00}\log(\nu_0)+n_{01}\log(1-\nu_0)+n_{10}\log(\nu_1)+n_{11}\log(1-\nu_1)\\
&\quad -\sum_{i=0}^{1}\sum_{j=1}^{n_{i1}}\log\left\{1+\rho[\exp\{\boldsymbol{\theta}^\top\boldsymbol{Q}(X_{ij})\}-1]\right\}+\sum_{j=1}^{n_{11}}\{\boldsymbol{\beta}^\top\boldsymbol{q}(X_{1j})\}. \qquad \text{(S1.8)}
\end{aligned}
$$

By (S1.6), we get

$$
\frac{\partial H(\hat{\boldsymbol{\nu}},\hat{\rho},\hat{\boldsymbol{\theta}})}{\partial\boldsymbol{\nu}}=\mathbf{0}\ \text{and}\ \frac{\partial H(\hat{\boldsymbol{\nu}},\hat{\rho},\hat{\boldsymbol{\theta}})}{\partial\boldsymbol{\theta}}=\mathbf{0}. \qquad \text{(S1.9)}
$$

From (S1.7), we can verify that

$$
\frac{\partial H(\hat{\boldsymbol{\nu}},\hat{\rho},\hat{\boldsymbol{\theta}})}{\partial\rho}=0. \qquad \text{(S1.10)}
$$

Then (S1.9) and (S1.10) together imply that $\hat{\boldsymbol{\eta}}$ satisfies

$$
\frac{\partial H(\hat{\boldsymbol{\eta}})}{\partial\boldsymbol{\eta}}=\mathbf{0}, \qquad \text{(S1.11)}
$$

which serves as the starting point of our proof for $\hat{\boldsymbol{\eta}}$.

Next, we apply the first-order Taylor expansion to $\partial H(\hat{\boldsymbol{\eta}})/\partial\boldsymbol{\eta}$ to find an approximation for $\hat{\boldsymbol{\eta}}$. In this process, the first and second derivatives of $H(\boldsymbol{\nu},\rho,\boldsymbol{\theta})$ play important roles. Their detailed forms are given below.

S1.1 First derivatives of $H(\boldsymbol{\nu},\rho,\boldsymbol{\theta})$

After some calculation, we find the first derivatives of $H(\boldsymbol{\nu},\boldsymbol{\theta},\rho)$ as follows:

$$
\begin{aligned}
\frac{\partial H(\boldsymbol{\nu},\rho,\boldsymbol{\theta})}{\partial\boldsymbol{\nu}} &= \left(\frac{\partial H(\boldsymbol{\nu},\rho,\boldsymbol{\theta})}{\partial\nu_0},\frac{\partial H(\boldsymbol{\nu},\rho,\boldsymbol{\theta})}{\partial\nu_1}\right)^\top=\left(\frac{n_{00}}{\nu_0}-\frac{n_{01}}{1-\nu_0},\frac{n_{10}}{\nu_1}-\frac{n_{11}}{1-\nu_1}\right)^\top,\\
\frac{\partial H(\boldsymbol{\nu},\rho,\boldsymbol{\theta})}{\partial\rho} &= -\sum_{ij}\frac{\omega(X_{ij};\boldsymbol{\theta})-1}{1+\rho\{\omega(X_{ij};\boldsymbol{\theta})-1\}}I(X_{ij}>0),\\
\frac{\partial H(\boldsymbol{\nu},\rho,\boldsymbol{\theta})}{\partial\boldsymbol{\theta}} &= \sum_{j=1}^{n_1}\boldsymbol{Q}(X_{1j})I(X_{1j}>0)-\sum_{ij}\frac{\rho\omega(X_{ij};\boldsymbol{\theta})}{1+\rho\{\omega(X_{ij};\boldsymbol{\theta})-1\}}\boldsymbol{Q}(X_{ij})I(X_{ij}>0).
\end{aligned}
$$

We evaluate the above derivatives at $\boldsymbol{\eta}^*$ and define

$$
\boldsymbol{S}_n=\frac{\partial H(\boldsymbol{\eta}^*)}{\partial\boldsymbol{\eta}}=\begin{pmatrix}\frac{\partial H(\boldsymbol{\eta}^*)}{\partial\boldsymbol{\nu}}\\ \frac{\partial H(\boldsymbol{\eta}^*)}{\partial\rho}\\ \frac{\partial H(\boldsymbol{\eta}^*)}{\partial\boldsymbol{\theta}}\end{pmatrix}=\begin{pmatrix}\boldsymbol{S}_{n,\boldsymbol{\nu}}\\ S_{n,\rho}\\ \boldsymbol{S}_{n,\boldsymbol{\theta}}\end{pmatrix}, \qquad \text{(S1.12)}
$$

where the corresponding entries are

$$
\begin{aligned}
\boldsymbol{S}_{n,\boldsymbol{\nu}} &= \left(\frac{n_{00}}{\nu_0^*}-\frac{n_{01}}{1-\nu_0^*},\frac{n_{10}}{\nu_1^*}-\frac{n_{11}}{1-\nu_1^*}\right)^\top,\\
S_{n,\rho} &= -\sum_{ij}\frac{\omega(X_{ij})-1}{h(X_{ij})}I(X_{ij}>0),\\
\boldsymbol{S}_{n,\boldsymbol{\theta}} &= \sum_{j=1}^{n_1}\boldsymbol{Q}(X_{1j})I(X_{1j}>0)-\sum_{ij}h_1(X_{ij})\boldsymbol{Q}(X_{ij})I(X_{ij}>0).
\end{aligned}
$$

S1.2 Second derivatives of $H(\boldsymbol{\nu}, \rho, \boldsymbol{\theta})$

We next calculate the second derivatives of $H(\boldsymbol{\nu}, \rho, \boldsymbol{\theta})$ and evaluate them at $\boldsymbol{\eta}^*$. This leads to

$$\frac{\partial^2 H(\boldsymbol{\eta}^*)}{\partial\boldsymbol{\eta}\partial\boldsymbol{\eta}^\top} = \begin{pmatrix} \frac{\partial^2 H(\boldsymbol{\eta}^*)}{\partial\boldsymbol{\nu}\partial\boldsymbol{\nu}^\top} & \frac{\partial^2 H(\boldsymbol{\eta}^*)}{\partial\boldsymbol{\nu}\partial\rho} & \frac{\partial^2 H(\boldsymbol{\eta}^*)}{\partial\boldsymbol{\nu}\partial\boldsymbol{\theta}^\top} \\ \frac{\partial^2 H(\boldsymbol{\eta}^*)}{\partial\rho\partial\boldsymbol{\nu}^\top} & \frac{\partial^2 H(\boldsymbol{\eta}^*)}{\partial\rho^2} & \frac{\partial^2 H(\boldsymbol{\eta}^*)}{\partial\rho\partial\boldsymbol{\theta}^\top} \\ \frac{\partial^2 H(\boldsymbol{\eta}^*)}{\partial\boldsymbol{\theta}\partial\boldsymbol{\nu}^\top} & \frac{\partial^2 H(\boldsymbol{\eta}^*)}{\partial\boldsymbol{\theta}\partial\rho} & \frac{\partial^2 H(\boldsymbol{\eta}^*)}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^\top} \end{pmatrix}, \tag{S1.13}$$

where

$$\begin{aligned}
\frac{\partial^2 H(\boldsymbol{\eta}^*)}{\partial\boldsymbol{\nu}\partial\boldsymbol{\nu}^\top} &= \operatorname{diag}\left\{-\frac{n_{00}}{\nu_0^{*2}} - \frac{n_{01}}{(1-\nu_0^*)^2}, -\frac{n_{10}}{\nu_1^{*2}} - \frac{n_{11}}{(1-\nu_1^*)^2}\right\}, \\
\frac{\partial^2 H(\boldsymbol{\eta}^*)}{\partial\rho^2} &= -\sum_{ij} \frac{-\{\omega(X_{ij})-1\}^2}{h(X_{ij})^2} I(X_{ij}>0), \\
\frac{\partial^2 H(\boldsymbol{\eta}^*)}{\partial\boldsymbol{\nu}\partial\rho} &= \left\{\frac{\partial^2 H(\boldsymbol{\eta}^*)}{\partial\rho\partial\boldsymbol{\nu}^\top}\right\}^\top = \mathbf{0}, \\
\frac{\partial^2 H(\boldsymbol{\eta}^*)}{\partial\boldsymbol{\theta}\partial\rho} &= \left\{\frac{\partial^2 H(\boldsymbol{\eta}^*)}{\partial\rho\partial\boldsymbol{\theta}^\top}\right\}^\top = -\sum_{ij} \frac{\omega(X_{ij})}{h(X_{ij})^2} \boldsymbol{Q}(X_{ij}) I(X_{ij}>0), \\
\frac{\partial^2 H(\boldsymbol{\eta}^*)}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^\top} &= -\sum_{ij} h_0(X_{ij}) h_1(X_{ij}) \{\boldsymbol{Q}(X_{ij})\boldsymbol{Q}(X_{ij})^\top\} I(X_{ij}>0), \\
\frac{\partial^2 H(\boldsymbol{\eta}^*)}{\partial\boldsymbol{\nu}\partial\boldsymbol{\theta}^\top} &= \left\{\frac{\partial^2 H(\boldsymbol{\eta}^*)}{\partial\boldsymbol{\theta}\partial\boldsymbol{\nu}^\top}\right\}^\top = \mathbf{0}.
\end{aligned}$$

S2 Some useful lemmas

In the proof of Theorem 1, we need the expectation of $\partial^2 H(\boldsymbol{\eta}^*)/(\partial\boldsymbol{\eta}\partial\boldsymbol{\eta}^\top)$ and the asymptotic property of $\boldsymbol{S}_n$. The following lemma is used to ease the calculation burden in our main proofs.

Lemma 1 *Suppose that f is an arbitrary vector-valued function. Let $E_0(\cdot)$ represent the expectation with respect to G_0 and X refer to a random variable from G_0. Then*

$$E\left\{\sum_{ij} f(X_{ij}) I(X_{ij}>0)\right\} = n\Delta^* E_0\{h(X)f(X)\}.$$

Proof Note that

$$\begin{aligned}
E\left\{\sum_{ij} f(X_{ij}) I(X_{ij}>0)\right\} &= \sum_{i=0}^{1} n_i E\{f(X_{i1}) I(X_{i1}>0)\} \\
&= n_0(1-\nu_0^*) E_0\{f(X)\} + n_1(1-\nu_1^*) E_0\{\omega(X) f(X)\},
\end{aligned}$$

where we use the DRM (S1.2) in the last step. Using the facts that $w = n_0/n$ and $1 - w = n_1/n$, we further have

$$E\left\{\sum_{ij} f(X_{ij}) I(X_{ij}>0)\right\} = nw(1-\nu_0^*) E_0\{f(X)\} + n(1-w)(1-\nu_1^*) E_0\{\omega(X) f(X)\}.$$

Recall the definitions of Δ^* and ρ^*. We then have

$$E\left\{\sum_{ij} f(X_{ij}) I(X_{ij}>0)\right\} = n\Delta^* E_0\{(1-\rho^*) f(X)\} + n\Delta^* E_0[\rho^* \omega(X) f(X)]$$

$$= n\Delta^* E_0\{h(X)f(X)\}.$$

This completes the proof. $\square$

With the help of Lemma 1, we calculate the expectation of $\partial^2 H(\boldsymbol{\eta}^*)/(\partial\boldsymbol{\eta}\partial\boldsymbol{\eta}^\top)$.

Lemma 2 *With the form of $\partial^2 H(\boldsymbol{\eta}^*)/(\partial\boldsymbol{\eta}\partial\boldsymbol{\eta}^\top)$ given in (S1.13), we have*

$$-\frac{1}{n}E\left\{\frac{\partial^2 H(\boldsymbol{\eta}^*)}{\partial\boldsymbol{\eta}\partial\boldsymbol{\eta}^\top}\right\} = \boldsymbol{A} = \begin{pmatrix} \boldsymbol{A}_{\boldsymbol{\nu}} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & -A_\rho & \boldsymbol{A}_{\rho,\boldsymbol{\theta}} \\ \mathbf{0} & \boldsymbol{A}_{\boldsymbol{\theta},\rho} & \boldsymbol{A}_{\boldsymbol{\theta}} \end{pmatrix},$$

where

$$\boldsymbol{A}_{\boldsymbol{\nu}} = diag\left\{\frac{w}{\nu_0^*(1-\nu_0^*)}, \frac{1-w}{\nu_1^*(1-\nu_1^*)}\right\},\ \boldsymbol{A}_{\boldsymbol{\theta}} = \Delta^*(1-\rho^*)E_0\left[h_1(X)\boldsymbol{Q}(X)\boldsymbol{Q}^\top(X)\right],$$

$$A_\rho = \Delta^* E_0\left\{\frac{\{\omega(X)-1\}^2}{h(X)}\right\} = \{\rho^*(1-\rho^*)\}^{-1}\left[\Delta^* - \{\rho^*(1-\rho^*)\}^{-1}\boldsymbol{e}^\top\boldsymbol{A}_{\boldsymbol{\theta}}\boldsymbol{e}\right],$$

$$\boldsymbol{A}_{\boldsymbol{\theta},\rho} = \boldsymbol{A}_{\rho,\boldsymbol{\theta}}^\top = \Delta^* E_0\left\{\frac{\omega(X)}{h(X)}\boldsymbol{Q}(X)\right\} = \{\rho^*(1-\rho^*)\}^{-1}\boldsymbol{A}_{\boldsymbol{\theta}}\boldsymbol{e}$$

with $\boldsymbol{e} = (1, \mathbf{0}_{d\times 1}^\top)^\top$.

Proof Note that $n_{00} \sim \text{Bin}(n_0, \nu_0)$ and $n_{10} \sim \text{Bin}(n_1, \nu_1)$, where "Bin" denotes the binomial distribution. Since $w = n_0/n$ and $1-w = n_1/n$, we can easily show that

$$-\frac{1}{n}E\left\{\frac{\partial^2 H(\boldsymbol{\eta}^*)}{\partial\nu\partial\nu^\top}\right\} = \boldsymbol{A}_{\boldsymbol{\nu}}.$$

Next, we apply Lemma 1 to find the remaining entries of $E\left\{\partial^2 H(\boldsymbol{\eta}^*)/(\partial\boldsymbol{\eta}\partial\boldsymbol{\eta}^\top)\right\}$. We use

$$E\left\{\frac{\partial^2 H(\boldsymbol{\eta}^*)}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^\top}\right\}$$

as an illustration. For the other entries, the idea is similar and we omit the details.

Note that

$$\begin{aligned}-\frac{1}{n}E\left\{\frac{\partial^2 H(\boldsymbol{\eta}^*)}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^\top}\right\} &= \frac{1}{n}E\left\{\sum_{ij} h_0(X_{ij})h_1(X_{ij})\boldsymbol{Q}(X_{ij})\boldsymbol{Q}(X_{ij})^\top I(X_{ij}>0)\right\} \\ &= \Delta^* E_0\left\{h(X)h_0(X)h_1(X)\boldsymbol{Q}(X)\boldsymbol{Q}(X)^\top\right\} \\ &= \Delta^*(1-\rho^*)E_0\left\{h_1(X)\boldsymbol{Q}(X)\boldsymbol{Q}(X)^\top\right\},\end{aligned}$$

where we have used Lemma 1 in the second step and the fact that $h(x)h_0(x) = 1-\rho^*$ in the third step. This completes the proof. $\square$

We now study the asymptotic properties of $\boldsymbol{S}_n$ defined in (S1.12). Recall that $\boldsymbol{W} = \left((1-\nu_0^*)^{-1}, -(1-\nu_1^*)^{-1}\right)$ and define $S = w^{-1} + (1-w)^{-1}$.

Lemma 3 *With the form of $\boldsymbol{S}_n$ in (S1.12), as $n\to\infty$*

$$n^{-1/2}\boldsymbol{S}_n \xrightarrow{d} N(\mathbf{0}, \boldsymbol{B}),$$

where

$$\boldsymbol{B} = \begin{pmatrix} \boldsymbol{A}_{\boldsymbol{\nu}} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & A_\rho & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \boldsymbol{A}_{\boldsymbol{\theta}} \end{pmatrix} + \begin{pmatrix} \mathbf{0} & -\rho^*(1-\rho^*)A_\rho\boldsymbol{W}^\top & \boldsymbol{W}^\top\boldsymbol{e}^\top\boldsymbol{A}_{\boldsymbol{\theta}} \\ -\rho^*(1-\rho^*)A_\rho\boldsymbol{W} & -S\{\rho^*(1-\rho^*)\}^2A_\rho^2 & S\rho^*(1-\rho^*)A_\rho\boldsymbol{e}^\top\boldsymbol{A}_{\boldsymbol{\theta}} \\ \boldsymbol{A}_{\boldsymbol{\theta}}\boldsymbol{e}\boldsymbol{W} & S\rho^*(1-\rho^*)A_\rho\boldsymbol{A}_{\boldsymbol{\theta}}\boldsymbol{e} & -S\boldsymbol{A}_{\boldsymbol{\theta}}\boldsymbol{e}(\boldsymbol{A}_{\boldsymbol{\theta}}\boldsymbol{e})^\top \end{pmatrix}.$$

Proof Using the results in Lemma 1, it is easy to show that $E(\boldsymbol{S}_n) = \mathbf{0}$; we omit the details.

Next, we verify that $\mathrm{Var}(\boldsymbol{S}_n) = \boldsymbol{B}$. For convenience, we write $\boldsymbol{B}$ as

$$\boldsymbol{B} = \begin{pmatrix} \boldsymbol{B}_{11} & \boldsymbol{B}_{12} & \boldsymbol{B}_{13} \\ \boldsymbol{B}_{21} & B_{22} & \boldsymbol{B}_{23} \\ \boldsymbol{B}_{31} & \boldsymbol{B}_{32} & \boldsymbol{B}_{33} \end{pmatrix}.$$

We concentrate on deriving $\boldsymbol{B}_{13}$; the other entries can be similarly obtained and we omit the details.

Note that $\boldsymbol{S}_{n,\boldsymbol{\nu}}$ and $\boldsymbol{S}_{n,\boldsymbol{\theta}}$ can be rewritten as

$$\begin{aligned}
\boldsymbol{S}_{n,\nu_0} &= \frac{n_{00}}{\nu_0^*} - \frac{n_{01}}{1-\nu_0^*} = -\frac{n_{01}}{\nu_0^*(1-\nu_0^*)} = -\frac{1}{\nu_0^*(1-\nu_0^*)}\sum_{j=1}^{n_0} I(X_{0j} > 0), \\
\boldsymbol{S}_{n,\nu_1} &= \frac{n_{10}}{\nu_1^*} - \frac{n_{11}}{1-\nu_1^*} = -\frac{n_{11}}{\nu_1^*(1-\nu_1^*)} = -\frac{1}{\nu_1^*(1-\nu_1^*)}\sum_{j=1}^{n_1} I(X_{1j} > 0), \\
\boldsymbol{S}_{n,\boldsymbol{\theta}} &= \sum_{j=1}^{n_1} \boldsymbol{Q}(X_{1j}) I(X_{1j} > 0) - \sum_{ij} h_1(X_{ij}) \boldsymbol{Q}(X_{ij}) I(X_{ij} > 0) \\
&= \sum_{j=1}^{n_1} h_0(X_{1j}) \boldsymbol{Q}(X_{1j}) I(X_{1j} > 0) - \sum_{j=1}^{n_0} h_1(X_{0j}) \boldsymbol{Q}(X_{0j}) I(X_{0j} > 0).
\end{aligned}$$

Then we have

$$\begin{aligned}
\frac{1}{n}\mathrm{Cov}(\boldsymbol{S}_{n,\nu_0}, \boldsymbol{S}_{n,\boldsymbol{\theta}}^\top) &= \frac{1}{n\nu_0^*(1-\nu_0^*)}\mathrm{Cov}\left\{\sum_{j=1}^{n_0} I(X_{0j} > 0), \sum_{j=1}^{n_0} h_1(X_{0j}) \boldsymbol{Q}(X_{0j})^\top I(X_{0j} > 0)\right\} \\
&= \frac{n_0}{n\nu_0^*(1-\nu_0^*)}\left[(1-\nu_0^*)E_0\left\{h_1(X)\boldsymbol{Q}(X)^\top\right\} - (1-\nu_0^*)^2 E_0\left\{h_1(X)\boldsymbol{Q}(X)^\top\right\}\right] \\
&= wE_0\left\{h_1(X)\boldsymbol{Q}(X)^\top\right\} \\
&= (1-\nu_0^*)^{-1}(\boldsymbol{A}_{\boldsymbol{\theta}}\boldsymbol{e})^\top.
\end{aligned}$$

Similarly,

$$\begin{aligned}
\frac{1}{n}\mathrm{Cov}(\boldsymbol{S}_{n,\nu_1}, \boldsymbol{S}_{n,\boldsymbol{\theta}}^\top) &= \frac{-1}{n\nu_1^*(1-\nu_1^*)}\mathrm{Cov}\left\{\sum_{j=1}^{n_1} I(X_{1j} > 0), \sum_{j=1}^{n_1} h_0(X_{1j}) \boldsymbol{Q}(X_{1j})^\top I(X_{1j} > 0)\right\} \\
&= \frac{-n_1}{n\nu_1^*(1-\nu_1^*)}\left[(1-\nu_1^*)E_0\left\{h_0(X)\omega(X)\boldsymbol{Q}(X)^\top\right\} - (1-\nu_1^*)^2 E_0\left\{h_0(X)\omega(X)\boldsymbol{Q}(X)^\top\right\}\right] \\
&= -(1-w)E_0\left\{\omega(X)h_0(X)\boldsymbol{Q}(X)^\top\right\} \\
&= -(1-w)\cdot\frac{1-\rho^*}{\rho^*}E_0\left\{h_1(X)\boldsymbol{Q}(X)^\top\right\} \\
&= -(1-\nu_1^*)^{-1}(\boldsymbol{A}_{\boldsymbol{\theta}}\boldsymbol{e})^\top.
\end{aligned}$$

Recall that $\boldsymbol{W} = \left((1-\nu_0^*)^{-1}, -(1-\nu_1^*)^{-1}\right)$. Then $\boldsymbol{B}_{13} = \boldsymbol{W}^\top \boldsymbol{e}^\top \boldsymbol{A}_{\boldsymbol{\theta}}$.

Note that $\boldsymbol{S}_n$ in (S1.12) is a sum of independent random vectors. Therefore, by the classical central limit theorem, we have as $n \to \infty$

$$n^{-1/2}\boldsymbol{S}_n \xrightarrow{d} N(\mathbf{0}, \boldsymbol{B}),$$

which completes the proof. □

S3 Proof of Theorem 1

With the preparation in Sections 1 and 2, we now move to the proof of Theorem 1.

Recall that $\hat{\boldsymbol{\eta}}$ satisfies

$$\frac{\partial H(\hat{\boldsymbol{\eta}})}{\partial \boldsymbol{\eta}} = \mathbf{0}.$$

Applying the first-order Taylor expansion to $\partial H(\hat{\boldsymbol{\eta}})/\partial \boldsymbol{\eta}$, and using (S1.12) and Lemma 2, we have

$$\begin{aligned}
\mathbf{0} &= \frac{\partial H(\boldsymbol{\eta}^*)}{\partial \boldsymbol{\eta}} + \left(\frac{\partial^2 H(\boldsymbol{\eta}^*)}{\partial \boldsymbol{\eta} \partial \boldsymbol{\eta}^\top}\right)(\hat{\boldsymbol{\eta}} - \boldsymbol{\eta}^*) + o_p(n^{1/2}) \\
&= \boldsymbol{S}_n - n\boldsymbol{A}(\hat{\boldsymbol{\eta}} - \boldsymbol{\eta}^*) + o_p(n^{1/2}).
\end{aligned}$$

Conditions C1–C4 in the main paper ensure that the matrix $\boldsymbol{A}$ is positive definite. Hence, we obtain an approximation for $\hat{\boldsymbol{\eta}} - \boldsymbol{\eta}^*$ as

$$\hat{\boldsymbol{\eta}} - \boldsymbol{\eta}^* = \begin{pmatrix} \hat{\boldsymbol{\nu}} - \boldsymbol{\nu}^* \\ \hat{\rho} - \rho^* \\ \hat{\boldsymbol{\theta}} - \boldsymbol{\theta}^* \end{pmatrix} = \frac{1}{n}\boldsymbol{A}^{-1}\boldsymbol{S}_n + o_p(n^{-1/2}). \tag{S3.14}$$

This together with the asymptotic property of $\boldsymbol{S}_n$ in Lemma 3 and Slutsky's theorem gives

$$n^{1/2}(\hat{\boldsymbol{\eta}} - \boldsymbol{\eta}^*) \xrightarrow{d} N(\mathbf{0}, \boldsymbol{A}^{-1}\boldsymbol{B}\boldsymbol{A}^{-1}),$$

as $n \to \infty$.

To find the explicit form of $\boldsymbol{A}^{-1}\boldsymbol{B}\boldsymbol{A}^{-1}$, we first identify the structure of $\boldsymbol{A}^{-1}$. We write

$$\begin{pmatrix} -A_\rho & \boldsymbol{A}_{\rho,\boldsymbol{\theta}} \\ \boldsymbol{A}_{\boldsymbol{\theta},\rho} & \boldsymbol{A}_{\boldsymbol{\theta}} \end{pmatrix}^{-1} = \begin{pmatrix} A^{11} & \boldsymbol{A}^{12} \\ \boldsymbol{A}^{21} & \boldsymbol{A}^{22} \end{pmatrix}.$$

Using the formula for the inverse of a 2×2 block matrix, we have

$$\begin{aligned}
A^{11} &= \{-A_\rho - (\boldsymbol{A}_{\rho,\boldsymbol{\theta}})\boldsymbol{A}_{\boldsymbol{\theta}}^{-1}(\boldsymbol{A}_{\boldsymbol{\theta},\rho})\}^{-1} \\
&= \left[\{\rho^*(1-\rho^*)\}^{-2}\boldsymbol{e}^\top \boldsymbol{A}_{\boldsymbol{\theta}}\boldsymbol{e} - \Delta^*\{\rho^*(1-\rho^*)\}^{-1} - \{\rho^*(1-\rho^*)\}^{-2}\boldsymbol{e}^\top \boldsymbol{A}_{\boldsymbol{\theta}}\boldsymbol{e}\right]^{-1} \\
&= -\frac{\rho^*(1-\rho^*)}{\Delta}, \\
\boldsymbol{A}^{12} &= (\boldsymbol{A}^{21})^\top = -A^{11}(\boldsymbol{A}_{\rho,\boldsymbol{\theta}})\boldsymbol{A}_{\boldsymbol{\theta}}^{-1} = \frac{\boldsymbol{e}^\top}{\Delta^*}, \\
\boldsymbol{A}^{22} &= \boldsymbol{A}_{\boldsymbol{\theta}}^{-1} + \boldsymbol{A}_{\boldsymbol{\theta}}^{-1}(\boldsymbol{A}_{\boldsymbol{\theta},\rho})A^{11}(\boldsymbol{A}_{\rho,\boldsymbol{\theta}})\boldsymbol{A}_{\boldsymbol{\theta}}^{-1} = \boldsymbol{A}_{\boldsymbol{\theta}}^{-1} - \frac{\boldsymbol{e}\boldsymbol{e}^\top}{\Delta^*\rho^*(1-\rho^*)}.
\end{aligned}$$

Hence, $\boldsymbol{A}^{-1}$ is given by

$$\boldsymbol{A}^{-1} = \begin{pmatrix} \boldsymbol{A}_{\boldsymbol{\nu}}^{-1} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & -\frac{\rho^*(1-\rho^*)}{\Delta^*} & \frac{\boldsymbol{e}^\top}{\Delta^*} \\ \mathbf{0} & \frac{\boldsymbol{e}}{\Delta^*} & \boldsymbol{A}_{\boldsymbol{\theta}}^{-1} - \frac{\boldsymbol{e}\boldsymbol{e}^\top}{\Delta^*\rho^*(1-\rho^*)} \end{pmatrix}. \tag{S3.15}$$

With the form of $\boldsymbol{A}^{-1}$ in (S3.15) and the form of $\boldsymbol{B}$ in Lemma 3, after some tedious algebra, we find that

$$\boldsymbol{\Lambda} = \boldsymbol{A}^{-1}\boldsymbol{B}\boldsymbol{A}^{-1} = \begin{pmatrix} \boldsymbol{A}_{\boldsymbol{\nu}}^{-1} & \rho^*(1-\rho^*)\boldsymbol{A}_{\boldsymbol{\nu}}^{-1}\boldsymbol{W}^\top & \mathbf{0} \\ \rho^*(1-\rho^*)\boldsymbol{W}\boldsymbol{A}_{\boldsymbol{\nu}}^{-1} & \rho^*(1-\rho^*)\{\frac{1}{\Delta^*} - S\rho^*(1-\rho^*)\} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \boldsymbol{A}_{\boldsymbol{\theta}}^{-1} - \frac{\boldsymbol{e}\boldsymbol{e}^\top}{\Delta^*\rho^*(1-\rho^*)} \end{pmatrix}.$$

Recall that $S = w^{-1} + (1-w)^{-1}$. Some algebra leads to

$$\frac{1}{\Delta^*} - S\rho^*(1-\rho^*) = \frac{1}{\Delta^*}\{\rho^*\nu_0^* + (1-\rho^*)\nu_1^*\}.$$

This completes the proof of Theorem 1. $\square$

S4 Proof of Theorem 2

Recall that we are interested in a class of general parameter vectors $\boldsymbol{\psi}$ of length p defined as

$$\boldsymbol{\psi} = \int_0^{\infty} \boldsymbol{u}(x;\boldsymbol{\nu},\boldsymbol{\theta}) dG_0(x),$$

where $\boldsymbol{u}(x;\boldsymbol{\nu},\boldsymbol{\theta}) = (u_1(x;\boldsymbol{\nu},\boldsymbol{\theta}),\dots,u_p(x;\boldsymbol{\nu},\boldsymbol{\theta}))^{\top}$ is a given $(p\times 1)$-dimensional function. The MELE of $\boldsymbol{\psi}$ is given by

$$\begin{aligned}
\hat{\boldsymbol{\psi}} &= \sum_{i=0}^{1}\sum_{j=1}^{n_{i1}} \hat{p}_{ij}\boldsymbol{u}(X_{ij};\hat{\boldsymbol{\nu}},\hat{\boldsymbol{\theta}})\\
&= \frac{1}{n_{01}+n_{11}} \sum_{i=0}^{1}\sum_{j=1}^{n_{i1}} \frac{\boldsymbol{u}(X_{ij};\hat{\boldsymbol{\nu}},\hat{\boldsymbol{\theta}})}{1+\hat{\rho}[\exp\{\hat{\alpha}+\hat{\boldsymbol{\beta}}^{\top}\boldsymbol{q}(X_{ij})\}-1]}\\
&= \frac{1}{nw(1-\hat{\nu}_0)+n(1-w)(1-\hat{\nu}_1)} \sum_{ij} \frac{\boldsymbol{u}(X_{ij};\hat{\boldsymbol{\nu}},\hat{\boldsymbol{\theta}})}{1+\hat{\rho}[\exp\{\hat{\alpha}+\hat{\boldsymbol{\beta}}^{\top}\boldsymbol{q}(X_{ij})\}-1]} I(X_{ij}>0).
\end{aligned}$$

Note that $\hat{\boldsymbol{\psi}}$ is a function of $\hat{\boldsymbol{\eta}}$, so we write $\hat{\boldsymbol{\psi}}$ as $\hat{\boldsymbol{\psi}}(\hat{\boldsymbol{\eta}})$. From Theorem 1, we have $\hat{\boldsymbol{\eta}} = \boldsymbol{\eta}^* + O_p(n^{-1/2})$. Applying the first-order Taylor expansion to $\hat{\boldsymbol{\psi}}(\hat{\boldsymbol{\eta}})$, we get

$$\hat{\boldsymbol{\psi}} = \hat{\boldsymbol{\psi}}(\boldsymbol{\eta}^*) + \left(\frac{\partial \hat{\boldsymbol{\psi}}(\boldsymbol{\eta}^*)}{\partial \boldsymbol{\eta}}\right)(\hat{\boldsymbol{\eta}}-\boldsymbol{\eta}^*) + o_p(n^{-1/2}).$$

For convenience, we write $\boldsymbol{u}(x) = \boldsymbol{u}(x;\boldsymbol{\nu}^*,\boldsymbol{\theta}^*)$. Note that

$$\frac{\partial \hat{\boldsymbol{\psi}}(\boldsymbol{\eta}^*)}{\partial \boldsymbol{\eta}} = \left(\frac{\partial \hat{\boldsymbol{\psi}}(\boldsymbol{\eta}^*)}{\partial \boldsymbol{\nu}}, \frac{\partial \hat{\boldsymbol{\psi}}(\boldsymbol{\eta}^*)}{\partial \rho}, \frac{\partial \hat{\boldsymbol{\psi}}(\boldsymbol{\eta}^*)}{\partial \boldsymbol{\theta}}\right)$$

where

$$\begin{aligned}
\frac{\partial \hat{\boldsymbol{\psi}}(\boldsymbol{\eta}^*)}{\partial \boldsymbol{\nu}} &= \frac{1}{n\Delta^{*2}} \sum_{ij} \left\{ \frac{\partial \boldsymbol{u}(X_{ij};\boldsymbol{\nu}^*,\boldsymbol{\theta}^*)/\partial\boldsymbol{\nu}}{h(X_{ij})}\Delta^* + (w,1-w)\otimes \frac{\boldsymbol{u}(X_{ij})}{h(X_{ij})}\right\} I(X_{ij}>0),\\
\frac{\partial \hat{\boldsymbol{\psi}}(\boldsymbol{\eta}^*)}{\partial \rho} &= -\frac{1}{n\Delta^*}\sum_{ij} \frac{\boldsymbol{u}(X_{ij})\{\omega(X_{ij})-1\}}{h(X_{ij})^2} I(X_{ij}>0),\\
\frac{\partial \hat{\boldsymbol{\psi}}(\boldsymbol{\eta}^*)}{\partial \boldsymbol{\theta}} &= \frac{1}{n\Delta^*}\sum_{ij} \frac{\{\partial \boldsymbol{u}(X_{ij};\boldsymbol{\nu}^*,\boldsymbol{\theta}^*)/\partial\boldsymbol{\theta}\}\cdot h(X_{ij}) - \boldsymbol{u}(X_{ij})\rho^*\omega(X_{ij})\boldsymbol{Q}(X)^{\top}}{h(X_{ij})^2} I(X_{ij}>0),
\end{aligned}$$

where $\otimes$ indicates the Kronecker product. By the law of large numbers and Lemma 1, we have

$$\frac{\partial \hat{\boldsymbol{\psi}}(\boldsymbol{\eta}^*)}{\partial \boldsymbol{\eta}} \xrightarrow{p} \boldsymbol{C}$$

as $n\to\infty$, where $\boldsymbol{C} = (\boldsymbol{C}_{\boldsymbol{\nu}}, \boldsymbol{C}_{\rho}, \boldsymbol{C}_{\boldsymbol{\theta}})$ with

$$\begin{aligned}
\boldsymbol{C}_{\boldsymbol{\nu}} &= E_0\left\{\frac{\partial \boldsymbol{u}(X;\boldsymbol{\nu}^*,\boldsymbol{\theta}^*)}{\partial \boldsymbol{\nu}}\right\} + (w,1-w)\otimes\frac{\boldsymbol{\psi}^*}{\Delta^*},\\
\boldsymbol{C}_{\rho} &= -E_0\left\{\frac{\boldsymbol{u}(X)\{\omega(X)-1\}}{h(X)}\right\} = \frac{\rho^*\boldsymbol{\psi}^* - E_0\{h_1(X)\boldsymbol{u}(X)\}}{\rho^*(1-\rho^*)},\\
\boldsymbol{C}_{\boldsymbol{\theta}} &= E_0\left[\frac{\{\partial \boldsymbol{u}(X;\boldsymbol{\nu}^*,\boldsymbol{\theta}^*)/\partial\boldsymbol{\theta}\}\cdot h(X) - \boldsymbol{u}(X)\rho^*\omega(X)\boldsymbol{Q}(X)^{\top}}{h(X)}\right]\\
&= E_0\left\{\frac{\partial \boldsymbol{u}(X;\boldsymbol{\nu}^*,\boldsymbol{\theta}^*)}{\partial \boldsymbol{\theta}}\right\} - E_0\left\{h_1(X)\boldsymbol{u}(X)\boldsymbol{Q}(X)^{\top}\right\}.
\end{aligned}$$

For convenience, we let

$$\mathbf{E}_{0\boldsymbol{u}} = E_0\{h_0(X)\boldsymbol{u}(X)\} \quad \text{and} \quad \mathbf{E}_{1\boldsymbol{u}} = E_0\{h_1(X)\boldsymbol{u}(X)\}.$$

Then $\mathbf{E}_{0\boldsymbol{u}} + \mathbf{E}_{1\boldsymbol{u}} = \boldsymbol{\psi}^*$ and

$$\boldsymbol{C}_\rho = \frac{\rho^*\boldsymbol{\psi}^* - \mathbf{E}_{1\boldsymbol{u}}}{\rho^*(1-\rho^*)}.$$

Recall from (S3.14) that $\hat{\boldsymbol{\eta}} - \boldsymbol{\eta}^* = n^{-1}\boldsymbol{A}^{-1}\boldsymbol{S}_n + o_p(n^{-1/2})$. Therefore, as $n \to \infty$, $n^{1/2}(\hat{\boldsymbol{\psi}} - \boldsymbol{\psi}^*)$ has the same limiting distribution as

$$n^{1/2}\left[\left\{\hat{\boldsymbol{\psi}}(\boldsymbol{\eta}^*) - \boldsymbol{\psi}^*\right\} + \boldsymbol{C}\boldsymbol{A}^{-1}\boldsymbol{S}_n/n\right]. \tag{S4.16}$$

It can easily be verified that (S4.16) has expectation zero. We will now decompose its asymptotic variance into three parts.

Note that the first term of (S4.16) involves

$$\hat{\boldsymbol{\psi}}(\boldsymbol{\eta}^*) = \frac{1}{n\Delta^*}\sum_{ij}\frac{\boldsymbol{u}(X_{ij})}{h(X_{ij})}I(X_{ij} > 0).$$

Then the variance of the first term in (S4.16) is

$$\begin{aligned}
\boldsymbol{\Gamma}_1 &= \frac{1}{\Delta^*}E_0\left\{\frac{\boldsymbol{u}(X)\boldsymbol{u}(X)^\top}{h(X)}\right\} - \frac{1}{w}E_0\{h_0(X)\boldsymbol{u}(X)\}E_0\{h_0(X)\boldsymbol{u}(X)^\top\} \\
&\quad - \frac{1}{1-w}E_0\{h_1(X)\boldsymbol{u}(X)\}E_0\{h_1(X)\boldsymbol{u}(X)^\top\} \\
&= \frac{1}{\Delta^*}E_0\left\{\frac{\boldsymbol{u}(X)\boldsymbol{u}(X)^\top}{h(X)}\right\} - \frac{1}{w}\mathbf{E}_{0\boldsymbol{u}}\mathbf{E}_{0\boldsymbol{u}}^\top - \frac{1}{1-w}\mathbf{E}_{1\boldsymbol{u}}\mathbf{E}_{1\boldsymbol{u}}^\top,
\end{aligned} \tag{S4.17}$$

where in the first step we have used the results in Lemma 1, and in the second step we have used the definitions of $\mathbf{E}_{0\boldsymbol{u}}$ and $\mathbf{E}_{1\boldsymbol{u}}$.

Next, we derive the variance of the second term in (S4.16):

$$\boldsymbol{\Gamma}_2 = n\mathrm{Var}(\boldsymbol{C}\boldsymbol{A}^{-1}\boldsymbol{S}_n/n) = \boldsymbol{C}\boldsymbol{\Lambda}\boldsymbol{C}^\top.$$

Together with the form of $\boldsymbol{\Lambda}$ in Theorem 1, we have

$$\begin{aligned}
\boldsymbol{\Gamma}_2 &= \boldsymbol{C}_{\boldsymbol{\nu}}\boldsymbol{A}_{\boldsymbol{\nu}}^{-1}\boldsymbol{C}_{\boldsymbol{\nu}}^\top + \rho^*(1-\rho^*)\boldsymbol{C}_{\boldsymbol{\nu}}\boldsymbol{A}_{\boldsymbol{\nu}}^{-1}\boldsymbol{W}^\top\boldsymbol{C}_\rho^\top + \rho^*(1-\rho^*)\boldsymbol{C}_\rho\boldsymbol{W}\boldsymbol{A}_{\boldsymbol{\nu}}^{-1}\boldsymbol{C}_{\boldsymbol{\nu}}^\top \\
&\quad + (\Delta^*)^{-1}\rho^*(1-\rho^*)\{\rho^*\nu_0^* + (1-\rho^*)\nu_1^*\}\boldsymbol{C}_\rho\boldsymbol{C}_\rho^\top + \boldsymbol{C}_{\boldsymbol{\theta}}\left\{\boldsymbol{A}_{\boldsymbol{\theta}}^{-1} - \frac{\boldsymbol{e}\boldsymbol{e}^\top}{\Delta^*\rho^*(1-\rho^*)}\right\}\boldsymbol{C}_{\boldsymbol{\theta}}^\top.
\end{aligned}$$

Note that

$$\boldsymbol{W}\boldsymbol{A}_{\boldsymbol{\nu}}^{-1}\boldsymbol{W}^\top = \frac{\rho^*\nu_0^* + (1-\rho^*)\nu_1^*}{\Delta^*\rho^*(1-\rho^*)}.$$

Then

$$\begin{aligned}
\boldsymbol{\Gamma}_2 &= \{\boldsymbol{C}_{\boldsymbol{\nu}} + \rho^*(1-\rho^*)\boldsymbol{C}_\rho\boldsymbol{W}\}\boldsymbol{A}_{\boldsymbol{\nu}}^{-1}\{\boldsymbol{C}_{\boldsymbol{\nu}} + \rho^*(1-\rho^*)\boldsymbol{C}_\rho\boldsymbol{W}\}^\top \\
&\quad - \frac{1}{\Delta^*\rho^*(1-\rho^*)}(\boldsymbol{C}_{\boldsymbol{\theta}}\boldsymbol{e})(\boldsymbol{C}_{\boldsymbol{\theta}}\boldsymbol{e})^\top + \boldsymbol{C}_{\boldsymbol{\theta}}\boldsymbol{A}_{\boldsymbol{\theta}}^{-1}\boldsymbol{C}_{\boldsymbol{\theta}}^\top.
\end{aligned} \tag{S4.18}$$

Lastly, we derive the covariance of the two terms in (S4.16). That is,

$$\boldsymbol{\Gamma}_3 = n\mathrm{Cov}[\boldsymbol{\psi}(\boldsymbol{\eta}^*), n^{-1}\{\boldsymbol{C}\boldsymbol{A}^{-1}\boldsymbol{S}_n\}^\top] = \mathrm{Cov}\{\boldsymbol{\psi}(\boldsymbol{\eta}^*), \boldsymbol{S}_n^\top\}\boldsymbol{A}^{-1}\boldsymbol{C}^\top.$$

For convenience, we write $\mathrm{Cov}\{\boldsymbol{\psi}(\boldsymbol{\eta}^*), \boldsymbol{S}_n^\top\} = (\boldsymbol{D}_{\boldsymbol{\nu}}, \boldsymbol{D}_\rho, \boldsymbol{D}_{\boldsymbol{\theta}})$.

We first look at

$$\mathrm{Cov}\{\boldsymbol{\psi}(\boldsymbol{\eta}^*), \boldsymbol{S}_{n,\nu_1}\}$$

$$
\begin{aligned}
&= \mathrm{Cov}\left\{\frac{1}{n\Delta^*}\sum_{ij}\frac{\boldsymbol{u}(X_{ij})}{h(X_{ij})}I(X_{ij}>0), \frac{-n_{11}}{\nu_1^*(1-\nu_1^*)}\right\}\\
&= \frac{-1}{n\Delta^*\nu_1^*(1-\nu_1^*)}\mathrm{Cov}\left\{\sum_{j=1}^{n_1}\frac{\boldsymbol{u}(X_{1j})}{h(X_{1j})}I(X_{1j}>0), \sum_{j=1}^{n_1}I(X_{1j}>0)\right\}\\
&= \frac{-n_1}{n\Delta^*\nu_1^*(1-\nu_1^*)}\left[(1-\nu_1^*)E_0\left\{\frac{\boldsymbol{u}(X)\omega(X)}{h(X)}\right\} - (1-\nu_1^*)^2E_0\left\{\frac{\boldsymbol{u}(X)\omega(X)}{h(X)}\right\}\right]\\
&= \frac{-(1-w)}{\Delta^*}E_0\left\{\frac{\boldsymbol{u}(X)\omega(X)}{h(X)}\right\}\\
&= -(1-\nu_1^*)^{-1}\mathbf{E}_{1\boldsymbol{u}}.
\end{aligned}
$$

Similarly, we find

$$
\mathrm{Cov}\{\boldsymbol{\psi}(\boldsymbol{\eta}^*), \boldsymbol{S}_{n,\nu_0}\} = -(1-\nu_0^*)^{-1}\mathbf{E}_{0\boldsymbol{u}}.
$$

Hence,

$$
\boldsymbol{D}_{\boldsymbol{\nu}} = \left(-(1-\nu_0^*)^{-1}\mathbf{E}_{0\boldsymbol{u}}, -(1-\nu_1^*)^{-1}\mathbf{E}_{1\boldsymbol{u}}\right).
$$

We can find $\boldsymbol{D}_\rho$ and $\boldsymbol{D}_{\boldsymbol{\theta}}$ in a similar manner. For $\boldsymbol{D}_\rho$,

$$
\begin{aligned}
\boldsymbol{D}_\rho &= \mathrm{Cov}\{\boldsymbol{\psi}(\boldsymbol{\eta}^*), S_{n,\rho}\}\\
&= -\frac{1}{n\Delta^*}\mathrm{Cov}\left\{\sum_{ij}\frac{\boldsymbol{u}(X_{ij})}{h(X_{ij})}I(X_{ij}>0), \sum_{ij}\frac{\omega(X_{ij})-1}{h(X_{ij})}I(X_{ij}>0)\right\}\\
&= C_\rho + \frac{\Delta^*}{w}E_0\{h_0(X)\boldsymbol{u}(X)\}E_0[h_0(X)\{\omega(X)-1\}]\\
&\quad + \frac{\Delta^*}{(1-w)}E_0\{h_1(X)\boldsymbol{u}(X)\}E_0[h_1(X)\{\omega(X)-1\}]\\
&= C_\rho - \Delta^*\boldsymbol{m}E_0[h_1(X)\{\omega(X)-1\}],
\end{aligned}
$$

where $\boldsymbol{m} = \boldsymbol{\psi}^*/w - SE_0\{h_1(X)\boldsymbol{u}(X)\} = \boldsymbol{\psi}^*/w - \mathbf{E}_{1\boldsymbol{u}}/\{w(1-w)\}$.

For $\boldsymbol{D}_{\boldsymbol{\theta}}$,

$$
\begin{aligned}
\boldsymbol{D}_{\boldsymbol{\theta}} &= \mathrm{Cov}\{\boldsymbol{\psi}(\boldsymbol{\eta}^*), \boldsymbol{S}_{n,\boldsymbol{\theta}}^\top\}\\
&= \frac{1}{n\Delta^*}\mathrm{Cov}\left\{\sum_{ij}\frac{\boldsymbol{u}(X_{ij})}{h(X_{ij})}I(X_{ij}>0), \sum_{j=1}^{n_1}\boldsymbol{Q}(X_{1j})^\top I(X_{1j}>0) - \sum_{ij}h_1(X_{ij})\boldsymbol{Q}(X_{ij})^\top I(X_{ij}>0)\right\}\\
&= \frac{1}{n\Delta^*}\mathrm{Cov}\left\{\sum_{ij}\frac{\boldsymbol{u}(X_{ij})}{h(X_{ij})}I(X_{ij}>0), \sum_{j=1}^{n_1}h_0(X_{1j})\boldsymbol{Q}(X_{1j})^\top I(X_{1j}>0)\right\}\\
&\quad -\frac{1}{n\Delta^*}\mathrm{Cov}\left\{\sum_{ij}\frac{\boldsymbol{u}(X_{ij})}{h(X_{ij})}I(X_{ij}>0), \sum_{j=1}^{n_0}h_1(X_{0j})\boldsymbol{Q}(X_{0j})^\top I(X_{0j}>0)\right\}\\
&= (1-\rho^*)\Delta^*\boldsymbol{m}E_0\{h_1(X)\boldsymbol{Q}(X)^\top\}\\
&= \boldsymbol{m}(\boldsymbol{A}_{\boldsymbol{\theta}}\boldsymbol{e})^\top.
\end{aligned}
$$

With the form of $(\boldsymbol{D}_{\boldsymbol{\nu}}, \boldsymbol{D}_\rho, \boldsymbol{D}_{\boldsymbol{\theta}})$ and the form of $\boldsymbol{A}^{-1}$ in (S3.15), $\boldsymbol{\Gamma}_3$ is given as

$$
\begin{aligned}
\boldsymbol{\Gamma}_3 &= (\boldsymbol{D}_{\boldsymbol{\nu}}, \boldsymbol{D}_\rho, \boldsymbol{D}_{\boldsymbol{\theta}})\boldsymbol{A}^{-1}\boldsymbol{C}^\top\\
&= \boldsymbol{D}_{\boldsymbol{\nu}}\boldsymbol{A}_{\boldsymbol{\nu}}^{-1}\boldsymbol{C}_{\boldsymbol{\nu}}^\top - \frac{\rho^*(1-\rho^*)}{\Delta^*}\boldsymbol{D}_\rho\boldsymbol{C}_\rho^\top + \boldsymbol{D}_\rho\frac{\boldsymbol{e}^\top}{\Delta^*}\boldsymbol{C}_{\boldsymbol{\theta}}^\top + \boldsymbol{D}_{\boldsymbol{\theta}}\frac{\boldsymbol{e}}{\Delta^*}\boldsymbol{C}_\rho^\top + \boldsymbol{D}_{\boldsymbol{\theta}}\left\{\boldsymbol{A}_{\boldsymbol{\theta}}^{-1} - \frac{\boldsymbol{e}\boldsymbol{e}^\top}{\Delta^*\rho^*(1-\rho^*)}\right\}\boldsymbol{C}_{\boldsymbol{\theta}}^\top
\end{aligned}
$$

$$= \boldsymbol{D}_{\boldsymbol{\nu}}\boldsymbol{A}_{\boldsymbol{\nu}}^{-1}\boldsymbol{C}_{\boldsymbol{\nu}}^{\top} + \boldsymbol{D}_{\boldsymbol{\theta}}\boldsymbol{A}_{\boldsymbol{\theta}}^{-1}\boldsymbol{C}_{\boldsymbol{\theta}}^{\top} + \frac{1}{\Delta^*}\left\{\boldsymbol{D}_{\boldsymbol{\theta}}\boldsymbol{e} - \rho^*(1-\rho^*)\boldsymbol{D}_{\rho}\right\}\boldsymbol{C}_{\rho}^{\top} + \left\{\frac{\boldsymbol{D}_{\rho}}{\Delta^*} - \frac{\boldsymbol{D}_{\boldsymbol{\theta}}\boldsymbol{e}}{\Delta^*\rho^*(1-\rho^*)}\right\}\boldsymbol{e}^{\top}\boldsymbol{C}_{\boldsymbol{\theta}}^{\top}.$$

With the forms of $\boldsymbol{D}_{\rho}$ and $\boldsymbol{D}_{\boldsymbol{\theta}}$, we have

$$\boldsymbol{D}_{\boldsymbol{\theta}}\boldsymbol{A}_{\boldsymbol{\theta}}^{-1} = \boldsymbol{m}\boldsymbol{e}^{\top} \text{ and } \frac{1}{\Delta^*}\left\{\boldsymbol{D}_{\boldsymbol{\theta}}\boldsymbol{e} - \rho^*(1-\rho^*)\boldsymbol{D}_{\rho}\right\} = \rho^*(1-\rho^*)\boldsymbol{m} - \rho^*(1-\rho^*)\boldsymbol{C}_{\rho}/\Delta^*.$$

Hence,

$$\begin{aligned}\boldsymbol{\Gamma}_3 &= \boldsymbol{D}_{\boldsymbol{\nu}}\boldsymbol{A}_{\boldsymbol{\nu}}^{-1}\boldsymbol{C}_{\boldsymbol{\nu}}^{\top} + \boldsymbol{m}\boldsymbol{e}^{\top}\boldsymbol{C}_{\boldsymbol{\theta}}^{\top} + \left(\boldsymbol{m} - \frac{\boldsymbol{C}_{\rho}}{\Delta^*}\right)\left\{\rho^*(1-\rho^*)\boldsymbol{C}_{\rho}^{\top} - \boldsymbol{e}^{\top}\boldsymbol{C}_{\boldsymbol{\theta}}^{\top}\right\} \\ &= \boldsymbol{D}_{\boldsymbol{\nu}}\boldsymbol{A}_{\boldsymbol{\nu}}^{-1}\boldsymbol{C}_{\boldsymbol{\nu}}^{\top} + \left(\boldsymbol{m} - \frac{\boldsymbol{C}_{\rho}}{\Delta^*}\right)\rho^*(1-\rho^*)\boldsymbol{C}_{\rho}^{\top} + \frac{1}{\Delta^*}\boldsymbol{C}_{\rho}\boldsymbol{e}^{\top}\boldsymbol{C}_{\boldsymbol{\theta}}^{\top}. \end{aligned} \tag{S4.19}$$

Substituting $\boldsymbol{\Gamma}_2$ into (S4.18) and $\boldsymbol{\Gamma}_3$ into (S4.19) and using the facts that $\boldsymbol{C}_{\boldsymbol{\theta}} = \mathcal{M}_3$,

$$\boldsymbol{C}_{\boldsymbol{\nu}} + \rho^*(1-\rho^*)\boldsymbol{C}_{\rho}\boldsymbol{W} + \boldsymbol{D}_{\boldsymbol{\nu}} = \mathcal{M}_1, \tag{S4.20}$$

and

$$-\frac{(\boldsymbol{C}_{\boldsymbol{\theta}}\boldsymbol{e})(\boldsymbol{C}_{\boldsymbol{\theta}}\boldsymbol{e})^{\top}}{\Delta^*\rho^*(1-\rho^*)} + \frac{1}{\Delta^*}\boldsymbol{C}_{\rho}\boldsymbol{e}^{\top}\boldsymbol{C}_{\boldsymbol{\theta}}^{\top} + \frac{1}{\Delta^*}\boldsymbol{C}_{\boldsymbol{\theta}}\boldsymbol{e}\boldsymbol{C}_{\rho}^{\top} - \frac{\boldsymbol{C}_{\rho}}{\Delta^*}\rho^*(1-\rho^*)\boldsymbol{C}_{\rho}^{\top} = -\frac{\mathcal{M}_2\mathcal{M}_2^{\top}}{\Delta^*\rho^*(1-\rho^*)},$$

we have

$$\begin{aligned}\boldsymbol{\Gamma}_2 + \boldsymbol{\Gamma}_3 + \boldsymbol{\Gamma}_3^{\top} &= (\mathcal{M}_1 - \boldsymbol{D}_{\boldsymbol{\nu}})\boldsymbol{A}_{\boldsymbol{\nu}}^{-1}(\mathcal{M}_1 - \boldsymbol{D}_{\boldsymbol{\nu}})^{\top} - \frac{\mathcal{M}_2\mathcal{M}_2^{\top}}{\Delta^*\rho^*(1-\rho^*)} + \mathcal{M}_3\boldsymbol{A}_{\boldsymbol{\theta}}^{-1}\mathcal{M}_3^{\top} \\ &\quad + \boldsymbol{D}_{\boldsymbol{\nu}}\boldsymbol{A}_{\boldsymbol{\nu}}^{-1}\boldsymbol{C}_{\boldsymbol{\nu}}^{\top} + \boldsymbol{C}_{\boldsymbol{\nu}}\boldsymbol{A}_{\boldsymbol{\nu}}^{-1}\boldsymbol{D}_{\boldsymbol{\nu}}^{\top} + \rho^*(1-\rho^*)(\boldsymbol{m}\boldsymbol{C}_{\rho}^{\top} + \boldsymbol{C}_{\rho}\boldsymbol{m}^{\top}) - \frac{\boldsymbol{C}_{\rho}}{\Delta^*}\rho^*(1-\rho^*)\boldsymbol{C}_{\rho}^{\top} \\ &= \mathcal{M}_1\boldsymbol{A}_{\boldsymbol{\nu}}^{-1}\mathcal{M}_1^{\top} - \frac{\mathcal{M}_2\mathcal{M}_2^{\top}}{\Delta^*\rho^*(1-\rho^*)} + \mathcal{M}_3\boldsymbol{A}_{\boldsymbol{\theta}}^{-1}\mathcal{M}_3^{\top} \\ &\quad + \boldsymbol{D}_{\boldsymbol{\nu}}\boldsymbol{A}_{\boldsymbol{\nu}}^{-1}(\boldsymbol{C}_{\boldsymbol{\nu}} - \mathcal{M}_1)^{\top} + (\boldsymbol{C}_{\boldsymbol{\nu}} - \mathcal{M}_1)\boldsymbol{A}_{\boldsymbol{\nu}}^{-1}\boldsymbol{D}_{\boldsymbol{\nu}}^{\top} + \boldsymbol{D}_{\boldsymbol{\nu}}\boldsymbol{A}_{\boldsymbol{\nu}}^{-1}\boldsymbol{D}_{\boldsymbol{\nu}}^{\top} \\ &\quad + \rho^*(1-\rho^*)(\boldsymbol{m}\boldsymbol{C}_{\rho}^{\top} + \boldsymbol{C}_{\rho}\boldsymbol{m}^{\top}) - \frac{\boldsymbol{C}_{\rho}}{\Delta^*}\rho^*(1-\rho^*)\boldsymbol{C}_{\rho}^{\top}. \end{aligned} \tag{S4.21}$$

Next we further simplify the form of $\boldsymbol{\Gamma}_2 + \boldsymbol{\Gamma}_3 + \boldsymbol{\Gamma}_3^{\top}$. Note that with (S4.20), we have

$$\begin{aligned}&\boldsymbol{D}_{\boldsymbol{\nu}}\boldsymbol{A}_{\boldsymbol{\nu}}^{-1}(\boldsymbol{C}_{\boldsymbol{\nu}} - \mathcal{M}_1)^{\top} + \boldsymbol{D}_{\boldsymbol{\nu}}\boldsymbol{A}_{\boldsymbol{\nu}}^{-1}\boldsymbol{D}_{\boldsymbol{\nu}}^{\top} + \rho^*(1-\rho^*)\boldsymbol{m}\boldsymbol{C}_{\rho}^{\top} - \frac{\boldsymbol{C}_{\rho}}{\Delta^*}\rho^*(1-\rho^*)\boldsymbol{C}_{\rho}^{\top} \\ &= -\rho^*(1-\rho^*)\boldsymbol{D}_{\boldsymbol{\nu}}\boldsymbol{A}_{\boldsymbol{\nu}}^{-1}\boldsymbol{W}^{\top}\boldsymbol{C}_{\rho}^{\top} + \rho^*(1-\rho^*)\boldsymbol{m}\boldsymbol{C}_{\rho}^{\top} - \frac{\boldsymbol{C}_{\rho}}{\Delta^*}\rho^*(1-\rho^*)\boldsymbol{C}_{\rho}^{\top} \\ &= \rho^*(1-\rho^*)\left(-\boldsymbol{D}_{\boldsymbol{\nu}}\boldsymbol{A}_{\boldsymbol{\nu}}^{-1}\boldsymbol{W}^{\top} + \boldsymbol{m} - \frac{\boldsymbol{C}_{\rho}}{\Delta^*}\right)\boldsymbol{C}_{\rho}^{\top}. \end{aligned}$$

With the forms of $\boldsymbol{D}_{\boldsymbol{\nu}}$, $\boldsymbol{A}_{\boldsymbol{\nu}}^{-1}$, and $\boldsymbol{W}$, we have

$$\begin{aligned}\boldsymbol{D}_{\boldsymbol{\nu}}\boldsymbol{A}_{\boldsymbol{\nu}}^{-1}\boldsymbol{W}^{\top} &= -\frac{\nu_0}{\Delta^*(1-\rho^*)}\boldsymbol{\psi}^* + \frac{\rho^*\nu_0 + (1-\rho^*)\nu_1}{\Delta^*\rho^*(1-\rho^*)}\mathbf{E}_{1\boldsymbol{u}} \\ &= -\frac{\nu_0}{\Delta^*(1-\rho^*)}\boldsymbol{\psi}^* + \left\{\frac{1}{\Delta^*\rho^*(1-\rho^*)} - S\right\}\mathbf{E}_{1\boldsymbol{u}} \\ &= \frac{1-\nu_0}{\Delta^*(1-\rho^*)}\boldsymbol{\psi}^* - S\mathbf{E}_{1\boldsymbol{u}} - \frac{1}{\Delta^*\rho^*(1-\rho^*)}\{\rho^*\boldsymbol{\psi}^* - \mathbf{E}_{1\boldsymbol{u}}\} \\ &= \boldsymbol{m} - \frac{\boldsymbol{C}_{\rho}}{\Delta^*}. \end{aligned}$$

Hence,

$$\boldsymbol{D}_{\boldsymbol{\nu}}\boldsymbol{A}_{\boldsymbol{\nu}}^{-1}(\boldsymbol{C}_{\boldsymbol{\nu}} - \mathcal{M}_1)^{\top} + \boldsymbol{D}_{\boldsymbol{\nu}}\boldsymbol{A}_{\boldsymbol{\nu}}^{-1}\boldsymbol{D}_{\boldsymbol{\nu}}^{\top} + \rho^*(1-\rho^*)\boldsymbol{m}\boldsymbol{C}_{\rho}^{\top} - \frac{\boldsymbol{C}_{\rho}}{\Delta^*}\rho^*(1-\rho^*)\boldsymbol{C}_{\rho}^{\top} = \boldsymbol{0}$$

and $\boldsymbol{\Gamma}_2 + \boldsymbol{\Gamma}_3 + \boldsymbol{\Gamma}_3^\top$ in (S4.21) becomes

$$\begin{aligned}\boldsymbol{\Gamma}_2 + \boldsymbol{\Gamma}_3 + \boldsymbol{\Gamma}_3^\top = \mathcal{M}_1 \boldsymbol{A}_{\boldsymbol{\nu}}^{-1} \mathcal{M}_1^\top - \frac{\mathcal{M}_2 \mathcal{M}_2^\top}{\Delta^* \rho^* (1-\rho^*)} + \mathcal{M}_3 \boldsymbol{A}_{\boldsymbol{\theta}}^{-1} \mathcal{M}_3^\top \\ + (\boldsymbol{C}_{\boldsymbol{\nu}} - \mathcal{M}_1) \boldsymbol{A}_{\boldsymbol{\nu}}^{-1} \boldsymbol{D}_{\boldsymbol{\nu}}^\top + \rho^*(1-\rho^*) \boldsymbol{C}_\rho \boldsymbol{m}^\top .\end{aligned} \tag{S4.22}$$

With $\boldsymbol{\Gamma}_1$ in (S4.17) and $\boldsymbol{\Gamma}_2 + \boldsymbol{\Gamma}_3 + \boldsymbol{\Gamma}_3^\top$ in (S4.22), to show that $\boldsymbol{\Gamma} = \boldsymbol{\Gamma}_1 + \boldsymbol{\Gamma}_2 + \boldsymbol{\Gamma}_3 + \boldsymbol{\Gamma}_3^\top$, we need to argue that

$$-\frac{\boldsymbol{\psi}^* \boldsymbol{\psi}^{*\top}}{\Delta^*} = (\boldsymbol{C}_{\boldsymbol{\nu}} - \mathcal{M}_1) \boldsymbol{A}_{\boldsymbol{\nu}}^{-1} \boldsymbol{D}_{\boldsymbol{\nu}}^\top + \rho^*(1-\rho^*) \boldsymbol{C}_\rho \boldsymbol{m}^\top - \frac{1}{w} \mathbf{E}_{0\boldsymbol{u}} \mathbf{E}_{0\boldsymbol{u}}^\top - \frac{1}{1-w} \mathbf{E}_{1\boldsymbol{u}} \mathbf{E}_{1\boldsymbol{u}}^\top . \tag{S4.23}$$

Note that

$$\boldsymbol{C}_{\boldsymbol{\nu}} = \mathcal{M}_1 + (w, 1-w) \otimes \frac{\boldsymbol{\psi}^*}{\Delta^*}.$$

Then

$$(\boldsymbol{C}_{\boldsymbol{\nu}} - \mathcal{M}_1) \boldsymbol{A}_{\boldsymbol{\nu}}^{-1} \boldsymbol{D}_{\boldsymbol{\nu}}^\top = -\frac{\nu_0}{\Delta^*} \boldsymbol{\psi} \mathbf{E}_{0\boldsymbol{u}}^\top - \frac{\nu_1}{\Delta^*} \boldsymbol{\psi} \mathbf{E}_{1\boldsymbol{u}}^\top = -\boldsymbol{\psi} \left(\frac{\nu_0}{\Delta^*} \mathbf{E}_{0\boldsymbol{u}}^\top + \frac{\nu_1}{\Delta^*} \mathbf{E}_{1\boldsymbol{u}} \right)^\top . \tag{S4.24}$$

Recall that

$$\rho^*(1-\rho^*) \boldsymbol{C}_\rho = \rho^* \boldsymbol{\psi}^* - \mathbf{E}_{1\boldsymbol{u}} = \rho^* \mathbf{E}_{0\boldsymbol{u}} - (1-\rho^*) \mathbf{E}_{1\boldsymbol{u}}$$

and

$$\boldsymbol{m} = \boldsymbol{\psi}^*/w - \mathbf{E}_{1\boldsymbol{u}}/\{w(1-w)\} = \mathbf{E}_{0\boldsymbol{u}}/w - \mathbf{E}_{1\boldsymbol{u}}/(1-w).$$

Then

$$\begin{aligned}&\rho^*(1-\rho^*) \boldsymbol{C}_\rho \boldsymbol{m}^\top - \frac{1}{w} \mathbf{E}_{0\boldsymbol{u}} \mathbf{E}_{0\boldsymbol{u}}^\top - \frac{1}{1-w} \mathbf{E}_{1\boldsymbol{u}} \mathbf{E}_{1\boldsymbol{u}}^\top \\ &= \{\rho^* \mathbf{E}_{0\boldsymbol{u}} - (1-\rho^*) \mathbf{E}_{1\boldsymbol{u}}\} \{\mathbf{E}_{0\boldsymbol{u}}/w - \mathbf{E}_{1\boldsymbol{u}}/(1-w)\}^\top - \frac{1}{w} \mathbf{E}_{0\boldsymbol{u}} \mathbf{E}_{0\boldsymbol{u}}^\top - \frac{1}{1-w} \mathbf{E}_{1\boldsymbol{u}} \mathbf{E}_{1\boldsymbol{u}}^\top \\ &= -\frac{1-\rho^*}{w} \mathbf{E}_{0\boldsymbol{u}} \mathbf{E}_{0\boldsymbol{u}}^\top - \frac{1-\rho^*}{w} \mathbf{E}_{1\boldsymbol{u}} \mathbf{E}_{0\boldsymbol{u}}^\top - \frac{\rho^*}{1-w} \mathbf{E}_{0\boldsymbol{u}} \mathbf{E}_{1\boldsymbol{u}}^\top - \frac{\rho^*}{1-w} \mathbf{E}_{1\boldsymbol{u}} \mathbf{E}_{1\boldsymbol{u}}^\top \\ &= -(\mathbf{E}_{0\boldsymbol{u}} + \mathbf{E}_{1\boldsymbol{u}}) \left(\frac{1-\rho^*}{w} \mathbf{E}_{0\boldsymbol{u}} + \frac{\rho^*}{1-w} \mathbf{E}_{1\boldsymbol{u}} \right)^\top \\ &= -\boldsymbol{\psi}^* \left(\frac{1-\nu_0}{\Delta^*} \mathbf{E}_{0\boldsymbol{u}} + \frac{1-\nu_1}{\Delta^*} \mathbf{E}_{1\boldsymbol{u}} \right)^\top .\end{aligned} \tag{S4.25}$$

Since $\mathbf{E}_{0\boldsymbol{u}} + \mathbf{E}_{1\boldsymbol{u}} = \boldsymbol{\psi}$, combining (S4.24) and (S4.25) gives

$$(\boldsymbol{C}_{\boldsymbol{\nu}} - \mathcal{M}_1) \boldsymbol{A}_{\boldsymbol{\nu}}^{-1} \boldsymbol{D}_{\boldsymbol{\nu}}^\top + \rho^*(1-\rho^*) \boldsymbol{C}_\rho \boldsymbol{m}^\top - \frac{1}{w} \mathbf{E}_{0\boldsymbol{u}} \mathbf{E}_{0\boldsymbol{u}}^\top - \frac{1}{1-w} \mathbf{E}_{1\boldsymbol{u}} \mathbf{E}_{1\boldsymbol{u}}^\top = -\frac{1}{\Delta} \boldsymbol{\psi} \boldsymbol{\psi}^{*\top},$$

which verifies (S4.23). Hence,

$$\boldsymbol{\Gamma} = \boldsymbol{\Gamma}_1 + \boldsymbol{\Gamma}_2 + \boldsymbol{\Gamma}_3 + \boldsymbol{\Gamma}_3^\top .$$

Applying Slutsky's theorem and the central limit theorem to (S4.16), we get as $n \to \infty$

$$n^{1/2} \left(\hat{\boldsymbol{\psi}} - \boldsymbol{\psi}^* \right) \xrightarrow{d} N(\mathbf{0}, \boldsymbol{\Gamma}).$$

This completes the proof of Theorem 2. □

S5 Proof of Corollary 1

Note that for $\boldsymbol{\psi} = \left(\boldsymbol{\psi}_0^\top, \boldsymbol{\psi}_1^\top\right)^\top$, $\boldsymbol{u}(x;\boldsymbol{\nu},\boldsymbol{\theta})$ can be written as

$$\boldsymbol{u}(X;\boldsymbol{\nu},\boldsymbol{\theta}) = \begin{pmatrix} (1-\nu_0)\mathbf{a}(x) \\ (1-\nu_1)\mathbf{a}(x)\omega(x;\boldsymbol{\theta}) \end{pmatrix}. \tag{S5.26}$$

We substitute this $\boldsymbol{u}(x;\boldsymbol{\nu},\boldsymbol{\theta})$ into $\boldsymbol{\Gamma}$ in Theorem 2 and obtain $\boldsymbol{\Gamma}_{sem}$.

Note that

$$E_0\left\{\frac{\boldsymbol{u}(X;\boldsymbol{\nu}^*,\boldsymbol{\theta}^*)}{\partial\boldsymbol{\nu}}\right\} = \begin{pmatrix} -\frac{\boldsymbol{\psi}_0}{1-\nu_0^*} & \mathbf{0} \\ \mathbf{0} & -\frac{\boldsymbol{\psi}_1}{1-\nu_1^*} \end{pmatrix}$$

and

$$E_0\left\{\frac{\partial\boldsymbol{u}(X;\boldsymbol{\nu}^*,\boldsymbol{\theta}^*)}{\partial\boldsymbol{\theta}}\right\} = \begin{pmatrix} \mathbf{0} \\ (1-\nu_1^*)E_0\{\mathbf{a}(X)\omega(X)\boldsymbol{Q}(X)^\top\} \end{pmatrix}.$$

Then we have $\mathcal{M}_1 = \operatorname{diag}\{-\boldsymbol{\psi}_0/(1-\nu_0^*), -\boldsymbol{\psi}_1/(1-\nu_1^*)\}$ and

$$\mathcal{M}_2 = \begin{pmatrix} -\rho^*\boldsymbol{\psi}_0 \\ (1-\rho^*)\boldsymbol{\psi}_1 \end{pmatrix}, \quad \mathcal{M}_3 = \begin{pmatrix} -w^{-1}\Delta^*(1-\rho^*)E_0\{h_1(X)\mathbf{a}(X)\boldsymbol{Q}(X)^\top\} \\ (1-w)^{-1}\Delta^*(1-\rho^*)E_0\{h_1(X)\mathbf{a}(X)\boldsymbol{Q}(X)^\top\} \end{pmatrix}.$$

Substituting $\mathcal{M}_1$, $\mathcal{M}_2$, and $\mathcal{M}_3$ into $\boldsymbol{\Gamma}$ and simplifying, we find that $\boldsymbol{\Gamma}_{sem}$ is

$$\boldsymbol{\Gamma}_{sem} = \frac{1}{\Delta^*}E_0\left\{\frac{\boldsymbol{u}(X)\boldsymbol{u}(X)^\top}{h(X)}\right\} - \begin{pmatrix} \frac{\boldsymbol{\psi}_0\boldsymbol{\psi}_0^\top}{w} & \mathbf{0} \\ \mathbf{0} & \frac{\boldsymbol{\psi}_1\boldsymbol{\psi}_1^\top}{1-w} \end{pmatrix} + \begin{pmatrix} \frac{1}{w^2} & -\frac{1}{w(1-w)} \\ -\frac{1}{w(1-w)} & \frac{1}{(1-w)^2} \end{pmatrix} \otimes \boldsymbol{D}_1,$$

where

$$\boldsymbol{D}_1 = \{\Delta^*(1-\rho^*)\}^2 E_0\{h_1(X)\mathbf{a}(X)\boldsymbol{Q}(X)^\top\}\boldsymbol{A}_{\boldsymbol{\theta}}^{-1}E_0\{h_1(X)\boldsymbol{Q}(X)\mathbf{a}(X)^\top\}.$$

Substituting (S5.26) into the first term of $\boldsymbol{\Gamma}_{sem}$, we find that

$$\begin{aligned}\frac{1}{\Delta^*}E_0\left\{\frac{\boldsymbol{u}(X)\boldsymbol{u}(X)^\top}{h(X)}\right\} &= \begin{pmatrix} w^{-1}(\mathbf{V}_0 + \boldsymbol{\psi}_0\boldsymbol{\psi}_0^\top) & \mathbf{0} \\ \mathbf{0} & (1-w)^{-1}(\mathbf{V}_1 + \boldsymbol{\psi}_1\boldsymbol{\psi}_1^\top) \end{pmatrix} \\ &\quad - \begin{pmatrix} \frac{1}{w^2} & -\frac{1}{w(1-w)} \\ -\frac{1}{w(1-w)} & \frac{1}{(1-w)^2} \end{pmatrix} \otimes \boldsymbol{D}_0,\end{aligned}$$

where

$$\boldsymbol{D}_0 = \Delta^*(1-\rho^*)E_0\{h_1(X)\mathbf{a}(X)\mathbf{a}(X)^\top\}.$$

Hence,

$$\boldsymbol{\Gamma}_{sem} = \boldsymbol{\Gamma}_{non} - \begin{pmatrix} \frac{1}{w^2} & -\frac{1}{w(1-w)} \\ -\frac{1}{w(1-w)} & \frac{1}{(1-w)^2} \end{pmatrix} \otimes (\boldsymbol{D}_0 - \boldsymbol{D}_1).$$

Recall that

$$\mathbf{d}(X) = \mathbf{a}(X) - \Delta^*(1-\rho^*)E_0\left\{h_1(X)\mathbf{a}(X)\boldsymbol{Q}(X)^\top\right\}\boldsymbol{A}_{\boldsymbol{\theta}}^{-1}\boldsymbol{Q}(X)$$

and

$$\boldsymbol{A}_{\boldsymbol{\theta}} = \Delta^*(1-\rho^*)E_0\left[h_1(X)\boldsymbol{Q}(X)\boldsymbol{Q}^\top(X)\right].$$

It can be verified that

$$E_0\left\{h_1(X)\mathbf{d}(X)\mathbf{d}(X)^\top\right\} = \frac{1}{\Delta^*(1-\rho^*)}(\boldsymbol{D}_0 - \boldsymbol{D}_1).$$

Therefore,

$$\boldsymbol{\Gamma}_{sem} = \boldsymbol{\Gamma}_{non} - \Delta^*(1-\rho^*)E_0\left\{h_1(X)\begin{pmatrix} w^{-1}\mathbf{d}(X) \\ -(1-w)^{-1}\mathbf{d}(X) \end{pmatrix}\begin{pmatrix} w^{-1}\mathbf{d}(X) \\ -(1-w)^{-1}\mathbf{d}(X) \end{pmatrix}^\top\right\},$$

as claimed in Corollary 1. This completes the proof. □

S6 Additional simulation study

In our setup, the basis function $\boldsymbol{q}(x)$ in the DRM (S1.2) needs to be prespecified. In this section, we provide an additional small simulation to study the impact of misspecification of $\boldsymbol{q}(x)$ on the performance of our proposed estimators and confidence intervals (CIs). Table S1 gives the parameter settings for the simulation study.

Table S1 Parameter settings for simulation studies: $G_0 = \mathcal{LN}(a_0, b_0)$ and $G_1 = \mathcal{LN}(a_1, b_1)$.

(ν_0,ν_1)	(a_0,a_1)	(b_0,b_1)	correctly-specified $\boldsymbol{q}(x)$	δ	σ_0^2	σ_1^2
(0.3,0.5)	(0.33,0.66)	(1,1)	$\log x$	0.99	7.43	11.29
		(1,1.25)	$(x, \log x)$	1.13	7.43	19.54
(0.3,0.3)	(0,0.5)	(1,1)	$\log x$	1.65	3.84	10.44
		(1,1.25)	$(x, \log x)$	1.87	3.84	18.53

For all the models listed in Table S1, we use $\boldsymbol{q}(x) = \log x$ to fit the DRM. Hence, this $\boldsymbol{q}(x)$ is misspecified for the settings with $(b_0, b_1) = (1, 1.25)$. For each model, we still consider four combinations of the sample sizes (n_0, n_1): $(50, 50)$, $(100, 100)$, $(50, 150)$, and $(150, 50)$. The number of replications is 10,000 for each configuration of the parameter settings. For the methods calibrated by the bootstrap method, 999 bootstrap samples are drawn from the original sample.

We consider all estimators, as discussed in Section 3.2 of the main paper, of the mean ratio δ and the population variances σ_0^2 and σ_1^2. Tables S2 and S3 present the biases and mean square errors (MSEs) of these estimators. When the DRM is fitted by the misspecified basis function $\boldsymbol{q}(x)$, the biases of the proposed estimators slightly increase but are still acceptable. The MSEs of the proposed estimators remain comparable to the fully nonparametric estimators. Moreover, the estimators $\hat{\delta}$ and $\hat{\sigma}_1^2$ always give smaller MSEs than those of the fully nonparametric estimators $\tilde{\delta}$ and $\tilde{\sigma}_1^2$, respectively.

Table S2 Bias and mean square error of point estimates for δ, σ_0^2, and σ_1^2 with $(\nu_0, \nu_1) = (0.3, 0.5)$ and $(a_0, a_1) = (0.33, 0.66)$.

		$\tilde{\delta}$		$\hat{\delta}$		$\tilde{\sigma}_0^2$		$\hat{\sigma}_0^2$		$\tilde{\sigma}_1^2$		$\hat{\sigma}_1^2$	
(b_0,b_1)	(n_0,n_1)	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE
(1, 1)	(50, 50)	0.06	0.17	0.03	0.11	-0.05	104.16	-0.03	54.97	0.19	426.17	-0.15	292.11
	(50, 150)	0.06	0.10	0.03	0.07	-0.05	118.37	0.19	36.01	0.11	120.63	-0.08	100.58
	(150, 50)	0.01	0.11	0.02	0.08	0.02	42.75	-0.12	23.86	-0.26	193.29	-0.16	135.72
	(100, 100)	0.03	0.08	0.02	0.05	-0.06	57.48	-0.10	21.76	-0.09	147.69	-0.21	108.06
(1, 1.25)	(50, 50)	0.06	0.26	-0.07	0.15	-0.02	101.56	2.51	157.28	-0.29	1,834.17	-3.25	1,183.53
	(50, 150)	0.07	0.15	-0.08	0.09	-0.02	101.56	4.58	208.71	0.34	1,168.17	-1.35	965.79
	(150, 50)	0.02	0.17	-0.10	0.11	-0.05	41.34	1.21	45.03	-0.87	1,112.45	-5.04	477.88
	(100, 100)	0.04	0.12	-0.09	0.08	-0.17	43.43	2.83	148.65	0.46	2,271.41	-2.74	1,588.47

Table S3 Bias and mean square error of point estimates for δ, σ_0^2, and σ_1^2 with $(\nu_0, \nu_1) = (0.3, 0.3)$ and $(a_0, a_1) = (0, 0.5)$.

		$\tilde{\delta}$		$\hat{\delta}$		$\tilde{\sigma}_0^2$		$\hat{\sigma}_0^2$		$\tilde{\sigma}_1^2$		$\hat{\sigma}_1^2$	
(b_0,b_1)	(n_0,n_1)	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE	Bias	MSE
(1, 1)	(50, 50)	0.10	0.37	0.06	0.27	0.02	33.28	0.04	9.29	-0.14	190.88	-0.41	160.50
	(50, 150)	0.09	0.24	0.05	0.17	-0.01	27.55	0.15	6.82	0.11	106.76	-0.03	97.86
	(150, 50)	0.03	0.22	0.02	0.15	-0.05	8.84	-0.05	4.72	-0.12	173.09	-0.35	109.84
	(100, 100)	0.04	0.17	0.03	0.12	0.02	18.88	0.00	3.54	-0.12	211.21	-0.22	196.79
(1, 1.25)	(50, 50)	0.11	0.55	-0.12	0.34	-0.01	27.13	1.73	32.99	-0.38	1,158.33	-2.36	920.49
	(50, 150)	0.11	0.35	-0.13	0.22	-0.01	27.13	2.59	35.37	0.29	1,117.63	-0.64	1,046.12
	(150, 50)	0.04	0.35	-0.15	0.23	-0.03	11.04	0.99	11.46	-0.69	779.72	-3.85	489.75
	(100, 100)	0.06	0.26	-0.15	0.18	-0.09	11.60	1.83	31.71	0.46	2,209.98	-1.51	1,915.74

We also examine the behavior of the 95% CIs of the mean ratio δ, as discussed in Section 3.3 of the main paper. The coverage probabilities (CPs) and average lengths (ALs) of CIs are included in Tables S4 and S5, respectively. The CIs $\mathcal{I}_{1B}$ and $\mathcal{I}_{2B}$ have similar and best performance among all the considered CIs in terms of the CP. The CIs $\mathcal{I}_4$ always have the lowest CPs and the shortest ALs. The CIs $\mathcal{I}_3$ and $\mathcal{I}_{4L}$ have similar CPs as the CIs $\mathcal{I}_1$ and $\mathcal{I}_2$ but slightly shorter ALs.

Table S4 Coverage probability (%) and average length of 95% CIs for $\delta$with $(\nu_0, \nu_1) = (0.3, 0.5)$ and $(a_0, a_1) = (0.33, 0.66)$.

		$\mathcal{I}_1$		$\mathcal{I}_{1B}$		$\mathcal{I}_2$		$\mathcal{I}_{2B}$		$\mathcal{I}_3$		$\mathcal{I}_4$		$\mathcal{I}_{4L}$	
(b_0, b_1)	(n_0, n_1)	CP	AL	CP	AL	CP	AL	CP	AL	CP	AL	CP	AL	CP	AL
(1,1)	(50, 50)	92.3	1.55	94.5	1.92	91.5	1.52	94.1	1.84	94.3	1.36	92.8	1.25	94.8	1.34
	(50, 150)	92.6	1.17	94.3	1.36	93.0	1.15	94.7	1.30	94.8	1.04	93.5	0.97	95.0	1.01
	(150, 50)	92.5	1.27	94.2	1.66	91.3	1.25	93.7	1.56	94.8	1.11	93.3	1.06	95.2	1.11
	(100, 100)	94.2	1.06	95.4	1.19	92.7	1.06	93.9	1.18	94.8	0.92	94.2	0.88	95.2	0.91
(1, 1.25)	(50, 50)	91.82	1.91	94.12	2.94	90.92	1.86	93.59	2.50	91.44	1.49	85.40	1.36	91.67	1.47
	(50, 150)	92.94	1.41	94.62	1.66	92.11	1.40	94.20	1.62	91.76	1.14	85.52	1.06	91.17	1.11
	(150, 50)	91.15	1.60	93.41	3.69	90.11	1.56	92.88	2.24	90.03	1.19	84.73	1.12	90.87	1.18
	(100, 100)	93.33	1.31	94.81	2.11	92.71	1.32	94.56	1.61	90.59	1.00	84.77	0.95	90.52	0.98

Table S5 Coverage probability (%) and average length of 95% CIs for δ with $(\nu_0, \nu_1) = (0.3, 0.3)$ and $(a_0, a_1) = (0, 0.5)$.

		$\mathcal{I}_1$		$\mathcal{I}_{1B}$		$\mathcal{I}_2$		$\mathcal{I}_{2B}$		$\mathcal{I}_3$		$\mathcal{I}_4$		$\mathcal{I}_{4L}$	
(b_0, b_1)	(n_0, n_1)	CP	AL	CP	AL	CP	AL	CP	AL	CP	AL	CP	AL	CP	AL
(1,1)	(50, 50)	92.6	2.23	94.5	2.66	91.4	2.24	93.8	2.65	94.0	1.98	92.6	1.85	94.0	1.95
	(50, 150)	92.6	1.79	94.2	2.07	91.5	1.77	93.5	2.00	94.8	1.61	93.8	1.53	94.5	1.58
	(150, 50)	92.7	1.77	94.5	2.68	92.2	1.78	94.2	2.14	94.4	1.54	92.9	1.47	94.6	1.52
	(100, 100)	93.5	1.56	95.0	1.75	92.8	1.56	94.4	1.72	94.3	1.37	93.5	1.30	94.5	1.34
(1, 1.25)	(50, 50)	92.14	2.76	94.25	3.74	91.33	2.72	93.85	3.44	90.83	2.23	84.83	2.05	90.62	2.18
	(50, 150)	92.66	2.14	94.41	2.51	91.67	2.13	93.80	2.45	92.03	1.80	86.37	1.69	91.32	1.76
	(150, 50)	92.07	2.24	93.82	4.84	91.29	2.23	93.50	2.98	88.56	1.71	82.98	1.61	88.67	1.68
	(100, 100)	93.56	1.92	95.00	2.44	92.95	1.94	94.52	2.27	89.73	1.53	84.21	1.44	89.22	1.49

In conclusion, the misspecification of $\boldsymbol{q}(x)$ may increase the biases of the proposed estimators under the DRM, however, the MSEs of the proposed estimators still remain comparable with the MSEs of the nonparametric counterparts. For the CIs, the problem of low CPs seems to arise under the DRM with misspecified basis function, and using the log transformation and the empirical-likelihood-ratio based method may relatively alleviate the problem. On the other hand, it seems that the ALs of the proposed CIs under the DRM are always shorter than the ALs of the fully nonparametric alternatives.